



Programa de Pós-Graduação em

# **Computação Aplicada**

---

Doutorado Acadêmico

Tiago Roberto Kautzmann

Os efeitos de usar estimativas de conhecimento do aluno em  
programação de computadores em modelos livres de sensores de  
detecção da emoção confusão

São Leopoldo, 2022



Tiago Roberto Kautzmann

**OS EFEITOS DE USAR ESTIMATIVAS DE CONHECIMENTO DO  
ALUNO EM PROGRAMAÇÃO DE COMPUTADORES EM MODELOS  
LIVRES DE SENSORES DE DETECÇÃO DA EMOÇÃO CONFUSÃO**

Tese apresentada como requisito parcial para  
obtenção do título de Doutor em Computação  
Aplicada, pelo Programa de Pós-Graduação em  
Computação Aplicada da Universidade do Vale  
do Rio dos Sinos - UNISINOS

Orientadora: Prof.<sup>a</sup> Dra. Patrícia Augustin Jaques Maillard

Coorientador: Prof. Dr. Gabriel de Oliveira Ramos

São Leopoldo

2022

K21e Kautzmann, Tiago Roberto.  
Os efeitos de usar estimativas de conhecimento do aluno em programação de computadores em modelos livres de sensores de detecção da emoção confusão / Tiago Roberto Kautzmann. – 2022.  
257 f. : il. ; 30 cm.

Tese (doutorado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, 2022.  
“Orientadora: Profa. Dra. Patrícia Augustin Jaques Maillard  
Coorientador: Prof. Dr. Gabriel de Oliveira Ramos.”

1. Ambiente computacional de aprendizagem. 2. Aprendizado de máquina. 3. Computação afetiva. 4. Detecção de confusão. 5. Programação de computadores. I. Título.

CDU 004

## **AGRADECIMENTOS À CAPES**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 / This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Dedico esta Tese ao meu pai, Ernesto (*in memoriam*), e a minha mãe, Silanda, por tudo que fizeram por mim e por darem toda a base de carinho e de amor que me é necessário para enfrentar os desafios da vida.

## AGRADECIMENTOS

Primeiramente, agradeço à minha orientadora, Patrícia Jaques, por tudo o que me ensinou sobre como fazer Ciência, nestes anos todos, desde a época do Mestrado. Agradeço por todo o apoio, paciência, dedicação, e competência que sempre empregou no seu trabalho. Agradeço pelas longas conversas sobre metodologia de pesquisa e sobre os desafios e direcionamentos das pesquisas que realizamos juntos. Não via o tempo passar. Agradeço também por todo o suporte emocional e pela gentileza de sempre. Serei sempre muito grato!

Ao colega doutorando Felipe de Moraes, que participou do trabalho da Tese com diversas colaborações, como no desenvolvimento, em parceria, do protocolo de anotações de emoções, e na aplicação de um dos testes piloto do ambiente de programação adaptado para a Tese, enquanto eu estava na França, no período do Doutorado Sanduíche. Agradeço também às inúmeras conversas que tivemos sobre questões técnicas e metodológicas.

Ao coorientador, Gabriel Ramos, por todo o suporte técnico e emocional nestes últimos dois anos. Agradeço também por possibilitar a coleta de dados da pesquisa nas suas turmas de programação. Obrigado pela competência e dedicação!

À Magalie Ochs, professora do *Laboratoire d'Informatique et Systèmes*, da *Aix Marseille Université*, que me orientou no período do Doutorado Sanduíche, entre os meses de agosto e dezembro de 2019, na cidade de Marselha, no sul da França.

À minha esposa, Vanessa, por todo o suporte, carinho e amor, tão importantes e necessários nessa longa jornada. Obrigado por toda a paciência!

À professora da Unisinos, Cassia Pereira Nino, por possibilitar a coleta de dados em sua turma de programação e por algumas sugestões ao trabalho.

Às instituições de ensino Unisinos, Colégio Luterano Concórdia de São Leopoldo (RS) e IENH - Instituição Evangélica de Novo Hamburgo (RS), por aceitarem participar da pesquisa e cederem seus espaços para a coleta de dados e aplicação de testes piloto.

Às funcionárias do PPG em Computação Aplicada, Luciana Aquino, Victória Zwetsch, Alice Borges e Bruna Severo, pelo suporte nas dúvidas sobre questões administrativas e pela competência e agilidade nos encaminhamentos.

Aos professores do PPG em Computação Aplicada e demais colegas doutorandos e mestrandos pelas discussões e ideias discutidas nas aulas, nos intervalos das aulas, nos corredores da instituição e na sala do café.

Aos familiares e amigos, por todo o apoio de sempre, e pela paciência pelos momentos em que não pude estar presente para me dedicar a este trabalho.

## RESUMO

A área de pesquisa em Computação Afetiva tem buscado formas de melhorar a detecção da confusão do aluno em ambientes computacionais de aprendizagem. Ambientes capazes de detectar a confusão do estudante podem utilizar diferentes estratégias pedagógicas, como intervir no ambiente e auxiliar o aluno a resolver sua confusão ou controlar a confusão do aluno para beneficiar sua aprendizagem. O autor tem interesse em contribuir com o estado da arte na detecção da confusão sem o uso de sensores (livre de sensores) no contexto de aprendizagem de programação de computadores. A Tese levantou a hipótese de que utilizar dados sobre estimativas de conhecimento do aluno, juntamente de dados de interação do aluno com o ambiente computacional, pode melhorar o desempenho de modelos de aprendizado de máquina livres de sensores na detecção da confusão do estudante em tarefas de aprendizagem de programação, quando comparado a modelos *baseline*. Os modelos *baseline* representam os trabalhos relacionados, que desenvolveram seus modelos utilizando dados de interação do aluno com o ambiente. A hipótese da Tese é justificada em teorias cognitivas de emoções, que relacionam a confusão com avaliações (*appraisals*) de incompatibilidade entre a informação que chega para o aluno e o modelo mental do estudante, como o modelo mental de conhecimento anterior. Para verificar sua hipótese, a trabalho gerou diversos modelos de aprendizado de máquina que representam a abordagem da Tese (hipótese da Tese) e a abordagem *baseline* (trabalhos relacionados), para diferentes configurações de janelas de tempo de observação (5, 10, 20, 40, 60, 90, 120, 180, 240 e 360 segundos e variável), e para diferentes algoritmos. Testes estatísticos compararam os resultados dos modelos de cada abordagem (Tese e *baseline*). Também foram aplicados métodos para verificar os dados mais relevantes para os modelos e verificar o poder de generalização para estudantes com características heterogêneas. Os modelos de aprendizado de máquina foram treinados e testados com amostras formadas por dados coletados de 62 alunos de ensino técnico e superior, durante cinco meses, enquanto solucionavam exercícios em um software de programação adaptado para a Tese. Testes estatísticos mostraram que os melhores modelos da Tese apresentaram acurácia preditiva superior e significativa em relação aos melhores modelos *baseline*, em todas as janelas de observação. Em uma relação dos dez atributos de dados mais relevantes para os melhores modelos da Tese, cinco foram atributos sobre interação com o ambiente e os outros cinco foram atributos sobre estimativas de conhecimento do aluno. Em relação ao poder de generalização para estudantes com características heterogêneas, foram encontradas diferenças significativas e consistentes entre as abordagens em janelas de observação de 5, 10 e 20 segundos. Nestas

janelas, o desempenho dos melhores modelos da Tese foi superior ao desempenho dos melhores modelos *baseline*. Os resultados apresentaram evidências positivas que suportam a hipótese da Tese, de que estimativas de conhecimento do aluno podem melhorar o desempenho de modelos de detecção de confusão livres de sensores nas tarefas de programação. A Tese apresenta discussões para outros resultados intermediários e sobre cenários onde a abordagem apresentada é mais vantajosa.

**Palavras-chave:** detecção de confusão; ambiente computacional de aprendizagem; programação de computadores; aprendizado de máquina; computação afetiva.

## ABSTRACT

The research area of Affective Computing has been looking for ways to improve the detection of student confusion in computer-based learning environments. Environments capable of detecting student confusion can use different pedagogical strategies, such as intervening in the environment and helping students resolve their confusion or controlling it to benefit their learning. The author is interested in contributing to state the art in detecting confusion without using physical sensors (sensor-free) in the context of computer programming learning. The Thesis hypothesized that using data on student knowledge estimates and data on student interaction with the computer-based learning environment can improve the performance of sensor-free machine learning models in detecting student confusion in tasks of programming learning compared to baseline models. Baseline models represent related works, which developed their models using only student environment interaction data. The Thesis hypothesis is justified in cognitive theories of emotions, which relate the confusion with appraisals of incompatibility between the information that comes to the student and the student's mental model, such as the mental model of prior knowledge. To verify its hypothesis, the Thesis generated several machine learning models that represent the Thesis approach (Thesis hypothesis) and the baseline approach (related works) for different configurations of observation time windows (5, 10, 20, 40, 60, 90, 120, 180, 240 and 360 seconds and variable) and different algorithms. Statistical tests compared the results of each approach (Thesis and baseline). Methods were also applied to verify the models' most relevant data and the generalization performance for students with heterogeneous characteristics. The machine learning models were trained and tested with samples formed by data collected from 62 technical and higher education students for five months while solving exercises in a programming software adapted for the Thesis. Statistical tests showed that the best models of the Thesis approach presented superior and significant predictive accuracy compared to the best baseline models in all observation windows. In a list of the ten most relevant data attributes for the best models of the Thesis approach, five were attributes about interaction with the environment, and the other five were attributes about estimates of student knowledge. Regarding the performance of generalization for students with heterogeneous characteristics, significant differences were found between the approaches only in observation windows of 5, 10, and 20 seconds. In these windows, the performance of the thesis approach's best models was superior to that of the best baseline models. The results presented positive evidence that supports the hypothesis raised that estimates of student knowledge can improve the

performance of sensor-free confusion detection models in computer programming tasks. The Thesis presents discussions for several other intermediate results and the scenarios where the Thesis approach is most advantageous.

**Keywords:** confusion detection; computer-based learning environment; computer programming; machine learning; affective computing.

## LISTA DE FIGURAS

Figura 1 – Modelo de dinâmicas afetivas de emoções relacionadas à aprendizagem.....	43
Figura 2 - Expressões faciais de confusão.....	47
Figura 3 – Fragmento da interface gráfica do Módulo do Aluno do <i>Feeper</i> .....	82
Figura 4 – Fragmento da interface gráfica do Módulo do Professor do <i>Feeper</i> .....	83
Figura 5 – Fragmento de tela do <i>Feeper</i> que mostra a edição de um enunciado de exercício .	83
Figura 6 – Fragmento de tela do <i>Feeper</i> que mostra a edição de um teste unitário .....	84
Figura 7 – Fluxograma de funcionamento do Juiz Online do <i>Feeper</i> .....	85
Figura 8 – Tela principal do <i>Feeper</i> Afetivo com menu de opções .....	87
Figura 9 – Tela de objetivos do exercício.....	88
Figura 10 – Envio da solução de objetivo para avaliação .....	90
Figura 11 – Tela do editor de código do <i>BlueJ</i> Afetivo .....	90
Figura 12 – Estrutura básica do modelo BKT .....	103
Figura 13 - Processo de submissão da solução do objetivo de exercício .....	113
Figura 14 – Fragmento de registros de <i>log</i> coletados e mantidos em banco de dados.....	123
Figura 15 – Botão na interface do <i>BlueJ</i> Afetivo para o autorrelato de confusão.....	125
Figura 16 – Diferentes configurações das amostras geradas.....	129
Figura 17 – Representação de intervalos de <i>logs</i> atrelados e não atrelados a objetivos.....	130
Figura 18 – Exemplo de sequência de <i>logs</i> atrelada a um objetivo de exercício .....	131
Figura 19 - Representação de sequências de <i>logs</i> para uma janela de 5 segundos atreladas ao rótulo de “confusão” .....	132
Figura 20 - Representação de sequências de <i>logs</i> para uma janela de 5 segundos atreladas ao rótulo de “não confusão” .....	133
Figura 21 - Representação de sequências de <i>logs</i> para uma janela variável atreladas ao rótulo de “confusão” .....	134
Figura 22 - Representação de sequências de <i>logs</i> para uma janela variável atreladas ao rótulo de “não confusão” .....	135
Figura 23 – Fragmento de registros de amostras .....	144
Figura 24 – Esquema de geração dos modelos da Tese e <i>baseline</i> .....	148
Figura 25 – <i>Pipeline</i> de geração e validação de um modelo de aprendizado de máquina .....	149
Figura 26 – Pseudocódigo de geração e validação dos modelos.....	150
Figura 27 – <i>Pipeline</i> de geração e validação de um modelo de aprendizado de máquina com seleção de atributos.....	151

Figura 28 – Pseudocódigo de geração e validação dos modelos com seleção de atributos ...	152
Figura 29 – Divisão dos <i>folds</i> no método de validação cruzada <i>k-fold</i> .....	158
Figura 30 – Tabela de contingência 2 x 2 para dois cenários de exemplo .....	161
Figura 31 – Tabelas de contingência usadas nas comparações dos modelos de janela variável .....	169
Figura 32 – Pseudocódigo de geração dos modelos para verificação de generalização.....	179
Figura 33 – Pseudocódigo de geração dos modelos para verificação de generalização com seleção de atributos de dados.....	180

## LISTA DE GRÁFICOS

Gráfico 1 - Desempenho dos melhores modelos com janelas fixas e sem seleção de atributos .....	165
Gráfico 2 - Desempenho dos melhores modelos com janelas fixas e com seleção de atributos .....	167
Gráfico 3 - Desempenhos dos melhores modelos com janela variável .....	170
Gráfico 4 – Os dez atributos de dados mais relevantes no melhor modelo da Tese .....	173
Gráfico 5 – Os dez atributos de dados mais relevantes no melhor modelo <i>baseline</i> .....	174
Gráfico 6 – Desempenho dos modelos de janela fixa que obtiveram melhor média de acurácia nos cenários .....	184
Gráfico 7 - Desempenho dos modelos de janela fixa que obtiveram melhor média de acurácia nos cenários com seleção de atributos de dados.....	187

## LISTA DE QUADROS

Quadro 1 – Componentes de emoção no modelo multicomponencial .....	40
Quadro 2 – Trabalhos relacionados .....	58
Quadro 3 – Questões do questionário sobre as experiências anteriores dos alunos .....	76
Quadro 4 – Componentes de conhecimento vinculados a objetivos de um exercício.....	100
Quadro 5 – Parâmetros do modelo BKT .....	104
Quadro 6 – Probabilidades dos parâmetros do modelo BKT .....	104
Quadro 7 – Dados mantidos em cada registro de <i>log</i> do <i>BlueJ Afetivo</i> .....	122
Quadro 8 – Atributos de dados gerados para as amostras .....	137
Quadro 9 – Valores candidatos para os hiperparâmetros do algoritmo <i>k-Nearest Neighbor</i> .	154

## LISTA DE TABELAS

Tabela 1 – Distribuição dos alunos participantes da pesquisa.....	73
Tabela 2 – Resultados sobre as medidas de competência emocional.....	75
Tabela 3 – Resultados sobre as experiências anteriores em programação .....	77
Tabela 4 – Idades dos alunos por nível de ensino .....	78
Tabela 5 - Escores para as afirmações do CSUQ.....	93
Tabela 6 – Escores nos grupos de afirmações do CSUQ .....	94
Tabela 7 – Componentes de conhecimento do modelo de conhecimento do aluno.....	99
Tabela 8 – Simulação do modelo BKT com 25 aplicações de componente de conhecimento .....	105
Tabela 9 - Distribuição das evidências de aplicação de componentes de conhecimento .....	113
Tabela 10 – Resultados da validação cruzada <i>k-fold</i> para o modelo BKT+ <i>forget</i> geral .....	115
Tabela 11 – Resultados da validação cruzada <i>k-fold</i> para o modelo BKT do ensino técnico	117
Tabela 12 – Resultados da validação cruzada <i>k-fold</i> para o modelo BKT do ensino superior .....	117
Tabela 13 – Média dos resultados dos modelos BKT+ <i>forget</i> geral e por nível de ensino .....	118
Tabela 14 – Parâmetros gerados para os componentes de conhecimento no ensino superior	119
Tabela 15 – Parâmetros gerados para os componentes de conhecimento no ensino técnico .	120
Tabela 16 - Exemplo de estimativas geradas para um aluno do ensino superior .....	121
Tabela 17 – Quantidade de registros de <i>log</i> coletada em cada turma.....	124
Tabela 18 – Quantidade de autorrelatos de confusão em cada turma .....	127
Tabela 19 – Quantidades de sequências de <i>logs</i> atreladas a objetivos de exercício.....	131
Tabela 20 – Quantidade de sequências de <i>logs</i> para as janelas de tamanho fixo .....	133
Tabela 21 – Quantidade de sequências de <i>logs</i> para as janelas variáveis.....	136
Tabela 22 – Quantidade de amostras geradas para janelas de tamanho fixo e variável.....	145
Tabela 23 – Desempenho e comparação dos melhores modelos com janelas fixas e sem seleção de atributos de dados .....	163
Tabela 24 – Desempenho e comparação dos melhores modelos com janelas fixas e com seleção de atributos de dados .....	166
Tabela 25 - Desempenho e comparação dos melhores modelos com janela variável.....	169
Tabela 26 – Atributos de dados selecionados nos melhores modelos.....	171
Tabela 27 – Informações sobre os atributos mais relevantes no melhor modelo da Tese.....	173

Tabela 28 – Os 10 atributos mais frequentes entre os 10 atributos mais relevantes nas janelas .....	175
Tabela 29 – Os melhores valores obtidos em cada métrica nos modelos da Tese e <i>baseline</i>	176
Tabela 30 – Desempenho dos modelos de janela fixa quando treinados com amostras de cada grupo de estudantes .....	182
Tabela 31 – Desempenho dos modelos de janela fixa que obtiveram melhor média de acurácia nos dois cenários.....	184
Tabela 32 – Modelos de janela fixa com melhor generalização com seleção de atributos ....	185
Tabela 33 – Desempenho dos modelos de janela fixa que obtiveram melhor média de acurácia nos cenários com seleção de atributos de dados.....	187
Tabela 34 - Modelos de janela variável com melhor desempenho de generalização.....	188

## LISTA DE SIGLAS

AB	<i>AdaBoost</i>
AU	<i>Action Unit</i>
BKT	<i>Bayesian Knowledge Tracing</i>
BROMP	<i>Baker Rodrigo Ocumpaugh Monitoring Protocol</i>
CAAE	Certificado de Apresentação para Apreciação Ética
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CC	Componente de Conhecimento
CEP	Comitê de Ética em Pesquisa
CERT	<i>Computer Expression Recognition Toolbox</i>
CONEP	Comissão Nacional de Ética em Pesquisa
CSUQ	<i>Computer System Usability Questionnaire</i>
DKT	<i>Deep Knowledge Tracing</i>
DT	<i>Decision Tree</i>
EC-TEL	<i>European Conference on Technology-Enhanced Learning</i>
EM	<i>Expectation-Maximization</i>
EmAP-ML	<i>Emotions Annotation Protocol for Machine Learning</i>
ESCQ	<i>Emotional Skills and Competence Questionnaire</i>
F	<i>Forget</i>
FACS	<i>Facial Action Coding System</i>
FN	Falso negativo
FP	Falso positivo
G	<i>Guess</i>
GRU	<i>Gated Recurrent Unit Neural Network</i>
GTB	<i>Gradient Tree Boosting</i>
HMM	<i>Hidden Markov Model</i>
IDE	<i>Integrated Development Environment</i>
JO	Juiz Online
k-NN	<i>k-Nearest Neighbor</i>
L	<i>Learning</i>
LOOCV	<i>Leave-One-Out Cross fold Validation</i>
LpOCV	<i>Leave-p-out student-level cross-validation</i>

LR	<i>Logistic Regression</i>
LSTM	<i>Long-Short Term Memory Network</i>
MLP	<i>Multilayer Perceptron</i>
MOOC	<i>Massive Open Online Course</i>
NB	<i>Naïve Bayes</i>
OE1	Objetivo Específico 1
OE2	Objetivo Específico 2
OE3	Objetivo Específico 3
OMM	<i>Observable Markov Model</i>
PPGCA	Programa de Pós-Graduação em Computação Aplicada
PSSUQ	<i>Post-Study System Usability Questionnaire</i>
QCE	Questionário de Competência Emocional
QP	Questão de Pesquisa
RF	<i>Random Forest</i>
RFE	<i>Recursive Feature Elimination</i>
RMSE	<i>Root Mean Square Error</i>
RNN	<i>Recurrent Neural Network</i>
S	<i>Slip</i>
SA	Atributos de dados
SBIE	Simpósio Brasileiro de Informática na Educação
SGD	<i>Stochastic Gradient Descent</i>
SQP1	Subquestão de Pesquisa 1
SQP2	Subquestão de Pesquisa 2
SQP3	Subquestão de Pesquisa 3
SVM	<i>Support Vector Machine</i>
T	<i>Transition</i>
TA	Termo de Assentimento
TCLE	Termo de Consentimento Livre e Esclarecido
TI	Tecnologia da Informação
UML	<i>Unified Modeling Language</i>
VN	Verdadeiro negativo
VP	Verdadeiro positivo
XGB	<i>XGBoost</i>

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>23</b>
<b>1.1 Tema .....</b>	<b>25</b>
<b>1.2 Trabalhos relacionados .....</b>	<b>27</b>
<b>1.3 Questão de pesquisa e hipótese.....</b>	<b>30</b>
<b>1.4 Objetivos.....</b>	<b>34</b>
<b>1.5 Contribuições da Tese .....</b>	<b>36</b>
<b>1.6 Organização do documento .....</b>	<b>37</b>
<b>2 A EMOÇÃO CONFUSÃO .....</b>	<b>39</b>
<b>2.1 Estados afetivos e emoções.....</b>	<b>39</b>
<b>2.2 <i>Appraisal</i> de emoções .....</b>	<b>42</b>
<b>2.3 Dinâmica das emoções na aprendizagem .....</b>	<b>43</b>
<b>2.4 Confusão .....</b>	<b>44</b>
2.4.1 O <i>appraisal</i> da confusão.....	45
2.4.2 As expressões de confusão .....	46
2.4.3 As facetas da confusão para resultados de aprendizagem .....	47
<b>2.5 Confusão em atividades de programação de computadores .....</b>	<b>48</b>
<b>3 DETECÇÃO DE CONFUSÃO EM AMBIENTES COMPUTACIONAIS DE APRENDIZAGEM DE PROGRAMAÇÃO DE COMPUTADORES .....</b>	<b>51</b>
<b>3.1 Detecção de emoções em ambientes computacionais de aprendizagem .....</b>	<b>51</b>
<b>3.2 Obtenção de rótulos de emoções para aprendizagem de máquina supervisionada ...</b>	<b>53</b>
<b>3.3 Detecção de emoções baseada em sensores <i>versus</i> livre de sensores .....</b>	<b>55</b>
<b>3.4 Trabalhos relacionados: detectores de confusão em ambientes de programação .....</b>	<b>56</b>
3.4.1 Método de busca e seleção dos trabalhos relacionados .....	56
3.4.2 Trabalhos selecionados .....	60
3.4.3 Discussão dos trabalhos relacionados.....	65
<b>4 TRABALHO DESENVOLVIDO .....</b>	<b>69</b>
<b>5 PARTICIPANTES DA PESQUISA.....</b>	<b>73</b>
<b>5.1 Competências emocionais .....</b>	<b>74</b>
<b>5.2 Experiências anteriores em programação de computadores.....</b>	<b>75</b>
<b>5.3 Questões éticas de pesquisa.....</b>	<b>79</b>
<b>6 AMBIENTE DE PROGRAMAÇÃO ADAPTADO PARA A TESE .....</b>	<b>81</b>
<b>6.1 Feeper .....</b>	<b>81</b>

6.1.1 Juiz Online.....	85
6.1.2 Cenário de utilização do <i>Feeper</i> .....	86
<b>6.2 Ambiente utilizado na Tese.....</b>	<b>86</b>
<b>6.3 <i>BlueJ</i> Afetivo.....</b>	<b>87</b>
<b>6.4 Avaliação de usabilidade do <i>BlueJ</i> Afetivo.....</b>	<b>92</b>
6.4.1 Resultados.....	93
<b>7 MODELO DE CONHECIMENTO DO ALUNO DESENVOLVIDO PARA A TESE 95</b>	
<b>7.1 Modelos de conhecimento do aluno.....</b>	<b>95</b>
<b>7.2 Componentes de conhecimento.....</b>	<b>98</b>
7.2.1 Relação entre os componentes de conhecimento e os objetivos de exercício.....	99
<b>7.3 Modelo <i>Bayesian Knowledge Tracing</i> (BKT).....</b>	<b>102</b>
7.3.1 Extensão do modelo com um parâmetro sobre “esquecimento”.....	108
7.3.2 Adequação dos parâmetros do modelo BKT.....	109
<b>7.4 Coleta de evidências de aplicação dos componentes de conhecimento.....</b>	<b>111</b>
<b>7.5 Treinamento e validação do modelo.....</b>	<b>114</b>
7.5.1 Resultados.....	115
7.5.2 Parâmetros gerados para os componentes de conhecimento.....	120
<b>8 GERAÇÃO DE AMOSTRAS E OBTENÇÃO DE RÓTULOS DE CONFUSÃO.....</b>	<b>122</b>
<b>8.1 Coleta de dados de interação com o <i>BlueJ</i> Afetivo.....</b>	<b>122</b>
<b>8.2 Coleta de rótulos de confusão.....</b>	<b>124</b>
8.2.1 <i>EmAP-ML</i> - Protocolo de anotação de emoções desenvolvido para a Tese.....	125
8.2.2 Mudança de método de obtenção de rótulos de emoções.....	125
<b>8.3 Geração de amostras para os modelos de aprendizado de máquina.....</b>	<b>128</b>
8.3.1 Identificação de sequências de <i>logs</i> .....	130
8.3.2 Identificação de janelas de <i>logs</i> .....	132
8.3.3 Geração de amostras.....	136
<b>9 EFEITOS DE ESTIMATIVAS DE CONHECIMENTO DO ALUNO EM DIFERENTES TAMANHOS DE JANELAS DE OBSERVAÇÃO.....</b>	<b>146</b>
<b>9.1 Método 1: geração e comparação dos modelos.....</b>	<b>146</b>
9.1.1 <i>Pipeline</i> de geração dos modelos.....	149
9.1.2 Algoritmos.....	153
9.1.3 Normalização das amostras.....	153
9.1.4 Otimização de hiperparâmetros.....	153
9.1.5 Métricas de avaliação.....	155

9.1.6 Seleção dos melhores atributos de dados .....	157
9.1.7 Validação dos modelos.....	158
9.1.8 Comparação da acurácia preditiva dos modelos da Tese e <i>baseline</i> .....	160
9.1.9 Ambiente de implementação e execução dos modelos .....	162
<b>9.2 Método 2: identificação dos atributos de dados mais relevantes .....</b>	<b>162</b>
<b>9.3 Resultados .....</b>	<b>163</b>
<b>10 EFEITOS DE ESTIMATIVAS DE CONHECIMENTO DO ALUNO NA GENERALIZAÇÃO DOS MODELOS PARA ALUNOS HETEROGÊNEOS .....</b>	<b>178</b>
<b>10.1 Método .....</b>	<b>178</b>
<b>10.2 Resultados .....</b>	<b>181</b>
<b>11 CONCLUSÃO.....</b>	<b>190</b>
<b>11.1 Limitações e trabalhos futuros .....</b>	<b>197</b>
<b>REFERÊNCIAS .....</b>	<b>201</b>
<b>APÊNDICE A – TODOS OS RESULTADOS DE VALIDAÇÃO CRUZADA DOS MODELOS NAS AMOSTRAS DE JANELAS FIXAS .....</b>	<b>216</b>
<b>APÊNDICE B – TODOS OS RESULTADOS DE VALIDAÇÃO CRUZADA DOS MODELOS NAS AMOSTRAS DE JANELAS FIXAS COM ATRIBUTOS DE DADOS SELECIONADOS .....</b>	<b>218</b>
<b>APÊNDICE C – TODOS OS RESULTADOS DE VALIDAÇÃO CRUZADA DOS MODELOS NAS AMOSTRAS DE JANELAS VARIÁVEIS.....</b>	<b>220</b>
<b>APÊNDICE D – OCORRÊNCIAS DOS ATRIBUTOS DE DADOS SELECIONADOS NOS MODELOS DA TESE .....</b>	<b>222</b>
<b>APÊNDICE E – OCORRÊNCIAS DOS ATRIBUTOS DE DADOS SELECIONADOS NOS MODELOS <i>BASELINE</i>.....</b>	<b>224</b>
<b>APÊNDICE F – ATRIBUTOS DE DADOS RELEVANTES NAS AMOSTRAS DA TESE .....</b>	<b>226</b>
<b>APÊNDICE G – ATRIBUTOS DE DADOS RELEVANTES NAS AMOSTRAS <i>BASELINE</i> .....</b>	<b>228</b>
<b>APÊNDICE H – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS - TREINAMENTO COM AMOSTRAS DO TÉCNICO E TESTES COM AMOSTRAS DO ENSINO SUPERIOR .....</b>	<b>230</b>
<b>APÊNDICE I – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS - TREINAMENTO COM AMOSTRAS DO TÉCNICO E TESTES COM</b>	

AMOSTRAS DO ENSINO SUPERIOR – COM ATRIBUTOS DE DADOS SELECIONADOS .....	232
APÊNDICE J – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS - TREINAMENTO COM AMOSTRAS DO ENSINO SUPERIOR E TESTES COM AMOSTRAS DO TÉCNICO.....	234
APÊNDICE K – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS - TREINAMENTO COM AMOSTRAS DO ENSINO SUPERIOR E TESTES COM AMOSTRAS DO TÉCNICO – COM ATRIBUTOS DE DADOS SELECIONADOS .....	236
APÊNDICE L – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS – MÉDIA DOS RESULTADOS.....	238
APÊNDICE M – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS – MÉDIA DOS RESULTADOS – COM ATRIBUTOS SELECIONADOS .....	240
APÊNDICE N – TODOS OS RESULTADOS SOBRE GENERALIZAÇÃO NOS MODELOS DE JANELAS VARIÁVEIS .....	242
APÊNDICE O – HIPERPARÂMETROS CANDIDATOS.....	244
APÊNDICE P – TERMO DE ASSENTIMENTO (TA).....	245
APÊNDICE Q – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE) PARA ALUNOS MENORES DE IDADE.....	247
APÊNDICE R – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE) PARA ALUNOS MAIORES DE IDADE.....	249
APÊNDICE S – PARECER DO CEP DA UNISINOS SOBRE O PROJETO DE PESQUISA RELACIONADO A ESTA TESE.....	251
ANEXO A - QUESTIONÁRIO DE COMPETÊNCIA EMOCIONAL (QCE).....	255



## 1 INTRODUÇÃO

As emoções possuem papel importante na aprendizagem de alunos, pois afetam processos cognitivos como a memorização e a tomada de decisão, podendo beneficiar ou prejudicar o processo de aquisição de conhecimento e de habilidades (PEKRUN, 2011). Quando o aluno está transcorrendo bem na direção dos objetivos de uma tarefa de aprendizagem, emoções como engajamento e orgulho podem beneficiar o aprendizado (FREDRICKSON, 1998). No entanto, quando o aluno fica travado em uma tarefa, sem conseguir avançar, emoções como confusão, frustração e tédio podem surgir e prejudicar o estudante, principalmente se ele permanecer travado por um período longo (D’MELLO; GRAESSER, 2014b).

Nas últimas décadas, computadores têm sido cada vez mais utilizados para promover a aprendizagem e as emoções têm sido consideradas neste contexto. Pesquisadores das áreas de psicologia educacional e computação afetiva têm se interessado nos aspectos emocionais existentes na interação de usuários com computadores. Na área de computação afetiva, um dos principais objetivos de suas pesquisas é desenvolver computadores que detectam e respondem às emoções dos seus usuários (PICARD, 2010). Diversos estudos apresentaram evidências sobre a frequência com que as emoções se manifestam na interação de usuários com computadores. Já foi encontrado que as emoções engajamento, tédio, confusão e frustração são mais frequentes do que as emoções básicas (felicidade, tristeza, medo, surpresa, raiva e nojo) quando usuários interagem com ambientes computacionais (D’MELLO; CALVO, 2013). No contexto de aprendizagem com computadores, uma meta-análise envolvendo 24 estudos mostrou que as seguintes emoções são mais frequentes: engajamento/*flow*<sup>1</sup>, tédio, confusão, curiosidade, felicidade e frustração (D’MELLO, 2013).

Não há consenso entre os pesquisadores de que fenômenos afetivos frequentes na aprendizagem, como a confusão, a frustração, o tédio e o engajamento sejam emoções. Isso ocorre porque estes estados afetivos compartilham algumas, mas não todas as características comumente atribuídas às emoções básicas (D’MELLO; CALVO, 2013). No entanto, estes estados afetivos são comumente referenciados como emoções. Alguns autores as chamam de **emoções não básicas** (D’MELLO; CALVO, 2013; D’MELLO; KORY, 2012) ou **emoções centradas na aprendizagem** (AZEVEDO; TAUB; MUDRICK, 2019). O presente trabalho usa o termo *emoção* para designar estes estados afetivos relacionados com a aprendizagem.

---

<sup>1</sup> Pode ser traduzido como “fluxo”.

A Tese tem interesse na emoção **confusão**, no contexto de tarefas de aprendizagem de programação de computadores, durante a interação de alunos com um ambiente computacional, como quando os alunos interagem com um software editor de programação para desenvolver uma tarefa de aprendizagem. Estudar a confusão de alunos na aprendizagem de programação de computadores é interessante, pois é uma emoção propensa de ocorrer em ambientes de aprendizagem complexa, como em programação de computadores (LEHMAN et al., 2013). Além de precisar entender o problema subjacente à tarefa de programação, o aluno precisa mobilizar esforços cognitivos para identificar e relacionar conhecimentos e estratégias de resolução de problemas e tomar decisões (CHI; OHLSSON, 2005; LEHMAN et al., 2013). Juntamente com engajamento, frustração, tédio e curiosidade, a confusão é uma das emoções mais frequentes durante a aprendizagem de programação (BOSCH; D’MELLO, 2017; BOSCH; D’MELLO; MILLS, 2013; COTO et al., 2021).

É possível tentar explicar a frequência com que a confusão é observada em ambientes de aprendizagem, através de teorias de emoções. A confusão pode ser abordada como uma emoção relacionada ao conhecimento, ou seja, uma emoção associada a metas de aprendizagem (SILVIA, 2010). Teorias de emoções que associam a confusão com o conhecimento, a consideram um fenômeno afetivo multicomponencial que inclui um componente cognitivo de *appraisal*<sup>2</sup> (a Seção 2.1 descreve esta abordagem multicomponencial das emoções). O *appraisal* é um processo interno de um indivíduo, que detecta e avalia a significância do ambiente (satisfação ou não satisfação de necessidades individuais, valores, metas e crenças) para o bem-estar pessoal (FRIJDA, 1988). Neste contexto teórico, a confusão é um estado afetivo que se supõe ocorrer quando o indivíduo avalia que a informação que chega não corresponde ao seu conhecimento anterior, ou quando a informação que chega não pode ser integrada ao modelo mental existente. Quando o aluno está confuso, ele entra em um estado de desequilíbrio cognitivo, resultante de contradições, conflitos ou informações erradas apresentadas a ele (D’MELLO; GRAESSER, 2014b). Quando este desequilíbrio ocorre, além de entrar em confusão, o indivíduo poderá experimentar um conjunto de outras emoções com efeitos negativos para a aprendizagem, como a frustração e o tédio, que o levam ao desengajamento (D’MELLO et al., 2014; SILVIA, 2010). Quando o equilíbrio é recuperado, mais emoções com efeitos positivos para a aprendizagem emergem, como o engajamento / *flow* (GRAESSER; D’MELLO, 2011). Quando resolvida, a confusão é benéfica para o estudante e pode levar a aprendizagem a níveis mais profundos. Isso ocorre porque o desequilíbrio

---

<sup>2</sup> Pode ser traduzido como “avaliação”.

cognitivo causado pela confusão pode levar o aluno a mobilizar esforços cognitivos, levando a uma aprendizagem mais profunda (D’MELLO; GRAESSER, 2014b). Mas quando o aluno não se engaja na resolução da confusão, os resultados para a aprendizagem podem ser negativos (D’MELLO; GRAESSER, 2012a). Outro aspecto que pode fazer a confusão levar o aluno a resultados de aprendizagem negativos é quando ele possui defasagem de habilidades metacognitivas e autorregulatórias. Ao apresentar estas defasagens, o estudante pode não conseguir se engajar para resolver os impasses enfrentados em tarefas atribuídas em ambientes mais complexos de aprendizagem (LEHMAN; D’MELLO; GRAESSER, 2012), como o de programação de computadores.

### 1.1 Tema

A Tese foi desenvolvida como uma pesquisa de computação aplicada, na qual ferramentas da computação são utilizadas para solucionar problemas de outras áreas, como a psicologia educacional e a educação. A primeira definição da Tese foi a definição do seguinte tema de pesquisa:

**Tema: modelos computacionais de detecção<sup>3</sup> da confusão de alunos durante tarefas de aprendizagem de programação de computadores.**

Uma primeira justificativa para este tema é que modelos computacionais que identificam a confusão dos alunos podem gerar alertas ao professor, tutor humano, ou ambiente computacional de aprendizagem, sobre o estado de confusão do estudante. Identificar a confusão de alunos em ambientes computacionais de aprendizagem complexa não é uma atividade fácil para um professor, principalmente quando ele é o único agente prestando assistência aos estudantes. A partir dos alertas do ambiente computacional, é possível tomar decisões instrucionais, como ajudar o aluno a sair do desequilíbrio cognitivo, através da entrega de alguma dica ou atribuir tarefas mais apropriadas para o conhecimento do estudante. O aluno também poderia ser encorajado a regular o seu estado de confusão, levando-o a refletir sobre sua confusão na tarefa e a mudar sua estratégia de solução do problema, como solicitar ajuda ao colega ou buscar respostas em materiais de estudo.

---

<sup>3</sup> O texto da Tese também utiliza o verbo “identificar”, como sinônimo de “detectar”.

A confusão é uma das emoções mais frequentes durante a interação de usuários em ambientes computacionais em geral (D'MELLO; CALVO, 2013), em ambientes computacionais de aprendizagem (D'MELLO, 2013) e, especificamente, em ambientes computacionais de aprendizagem de programação de computadores (BOSCH; D'MELLO, 2017; BOSCH; D'MELLO; MILLS, 2013). Estudos já mostraram uma correlação negativa da confusão de programadores novatos no seu desempenho de aprendizagem durante a codificação (BOSCH; D'MELLO; MILLS, 2013) e em testes de conhecimento (GRAFSGAARD et al., 2012; RODRIGO et al., 2009). No entanto, há cada vez mais evidências mostrando que a confusão precede o sucesso de aprendizagem (D'MELLO et al., 2014; RICHEY et al., 2019; VANLEHN et al., 2003). Na aprendizagem complexa, como em programação de computadores, o aluno precisa diagnosticar e resolver problemas, fazer comparações de conceitos e inferências, aplicar e transferir conhecimentos prévios adquiridos e gerar explicações coerentes. Nestes cenários, é provável que a confusão ocorra e que seja mais norma do que exceção. Quando o aluno se engaja para resolver estas adversidades e avança na solução, é provável que o estudante experimente níveis mais profundos de compreensão do conteúdo. No entanto, quando o aluno fica travado na tarefa e a confusão não é resolvida, ele pode entrar em um ciclo de emoções correlacionadas negativamente com a aprendizagem, como a frustração e o tédio (D'MELLO et al., 2014; LEHMANN; GRAESSER, 2015).

Quando o aluno recebe assistência para resolver a confusão gerada pelas contradições e conflitos, a confusão pode ser benéfica para a aprendizagem. Identificar a confusão do aluno no andamento das tarefas de aprendizagem e, a partir daí, ajudá-lo a regular suas emoções e a sair do desequilíbrio cognitivo, permite encorajá-lo a enfrentar tarefas de aprendizagem complexas que exigem processos cognitivos complexos (D'MELLO et al., 2014). Isto não significa que o ambiente de aprendizagem deva evitar a confusão dos alunos. A confusão pode levar o aluno a engajar esforços, manter foco e atenção na tarefa e a buscar o conhecimento (D'MELLO; GRAESSER, 2014b). Uma discussão apresentada por Richey et al. (2019) sugere que ao detectar que o aluno se mantém muito tempo sem estar confuso, o sistema poderia realizar ações instrucionais que levassem o estudante a entrar em um estado de confusão benéfica para a aprendizagem, como apresentar exemplos com contradições. O ambiente também poderia não agir por determinado tempo, ao detectar a confusão do aluno, levando-o a experimentar um período benéfico de confusão.

## 1.2 Trabalhos relacionados

Uma revisão da literatura foi realizada com o objetivo de encontrar trabalhos relacionados ao tema da Tese. Todos os trabalhos relacionados (BOSCH; CHEN; D’MELLO, 2014; FELIPE et al., 2012; GRAFSGAARD; BOYER; LESTER, 2011; LEE et al., 2011; TIAM-LEE; SUMI, 2018, 2019; VEA; RODRIGO, 2017a) utilizaram a mesma abordagem computacional para os modelos de detecção de confusão de alunos em tarefas de aprendizagem de programação de computadores: modelos compostos por algoritmos de aprendizado de máquina supervisionada. Estes modelos visualizam um conjunto de amostras de dados de entrada que refletem determinados comportamentos do problema e tentam aprender padrões relacionados a classes de saída pré-definidas. Após aprender os padrões, os modelos visualizam um outro conjunto de amostras de dados, não utilizado no aprendizado, e tentam identificar corretamente as classes de interesse (KUBAT, 2017). No contexto da Tese, os modelos dos trabalhos relacionados visualizam amostras de dados sobre o comportamento dos alunos no ambiente das tarefas de programação de computadores e tentam aprender padrões relacionados à confusão e à não confusão dos estudantes.

Os trabalhos relacionados utilizaram as seguintes modalidades de dados no aprendizado dos padrões sobre a confusão (alguns buscaram identificar também outras emoções, além da confusão): expressões faciais (BOSCH; CHEN; D’MELLO, 2014; GRAFSGAARD; BOYER; LESTER, 2011; TIAM-LEE; SUMI, 2018, 2019); informações de compilação de código (LEE et al., 2011; TIAM-LEE; SUMI, 2018, 2019); pressionamento de teclas (FELIPE et al., 2012; TIAM-LEE; SUMI, 2018, 2019; VEA; RODRIGO, 2017b); movimentos de mouse (VEA; RODRIGO, 2017b); orientação da cabeça, posição do nariz, tamanho do rosto e visibilidade de nariz, boca, olhos e ouvidos (BOSCH; CHEN; D’MELLO, 2014); informações textuais (EBERT et al., 2017; GRAFSGAARD; BOYER; LESTER, 2011); e status da tarefa (correta, incorreta, incompleta e mistura de ambas) (GRAFSGAARD; BOYER; LESTER, 2011). Todos os trabalhos relacionados utilizaram conjuntos de dados compostos por diferentes modalidades de dados, como o trabalho de Tiam-Lee e Sumi (2018), que utilizou dados de expressões faciais, de digitação de teclas e de compilações de código-fonte.

Os trabalhos relacionados que utilizaram dados sobre expressões faciais (BOSCH; CHEN; D’MELLO, 2014; GRAFSGAARD; BOYER; LESTER, 2011; TIAM-LEE; SUMI, 2018, 2019) e dados de orientação da cabeça, posição do nariz, tamanho do rosto e visibilidade de nariz, boca, olhos e ouvidos (BOSCH; CHEN; D’MELLO, 2014), utilizaram câmeras instaladas em computadores para a obtenção desses dados. Estes modelos podem ser chamados

de *modelos baseados em sensores*, pois necessitam de algum equipamento periférico (o sensor), como câmeras, sensores de condutância de pele e de postura, para a obtenção dos dados (BOTELHO; BAKER; HEFFERNAN, 2017). Uma conhecida dificuldade relacionada ao uso de sensores em ambientes computacionais de aprendizagem é o custo extra necessário para a aquisição de equipamentos periféricos para utilização em salas de informática. Em função disso, sistemas que utilizam estes sensores são mais difíceis de serem aplicados em escala e muitas vezes inviáveis de serem aplicados em instituições de ensino com baixo orçamento (YANG et al., 2019). Além disso, estes detectores são mais intrusivos, podendo afetar o curso normal da aprendizagem (BOTELHO et al., 2018). O uso de sensores também resulta em preocupações relacionadas à privacidade dos usuários (REYNOLDS; PICARD, 2004).

Os demais trabalhos relacionados (FELIPE et al., 2012; LEE et al., 2011; VEA; RODRIGO, 2017b) avaliaram modelos detectores de confusão que não utilizaram equipamentos periféricos para a coleta de dados sobre os estudantes. Estes modelos podem ser chamados de *modelos livres de sensores*. Modelos livres de sensores utilizam apenas dados relacionados à atividade do usuário no sistema, como ações de teclado e mouse e demais interações com a interface gráfica (BOTELHO; BAKER; HEFFERNAN, 2017). Estas interações com o sistema podem, ainda, derivar outros tipos de dados, como no caso de sistemas tutores inteligentes, que mantêm um modelo de conhecimento do aluno a partir de dados sobre o desempenho do aluno nas tarefas de aprendizagem (BAKER et al., 2012). A Tese tem interesse em apresentar contribuições relacionadas a modelos livres de sensores.

Os três trabalhos relacionados que utilizaram modelos livres de sensores (FELIPE et al., 2012; LEE et al., 2011; VEA; RODRIGO, 2017b) apresentaram algumas restrições importantes. O trabalho de Felipe et al. (2012) avaliou um modelo detector de confusão e tédio que utilizou dados derivados do pressionamento de teclas durante a codificação<sup>4</sup>, como velocidade média de digitação e quantidade de uso de *backspaces*. Apesar de ter obtido um bom desempenho ( $kappa^5 = 0,856$ ), este trabalho utilizou dados de uma quantidade pequena de alunos (6) no treinamento dos modelos. O trabalho de Vea e Rodrigo (2017) também utilizou dados de pressionamento de teclas, e acrescenta atributos de dados derivados de dinâmicas de uso do mouse. Com o objetivo de classificar confusão, tédio e frustração, este trabalho obteve

---

<sup>4</sup> A partir daqui, sempre que o termo “codificação” for utilizado, considere como uma tarefa de programação de computadores realizada pelo aluno.

<sup>5</sup> O kappa de Cohen é uma medida de concordância entre dois juízes. É uma medida baseada na acurácia, mas faz um ajuste na medição ao considerar a possibilidade de a concordância ter ocorrido por acaso (COHEN, 1960). Tem sido utilizado em trabalhos de detecção de emoções para verificar a concordância entre um juiz humano e o detector de emoções.

um valor moderado de desempenho ( $kappa = 0,572$ ), mas ainda abaixo de uma concordância esperada para juízes humanos ( $kappa = 0,6$ ) (OCUMPAUGH; BAKER; OCUMPAUGH, J., & BAKER, 2014). O trabalho de Lee et al. (2011) buscou detectar a confusão através da utilização de dados sobre compilações de códigos do aluno, como a quantidade de erros na compilação e a média de tempo entre compilações com e sem erros. Apesar de ter encontrado um bom desempenho ( $kappa = 0,86$ ), o artigo de Baker et al. (2012) apresentou um questionamento sobre os resultados deste trabalho relacionado. Os autores questionaram se, através do uso de dados de compilação, este trabalho estava realmente avaliando a confusão do aluno. Talvez o principal questionamento sobre este trabalho seja como utilizar dados de compilação para detectores de confusão em tarefas de programação, em tempo de codificação. Dados sobre compilações são obtidos somente após o aluno se engajar na codificação e submeter seu código para avaliação. A detecção de confusão, neste contexto, poderia ser tardia e o ambiente computacional de aprendizagem poderia não se beneficiar das inferências desse tipo de modelo que considera dados de compilações, obtidos pós-codificação.

Outra limitação importante dos trabalhos relacionados é que nenhum deles teve objetivo explícito de verificar o poder de generalização dos modelos para estudantes com características heterogêneas. Modelos de classificação *data-driven*, como os dos trabalhos relacionados, que usam dados sem relação causal ou relação conhecida com alguma teoria de emoções, como dados sobre pressionamento de teclas, podem ter dificuldade para transferir seu desempenho para outros grupos de sujeitos (D'MELLO; GRAESSER, 2014b). Um modelo com bom poder de generalização poderia ser treinado com amostras de determinado grupo de estudantes e utilizado para detecção de confusão em amostras de outros grupos de estudantes. Dois dos trabalhos relacionados (LEE et al., 2011; VEA; RODRIGO, 2017b) validaram seus modelos com validação cruzada a nível de estudantes. Em linhas gerais, este método de validação avalia, em certo nível, o poder de generalização para diferentes estudantes. Este método seleciona aleatoriamente estudantes que terão amostras utilizadas no treino dos modelos e estudantes que terão amostras utilizadas nos testes. O aluno que é selecionado para o treino do modelo não terá amostras de dados selecionadas nos testes. Isto garante um determinado nível de generalização, mas não garante que alunos com determinadas características sejam usadas no treino e estudantes com outras características sejam utilizadas nos testes. Se assumirmos, por exemplo, que estudantes de programação de cursos de graduação possuem características distintas de estudantes de programação de cursos técnicos, a validação cruzada dos modelos dos trabalhos relacionados não verificaria a generalização dos modelos para estudantes com características

distintas. A Seção 3.4 apresenta a descrição completa da revisão de literatura e dos trabalhos relacionados.

### 1.3 Questão de pesquisa e hipótese

A revisão de literatura mostrou que modelos de detecção livres de sensores da confusão de alunos em tarefas de aprendizagem de programação de computadores ainda têm sido pouco explorados. Os três principais trabalhos relacionados apresentaram limitações importantes, descritas anteriormente. Resolver estas limitações já poderia trazer novas evidências para a área ou reforçar as já existentes. A questão de pesquisa levantada pela Tese buscou sugerir uma nova abordagem a ser investigada para o problema. A questão surgiu a partir de discussões encontradas na literatura envolvendo detectores de emoções que utilizam uma **abordagem dirigida a dados**<sup>6</sup> e detectores que utilizam uma **abordagem guiada à teoria**<sup>7</sup>. Modelos detectores de emoções que utilizam uma **abordagem dirigida a dados** buscam encontrar correlações desconhecidas entre *atributos de dados*<sup>8</sup> e julgamentos de emoções realizados por humanos (D'MELLO; GRAESSER, 2014b). Estes julgamentos podem ser realizados pelos participantes da pesquisa (autorrelatos) ou por anotadores especialistas externos que observam os comportamentos dos participantes. A abordagem dirigida a dados utiliza atributos de dados que não são justificados por teorias de emoções. Por um lado, esta abordagem é interessante para encontrar correlações desconhecidas entre atributos de dados e emoções. Por outro lado, por não terem uma relação causal e/ou teórica com as emoções, modelos que utilizam a abordagem dirigida a dados podem apresentar problemas de falso-positivo, embora o risco possa ser tratado com métodos de validação cruzada (D'MELLO; GRAESSER, 2014b). Esta problemática do mapeamento de dados com determinados construtos emocionais já foi alertada por alguns autores, no contexto de detecção de confusão em programação de computadores (BAKER et al., 2012; BAKER; OCUMPAUGH, 2015; FELIPE et al., 2012). A **abordagem guiada à teoria**, por sua vez, considera a utilização de atributos de dados com relação mais direta com teorias de emoções, como alguma relação causal conhecida (D'MELLO; GRAESSER, 2014b).

É conhecido na literatura o forte relacionamento entre emoções básicas e expressões faciais (EKMAN; FRIESEN, 1987; ELFENBEIN; AMBADY, 2002). Modelos que detectam

---

<sup>6</sup> Do inglês *data-driven*.

<sup>7</sup> Do inglês *theory-guided*.

<sup>8</sup> Refere-se ao conceito de “*feature*”, na literatura de aprendizado de máquina.

emoções básicas a partir de expressões faciais poderiam ser chamados de modelos guiados à teoria. No entanto, os trabalhos que avaliaram detectores de confusão a partir do uso de dados faciais, com o uso de câmeras, são considerados detectores baseados em sensores. Todos os trabalhos relacionados à Tese que construíram modelos de aprendizagem de máquina supervisionada livres de sensores utilizaram uma abordagem dirigida a dados.

Um artigo de D’Mello (2020), sobre o uso de dados em ambientes computacionais de aprendizagem, apresenta uma crença do autor de que modelos preditivos seriam mais fortes se combinassem uma abordagem dirigida a dados e uma abordagem guiada à teoria. Segundo o autor, esta combinação de abordagens é raramente realizada e poderia gerar modelos preditivos com melhor relação causal. A partir das discussões apresentadas envolvendo as abordagens dirigida a dados e guiada à teoria, a Tese levantou a seguinte questão de pesquisa:

**Questão de Pesquisa (QP): modelos livres de sensores que utilizam uma abordagem mista guiada à teoria e dirigida a dados, poderiam apresentar um desempenho superior a modelos livres de sensores que utilizam uma abordagem dirigida a dados, na detecção da confusão do aluno durante tarefas de aprendizagem de programação de computadores?**

A Tese também levanta uma hipótese para a questão de pesquisa:

**Hipótese: a utilização de dados sobre estimativas de conhecimento do aluno no domínio de programação de computadores, juntamente de dados sobre a interação do estudante com o ambiente computacional de aprendizagem de programação, pode melhorar o desempenho dos modelos livres de sensores na detecção da confusão do aluno nas tarefas de aprendizagem de programação de computadores, quando comparado a utilizar apenas os dados de interação do estudante com o ambiente de aprendizagem.**

Em outras palavras, a Tese levanta a hipótese de que ao adicionar dados sobre estimativas de conhecimento dos alunos no domínio de programação de computadores aos dados já utilizados pelos modelos livres de sensores dos trabalhos relacionados, será possível verificar uma melhora da capacidade preditiva dos modelos, em comparação com os modelos que não utilizam as estimativas de conhecimento dos alunos. Esta é uma abordagem mista, pois mistura uma abordagem de dados guiados por teoria (as estimativas de conhecimento do aluno) e uma abordagem dirigida a dados dos trabalhos relacionados (teclado, mouse, interação com interface gráfica, entre outros). Os próximos parágrafos apresentam uma justificativa para a

hipótese da Tese, a partir de uma descrição da relação conhecida na literatura entre o conhecimento do aluno e a emoção confusão.

A confusão pode ser abordada como uma emoção não básica relacionada ao conhecimento (SILVIA, 2010). Esta ideia remete à um modelo multicomponencial das emoções (o Capítulo 2 detalha este modelo), no qual o *appraisal* (avaliação) é um componente cognitivo das emoções. Segundo as teorias de *appraisal* de emoções, a confusão surge quando o indivíduo é confrontado com uma contradição, como quando a informação que chega não pode ser integrada ao mapa mental existente, ou quando está incerto do que fazer na sequência de uma tarefa (D'MELLO; GRAESSER, 2014). O trabalho de Paul Silvia (2010) descreve emoções relacionadas ao conhecimento como emoções causadas pelas crenças dos indivíduos sobre seus próprios pensamentos e conhecimentos, e decorrem de objetivos associados ao aprendizado. O trabalho de Paul Silvia encontrou evidências positivas para a hipótese de que a confusão está associada com *appraisals* de novidade (algo não familiar) e de baixa compreensibilidade, no contexto da informação que chega para o indivíduo (SILVIA, 2010). Outros estudos também encontraram evidências de uma correlação da confusão com a falta de conhecimento (D'MELLO; PERSON; LEHMAN, 2009; LEE et al., 2011; RODRIGO; BAKER; NABOS, 2010) e que a confusão é um bom preditor de resultados de teste de conhecimento (GRAESSER et al., 2007).

Outras discussões também sugerem efeitos do conhecimento prévio do aluno sobre a aprendizagem e as contradições que levam à confusão. Algumas discussões sugerem que a relação de quanto mais conhecimento/menos confusão e quanto menos conhecimento/mais confusão não seja tão evidente. Níveis mais intermediários de conhecimento podem ser mais preditores de confusão, quando o aluno já tem algum conhecimento, mas não o domina (D'MELLO et al., 2014). Um exemplo deste cenário é quando um aluno tem algum conhecimento para aplicar um determinado conhecimento em uma tarefa, mas ainda lhe falta um entendimento mais completo. Se ele se deparar com algum insucesso ou contradição na tarefa, a situação pode levá-lo a ficar confuso, talvez precedido de surpresa (SILVIA, 2010). Estudos também mostraram que alunos com baixo conhecimento prévio em determinado conteúdo, poderiam falhar em identificar contradições e nem experimentar a confusão que surge de contradições (D'MELLO et al., 2014; VANLEHN et al., 2003). Em ambos os cenários descritos, as bases teóricas indicam algum nível de relação intrínseca entre o conhecimento do aluno e a confusão.

A hipótese também considera que modelos que utilizam atributos de dados sobre estimativas de conhecimento do aluno poderiam generalizar melhor seus resultados para alunos

com características heterogêneas. Nenhum dos trabalhos relacionados verificou diretamente este comportamento. Apesar de terem feito isso a algum nível, quando utilizaram validação cruzada a nível de estudante, os trabalhos não verificaram o poder de generalização para grupos de estudantes com características heterogêneas. Diferentes turmas de programação, como turmas de alunos do ensino técnico e do ensino superior, podem ser compostas por alunos com diferentes características, como idade ou experiências anteriores de programação. Turmas de programação na graduação poderiam ser compostas por alunos que já atuam como desenvolvedores de software em empresas. Turmas de ensino técnico, por sua vez, poderiam ser formadas por alunos mais jovens e sem experiência anterior.

É conhecido que modelos de classificação que utilizam atributos de dados com boa correlação com as classes de saída, evitam efeitos de sobreajuste<sup>9</sup>, generalizando melhor seus resultados (BAKER; OCUMPAUGH, 2015). Nesse sentido, D’Mello e Graesser (D’MELLO; GRAESSER, 2014b) indicam que modelos que utilizam dados não alinhados com uma teoria de emoções poderiam cometer erros de falso-positivo. É possível supor que modelos *dirigidos a dados* podem também falhar para prever a confusão de estudantes com características diferentes daqueles alunos considerados no treinamento dos modelos. Isto pela falta de uma relação mais causal e intrínseca dos dados com o construto emocional. Ou seja, os modelos treinados poderiam funcionar de uma forma para determinado grupo de estudantes, mas funcionar de uma forma diferente para outros grupos de estudantes. A suposição da Tese é que modelos mais guiados à teoria, que se utilizam de dados com relação mais intrínseca e causal com o construto emocional, poderiam ter bom desempenho para estudantes com características heterogêneas, de forma que o modelo poderia ser treinado com amostras de um determinado grupo de estudantes e ser utilizado para predição de confusão em amostras de outro grupo de estudantes, com características heterogêneas, que não foram usadas no treino dos modelos.

Por fim, a Tese ainda definiu três subquestões de pesquisa, relacionadas a janelas de observação de dados e à generalização dos resultados. A Tese considera uma janela de observação de dados como o tempo que o modelo observa o comportamento do aluno para identificar sua emoção (confuso ou não confuso).

As três subquestões de pesquisa levantadas são as seguintes:

---

<sup>9</sup> Do inglês *overfitting*.

**Subquestão de pesquisa 1 (SQP1): qual o desempenho dos modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente e dos modelos que utilizam apenas dados sobre interação com o ambiente, em diferentes janelas de observação de tamanho fixo?**

**Subquestão de pesquisa 2 (SQP2): qual o desempenho dos modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente e dos modelos que utilizam apenas dados sobre interação com o ambiente, em uma janela de observação de tamanho variável?**

**Subquestão de pesquisa 3 (SQP3): qual o desempenho dos modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente e dos modelos que utilizam apenas dados sobre interação com ambiente na generalização de seus resultados para estudantes com características heterogêneas?**

As duas primeiras subquestões de pesquisa (SQP1 e SQP2) são de natureza exploratória, pois a Tese não levantou nenhuma hipótese para elas. Os três trabalhos relacionados consideraram janelas com tempo fixo: 15 segundos (VEA; RODRIGO, 2017b), 20 segundos (FELIPE et al., 2012) e 8 compilações seguidas (LEE et al., 2011). A Tese gostaria de verificar o desempenho dos modelos em diferentes janelas de observação de tempo fixo e variável. Para a terceira subquestão de pesquisa (SQP3), a Tese levanta a hipótese que a utilização de dados sobre estimativas de conhecimento dos alunos, justificada por teorias de emoções, poderá modelar uma melhor relação das amostras com a confusão (ou não confusão) dos alunos. Estes modelos teriam um melhor poder de generalização dos seus resultados para estudantes com características heterogêneas, em comparação com modelos apenas dirigidos à dados.

#### **1.4 Objetivos**

O objetivo geral da Tese é verificar a sua hipótese, ou seja, verificar se **a utilização de dados sobre estimativas de conhecimento do aluno no domínio de programação de computadores, juntamente de dados sobre a interação do estudante com o ambiente computacional de aprendizagem de programação, pode melhorar o desempenho dos modelos livres de sensores na detecção da confusão do aluno nas tarefas de aprendizagem**

**de programação de computadores, quando comparado a utilizar apenas os dados de interação do estudante com o ambiente de aprendizagem.**

Também foram definidos alguns objetivos específicos, que buscam encontrar respostas para as subquestões de pesquisa e apresentar evidências para a hipótese:

**Objetivo específico 1 (OE1): verificar se modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente possuem desempenho superior a modelos que utilizam apenas dados sobre interação com o ambiente, quando são consideradas diferentes janelas de observação de tamanho fixo.**

**Objetivo específico 2 (OE2): verificar se modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente possuem desempenho superior a modelos que utilizam apenas dados sobre interação com o ambiente, quando é considerada uma janela de observação de tamanho variável.**

**Objetivo específico 3 (OE3): verificar se modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente generalizam bem seus resultados para estudantes com características heterogêneas, comparado a modelos que utilizam apenas dados sobre interação com o ambiente.**

Para alcançar seus objetivos, responder as questões de pesquisa e verificar sua hipótese, a Tese treinou e avaliou diversos modelos de aprendizado de máquina que representam a abordagem da hipótese da Tese (modelos treinados com dados sobre estimativas de conhecimento do aluno + dados de interação com o ambiente) e a abordagem dos trabalhos relacionados (modelos *baseline*, treinados apenas com dados de interação com o ambiente). Os modelos foram treinados e avaliados para diferentes configurações de janelas de observação (tamanho fixo e variável), e para diferentes algoritmos. Foram obtidas métricas de desempenho (acurácia, AUC ROC, *kappa*, precisão, *recall* e F1) para cada modelo gerado e foram aplicados testes estatísticos que compararam a acurácia preditiva dos modelos de cada abordagem (Tese e *baseline*). A hipótese é que os modelos da Tese superam o desempenho dos modelos *baseline*.

Também foram aplicados métodos para identificar os dados mais relevantes na detecção da confusão pelos modelos. A hipótese é que as estimativas de conhecimento do aluno irão compor o conjunto de dados mais relevantes, juntamente dos dados de interação. O trabalho também aplicou um método de verificação do poder de generalização dos modelos para estudantes com características heterogêneas. Todos os modelos foram, em determinado momento, treinados com os dados de um grupo de estudantes e testados com os dados do outro grupo de estudantes. A hipótese é que os modelos da Tese superam o desempenho dos modelos *baseline* quando testados com amostras de dados não usadas nos treinos. Os métodos também buscam identificar em quais configurações de janelas de observação, os modelos são mais vantajosos. Os modelos foram treinados e testados com amostras compostas por dados coletados de 62 alunos de ensino técnico e superior, durante cinco meses, enquanto resolviam exercícios de programação atribuídos por professores, em um software de programação adaptado para a Tese.

## 1.5 Contribuições da Tese

A Tese contribui para o estado da arte ao apresentar uma pesquisa original que avalia uma abordagem mista, que envolve treinar modelos de detecção de confusão livres de sensores no contexto de tarefas de aprendizagem de programação, utilizando dados justificados por teorias de emoções (estimativas de conhecimento do aluno) e dados não justificados por teorias de emoções (dados de interação com o ambiente). Também apresenta uma contribuição ao verificar o poder de generalização desta abordagem para estudantes com características heterogêneas. A Tese também apresenta indicações de cenários onde esta abordagem pode ser mais vantajosa e os algoritmos com melhor desempenho para o problema apresentado.

Como será visto no decorrer do documento, a Tese apresenta ainda outras contribuições que poderão ser úteis aos pesquisadores interessados no tema da pesquisa: um protocolo de anotações de emoções desenvolvido e avaliado para a Tese, para obtenção de rótulos de emoções para modelos de aprendizado de máquina; um modelo de inferência de estimativas de conhecimento do aluno desenvolvido e avaliado para a Tese, que apesar de não ser um trabalho original, pode auxiliar implementadores de modelos de conhecimento no domínio de programação; um ambiente de programação de computadores adaptado e avaliado para a Tese, que apresenta diversas decisões de *design* para um ambiente de codificação de exercícios, de coleta de dados e de avaliação da solução dos exercícios; e a descrição do método usado na Tese para a coleta de dados e geração das amostras, que pode ser útil para implementadores de modelos de detecção de confusão no contexto da Tese.

## 1.6 Organização do documento

Este documento foi organizado nos seguintes capítulos:

**Capítulo 2** – Apresenta a fundamentação teórica subjacente sobre a emoção confusão, suas formas de expressão e facetas para os resultados de aprendizagem em programação.

**Capítulo 3** – Apresenta os trabalhos relacionados e a fundamentação teórica sobre detecção de emoções em ambientes computacionais, detectores de emoções baseados e livres de sensores e sobre a obtenção de rótulos de emoções para modelos de aprendizado de máquina.

**Capítulo 4** – Descreve uma visão geral do trabalho realizado para alcançar os objetivos da Tese.

**Capítulo 5** – Descreve os participantes da pesquisa da Tese.

**Capítulo 6** – Descreve o ambiente de programação adaptado para a Tese.

**Capítulo 7** – Descreve o modelo de conhecimento do aluno desenvolvido para a Tese. Este modelo infere e mantém as estimativas de conhecimento do aluno em programação.

**Capítulo 8** – Descreve a geração das amostras e a obtenção dos rótulos de confusão para os modelos de aprendizado de máquina.

**Capítulos 9 e 10** – Detalham os métodos utilizados pela Tese para alcançar seus objetivos e apresenta os resultados encontrados.

**Capítulo 11** – Apresenta a conclusão da Tese.



## 2 A EMOÇÃO CONFUSÃO

Este capítulo apresenta definições sobre construtos e teorias subjacentes à Tese, relacionadas às emoções e, mais especificamente, à emoção confusão. A Seção 2.1 apresenta definições gerais sobre afeto, estados afetivos e emoções. A Seção também descreve a teoria utilizada pela Tese para explicar as emoções: o modelo multicomponencial de emoções. A Seção 2.2 descreve o componente cognitivo de *appraisal* das emoções, utilizado como base teórica para justificar a hipótese da Tese. A Seção 2.3 descreve a dinâmica das emoções na aprendizagem. A Seção 2.4 aborda especificamente a emoção confusão, descrevendo o componente cognitivo de *appraisal* responsável pela confusão, as expressões relacionadas à confusão e como esta emoção interfere na aprendizagem. Por fim, a Seção 2.5 descreve as facetas da confusão em atividades de programação de computadores.

### 2.1 Estados afetivos e emoções

Antes de traçar qualquer definição sobre emoção, é necessário compreender o conceito de afeto. Na perspectiva psicológica, o afeto é um sentimento subjetivo que afeta a cognição do indivíduo (MELLO et al., 2014). A emoção é um tipo de afeto, referenciado na literatura como um estado afetivo, ou fenômeno afetivo. Estado afetivo tem sido utilizado na literatura como um termo mais abrangente para designar fenômenos afetivos como emoção, humor, atitudes, disposições afetivas e postura interpessoal, os quais podem ser diferenciados em função da duração, intensidade, impacto comportamental, eventos-gatilho e respostas (SCHERER, 2005).

A Tese está interessada no estado afetivo de emoção, um construto que vem sendo tema de estudos e reflexão de filósofos e cientistas há milhares de anos (JAQUES; VICARI, 2005; SCHERER, 1987). Acredita-se que existam expressões universais, inatas e interculturais para seis emoções, chamadas de **emoções básicas**, a saber: tristeza, alegria, raiva, surpresa, medo e nojo (EKMAN; FRIESEN, 1987).

Não há consenso entre os pesquisadores sobre um conceito para emoção, pois trata-se de um construto complexo, que já foi estudado sob diferentes perspectivas: expressão, fisiológica, cognitiva e social. A neurociência também tem contribuído recentemente para a compreensão sobre como funciona o circuito neural relacionado às experiências emocionais (CALVO; D'MELLO, 2010). Um exemplo de abordagem de emoção é a das emoções básicas de Ekman (1987), a partir da perspectiva das expressões. Nesta perspectiva, surgiu o FACS (*Facial Action Coding System*), um modelo que caracteriza as expressões faciais causadas pelas

contrações musculares do rosto, chamadas de AUs (*Action Units*). O FACS tem sido amplamente usado em trabalhos de detecção de emoções básicas em fotos e vídeos (OLIVEIRA; JAQUES, 2008).

Quadro 1 – Componentes de emoção no modelo multicomponencial

Componente de emoção	Função	Subsistema organísmico
Componente cognitivo ( <i>appraisal</i> )	Avaliação de objetos e eventos	Processamento de informação (sistema nervoso central)
Componente neurofisiológico (sintomas corporais)	Regulação de sistema	Suporte (sistema nervoso central, sistema neuroendócrino, sistema nervoso autônomo)
Componente motivacional (tendências de ação)	Preparação e direção de ação	Executivo (sistema nervoso central)
Componente de expressão motora (expressão vocal e facial)	Comunicação da reação e intenção comportamental	Ação (sistema nervoso somático)
Componente de sentimento subjetivo (experiência emocional)	Monitoramento do estado interno e da interação do organismo com o ambiente	Monitor (sistema nervoso central)

Fonte: Adaptado de (SCHERER, 2005).

A Tese utiliza, como teoria subjacente sobre emoções, uma perspectiva para uma teoria de emoções, desenvolvida pelo pesquisador suíço Klaus Scherer, no qual diferentes perspectivas sobre as emoções são integradas em um **modelo multicomponencial de emoções** (MOORS et al., 2013; SCHERER, 2005). A teoria de Scherer tem sido amplamente aceita pelos pesquisadores em psicologia cognitiva e computação afetiva. No seu modelo de componentes de emoções, Scherer apresenta uma definição para emoção como sendo **um episódio de mudanças sincronizadas e interrelacionadas nos estados de todos ou quase todos os cinco subsistemas do organismo em resposta à avaliação de um evento de estímulo interno ou externo de grande importância para o organismo** (SCHERER, 2005). O Quadro 1 apresenta os relacionamentos entre os componentes de emoção, suas funções emocionais e os subsistemas organísmicos. Neste modelo, o sentimento não pode ser confundido com a emoção, em si, pois ele é apenas um dos componentes de emoção, um componente de experiência emocional subjetiva (SCHERER, 2005).

Nesta perspectiva multicomponencial, Scherer (2005) descreve algumas características de emoções que permitem discernir a emoção de outros estados afetivos.

**Evento foco** - A teoria de Scherer (2005) sugere que as emoções são geralmente provocadas por eventos de estímulo. O que o autor quer dizer é que algo (evento) acontece para o organismo que estimula ou desencadeia uma resposta após a significância do evento ter sido

avaliada pelo indivíduo. Estes eventos podem ser externos ao organismo, como fenômenos naturais (trovoadas, ventanias, entre outros), o comportamento de outras pessoas, ou o comportamento da própria pessoa. Eventos internos também são provocadores de emoções, como mudanças fisiológicas ou memórias que surgem na mente do indivíduo.

**Dirigido à *appraisal*** - Um aspecto central da definição multicomponencial de Scherer (2005) é que o evento provocador da emoção e suas consequências deve ser relevante para as principais preocupações do organismo. A teoria sugere que diferentes emoções são produzidas por avaliações (*appraisals*) de eventos de estímulos internos ou externos. São estas avaliações sobre a relevância do evento de estímulo para o bem-estar pessoal que Scherer chama de *appraisals*. Discussões sobre o *appraisal* de emoções são ampliadas na Seção 2.2. A Seção 2.4.1 descreve os *appraisals* que levam à confusão.

**Sincronização de resposta** – Segundo Scherer (2005), se as emoções ocasionam respostas a eventos de estímulo, os padrões de resposta nos sistemas orgânicos devem corresponder à análise de *appraisal* das implicações presumidas do evento. Dependendo do evento provocador da emoção, todos ou quase todos os subsistemas orgânicos devem contribuir para a preparação da resposta.

**Rapidez de mudança** – Os eventos de estímulo e, conseqüentemente, o *appraisal* sobre esses eventos, pode mudar rapidamente, considerando mudanças no ambiente ou, até mesmo, reavaliações (*reappraisals*) dos eventos de estímulo (SCHERER, 2005).

**Impacto comportamental** – As emoções têm forte efeito sobre o comportamento conseqüente dessas emoções, e frequentemente interrompem seqüências de comportamento em andamento. Além disso, o componente de expressão motora do modelo de Scherer (2005) tem forte impacto na comunicação, podendo ter fortes conseqüências nas interações sociais.

**Intensidade** – Scherer (2005) sugere que a intensidade dos padrões de resposta das emoções e da experiência emocional são relativamente altas, e que esta é uma importante característica para distinguir emoções de outros estados afetivos, como o humor, por exemplo.

**Duração** – Uma discussão de Scherer (2005) sugere que a duração das emoções deve ser relativamente curta, em função de apresentarem uma forte resposta e sincronização dos subsistemas orgânicos que pode sobrecarregá-los. O autor utiliza como exemplo o estado afetivo de humor, que possui baixa intensidade e pouco impacto sobre o comportamento, de forma que pode ser mantido por um período mais longo até apresentar efeitos adversos.

Scherer (2005) defende que estas sete características são importantes para discernir as emoções de outros estados afetivos como as atitudes, o humor, as disposições afetivas e as posturas interpessoais. Um exemplo é o humor, que é caracterizado por ser mais duradouro,

podendo durar horas ou dias, mas possui baixa intensidade e pode emergir sem uma causa aparente. As emoções, por sua vez, são mais curtas, possuem geralmente intensidade mais alta e precisam de algum evento de estímulo.

## **2.2 Appraisal de emoções**

Um dos componentes de emoção presente na teoria de componentes de Scherer, e que tem sido fortemente considerado nos estudos dos pesquisadores em psicologia cognitiva é o componente cognitivo das emoções, responsável pelo *appraisal*, nome dado a uma avaliação cognitiva de situações e eventos que precedem uma emoção (SCHERER, 1999). *Appraisal* é um processo que avalia a significância do ambiente para o bem-estar pessoal. Significância para o bem-estar pode ser compreendida como a satisfação ou a não satisfação de necessidades pessoais, metas, crenças ou tudo aquilo com que o indivíduo se importa (MOORS et al., 2013). Scherer descreve que o *appraisal* pode ser intrínseco ou extrínseco (SCHERER, 2005). O *appraisal* é intrínseco quando ele avalia a característica de um objeto ou pessoa independente das metas ou necessidades do indivíduo, mas baseado em preferências genéticas ou aprendidas. O *appraisal* extrínseco avalia um evento de estímulo em relação a quanto ele vai de encontro com as necessidades, desejos e metas pessoais.

Segundo Moors e Scherer (2013), o termo *appraisal* foi usado pela primeira vez nos estudos de Magda Arnold nos anos 1960 (ARNOLD, 1960), mas foi o trabalho de Lazarus (1966) que teve influência direta para o desenvolvimento de uma teoria de *appraisal*. Nas décadas seguintes, trabalhos que deram continuidade ao desenvolvimento deste conceito passaram a ser referenciados como teorias de *appraisal* (BARRETT, 2006; ORTONY; CLORE; COLLINS, 1988; RUSSELL, 2003; SCHERER, 2009).

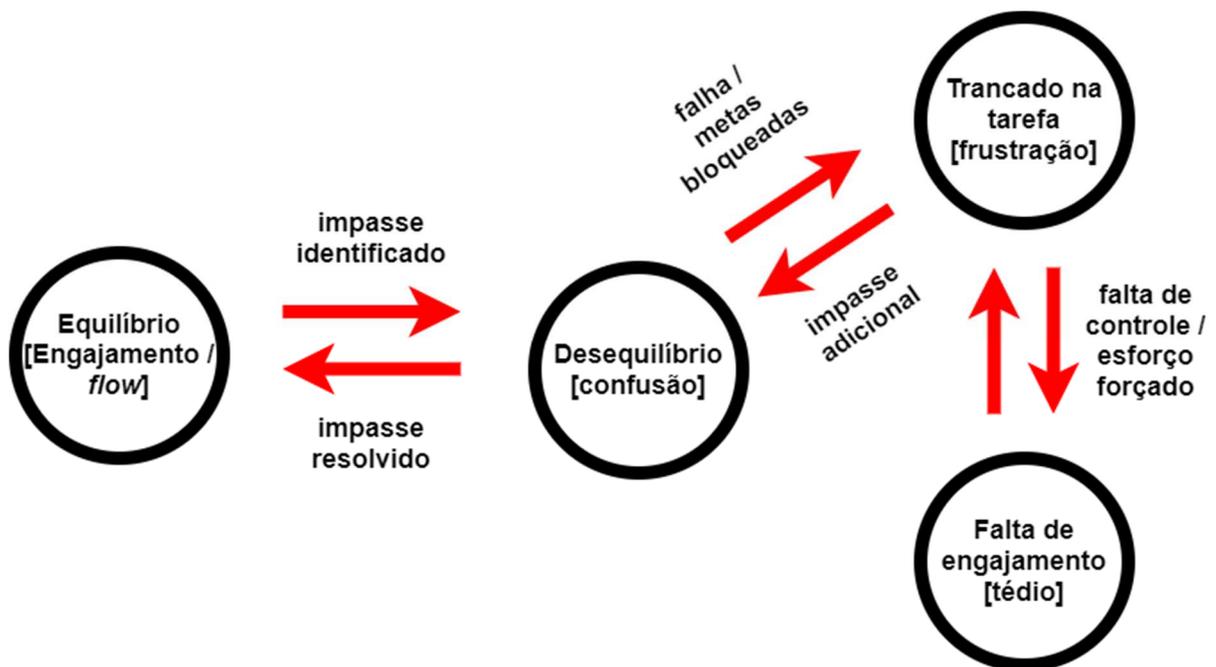
Diferentes emoções são produzidas por diferentes *appraisals* de emoções (SCHERER, 2005). Uma das teorias mais consideradas nos estudos de emoções envolvendo computadores é o modelo OCC (ORTONY; CLORE; COLLINS, 1988), que assume que as emoções surgem como resultado da forma como o indivíduo avalia as situações. O modelo OCC resulta em um modelo psicológico que explica a origem de 22 tipos de emoções.

Esta tese utiliza o conceito de *appraisal* no contexto da teoria multicomponencial das emoções, desenvolvida por Scherer (2005), amplamente aceita na comunidade de psicologia educacional. Moors e Scherer (2013) utilizam o termo *appraisal*, no singular, para referenciar o processo de *appraisal*, enquanto o termo *appraisals*, no plural, é utilizado para referenciar as saídas desse processo.

### 2.3 Dinâmica das emoções na aprendizagem

Os pesquisadores têm se preocupado em entender a dinâmica temporal das emoções na aprendizagem. O trabalho de D’Mello e Graesser (2012a) avaliou um modelo sobre as dinâmicas afetivas (ver Figura 1) em atividades de aprendizagem complexa. Segundo o modelo, quando o aluno está em um estado de engajamento / *flow*<sup>10</sup> e enfrenta impasses, como contradições, incongruências e obstáculos para alcançar suas metas, ele poderá experimentar um desequilíbrio cognitivo que resultará em confusão. Quando o aluno consegue resolver esses impasses e o equilíbrio é restaurado, o aluno retorna a um estado de engajamento / *flow*. No entanto, se o aluno falha para restaurar o equilíbrio e fica bloqueado para alcançar suas metas, o estudante entra em um estado de frustração que, se não resolvida, pode levá-lo a um estado de tédio.

Figura 1 – Modelo de dinâmicas afetivas de emoções relacionadas à aprendizagem



Fonte: Traduzido de (D’MELLO; GRAESSER, 2012a).

Sobre características da confusão na dinâmica das emoções, um estudo encontrou que alunos com pouco conhecimento em álgebra e com alto nível de neuroticismo<sup>11</sup> não

<sup>10</sup> *Flow* é um estado de concentração profunda de uma pessoa em uma atividade (CSIKSZENTMIHALYI; CSIKSZENTMIHALY, 1990).

<sup>11</sup> De acordo com o modelo “*Big Five*” de fatores de personalidade, são exemplos de traços associados aos neuroticismo ser ansioso, depressivo, envergonhado, emocional, preocupado e inseguro (BARRICK; MOUNT, 1991).

conseguiram lidar muito bem com a confusão, passando menos tempo neste estado afetivo e entrando mais rapidamente para os estados de frustração e tédio (REIS et al., 2018). O tempo com que o aluno permanece na confusão também pode estar relacionado com suas habilidades metacognitivas autorregulatórias. Quando alunos com boas habilidades autorregulatórias entram em confusão, eles podem demorar mais tempo para passarem para os estados de tédio e frustração, ou, até mesmo, não chegarem a estes estados. Isso porque estes alunos conseguem mobilizar mais estratégias para resolver os impasses e contradições encontradas em uma tarefa de aprendizagem (DI LEO; MUIS, 2020).

## 2.4 Confusão

Da mesma forma que não há um consenso entre os pesquisadores sobre o que é uma emoção (ver Seção 2.1), a confusão também tem recebido diferentes conceitualizações (D'MELLO; GRAESSER, 2014b). A confusão já foi classificada como uma emoção genuína (ROZIN; COHEN, 2003), uma emoção relacionada ao conhecimento (SILVIA, 2010), uma emoção epistêmica (PEKRUN; STEPHENS, 2012) e uma “não emoção” (HESS, 2003; KELTNER; SHIOTA, 2003).

O trabalho de D’Mello e Graesser (2014b) apresenta uma discussão interessante sobre o que é a confusão. Segundo os autores, a falta de uma clara definição de emoção, as múltiplas perspectivas sobre o construto de emoção e a escassez de pesquisa básica sobre a confusão, leva à falta de uma definição mais precisa sobre a confusão e se a confusão é uma emoção. Com o objetivo de encontrar uma correspondência entre o que se sabe sobre a confusão e as características de uma emoção, os autores sugerem iniciar a discussão a partir de um estudo de Carrol Izard (2010), que buscou identificar as características que definem uma emoção. Este estudo solicitou respostas a 37 pesquisadores de emoções para seis questões, entre elas, “o que é uma emoção”. O resultado do estudo descreve a emoção como um estado afetivo que envolve seis características, a saber: a) circuitos neurais parcialmente dedicados ao processamento emocional, b) ativação de sistemas de respostas em preparação para alguma ação, c) ter estados de sentimentos distintos, d) desempenhar um papel no comportamento expressivo, e) surgir a partir de resultados de processos de *appraisal*, e f) poder envolver interpretação cognitiva de sentimentos. Segundo D’Mello e Graesser (2014b), a confusão compartilha pelo menos quatro (*a*, *c*, *d* e *e*) das seis principais características de emoção identificadas pelos pesquisadores. Em relação ao item *a*, já foram encontradas evidências de que a confusão está ligada a atividades neurais, como no trabalho descrito por Halgren et al. (2002), que encontrou um alto valor para

N400<sup>12</sup> em atividades de eletroencefalograma, quando um indivíduo foi confrontado com uma sentença que apresentou uma anomalia semântica, como “Maria estava com fome, então ela foi até uma loja comprar algumas cadeiras”. Em relação a *c*, já foi descrito um estado de sentimento distinto relacionado a confusão (ROZIN; COHEN, 2003). Em relação a *d*, já foi encontrado um componente expressivo para a confusão, a sobrancelha franzida (CRAIG et al., 2008a; GRAFSGAARD; BOYER; LESTER, 2011; MCDANIEL et al., 2007). Em relação ao item *e*, foram encontradas evidências de que a confusão surge a partir de um processo de *appraisal* cognitivo relacionado a uma falta de correspondência entre a informação que chega ao indivíduo e o seu conhecimento anterior relativo à informação (D’MELLO et al., 2014; SILVIA, 2010). Sobre as demais características de emoções apontadas por Izard (IZARD, 2010), não se sabe ainda muito sobre mudanças corporais associadas com a confusão (característica *b*), mas o trabalho apresentado Alzoubi et al. (2012) conseguiu algum sucesso em distinguir a confusão do estado neutro e outras emoções, a partir de dados fisiológicos. Assim, a Tese qualifica a confusão como uma emoção, acompanhando o mesmo entendimento de D’Mello e Graesser (2014b), que consideraram evidências científicas relacionadas a características vinculadas a emoções, como ser um estado afetivo que surge de interações neurais, que tem um componente expressivo, que tem um estado de sentimento distinto e que está relacionado a um componente cognitivo de *appraisal*.

#### 2.4.1 O *appraisal* da confusão

Ao ser caracterizada como uma emoção epistêmica (PEKRUN; STEPHENS, 2012), por ser causada pelo processamento cognitivo de informação como parte de uma tarefa de aprendizagem (ARGUEL et al., 2019a), e uma emoção relacionada ao conhecimento (SILVIA, 2010), é importante para a Tese destacar a relação desta emoção com o conhecimento do indivíduo, para fundamentar a hipótese da Tese. Como já descrito na Seção 2.2, o *appraisal* é um processo interno de avaliação de um indivíduo sobre como um evento de estímulo se relaciona com as próprias metas, valores, conhecimentos e habilidades. O *appraisal* é o componente cognitivo do modelo multicomponencial de emoções de Scherer (1999). O trabalho de Silvia (2010) descreve emoções relacionadas ao conhecimento como emoções causadas pelas crenças dos indivíduos sobre seus próprios pensamentos e conhecimentos e decorrem de objetivos associados ao aprendizado, como o conhecimento, o pensamento e a

---

<sup>12</sup> Sinal eletrofisiológico relacionado a uma informação semanticamente anômala (KUTAS; HILLYARD, 1980).

compreensão. Seu trabalho encontrou evidências positivas para a hipótese de que a confusão está associada com *appraisals* de novidade (familiar) e de baixa compreensibilidade, no contexto da informação que chega para o indivíduo. Neste sentido, a teoria de discrepância de Macdowell e Mandler (1989) descreve que os indivíduos estão constantemente assimilando novas informações em um modelo mental de conhecimento. Quando uma informação discrepante é identificada, como um conflito com o conhecimento prévio, ocorre uma excitação no sistema nervoso autônomo, e o indivíduo experimenta uma variedade de possíveis emoções. A emoção experimentada depende do contexto e do processo de *appraisal*. O trabalho de Silvia (2010) encontrou que a surpresa pode preceder a confusão quando um estímulo inesperado é avaliado no processo de *appraisal* como algo incompreensível. Com isso, os estudos que consideram o componente emocional de *appraisal* descrevem a confusão como uma emoção que ocorre quando há uma falta de correspondência entre a informação que chega e o conhecimento anterior, ou quando essa informação não pode ser integrada aos modelos mentais existentes, iniciando um desequilíbrio cognitivo (D'MELLO; GRAESSER, 2014b).

#### 2.4.2 As expressões de confusão

Os pesquisadores têm utilizado duas abordagens para identificar componentes expressivos de confusão: **guiada a teoria** (*theory-guided*) e **dirigida a dados** (*data-driven*) (D'MELLO; GRAESSER, 2014b). A **abordagem guiada a teoria** se foca em expressões ou ações que possuem alguma base em teorias de emoções. Um exemplo são os trabalhos que consideram expressões faciais para identificar a confusão, a partir das justificativas teóricas sobre a relação das expressões faciais e as emoções. A segunda abordagem, **dirigida a dados**, utiliza técnicas de mineração de dados para computar grandes conjuntos de dados para tentar encontrar correlações entre dados e julgamentos de confusão, como autorrelatos e anotações de emoções por anotadores externos.

Segundo Silvia (2010), a primeira discussão sobre expressões faciais da confusão foi realizada por Darwin (1872), no contexto de obstáculos para alcançar metas cognitivas. Darwin descreve o seguinte: “Um homem pode ser absorvido pelo pensamento mais profundo, e sua sobrancelha permanecerá serena até encontrar algum obstáculo em sua linha de raciocínio, ou ser interrompido por algum distúrbio, e então uma carranca passa como uma sombra sobre sua sobrancelha.”. Utilizando uma abordagem guiada à teoria, o trabalho de Craig et al. (2008)

encontrou uma correlação da confusão com as *action units*<sup>13</sup> (AUs) AU4 (sobrancelha abaixada) (Figura 2 *a, b e c*), AU7 (pálpebras esticadas), AU4 + AU7 (Figura 2 *d*) e AU12 (canto dos lábios puxados), do sistema FACS (*Facial Acting Coding System*). O trabalho de McDaniel et al. (2007) também encontrou uma correlação da confusão com AU4.

Figura 2 - Expressões faciais de confusão



Fonte: (GRAFSGAARD; BOYER; LESTER, 2011; MCDANIEL et al., 2007).

Em relação a outros modos de expressão, estudos já encontraram evidências de correlação da confusão com a postura do corpo (D'MELLO; DALE; GRAESSER, 2012; K. D'MELLO; GRAESSER, 2010), mas esta forma de expressão necessita de mais investigações. Em relação a dados fisiológicos, a classificação da confusão como uma emoção relacionada ao conhecimento pode trazer o questionamento se realmente existem características fisiológicas com boa correlação com a confusão. O estudo descrito em (ALZOUBI; D'MELLO; CALVO, 2012) tentou encontrar estas características, como na atividade elétrica do coração (eletrocardiograma), para diferenciar emoções como confusão, curiosidade e o estado neutro. Os resultados foram apenas moderados. Também é bastante conhecido que conteúdo emocional pode ser expressado na linguagem, como nos trabalhos sobre análise de sentimentos em textos (PANG; LEE, 2008). Estudos já mostraram que respostas textuais de alunos e respostas em conversações podem carregar padrões que transmitem o estado de confusão (D'MELLO et al., 2008; D'MELLO; GRAESSER, 2014b).

#### 2.4.3 As facetas da confusão para resultados de aprendizagem

Apesar de a confusão ser comumente descrita como uma emoção negativa para o processo de aprendizagem, ela não deveria ser evitada, pois a partir dela o aluno pode engajar esforços para manter foco e atenção e buscar o conhecimento que lhe falta em materiais ou através da ajuda de colegas e professores (D'MELLO; GRAESSER, 2012b). Trabalhos já

<sup>13</sup> Expressões faciais causadas por contrações musculares, catalogadas no sistema FACS (*Facial Acting Coding System*) (EKMAN; FRIESEN, 1987).

encontraram uma correlação da confusão com a aprendizagem (CRAIG et al., 2004; D'MELLO; GRAESSER; PICARD, 2007), inclusive em ambientes de programação de computadores (LEE et al., 2011). O estudo de Graesser et al. (2007) encontrou que a confusão foi a única emoção entre as investigadas (incluíram as demais emoções relacionadas a aprendizagem: frustração, tédio e engajamento) que apresentou uma correlação positiva com os resultados de testes de conhecimento, de forma que a confusão foi um bom preditor do sucesso ou fracasso da aprendizagem e um preditor de conhecimento. Outros trabalhos também já encontraram evidências de uma correlação da confusão com a falta de conhecimento (D'MELLO; PERSON; LEHMAN, 2009; LEE et al., 2011; RODRIGO; BAKER; NABOS, 2010). Os estudos descritos na Seção 2.3 mostraram que quando o aluno passa muito tempo no estado de confusão, seu estado emocional pode mudar para frustração e tédio, emoções consideradas negativas para o processo de aprendizagem (D'MELLO; GRAESSER, 2012a). No entanto, quando resolvida a tempo, a confusão pode levar o aluno a adquirir níveis mais profundos de conhecimento (ARGUEL et al., 2019b; D'MELLO et al., 2014; LEHMAN et al., 2013; LEHMANN; GRAESSER, 2015), inclusive em ambientes de programação de computadores (LEE et al., 2011).

## **2.5 Confusão em atividades de programação de computadores**

O trabalho apresentado por Kinnunen e Simon (2010) coletou relatos de estudantes de uma disciplina introdutória de programação de uma universidade, com o objetivo de entender as experiências emocionais envolvidas. O estudo encontrou que quando o aluno se depara pela primeira vez com enunciados de tarefas de programação, a falta de entendimento pode levá-lo de uma completa desorientação a um estado de confusão. Neste momento, o aluno pode se sentir travado, sem saber o que fazer. Os alunos que conseguiram obter um melhor entendimento dos enunciados foram aqueles que leram os enunciados mais de uma vez e discutiram as instruções das tarefas com um colega. Quando passa por essa dificuldade inicial, em que começa a entender o problema, o aluno pode ainda não saber como prosseguir ou onde começar a programar. Isso pode levá-lo novamente a um estado de confusão. Nesta situação, os alunos que obtiveram sucesso nas tarefas usaram estratégias como procurar por exemplos em livros, fazer pesquisa na internet ou começar a escrever algum código, mesmo sem saber exatamente o que estava fazendo. Na medida que os alunos foram programando e cometendo os primeiros erros em sequência, foram observadas fortes emoções como confusão, frustração e raiva. Estes alunos também passaram a refletir e a questionar sobre a própria capacidade de

resolverem problemas (autoeficácia). Estes relatos mostram que o contexto de tarefas de programação de computadores é fortemente emocional e propício para o desequilíbrio cognitivo e a confusão.

Estudos já mostraram que a confusão pode ser prejudicial, quando não resolvida a tempo, mas também pode ser benéfica, levando a níveis mais profundos de aprendizagem, quando os impasses que levam ao desequilíbrio cognitivo são resolvidos (D'MELLO et al., 2014; LEHMANN; GRAESSER, 2015). Estes resultados também parecem ser verdadeiros no contexto de programação de computadores. O trabalho de Lee et al. (2011) encontrou que a confusão prolongada em atividades de programação levou a uma piora de desempenho, mas quando resolvida, levou a resultados melhores do que se o aluno não tivesse estado confuso.

Sobre as transições emocionais no desenvolvimento de algoritmos, o trabalho de Bosch e D'Mello (2017) testou o modelo de dinâmicas afetivas (ver Seção 2.3) no contexto de tarefas de programação de computadores com novatos (D'MELLO; GRAESSER, 2012a). Foram encontradas as mesmas transições do modelo de comparação. Segundo os autores (BOSCH; D'MELLO, 2017), os impasses que levam à confusão costumam surgir quando os alunos se deparam com problemas não familiares, erros de sintaxe e saídas não esperadas dos algoritmos. Estes impasses podem, então, levar à frustração, que pode transitar para um estado de tédio.

Em relação ao comportamento dos alunos em tarefas de programação de computadores, um trabalho que identificou a confusão, o engajamento, a frustração e o tédio como as emoções mais proeminentes neste contexto, encontrou que quando o estudante está confuso, ele fica mais tempo ocioso no sistema (BOSCH; D'MELLO; MILLS, 2013).



### **3 DETECÇÃO DE CONFUSÃO EM AMBIENTES COMPUTACIONAIS DE APRENDIZAGEM DE PROGRAMAÇÃO DE COMPUTADORES**

O principal objetivo deste capítulo é apresentar os trabalhos relacionados ao tema da Tese, ou seja, trabalhos que apresentaram e avaliaram **modelos computacionais de detecção livre de sensores da confusão de alunos enquanto utilizam computadores para a realização de tarefas de aprendizagem de programação de computadores.**

Antes de apresentar os trabalhos relacionados, a Seção 3.1 apresenta uma fundamentação teórica sobre detecção de emoções em ambientes computacionais de aprendizagem. A Seção 3.2 traz discussões sobre as formas utilizadas na literatura para a obtenção de rótulos de emoções. Estes rótulos são importantes no contexto da Tese, pois a principal abordagem computacional utilizada pelos trabalhos relacionados para a detecção de emoções é a de aprendizagem de máquina supervisionada. Estes modelos são treinados com exemplos de dados sobre aspectos comportamentais dos alunos, vinculados a rótulos de emoções (KUBAT, 2017). A Seção 3.3 apresenta algumas discussões comparativas sobre detectores baseados em sensores e detectores livres de sensores. A Seção 3.4 apresenta os trabalhos relacionados. São apresentados também o método de busca e seleção dos trabalhos e uma discussão sobre eles.

#### **3.1 Detecção de emoções em ambientes computacionais de aprendizagem**

O principal objetivo das pesquisas em computação afetiva é descobrir formas de tornar os computadores habilidosos em reconhecer os estados afetivos (como as emoções) dos seus usuários, em responder apropriadamente a estes estados e em expressar afetividade (PICARD, 1995). No entanto, detectar e responder às emoções humanas são atividades complexas e desafiadoras que vão além da ciência da computação, envolvendo também campos de pesquisa em neurociência, fisiologia e psicologia (PICARD, 2010).

A detecção de estados afetivos é um tipo de problema de reconhecimento de padrões que envolve o desenvolvimento de detectores de fenômenos mal definidos, como os estados afetivos, a partir de sinais observáveis (D'MELLO; KORY, 2015). Este é um tipo de problema complexo e desafiador porque os estados afetivos são construtos psicológicos que não podem ser observados diretamente. Estes construtos são observados indiretamente, através de expressões de comportamentos que são sensíveis ao contexto, a um mascaramento social e a

variabilidade cultural e individual (D’MELLO; KORY, 2015; RUSSELL, 1994; RUSSELL; BACHOROWSKI; FERNÁNDEZ-DOLS, 2003).

Os pesquisadores têm usado dados de interação de alunos em ambientes computacionais de ensino para estudar como estes processos afetivos se desenrolam durante a aprendizagem envolvendo tecnologias (KOEDINGER et al., 2015; WINNE; BAKER, 2013). Os pesquisadores costumam usar algoritmos de aprendizado de máquina supervisionada para modelar os estados afetivos durante a aprendizagem (KOEDINGER et al., 2015). Estes modelos de aprendizado de máquina supervisionada são treinados com amostras de dados formadas por atributos de dados<sup>1</sup> (dados sobre expressões faciais, interações com o ambiente computacional, entre outros) vinculados a rótulos de estados afetivos, como as emoções (“confuso”, “frustrado”, “entediado”, etc), fornecidos (os rótulos) através de julgamentos humanos. A partir da visualização de amostras de exemplos, vinculados aos rótulos, os algoritmos modelam (aprendem) relacionamentos entre os atributos de dados e os rótulos de estados afetivos. Após treinados, os modelos visualizam amostras de exemplos (sem os rótulos) não usadas no treinamento e tentam identificar (detectar) os rótulos de estados afetivos das amostras (BAKER, 2010), ou seja, conseguem identificar os estados afetivos sobre novos dados dos estudantes, não vistos anteriormente no treino.

Diferentes tipos de atributos de dados têm sido considerados no treinamento de modelos de detecção de emoções. D’Mello e Graesser (2009) investigaram a qualidade preditiva de atributos de dados sobre a posição e a inquietação corporal durante interações de alunos com o sistema tutor inteligente AutoTutor. Os rótulos afetivos foram obtidos de três fontes: anotações do próprio aprendiz, de um colega e de dois juízes treinados. Bosch et al. (2016) utilizaram amostras de atributos de dados sobre expressões faciais e padrões de interação dos alunos em um jogo educacional de Física para treinar a detecção das emoções tédio, confusão, engajamento / *flow*, frustração, felicidade / prazer e ansiedade. Os rótulos de emoções foram obtidos através da aplicação do protocolo BROMP (*Baker Rodrigo Ocumpaugh Monitoring Protocol*) (OCUMPAUGH; BAKER; OCUMPAUGH, J., & BAKER, 2014), no qual juízes treinados observam alunos durante a interação com o ambiente computacional, em salas com computadores, e fazem julgamentos sobre os estados afetivos dos estudantes. Os pesquisadores também têm utilizado abordagens multimodais para detectores afetivos, como o trabalho de D’Mello e Graesser (2010), que utilizou atributos de dados sobre expressões faciais, dados conversacionais e linguagem corporal, obtidos em sessões de aprendizagem com o sistema tutor

---

<sup>1</sup> Do inglês *feature*.

inteligente AutoTutor. Os rótulos de emoções foram obtidos através de dois juízes treinados que observaram vídeos do rosto e da tela dos alunos nas interações com o AutoTutor.

Os trabalhos citados anteriormente utilizaram atributos de dados sobre expressões faciais, posição e inquietação corporal, linguagem corporal, dados de conversação e de interação com o sistema. A literatura apresenta ainda trabalhos que utilizaram outras modalidades de atributos de dados, como dados vocais, fisiológicos, textuais, provenientes de imagens cerebrais, além de abordagens que consideraram múltiplas modalidades (CALVO; D'MELLO, 2010). Sobre os rótulos de emoções, a próxima seção apresenta discussões sobre como os trabalhos têm obtido rótulos para as amostras dos modelos de aprendizagem de máquina supervisionada.

### **3.2 Obtenção de rótulos de emoções para aprendizagem de máquina supervisionada**

Para que algoritmos de aprendizado de máquina supervisionada tenham bom desempenho em identificar emoções dos estudantes em ambientes computacionais de aprendizagem, eles precisam ser treinados com rótulos confiáveis de emoções. Os trabalhos que apresentaram detectores de emoções de estudantes, integrados em ambientes computacionais de aprendizagem, usaram os seguintes métodos para obter rótulos para modelos de aprendizagem de máquina: métodos baseados na observação dos alunos por juízes humanos diretamente no ambiente de aprendizagem (OCUMPAUGH; BAKER; OCUMPAUGH, J., & BAKER, 2014; WOOLF et al., 2009) ou através de vídeos (BIXLER; D'MELLO, 2013; CRAIG et al., 2008c; MOTA; PICARD, 2003); métodos de autorrelato dos próprios estudantes (BIXLER; D'MELLO, 2013; BOSCH; D'MELLO; MILLS, 2013; CRAIG et al., 2008c; WOOLF et al., 2009); e métodos baseados em anotação de arquivos de *logs* (LEE et al., 2011).

Os métodos de autorrelato encontrados na literatura podem ser métodos de julgamento retrospectivo, após a sessão de aprendizagem (BIXLER; D'MELLO, 2013; BOSCH; D'MELLO; MILLS, 2013) ou concorrente, durante a sessão de aprendizagem (CRAIG et al., 2008c). Nos métodos concorrentes, os alunos relatam suas próprias emoções durante as tarefas de aprendizagem, como no método *emote-aloud*, no qual o aluno relata suas emoções verbalmente (PORAYSKA-POMSTA et al., 2013). Estes autorrelatos podem ser realizados pelos alunos como respostas livres, quando são livres para falar sobre suas emoções enquanto realizam as tarefas (MELLO; SCIENCE; HALL, 2006), ou através de relato forçado, como quando o sistema abre uma janela em determinado momento ou de tempos em tempos, solicitando que o aluno relate sua emoção (CONATI; MACLAREN, 2009). O problema do

relato forçado é que exige do aluno um relato de emoção mesmo quando ele não se sente confortável para isso. Nestes casos, o aluno pode simplesmente negligenciar a janela de relato forçado, relatar incorretamente sua emoção para se livrar da atividade de relato ou, ainda pior, ficar irritado com o sistema. Os autorrelatos também podem ser retrospectivos, quando os alunos participam das tarefas de aprendizagem em um primeiro momento, para só depois relatar suas emoções, como nos estudos em que os estudantes assistiram seus vídeos de aprendizagem e relataram retrospectivamente suas emoções (MILLS; D’MELLO, 2012).

A qualidade dos rótulos obtidos através de métodos de autorrelato está relacionada com alguns fatores, como a idade dos participantes (CONATI; MACLAREN, 2009; SAYFAN; LAGATTUTA, 2008), aspectos culturais (ELFENBEIN; AMBADY, 2003) e habilidades metacognitivas (PORAYSKA-POMSTA et al., 2013). Além disso, autorrelatos concorrentes possuem algumas desvantagens (PORAYSKA-POMSTA et al., 2013): aumentam a carga cognitiva dos alunos; interferem na tarefa primária de aprendizagem; o aluno pode se sentir desconfortável em demonstrar suas fraquezas; podem influenciar as emoções do aluno durante a tarefa de aprendizagem; requerem um engajamento extra para a tarefa primária de aprendizagem e a tarefa de relato das emoções; e a subjetividade dos relatos. Os autorrelatos retrospectivos, também possuem suas desvantagens (PORAYSKA-POMSTA et al., 2013): a distância de tempo entre a tarefa e o autorrelato; requerem tempo extra dos alunos para participação na pesquisa (tempo do aluno nas tarefas de aprendizagem mais o tempo de assistir sua aprendizagem e fazer os autorrelatos); requerem habilidades metacognitivas dos alunos; e a subjetividade dos relatos.

Foi devido às restrições nos métodos de autorrelato de emoções que surgiram protocolos que usam juízes humanos treinados em relatar as emoções dos alunos (OCUMPAUGH; BAKER; OCUMPAUGH, J., & BAKER, 2014). Os juízes podem anotar as emoções dos alunos durante as sessões de aprendizagem (observação *in loco*) ou retrospectivamente, assistindo a vídeos com os rostos e ações dos alunos. No entanto, julgar as emoções dos alunos apenas pela observação do rosto dos estudantes pode ser uma tarefa desafiadora. As expressões faciais de algumas emoções costumam ser mais sutis, como na confusão e tédio, devido a conotações sociais negativas associadas a expressões de emoções, o que pode levar os estudantes a controlar as expressões de suas emoções (D’MELLO; GRAESSER, 2010; OCUMPAUGH; BAKER; OCUMPAUGH, J., & BAKER, 2014).

No método de observação de emoções através de arquivos de *logs*, os anotadores criam rótulos de emoções através da anotação de amostras de *log* assistidas retrospectivamente (LEE et al., 2011). O problema deste método é que pode ser difícil encontrar sinais de emoções apenas

através da observação de registros de *log*, não sendo suficiente para a tomada de decisão das anotações. Métodos multimodais poderiam ser necessários para um anotador humano observar diferentes expressões associadas com determinadas emoções para a tomada de decisão de anotação e obtenção de rótulos (D’MELLO; GRAESSER, 2010).

O autor da Tese, em conjunto com colegas do Programa de Pós-graduação em Computação Aplicada da Unisinos (PPGCA), implementaram um método de obtenção de rótulos de emoções, a partir de observações retrospectivas de vídeos por anotadores humanos (MORAIS et al., 2019). Os vídeos são compostos pelo rosto e pela tela dos alunos interagindo com o sistema. No entanto, em função da pandemia de COVID-19 e a impossibilidade de realização de aulas presenciais, não seria possível coletar os vídeos dos alunos para posterior anotação retrospectiva dos rótulos de emoções. Desta forma, a Tese utilizou o método de autorrelato de emoções, o qual será abordado na Seção 8.2.

### 3.3 Detecção de emoções baseada em sensores *versus* livre de sensores

A Seção 3.1 mostrou que modelos que buscaram detectar emoções em ambientes computacionais consideraram uma ampla variedade de modalidades de dados, inclusive abordagens multimodais. As diferentes modalidades de dados podem estar atreladas ainda a uma de duas outras classificações de detectores: **baseados em sensores** e **livres de sensores**. Os detectores baseados em sensores utilizam para a coleta de dados algum tipo de dispositivo periférico físico conectado a um computador, como: câmeras, em sensores de expressões faciais (BOSCH et al., 2015), de voz (ZENG et al., 2009) e de postura (D’MELLO; GRAESSER, 2010); sensores de condutância de pele (fisiológico) (SHU et al., 2018); e eletroencefalogramas (ERWIANDA et al., 2019; ZHOU et al., 2018). Os detectores livres de sensores utilizam dados derivados da interação do usuário com o ambiente computacional, como dados de digitação e de mouse (YANG et al., 2019).

Em comparação com os detectores livres de sensores, os detectores baseados em sensores são mais intrusivos (ARROYO et al., 2009; BOTELHO et al., 2018; MELLO et al., 2005) e com custo mais elevado, sendo mais difíceis de serem aplicados em escala (YANG et al., 2019) e em escolas com limitações econômicas. Sensores físicos também costumam ser mais difíceis de manterem a privacidade dos usuários (REYNOLDS; PICARD, 2004). Dados de interação, por sua vez, são mais fáceis de serem anonimizados. Outra dificuldade encontrada em sensores físicos, como os de expressões faciais e de posicionamento de cabeça, é que em salas com computadores, nem sempre os alunos ficam adequadamente posicionados em seus

lugares, como ocorreu no trabalho de Tiam-Lee e Sumi (2019), no contexto de programação de computadores.

### 3.4 Trabalhos relacionados: detectores de confusão em ambientes de programação

Esta seção apresenta os trabalhos relacionados da Tese. A Seção 3.4.1 descreve o método utilizado na busca e seleção dos trabalhos. A Seção 3.4.2 apresenta os trabalhos selecionados, e uma discussão é apresentada na Seção 3.4.3.

#### 3.4.1 Método de busca e seleção dos trabalhos relacionados

O Quadro 2 apresenta a relação dos trabalhos relacionados, obtidos através de buscas por artigos nas principais bases das áreas de psicologia educacional, informática na educação, interação humano-computador, mineração de dados educacionais, inteligência artificial e computação afetiva: *Springer*<sup>2</sup>, *Elsevier Science Direct*<sup>3</sup>, *IEEE Xplore*<sup>4</sup>, *ACM Digital Library*<sup>5</sup>, *Web of Science*<sup>6</sup>, *Scopus*<sup>7</sup>, *Taylor & Francis*<sup>8</sup> e *Google Scholar*<sup>9</sup>. As buscas foram realizadas nos meses de janeiro e fevereiro de 2019. Em março de 2022, as buscas foram realizadas novamente, para verificar a existência de trabalhos publicados entre 2019 e 2022.

Foram utilizadas diversas *strings* de busca com diferentes arranjos envolvendo os termos *predict*, *detect*, *confusion* e *programming*. As *strings* foram modificadas para se adequarem aos mecanismos de busca de cada base de publicações. Os termos das *strings* de busca foram definidos com o objetivo de encontrar trabalhos relacionados ao tema da tese, ou seja, trabalhos que apresentaram e avaliaram **modelos computacionais de detecção livre de sensores da confusão de alunos enquanto utilizam computadores para a realização de tarefas de aprendizagem de programação de computadores**.

A partir do primeiro resultado das buscas, foi realizada a leitura do título e do resumo de cada artigo, com o objetivo de identificar se cada documento atendia os critérios de inclusão e de exclusão. Nos casos em que havia dúvida sobre o atendimento dos critérios, foi realizada a leitura de outras partes do artigo, como as seções de introdução, resultados e conclusão. Para

<sup>2</sup> Pesquisa realizada em <https://link.springer.com/>

<sup>3</sup> Pesquisa realizada em <https://www.sciencedirect.com/>

<sup>4</sup> Pesquisa realizada em <https://ieeexplore.ieee.org/>

<sup>5</sup> Pesquisa realizada em <https://dl.acm.org/>

<sup>6</sup> Pesquisa realizada via Portal de Periódicos da Capes, em <http://periodicos.capes.gov.br>

<sup>7</sup> Pesquisa realizada via Portal de Periódicos da Capes, em <http://periodicos.capes.gov.br>

<sup>8</sup> Pesquisa realizada em <https://www.tandfonline.com/>

<sup>9</sup> Pesquisa realizada em <https://scholar.google.com/>

um artigo ser incluído na seleção, nenhum critério de exclusão deveria ser atendido e todos os critérios de inclusão deveriam ser atendidos. Os **critérios de exclusão** foram os seguintes: artigos não escritos em inglês; artigos duplicados; artigos que apresentam os mesmos resultados de avaliação de modelo de detecção de confusão (o artigo mais completo e mais recente é selecionado); trabalhos publicados como pôster, capítulo de livro, resumo, tese e relatório técnico; artigos que apresentam estudos secundários, como mapeamentos e revisões sistemáticas. Os **critérios de inclusão** foram os seguintes: artigos que apresentam um modelo de detecção de confusão em ambientes computacionais de programação de computadores; artigos que descrevem como o modelo de predição de confusão foi implementado, como os rótulos para modelos de aprendizagem de máquina supervisionada foram obtidos, quais dados de treinamento foram utilizados, como o treinamento e a validação foram realizados, e quais os resultados obtidos.

Além da busca nas bases de dados, foi realizado um processo recursivo de mapeamento de todos os trabalhos relacionados citados pelos estudos selecionados, que também passaram pelos critérios de exclusão e inclusão. Também foram mapeados estudos dos principais grupos de pesquisa que trabalham com afetividade em ambientes computacionais de aprendizagem, como os grupos de pesquisa dos pesquisadores Sidney D’Mello, Ryan Baker, Ivon Arroyo, Rosalind Picard, Jennifer Sabourin, Cristina Conati e Beverly Park Woolf e o programa de pesquisa sobre o Cognitive Tutor, da Universidade Carnegie Melon (ALEVEN et al., 2006).

Os critérios de inclusão e os termos usados nas *strings* de busca não especificaram a abordagem utilizada quanto aos sensores (com sensores ou livres de sensores), pois foi intenção do autor da Tese encontrar e relatar trabalhos que usassem as duas abordagens, ampliando a visão sobre os trabalhos. Então, primeiramente, na Seção 3.4.2, são apresentados trabalhos que utilizaram as duas abordagens: com sensores e sem sensores. A Seção 3.4.3 apresenta discussões específicas sobre os trabalhos que usaram a abordagem livre de sensores, os mais interessantes para a Tese.

Quadro 2 – Trabalhos relacionados

Referência	Emoções alvo	Dados usados no treinamento dos modelos	Formas de obtenção de rótulos de emoção	Algoritmos utilizados	Janelas de dados	É livre de sensores?	Ambiente	Alunos	Melhores resultados
(TIAM-LEE; SUMI, 2019)	Confusão, engajamento, frustração, tédio e neutro.	Digitação de teclas, compilação, postura da cabeça e dados faciais.	Anotação retrospectiva de emoções e ações, pelos próprios alunos.	<i>C4.5, SVM, MLP e Naive bayes.</i>	Obtidos em pontos fixos. Janelas de tempo variado, com média de 17,24 segundos.	Não.	Sistema tutor inteligente. Java.	73.	Acurácia de 0,77 e <i>Kappa</i> de 0,18.
(TIAM-LEE; SUMI, 2018)	Confusão e engajamento.	Expressões faciais (AUs do FACS), digitação de teclas (digitando ou não digitando), compilações (sucesso ou falha).	Anotação retrospectiva de emoções, pelos próprios alunos que assistiram <i>logs</i> e vídeos.	<i>HMM.</i>	Sequências de estados discretos (digitando, compilando, entre outros) de um rótulo de emoção.	Não.	Ambiente próprio. Java.	11 no treino e 35 no teste.	<i>Kappa</i> de 0,41.
(VEA; RODRIGO, 2017b)	Confusão, tédio e frustração.	Teclado (velocidade de digitação, quantidade de <i>backspaces</i> , entre outros) e mouse (quantidade de movimentos, <i>scrolls</i> , entre outros).	Anotação retrospectiva de emoções, por dois anotadores (os próprios autores do trabalho).	<i>J48, Decision Tree e Random Forest.</i>	Obtidos em pontos fixos. Janelas de tempo fixo, de 15 segundos.	Sim.	IDE Dev-C++.	55.	<i>Kappa</i> de 0,572 e precisão de 0,74.

Referência	Emoções alvo	Dados usados no treinamento dos modelos	Formas de obtenção de rótulos de emoção	Algoritmos utilizados	Tamanho das janelas de dados	É livre de sensores?	Ambiente	Alunos	Melhores resultados
(BOSCH; CHEN; D'MELLO, 2014)	Confusão, tédio, <i>flow</i> / engajamento, frustração e neutro.	AUs do FACS, orientação da cabeça, posição do nariz, tamanho do rosto e visibilidade de nariz, boca, olhos e ouvidos.	Anotação retrospectiva de emoções, pelos próprios alunos.	15 modelos, mas cita apenas um classificador logístico e <i>Naïve bayes</i> .	Obtidos em pontos não fixos. Janelas de tempo fixo, de 2 segundos para a confusão.	Não.	Ambiente próprio. <i>Python</i> .	99.	<i>Kappa</i> de 0,221.
(FELIPE et al., 2012)	Confusão e tédio.	Pressionamento de teclas (velocidade média, desvio padrão de digitação, tempo ocioso e total de <i>backspaces</i> ).	Anotação retrospectiva de emoções, por dois anotadores que não os alunos.	<i>J48</i> , <i>Random Tree</i> e <i>BF Tree</i> .	Obtidos em pontos fixos. Janelas de tempo fixo, de 20 segundos.	Sim.	IDE Visual Studio 6.0. C++.	6.	<i>Kappa</i> de 0,856.
(GRAFSGAARD; BOYER; LESTER, 2011)	AU4 (AU relacionada à confusão).	Diálogo textual, ações em tarefas e histórico de expressões faciais.	Anotação retrospectiva de AU4, por dois anotadores certificados FACS.	<i>HMM</i> e <i>OMM</i> .	Não explica como realizou.	Não.	Sistema tutor inteligente.	14.	Acurácia de 0,907.
(LEE et al., 2011)	Confusão.	Dados sobre compilações (com erro, sem erro, média de tempo entre compilações com e sem erros, entre outros).	Anotação retrospectiva de emoções, por dois anotadores, que não os alunos.	<i>J48 (C4.5)</i> .	Conjunto de 8 compilações.	Sim.	IDE <i>BlueJ</i> . Java.	149.	<i>Kappa</i> de 0,860.

Fonte: Elaborado pelo autor.

### 3.4.2 Trabalhos selecionados

Após a aplicação dos critérios de inclusão e exclusão, foram encontrados sete trabalhos relacionados, sendo quatro trabalhos com sensores e três trabalhos livres de sensores. O Quadro 2 apresenta um resumo dos estudos relacionados.

O trabalho de Tiam-Lee e Sumi (2018) descreve um ambiente computacional para prática de programação de computadores que modifica suas ações (complexidade dos exercícios atribuídos e atribuição de orientações) a partir da detecção da confusão do aluno. Um modelo HMM (*Hidden Markov Model*) foi construído com base em dados coletados de 11 alunos de uma turma de introdução à programação, em uma universidade japonesa. Os alunos já tinham alguma experiência de programação (em torno de 2 meses) e participaram de uma única sessão do estudo, com duração de 45 minutos. O ambiente de programação usado pelos alunos foi desenvolvido especialmente para o estudo. O ambiente atribuiu exercícios de programação com dificuldade crescente. Em cada exercício, os alunos deveriam escrever o corpo de uma função na linguagem Java. Os exercícios cobriam conceitos de programação como variáveis, expressões, condicionais, iteração e *arrays*. Durante a sessão, foram gravados vídeos com o rosto dos alunos e *logs* contendo dados de digitação e compilações. Após participarem da sessão de aprendizagem, cada aluno assistiu sua própria sessão (*logs* e vídeos) e anotou suas emoções. Os alunos tinham liberdade para definir o momento da anotação e a emoção anotada (engajado, confusão, frustração e tédio). Como os estados afetivos engajamento e confusão obtiveram mais ocorrências, apenas estes foram considerados no treinamento. O estado afetivo engajamento foi usado como um estado “não confuso”. Os dados relacionados aos demais estados afetivos anotados (tédio e frustração) e a intervalos não anotados, foram usados no treinamento do sistema sob o rótulo de estados inativos. O sistema coletou os seguintes dados do aluno: digitação (“digitando” ou “inativo”), compilação (“positiva” ou “negativa”) e *action units* (AUs) do FACS. O trabalho tratou cada intervalo de dados como um estado (digitando, não digitando, compilando, etc). O estado de digitação considerou um limite de 3 segundos, ou seja, quando o aluno não estava digitando por 3 segundos, foi considerado “inativo”. As anotações e os dados dos alunos foram utilizados para treinar modelos HMM (*Hidden Markov Model*). Como próximo passo, o melhor modelo de detecção foi utilizado para detectar automaticamente a confusão. O ambiente detecta a presença de confusão (ou não confusão) a cada 10 segundos. Se a confusão é detectada, o sistema oferece um guia ao aluno. Se aceito, o sistema mostra ao aluno um fluxograma com os passos da solução

do exercício. O aluno pode selecionar um passo do fluxograma para receber uma dica de texto sobre como implementar aquele passo. Na etapa de validação cruzada (*leave one out cross fold validation*), o melhor modelo HMM alcançou *Kappa* de 0,41. Para testar o algoritmo, participaram 35 alunos de diferentes níveis de conhecimento de uma universidade, de novatos a graduados. Os alunos foram divididos em dois grupos, de forma que existisse uma representação balanceada de idade, gênero e meses de experiência com programação. O grupo A (17 alunos) não recebeu nenhuma orientação ou adaptação na atribuição de exercícios, enquanto o grupo B (18 alunos) recebeu. Os alunos usaram o sistema durante 40 minutos. Para avaliar a detecção da confusão, os alunos assistiram trechos das sessões em que o sistema detectou confusão, iniciando sempre 30 segundos antes da detecção. Os alunos tiveram que responder “muito confuso”, “pouco confuso” e “não confuso” para cada trecho indicado pelo sistema. Os alunos não sabiam que estes trechos selecionados eram relacionados à identificação de confusão. Em 77,8% dos trechos que o sistema detectou confusão, os alunos relataram terem estado confusos.

Um outro trabalho de Tiam-Lee e Sumi (2019) tentou detectar as emoções confusão, engajamento, frustração e tédio, utilizando atributos de dados derivados de dados de digitação, de compilação e de rosto dos alunos. Da mesma forma como no trabalho anterior dos mesmos autores, os alunos utilizaram um ambiente de programação em que deveriam escrever um corpo de uma função que atendesse os requisitos do enunciado do problema, em uma sessão única de 45 minutos. Participaram 73 estudantes. Após a sessão, cada aluno assistiu a sua própria sessão (vídeos e *logs*). A sessão foi particionada em intervalos baseados no início e fim de uma série de digitação, momento de compilação de código, submissão de código e início de um novo problema. Para cada intervalo, o aluno deveria anotar um estado afetivo (engajado, confuso, entediado, frustrado ou neutro) e um rótulo de ação (lendo, pensando, escrevendo, encontrando, resolvendo, desfocando, entre outros). A média de tempo dos intervalos ficou em 17,24 segundos. O trabalho usou a técnica *OpenFace* para extrair informações sobre a postura de cabeça, que junto dos dados de digitação e de compilação, foram criados diferentes conjuntos de atributos de dados para treinar modelos de classificação dos estados afetivos. Os modelos foram avaliados com validação cruzada *k-fold* (*10-fold cross validation*). Os melhores resultados encontrados para a detecção de confusão foi uma acurácia de 0,77 e *Kappa* de 0,18.

O trabalho de Veia e Rodrigo (2017) descreve e avalia modelos de detecção das emoções tédio, confusão e frustração de alunos novatos em programação. Para o treinamento dos modelos, o estudo utilizou atributos oriundos de atividades de teclado e mouse. Foram realizadas três sessões com 60 alunos de três turmas de uma disciplina introdutória de

programação. No entanto, em função de problemas técnicos, apenas os dados de 55 alunos foram considerados. Durante as sessões, os alunos usaram o ambiente Dev-C++ e tiveram dados de *log* de teclado e mouse coletados, através de uma aplicação externa ao ambiente de programação. Uma câmera também gravou os rostos dos alunos e seus movimentos de corpo. Os *logs* referentes às atividades em teclado e mouse foram agrupados em segmentos fixos de 15 segundos. Os vídeos dos alunos também foram agrupados em segmentos fixos de 15 segundos e sincronizados com os respectivos segmentos de *logs*. Os rótulos de estados afetivos relativos aos segmentos foram obtidos através de sessões de anotação de estados afetivos, a partir de um esquema de codificação (DRAGON et al., 2008; RODRIGO et al., 2009). O estudo utilizou alguns classificadores de árvore (*J48*, *Decision Tree* e *Random Forest*) e seus resultados foram analisados através de medidas de precisão e *Kappa*. O estudo também realizou validação cruzada a nível de aluno. Os modelos também foram analisados para encontrar atributos de dados no reconhecimento de estados afetivos. O algoritmo *Decision Tree* apresentou os melhores resultados. Os resultados mostraram um acordo moderado (VIERA; GARRETT, 2005). Os conjuntos de dados que apresentaram melhores resultados foram a verbosidade do pressionamento de teclas (*Kappa* = 0,569 e precisão = 74,23), todos os atributos de dados de pressionamento de tecla (*Kappa* = 0,568 e precisão = 74,28) e atributos de dados combinados de pressionamento de teclas e mouse (*Kappa* = 0,572 e precisão = 74,37).

O trabalho de Bosch, Chen e D’Mello (2014) apresenta um modelo de detecção de emoções de alunos novatos aprendendo programação de computadores. Este trabalho utiliza atributos de dados derivados de vídeos do rosto dos aprendizes (AUs do FACS). Os dados foram coletados de 99 alunos novatos de uma turma de programação que, em uma única sessão, resolveram exercícios de programação na linguagem *Python*. Da mesma forma como nos trabalhos relacionados anteriores, após a sessão de codificação, os alunos assistiram vídeos sincronizados de seus rostos e telas. Os alunos fizeram anotações de seus estados afetivos em pontos fixos. Estes períodos foram escolhidos para corresponder com eventos de interação, como quando o aluno está em digitação ou visualizando um novo exercício. Também foram considerados períodos ociosos, quando o aluno não interage com o sistema por mais de 15 segundos. Também foi permitido que os alunos fizessem anotações espontâneas de estados afetivos em qualquer ponto dos vídeos. A ferramenta *Computer Expression Recognition Toolbox* (CERT) foi usada para detectar 19 unidades de ação (AUs) do FACS e algumas combinações de AUs. As AUs (a nível de frame) foram agregadas em cada segmento de vídeo usando a mediana, máximo, e desvio padrão. Orientação de cabeça, posição de nariz, tamanho do rosto e visibilidade de nariz, boca, olhos e orelhas foram acrescentados aos segmentos de

vídeo. Foram considerados 15 classificadores e 6 diferentes tamanhos de segmentos de vídeo (2, 3, 6, 9, 12 e 15 segundos). As anotações em tempo fixo renderam mais dados para as amostras, mas não apresentaram resultados tão satisfatórios, com valores de *Kappa* um pouco acima de 0,0 para *flow* / engajamento, tédio, confusão, frustração e neutro (respectivamente 0,112, 0,038, 0,064, 0,083 e 0,070). Apenas confusão e frustração tiveram pelo menos 100 anotações espontâneas. Em função disso, apenas estas duas emoções foram consideradas. Julgamentos espontâneos permitiram classificar melhor os estados de confusão e frustração (0,221 e 0,232, respectivamente) em segmentos curtos (2 segundos para confusão e 3 segundos para frustração).

O trabalho de Felipe et al. (2012) apresenta um modelo de detecção de confusão e tédio em um ambiente de programação, na linguagem C++. Uma primeira avaliação (FELIPE et al., 2012) descreve um estudo piloto que usou dados de pressionamento de teclas de 6 alunos. Os rótulos de estados afetivos foram obtidos através de anotações realizadas pelos próprios pesquisadores em frames de vídeo de 20 segundos cada. Um total de seis alunos participaram de três sessões de programação de 1 hora e 30 minutos cada. O trabalho obteve um *Kappa* de 0,856, usando o algoritmo *J48*. A principal limitação do estudo é a pequena quantidade de alunos envolvidos nas amostras, o que reduz a força e a capacidade de generalização dos resultados. O estudo de Veal et al. (2017) descreve um sistema que monitora estados afetivos dos alunos através da rede de computadores e usa o mesmo modelo do trabalho relacionado de Felipe et al. (2012). O trabalho foi avaliado com 40 alunos e raramente detectou confusão. Este estudo de 2017 não foi considerado um trabalho relacionado em função de não ter apresentado detalhes sobre a avaliação do modelo. No entanto, o artigo faz uma crítica ao estudo de Felipe et al. (2012) sobre a pequena quantidade de alunos participantes e ainda sugere, sem apresentar detalhes sobre sua suposição, que o uso de dados de pressionamento de teclas não seria suficiente para detectar estados afetivos.

O trabalho de Grafsgaard, Boyer e Lester (2011) descreve um modelo de detecção de confusão que usa o corpus do diálogo textual entre aluno e tutor, além de dados da tarefa (correta, incompleta, incorreta ou mistura de ambas) e do histórico de expressões faciais do aluno, para identificar a presença de AU4, uma unidade de ação (AU) do FACS correlacionada com a confusão (D'MELLO; LEHMAN; PERSON, 2010). Um total de 14 alunos resolveram um problema introdutório de programação e, durante esta atividade, participaram de um diálogo textual com um tutor humano, via computador. Além do corpus relativo à conversação, foram registradas informações sobre o progresso do aluno na tarefa e o histórico de ocorrência de AU4, anotados por dois codificadores treinados no FACS. Sequências das observações foram

utilizadas para treinar modelos HMM, OMM (*Observable Markov Model*) e uma *baseline*. Um design *leave-one-out* foi utilizado para validar os modelos. Os resultados do HMM foram significativamente superiores aos do OMM e da *baseline*, alcançando uma acurácia de 0,907 nos testes.

O trabalho de Lee et al. (2011) estudou relacionamentos entre confusão e desempenho de alunos novatos de programação no ambiente de aprendizagem *BlueJ*, na linguagem Java. Dois anotadores humanos anotaram amostras de *logs* como sendo relacionadas a alunos confusos ou não confusos. A partir das anotações de uma parte das amostras, foi possível treinar um modelo de detecção de confusão para rotular o conjunto completo de *logs*. Foram considerados os seguintes dados de *logs* relativos a compilações: tempo médio entre compilações; tempo máximo entre compilações; média de tempo entre compilações com erros; tempo máximo entre compilações com erros; número de compilações com erros; número de pares de compilações seguidas com o mesmo erro. Participaram 149 alunos calouros e de segundo ano de um curso. Um desafio do trabalho foi a definição do nível de granularidade dos quais as compilações foram agrupadas para o treinamento do algoritmo de aprendizado de máquina. Os pesquisadores não conseguiram, por exemplo, dizer se um aluno estava confuso baseado em uma única compilação. Também não era interessante para o estudo classificar o aluno como confuso ou não confuso considerando os dados de uma sessão inteira (foram 5 sessões, no total), pois os pesquisadores poderiam querer utilizar futuramente o algoritmo para detectar confusão enquanto o aluno escreve código. A decisão dos pesquisadores foi uma solução intermediária: organizar segmentos (o trabalho chama de *clips*) de oito compilações cada (quantidade definida por escolha arbitrária). Uma pequena aplicação foi escrita para a realização das anotações por dois codificadores. A aplicação mostra ao codificador um clip por vez, apresentando dados de oito compilações e disponibiliza botões para o codificador anotar “confuso”, “não confuso” ou “*clip* ruim” (quando o codificador identifica que o aluno não apresenta os requisitos mínimos para resolver um problema de programação). Os anotadores apresentaram desempenho moderado de concordância (*Kappa* de 0,77). Os anotadores anotaram apenas um subconjunto de segmentos. Os demais foram anotados a partir de um detector de confusão treinado a partir destas primeiras amostras anotadas. Para isso, foi utilizado o algoritmo de classificação *J48* (uma implementação do algoritmo *C4.5*). Uma validação *10-fold cross-validation* resultou em um *Kappa* de 0,86. O trabalho também concluiu que confusão prolongada está associada à baixo desempenho dos alunos. Além disso, compartilhando a mesma percepção dos autores do trabalho de Baker et al. (2012) e Baker e

Ocuppaugh (2015), não ficou claro se o detector estava detectando o estado afetivo de confusão ou apenas o aluno tendo dificuldade com o conteúdo no ambiente de desenvolvimento *BlueJ*.

### 3.4.3 Discussão dos trabalhos relacionados

Uma das principais preocupações dos pesquisadores da área de mineração de dados é a quantidade de dados disponíveis para o treinamento de modelos de aprendizado de máquina. Em geral, quanto mais complexo o modelo, mais amostras de dados serão necessárias para o modelo gerar boas fronteiras de classificação das classes de saída, e apresentar bom desempenho após o treinamento (DECATUR; GOLDREICH; RON, 1997). Neste sentido, o trabalho de Tiam-Lee e Sumi (2018) considerou apenas 11 alunos e 20 intervalos de dados com rótulo “confuso” e 24 intervalos com rótulo “engajado”, na etapa de treinamento. Um total de 35 alunos participaram da etapa de teste do modelo. O tempo de coleta também foi curto, em uma única sessão de aprendizagem com 45 minutos de duração. A quantidade reduzida de alunos participantes e de tempo de coleta de dados pode ter contribuído para os resultados deste trabalho não terem sido tão fortes (*Kappa* de 0,41).

Todos os trabalhos relacionados livres de sensores treinaram modelos de aprendizado de máquina através de uma abordagem dirigida a dados (*data-driven*), no qual algoritmos de aprendizado de máquina supervisionado tentaram identificar correlações entre conjuntos de dados e rótulos de julgamentos emocionais realizados por humanos (D’MELLO; GRAESSER, 2014b). Estes julgamentos podem ser realizados pelos próprios alunos (autorrelatos) ou por anotadores que observam os comportamentos dos estudantes. A Seção 3.2 apresentou diferentes formas de julgamento de emoções humanas para obtenção de rótulos emocionais para algoritmos de aprendizado de máquina. Uma desvantagem da abordagem *data-driven* é que apesar de ser interessante encontrar correlações não conhecidas entre dados e rótulos, o uso desses dados não se justifica em alguma teoria de emoções. Assim não se sabe o nível de relação entre os dados e os construtos de emoções. Um modelo de classificação *data-driven* que usa atributos derivados sem relação conhecida com alguma teoria de emoções, como dados sobre pressionamento de teclas, pode ter dificuldade para transferir seu desempenho para outros grupos de sujeitos (D’MELLO; GRAESSER, 2014b). Este tipo de problema no mapeamento de dados com determinados construtos emocionais já foi alertado por pesquisadores (BAKER et al., 2012; BAKER; OCUMPAUGH, 2015; FELIPE et al., 2012; VEA et al., 2017). A mesma reflexão não pode ser feita para os trabalhos que usaram dados faciais para detectar emoções, pois já é conhecida a forte correlação entre atributos faciais e estados afetivos (MCDANIEL et

al., 2007). No entanto, estes trabalhos não são livres de sensores. Nenhum dos trabalhos relacionados livres de sensores consideraram atributos de dados com relação conhecida com alguma teoria de emoções.

Outra discussão é sobre como os estudos relacionados trabalharam com as janelas de dados. Janela de dados é um conjunto de dados sobre o comportamento do aluno no ambiente em determinado intervalo de tempo. Cada conjunto de dados é uma amostra vinculada a um rótulo de emoção (exemplo: “confuso” e “não confuso”) para os modelos de aprendizado de máquina supervisionada. Cada amostra e o respectivo rótulo de emoção podem ser utilizados no treinamento de modelos de aprendizado de máquina supervisionada. Na etapa de detecção, quando o modelo de confusão já está treinado, uma amostra ainda não vista pelo modelo é usada como entrada do modelo, que, por sua vez, retorna como saída uma classe de emoção identificada. Há dois questionamentos a fazer sobre a criação de janelas de dados nos trabalhos relacionados: a) qual o momento da criação das janelas, se criadas em pontos fixos de tempo ou em pontos arbitrários; e b) qual o tamanho das janelas em relação ao tempo de observação do comportamento do aluno: fixo ou variável.

O trabalho apresentado por Tiam-Lee e Sumi (2019) utilizou janelas de dados de tempo variado na etapa de treinamento. Cada janela, e seu tamanho, foi criada com base na interação do aluno com o sistema, como nos estados “digitando”, “não digitando” e “compilando”. Cada uma dessas janelas recebeu um rótulo de emoção na etapa de anotação de emoções. O autor da Tese supõe um problema nessa abordagem. Criar janelas de dados a partir de eventos de interação do aluno com o sistema e não a partir de eventos relacionados às emoções observadas pode não permitir um vínculo preciso dos rótulos de emoções com os dados. Além disso, o trabalho relacionado utilizou dados de apenas uma única sessão de aprendizagem, com duração de 45 minutos, o que pode ser considerado pouco tempo para modelar um construto complexo como o da confusão.

O trabalho apresentado por Bosch, Chen e D’Mello (2014) criou segmentos de dados a partir da anotação em pontos não fixos da etapa de anotação retrospectiva, mas manteve um tamanho fixo para as janelas de dados, de 2 segundos, após testar diferentes tamanhos. Outros trabalhos relacionados definiram janelas de dados em momentos fixos (FELIPE et al., 2012), o que também poderia prejudicar a representatividade dos dados vinculados aos rótulos afetivos, pois estados afetivos como a confusão não ocorrem em momentos fixos de tempo, mas a partir de avaliações do ambiente e situações (D’MELLO; GRAESSER, 2014b). Com janelas criadas em momentos fixos, um estado afetivo poderia iniciar no final de uma janela e finalizar em outra janela. Já foram apresentadas evidências que a criação desses segmentos de dados em

pontos fixos de tempo pode reduzir o desempenho dos classificadores emocionais (BOSCH; CHEN; D'MELLO, 2014). Todos os trabalhos relacionados livres de sensores criaram janelas de dados em pontos fixos e com tamanho fixo (FELIPE et al., 2012; LEE et al., 2011; VEA; RODRIGO, 2017b).

Sobre os resultados dos três trabalhos relacionados livres de sensores, dois deles apresentaram bons valores de correspondência entre as detecções dos modelos e os julgamentos humanos, com valores de *Kappa* de 0,856 (FELIPE et al., 2012) e 0,860 (LEE et al., 2011). No entanto, um dos trabalhos (FELIPE et al., 2012) foi treinado com dados de apenas seis alunos. O estudo também não utilizou alunos com diferentes características para verificar o poder de generalização do modelo para diferentes contextos. O outro trabalho com bom valor de *Kappa* (LEE et al., 2011) usou dados de compilação, como média de tempo entre compilações com e sem erros e quantidade de compilações com erro e sem erros. Além disso, usou janelas fixas compostas por sequências de 8 compilações. Como este tipo de dado não se refere ao andamento da codificação (em tempo de codificação), mas após a etapa de codificação, quando o aluno submete seu código para compilação, este modelo não poderia ser utilizado para detectar a confusão em tempo de codificação. Um dos preceitos para ambientes computacionais adaptativos de aprendizagem terem sucesso nos seus objetivos educacionais é acompanhar e inferir características dos estudantes durante o processo da aprendizagem. Por esse motivo, considera-se que este trabalho relacionado (LEE et al., 2011) possui importantes restrições. O terceiro trabalho relacionado (VEA; RODRIGO, 2017c) com os melhores resultados apresentou um valor de *Kappa* de 0,572, ainda abaixo do valor considerado ideal para classificadores de emoções, que é de 0,6 para o *Kappa* (OCUMPAUGH; BAKER; OCUMPAUGH, J., & BAKER, 2014). Este estudo relacionado apresenta um modelo composto por janelas de tempo fixo, com rótulos obtidos em tempo fixo e dados não justificados em nenhuma teoria de emoções (dados de pressionamento de teclas e uso do mouse). O autor da Tese supõe que estes resultados podem ser melhorados com um modelo composto por janelas de dados obtidos em tempo não fixo (decisão arbitrária do juiz humano), e utilizando dados com forte aderência às teorias cognitivas de *appraisal* de emoções.

Os trabalhos relacionados também não avaliaram o poder de generalização dos detectores de confusão livres de sensores, em tarefas de programação de computadores, para grupos de estudantes com características heterogêneas. Considerando a suposição que detectores *data-driven* que usam dados não justificados em alguma teoria de emoção, podem apresentar problemas de falso-positivo (D'MELLO; GRAESSER, 2014b), é possível supor que esses dados podem também falhar, no seu poder de generalização, para prever a confusão de

alunos com características diferentes daqueles alunos considerados no treinamento dos modelos. Alguns trabalhos relacionados (BOSCH; CHEN; D'MELLO, 2014; LEE et al., 2011; VEA; RODRIGO, 2017a) utilizaram validação cruzada *k-fold* a nível de estudante. Dois destes trabalhos são referentes a modelos livres de sensores (LEE et al., 2011; VEA; RODRIGO, 2017a). Neste tipo de validação cruzada, os dados de determinado estudante ou são usados no treino ou no teste do modelo. Este tipo de validação cruzada valida o poder de generalização do modelo para diferentes alunos, mas não garante que os alunos selecionados para o treino e o teste possuem características distintas. Para isso ocorrer, o trabalho poderia agrupar os alunos com características heterogêneas em dois grupos e só usar os dados dos alunos de um grupo no treino e os dados dos alunos de outro grupo no teste, e vice-versa.

## 4 TRABALHO DESENVOLVIDO

Este breve capítulo apresenta uma descrição geral do trabalho realizado, dos métodos implementados e da abordagem computacional utilizada pela Tese para alcançar seus objetivos, dar respostas às suas questões de pesquisa e verificar sua hipótese. O detalhamento do trabalho realizado e dos métodos implementados é apresentado nos capítulos seguintes.

O objetivo geral da Tese é verificar a sua hipótese, ou seja, verificar se a utilização de dados sobre estimativas de conhecimento do aluno no domínio de programação de computadores, juntamente de dados sobre a interação do estudante com o ambiente computacional de aprendizagem de programação, pode melhorar o desempenho dos modelos livres de sensores na detecção da confusão do aluno nas tarefas de aprendizagem de programação de computadores, quando comparado a utilizar apenas os dados de interação do estudante com o ambiente de aprendizagem.

A Tese também define três objetivos específicos, a lembrar: verificar se modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente possuem desempenho superior a modelos que utilizam apenas dados sobre interação com o ambiente, quando são consideradas diferentes janelas de observação de tamanho fixo (OE1); verificar se modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente possuem desempenho superior a modelos que utilizam apenas dados sobre interação com o ambiente, quando é considerada uma janela de observação de tamanho variável (OE2); e verificar se modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente generalizam bem seus resultados para estudantes com características heterogêneas, comparado a modelos que utilizam apenas dados sobre interação com o ambiente (OE3). Ao alcançar seus objetivos (geral e específicos), a Tese busca responder suas questões de pesquisa e apresentar evidências para aceitar ou rejeitar sua hipótese.

A Tese foi desenvolvida como uma pesquisa de computação aplicada, no qual técnicas computacionais são aplicadas para resolver o problema de detecção da confusão de alunos no contexto de aprendizagem de programação de computadores em um ambiente computacional, como um software editor de códigos de programação ou até mesmo um ambiente adaptativo de aprendizagem. Os objetivos indicam que o tipo de problema abordado, no âmbito computacional, é um problema de classificação. Na ciência da computação, uma abordagem tradicional para este tipo de problema, também utilizada em todos os trabalhos relacionados, é

a utilização de modelos de aprendizado de máquina supervisionada. Estes modelos utilizam algoritmos computacionais que funcionam, basicamente, da seguinte forma: o algoritmo visualiza um conjunto de amostras de dados, chamado de amostras ou exemplos de treino, que refletem determinados comportamentos do problema representado, e tentam encontrar padrões relacionados a classes de interesse pré-definidas; após aprender os padrões, o algoritmo visualiza um outro conjunto de amostras de dados, chamado de amostras ou exemplos de teste, que não foi utilizado no aprendizado dos padrões, e tentam identificar corretamente as classes de interesse, baseado no que aprenderam (KUBAT, 2017). No contexto da Tese, o algoritmo visualiza amostras de dados sobre o comportamento do aluno nas tarefas de programação de computadores e tenta aprender padrões relacionados ao estado de confusão e de não confusão do estudante. Após, o algoritmo visualiza um outro conjunto de amostras, não usado no treinamento, para detectar a confusão (ou não confusão) do estudante. Espera-se um bom desempenho dos modelos nas amostras não utilizadas no treino, ou seja, espera-se que, após treinados, os modelos identifiquem corretamente a confusão (ou não confusão) do aluno durante suas tarefas de programação.

A Tese desenvolveu e avaliou diversos modelos de aprendizado de máquina supervisionada que identificam a confusão do aluno (e a não confusão) em amostras de dados que refletem o comportamento do estudante em um ambiente computacional de programação de computadores, enquanto realizam tarefas de programação. A Tese verifica o desempenho dos modelos quando são treinados com amostras compostas por estimativas de conhecimento do aluno em programação, juntamente dos demais dados da abordagem dirigida a dados dos trabalhos relacionados (dados sobre o uso de mouse, teclado, erros de codificação, entre outros). Estes modelos são identificados como **modelos da Tese**, pois representam a hipótese da Tese. O trabalho também verifica o desempenho dos modelos quando são treinados com amostras compostas apenas por dados da abordagem dirigida a dados dos trabalhos relacionados (dados sobre o uso de mouse, teclado, erros de codificação, entre outros). Estes modelos são identificados como **modelos *baseline***, e representam a abordagem dos trabalhos relacionados.

O principal método do trabalho é comparar o desempenho dos melhores modelos da Tese com os melhores modelos *baseline*. Desempenhos superiores dos modelos da Tese são consideradas evidências para a hipótese da Tese. Esta abordagem foi escolhida, porque seria inapropriado comparar os resultados da Tese diretamente com os resultados apresentados nos artigos dos trabalhos relacionados, pois estes trabalhos utilizaram bases de dados diferentes dos utilizados na Tese. A Tese representou a abordagem dos trabalhos relacionados (dados de

interação com o ambiente) nos modelos *baseline* e comparou seus resultados com os resultados dos modelos que representam a hipótese da Tese.

Serão geradas centenas de modelos de aprendizagem de máquina, que refletem diferentes algoritmos de aprendizado de máquina recomendados pela literatura e utilizados pelos trabalhos relacionados, e que refletem diferentes configurações de dados nas amostras de treinamento e testes dos modelos. Além das amostras refletirem as duas abordagens testadas, elas também refletem diferentes janelas de observação de dados de comportamento do aluno. Uma janela de dados é o tempo que o aluno é observado para sintetizar um determinado dado de uma amostra. Por exemplo, em uma janela de dados de 5 segundos, um dado de comportamento como a “quantidade de teclas digitadas” refletirá a quantidade de teclas digitadas pelo aluno em 5 segundos. As subquestões de pesquisa SQP1 e SQP2 são questões exploratórias que buscam entender qual o melhor tempo de observação do comportamento do aluno no contexto da Tese, se através de janelas de dados de tempo fixo ou janelas de dados de tempo aleatório. Além de identificar a melhor dinâmica nas janelas de observação, a Tese busca verificar se os modelos da Tese apresentam desempenho superior aos modelos *baseline* nas diferentes configurações de janelas de observação.

É importante reforçar que a Tese utilizou uma abordagem livre de sensores na geração das amostras usadas nos modelos de aprendizagem de máquina, pois os dados foram obtidos sem a utilização de sensores (*webcams*, sensores de postura, entre outros). Os dados foram obtidos somente pela interação do aluno com o ambiente de programação de computadores, como ações de teclado e mouse, e demais inferências realizadas a partir dessas interações, como a identificação de erros de codificação nas tarefas e avaliações sobre as soluções dos alunos para as tarefas de programação.

Os dados utilizados para compor as amostras usadas nos treinamentos e testes dos modelos de aprendizado de máquina foram obtidos a partir da interação de alunos com um ambiente de programação de computadores, durante a realização e tarefas de programação em turmas de instituições de ensino. Participaram alunos de duas modalidades de ensino: ensino técnico em Informática e ensino superior (graduação) em Tecnologia da Informação (TI).

O Capítulo 5 descreve as características dos participantes envolvidos, incluindo características relacionadas a competências emocionais e experiências prévias desses estudantes em programação de computadores. Um dos objetivos desse capítulo é verificar se os alunos participantes possuem características heterogêneas. Esta informação é importante para o objetivo específico OE3, que verifica quão bem os modelos implementados (Tese e *baseline*) generalizam para alunos com características heterogêneas.

O capítulo 6 descreve o ambiente de programação de computadores adaptado pela Tese para a coleta dos dados dos alunos participantes. O ambiente é uma extensão para um conhecido ambiente de programação utilizado em disciplinas introdutórias de programação, o *BlueJ*. O *BlueJ* foi estendido para fazer a coleta dos dados de interação dos alunos com o ambiente. O capítulo também apresenta uma avaliação de usabilidade do ambiente, com o objetivo de verificar se algum problema de usabilidade possa ter impactado na coleta dos dados.

Os dados usados nas amostras dos modelos da Tese sobre estimativas do conhecimento do aluno em programação foram obtidos através de um modelo de inferência implementado para essa Tese pelo próprio autor. O modelo e uma avaliação dele são apresentados no Capítulo 7.

O Capítulo 8 descreve a geração das amostras de dados para os modelos da Tese e *baseline*. São apresentadas todas as etapas de desenvolvimento dessas amostras, como a coleta dos dados no ambiente de programação e a transformação em amostras de entrada para os modelos. Também é descrito como foram obtidos os rótulos de emoções (confusão e não confusão) vinculados às amostras.

O Capítulo 9 detalha os métodos utilizados pela Tese para alcançar os objetivos específicos OE1 e OE2, e descreve os resultados obtidos. O Capítulo 10 descreve o método utilizado pela Tese para alcançar o objetivo específico OE3, e descreve os resultados obtidos.

O Capítulo 11 apresenta a conclusão da Tese, como discussões finais sobre os resultados, as contribuições, as limitações, e possibilidades de trabalhos futuros.

## 5 PARTICIPANTES DA PESQUISA

Este capítulo descreve características dos participantes da pesquisa. O objetivo específico OE3 é sobre verificar o poder de generalização dos modelos para alunos com características heterogêneas. Desta forma, a Tese buscou convidar alunos participantes com características heterogêneas. O autor da Tese convidou turmas de ensino técnico e de ensino superior (graduação), supondo que os alunos destes diferentes níveis de ensino pudessem apresentar características heterogêneas, como possuir diferenças de idades, diferentes características emocionais e diferentes experiências prévias de programação de computadores. Este capítulo busca evidenciar características heterogêneas dos participantes.

Tabela 1 – Distribuição dos alunos participantes da pesquisa

<i>Turmas</i>	<i>Instituição de ensino</i>	<i>Nível de ensino</i>	<i>Quantidade de alunos</i>	<i>Quantidade Masculino</i>	<i>Quantidade Feminino</i>
Turma 1	Instituição 1	Técnico	18	18	0
Turma 2	Instituição 2		13	8	5
		<b>Total Técnico</b>	<b>31</b>	<b>26</b>	<b>5</b>
Turma 3	Instituição 3	Superior	3	2	1
Turma 4	Instituição 3		20	17	3
Turma 5	Instituição 3		8	8	0
		<b>Total Superior</b>	<b>31</b>	<b>27</b>	<b>4</b>
		<b>Total geral</b>	<b>62</b>	<b>53</b>	<b>9</b>

Fonte: Elaborado pelo autor.

Os participantes foram alunos matriculados em disciplinas introdutórias de programação de computadores de cursos de ensino técnico e de ensino superior (graduação) de três instituições de ensino da região sul do Brasil. O autor da Tese foi o professor nas turmas participantes do ensino técnico. A participação dos alunos consistiu em utilizar um ambiente de programação de computadores adaptado para a Tese (descrito no Capítulo 6), durante todo o semestre de aulas, entre os meses de fevereiro e julho de 2021. O ambiente de programação da Tese foi a principal ferramenta para o desenvolvimento dos exercícios de programação nestas disciplinas. Enquanto os alunos interagem com o ambiente para resolver os exercícios, o ambiente coletou dados de interação que foram utilizados para gerar as amostras de treino e testes dos modelos da Tese e *baseline* e coletou os rótulos de emoções dos estudantes (descrito no Capítulo 8). A Tabela 1 mostra a distribuição dos alunos nas turmas participantes e nos respectivos níveis de ensino. As turmas participantes foram compostas por um total de 93 alunos. No entanto, a Tese considerou como participantes da pesquisa apenas os alunos que, além de utilizarem o ambiente

de programação da Tese, também entregaram os termos de consentimento e/ou assentimento assinados (ver Seção 5.3). Nestes termos, a quantidade total de alunos participantes foi 62.

As evidências de características heterogêneas entre os alunos participantes de ensino superior e técnico foram obtidas em relação a aspectos emocionais, de experiências anteriores em computação e programação, de sexo e de idade. Os alunos responderam a dois questionários para levantamento de evidências sobre experiências anteriores em programação e sobre competências emocionais. As próximas Seções descrevem as evidências obtidas.

### 5.1 Competências emocionais

Com o objetivo de verificar suas características emocionais, os alunos responderam ao **Questionário de Competência Emocional (QCE)**. O QCE é uma versão traduzida e validada para o português de Portugal (FARIA; SANTOS, 2017) do questionário *Emotional Skills and Competence Questionnaire (ESCQ)* (TAKSIC, 2000). O QCE é uma medida de autorrelato que inclui um total de 45 itens, respondidos em uma escala Likert de 6 pontos, variando entre “nunca” e “sempre” (1 – nunca; 2 – raramente; 3 - poucas vezes; 4 - algumas vezes; 5 – frequentemente; e 6 - sempre). Valores próximos de 1 indicam “baixa competência emocional”, enquanto valores próximos de 6 indicam “elevada competência emocional”. Os 45 itens representam três dimensões emocionais: **percepção emocional**, com 15 itens (exemplo: “Consigo perceber a diferença, se os meus amigos estão tristes ou decepcionados”); **expressão emocional**, com 14 itens (exemplo: “Consigo expressar os meus sentimentos e emoções em palavras”); e **capacidade de lidar com a emoção**, com 16 itens (exemplo: “Consigo manter-me de bom humor, mesmo quando os que me rodeiam estão de mau humor”). O QCE destina-se a adolescentes (a partir de 15 anos), jovens adultos e adultos, e pode ser utilizado em vários contextos, entre eles o acadêmico (escolar ou universitário). Por ser uma medida de autorrelato, o QCE mede a percepção sobre a competência emocional pessoal (FARIA; SANTOS, 2017). Os participantes responderam ao questionário entre os meses de fevereiro e março de 2021.

A Tabela 2 mostra os resultados da aplicação do questionário QCE. O questionário foi disponibilizado aos alunos através de um documento do *Google Forms*, pois a participação dos alunos ocorreu em um período de aulas remotas, em função da pandemia de COVID-19. O ANEXO A mostra o questionário de competência emocional original, traduzido para o português, disponível no artigo de Faria e Santos (2017), usado como base para a geração do documento do *Google Forms* respondido pelos alunos. Não foram alterados os textos originais dos 45 itens. Nem todos os alunos que participaram da pesquisa submeteram o formulário com

suas respostas. Todos os 31 alunos participantes do ensino técnico submeteram suas respostas, mas apenas 17 participantes do ensino superior submeteram, totalizando 48 respostas (77,4% do total de participantes). Uma hipótese para mais participação dos alunos do ensino técnico é que, nestas turmas, o autor da Tese foi o professor, e podem ter sido mais incentivados a participar do que nas turmas de graduação.

Tabela 2 – Resultados sobre as medidas de competência emocional

<i>Dimensão emocional</i>	<i>Ensino Técnico</i>		<i>Graduação</i>		<i>Mann-Whitney U</i> <i>p</i>
	<i>Média</i>	<i>Desvio</i>	<i>Média</i>	<i>Desvio</i>	
Percepção emocional	4,07096	1,26547	3,92156	1,20526	,06378
Expressão emocional	4,01152	1,37869	3,93697	1,19883	,15031
Capacidade de lidar com a emoção	4,29637	1,39947	4,30882	1,32017	,52182
<b>Total</b>	<b>4,13261</b>	<b>1,35532</b>	<b>4,06405</b>	<b>1,25862</b>	<b>,06952</b>

Fonte: Elaborado pelo autor.

Os resultados mostram que nas dimensões de “percepção emocional”, “expressão emocional” e na escala total, os valores de média dos alunos do ensino técnico foram superiores aos valores dos alunos do ensino superior. Apenas na dimensão “capacidade de lidar com a emoção”, o valor médio dos alunos da graduação foi superior. Para verificar se a diferença encontrada entre as médias foi significativa, foi aplicado o teste estatístico *Mann-Whitney U*. O teste é utilizado para comparar as diferenças entre dois grupos independentes (alunos do técnico e alunos de graduação) quando a variável dependente (medidas de competência emocional) é ordinal (é o caso das medidas obtidas) ou contínuas, mas não normalmente distribuídas. Testes de *Shapiro-Wilk* mostraram que nenhuma das medidas para os 45 itens são normalmente distribuídas. Os testes estatísticos mostraram que nas três dimensões e na escala total, as diferenças das médias não são significativas, com valores de *p* acima do nível de significância ( $\alpha = 0,05$ ). Na dimensão “percepção emocional” e na escala total, a diferença é significativa apenas marginalmente, com valores de *p* próximos de 0,05. Os resultados indicam que não foi encontrada diferença significativa entre os alunos do ensino técnico e de graduação nas suas percepções sobre suas competências emocionais.

## 5.2 Experiências anteriores em programação de computadores

Com o objetivo de verificar as experiências anteriores dos alunos, foi aplicado um questionário, criado pela Tese, com questões sobre experiências profissionais e não profissionais com programação de computadores e a linguagem de programação Java

(linguagem usada para a codificação dos exercícios de programação). Os participantes responderam ao questionário entre os meses de fevereiro e março de 2021.

Quadro 3 – Questões do questionário sobre as experiências anteriores dos alunos

#	Questão	Tipo de resposta
1	Há quantos anos você possui alguma experiência (profissional e/ou não profissional) com programação de computadores (qualquer linguagem de programação)?	Valor inteiro livre.
2	Há quantos anos você possui alguma experiência (profissional e/ou não profissional) com a linguagem de programação JAVA?	
3	Avalie em qual dos níveis abaixo você se enquadra melhor sobre seu conhecimento em programação de computadores (qualquer linguagem de programação). Nível 1: Avalio que não possuo qualquer conhecimento de programação de computadores; Nível 2: Avalio que possuo algum conhecimento, mas menos que conhecimento básico, em programação de computadores; Nível 3: Avalio que possuo conhecimento básico em programação de computadores; Nível 4: Avalio que possuo conhecimento acima do básico em programação de computadores; Nível 5: Avalio que possuo domínio em programação de computadores.	Valor ordinal de 1 a 5.
4	Avalie em qual dos níveis abaixo você se enquadra melhor sobre seu conhecimento na linguagem de programação JAVA. Nível 1: Avalio que não possuo qualquer conhecimento na linguagem de programação JAVA; Nível 2: Avalio que possuo algum conhecimento, mas menos que conhecimento básico, na linguagem de programação JAVA; Nível 3: Avalio que possuo conhecimento básico na linguagem de programação JAVA; Nível 4: Avalio que possuo conhecimento acima do básico na linguagem de programação JAVA; Nível 5: Avalio que possuo domínio na linguagem de programação JAVA.	
5	Abaixo, aparecerá nomes de componentes de conhecimento de programação, suportados por diversas linguagens de programação, não somente na linguagem Java. Avalie seu nível de conhecimento em cada um dos componentes de conhecimento abaixo, com respostas ENTRE 1 e 5, sendo que 1 significa que você avalia que não possui qualquer conhecimento no componente de conhecimento e 5 que você avalia que domina o componente de conhecimento.	
5.1	Classes	
5.2	Atributos	
5.3	Métodos	
5.4	Métodos construtores	
5.5	Invocação de métodos (chamada de métodos)	
5.6	Instanciação de classe (criação de objeto)	
5.7	Criação e manipulação de variáveis	
5.8	Expressões com operadores de atribuição (=, +=, etc)	
5.9	Expressões com operadores aritméticos (+, -, *, etc)	
5.10	Expressões com operadores relacionais (>, >=, <, etc)	
5.11	Expressões com operadores lógicos (AND, OR, etc)	
5.12	Estrutura de seleção If	
5.13	Estrutura de seleção switch case	
5.14	Estrutura de repetição do...while	
5.15	Estrutura de repetição while	
5.16	Estrutura de repetição for	
5.17	Estrutura de repetição foreach (usado com arrays)	
5.18	Criação e manipulação de arrays (vetores)	
5.19	Herança (classes e subclasses)	

Fonte: Elaborado pelo autor.

O Quadro 3 apresenta as questões do questionário. As duas primeiras são sobre o tempo de experiência profissional e/ou não profissional do aluno com programação de computadores e com a linguagem de programação Java. As respostas para estas questões são valores inteiros livres, referentes a uma quantidade de anos. A terceira e quarta questões questionam o aluno sobre sua percepção sobre seu nível de conhecimento em programação de computadores e na linguagem Java. As respostas são valores ordinais de 1 a 5, no qual 1 corresponde a não possuir qualquer conhecimento e 5 corresponde a dominar o conhecimento. Nas demais questões, de 5.1 a 5.19, o aluno deve avaliar, através de valores ordinais de 1 a 5, sua percepção sobre seu nível de conhecimento em diversas estruturas de programação específicas.

Tabela 3 – Resultados sobre as experiências anteriores em programação

<i>Questões</i>	<i>Ensino Técnico</i>		<i>Ensino Superior</i>		<i>Mann-Whitney U</i>
	<i>Média</i>	<i>Desvio</i>	<i>Média</i>	<i>Desvio</i>	<i>p</i>
1	0,42857	0,87891	1,33333	1,99999	,01694
2	0,25000	0,70052	0,44444	1,36813	,25308
3	1,53571	0,63724	2,88888	1,08604	<,00001
4	1,35714	0,62148	2,14814	1,06351	,00070
5.1	2,32142	1,24880	3,07407	1,41219	,02219
5.2	1,53571	0,88116	3,00000	1,35873	,00002
5.3	2,07142	1,15240	2,92592	1,23804	,00565
5.4	1,25000	0,58531	2,77777	1,21949	<,00001
5.5	1,96428	1,23174	2,88888	1,31070	,00362
5.6	1,42857	0,74179	3,29629	1,40916	<,00001
5.7	2,00000	1,21716	3,62962	1,33440	,00002
5.8	2,25000	1,26564	4,03703	1,25518	<,00001
5.9	2,46428	1,26145	4,25925	1,12975	<,00001
5.10	1,92857	0,97860	4,07407	1,20658	<,00001
5.11	1,60714	0,83174	3,74074	1,19590	<,00001
5.12	1,32142	0,66963	3,62962	1,39085	<,00001
5.13	1,17857	0,54796	3,18518	1,56983	<,00001
5.14	1,21428	0,56811	3,14814	1,56164	<,00001
5.15	1,21428	0,56811	3,48148	1,47727	<,00001
5.16	1,21428	0,49867	3,48148	1,50308	<,00001
5.17	1,07142	0,37796	1,88888	1,52752	,00446
5.18	1,17857	0,66963	2,37037	1,66752	,00034
5.19	1,35714	0,67846	2,11111	1,45002	,02682
Total 5.x	1,60902	0,97975	3,21052	1,49805	<,00001

Fonte: Elaborado pelo autor.

A Tabela 3 apresenta os resultados de média e desvio padrão sobre as experiências anteriores em programação, em ambos os níveis de ensino. Também foram realizados testes estatísticos *Mann-Whitney U* para verificar se as diferenças foram significativas.

Para as questões sobre tempo, em anos, de experiência profissional e/ou não profissional com programação e a linguagem Java (questões 1 e 2), a média dos alunos de graduação foi

superior em ambas, mas significativa ( $p < 0,05$ ) apenas para a questão 1, sobre o tempo de experiência anterior com programação de computadores. Não foi significativa a diferença em relação ao tempo de experiência anterior na linguagem Java.

Nas questões de percepção dos alunos sobre seu nível de conhecimento em programação e em Java (questões 3 e 4), as médias do grupo de alunos de graduação foram todas significativamente superior. Nas questões de percepção dos alunos sobre seu nível de conhecimento em estruturas de programação específicas (questões 5.1 a 5.19), todas as médias do grupo de graduação foram significativamente superior. Na última linha da tabela (Total 5.x), que apresenta o resultado total sobre todas as percepções em todas as estruturas de programação, a média para os alunos da graduação também foi superior e significativa. Outro resultado obtido é que, com exceção das questões 5.8 e 5.9, os valores de desvio padrão dos alunos de graduação foram superiores aos dos alunos do ensino técnico. Isso indica que há uma variação menor nos valores dos alunos de ensino técnico, o que poderia indicar um maior nivelamento de experiências anteriores entre os alunos do ensino técnico.

A Tese também verificou se a diferença de idade dos alunos participantes é significativa entre os grupos. Apenas um aluno não respondeu ao pesquisador sobre sua idade, de forma que foram obtidas as idades de 61 alunos participantes. Os resultados são apresentados na Tabela 4. A média de idade dos alunos de graduação é significativamente ( $p < 0,05$ ) superior à média de idade dos alunos de ensino técnico.

Tabela 4 – Idades dos alunos por nível de ensino

<i>Ensino Técnico</i>				<i>Graduação</i>				<i>Mann-Whitney U</i>
<i>Média</i>	<i>Desvio</i>	<i>Menor</i>	<i>Maior</i>	<i>Média</i>	<i>Desvio</i>	<i>Menor</i>	<i>Maior</i>	<i>p</i>
18,51612	5,84681	15	45	20,26666	1,91369	18	25	<,00001

Fonte: Elaborado pelo autor.

Os resultados não evidenciaram uma diferença significativa entre os alunos de ensino técnico e os alunos de ensino superior em relação à percepção de competência emocional, mas evidenciaram uma diferença significativa da percepção dos alunos sobre seu conhecimento anterior e sua experiência anterior (profissional ou não profissional) em programação de computadores. A idade dos alunos dos dois grupos também é significativamente diferente. A próxima seção apresenta algumas considerações sobre algumas questões éticas na participação dos alunos na pesquisa.

### 5.3 Questões éticas de pesquisa

A Tese utilizou dados coletados de interações de alunos com um ambiente de programação. Todos os projetos de pesquisa desenvolvidos no Brasil que envolvam seres humanos devem atender às resoluções 466/12 e 510/16, do CEP/CONEP (Comissão Nacional de Ética em Pesquisa, do Conselho Nacional de Saúde), e serem aprovados pelo CEP (Comité de Ética e Pesquisa) da instituição de ensino onde o pesquisador realiza a pesquisa. Com o objetivo de atender as resoluções, foi desenvolvido um projeto de pesquisa, referente à Tese, para apreciação do CEP da Universidade do Vale do Rio dos Sinos (Unisinos). O projeto foi submetido à Plataforma Brasil<sup>1</sup>, recebendo o CAAE (Certificado de Apresentação para Apreciação Ética) de número 24435519.2.0000.5344. O projeto foi, posteriormente, aprovado pelo CEP da Unisinos, através do parecer de número 3.707.158, de 18 de novembro de 2019. Posteriormente, foram aprovadas emendas relativas à coleta de dados à distância, em função da pandemia de COVID-19. Para a coleta e uso de dados pela Tese, foram utilizados somente dados de alunos que retornaram assinados os termos aprovados pelo CEP da Unisinos. Alunos maiores de idade retornaram assinado pelo próprio aluno um termo de consentimento livre e esclarecido (TCLE). Alunos menores de idade retornaram um termo de assentimento (TA) assinado pelo próprio aluno e um termo de consentimento livre e esclarecido (TCLE) assinado por um adulto responsável. Os termos e o parecer do CEP estão disponíveis nos apêndices P, Q, R e S.

---

<sup>1</sup> Disponível em <http://plataformabrasil.saude.gov.br/>



## 6 AMBIENTE DE PROGRAMAÇÃO ADAPTADO PARA A TESE

Este capítulo apresenta o ambiente de programação de computadores adaptado para a Tese, chamado a partir daqui como **BlueJ Afetivo**. A principal importância do *BlueJ Afetivo* para os objetivos da Tese foi permitir a coleta de dados de interação dos alunos com o ambiente, enquanto desenvolvem soluções para exercícios de programação de computadores atribuídos pelo professor. O *BlueJ Afetivo* é uma versão modificada do *BlueJ*<sup>1</sup>, um ambiente de desenvolvimento de *software* projetado para ser utilizado no ensino de programação de computadores (KÖLLING et al., 2003). Para esta coleta ocorrer, diversas funcionalidades foram adicionadas ao *BlueJ*, as quais são apresentadas neste capítulo.

Outra característica implementada foi a conectividade do *BlueJ* com um sistema *web* de programação de computadores chamado **Feeper**, apresentado na Seção 6.1, que fornece funcionalidades ao *BlueJ Afetivo*, como a disponibilização dos exercícios atribuídos pelos professores e o acesso a um Juiz Online<sup>2</sup> (JO) que avalia a solução do aluno. O sistema *Feeper* também foi adaptado pelo autor da Tese para atender requisitos técnicos de conectividade com o *BlueJ Afetivo* e de coleta de dados para a Tese. As Seções 6.2 e 6.3 descrevem o *BlueJ Afetivo*. A Seção 6.4 descreve a aplicação e os resultados de uma avaliação de usabilidade do *BlueJ Afetivo*.

### 6.1 Feeper

O *Feeper* é um sistema *web* projetado para apoiar alunos e professores em disciplinas de programação de computadores. O ambiente foi implementado por Fábio Alves, em seu trabalho de conclusão de graduação na Unisinos (ALVES; JAQUES, 2014). Foi posteriormente estendido por Gilvani Schneider, que acrescentou um mecanismo de *feedback*, também em um trabalho de conclusão de graduação na Unisinos (SCHNEIDER; JAQUES, 2016). O ambiente foi novamente estendido por Rodrigo Smiderle, em seu trabalho de mestrado, na Unisinos, no qual acrescentou estratégias de gamificação ao ambiente (SMIDERLE, 2019). A arquitetura do *Feeper* é composta por três módulos: a) **Módulo do Aluno**; b) **Módulo do Professor**; e c) **Juiz Online**. Os próximos parágrafos descrevem estes módulos e explicam como a Tese utilizou estes módulos.

---

<sup>1</sup> Disponível em <https://www.bluej.org/>.

<sup>2</sup> Sistema de avaliação automática de tarefas de programação (KURNIA; LIM; CHEANG, 2001).

O **Módulo do Aluno** do *Feeper* possui uma interface *web*, acessível através de um navegador *web*<sup>3</sup>, no qual o aluno visualiza exercícios atribuídos pelo professor e codifica soluções algorítmicas para os exercícios em um editor de código-fonte disponível na própria interface *web*. O aluno também pode enviar sua solução para avaliação do módulo Juiz Online (descrito na Seção 6.1.1), que retorna ao aluno um *feedback* sobre sua solução. A Figura 3 mostra um fragmento da interface gráfica do Módulo do Aluno do *Feeper* onde o aluno codifica uma solução para um exercício.

Figura 3 – Fragmento da interface gráfica do Módulo do Aluno do *Feeper*



Fonte: (ALVES; JAQUES, 2014).

O **Módulo do Professor** do *Feeper* disponibiliza funcionalidades aos professores, como a manutenção de registros de exercícios, a atribuição de exercícios à turma e o acompanhamento do desempenho dos alunos nos exercícios.

A Figura 4 mostra um fragmento de tela do Módulo do Professor do *Feeper* que visualiza uma tabela contendo o histórico de desempenho dos alunos nos exercícios atribuídos para a turma. O ícone verde indica que determinado exercício foi implementado com sucesso pelo aluno. O ícone em laranja indica que o aluno finalizou o exercício parcialmente. O ícone em vermelho indica erro de compilação na solução para um exercício. A falta de ícone indica que o aluno não submeteu nenhuma solução para determinado exercício.

<sup>3</sup> Disponível em <http://feeper.unisinos.br/>.

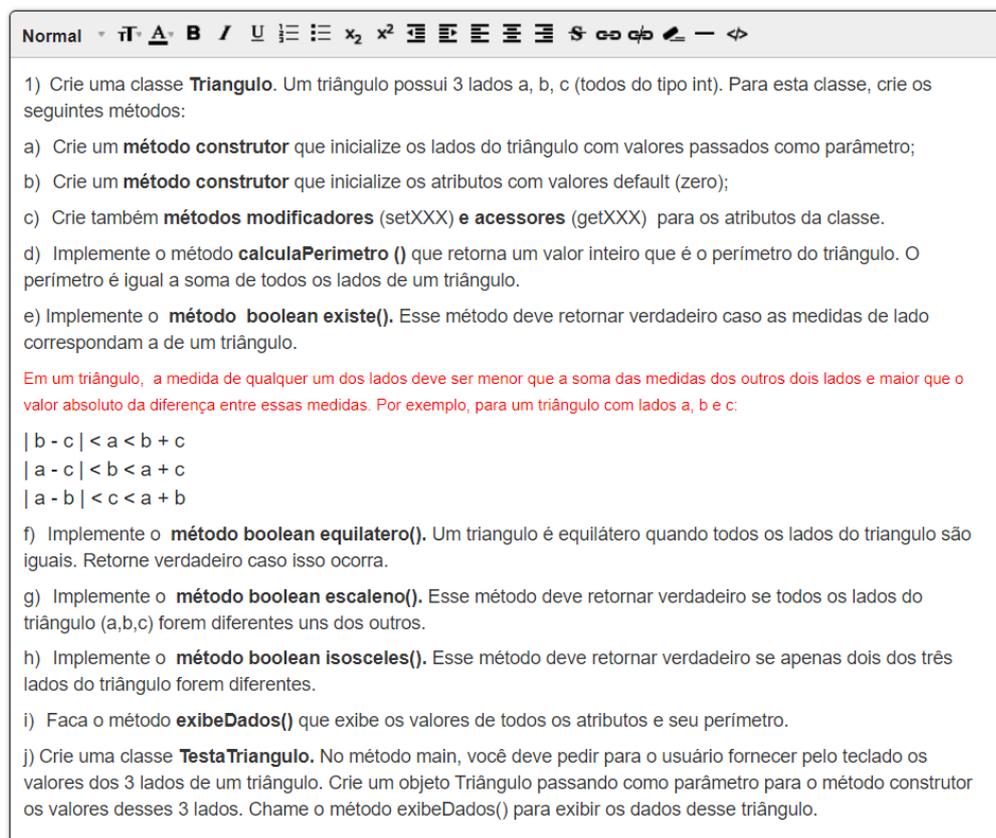
Figura 4 – Fragmento da interface gráfica do Módulo do Professor do *Feeper*



Fonte: Elaborado pelo autor.

O cadastro de um exercício de programação é realizado pelo professor em duas etapas: a) cadastro do enunciado do exercício; e b) cadastro de testes unitários. O enunciado de cada exercício é composto por vários objetivos. Este é um requisito importante da Tese. A Seção 6.3 e o Capítulo 7 trazem mais detalhes sobre os objetivos de exercícios. A Figura 5 apresenta um fragmento de tela do *Feeper* que mostra o enunciado de um exercício, com seus objetivos.

Figura 5 – Fragmento de tela do *Feeper* que mostra a edição de um enunciado de exercício

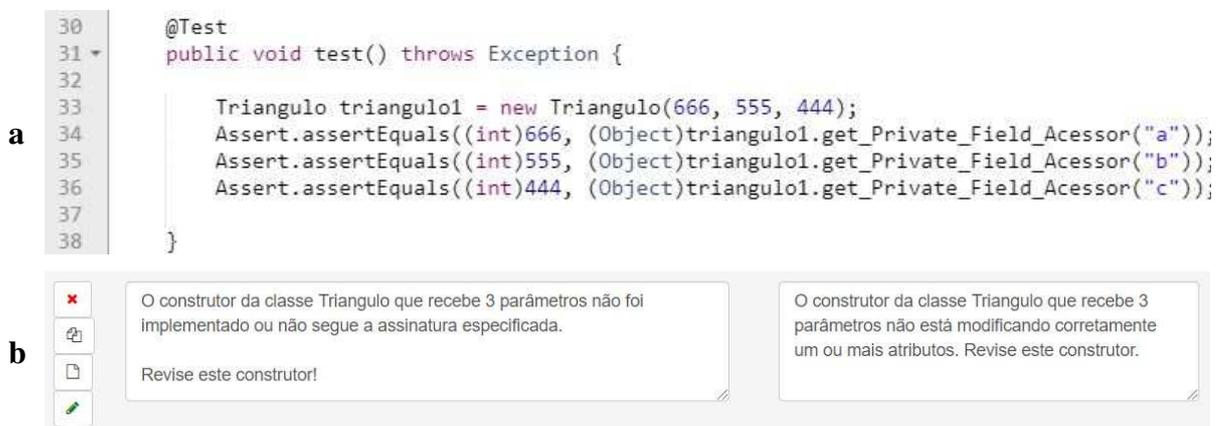


Fonte: (ALVES; JAQUES, 2014).

A segunda etapa no cadastro de um exercício é o registro de testes unitários. Testes unitários são utilizados pelo Juiz Online para a avaliação da solução do aluno. Na Engenharia de Software, um teste unitário é um processo que testa subprogramas individuais de um *software*. Ao invés de testar o *software* como um todo, os testes unitários testam blocos de construção de um programa (MYERS; SANDLER; BADGETT, 2012). No contexto do *Feeper*, cada teste unitário testa um determinado aspecto de um objetivo de exercício. O professor registra, no mínimo, um teste unitário para cada objetivo do exercício.

No *Feeper*, cada teste unitário é implementado com o auxílio do framework *Java JUnit*<sup>4</sup>. A Figura 6 mostra um fragmento de tela de cadastro de um teste unitário. O teste unitário da figura é sobre o objetivo de exercício apresentado no item *a* da Figura 5, em que o aluno deve implementar um método construtor. O teste unitário faz uma chamada ao método construtor que o aluno deveria implementar. Se o aluno implementou o método construtor corretamente, o teste passará com sucesso. Caso contrário, o teste unitário falhará. É possível registrar mensagens de *feedback* para apresentar ao estudante (item *b* da Figura 6).

Figura 6 – Fragmento de tela do *Feeper* que mostra a edição de um teste unitário



Fonte: Elaborado pelo autor.

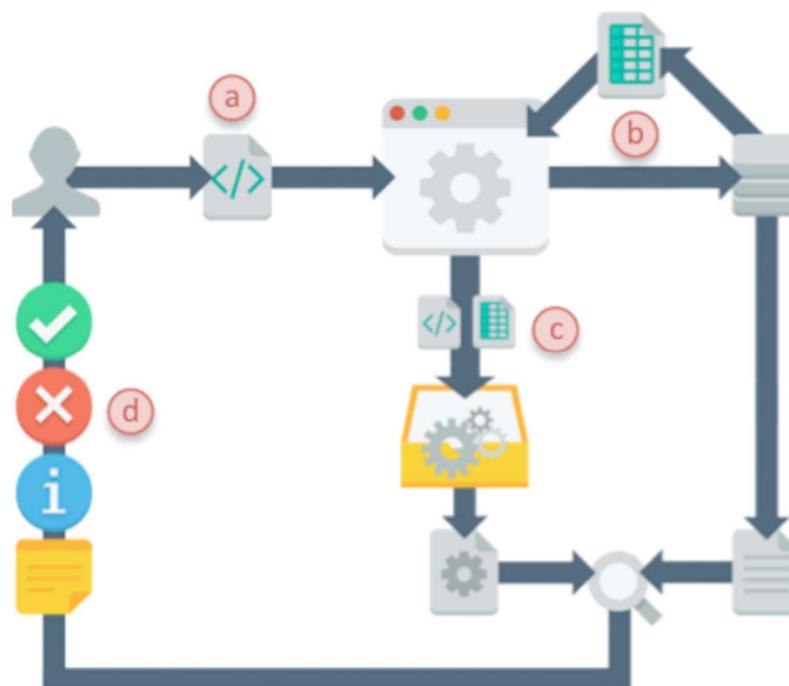
O *Feeper* também possui um módulo que implementa a funcionalidade de **Juiz Online** (JO), que avalia as soluções implementadas pelos alunos para os exercícios e retorna *feedbacks* sobre as soluções. A linguagem de programação suportada pelo *Feeper*, para a codificação dos exercícios, é a linguagem Java.

<sup>4</sup> Disponível em <https://junit.org/>.

### 6.1.1 Juiz Online

O Juiz Online do *Feeper* recebe a solução do aluno para um exercício e aplica os testes unitários que avaliam se a solução para os objetivos do exercício foi corretamente implementada. Uma solução é considerada correta quando todos os testes unitários atribuídos ao exercício **tiveram seus testes satisfeitos**, ou seja, tiveram os resultados esperados. A Figura 7 ilustra o fluxo de processamento do Juiz Online do *Feeper* (ALVES; JAQUES, 2014). O processo inicia quando o aluno submete o código-fonte da sua solução para o exercício (a). Como próximo passo, o JO solicita ao banco de dados os testes unitários cadastrados para o exercício e o banco de dados retorna os testes unitários ao JO (b). Em seguida, o JO reúne o código-fonte do aluno e os testes unitários e executa cada teste unitário, comparando as respostas geradas pela solução do aluno com as respostas esperadas para cada teste unitário (c). Havendo falha na execução de algum teste unitário, como no caso de haver um erro no registro de algum teste unitário (o professor cometeu um erro ao cadastrá-lo), uma mensagem de erro é exibida ao aluno. Caso não ocorra nenhum erro de execução dos testes unitários, o JO verifica se há alguma divergência entre a resposta esperada pelo teste unitário e a resposta gerada pela solução do aluno. Por fim, o juiz online retorna ao *Feeper* o resultado da avaliação (correta ou incorreta) e as mensagens de *feedback* geradas (d).

Figura 7 – Fluxograma de funcionamento do Juiz Online do *Feeper*



Fonte: (ALVES; JAQUES, 2014).

### 6.1.2 Cenário de utilização do *Feeper*

Um exemplo de cenário de utilização do *Feeper* é descrito a seguir. O professor acessa o sistema com usuário e senha, acessa o Módulo do Professor, cria uma turma, cadastra exercícios e alunos e vincula os exercícios e os alunos à turma. O aluno acessa o Módulo do Aluno com usuário e senha e tenta resolver algum exercício. A solução do exercício é submetida para avaliação. O Juiz Online corrige a solução e envia o resultado para a interface do *Feeper*, que apresenta o *feedback* para o aluno. O professor acompanha o histórico de desenvolvimento e desempenho dos alunos nos exercícios, através do Módulo do Professor (ver Figura 4).

## 6.2 Ambiente utilizado na Tese

No primeiro ano da pesquisa da Tese, avaliou-se a possibilidade de utilizar o Módulo do Aluno do *Feeper* como ambiente de programação e de coleta de dados dos alunos. O módulo chegou a ser adaptado para coletar dados de interação dos alunos, e testes piloto foram aplicados em turmas introdutórias de programação de computadores. Nos testes piloto, foi observado que diversos alunos demonstraram falta de interesse na utilização do editor de código da interface *web* do Módulo do Aluno do *Feeper*. Alguns alunos comentavam que preferiam utilizar um editor de código externo, em função de limitações do editor do *Feeper*, não possuindo características de um ambiente de desenvolvimento integrado de software (*IDE*), como apresentar marcas de erros e sugestões de código durante a digitação. Os professores dessas turmas também manifestaram falta de interesse em utilizar o Módulo do Aluno do *Feeper* em suas explanações e escolhiam outros *softwares*, como o *BlueJ* ou o *Eclipse*<sup>5</sup>. O autor da Tese então decidiu não utilizar o Módulo do Aluno do *Feeper* para a coleta de dados. Foi então implementada uma versão modificada da *IDE BlueJ* (versão 4.1.4), chamada pela Tese como ***BlueJ Afetivo***, que substituiu o *Feeper* como ferramenta utilizada pelos alunos para codificar seus exercícios e para a coleta dos dados de interação dos alunos. No entanto, o Módulo do Professor e o Juiz Online do *Feeper* foram ainda utilizados pela Tese. Como o *BlueJ Afetivo* e o *Feeper* são sistemas diferentes, foi necessário implementar a conectividade do *BlueJ Afetivo* com o *Feeper*, para que o *BlueJ Afetivo* pudesse acessar os exercícios cadastrados no Módulo do Professor do *Feeper* e pudesse se comunicar com o Juiz Online do *Feeper* para a avaliação dos exercícios dos alunos.

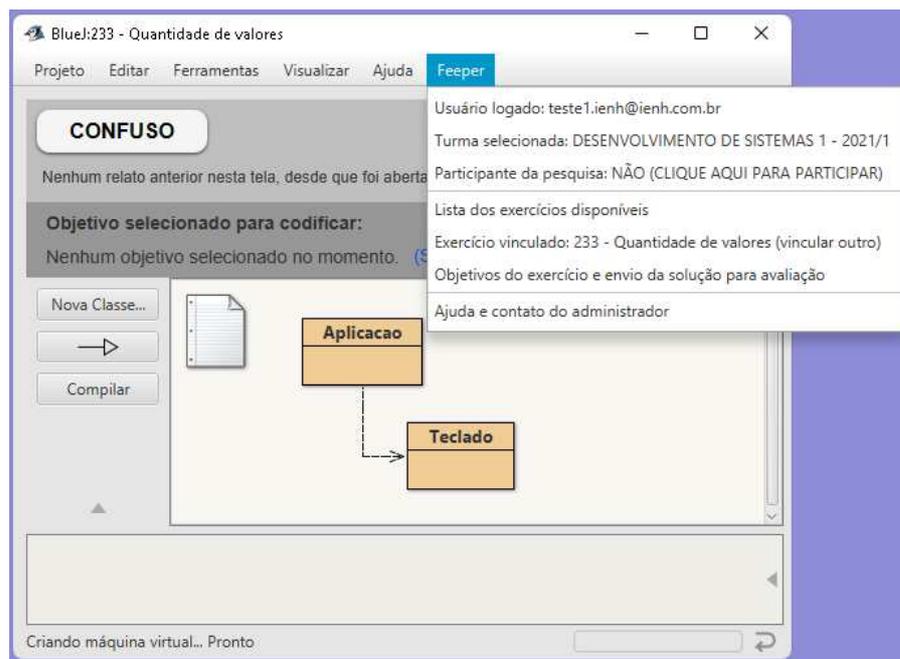
---

<sup>5</sup> Disponível em <https://www.eclipse.org/>.

### 6.3 BlueJ Afetivo

Ambientes computacionais de aprendizagem oferecem diferentes níveis de flexibilidade em relação às possíveis interações dos usuários e como essas interações são orientadas. Por exemplo, alguns ambientes implementam uma abordagem *mastery learning*, no qual o ambiente automaticamente seleciona conteúdos de aprendizagem para o estudante, baseado em medidas sobre o que o aluno já sabe sobre o conteúdo (HEFFERNAN; HEFFERNAN, 2014; RITTER et al., 2016; VANLEHN, 2006). Para a Tese, foi definida uma abordagem *open-ended*, no qual os alunos têm flexibilidade para decidir quais tarefas desejam resolver e a ordem para resolvê-las (BOSCH et al., 2021). Como descrito anteriormente, a Tese implementou o *BlueJ Afetivo*, uma versão modificada da *IDE BlueJ*, um ambiente *open-ended* de programação, desenvolvido para o ensino de programação na linguagem Java (KÖLLING et al., 2003). O *BlueJ Afetivo* coleta dados de interação dos alunos durante a codificação dos exercícios, para a geração de amostras de treino e teste dos modelos de aprendizado de máquina da Tese e *baseline* (a coleta de dados no *BlueJ Afetivo* é detalhada na Seção 8.1). O *BlueJ Afetivo* se conecta ao *Feeper* para acessar os exercícios de programação cadastrados no Módulo do Professor, e acessar o Juiz Online, para a avaliação dos exercícios. Os próximos parágrafos descrevem as principais telas implementadas pela Tese para o *BlueJ Afetivo*.

Figura 8 – Tela principal do *Feeper Afetivo* com menu de opções



Fonte: Elaborado pelo autor.

A Figura 8 mostra a tela principal do *BlueJ Afetivo*, exibida após a autenticação do aluno (conexão do *BlueJ Afetivo* com o *Feeper*). A figura mostra um menu aberto que mostra o *status* de *login*, a turma do aluno e se o aluno é participante da pesquisa. Exibe também opções para o estudante exibir uma lista de exercícios disponíveis. Se o aluno está em meio ao desenvolvimento de um exercício, o menu exibe qual exercício está selecionado. Se não houver nenhum exercício selecionado, o aluno pode fazer o vínculo através do mesmo menu. Também é disponibilizada uma opção para o aluno abrir a tela com a lista de objetivos do exercício. Cada exercício de programação foi disponibilizado ao aluno como uma lista de objetivos textuais, de forma que um exercício só é finalizado pelo aluno depois que ele enviou soluções sintática e funcionalmente corretas para todos os objetivos do exercício.

Figura 9 – Tela de objetivos do exercício



Fonte: Elaborado pelo autor.

A Figura 9 mostra a tela que exibe a lista de objetivos do exercício selecionado. Para cada objetivo da lista, o aluno poderá indicar, através do botão “Vou codificar agora”, que irá tentar implementar uma solução para o objetivo. Quando um objetivo está selecionado para codificação, o aluno poderá cancelar o objetivo selecionado, através do botão “Cancelar seleção

do objetivo”, ou até mesmo enviar a sua solução do objetivo para avaliação do Juiz Online do *Feeper*, através do botão “Enviar solução para avaliação”. O aluno não poderá codificar sem que um objetivo de exercício esteja selecionado. Esta é uma decisão de *design* importante para a Tese, pois através deste mecanismo de seleção de objetivo para codificação, é possível acompanhar qual objetivo de exercício é codificado pelo aluno em determinado momento. Cada objetivo de exercício é vinculado a um ou dois **componentes de conhecimento**, um conceito que será detalhado no Capítulo 7. Em linhas gerais, um componente de conhecimento representa uma unidade de conhecimento do aluno em programação. Identificar qual objetivo de exercício o aluno está codificando permite acompanhar com quais conhecimentos o aluno estava engajado em determinado momento. Este acompanhamento é fundamental para a geração dos atributos de dados sobre o conhecimento do aluno, que compõe as amostras dos modelos de aprendizado de máquina da Tese. A etapa de geração das amostras é detalhada no Capítulo 8.

A decisão de *design* que acrescentou o botão “Vou codificar agora”, descrito anteriormente, possibilitou a identificação dos componentes de componentes que o aluno estava engajado nos exercícios de programação. Este mecanismo do botão substituiu um mecanismo implementado anteriormente no trabalho da Tese, que identifica automaticamente os componentes de conhecimento de programação, durante a codificação. Este mecanismo, chamado de **“modelo de identificação de unidades de conhecimento de programação em processo de aplicação durante a codificação”**, foi descrito em um artigo publicado no SBIE 2020 (Simpósio Brasileiro de Informática na Educação) (KAUTZMANN; JAQUES, 2020). O modelo, baseado em análise sintática, acompanha a codificação do aluno no editor de código, a cada tecla digitada, e identifica os componentes de conhecimento de programação em processo de aplicação. Uma avaliação do modelo mostrou uma boa concordância entre as inferências do modelo e os julgamentos de especialistas, mesmo no cenário mais complexo de identificação, quando o código estava inacabado e com erros ( $kappa = 0,741$ ). No entanto, o autor da Tese decidiu não integrar este mecanismo na versão do *BlueJ Afetivo* utilizada na coleta de dados da Tese. Testes piloto do *BlueJ Afetivo*, com a participação de alunos, realizados no segundo semestre de 2020, mostraram que o modelo de identificação dos componentes de conhecimento acrescentava uma carga de processamento significativa ao editor de códigos do *BlueJ Afetivo*, que já possuía o mecanismo de coleta de teclas digitadas integrado. Para computadores com capacidade intermediária de processamento, em um laboratório de informática, a utilização dos dois mecanismos, em conjunto, já gerava um atraso (*lag*) perceptível pelo usuário no tempo de resposta das teclas pressionadas, o que gerou alguns

comentários negativos dos alunos participantes. Com a declaração da pandemia de COVID-19, e a necessidade de aplicação da coleta de dados da Tese no regime de aulas remotas, os alunos participantes da pesquisa tiveram que instalar o *BlueJ* Afetivo em seus computadores pessoais. Como não seria possível garantir uma boa capacidade de processamento nos computadores dos alunos participantes, foi avaliado que a experiência dos alunos na codificação dos exercícios de programação, ao acrescentar os dois mecanismos (coleta de teclas pressionadas + modelo de identificação de componentes de conhecimento) ao editor de códigos, seria inviável.

A Figura 10 mostra um fragmento da mesma tela apresentada na Figura 9, mas após o aluno ter enviado para avaliação do Juiz Online a sua solução para o objetivo selecionado. No exemplo da figura, a solução do aluno foi enviada para avaliação do Juiz Online, e a resposta obtida foi de que o objetivo não foi alcançado. O *BlueJ* Afetivo exibe um *feedback* geral ao aluno, sem exibir qual erro específico foi cometido pelo estudante.

Figura 10 – Envio da solução de objetivo para avaliação

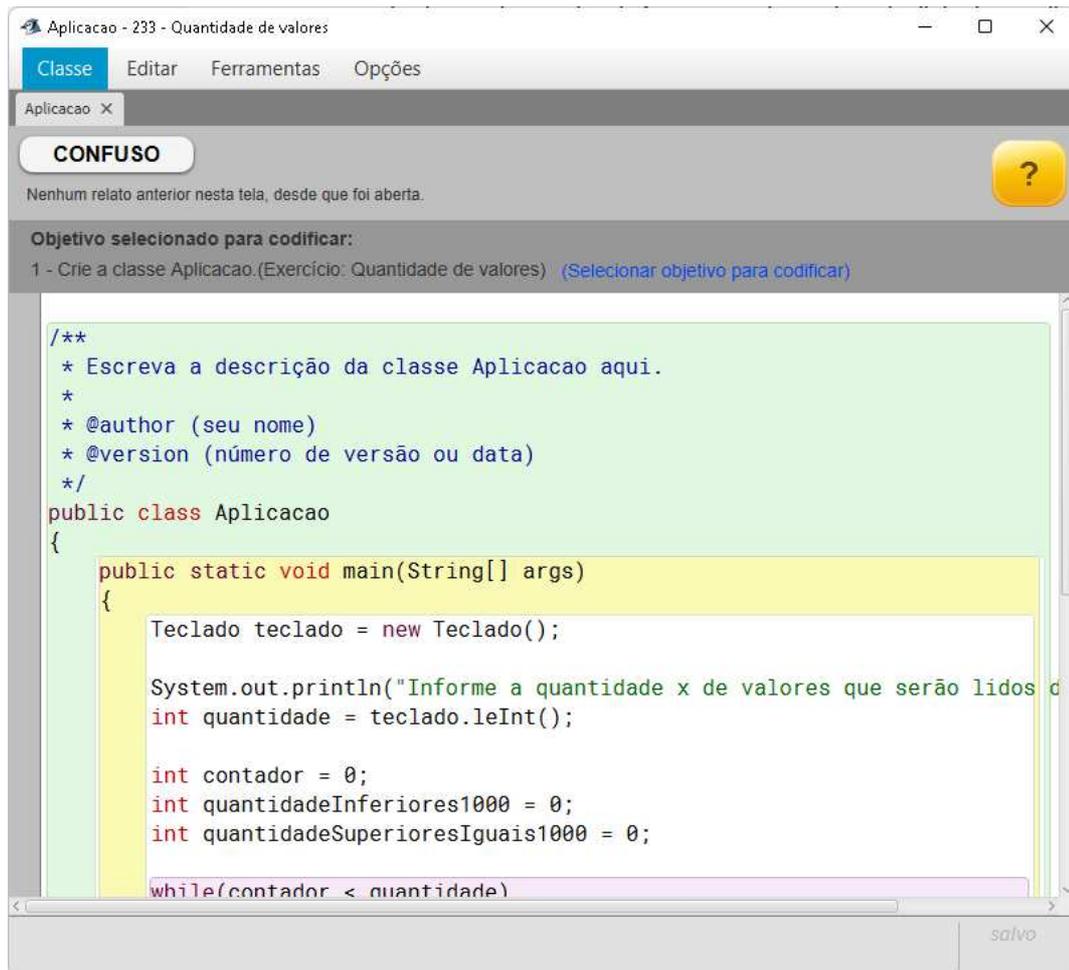


Fonte: Elaborado pelo autor.

Como será detalhado no Capítulo 7, é no momento que o aluno submete para avaliação do Juiz Online sua solução para o objetivo de exercício que os componentes de conhecimento atrelados ao objetivo de exercício são atualizados pelo modelo de inferência do conhecimento do aluno.

A Figura 11 mostra a tela do editor de código do *BlueJ* Afetivo, onde o aluno codifica sua solução para o exercício, na linguagem Java. Nesta tela, foi implementado um mecanismo que coleta informações sobre cada tecla digitada no editor e cada ação do *mouse*.

Figura 11 – Tela do editor de código do *BlueJ* Afetivo



Fonte: Elaborado pelo autor.

Em cada tela do *BlueJ Afetivo*, na parte superior, foram disponibilizados botões “CONFUSO” para os alunos fazerem o autorrelato de confusão, durante a interação com o ambiente. Estes autorrelatos são utilizados para a geração dos rótulos de emoção que são vinculados às amostras de dados dos modelos de aprendizado de máquina supervisionada. Este método de coleta de relatos de confusão é detalhado no Capítulo 8. Além disso, em cada tela, foi disponibilizado um painel que destaca qual o objetivo selecionado para codificação. Com um objetivo selecionado, o painel disponibiliza uma opção para o aluno selecionar outro objetivo (troca de objetivo), se desejar. Se nenhum objetivo estiver selecionado em dado momento, o aluno terá neste painel uma opção para selecionar o objetivo para codificação. O ambiente não permite o aluno codificar sem ter um objetivo de exercício selecionado. Em cada tela, também é disponibilizado um botão de ajuda “?” em que o aluno poderá rever as orientações específicas para participação na pesquisa, que foram explicadas nos termos de consentimento e assentimento, e através de vídeo disponibilizado pelo autor da Tese.

#### 6.4 Avaliação de usabilidade do *BlueJ* Afetivo

O autor da Tese solicitou aos alunos participantes da pesquisa que respondessem a um questionário que avalia a usabilidade percebida do *BlueJ* Afetivo. Os alunos responderam ao questionário no final da sua participação na pesquisa, nos meses de junho e julho de 2021. A aplicação do questionário buscou verificar se ocorreram eventuais problemas de usabilidade que poderiam prejudicar a experiência dos participantes e, de alguma forma, afetar a confiabilidade dos dados coletados.

Foi aplicado o questionário CSUQ (*Computer System Usability Questionnaire*), um questionário padronizado projetado para a avaliação da usabilidade percebida por usuários de sistemas de computador (LEWIS, 1995; SAURO; LEWIS, 2012). O CSUQ é composto por 19 afirmações positivas sobre aspectos de usabilidade do sistema de computador. Suas alternativas de resposta são baseadas em uma escala de 7 pontos em que os participantes avaliam sua concordância ou discordância com cada afirmação, em um intervalo de escores inteiros entre 1 e 7. Uma resposta com o escore 1 indica que o usuário concorda fortemente com a afirmação e uma resposta com o escore 7 indica que o usuário discorda fortemente da afirmação. O escore 4 indica que o usuário não concorda e nem discorda da afirmação. Escores mais baixos, mais próximos de 1, indicam a satisfação do usuário com a usabilidade do sistema. Escores mais altos, mais próximos de 7, indicam insatisfação (LEWIS, 1995; SAURO; LEWIS, 2012). Para cada uma das afirmações, o usuário é encorajado a escrever algum comentário sobre sua resposta (LEWIS, 1995). O CSUQ também prevê como alternativa de resposta a opção NA (não se aplica), quando o usuário avalia não ter condições de dar uma resposta à afirmação (LEWIS, 1995). O CSUQ é um dos métodos mais utilizados na literatura para medir a usabilidade de sistemas de computador (LEWIS, 2018). Ele é baseado no questionário PSSUQ (*Post-Study System Usability Questionnaire*), sendo ambos praticamente idênticos. No entanto, o CSUQ foi escolhido por ser indicado para ser aplicado de forma *online* e para diferentes grupos de usuários (SAURO; LEWIS, 2012).

A Tabela 5 apresenta as 19 afirmações do questionário CSUQ. As afirmações foram traduzidas pelo autor da Tese para a língua portuguesa, uma vez que não foi encontrada uma versão do questionário em português. No entanto, essa versão em português não foi validada.

Em função de um erro do autor da Tese na elaboração do questionário no *Google Forms*, a opção NA (não se aplica), prevista para ser fornecida como alternativa de resposta para cada afirmação, não foi disponibilizada.

#### 6.4.1 Resultados

Dos 62 alunos participantes da pesquisa, 47 responderam ao questionário CSUQ (75,8%). A Tabela 5 mostra os escores individuais obtidos para cada uma das 19 afirmações do questionário. Cada escore é uma média das respostas de todos os alunos respondentes. Em 16 afirmações, os escores ficaram abaixo de 3,0, indicando que, em média, os alunos concordam com estas afirmações positivas sobre a usabilidade do sistema. Em apenas três afirmações, os escores ficaram entre 3,0 e 4,0, ainda uma concordância com as afirmações, mas não tão fortes, próximas do escore 4 (não concorda e nem discorda). Nenhuma afirmação obteve escore de discordância.

Tabela 5 - Escores para as afirmações do CSUQ

#	<i>Afirmção</i>	<i>Escore</i>
1	Em geral, estou satisfeito com a facilidade de utilização do sistema.	2,53191
2	O sistema é simples de utilizar.	1,87234
3	Consigo completar os exercícios usando este sistema.	1,85106
4	Consigo completar rapidamente os exercícios usando este sistema.	2,93617
5	Consigo completar os exercícios com eficiência usando este sistema.	2,63830
6	Me sinto confortável usando este sistema.	2,46809
7	Foi fácil aprender a utilizar este sistema.	1,76596
8	Acredito que me tornei eficiente rapidamente ao usar este sistema.	3,00000
9	O sistema dá mensagens de erro que me indicam claramente como resolver problemas.	3,63830
10	Sempre que eu cometi um erro usando o sistema, consegui recuperar de forma rápida e fácil.	3,12766
11	As informações fornecidas pelo sistema (como nas mensagens na tela) são claras.	2,42553
12	É fácil encontrar as informações que eu preciso.	2,48936
13	As informações fornecidas pelo sistema são fáceis de entender.	2,44681
14	As informações fornecidas pelo sistema são efetivas em me ajudar a completar os exercícios.	2,63830
15	A organização das informações nas telas do sistema é clara.	2,17021
16	A interface do sistema é agradável. Observação: a interface inclui tudo aquilo que você usa para interagir com o sistema, como o teclado, o mouse e as telas.	2,59574
17	Eu gosto de usar a interface deste sistema.	2,89362
18	O sistema tem todas as funções e capacidades que eu esperava.	2,55319
19	Em geral, estou satisfeito com este sistema.	2,38298

Fonte: Elaborado pelo autor.

Além dos escores individuais para cada afirmação, o CSUQ também produz quatro escores sobre conjuntos de afirmações. O primeiro escore, o GERAL, é sobre a satisfação geral do usuário com o sistema, e é calculado como a média dos escores de todas as afirmações (1 a 19). O segundo escore, o SISUTIL, é sobre a utilidade do sistema, e é calculado como a média dos escores das afirmações 1 a 8. O terceiro escore, o QUALINFO, é sobre a qualidade da informação fornecida pelo sistema, e é calculado como a média dos escores das afirmações 9 a

15. O quarto escore, o QUALINTER, é sobre a qualidade da interface do sistema, e é calculado como a média dos escores das afirmações 16 a 18.

Tabela 6 – Escores nos grupos de afirmações do CSUQ

<i>Tipo</i>	<i>Escore</i>
GERAL	2,54871
SISUTIL	2,38298
QUALINFO	2,70517
QUALINTER	2,68085

Fonte: Elaborado pelo autor.

A Tabela 6 mostra os resultados para os quatro escores. Todos os escores obtiveram valor entre 2,0 e 3,0, indicando que, em média, os alunos concordaram com as afirmações positivas para cada um dos escores. O escore que obteve pior resultado foi o QUALINFO, sobre a qualidade da informação fornecida pelo sistema.

## 7 MODELO DE CONHECIMENTO DO ALUNO DESENVOLVIDO PARA A TESE

Este capítulo descreve um modelo de inferência de estimativas do conhecimento do aluno em programação de computadores, desenvolvido e avaliado pela Tese, chamado de **modelo de conhecimento do aluno**. As estimativas de conhecimento do aluno, geradas pelo modelo, são utilizadas para derivar atributos de dados sobre o conhecimento do aluno, usadas nas amostras de treinamento e teste dos modelos de aprendizado de máquina da Tese. A derivação desses atributos de dados para as amostras é descrita mais adiante, no Capítulo 8.

A Seção 7.1 apresenta apontamentos sobre modelos de conhecimento do aluno descritos na literatura e traz justificativas para o tipo de modelo selecionado para implementação na Tese. A Seção 7.2 apresenta o conceito de “componente de conhecimento”, fundamental para a compreensão sobre quais aspectos do conhecimento do aluno são modelados. A Seção 7.3 descreve o *Bayesian Knowledge Tracing* (BKT), o tipo de modelo de conhecimento do aluno implementado na Tese. A Seção 7.4 descreve como foi realizada a coleta de evidências de aplicação de conhecimento do aluno, no *BlueJ* Afetivo. Estas evidências são utilizadas pelo modelo BKT para atualizar as estimativas de conhecimento dos estudantes nos componentes de conhecimento. O treinamento e validação do modelo de conhecimento do aluno são descritos na Seção 7.5. Os resultados da validação são apresentados na Seção 7.5.1. Os parâmetros gerados no treinamento do modelo são apresentados na Seção 7.5.2.

### 7.1 Modelos de conhecimento do aluno

A importância da aprendizagem individualizada é reconhecida desde os estudos de Bloom (1984), que mostraram que os alunos aprendem mais numa condição *one-to-one-tutoring* (um tutor para cada aluno). Mais recentemente, estudos mostraram que sistemas tutores inteligentes<sup>1</sup> são tão efetivos quanto tutores humanos em uma instrução individualizada (MA et al., 2014; VANLEHN, 2011). Ambientes adaptativos de aprendizagem, como os sistemas tutores inteligentes, possuem um componente chamado **Modelo de Aluno**. Este modelo tem a função de modelar diferentes características do aprendiz, como seu conhecimento, habilidades e estados afetivos (VANLEHN, 2006; WOOLF, 2009). Ele também pode inferir e manter informações sobre o desempenho do aluno, como escores sobre testes, quantidade de erros,

---

<sup>1</sup> Sistemas tutores inteligentes são softwares educacionais que simulam capacidades cognitivas do aluno, mantém um Modelo do Aluno, e adaptam as decisões pedagógicas de acordo com as necessidades do aluno. O sistema pode ajustar sua interface, mensagens, tipos e complexidades dos exercícios ofertados, adaptando as experiências de aprendizagem dos alunos (VANLEHN, 2006; WOOLF et al., 2009).

acertos ou ajudas solicitadas em tarefas e a duração de tempo que ele se dedicou nas tarefas (VANLEHN, 2006). Modelar características do aluno, como suas emoções e seu conhecimento no domínio, é uma característica fundamental para que esses ambientes consigam se adaptar às características dos alunos e proporcionar um ensino individualizado.

A Tese tem interesse em modelar uma das características do Modelo de Aluno: o conhecimento do estudante. A hipótese da Tese é que estimativas sobre o conhecimento do estudante, ao longo do desenvolvimento de tarefas de aprendizagem de programação de computadores, podem ajudar modelos computacionais de detecção de confusão a identificar a confusão do estudante no decorrer das tarefas de aprendizagem. No entanto, modelar o conhecimento do aprendiz durante a realização de tarefas de aprendizagem é um objetivo desafiador. Uma das dificuldades é que o conhecimento do aluno é avaliado à medida que ele vai adquirindo conhecimento, durante a realização das tarefas. Isso é mais desafiador do que inferir o conhecimento em testes de conhecimento realizados após o processo de aquisição de conhecimento. Este desafio é ainda maior se o modelo de conhecimento do aluno é integrado em ambientes adaptativos de aprendizagem, onde as estratégias de ensino entregues pelo ambiente podem mudar à medida que o aluno aprende (VANLEHN, 2006).

A Tese implementou um modelo de conhecimento do aluno em programação de computadores baseado no modelo *Bayesian Knowledge Tracing* (BKT). Diversos autores concordam que o BKT tem sido o modelo mais empregado para modelar o conhecimento do aluno e possui bons resultados (AGARWAL; BAKER, 2019; EAGLE et al., 2017; SLATER; BAKER, 2018). O BKT é um tradicional método de inferência de conhecimento do aluno (CORBETT; ANDERSON, 1995) que tem sido usado para modelar o conhecimento de estudantes em sistemas tutores inteligentes há mais de 20 anos (AGARWAL; BAKER, 2019; CORBETT; ANDERSON, 1995; KOEDINGER; CORBETT, 2006; SLATER; BAKER, 2018; YUDELSON; KOEDINGER; GORDON, 2013) em domínios como matemática (KOEDINGER; CORBETT, 2006), leitura (BECK; CHANG, 2007), genética (CORBETT et al., 2010), e programação de computadores (CORBETT; ANDERSON, 1995). O modelo apresenta boa correlação com resultados de testes de desempenho do aluno (BAKER et al., 2010; PARDOS; HEFFERNAN, 2011) e sua capacidade preditiva é comparável a modelos de inferência de conhecimento recentes e complexos (BAKER et al., 2011; GONG; BECK; HEFFERNAN, 2010; GOWDA et al., 2011; KHAJAH; LINDSEY; MOZER, 2016; LIN; CHI, 2018), como os modelos DKT (*Deep Knowledge Tracing*), baseados em redes neurais (KHAJAH; LINDSEY; MOZER, 2016). Um estudo mostrou que quando comparado ao BKT clássico, o DKT tende a apresentar desempenho superior, mas quando extensões são

adicionadas ao BKT clássico, com o objetivo de tratar suas limitações, o desempenho do BKT pode ser similar, e em alguns casos até mesmo superior ao DKT (KHAJAH; LINDSEY; MOZER, 2016).

Uma das limitações não tratadas pelo BKT clássico é o efeito de recência. O modelo BKT clássico não considera o esquecimento do aluno. Uma vez que o BKT infere que o aluno aprendeu um componente de conhecimento, mesmo que o aluno apresente uma sequência de aplicações incorretas de determinado componente de conhecimento, a estimativa de conhecimento do aluno no componente de conhecimento não regride. Incorporar o esquecimento ao modelo BKT poderia tornar o modelo sensível a eventos recentes, o que aparentemente já é tratado pelos modelos de redes neurais DKT (KHAJAH; LINDSEY; MOZER, 2016). A implementação do modelo BKT usada na Tese considera uma extensão que modela o esquecimento do aluno e seus efeitos de recência.

Os pesquisadores em modelagem de conhecimento destacam uma vantagem do BKT em relação aos modelos DKT: o BKT proporciona uma melhor interpretação do processo de inferência do conhecimento em relação a modelos baseados em redes neurais, que funcionam como “caixas pretas” (LIN; CHI, 2018; SLATER; BAKER, 2018). Isso porque todas as etapas de inferência do modelo BKT são transparentes. O modelo BKT é tão simples que suas inferências podem ser obtidas através de fórmulas em uma planilha do Excel. Além disso, os parâmetros que compõe o modelo BKT são fáceis de interpretar. O modelo BKT clássico (sem extensões) possui quatro parâmetros, a saber: a) **a probabilidade de o aluno dominar o conhecimento antes da primeira oportunidade de aplicá-lo**; b) **a probabilidade de o aluno transitar do estado de não domínio para o estado de domínio, após uma oportunidade de aplicar o conhecimento**; c) **a probabilidade de o aluno aplicar incorretamente o conhecimento quando estiver no estado de domínio do conhecimento**; e d) **a probabilidade de o aluno aplicar corretamente o conhecimento quando estiver no estado de não domínio do conhecimento**. Estes parâmetros são detalhados na Seção 7.3. A facilidade de explicar o modelo e seus parâmetros, juntamente do seu bom desempenho, foram as principais justificativas para a escolha desse modelo para a Tese. As informações obtidas pelo modelo BKT e seus parâmetros permitem derivar diversos atributos de dados sobre estimativas de conhecimento do aluno, que serão utilizadas nas amostras de treino e testes dos modelos da Tese, como será explicado no Capítulo 8.

Os parâmetros do modelo BKT precisam ser ajustados antes de o modelo fazer suas inferências. A adequação dos parâmetros pode ser realizada a nível do conhecimento, como no modelo padrão de BKT, ou a nível de aluno (individualização). Na primeira abordagem, os

parâmetros ajustados são atribuídos a todos os alunos, sem individualização, o que presume que os alunos possuem características homogêneas. Na segunda abordagem (individualização dos parâmetros), os parâmetros são ajustados para cada aluno, individualmente, o que presume que os estudantes possuem características heterogêneas (SLATER; BAKER, 2018).

## 7.2 Componentes de conhecimento

Uma definição importante para a compreensão do modelo de conhecimento do aluno é o de **componente de conhecimento (CC)**. A Tese considera a conceitualização apresentada por Koedinger, Corbett e Perfetti (2012) para uma noção de componente de conhecimento. Os autores definem **um componente de conhecimento como uma unidade adquirida de função ou estrutura cognitiva e que pode ser inferida através do desempenho observado em um conjunto de tarefas** (KOEDINGER; CORBETT; PERFETTI, 2012). São exemplos de componentes de conhecimento modelados na Tese, no domínio de programação, os componentes de “Repetição”, “Seleção de caso” e “Classe”. Estimativas sobre estes componentes de conhecimento são inferidos pelo modelo de conhecimento do aluno, a partir do desempenho observado em tarefas de programação no *BlueJ* Afetivo. As próximas seções descrevem os demais componentes de conhecimento modelados na Tese e a forma como o modelo de conhecimento do aluno infere suas estimativas.

A Tabela 7 mostra os componentes de conhecimento definidos pela Tese para o modelo de conhecimento do aluno. Na literatura, os trabalhos que modelaram componentes de conhecimento no domínio de programação de computadores definiram componentes de conhecimento vinculados a estruturas, conceitos ou regras de programação, como estruturas de seleção, estruturas de repetição, atribuição de valores a variáveis, entre outros (KASURINEN; NIKULA, 2009; PENMETSA; SHI; PRICE, 2021; RAPOSO; MARANHÃO; SOARES NETO, 2019; WANG et al., 2017). Da mesma forma como nos trabalhos relacionados, a Tese definiu componentes de conhecimento relacionados a conceitos de programação. Os componentes foram definidos apenas após os professores das turmas participantes da pesquisa definirem todos os exercícios (e seus objetivos) atribuídos às suas turmas. O autor da Tese, que possui experiência de mais de 20 anos em programação e possui certificação oficial na linguagem Java, identificou os conceitos de programação existentes nos exercícios definidos pelos professores. A definição também foi baseada nas ementas das disciplinas das turmas participantes. Todas as disciplinas das turmas participantes abordam conteúdos introdutórios de programação de computadores. A Tabela 7 também mostra a quantidade de objetivos dos

exercícios atribuídos pelos professores das turmas participantes são vinculados a cada componente de conhecimento identificado pela Tese. Esta informação é importante, pois componentes de conhecimento com poucos objetivos atrelados poderão ter problemas de ajuste, pois terão menos evidências de aplicação de conhecimento. Uma discussão nesse sentido é apresentada na Seção 7.5.1, quando são apresentados os resultados da avaliação do modelo.

Tabela 7 – Componentes de conhecimento do modelo de conhecimento do aluno

#	Componente de conhecimento	Objetivos vinculados
1	Atributo de classe	83
2	Classe	214
3	Método construtor com inicialização de atributo	52
4	Método construtor com inicialização de atributo do tipo array	6
5	Método construtor com inicialização de atributo herdado	9
6	Método construtor padrão	15
7	Método de classe com parâmetro e com retorno	180
8	Método de classe com parâmetro e sem retorno	14
9	Método de classe sem parâmetro e com retorno	124
10	Método de classe sem parâmetro e sem retorno	18
11	Método principal	96
12	Métodos de modificação e de acesso	51
13	Subclasse	15
14	Método toString	8
15	Entrada de array, instanciação, invocação de método e saída de array	10
16	Entrada, instanciação, invocação de método e saída	41
17	Entrada, invocação de método e saída	25
18	Escrita de array	32
19	Algoritmo formal	1
20	Incremento/decremento de atributo	21
21	Invocação de método	8
22	Leitura de array de tipo não primitivo	26
23	Leitura de array de tipo primitivo	35
24	Leitura de atributo	4
25	Operação aritmética	96
26	Repetição	19
27	Seleção de caso	44
28	Seleção de caso com operação aritmética	8
29	Verificação de tipo de objeto	3

Fonte: Elaborado pelo autor.

### 7.2.1 Relação entre os componentes de conhecimento e os objetivos de exercício

Esta Seção descreve a relação entre os componentes de conhecimento e os objetivos dos exercícios atribuídos aos alunos no *BlueJ* Afetivo. Como descrito no capítulo anterior, os enunciados dos exercícios foram organizados em objetivos. Cada objetivo de exercício é vinculado a um ou, no máximo, dois componentes de conhecimento.

Quadro 4 – Componentes de conhecimento vinculados a objetivos de um exercício

Objetivo de exercício	CC primário	CC Secundário										
Crie a classe IMC.	Classe											
Na classe IMC, crie os seguintes atributos privados: peso e altura, ambos do tipo double. O atributo peso representa o peso de uma pessoa, em quilos. O atributo altura representa a altura de uma pessoa, em metros.	Atributo de classe											
Na classe IMC, crie um método construtor com dois parâmetros, ambos do tipo double. O primeiro parâmetro representa o peso de uma pessoa e o segundo parâmetro representa a altura da mesma pessoa. O construtor deverá atribuir o valor do primeiro parâmetro ao atributo peso e o valor do segundo parâmetro ao atributo altura.	Método construtor com inicialização de atributo											
Na classe IMC, crie um método público com o nome classificacao. O método não possui parâmetros e tem retorno do tipo String. O método deverá calcular o IMC ( $\text{peso} / (\text{altura} * \text{altura})$ ) com os valores dos atributos e retornar uma String com a classificação de IMC, de acordo com a tabela abaixo.	Método de classe sem parâmetro e com retorno	Seleção de caso com operação aritmética										
<table border="1"> <thead> <tr> <th>IMC</th> <th>Classificação</th> </tr> </thead> <tbody> <tr> <td>entre 0,0 (inclusive) e 18,5 (exclusive)</td> <td>magro</td> </tr> <tr> <td>entre 18,5 (inclusive) e 25,0 (exclusive)</td> <td>normal</td> </tr> <tr> <td>entre 25,0 (inclusive) e 30,0 (exclusive)</td> <td>sobrepeso</td> </tr> <tr> <td>a partir de 30,0 (inclusive)</td> <td>obeso</td> </tr> </tbody> </table>			IMC	Classificação	entre 0,0 (inclusive) e 18,5 (exclusive)	magro	entre 18,5 (inclusive) e 25,0 (exclusive)	normal	entre 25,0 (inclusive) e 30,0 (exclusive)	sobrepeso	a partir de 30,0 (inclusive)	obeso
IMC			Classificação									
entre 0,0 (inclusive) e 18,5 (exclusive)			magro									
entre 18,5 (inclusive) e 25,0 (exclusive)			normal									
entre 25,0 (inclusive) e 30,0 (exclusive)	sobrepeso											
a partir de 30,0 (inclusive)	obeso											
Observação: O método deve retornar uma String com a classificação do IMC, exatamente como aparece na tabela. Exemplo: retornar a String "sobrepeso".												
Crie uma classe Teste.	Classe											
Na classe Teste, crie um método main (public static void main(String[] args)). No escopo do método main, implemente o que se pede: faça a leitura de 2 valores do teclado do usuário (use a classe Teclado), ambos do tipo double, sendo o primeiro valor referente ao peso de uma pessoa e o segundo valor referente à altura de uma pessoa; instancie a classe IMC (crie um objeto IMC), e guarde a referência em uma variável; invoque o método classificacao do objeto instanciado, obtenha seu retorno e o escreva na saída do sistema (System.out.println).	Método principal	Entrada, instanciação de objeto, invocação de método e saída										

Fonte: Elaborado pelo autor.

A limitação da quantidade de componentes de conhecimento atribuídos a cada objetivo de exercício tem relação com a forma como o Juiz Online do *Feeper* avalia uma solução de objetivo de exercício. A técnica utilizada pelo Juiz Online para avaliação é a de teste unitário. Testes unitários avaliam apenas a funcionalidade do código (a solução do aluno), ou seja, avaliam apenas se a saída do código é correta para determinada entrada. No entanto, testes unitários não avaliam a estrutura interna dos códigos. Isso explica por que os enunciados dos

exercícios foram fragmentados em objetivos, para que cada parte (objetivo) do exercício fosse avaliado individualmente por testes unitários. Cada teste unitário avalia um componente de conhecimento. Se o teste unitário falhar, obtém-se uma evidência de aplicação incorreta do componente de conhecimento. Se o teste unitário apresentar sucesso, obtém-se uma evidência de aplicação correta do componente de conhecimento. Em outras palavras, fragmentar o exercício em partes menores (objetivos) foi uma decisão importante porque um exercício de programação pode envolver a aplicação de diversos componentes de conhecimento. Avaliar cada parte do exercício possibilitaria identificar evidências de aplicação de componentes de conhecimento individuais. Tecnicamente, foi possível vincular cada objetivo de exercício a, pelo menos, um componente de conhecimento. No entanto, verificou-se que, para alguns objetivos de exercício, seria possível vincular dois componentes de conhecimento, o **componente de conhecimento primário** e o **componente de conhecimento secundário**.

O autor da Tese acredita que a quebra dos exercícios em objetivos, uma decisão técnica importante para a implementação do modelo de conhecimento do aluno, não traria efeitos negativos para a aprendizagem dos estudantes no *BlueJ Afetivo*, baseado no princípio de segmentação, apresentado por Richard Mayer, no contexto de tarefas multimídia (MAYER et al., 2018). O princípio define que as pessoas aprendem melhor quando uma lição é apresentada ao aluno em segmentos, no ritmo do usuário, ao invés de ser apresentado como uma única unidade de apresentação (MAYER et al., 2018).

O Quadro 4 mostra os enunciados de objetivos de um exercício atribuído por um professor em uma das turmas participantes. Cada objetivo pode estar vinculado a um ou, no máximo, dois componentes de conhecimento. Cada objetivo de exercício está atrelado a um ou mais testes unitários. Cada teste unitário é atrelado a determinado componente de conhecimento do objetivo. Se o teste unitário para determinado componente de conhecimento falhar, obtém-se uma evidência de aplicação incorreta do componente de conhecimento. Senão, obtém-se uma evidência de aplicação correta.

Ainda sobre os enunciados dos objetivos de exercícios, é necessário trazer uma discussão sobre um importante aspecto tratado pelo trabalho: a padronização dos enunciados. Ao definir os enunciados dos objetivos de exercício, com o auxílio dos professores das turmas participantes, o pesquisador buscou manter um padrão sobre o nível de definição dos problemas e os domínios envolvidos nos enunciados. Os padrões seguiram as definições sobre “problemas mal definidos” e “domínios mal definidos” apresentados por Lynch et al (2009). O objetivo foi gerar enunciados que não seguissem as definições de “problemas mal definidos” e “domínios mal definidos”, de forma que tanto os problemas quanto os domínios dos enunciados fossem

“bem definidos”. Por exemplo, buscou-se deixar bastante claro cada objetivo de exercício, de forma que a solução do aluno deveria atender uma saída esperada e claramente definida no enunciado, como em “Crie a classe IMC” e “O construtor deverá atribuir o valor do primeiro parâmetro ao atributo peso e o valor do segundo parâmetro ao atributo altura.” (ver Quadro 4). Este exemplo corresponde a um critério adotado para “problema bem definido”. Um exemplo de critério adotado para “domínio bem definido” é que sempre que o enunciado trouxesse algum conceito do domínio do problema, uma definição clara para o conceito precisaria ser apresentada no enunciado. Por exemplo, se o objetivo de exercício fosse sobre a solução de código apresentar a classificação de IMC (índice de massa corporal) na tela do usuário, o enunciado deveria trazer detalhes sobre as regras de classificação de IMC (ver Quadro 4), não deixando ao aluno a tarefa de encontrar uma caracterização para o domínio (IMC). A padronização dos enunciados dos objetivos de exercícios buscou evitar que o aluno ficasse confuso em função da leitura do enunciado de objetivo de exercício, pois o trabalho buscou identificar padrões de confusão durante a codificação dos objetivos de exercício, e não durante a leitura dos enunciados dos objetivos. Essa estratégia buscou evitar, por exemplo, que um aluno que domina os conhecimentos para codificar um objetivo de exercício, ficasse confuso por não ter entendido um enunciado mal definido, o que certamente geraria algum ruído para os modelos classificadores de confusão.

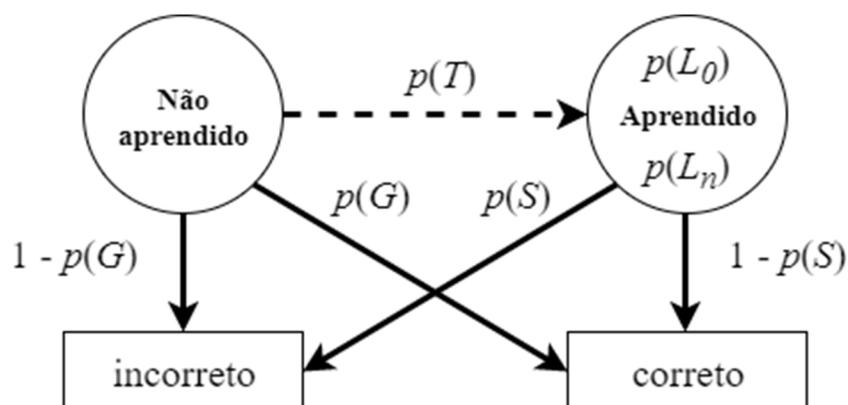
### **7.3 Modelo *Bayesian Knowledge Tracing* (BKT)**

O modelo BKT, do inglês *Bayesian Knowledge Tracing* (CORBETT; ANDERSON, 1995), é um método para modelagem de conhecimento largamente usado em sistemas adaptativos inteligentes, como no sistema tutor inteligente Cognitive Tutor, utilizado por centenas de milhares de alunos a cada ano nos Estados Unidos (KOEDINGER; CORBETT, 2006). O BKT é considerado um modelo efetivo para modelar o conhecimento dos estudantes (BAKER et al., 2010; CORBETT; ANDERSON, 1995; PARDOS; HEFFERNAN, 2011) e tem sido aplicado, juntamente de suas variações, em diversos trabalhos (BECK; CHANG, 2007; CORBETT et al., 2010; CORBETT; ANDERSON, 1995; KOEDINGER; CORBETT, 2006; SLATER; BAKER, 2018). O BKT funciona como um modelo HMM (*Hidden Markov Model*), que tenta estimar uma única variável binária latente (o indivíduo domina (1) ou não domina (0)) para cada componente de conhecimento, ou seja, um HMM com dois estados escondidos. A variável latente é estimada com base em ocorrências de uma variável observável também binária (o aluno aplicou corretamente o conhecimento quando teve a oportunidade (1) ou não

(0)). O BKT calcula a probabilidade de o aluno dominar um determinado conhecimento ou habilidade a partir de cada oportunidade de o estudante aplicá-lo. Nesta Tese, considere “um determinado conhecimento ou habilidade” como um componente de conhecimento (descrito na Seção 7.2). O BKT assume que o conhecimento do aluno é um conjunto de variáveis binárias latentes, uma por componente de conhecimento. Como se trata de um modelo probabilístico, as saídas do modelo são valores que representam probabilidades, ou seja, valores reais entre 0,0 e 1,0. Sistemas tutores inteligentes costumam usar um valor limiar de 0,95 para inferir se o aluno domina ou não domina o componente de conhecimento. Um valor superior ou igual ao limiar indica que o aluno domina, enquanto um valor inferior ao limiar indica que o aluno não domina o componente de conhecimento (KOEDINGER; CORBETT, 2006).

A Figura 12 apresenta a estrutura básica do modelo BKT. Quando o aluno está em um estado de “não aprendido” (“não domínio”), para determinado componente de conhecimento, há uma probabilidade de ele adivinhar (*guess*) a resposta correta ( $p(G)$ ) e de não adivinhar a resposta correta ( $1 - p(G)$ ). Quando o aluno está em um estado de “aprendido” (“domínio”), para determinado componente de conhecimento, há uma probabilidade de ele cometer um deslize (*slip*) e errar a resposta ( $p(S)$ ) e de não cometer o deslize e acertar a resposta ( $1 - p(S)$ ). A probabilidade de o aluno transitar (*transition*) do estado de “não aprendido” para “aprendido”, após uma oportunidade de aplicar o conhecimento, também é representada no modelo pela probabilidade  $p(T)$ . O modelo também considera a probabilidade inicial de o aluno estar no estado “aprendido”, para determinado componente de conhecimento, antes de tentar aplicar o componente de conhecimento pela primeira vez ( $p(L_0)$ ). A probabilidade de o aluno dominar o componente de conhecimento após aplicá-lo na  $n$ ésima oportunidade é representado por  $p(L_n)$ .

Figura 12 – Estrutura básica do modelo BKT



Fonte: Adaptado de (PELÁNEK, 2017a; SLATER; BAKER, 2018).

No modelo padrão do BKT, os quatro parâmetros ( $p(L_0)$ ,  $p(G)$ ,  $p(S)$  e  $p(T)$ ) são estimados (ajustados) separadamente para cada componente de conhecimento, sem individualização para os alunos (SLATER; BAKER, 2018). O Quadro 3 descreve os parâmetros do modelo BKT e o Quadro 4 apresenta os parâmetros em forma de matriz.

Quadro 5 – Parâmetros do modelo BKT

Parâmetro	Descrição
$p(L_0^k)$	A probabilidade inicial de aprendizagem. O $L$ significa “aprendizagem”, do inglês “ <i>Learning</i> ”. A probabilidade de o aluno dominar o componente de conhecimento $k$ antes de tentar aplicá-lo pela primeira vez ( $n = 0$ ).
$p(G^k)$	A probabilidade de o aluno aplicar corretamente o componente de conhecimento $k$ estando no estado de “não domínio” de $k$ . O $G$ significa “adivinhar”, do inglês “ <i>Guess</i> ”.
$p(S^k)$	A probabilidade de o aluno aplicar incorretamente o componente de conhecimento $k$ estando no estado de “domínio” de $k$ . O $S$ significa “deslize”, do inglês “ <i>Slip</i> ”.
$p(T^k)$	A probabilidade de o aluno mover de um estado de “não domínio” para o estado de “domínio” do componente de conhecimento $k$ , após uma oportunidade de aplicar o componente de conhecimento $k$ . O $T$ significa “transição”, do inglês “ <i>Transition</i> ”.

Fonte: (CORBETT; ANDERSON, 1995).

O cálculo de  $p(L_n)$  é baseado nos quatro parâmetros do modelo ( $p(L_0)$ ,  $p(G)$ ,  $p(S)$  e  $p(T)$ ). Em cada oportunidade  $n$  de o aluno aplicar o componente de conhecimento  $k$ , o BKT atualiza dois indicadores: a) a estimativa de o aluno dominar o componente de conhecimento  $k$  **antes** de aplicar o conhecimento na  $n$ ésima oportunidade ( $p(L_{n-1}^k)$ ); e b) a estimativa de o aluno dominar o componente de conhecimento  $k$  **após** aplicar o conhecimento na  $n$ ésima oportunidade ( $p(L_n^k)$ ).

Quadro 6 – Probabilidades dos parâmetros do modelo BKT

a) anteriores		b) transições		c) observações			
dominado	$p(L_0^k)$		para dominado	para não dominado		correto	errado
não dominado	$1 - p(L_0^k)$	de dominado	1	0	dominado	$1 - p(S^k)$	$p(S^k)$
		de não dominado	$p(T^k)$	$1 - p(T^k)$	não dominado	$p(G^k)$	$1 - p(G^k)$

Fonte: Adaptado e traduzido de (YUDELSON; KOEDINGER; GORDON, 2013).

A Fórmula 7.1 atualiza o valor de  $p(L_{n-1}^k)$ , referente à estimativa de o aluno ter dominado o componente de conhecimento  $k$  **antes** de aplicar o conhecimento na  $n$ ésima oportunidade. Esta estimativa é atualizada a partir do desempenho do aluno na oportunidade  $n$

de praticar o componente de conhecimento  $k$ . A variável *correta* recebe o valor 1 quando o aluno aplica corretamente o componente de conhecimento  $k$  na oportunidade  $n$ , de forma a considerar o termo correspondente, na fórmula, ou 0, quando aplica incorretamente, de forma a anular o termo correspondente, na fórmula. O mesmo ocorre com a variável *incorreta*, que recebe o valor 1 quando o aluno aplica incorretamente o componente de conhecimento  $k$  na oportunidade  $n$ , de forma a considerar o termo correspondente, na fórmula, ou 0, quando o aluno aplica corretamente, de forma a anular o termo correspondente, na fórmula.

$$p(L_{n-1}^k | resposta_n^k) = correta_n^k * \frac{p(L_{n-1}^k) * (1 - p(S^k))}{p(L_{n-1}^k) * (1 - p(S^k)) + (1 - p(L_{n-1}^k)) * p(G^k)} + \\ incorreta_n^k * \frac{p(L_{n-1}^k) * p(S^k)}{p(L_{n-1}^k) * p(S^k) + (1 - p(L_{n-1}^k)) * (1 - p(G^k))} \quad (7.1)$$

A partir de  $p(L_{n-1}^k | resposta_n^k)$ , é calculada a probabilidade de o aluno dominar o componente de conhecimento  $k$  **após** a oportunidade de aplicar o conhecimento na  $n$ ésima oportunidade ( $p(L_n^k)$ ) (ver Fórmula 7.2). A probabilidade atualizada do aluno dominar o componente de conhecimento antes da oportunidade de aplicar o conhecimento é somada à probabilidade de o aluno passar a dominar o componente de conhecimento após aplicá-lo, caso ele não dominasse antes da oportunidade de aplicá-lo.

$$p(L_n^k) = p(L_{n-1}^k | resposta_n^k) + (1 - p(L_{n-1}^k | resposta_n^k)) * p(T^k) \quad (7.2)$$

Para inferir que o aluno domina o componente de conhecimento  $k$ , os sistemas tutores costumam usar um limiar de  $p(L_n^k) > 0,95$  (KOEDINGER; CORBETT, 2006). Uma característica do modelo padrão do BKT é que ele considera que o aluno não “esquece” o componente de conhecimento, a partir do momento em que ele o domina, de forma que quando  $p(L_n^k) = 1$  (100%), este valor não regride mais (YUDELSON; KOEDINGER; GORDON, 2013).

Tabela 8 – Simulação do modelo BKT com 25 aplicações de componente de conhecimento

$n$	correto	incorreto	$p(L_{n-1}^k   resposta_n^k)$	$p(L_n^k)$	$p(L_n^k) > 0,95$	$p(\text{correto}_{n+1}^k)$
1	0	1	0,03571	0,18035	0	0,24428
2	0	1	0,02386	0,17028	0	0,23622
3	1	0	0,64876	0,70145	0	0,66116
4	1	0	0,95484	0,96161	1	0,86929

5	0	1	0,73571	0,77535	0	0,72028
6	1	0	0,96881	0,97349	1	0,87879
7	1	0	0,99698	0,99743	1	0,89794
8	1	0	0,99971	0,99975	1	0,89980
9	0	1	0,99781	0,99814	1	0,89851
10	1	0	0,99979	0,99982	1	0,89985
11	1	0	0,99998	0,99998	1	0,89998
12	1	0	0,99999	0,99999	1	0,89999
13	1	0	0,99999	0,99999	1	0,89999
14	1	0	0,99999	0,99999	1	0,89999
15	1	0	0,99999	0,99999	1	0,89999
16	1	0	0,99999	0,99999	1	0,89999
17	1	0	1,00000	1,00000	1	0,90000
18	1	0	1,00000	1,00000	1	0,90000
19	1	0	1,00000	1,00000	1	0,90000
20	1	0	1,00000	1,00000	1	0,90000
21	1	0	1,00000	1,00000	1	0,90000
22	0	1	1,00000	1,00000	1	0,90000
23	0	1	1,00000	1,00000	1	0,90000
24	1	0	1,00000	1,00000	1	0,90000
25	1	0	1,00000	1,00000	1	0,90000

Fonte: Elaborado pelo autor.

Através do modelo BKT, também é possível descobrir a probabilidade de que o aluno irá aplicar corretamente o componente de conhecimento  $k$  em uma próxima oportunidade ( $n + 1$ ). Isso é feito através da Fórmula 7.3, que soma a probabilidade de o aluno não dominar o conhecimento  $k$  e acertar a resposta com a probabilidade de o aluno dominar o conhecimento  $k$  e não cometer um deslize.

$$p(\text{correto}_{n+1}^k) = (1 - p(L_n^k)) * p(G^k) + p(L_n^k) * (1 - p(S^k)) \quad (7.3)$$

A Tabela 8 apresenta uma simulação com 25 oportunidades de aplicação de um componente de conhecimento  $k$  por um determinado aluno. A tabela é interessante para compreender a dinâmica de atualização das variáveis  $p(L_n^k)$  e  $p(\text{correto}_{n+1}^k)$  a cada oportunidade de o aluno aplicar um componente de conhecimento em uma tarefa de aprendizagem.

Para a simulação da Tabela 8, foram considerados os seguintes valores para os quatro parâmetros do modelo, escolhidos arbitrariamente:  $p(L_0^k) = 0,25$ ;  $p(G^k) = 0,10$ ;  $p(S^k) = 0,10$ ; e  $p(T^k) = 0,15$ . Veja que estes quatro parâmetros são definidos antes da primeira oportunidade de o aluno aplicar um componente de conhecimento. Já as variáveis  $p(L_n^k)$  e  $p(\text{correto}_{n+1}^k)$  são atualizadas a cada oportunidade de o aluno aplicar um componente de conhecimento. Na simulação, quando  $n = 1$ , então  $p(L_{n-1}^k) = p(L_0^k) = 0,25$ , mas como o aluno errou a primeira oportunidade de aplicar o conhecimento,  $p(L_{n-1}^k)$  é atualizado para 0,03571. Na simulação, o modelo infere que o aluno atingiu o domínio (“aprendido”) no componente de conhecimento  $k$

(quando  $p(L_n^k) > 0,95$  recebe 1 (aprendido), senão 0 (não aprendido)) na quarta oportunidade de aplicar o conhecimento ( $n = 4$ ). Na quinta oportunidade ( $n = 5$ ), o aluno comete um erro e retorna ao estado “não aprendido”. O estado “aprendido” é obtido novamente com um novo acerto na sexta oportunidade ( $n = 6$ ). A partir de  $n = 17$ , o valor de  $p(L_n^k)$  passa a ser 1,0 (100%), de forma que este indicador não regredirá mais nas próximas oportunidades de aplicação de conhecimento, mesmo que o aluno erre, como ocorre em  $n = 22$  e  $n = 23$ .

Nesta simulação, os valores para os quatro parâmetros foram atribuídos de forma arbitrária. No entanto, existem recomendações sobre os valores atribuídos para estes parâmetros. O trabalho de Baker, Corbett e Alevén (2008) propõe limitar valores de  $p(G)$  e  $p(S)$ , com o objetivo de evitar problemas de degeneração do modelo. A degeneração do modelo ocorre quando aplicar incorretamente um componente de conhecimento se torna mais evidência para o domínio de conhecimento do que quando a aplicação é correta. Quando  $p(S) > 0,5$ , está sendo modelado que quando o aluno domina o componente de conhecimento, ele teria mais chance de cometer um deslize na aplicação do conhecimento do que acertar. E quando  $p(G) > 0,5$ , está sendo modelado que um aluno que não domina o componente de conhecimento tem mais chance de “adivinhar” uma aplicação correta do conhecimento do que errar. Para evitar a degeneração do modelo, autores sugerem valores de  $p(S)$  e  $p(G)$  menores que 0,5 (BAKER; CORBETT; ALEVEN, 2008) ou menores de 0,25 (SLATER; BAKER, 2018).

Mais considerações podem ser descritas sobre os parâmetros do modelo. Valores baixos, próximos de 0, para  $p(L_0)$ , indicam que o aluno não tem familiaridade com o conteúdo no início das tarefas. No entanto, valores altos, próximos de 1, indicam que o aluno não teria o que aprender e apenas  $p(S)$  teria relevância para o modelo. Valores de  $p(T)$  têm relação com a velocidade de aprendizagem, ou seja, quantas aplicações do conhecimento são necessárias para o modelo inferir o domínio do aluno no componente. Quanto maior o valor  $p(T)$ , mais rapidamente é alcançado o domínio de conhecimento, a cada oportunidade de aplicá-lo.

Esta Seção também buscou evidenciar a natureza explicativa do modelo BKT clássico usado na Tese, pois seus quatro parâmetros e as variáveis inferidas ( $p(L_n)$  e  $p(\text{correto})$ ) modelam diferentes aspectos do processo de aquisição de conhecimento do aluno. Estas informações do modelo são utilizadas pela Tese na derivação de atributos de dados sobre o conhecimento do aluno para as amostras dos modelos de detecção de confusão da Tese.

### 7.3.1 Extensão do modelo com um parâmetro sobre “esquecimento”

Anteriormente foi apresentada uma discussão sobre uma limitação do modelo BKT clássico: o modelo assume que o aluno não “esquece” o componente de conhecimento, a partir do momento que ele domina o conhecimento. Na simulação da Tabela 8, a partir da oportunidade 17 de aplicação de conhecimento ( $n = 17$ ), o valor de  $p(L_n^k)$  se torna 1.0, e nas próximas oportunidades de aplicação, mesmo que o aluno aplique incorretamente o componente de conhecimento, o valor de  $p(L_n^k)$  não regride. Esta característica do modelo BKT clássico poderia trazer um viés aos atributos de dados sobre estimativas de conhecimento do aluno e diminuir o desempenho dos modelos na detecção da confusão.

O modelo BKT clássico foi estendido para considerar o “esquecimento” do aluno nos componentes de conhecimento, através da inclusão de um quinto parâmetro, o  $p(F)$ , que corresponde à **probabilidade de esquecimento no componente de conhecimento**. Em  $p(F)$ , o F significa *forgetting* (esquecimento, em inglês). O parâmetro  $p(F)$  foi modelado para ter comportamento contrário ao parâmetro  $p(T)$ . Enquanto o  $p(T)$  indica a probabilidade de o aluno transitar do estado “não aprendido” para “aprendido”, o  $p(F)$  indica a probabilidade de o aluno transitar do estado “aprendido” para o “não aprendido”. No modelo BKT clássico, este parâmetro é inexistente, ou seja,  $p(F) = 0$  (zero). (KHAJAH; LINDSEY; MOZER, 2016).

A Tese implementou o modelo BKT com a biblioteca Python *pyBKT* (BADRINATH; WANG; PARDOS, 2021)<sup>2</sup>. A biblioteca fornece diversas rotinas para a implementação de modelos BKT, geração de dados, ajustes dos parâmetros, predição e validação cruzada. A biblioteca também fornece extensões para o modelo clássico do BKT, incluindo uma extensão que inclui o parâmetro de “esquecimento” do aluno. Esta extensão é chamada de *BKT+forget*.

A biblioteca não descreve detalhes sobre como adiciona o parâmetro  $p(F)$  ao modelo BKT clássico. Não há uma indicação sobre como as fórmulas do modelo BKT apresentadas anteriormente são modificadas para receber o parâmetro de esquecimento. No entanto, a biblioteca sugere seguir a implementação descrita no trabalho de Khajah, Lindsey e Mozer (2016). O trabalho referenciado também não apresenta detalhes de implementação do parâmetro no modelo BKT clássico, mas descreve o comportamento do parâmetro no modelo. Os autores basearam a implementação em um outro trabalho (QIU et al., 2011), que modelou o esquecimento do aluno de um dia para o outro. Quando o aluno tentava aplicar o mesmo componente de conhecimento em uma mesma tarefa no mesmo dia, o modelo considerava o

---

<sup>2</sup> Disponível em <https://github.com/CAHLR/pyBKT>.

valor 0 (zero) para o parâmetro  $p(F)$ . No entanto, quando o componente de conhecimento era aplicado no dia posterior, o parâmetro era adequado para prever o esquecimento do aluno de um dia para o outro. No entanto, Khajah e colegas consideraram que o esquecimento poderia ocorrer em um período de tempo mais curto, na seguinte forma: considerando uma sequência de tarefas  $A_1-A_2-B_1-A_3-B_2-B_3-A_4$ , onde as letras referem-se a determinado componente de conhecimento (A e B), os autores trataram a quebra de sequência de tarefas do mesmo componente de conhecimento, de forma que a probabilidade de esquecimento do componente de conhecimento entre  $A_1$  e  $A_2$  é  $p(F)$ , mas a probabilidade de esquecimento entre  $A_2$  e  $A_3$  é  $1 - (1 - p(F))^2$  e entre  $A_3$  e  $A_4$  é  $1 - (1 - p(F))^3$ . Ou seja, após uma tarefa ser realizada para um componente de conhecimento A, quanto mais tarefas forem realizadas para outros componentes de conhecimento que não A, antes de ocorrer a próxima tarefa para A, maior será a probabilidade de esquecimento de A numa próxima tarefa para A. A forma como a probabilidade de esquecimento é integrada às fórmulas do modelo BKT clássico não foi apresentada pelos autores (KHAJAH; LINDSEY; MOZER, 2016).

Apesar de os autores já terem mostrado o bom desempenho do modelo BKT+*forget*, em comparação com o modelo clássico e com modelos DKT (KHAJAH; LINDSEY; MOZER, 2016), o autor da Tese realizou testes piloto com as duas versões do BKT (clássico e BKT+*forget*). O BKT+*forget* superou o desempenho do modelo clássico.

### 7.3.2 Adequação dos parâmetros do modelo BKT

Na simulação apresentada na Tabela 8, foram definidos valores arbitrários para os quatro parâmetros ( $p(L_0)$ ,  $p(T)$ ,  $p(S)$ , e  $p(G)$ ) do modelo, relativos ao componente de conhecimento  $k$ . Estes valores são ajustados antes da aplicação do modelo BKT. No modelo clássico do BKT, estes valores são ajustados para cada componente de conhecimento e são compartilhados para todos os estudantes. Esta abordagem funciona bem quando os alunos possuem características homogêneas, como quando possuem o mesmo conhecimento prévio antes da primeira tentativa de aplicar o conhecimento, e a mesma taxa de aprendizagem. Quando os estudantes possuem características heterogêneas, pode ser interessante individualizar a adequação dos parâmetros, ajustando os parâmetros para cada aluno. Esta segunda forma de adequação é também conhecida como individualização do modelo BKT, pois considera a variabilidade do fator aluno na estimação do conhecimento do aprendiz (YUDELSON, 2016). Grupos heterogêneos podem ser compostos por alunos que tiveram diferentes experiências prévias de aprendizagem no domínio de conhecimento. Além disso,

diferentes alunos podem possuir habilidades cognitivas e metacognitivas distintas. O trabalho de Fancsali et al. (2013) testou o desempenho do modelo clássico de BKT. Os resultados mostraram que para conjuntos de dados heterogêneos, este modelo clássico (sem individualização dos parâmetros) perde em desempenho, aumentando a proporção de alunos identificados prematuramente em estado de domínio do conhecimento, ou identificados incorretamente no estado de não domínio. Os resultados do trabalho de Lee e Brunskill (2012) ajudam a explicar essa perda de desempenho do modelo clássico. Os autores compararam as demandas de práticas de aprendizagem dos estudantes em relação às previsões de modelos com parâmetros treinados a nível da população e a nível individual. O estudo encontrou que os alunos precisam praticar determinado componente de conhecimento duas vezes mais quando o modelo de parâmetros treinados a nível de população indica o domínio do aluno, comparado com o modelo de parâmetros treinados a nível de aluno. Considerar as variações individuais dos estudantes pode ter implicações importantes para as decisões instrucionais.

No contexto da Tese, um modelo preciso de conhecimento do aluno é fundamental para a Tese alcançar o objetivo de verificar a capacidade preditiva de atributos de dados derivados do conhecimento do aluno para a detecção da confusão. No entanto, treinar parâmetros individuais para cada estudante pode ser inviável em alguns contextos, pois exige mais interação individual do estudante com determinado componente de conhecimento, ou seja, exige que o professor atribua uma quantidade suficiente de tarefas de aprendizagem sobre determinado componente de conhecimento, para que os parâmetros individuais sejam adequadamente ajustados (PELÁNEK, 2017a). O autor da Tese realizou testes preliminares de adequação dos parâmetros à nível de aluno, mas a individualização dos parâmetros para cada estudante se mostrou inviável. Muitos alunos tiveram poucas oportunidades de aplicação de determinados componentes de conhecimento, em função da quantidade de exercícios disponibilizados e a liberdade que os alunos tinham de escolher os exercícios a serem resolvidos. Para estes alunos, ficou inviável validar os modelos com validação cruzada *k-fold* com  $k = 10$  ou  $k = 5$ . Os resultados destes testes levaram a Tese a validar o modelo de conhecimento em dois cenários: 1) treinar parâmetros gerais, ou seja, treinar os cinco parâmetros ( $p(L_0)$ ,  $p(T)$ ,  $p(S)$ ,  $p(G)$  e  $p(F)$ ) do modelo BKT+*forget* para cada componente de conhecimento para todos os alunos participantes da pesquisa; e 2) treinar os parâmetros à nível de ensino (ensino técnico e graduação), ou seja, treinar os cinco parâmetros ( $p(L_0)$ ,  $p(T)$ ,  $p(S)$ ,  $p(G)$  e  $p(F)$ ) do modelo BKT+*forget* para cada componente de conhecimento para os alunos de ensino técnico e depois treinar os mesmos parâmetros para cada componente de conhecimento para os alunos de graduação. Os resultados são apresentados na Seção 7.5.

Diversas técnicas já foram utilizadas na literatura para adaptar os parâmetros do modelo BKT, como minimização de função (SLATER; BAKER, 2018), *Expectation-Maximization* (BECK; CHANG, 2007; LEE; BRUNSKILL, 2012), *Discretized Brute-Force Search* (YUDELSON; KOEDINGER; GORDON, 2013), otimização (YUDELSON; KOEDINGER; GORDON, 2013), *Contextual Guess-and-Slip* (BECK; CHANG, 2007), *Individual Difference Weights* (CORBETT; ANDERSON, 1995), divisão de parâmetros e modelo hierárquico (YUDELSON, 2016). O trabalho descrito por Pardos e Heffernan (2010a) utilizou três estratégias de adaptação a nível de aluno: a) definir valores iniciais randômicos; b) definir valores iniciais baseados na primeira resposta ao problema; e c) encontrar parâmetros compartilhados entre as habilidades/conhecimentos para cada estudante.

A Tese utilizou o algoritmo *expectation-maximization* (MOON, 1996) para ajustar os cinco parâmetros ( $p(L_0)$ ,  $p(T)$ ,  $p(S)$ ,  $p(G)$  e  $p(F)$ ) do modelo BKT+*forget*. O *expectation-maximization* (EM) é um algoritmo eficiente e bastante utilizado na adequação de parâmetros de modelos BKT (PARDOS; HEFFERNAN, 2010b; PELÁNEK, 2017b). O EM é um algoritmo iterativo que tenta encontrar o conjunto de parâmetros que melhor se ajustam aos dados (PARDOS; HEFFERNAN, 2010b). No trabalho da Tese, o EM ajustou os parâmetros sobre os dados coletados (evidências de aplicação de conhecimento) durante todo o período de coleta da Tese, de fevereiro a julho de 2021. No cenário 1, foram ajustados os parâmetros para todos os alunos, em cada componente de conhecimento, considerando todas as evidências de aplicação de conhecimento em todo o período de coleta de dados. No cenário 2, primeiramente foram ajustados os parâmetros para o ensino técnico, em cada componente de conhecimento, considerando todas as evidências de aplicação de conhecimento dos alunos do ensino técnico em todo o período de coleta. Depois foram ajustados os parâmetros para o ensino superior, em cada componente de conhecimento, sobre todas as evidências de aplicação de conhecimentos dos alunos do ensino superior em todo o período de coleta.

#### **7.4 Coleta de evidências de aplicação dos componentes de conhecimento**

O modelo de conhecimento do aluno BKT+*forget* é atualizado a cada evidência de aplicação correta ou incorreta de um componente de conhecimento. Quando a evidência é obtida, o modelo atualiza as probabilidades  $p(L_n^k)$  (probabilidade de o aluno dominar o componente de conhecimento  $k$ ) e  $p(\text{correto}_{n+1}^k)$  (probabilidade de o aluno aplicar corretamente o componente de conhecimento  $k$  em uma próxima oportunidade). A evidência de aplicação correta ou incorreta do componente é obtida quando o aluno submete sua solução para

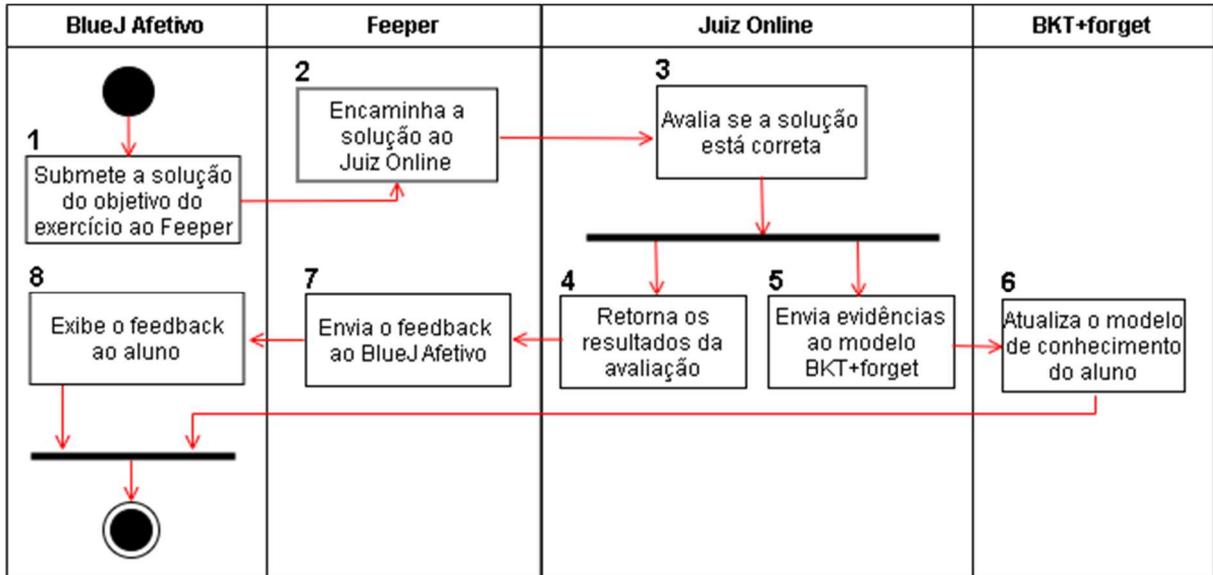
determinado objetivo de exercício, no *BlueJ* Afetivo. O Capítulo 6 mostrou o processo de submissão de solução de objetivo de exercício pelo aluno, na interface do *BlueJ* Afetivo. Cada submissão é avaliada pelo Juiz Online do *Feeper* como correta ou incorreta, de forma que cada submissão resulta em uma evidência de aplicação correta ou incorreta dos componentes de conhecimento atrelados ao objetivo de exercício. Cada objetivo de exercício pode ser atrelado a mais de um componente de conhecimento. Desta forma, quando o aluno submete para avaliação sua solução para um objetivo de exercício, valores de  $p(L_n^k)$  e  $p(\text{correto}_{n+1}^k)$  podem ser atualizados para um ou mais componentes de conhecimento.

A Figura 13 apresenta um diagrama de atividades<sup>3</sup> que mostra o fluxo de atividades do processo de submissão da solução de objetivo pelo aluno, no *BlueJ* Afetivo, e consequentemente a atualização do modelo BKT+*forget*. Os números presentes na figura não representam a ordem em que as atividades ocorrem no processo, são apenas para facilitar a identificação das atividades no texto. Primeiramente, o *BlueJ* Afetivo submete ao *Feeper* a solução do aluno para um objetivo de exercício (1). O *Feeper* então encaminha a solução ao Juiz Online (2). O Juiz Online avalia se a solução está correta (3). A solução estará incorreta se não compilar ou se o código não passar nos testes unitários vinculados ao objetivo de exercício. A solução estará correta se o código não tiver erros de compilação e passar nos testes unitários do objetivo de exercício. Na sequência do processo há uma bifurcação (linha preta). Em diagramas de atividade, uma bifurcação indica a quebra do fluxo em dois ou mais fluxos paralelos. No caso do processo descrito, são dois fluxos paralelos que bifurcam. Em um dos fluxos paralelos, o Juiz Online envia ao modelo BKT+*forget* as evidências sobre a solução do aluno para o objetivo de exercício (5). Se a solução estiver correta, será enviada uma evidência de aplicação correta para todos os componentes de conhecimento vinculados ao objetivo de exercício. Senão, é enviada uma evidência de aplicação incorreta para todos os componentes de conhecimento vinculados ao objetivo de exercício. Finalizando o fluxo paralelo, o modelo BKT+*forget* atualiza  $p(L_n^k)$  e  $p(\text{correto}_{n+1}^k)$  do aluno nos  $k$  componentes de conhecimento relacionados ao objetivo de exercício. O outro fluxo paralelo da bifurcação começa com o Juiz Online retornando ao *Feeper* o resultado da avaliação da solução do aluno para o objetivo de exercício (4). O *Feeper* então prepara um *feedback* para o aluno e o envia para o *BlueJ* Afetivo (7). O *BlueJ* Afetivo exibe o *feedback* (8). O processo termina quando os dois fluxos paralelos se encontram em uma junção, também representado no diagrama com uma linha preta.

---

<sup>3</sup> Diagrama de atividades da UML (*Unified Modeling Language*).

Figura 13 - Processo de submissão da solução do objetivo de exercício



Fonte: Elaborado pelo autor.

Como descrito no Capítulo 5, a coleta de dados da Tese foi realizada com 93 alunos de cinco turmas introdutórias de programação de computadores, mas foram considerados apenas os dados de 62 alunos. A Tabela 9 mostra a distribuição das ocorrências de evidências de aplicação de conhecimento para cada componente de conhecimento, nas turmas de ensino técnico e superior. É possível verificar que muitos dos componentes de conhecimento não foram abordados nas turmas de ensino técnico, por possuírem uma ementa diferente e mais simplificada do que no ensino superior. No entanto, esta característica não é um problema para o treinamento dos parâmetros do modelo BKT, pois os parâmetros são gerados para cada componente de conhecimento em cada nível de ensino. Os componentes de conhecimento que não ocorreram no ensino técnico não possuem parâmetros ajustados. Espera-se que quanto mais evidências de aplicação de conhecimento em um componente de conhecimento, melhor o ajuste dos parâmetros para este componente de conhecimento.

Tabela 9 - Distribuição das evidências de aplicação de componentes de conhecimento

#	Componente de conhecimento	Sup.	Téc.	Total
1	Atributo de classe	1443	-	1443
2	Classe	2229	1434	3663
3	Método construtor com inicialização de atributo	827	-	827
4	Método construtor com inicialização de atributo do tipo array	202	-	202
5	Método construtor com inicialização de atributo herdado	147	-	147
6	Método construtor padrão	213	-	213
7	Método de classe com parâmetro e com retorno	4155	2265	6420

8	Método de classe com parâmetro e sem retorno	373	153	526
9	Método de classe sem parâmetro e com retorno	2674	-	2674
10	Método de classe sem parâmetro e sem retorno	511	-	511
11	Método principal	1298	1240	2538
12	Métodos de modificação e de acesso	1016	-	1016
13	Subclasse	297	-	297
14	Método toString	181	-	181
15	Entrada de array, instanciação, invocação de método e saída de array	195	2	197
16	Entrada, instanciação, invocação de método e saída	659	-	659
17	Entrada, invocação de método e saída	27	801	828
18	Escrita de array	1672	-	1672
19	Algoritmo formal	32	-	32
20	Incremento/decremento de atributo	403	-	403
21	Invocação de método	177	-	177
22	Leitura de array de tipo não primitivo	1276	-	1276
23	Leitura de array de tipo primitivo	643	239	882
24	Leitura de atributo	108	-	108
25	Operação aritmética	1322	1292	2614
26	Repetição	344	261	605
27	Seleção de caso	791	833	1624
28	Seleção de caso com operação aritmética	89	182	271
29	Verificação de tipo de objeto	72	-	72

Fonte: Elaborado pelo autor.

Cada registro de evidência de aplicação de conhecimento possui os seguintes dados: a data e hora do registro de evidência (data e hora da submissão da solução do objetivo de exercício pelo aluno); a descrição do componente de conhecimento atrelado à evidência; e a informação se o registro de evidência é relativo a uma aplicação correta ou incorreta do componente de conhecimento.

## 7.5 Treinamento e validação do modelo

Modelos de conhecimento do aluno podem ser validados, com o objetivo de verificar o seu desempenho em estimar o conhecimento do aluno, de duas maneiras: **avaliação interna** e **avaliação externa**. A **avaliação interna** compara as predições do modelo para  $p(\text{correto}_{n+1}^k)$  com cada próxima aplicação de conhecimento do aluno em tarefas de aprendizagem, como realizado por Slater e Baker (2018). A **avaliação externa** compara as estimativas do modelo com pós-testes de desempenho. A Tese utilizou o método de **avaliação interna** com validação cruzada *k-fold* ( $k=10$  e  $k=5$ ), um método sólido para avaliação de modelos de conhecimento de alunos (PELÁNEK, 2017a). O método é interessante por evitar o super ajuste dos parâmetros do modelo aos dados de treino, de forma que as evidências de aplicação de conhecimento utilizadas para ajustar os parâmetros do modelo com *expectation-maximization*, não são as mesmas utilizadas para testar o modelo. Com  $k=10$  (10 *folds*), as evidências de aplicação de

conhecimento são agrupadas em 10 grupos de amostras (*folds*). Através de um processo iterativo, o método realiza 10 rodadas de treino e teste do modelo, sendo que em cada rodada são utilizados 9 diferentes grupos de evidências no treino, e o grupo restante é usado para testar o modelo. Após as 10 iterações, é gerada uma média para cada uma das métricas consideradas. A Tese utilizou como métricas a acurácia e o RMSE (raiz de erro quadrático médio, do inglês *root mean squared error*).

A acurácia mede a habilidade do modelo em prever o desempenho futuro do aluno e quanto maior for o seu valor, melhor o desempenho do modelo. Nesta métrica, o desempenho do aluno em um objetivo de exercício (correto ou incorreto) é comparado com a previsão do modelo. A previsão do modelo é obtida através de um valor absoluto obtido na estimativa do modelo para  $p(\text{correto}_{n+1}^k)$  (correto quando  $p(\text{correto}_{n+1}^k) \geq 0,5$ , e incorreto quando  $p(\text{correto}_{n+1}^k) < 0,5$ ). A acurácia é o percentual de predições corretas em relação ao total de predições (PELÁNEK, 2015). O RMSE mede a distância média quadrática do desempenho do aluno no objetivo de um exercício (1 para correto e 0 para incorreto) em relação ao valor de  $p(\text{correto}_{n+1}^k)$ . Quanto menor o valor de RMSE, melhor o desempenho do modelo (PELÁNEK, 2015).

### 7.5.1 Resultados

Esta seção apresenta os resultados da validação dos modelos BKT+*forget* usados na Tese. Foram realizadas validações dos modelos BKT+*Forget* em dois cenários: 1) modelo que gera parâmetros de cada componente de conhecimento para **todos os alunos** (modelo geral); e 2) modelo que gera parâmetros de cada componente de conhecimento **para os alunos de cada nível de ensino** (individualizado por nível de ensino). Como evidenciado no Capítulo 5, alunos de ensino superior e de ensino técnico possuem características heterogêneas em relação às suas experiências anteriores e conhecimentos prévios em programação. Sabendo disso, um modelo BKT+*forget* geral, que usa os mesmos parâmetros para todos os estudantes, poderia potencializar erros de estimativas de conhecimento dos alunos. Dessa forma, seria esperado que os modelos do segundo cenário apresentassem resultados superiores ao modelo do primeiro cenário.

Tabela 10 – Resultados da validação cruzada *k-fold* para o modelo BKT+*forget* geral

#	Componente de conhecimento	Acurácia		RMSE	
		k	k	k	k
		5	10	5	10

1	Atributo de classe	,898	,900	,297	,295
2	Classe	,968	,968	,171	,166
3	Método construtor com inicialização de atributo	,882	,861	,309	,308
4	Método construtor com inicialização de atributo do tipo array	,796	,840	,368	,339
5	Método construtor com inicialização de atributo herdado	,842	,866	,329	,280
6	Método construtor padrão	,920	,929	,257	,207
7	Método de classe com parâmetro e com retorno	,994	,994	,053	,049
8	Método de classe com parâmetro e sem retorno	1,00	1,00	,084	,043
9	Método de classe sem parâmetro e com retorno	,999	,999	,012	,009
10	Método de classe sem parâmetro e sem retorno	,995	,994	,034	,028
11	Método principal	,997	,997	,046	,039
12	Métodos de modificação e de acesso	,845	,859	,353	,338
13	Subclasse	,874	,881	,329	,305
14	Método toString	,812	,808	,374	,365
15	Entrada de array, instanciação, invocação de método e saída de array	,672	,632	,474	,477
16	Entrada, instanciação, invocação de método e saída	,774	,773	,415	,406
17	Entrada, invocação de método e saída	,759	,767	,425	,405
18	Escrita de array	,698	,686	,446	,453
19	Algoritmo formal	,738	,875	,473	,401
20	Incremento/decremento de atributo	,830	,825	,363	,367
21	Invocação de método	,731	,733	,438	,419
22	Leitura de array de tipo não primitivo	,717	,735	,440	,419
23	Leitura de array de tipo primitivo	,720	,719	,446	,444
24	Leitura de atributo	,799	,740	,399	,383
25	Operação aritmética	,747	,743	,431	,432
26	Repetição	,624	,620	,483	,482
27	Seleção de caso	,709	,704	,448	,448
28	Seleção de caso com operação aritmética	,718	,672	,425	,452
29	Verificação de tipo de objeto	,743	,796	,427	,389

Fonte: Elaborado pelo autor.

Todos os modelos passaram por duas validações cruzadas *k-fold*, sendo uma com  $k = 5$  e outra com  $k = 10$ . A justificativa para a realização de validações com duas configurações de *folds* (5 e 10) é que para alguns componentes de conhecimento, não havia amostras de evidências de aplicação de conhecimento suficientes para a distribuição dessas amostras em 10 *folds*, então também foi executada uma rodada de validação cruzada com  $k = 5$ . Embora a validação cruzada com 10 *folds* apresente resultados mais sólidos quando há dados suficientes disponíveis, a validação cruzada com 5 *folds* também é considerada sólida na maior parte dos casos (MARCOT; HANEA, 2021), então também foi executada para obtermos uma métrica de avaliação para componentes de conhecimento que não puderam ser avaliados com  $k = 10$ .

No primeiro cenário, foi avaliado um modelo BKT+*forget* que utilizou os dados de todos os alunos para ajustar os cinco parâmetros ( $p(L_0)$ ,  $p(T)$ ,  $p(S)$ ,  $p(G)$  e  $p(F)$ ) de cada componente de conhecimento. A Tese chama este modelo de **BKT+Forget geral**. A Tabela 10 mostra os resultados para cada componente de conhecimento. Para seis CCs a acurácia ficou próxima de 1,0 ( $> 0,95$ ) e a maioria obteve acurácia superior a 0,8 e RMSE próximo de 0,3. Os

demais componentes de conhecimento que obtiveram acurácia inferiores a 0,8 e RMSE acima de 0,4, podem ter sofrido com a alta granularidade da modelagem dos componentes de conhecimento.

No segundo cenário, foi realizada uma tentativa de ajustar parâmetros para cada nível de ensino dos participantes (técnico e graduação). A Tabela 11 mostra os resultados do ajuste de parâmetros no ensino técnico. A tabela mostra que alguns CCs não obtiveram resultados para as métricas, pela pequena quantidade de evidências de aplicação de conhecimento, não suficientes para a aplicação da validação cruzada com *folds* de 5 e 10.

Tabela 11 – Resultados da validação cruzada *k-fold* para o modelo BKT do ensino técnico

#	Componente de conhecimento	Acurácia		RMSE	
		k	k	k	k
		5	10	5	10
2	Classe	,945	,942	,220	,221
7	Método de classe com parâmetro e com retorno	,985	,979	,081	,083
8	Método de classe com parâmetro e sem retorno	1,00	-	,052	-
11	Método principal	,992	,990	,076	,060
15	Entrada de array, instanciação, invocação de método e saída de array	-	-	-	-
17	Entrada, invocação de método e saída	,776	,763	,410	,411
23	Leitura de array de tipo primitivo	,751	,657	,429	,469
25	Operação aritmética	,734	,716	,439	,449
26	Repetição	,664	,688	,476	,476
27	Seleção de caso	,706	,688	,436	,447
29	Seleção de caso com operação aritmética	,754	,889	,423	,329

Fonte: Elaborado pelo autor.

A Tabela 12 mostra os resultados do ajuste de parâmetros para o ensino superior. É possível verificar em ambas as tabelas que a quantidade de componentes de conhecimento estudados na graduação foi superior à do ensino técnico.

Tabela 12 – Resultados da validação cruzada *k-fold* para o modelo BKT do ensino superior

#	Componente de conhecimento	Acurácia		RMSE	
		k	k	k	k
		5	10	5	10
1	Atributo de classe	,897	,899	,297	,294
2	Classe	,987	,988	,104	,094
3	Método construtor com inicialização de atributo	,861	,870	,313	,305
4	Método construtor com inicialização de atributo do tipo array	,794	,840	,368	,340
5	Método construtor com inicialização de atributo herdado	,842	,858	,322	,288
6	Método construtor padrão	,919	,928	,257	,209
7	Método de classe com parâmetro e com retorno	1,00	1,00	,000	,000
8	Método de classe com parâmetro e sem retorno	1,00	1,00	,000	,000
9	Método de classe sem parâmetro e com retorno	,999	,999	,011	,008
10	Método de classe sem parâmetro e sem retorno	,995	,993	,033	,028
11	Método principal	1,00	1,00	,000	,000

12	Métodos de modificação e de acesso	,845	,858	,353	,338
13	Subclasse	,874	,880	,328	,307
14	Método toString	,805	,808	,375	,365
15	Entrada de array, instanciação, invocação de método e saída de array	,626	,630	,482	,486
16	Entrada, instanciação, invocação de método e saída	,746	,754	,413	,412
17	Entrada, invocação de método e saída	,670	,810	,449	,330
18	Escrita de array	,696	,691	,447	,450
19	Algoritmo formal	,741	,862	,472	,405
20	Incremento/decremento de atributo	,825	,826	,362	,370
21	Invocação de método	,730	,733	,433	,417
22	Leitura de array de tipo não primitivo	,706	,731	,440	,419
23	Leitura de array de tipo primitivo	,693	,698	,460	,454
24	Leitura de atributo	,722	,751	,397	,387
25	Operação aritmética	,760	,768	,412	,404
26	Repetição	,601	,636	,485	,480
27	Seleção de caso	,677	,708	,466	,451
28	Seleção de caso com operação aritmética	,618	,676	,492	,468
29	Verificação de tipo de objeto	,743	,781	,427	,396

Fonte: Elaborado pelo autor.

A Tabela 13 apresenta a média dos resultados dos modelos BKT+*forget* geral e por nível de ensino. Os resultados de acurácia e RMSE são similares aos melhores resultados da literatura para modelos BKT, com valores de acurácia próximos de 0,8 e de RMSE próximos de 0,3 (LIN; CHI, 2016; NAGATANI et al., 2019; WANG et al., 2017; YUDELSON; KOEDINGER; GORDON, 2013). Os melhores resultados são destacados em negrito e mostram que todas as melhores médias foram obtidas pelos modelos gerados por nível de ensino. Não foram realizados testes estatísticos para comparar os resultados dos cenários (geral e por nível de ensino). Aparentemente, os modelos gerados para cada nível de ensino não obtiveram resultados tão superiores ao modelo geral.

Tabela 13 – Média dos resultados dos modelos BKT+*forget* geral e por nível de ensino

Modelo BKT+ <i>forget</i>	Acurácia		RMSE	
	k 5	k 10	k 5	k 10
Modelo geral	,821	,825	,326	,314
Modelo do ensino técnico	<b>,831</b>	,812	<b>,299</b>	,327
Modelo do ensino superior	,806	<b>,827</b>	,325	<b>,307</b>

Fonte: Elaborado pelo autor.

Quando são observados os resultados individuais para cada componente de conhecimento, alguns componentes apresentaram acurácias próximas de 1,0 e erros próximos de 0,0, como os CCs de “Classe” e “Método principal”. Outros componentes, no entanto, apresentaram acurácias próximas de 0,6 e erros próximos de 0,5, como os CCs de “Repetição” e “Escrita de arrays”. Erros mais proeminentes são esperados em componentes de conhecimento

de alta granularidade e menos esperados em componentes de menor granularidade, que consistem em conceitos mais atômicos (PELÁNEK, 2017b; WANG et al., 2021). Pode ser mais difícil diminuir a granularidade de componentes de conhecimento em domínios com problemas mais complexos, como programação de computadores (WANG et al., 2021). Componentes de conhecimento como “Repetição” e “Escrita de arrays” podem ser interpretados como componentes de granularidade alta, pois provavelmente o domínio do componente “Repetição” envolve outras habilidades, como lógica condicional, escrita de expressões e delimitação de escopo. O mesmo pode ser dito sobre o componente “Escrita de arrays”, que provavelmente compreende outras habilidades, como estruturas de repetição e expressões relacionais e lógica. Uma dificuldade encontrada para diminuir a granularidade dos componentes de conhecimento é a forma como o Juiz Online do *Feep* realiza suas avaliações para as soluções dos alunos, através de testes unitários. Testes unitários são capazes de avaliar as características funcionais de códigos, mas não sua estrutura interna. Outro desafio bastante conhecido na literatura sobre a diminuição da granularidade dos componentes de conhecimento é que componentes de conhecimentos mais granulares demandam mais quantidade de evidências disponíveis para os ajustes dos parâmetros (PELÁNEK, 2017b).

Tabela 14 – Parâmetros gerados para os componentes de conhecimento no ensino superior

#	Componente de conhecimento	$p(L_0)$	$p(T)$	$p(G)$	$p(S)$	$p(F)$
1	Atributo de classe	,77123	,64492	,12642	,05961	,03102
2	Classe	,90428	,63936	,04784	,00479	,00493
3	Método construtor com inicialização de atributo	,81325	,25284	,17478	,03994	,03234
4	Método construtor com inicial... de atributo do tipo array	,57983	,18446	,02029	,05156	,01956
5	Método construtor com inicialização de atributo herdado	,94149	,23775	,24781	,05083	,04881
6	Método construtor padrão	,80975	,48620	,43112	,02437	,02356
7	Método de classe com parâmetro e com retorno	,99917	,92617	1,0000	0,0000	,01683
8	Método de classe com parâmetro e sem retorno	,98572	,47261	1,0000	,00000	,01642
9	Método de classe sem parâmetro e com retorno	,99688	,84586	,97571	,00033	,01503
10	Método de classe sem parâmetro e sem retorno	,99952	,53829	,48420	,00000	,00439
11	Método principal	,69313	,63562	1,0000	,00000	,26022
12	Métodos de modificação e de acesso	,64496	,55480	,21537	,04899	,07837
13	Subclasse	,86500	,97324	,11482	,04105	,12524
14	Método toString	,68164	,22221	,00948	,03871	,04916
15	Entrada de array, instanciação, invocação ... e saída de array	,75828	,25042	,13000	,03183	,33401
16	Entrada, instanciação, invocação de método e saída	,71225	,33035	,21355	,04502	,13205
17	Entrada, invocação de método e saída	,75563	,10012	,19840	,03929	,67075
18	Escrita de array	,60570	,12577	,10839	,26978	,23496
19	Algoritmo formal	,47217	,06353	,30892	,02136	,02603
20	Incremento/decremento de atributo	,81443	,24873	,21721	,01848	,08041
21	Invocação de método	,80265	,49263	,12374	,02139	,20420

22	Leitura de array de tipo não primitivo	,03841	,11244	,07153	,20794	,17787
23	Leitura de array de tipo primitivo	,71675	,28033	,11993	,11455	,15244
24	Leitura de atributo	,75258	,24775	,01261	,04542	,03022
25	Operação aritmética	,89676	,19561	,09625	,08831	,08274
26	Repetição	,72562	,21399	,09653	,15403	,29299
27	Seleção de caso	,66099	,18051	,07835	,20904	,09525
28	Seleção de caso com operação aritmética	,90497	,11180	,10067	,25908	,08843
29	Verificação de tipo de objeto	,72739	,46574	,08915	,07920	,04342

Fonte: Elaborado pelo autor.

### 7.5.2 Parâmetros gerados para os componentes de conhecimento

O modelo BKT+*forget* escolhido para gerar as estimativas de conhecimento do aluno, utilizadas nas amostras dos modelos de aprendizado de máquina da Tese, foi o modelo com os parâmetros ajustados por nível de ensino (técnico e graduação). A Tabela 14 apresenta os valores gerados para os parâmetros de cada componente de conhecimento no ensino superior. A Tabela 15 apresenta os valores gerados para os parâmetros do ensino técnico.

Tabela 15 – Parâmetros gerados para os componentes de conhecimento no ensino técnico

#	Componente de conhecimento	$p(L_0)$	$p(T)$	$p(G)$	$p(S)$	$p(F)$
2	Classe	,93081	,50787	,13794	,03104	,01485
7	Método de classe com parâmetro e com retorno	,79840	,13237	,26525	,00087	,00011
8	Método de classe com parâmetro e sem retorno	,66297	,46265	1,0000	,00000	,18682
11	Método principal	,86032	,31174	,35144	,00000	,00040
15	Entrada de array, instanciação, invocação ... e saída de array	,25760	,04531	,00000	1,0000	,85909
17	Entrada, invocação de método e saída	,82517	,13323	,10529	,06572	,09880
23	Leitura de array de tipo primitivo	,82540	,02593	,03303	,34787	,16046
25	Operação aritmética	,96062	,14030	,03217	,16477	,14464
26	Repetição	,66456	,15721	,14195	,14775	,33249
27	Seleção de caso	,89378	,08600	,04598	,39000	,11102
28	Seleção de caso com operação aritmética	,98860	,06090	,03062	,16220	,34225

Fonte: Elaborado pelo autor.

A Tabela 16 apresenta as estimativas geradas para um mesmo aluno do ensino superior no componente de conhecimento “Seleção de caso”<sup>4</sup>.

<sup>4</sup> Componente de conhecimento relacionado à estrutura de programação IF...ELSE.

Tabela 16 - Exemplo de estimativas geradas para um aluno do ensino superior

<i>n</i>	Correto	CC	Data e hora	<i>p</i> (Correto)	<i>p</i> ( <i>L<sub>n</sub></i> )
1	Sim	Seleção de caso	16/04/2021 14:01:15.137	,54938	,66099
2	Não	Seleção de caso	16/04/2021 14:24:17.826	,69812	,86972
3	Sim	Seleção de caso	16/04/2021 14:56:00.704	,51781	,61669
4	Sim	Seleção de caso	16/04/2021 15:50:23.241	,69314	,86273
5	Sim	Seleção de caso	28/04/2021 16:46:43.280	,71507	,89350
6	Sim	Seleção de caso	10/05/2021 18:58:56.618	,71705	,89629
7	Sim	Seleção de caso	10/05/2021 19:05:43.857	,71723	,89653
8	Sim	Seleção de caso	10/05/2021 19:30:12.695	,71724	,89655
9	Não	Seleção de caso	10/05/2021 19:44:11.235	,71724	,89656
10	Sim	Seleção de caso	10/05/2021 19:46:37.254	,54907	,66055
11	Não	Seleção de caso	14/05/2021 16:10:51.931	,69808	,86966
12	Sim	Seleção de caso	14/05/2021 16:12:44.090	,51774	,61659
13	Não	Seleção de caso	16/06/2021 18:16:10.605	,69313	,86272
14	Não	Seleção de caso	16/06/2021 18:18:56.981	,51029	,60614
15	Não	Seleção de caso	16/06/2021 18:21:35.391	,34052	,36790
-	-	-	-	,26717	,26497

Fonte: Elaborado pelo autor.

Os dados apresentados na Tabela 16 são reais e mostram que o aluno fez um total de 15 submissões ( $n = 15$ ) de soluções de objetivos vinculados ao componente. Após cada submissão, os valores de  $\mathbf{p}(\text{Correto})$  e  $\mathbf{p}(\mathbf{L}_n)$  foram atualizados. Veja que o valor de  $\mathbf{p}(\mathbf{L}_n)$  na primeira submissão ( $n = 1$ ) foi de ,66099, que é a probabilidade de o aluno dominar o componente de conhecimento antes da primeira tentativa, ou seja, esse é o valor do parâmetro  $\mathbf{p}(\mathbf{L}_0)$  obtido para os alunos de ensino superior no componente de conhecimento “Seleção de caso” (ver Tabela 14). À medida que o aluno realiza outras submissões de objetivos atrelados ao componente de conhecimento, algumas corretas e outras incorretas, os valores de  $\mathbf{p}(\text{Correto})$  e  $\mathbf{p}(\mathbf{L}_n)$  são atualizados pelo modelo de conhecimento do aluno  $\text{BKT}+\text{forget}$ . A última linha indica qual seria o valor de  $\mathbf{p}(\text{Correto})$  e  $\mathbf{p}(\mathbf{L}_n)$  se o aluno submetesse mais uma solução de objetivo atrelado ao componente “Seleção de caso”.

## 8 GERAÇÃO DE AMOSTRAS E OBTENÇÃO DE RÓTULOS DE CONFUSÃO

Para alcançar seus objetivos, a Tese treinou diferentes modelos de aprendizado de máquina supervisionada utilizando amostras de dados compostas por dados de interação dos alunos com o ambiente de programação + dados sobre estimativas de conhecimento do aluno (modelos da Tese), e diferentes modelos de aprendizado de máquina supervisionada utilizando amostras de dados compostas apenas por dados de interação dos alunos com o ambiente de programação (modelos *baseline*). Os modelos *baseline* representam a abordagem dos trabalhos relacionados (*data-driven*), enquanto os modelos da Tese representam a abordagem apresentada na hipótese da Tese (*data-driven + theory-guided*). A única diferença entre os modelos *baseline* e os modelos da Tese é que as amostras usadas nos modelos da Tese possuem atributos de dados sobre estimativas de conhecimento do aluno, enquanto os modelos *baseline*, não.

Este capítulo tem o objetivo de descrever as características das amostras dos modelos da Tese e dos modelos *baseline*, e como foram geradas. O capítulo também descreve como foram obtidos os rótulos de emoções (confusão e não confusão), que foram atrelados às amostras. As **amostras + rótulos de emoções** foram usados no treinamento e validação dos modelos de aprendizado de máquina da Tese e *baseline*. A Seção 8.1 descreve o início do processo de geração das amostras, a obtenção dos dados de interação dos alunos com o *BlueJ Afetivo*. A Seção 8.2 descreve como foram obtidos os rótulos de confusão (confusão e não confusão). A Seção 8.3 descreve a etapa final de geração das amostras. São apresentados detalhes sobre a geração de cada atributo de dado que compõe as amostras da Tese e *baseline*.

### 8.1 Coleta de dados de interação com o *BlueJ Afetivo*

A coleta de dados de interação dos alunos participantes com o *BlueJ Afetivo* foi realizada entre os meses de fevereiro e julho de 2021, enquanto os alunos resolviam os exercícios de programação atribuídos pelos seus professores. Em função da pandemia de COVID-19, os alunos utilizaram o *BlueJ Afetivo* em seus computadores pessoais, conectados à internet. Os dados de interação dos alunos foram mantidos como registros de *log* em um banco de dados interno do *BlueJ Afetivo*, e eram enviados a um servidor externo, via conexão com a internet, a cada 1 minuto. Quando a conexão com a internet do aluno ficava indisponível, o *BlueJ Afetivo* represava os *logs* e os enviava ao servidor externo assim que a conexão com a internet fosse disponibilizada.

Quadro 7 – Dados mantidos em cada registro de *log* do *BlueJ Afetivo*

Dado	Descrição
Id do log	Identificação do <i>log</i> .
Id do aluno	Identificação do aluno.
Id do exercício	Identificação do exercício.
Id da turma	Identificação da turma.
Tipo de evento	Identificação do tipo de evento que gerou o <i>log</i> , como digitação de tecla, movimento de mouse, submissão de solução de objetivo de exercício para avaliação, <i>feedback</i> da avaliação da solução para um objetivo de exercício, erro de sintaxe na codificação e relato de confusão.
Data e hora	Data e hora da ocorrência do evento que gerou o <i>log</i> .
Tecla pressionada	Identificação da tecla pressionada. Gerado apenas em eventos de teclado no editor de código-fonte.
Evento de teclado	Identificação de evento de teclado. Se o evento de teclado foi de pressionamento ou liberação de tecla. Gerado apenas em eventos de teclado no editor de código-fonte.
Evento de mouse	Identificação de evento de mouse. Se o evento de mouse foi de início de movimento, parada de movimento ou de clique de mouse. Os eventos de movimentos são gerados apenas sobre janelas do <i>BlueJ Afetivo</i> . Os eventos de clique do mouse são gerados apenas em eventos no editor de código-fonte.
Eventos no sistema	Identificação de demais eventos no sistema, como abertura de código-fonte, autenticação no <i>BlueJ Afetivo</i> e registro de <i>log</i> no banco de dados.
Geral	Um dado genérico usado na implementação para diversas anotações úteis para as etapas posteriores de geração das amostras de dados.

Fonte: Elaborado pelo autor.

Cada registro de *log* foi gerado a partir de um dos seguintes eventos no *BlueJ Afetivo*: a) tecla pressionada ou liberada pelo aluno no editor de código-fonte; b) clique do *mouse* pelo aluno no editor de código-fonte; c) movimento do *mouse* (início ou parada) pelo aluno sobre qualquer tela; e) erro de sintaxe gerado durante a digitação de código pelo aluno no editor de código-fonte; f) submissão de solução pelo aluno para avaliação do Juiz Online; g) resposta da avaliação do Juiz Online sobre a submissão de solução do aluno; h) autorrelato de confusão do aluno ao pressionar o botão “CONFUSO”; i) eventos gerais no sistema: abertura de arquivo de código-fonte para edição, seleção de arquivo de código-fonte para edição; criação de arquivo de código-fonte; fechamento de arquivo de código-fonte; exclusão de arquivo de código-fonte; abertura e fechamento de janelas do sistema; ativação de foco de tela; abertura e fechamento do *BlueJ Afetivo*; autenticação no *BlueJ Afetivo*; falha de autenticação no *BlueJ Afetivo*; registro de *log* no banco de dados. O Quadro 7 apresenta uma descrição dos dados mais importantes que compõem cada registro de *log*. A Figura 14 mostra um fragmento dos registros de *log* coletados, mantidos em um banco de dados, ordenados pela data e hora dos eventos geradores.

Figura 14 – Fragmento de registros de *log* coletados e mantidos em banco de dados

id	user	exercise_id	class_id	type	date_time	key_value	key_event	mouse_event
11374135		227	40	MouseStartMoveAtCode	2021-05-30 17:22:50.516	NULL	NULL	MOUSE_START_MOVE_AT_CODE
11374136		227	40	MouseEndMoveAtCode	2021-05-30 17:22:50.687	NULL	NULL	MOUSE_END_MOVE_AT_CODE
11374137		227	40	MouseStartMoveAtCode	2021-05-30 17:22:50.756	NULL	NULL	MOUSE_START_MOVE_AT_CODE
11374138		227	40	MouseEndMoveAtCode	2021-05-30 17:22:51.536	NULL	NULL	MOUSE_END_MOVE_AT_CODE
11374139		227	40	MouseStartMoveAtCode	2021-05-30 17:22:54.105	NULL	NULL	MOUSE_START_MOVE_AT_CODE
11374140		227	40	MouseEndMoveAtCode	2021-05-30 17:22:54.256	NULL	NULL	MOUSE_END_MOVE_AT_CODE
11374141		227	40	MouseStartMoveAtCode	2021-05-30 17:22:54.290	NULL	NULL	MOUSE_START_MOVE_AT_CODE
11374142		227	40	MouseEndMoveAtCode	2021-05-30 17:22:54.704	NULL	NULL	MOUSE_END_MOVE_AT_CODE
11374143		227	40	KeyStrokeAtCode	2021-05-30 17:22:58.175	f	KEY_PRESSED	NULL
11374144		227	40	KeyStrokeAtCode	2021-05-30 17:22:58.348	f	KEY_RELEASED	NULL
11374145		227	40	KeyStrokeAtCode	2021-05-30 17:22:58.364	o	KEY_PRESSED	NULL
11374146		227	40	KeyStrokeAtCode	2021-05-30 17:22:58.484	r	KEY_PRESSED	NULL
11374147		227	40	KeyStrokeAtCode	2021-05-30 17:22:58.504	o	KEY_RELEASED	NULL
11374148		227	40	KeyStrokeAtCode	2021-05-30 17:22:58.594	r	KEY_RELEASED	NULL
11374149		227	40	KeyStrokeAtCode	2021-05-30 17:22:58.811	Space	KEY_PRESSED	NULL
11374150		227	40	KeyStrokeAtCode	2021-05-30 17:22:58.894	Space	KEY_RELEASED	NULL
11374151		227	40	KeyStrokeAtCode	2021-05-30 17:22:59.701	Shift	KEY_PRESSED	NULL
11374152		227	40	KeyStrokeAtCode	2021-05-30 17:22:59.876	(	KEY_PRESSED	NULL
11374153		227	40	KeyStrokeAtCode	2021-05-30 17:22:59.945	(	KEY_RELEASED	NULL

Fonte: Elaborado pelo autor.

A Tabela 17 apresenta a quantidade de registros de *log* coletados para cada turma participante, considerando apenas os registros dos 62 alunos participantes da pesquisa. Foi coletado um total de 13,526664 milhões de registros.

Tabela 17 – Quantidade de registros de *log* coletada em cada turma

Turma	Quantidade de logs	Percentual de logs
Técnico 1	1825548	0,13495
Técnico 2	3195135	0,23621
Superior 1	300319	0,02220
Superior 2	5998318	0,44344
Superior 3	2207344	0,16318
<b>Total</b>	<b>13526664</b>	<b>1</b>

Fonte: Elaborado pelo autor.

## 8.2 Coleta de rótulos de confusão

Esta seção descreve como foram obtidos os rótulos de confusão, usados como rótulos *ground-truth* nos modelos de aprendizado de máquina supervisionada da Tese e *baseline*.

Os trabalhos relacionados geraram rótulos de emoções através de autorrelato retrospectivo de emoções dos próprios estudantes (BOSCH; CHEN; D’MELLO, 2014; TIAM-LEE; SUMI, 2018, 2019) ou através de codificadores humanos que observaram os estudantes retrospectivamente, assistindo vídeos de alunos interagindo com o sistema (FELIPE et al., 2012; VEA; RODRIGO, 2017c) ou analisando registros de *log* das atividades dos aprendizes no sistema (LEE et al., 2011).

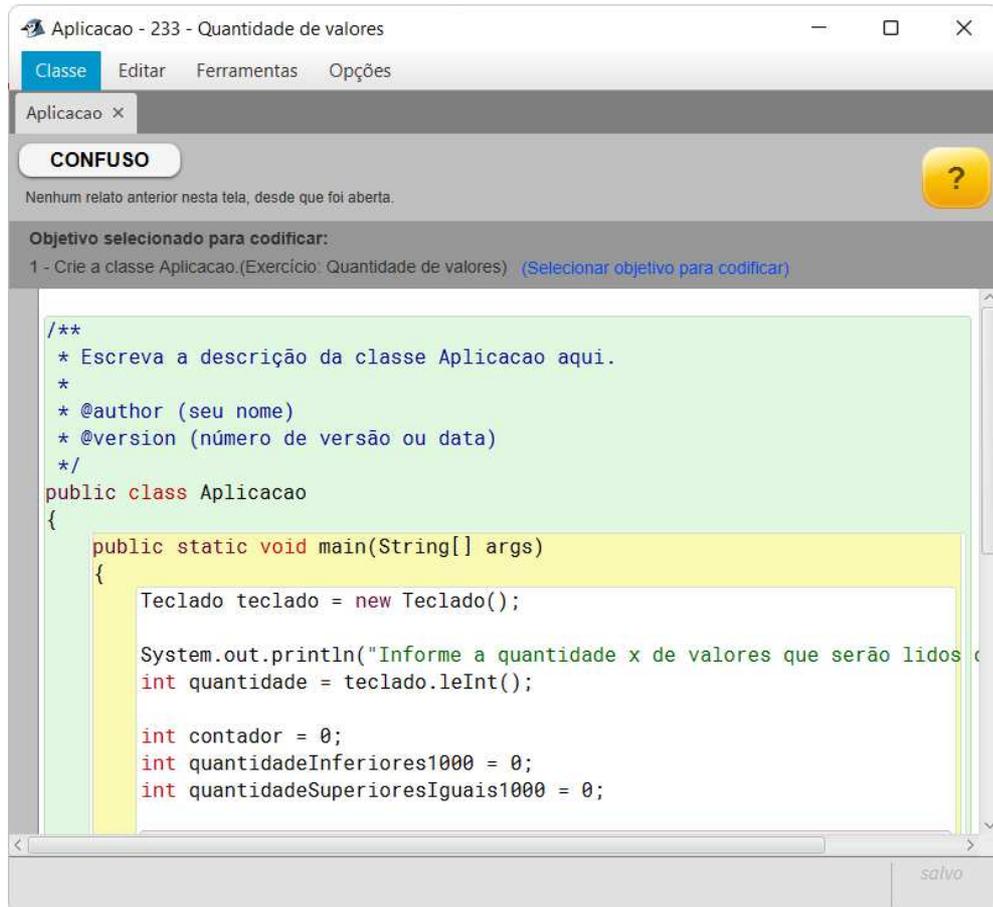
### 8.2.1 *EmAP-ML* - Protocolo de anotação de emoções desenvolvido para a Tese

A Seção 3.2 descreve as restrições conhecidas na literatura, relativas aos métodos de obtenção de rótulos de emoções, usados nos trabalhos relacionados. O autor da Tese e colegas de pesquisa, desenvolveram o protocolo *EmAP-ML*, do inglês *Emotions Annotation Protocol for Machine Learning* (MORAIS et al., 2019). O objetivo foi desenvolver um método que possibilitasse a obtenção de rótulos *ground-truth* confiáveis para modelos de aprendizado de máquina. Neste protocolo, juízes humanos anotam as emoções e comportamentos de aprendizagem dos estudantes a partir da observação retrospectiva de vídeos compostos por imagens do rosto do aluno e da tela do computador, além do áudio ambiente. Os registros de áudio ambiente e da tela e rosto do aluno fornecem sinais que ajudam o juiz a inferir o *appraisal* das emoções dos estudantes, auxiliando na identificação das emoções, mesmo quando as expressões faciais são sutis. O protocolo apresenta orientações sobre a preparação da coleta de rótulos e orienta sobre as etapas de treinamento dos juízes e de teste das anotações. O protocolo também apresenta diretrizes para a fase de anotações das emoções. Uma característica importante do protocolo são as orientações aos juízes sobre a identificação de eventos de *appraisal* de emoções e expressões comportamentais comuns relacionadas às emoções. O método foi desenvolvido durante o primeiro e segundo anos do Doutorado e um artigo sobre o método foi publicado em uma conferência (MORAIS et al., 2019).

### 8.2.2 Mudança de método de obtenção de rótulos de emoções

Até dois meses antes do início da coleta de dados nas turmas participantes, entre os meses de fevereiro e julho de 2021, o método selecionado pela Tese para a coleta dos rótulos de confusão foi o *EmAP-ML*. No entanto, com a manutenção das aulas no regime remoto nas instituições participantes, a aplicação do protocolo se tornou inviável, pois depende da coleta de vídeo e áudio gravados em salas de aula com computadores onde os alunos realizam as tarefas de programação. Chegou a ser avaliada a possibilidade de os próprios alunos gravarem os vídeos e áudios em seus computadores pessoais, mas isso exigiria dos alunos a disponibilidade de *webcams* e recursos computacionais compatíveis. Também demandaria atividades extras de colaboração dos alunos com a pesquisa.

Figura 15 – Botão na interface do *BlueJ* Afetivo para o autorrelato de confusão



Fonte: Elaborado pelo autor.

Com a inviabilidade de aplicação do protocolo *EmAP-ML*, o método escolhido para a coleta de rótulos de confusão foi o de autorrelato dos próprios alunos, realizado de forma concorrente durante as tarefas de aprendizagem. Foi avaliado que esta seria uma abordagem viável de ser implementada nas aulas remotas. Para implementar o autorrelato dos alunos, foi necessário fazer alterações na interface gráfica do *BlueJ Afetivo*. Todas as telas do *BlueJ Afetivo* receberam um botão de relato de confusão. A Figura 15 exibe a tela de codificação do *BlueJ Afetivo* com um botão de relato de confusão na área superior esquerda. Para relatar sua confusão, o aluno precisa clicar no botão CONFUSO. Foi possibilitada ainda uma opção na interface para cancelamento do último relato de confusão, para o caso de o aluno ter apertado o botão por engano ou para “testar” a funcionalidade. Os alunos participantes foram orientados a apertar o botão apenas quando estivessem sentindo a confusão, e não apertar o botão quando a confusão já tivesse passado.

Apesar da viabilidade de aplicação deste método no contexto da pandemia, com os alunos utilizando o *BlueJ Afetivo* em seus computadores pessoais, a utilização deste método gerou uma dúvida principal: os alunos se engajariam na tarefa de relatar sua confusão? Como

descrito anteriormente, na Seção 3.2, uma das desvantagens dos métodos de autorrelato de emoções é que os alunos acabam precisando se engajar tanto na tarefa de aprendizagem quanto na tarefa de relato da emoção, aumentando a carga cognitiva. Além disso, os alunos podem não se engajar nas tarefas de autorrelato (PORAYSKA-POMSTA et al., 2013). A suposição do autor da Tese é que quando o aluno estivesse experimentando uma confusão, mas ainda estivesse suficientemente engajado na tarefa principal, ele poderia esquecer de relatar sua confusão, e muitos momentos de confusão não seriam relatados. No entanto, se o aluno estivesse enfrentando uma confusão mais forte que o deixaria travado na tarefa, ou uma confusão mais duradoura, o aluno poderia desengajar da tarefa principal e lembrar de relatar sua confusão. Esta suposição não foi verificada na Tese. A subnotificação de confusão pelos alunos poderia diminuir o desempenho dos modelos de detecção de confusão, pois os modelos poderiam receber amostras de treinamento vinculados a rótulos de não confusão quando, na verdade, fossem amostras relativas à confusão.

Com o objetivo de lembrar os alunos a relatarem sua confusão, foi orientado aos professores das turmas participantes que a cada início de aula os alunos fossem lembrados sobre o relato da confusão durante as tarefas de aprendizagem. Além disso, no início da coleta de dados, nos meses de fevereiro e março de 2021, o pesquisador participou de uma aula em cada turma participante e promoveu uma discussão com os alunos sobre o que é a emoção confusão e sobre como reconhecê-la no contexto de tarefas de programação de computadores. As discussões foram baseadas nas teorias subjacentes da Tese sobre emoções. O pesquisador também disponibilizou um vídeo aos alunos, antes do início da participação deles (também disponível na interface do *BlueJ* Afetivo), com orientações sobre o que é a emoção confusão e sobre como identificá-la. Também foi implementado um sinal piscante no botão, a cada 10 minutos, para lembrar os alunos para relatarem a confusão, quando a sentissem.

Uma característica interessante da abordagem adotada é que os relatos de confusão são obtidos de maneira espontânea. Já foi verificado que quando a coleta de rótulos *ground-truth* é realizada em momentos fixos, de forma forçada, apesar de obter mais rótulos de emoção, o desempenho dos modelos classificadores tende a piorar em comparação quando são treinados com rótulos obtidos espontaneamente (BOSCH; CHEN; D’MELLO, 2014).

Tabela 18 – Quantidade de autorrelatos de confusão em cada turma

<b>Turma</b>	<b>Quantidade de relatos de confusão</b>	<b>Percentual de relatos de confusão</b>
Técnico 1	162	,14123

Técnico 2	433	,37750
Superior 1	51	,04446
Superior 2	206	,17959
Superior 3	295	,25719
<b>Total</b>	<b>1147</b>	<b>1</b>

Fonte: Elaborado pelo autor.

A Tabela 18 mostra a quantidade de autorrelatos de confusão obtidos em cada turma. Foi obtido um total de 1147 anotações de confusão. É possível verificar que foram obtidos mais relatos de confusão nas turmas de ensino técnico. Isso pode ter ocorrido em função de um viés na coleta de dados. O autor da Tese foi o professor nas turmas de ensino técnico, de forma que os alunos podem ter sido mais lembrados sobre a tarefa de relato da confusão, do que nas turmas de ensino superior, em que o autor não foi o professor. Uma outra suposição é que para estes alunos, com menos experiência anterior de programação, como evidenciado na Seção 5.2, os exercícios podem ter sido mais desafiadores e pode ter ocorrido mais incidência de confusão. Nem todos os relatos de confusão da Tabela 18 foram utilizados como rótulos *ground-truth* para serem vinculados às amostras dos modelos de aprendizado de máquina, por motivos que serão descritos na próxima Seção.

### 8.3 Geração de amostras para os modelos de aprendizado de máquina

Esta Seção descreve como foi realizada a geração das amostras para os modelos de aprendizado de máquina da Tese e *baseline*. Os atributos de dados das amostras foram gerados a partir de duas fontes: dados de interação dos alunos com o *BlueJ* Afetivo, descritos na Seção 8.1; dados do modelo de conhecimento do aluno *BKT+forget*.

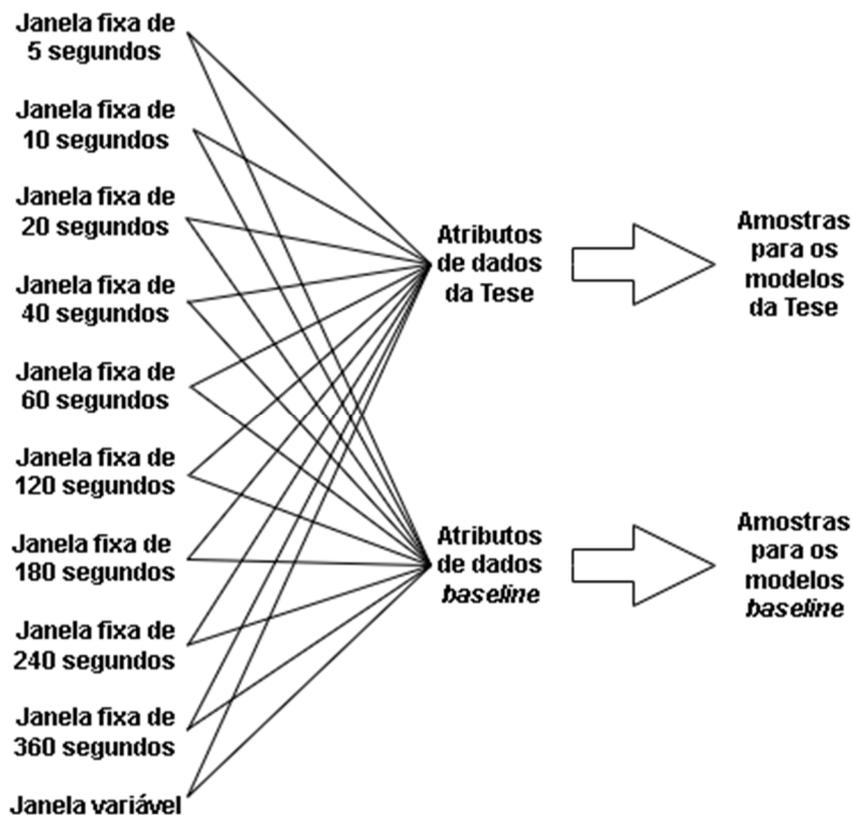
As amostras dos modelos da Tese e dos modelos *baseline* se diferenciam da seguinte forma: os **modelos da Tese**, que representam a abordagem apresentada na hipótese da Tese, utilizam amostras compostas por atributos de dados sobre a interação do aluno com o ambiente de programação + dados sobre estimativas de conhecimento do aluno; os **modelos *baseline***, que representam os trabalhos relacionados, utilizam amostras compostas apenas por atributos de dados sobre a interação dos alunos com o ambiente de programação.

Cada atributo de dado gerado para uma amostra representa algum aspecto de comportamento do aluno, como a velocidade de digitação, a quantidade de movimentos de mouse ou a estimativa de conhecimento do aluno nos componentes de conhecimento atrelados ao objetivo de exercício que o aluno estava codificando. Outra característica de um atributo de dado é que ele sintetiza aquele comportamento em determinado intervalo de tempo. As

subquestões de pesquisa SQP1 e SQP2 têm interesse em verificar os efeitos de diferentes tamanhos de janelas de observação do comportamento do aluno. Foram modeladas janelas de tamanho fixo (5, 10, 20, 40, 60, 90, 120, 180, 240 e 360 segundos) e variável. Os trabalhos relacionados consideraram janelas de, no máximo, 20 segundos, mas a Tese gostaria de testar diferentes configurações, incluindo tamanhos de tempo superiores a 20 segundos.

A Figura 16 apresenta as 20 diferentes configurações dos conjuntos de amostras geradas, sendo 10 diferentes configurações de tamanhos de janelas para os atributos de dados que refletem a hipótese da Tese e outras 10 diferentes configurações de tamanhos de janelas para os atributos de dados que refletem os trabalhos relacionados (*baseline*).

Figura 16 – Diferentes configurações das amostras geradas



Fonte: Elaborado pelo autor.

O processo de geração das amostras foi realizado em três etapas: a) **identificação de sequências de logs**; b) **geração de janelas**; e c) **geração de amostras**. As três etapas são descritas nas próximas Seções.

### 8.3.1 Identificação de sequências de *logs*

A primeira etapa no processo de geração das amostras consistiu em identificar sequências de *logs* do *BlueJ* Afetivo quando havia algum objetivo de exercício selecionado pelo aluno. Durante a interação do aluno com o *BlueJ* Afetivo, o aluno deveria selecionar o objetivo de exercício a ser solucionado, antes de qualquer codificação. Desta forma, alguns *logs* foram coletados quando nenhum objetivo de exercício estava selecionado. Ao estudo apenas interessam os *logs* gerados em momentos em que um objetivo de exercício estava selecionado para resolução pelo aluno, pois somente assim seria possível identificar os componentes de conhecimento que o aluno estava engajado (cada objetivo de exercício pode estar vinculado a, no máximo, dois componentes de conhecimento). Além disso, um objetivo de exercício selecionado na interface do *BlueJ* indica que naquele momento o aluno estava engajado na solução de um exercício.

Figura 17 – Representação de intervalos de *logs* atrelados e não atrelados a objetivos



Fonte: Elaborado pelo autor.

A Figura 17 mostra uma representação sobre como as sequências de *logs* foram identificadas. Considere que cada ponto na figura representa um *log* e os espaços entre eles o intervalo de tempo entre eles. Os pontos pertencentes aos intervalos em vermelhos representam *logs* gerados quando não havia um objetivo selecionado pelo aluno, e são desconsiderados na pesquisa. Os pontos pertencentes aos intervalos em verde representam *logs* gerados quando havia um objetivo selecionado, e são considerados na pesquisa. Foram identificadas 15,571 mil sequências de *logs* atreladas a objetivos de exercícios e outras 15,593 mil não atreladas. A Tabela 19 mostra a quantidade de sequências de *logs* atreladas a objetivos de exercício, em cada turma participante.

Tabela 19 – Quantidades de seqüências de *logs* atreladas a objetivos de exercício

<b>Turma</b>	<b>Quantidade de seqüências de logs</b>	<b>Percentual</b>
Técnico 1	1495	,09601
Técnico 2	2504	,16081
Superior 1	651	,04180
Superior 2	7816	,50195
Superior 3	3105	,19940
<b>Total</b>	<b>15571</b>	<b>1</b>

Fonte: Elaborado pelo autor.

A Figura 18 mostra uma seqüência de *logs* atrelada a um determinado objetivo de exercício. Para simplificação, foram omitidas algumas colunas e *logs* intermediários.

Figura 18 – Exemplo de seqüência de *logs* atrelada a um objetivo de exercício

id	id_user	type	system_action	date_time
12601706	507	ActionAtTheSystem	SelectObjectiveForCoding	2021-06-04 14:32:17.373
12601707	507	MouseStartMoveAtObjectives	NULL	2021-06-04 14:32:17.774
12601708	507	MouseStartMoveAtCode	NULL	2021-06-04 14:32:18.344
12601709	507	MouseEndMoveAtObjectives	NULL	2021-06-04 14:32:18.444
12601710	507	MouseEndMoveAtCode	NULL	2021-06-04 14:32:18.808
12601711	507	MouseStartMoveAtCode	NULL	2021-06-04 14:32:18.911
12601712	507	ActionAtTheSystem	SelectClass	2021-06-04 14:32:19.456
12601713	507	MouseEndMoveAtCode	NULL	2021-06-04 14:32:19.556
12601714	507	MouseClickedAtCode	NULL	2021-06-04 14:32:19.640
12601715	507	MouseStartMoveAtCode	NULL	2021-06-04 14:32:19.743
12601716	507	MouseEndMoveAtCode	NULL	2021-06-04 14:32:19.959
12601717	507	KeyStrokeAtCode	NULL	2021-06-04 14:32:20.850
12601718	507	KeyStrokeAtCode	NULL	2021-06-04 14:32:20.983
.....				
12612200	507	MouseStartMoveAtObjectives	NULL	2021-06-04 15:40:02.333
12612201	507	MouseEndMoveAtObjectives	NULL	2021-06-04 15:40:02.536
12612204	507	ActionAtTheSystem	ClickButtonSendEvaluation	2021-06-04 15:40:03.048
12612206	507	ActionAtTheSystem	ClickButtonConfirmSendE...	2021-06-04 15:40:04.776
12612207	507	ActionAtTheSystem	AttemptToSubmitForEval...	2021-06-04 15:40:04.893
12612208	507	ActionAtTheSystem	AttemptToSubmitForEval...	2021-06-04 15:40:05.009
12612209	507	MouseStartMoveAtObjectives	NULL	2021-06-04 15:40:05.010
12612212	507	CaseTestSuccess	NULL	2021-06-04 15:40:05.197
12612213	507	CaseTestSuccess	NULL	2021-06-04 15:40:05.197
12612214	507	CaseTestSuccess	NULL	2021-06-04 15:40:05.197
12612215	507	CaseTestSuccess	NULL	2021-06-04 15:40:05.197
12612216	507	CaseTestSuccess	NULL	2021-06-04 15:40:05.197
12612216	507	CaseTestSuccess	NULL	2021-06-04 15:40:05.197

Fonte: Elaborado pelo autor.

A seqüência de *logs* inicia quando o aluno seleciona um objetivo de exercício na interface gráfica do *BlueJ* Afetivo. O primeiro *log* registra esta seleção de objetivo. Estas

sequências podem encerrar em algumas ocasiões: quando o aluno fecha o *BlueJ* Afetivo; quando o aluno submete para avaliação uma solução correta para o objetivo; quando o aluno cancela a seleção de objetivo de exercício. No exemplo da Figura 18, a sequência é encerrada nos *logs* que registram que a solução do aluno para o objetivo de exercício obteve sucesso para todos os casos de teste executados pelo Juiz Online.

### 8.3.2 Identificação de janelas de *logs*

Esta etapa consiste na identificação de sequências de *logs* que se enquadram dentro de determinados intervalos de tempo, também chamados de janelas. As sequências de *logs* foram identificadas para as seguintes configurações de janelas de tempo fixo: 5, 10, 20, 40, 60, 120, 180, 240 e 360 segundos. Nesta etapa, os rótulos de “confusão”, obtidos através do autorrelato dos alunos (ver Seção 8.2.2), e os rótulos de “não confusão”, são vinculados aos dados. Os rótulos são vinculados às sequências de *logs*.

Figura 19 - Representação de sequências de *logs* para uma janela de 5 segundos atreladas ao rótulo de “confusão”

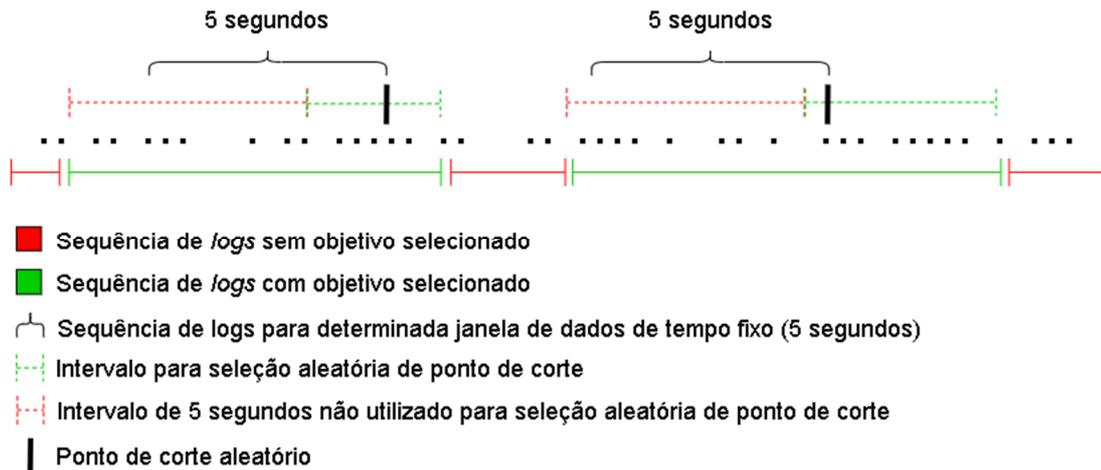


Fonte: Elaborado pelo autor.

A Figura 19 mostra uma representação sobre como foram identificadas as sequências de *logs* para uma janela de tempo fixo de 5 segundos, atreladas ao rótulo de “confusão”. Primeiramente, são identificadas todas as sequências de *logs* com objetivos atrelados que possuem pelo menos um *log* de relato de confusão. São desconsideradas sequências de *log* que possuem duração inferior ao tamanho de janela alvo, que aqui é de 5 segundos. Em cada sequência de *log* identificada, é verificado se, a partir do relato de confusão (“C”), há uma sequência anterior de 5 segundos de *logs*. Se a resposta é positiva, todos os *logs* identificados naqueles 5 segundos anteriores ao relato de confusão compõem uma sequência de *logs* atrelada

a uma janela de 5 segundos e atrelada a um rótulo de “confusão”. O mesmo procedimento foi realizado para os demais tamanhos de janelas (10, 20, 40, 60, 90, 120, 180, 240 e 320 segundos).

Figura 20 - Representação de sequências de *logs* para uma janela de 5 segundos atreladas ao rótulo de “não confusão”



Fonte: Elaborado pelo autor.

A Figura 20 mostra uma representação sobre como foram identificadas as sequências de logs para uma janela de tempo fixo de 5 segundos, atreladas ao rótulo de “não confusão”. É necessário lembrar que os autorrelatos dos alunos na interface do *BlueJ Afetivo* não geraram rótulos de “não confusão”, apenas rótulos de “confusão”, pois não foi solicitado ao aluno que indicasse quando não estava confuso. O trabalho assume que o aluno não estava confuso quando não relatou sua confusão. Primeiramente, são identificadas todas as sequências de *log* com objetivos atrelados que não possuem qualquer *log* de relato de confusão. São desconsideradas sequências de *log* que possuem duração inferior ao tamanho de janela alvo, que aqui é de 5 segundos. Em cada sequência de *log* identificada, foi selecionado, aleatoriamente, um ponto de corte. O ponto de corte fica localizado dentro da subsequência de *logs* que não pertence aos 5 primeiros segundos da sequência de *log*. Todos os *logs* identificados nos 5 segundos anteriores ao ponto de corte, compõe uma sequência de *logs* atrelada a uma janela de 5 segundos e atrelada a um rótulo de “não confusão”. O mesmo procedimento foi realizado para todos os demais tamanhos de janelas (10, 20, 40, 60, 90, 120, 180, 240 e 320 segundos).

Tabela 20 – Quantidade de sequências de *logs* para as janelas de tamanho fixo

Tamanho da janela de observação	Nível de ensino	Quantidade “confusão”	Quantidade “não confusão”
5	Superior	466	466
	Técnico	497	497
10	Superior	453	453
	Técnico	496	496
20	Superior	450	450
	Técnico	494	494
40	Superior	445	445
	Técnico	488	488
60	Superior	440	440
	Técnico	484	484
90	Superior	430	430
	Técnico	472	472
120	Superior	418	418
	Técnico	465	465
180	Superior	392	392
	Técnico	436	436
240	Superior	357	357
	Técnico	414	414
360	Superior	304	304
	Técnico	385	385
<b>Total</b>		<b>8786</b>	<b>8786</b>

Fonte: Elaborado pelo autor.

A Tabela 20 mostra a quantidade de sequências de *logs* identificadas para cada configuração de janela de tamanho fixo, em cada nível de ensino. Para cada nível de ensino e configuração de janela, foi possível balancear a quantidade de sequências de *logs* para cada rótulo de emoção (“confusão” e “não confusão”), uma característica que vai se refletir na geração das amostras e que é importante para o bom funcionamento dos modelos.

A Figura 21 mostra uma representação sobre como foram identificadas as sequências de *logs* para uma janela de tempo variável, atreladas ao rótulo de “confusão”. Primeiramente, são identificadas todas as sequências de *log* com objetivos atrelados que possuem pelo menos um *log* de relato de confusão. São desconsideradas sequências de *log* que possuem duração inferior a 5 segundos, o tempo mínimo das janelas de tempo variável. Em cada sequência de *log* identificada, é verificado se, a partir do relato de confusão (“C”), há uma sequência anterior de 5 segundos de *logs*. Se a resposta é positiva, toda a sequência de *logs* até o relato de confusão, compõe uma sequência de *logs* atrelada a uma janela de tempo variável e atrelada a um rótulo de “confusão”. No exemplo da figura, foram formadas duas sequências de *logs* de janela variável, uma de 6 segundos, e outra de 7 segundos, ambas atreladas ao rótulo de “confusão”.

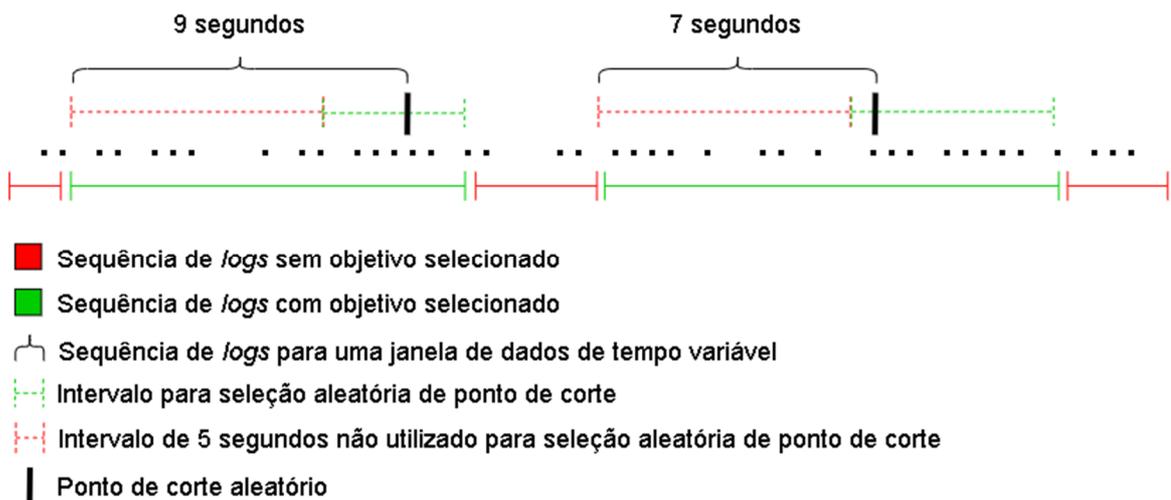
Figura 21 - Representação de sequências de *logs* para uma janela variável atreladas ao rótulo de “confusão”



Fonte: Elaborado pelo autor.

A Figura 22 mostra uma representação sobre como foram identificadas as sequências de logs para uma janela de tempo variável, atreladas ao rótulo de “não confusão”. Primeiramente, são identificadas todas as sequências de log com objetivos atrelados que não possuem qualquer log de relato de confusão. São desconsideradas sequências que possuem duração inferior a 5 segundos, o tempo mínimo das janelas de tempo variável. Em cada sequência de log, foi selecionado, aleatoriamente, um ponto de corte. O ponto de corte fica localizado dentro da subsequência de logs que não pertence aos 5 primeiros segundos da sequência. Os logs identificados do início da sequência até o ponto de corte, compõem uma sequência de logs atrelada a uma janela variável e atrelada a um rótulo de “não confusão”.

Figura 22 - Representação de sequências de logs para uma janela variável atreladas ao rótulo de “não confusão”



Fonte: Elaborado pelo autor.

No exemplo da Figura 22, foram formadas duas sequências de logs de janela variável, uma de 9 segundos, e outra de 7 segundos, ambas atreladas ao rótulo de “não confusão”. A

Tabela 21 mostra a quantidade de sequências de *logs* de janela de tamanho variável identificadas para cada nível de ensino. Foi possível balancear a quantidade de sequências de *logs* para cada rótulo de emoção (“confusão” e “não confusão”), como também foi feito na identificação das sequências de janela fixa (ver Tabela 20).

Tabela 21 – Quantidade de sequências de *logs* para as janelas variáveis

Nível de ensino	Quantidade “confusão”	Quantidade “não confusão”
Superior	466	466
Técnico	497	497
<b>Total</b>	<b>963</b>	<b>963</b>

Fonte: Elaborado pelo autor.

Após a identificação das sequências de *logs*, é possível observar que foram desconsiderados todos os relatos de confusão realizados pelo aluno quando não havia um objetivo de exercício selecionado no *BlueJ* Afetivo. Assim, do total de 1147 relatos de confusão, foram considerados, no máximo, 963, nas janelas de 5 segundos e de tempo variável.

### 8.3.3 Geração de amostras

Esta Seção descreve a etapa final do processo de geração de amostras para os modelos de aprendizado de máquina. É gerada uma amostra para cada sequência de *logs* descrita na Seção anterior. Cada amostra é composta por um conjunto de atributos de dados. Cada atributo de dado sintetiza um determinado comportamento observado na sequência de *logs* ou estimativa de conhecimento do aluno. O Quadro 8 descreve cada um dos 65 atributos gerados. Para cada atributo de dado, é apresentada sua descrição, classificação, para qual amostra foi atribuído (Tese e/ou *baseline*), e como foi gerado. Os atributos são assim classificados: a) **teclado**, derivados de dados sobre pressionamento de tecla (8 atributos); b) **mouse**, derivados de dados sobre cliques e movimentos de mouse (6); c) **erro de sintaxe**, derivados de dados sobre erros de sintaxe na codificação (1); d) **submissão de solução**, derivados de dados sobre as submissões de solução de objetivos para o Juiz Online (5); e) interações gerais (3); e f) **conhecimento do aluno**, derivados das estimativas de conhecimento do aluno, inferidas pelo modelo BKT+*forget* (42).

Quadro 8 – Atributos de dados gerados para as amostras

#	Descrição do atributo de dado	Classificação de Tipo	Amostras em que é atribuído	Como foi gerado
1	Quantidade de teclas pressionadas	Teclado	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem das teclas pressionadas no editor de código, na janela de dados.
2	Quantidade de teclas de <i>backspace</i> pressionadas	Teclado	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem das teclas de <i>backspace</i> pressionadas no editor de código, na janela de dados.
3	Quantidade de teclas <i>delete</i> pressionadas	Teclado	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem das teclas de <i>delete</i> pressionadas no editor de código, na janela de dados.
4	Quantidade de teclas de movimento de cursor pressionadas	Teclado	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem das teclas de movimento de cursor (esquerda, direita, cima, baixo) pressionadas no editor de código, na janela de dados.
5	Quantidade de intervalos de digitação	Teclado	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem dos intervalos de digitação, na janela de dados. Um intervalo de digitação compreende um conjunto de teclas pressionadas em sequência, com intervalo entre as teclas menor que 3 segundos.
6	Velocidade de digitação baixa	Teclado	Tese e <i>baseline</i>	Gerados a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Primeiramente, foi calculado o tempo médio entre teclas dentro dos intervalos de digitação, na janela de dados. Após, foi aplicado o algoritmo de clusterização <i>k-means</i> para agrupamento das médias em 3 grupos ( $k = 3$ ). Cada grupo corresponde, respectivamente, à velocidade de digitação baixa, média e alta. Os atributos 6, 7 e 8 representam os três grupos e seus valores são definidos com a abordagem <i>one-hot-encoding</i> . Quando não há pressionamento de teclas na janela de dados, ambos os atributos recebem o valor 0 (zero).
7	Velocidade de digitação média	Teclado	Tese e <i>baseline</i>	
8	Velocidade de digitação alta	Teclado	Tese e <i>baseline</i>	
9	Quantidade de movimentos de mouse	Mouse	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem dos movimentos de mouse, na janela de dados.
10	Quantidade de cliques no editor de código	Mouse	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem dos cliques de mouse no editor de código, na janela de dados.
11	Tempo total de movimentos de mouse	Mouse	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Um somatório dos tempos de movimento de mouse, na janela de dados.

12	Velocidade de movimentos de mouse baixa	Mouse	Tese e <i>baseline</i>	Gerados a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Primeiramente, foi calculado o tempo médio das movimentações de mouse (tempo do fim do movimento subtraído com o tempo do início do movimento), na janela de dados. Após, foi aplicado o algoritmo de clusterização <i>k-means</i> para agrupamento das médias em 3 grupos ( $k = 3$ ). Cada grupo corresponde, respectivamente, à velocidade de movimentos baixa, média e alta. Os atributos 12, 13 e 14 representam os três grupos e seus valores são definidos com a abordagem <i>one-hot-encoding</i> . Quando não há movimentos de mouse na janela de dados, ambos os atributos recebem o valor 0 (zero).
13	Velocidade de movimentos de mouse média	Mouse	Tese e <i>baseline</i>	
14	Velocidade de movimentos de mouse alta	Mouse	Tese e <i>baseline</i>	
15	Quantidade de erros de sintaxe no código	Erro de sintaxe	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem dos erros de sintaxe no editor de código, na janela de dados.
16	Quantidade de submissões de soluções	Submissões	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem das submissões de soluções de objetivo de exercício para avaliação do Juiz Online, na janela de dados.
17	Quantidade de submissões de soluções com erro	Submissões	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem das submissões de soluções de objetivo de exercício que foram avaliados pelo Juiz Online como soluções com erro (compilação ou testes unitários), na janela de dados.
18	Quantidade de submissões de soluções com sucesso	Submissões	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem das submissões de soluções de objetivo de exercício que foram avaliados pelo Juiz Online como soluções corretas para todos os testes unitários, na janela de dados.
19	Quantidade de submissões de soluções com erro de compilação	Submissões	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem das submissões de soluções de objetivo de exercício que foram avaliados pelo Juiz Online como soluções com erro de compilação, na janela de dados.
20	Quantidade de submissões de soluções com erro de objetivo	Submissões	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem das submissões de soluções de objetivo de exercício avaliados pelo Juiz Online como soluções cujos testes unitários falharam total ou parcialmente, na janela de dados.
21	Tempo ocioso total	Interações	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Um somatório do tempo dos intervalos ociosos, na janela de dados. Um intervalo ocioso inicia quando nenhum <i>log</i> é registrado na janela de dados por mais de 3 segundos.
22	Quantidade de intervalos ociosos	Interações	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem dos intervalos ociosos, na janela de dados.
23	Quantidade de visualizações do enunciado do exercício	Interações	Tese e <i>baseline</i>	Gerado a partir dos <i>logs</i> do <i>BlueJ</i> Afetivo. Uma contagem da quantidade de vezes que a tela com os objetivos do exercício é aberta ou colocada em foco, na janela de dados.

24	Probabilidade de aplicar corretamente o componente de conhecimento primário do objetivo em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Estimativa $p(\textit{Correto})$ do componente de conhecimento primário vinculado ao objetivo de exercício em desenvolvimento, na janela de dados.
25	Probabilidade de dominar o componente de conhecimento primário do objetivo em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Estimativa $p(L)$ do componente de conhecimento primário vinculado ao objetivo de exercício em desenvolvimento, na janela de dados.
26	Probabilidade de transitar do estado não aprendido para aprendido no componente de conhecimento primário do objetivo em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Estimativa $p(T)$ do componente de conhecimento primário vinculado ao objetivo de exercício em desenvolvimento, na janela de dados.
27	Probabilidade de aplicar incorretamente o componente de conhecimento primário do objetivo em andamento, ao dominá-lo	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Estimativa $p(S)$ do componente de conhecimento primário vinculado ao objetivo de exercício em desenvolvimento, na janela de dados.
28	Probabilidade de aplicar corretamente o componente de conhecimento primário do objetivo em andamento, ao não dominá-lo	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Estimativa $p(G)$ do componente de conhecimento primário vinculado ao objetivo de exercício em desenvolvimento, na janela de dados.
29	Probabilidade de ter dominado o componente de conhecimento primário do objetivo em andamento, antes da primeira oportunidade de aplicá-lo	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Estimativa $p(L_0)$ do componente de conhecimento primário vinculado ao objetivo de exercício em desenvolvimento, na janela de dados. Se o aluno estiver aplicando o componente de conhecimento pela primeira vez, então $p(L_0) = p(L)$ .
30	Probabilidade de esquecer o componente de conhecimento primário do objetivo em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Estimativa $p(F)$ do componente de conhecimento primário vinculado ao objetivo de exercício em desenvolvimento, na janela de dados.
31	Média das probabilidades de aplicar corretamente os componentes de conhecimento primário e secundário no objetivo em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(\textit{Correto})$ dos componentes de conhecimento primário e secundário (se houver) do objetivo de exercício em desenvolvimento, na janela de dados. Quando não houver CC secundário para o objetivo, a média será igual ao valor de $p(\textit{Correto})$ do CC primário.
32	Média das probabilidades de dominar os componentes de conhecimento primário e secundário do objetivo em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(L)$ dos componentes de conhecimento primário e secundário (se houver) do objetivo de exercício em desenvolvimento, na janela de dados. Quando não houver CC secundário para o objetivo, a média será igual ao valor de $p(L)$ do CC primário.
33	Média das probabilidades de transitar do estado não aprendido para aprendido nos	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(T)$ dos componentes de conhecimento primário e secundário (se houver) do

	componentes de conhecimento primário e secundário do objetivo em andamento			objetivo de exercício em desenvolvimento, na janela de dados. Quando não houver CC secundário para o objetivo, a média será igual ao valor de $p(T)$ do CC primário.
34	Média das probabilidades de aplicar incorretamente os componentes de conhecimento primário e secundário do objetivo em andamento, ao dominá-los	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(S)$ dos componentes de conhecimento primário e secundário (se houver) do objetivo de exercício em desenvolvimento, na janela de dados. Quando não houver CC secundário para o objetivo, a média será igual ao valor de $p(S)$ do CC primário.
35	Média das probabilidades de aplicar corretamente os componentes de conhecimento primário e secundário do objetivo em andamento, ao não dominá-los	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(G)$ dos componentes de conhecimento primário e secundário (se houver) do objetivo de exercício em desenvolvimento, na janela de dados. Quando não houver CC secundário para o objetivo, a média será igual ao valor de $p(G)$ do CC primário.
36	Média das probabilidades de ter dominado os componentes de conhecimento primário e secundário do objetivo em andamento, antes da primeira oportunidade de aplicá-los	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(L_0)$ dos componentes de conhecimento primário e secundário (se houver) do objetivo de exercício em desenvolvimento, na janela de dados. Quando não houver CC secundário para o objetivo, a média será igual ao valor de $p(L_0)$ do CC primário. Se o aluno estiver aplicando o componente de conhecimento pela primeira vez, então $p(L_0) = p(L)$ .
37	Média das probabilidades de esquecer os componentes de conhecimento primário e secundário do objetivo em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(F)$ dos componentes de conhecimento primário e secundário (se houver) do objetivo de exercício em desenvolvimento, na janela de dados. Quando não houver CC secundário para o objetivo, a média será igual ao valor de $p(F)$ do CC primário.
38	Média das probabilidades de aplicar corretamente todos os componentes de conhecimento do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(Correto)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
39	Média das probabilidades de dominar todos os componentes de conhecimento do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(L)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
40	Média das probabilidades de ter dominado todos os componentes de conhecimento do exercício em andamento, antes da primeira oportunidade de aplicá-los	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(L_0)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.

41	Média das probabilidades de aplicar incorretamente todos os componentes de conhecimento do exercício em andamento, quando os domina	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(S)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
42	Média das probabilidades de aplicar corretamente todos os componentes de conhecimento do exercício em andamento, quando não os domina	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(G)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
43	Média das probabilidades de transitar do estado não aprendido para aprendido em todos os componentes de conhecimento do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(T)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
44	Média das probabilidades de esquecimento de cada componente de conhecimento do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(F)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
45	Desvio padrão das probabilidades de aplicar corretamente todos os componentes de conhecimento do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(Correto)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
46	Desvio padrão das probabilidades de dominar todos os componentes de conhecimento do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(L)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
47	Desvio padrão das probabilidades de ter dominado todos os componentes de conhecimento do exercício em andamento, antes da primeira oportunidade de aplicá-los	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(L_0)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
48	Desvio padrão das probabilidades de aplicar incorretamente todos os componentes de conhecimento do exercício em andamento, quando os domina	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(S)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
49	Desvio padrão das probabilidades de aplicar corretamente todos os componentes de conhecimento do exercício em andamento, quando não os domina	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(G)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.

50	Desvio padrão das probabilidades de transitar do estado não aprendido para aprendido em todos os componentes de conhecimento do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(T)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
51	Desvio padrão das probabilidades de esquecimento de cada componente de conhecimento do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(F)$ de todos os componentes de conhecimento dos objetivos do exercício em andamento.
52	Média das probabilidades de aplicar corretamente todos os componentes de conhecimento primário do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(Correto)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
53	Média das probabilidades de dominar todos os componentes de conhecimento primário do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(L)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
54	Média das probabilidades de ter dominado todos os componentes de conhecimento primário do exercício em andamento, antes da primeira oportunidade de aplicá-los	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(L_0)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
55	Média das probabilidades de aplicar incorretamente todos os componentes de conhecimento primário do exercício em andamento, quando os domina	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(S)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
56	Média das probabilidades de aplicar corretamente todos os componentes de conhecimento primário do exercício em andamento, quando não os domina	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(G)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
57	Média das probabilidades de transitar do estado não aprendido para aprendido em todos os componentes de conhecimento primário do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(T)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
58	Média das probabilidades de esquecimento de cada componente de conhecimento primário do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Média das estimativas $p(F)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.

59	Desvio padrão das probabilidades de aplicar corretamente todos os componentes de conhecimento primário do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(\textit{Correto})$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
60	Desvio padrão das probabilidades de dominar todos os componentes de conhecimento primário do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(L)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
61	Desvio padrão das probabilidades de ter dominado todos os componentes de conhecimento primário do exercício em andamento, antes da primeira oportunidade de aplicá-los	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(L_0)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
62	Desvio padrão das probabilidades de aplicar incorretamente todos os componentes de conhecimento primário do exercício em andamento, quando os domina	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(S)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
63	Desvio padrão das probabilidades de aplicar corretamente todos os componentes de conhecimento primário do exercício em andamento, quando não os domina	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(G)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
64	Desvio padrão das probabilidades de transitar do estado não aprendido para aprendido em todos os componentes de conhecimento primário do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(T)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.
65	Desvio padrão das probabilidades de esquecimento de cada componente de conhecimento primário do exercício em andamento	Conhecimento do aluno	Tese	Gerado a partir do modelo de conhecimento do aluno. Desvio padrão das estimativas $p(F)$ de todos os componentes de conhecimento primário dos objetivos do exercício em andamento.

Fonte: Elaborado pelo autor.

Todos os 65 atributos de dados compõem as amostras dos modelos da Tese. As amostras dos modelos *baseline* não recebem os atributos sobre estimativas de conhecimento do aluno.

No processo de geração dos atributos de dados, diversos outros atributos de dados acabaram não sendo considerados, por apresentarem valores nulos em diversas amostras. Por exemplo, foi gerado um atributo de dados sobre a velocidade média de digitação, que recebia o tempo médio dos intervalos entre teclas. No entanto, para muitas amostras, principalmente nas amostras sobre janelas de tempo de observação menores, como 5 ou 10 segundos, não havia ocorrência de digitação. Para possibilitar que houvesse uma representação de atributo de dado sobre a velocidade de digitação, a estratégia adotada foi gerar três atributos de dados sobre a velocidade de digitação (baixa, média e alta), como descrito no Quadro 8, utilizando uma técnica de clusterização (*k-means*) e *one-hot-encoding*. O mesmo tipo de problema e tratamento foi adotado para possibilitar uma representação de atributo de dado sobre a velocidade de movimentos de mouse.

Figura 23 – Fragmento de registros de amostras

aluno	janela	confuso	nao_confuso	quant_teclas	quant_movimentos_mouse	quant_errores_sintaxe	media_pcorreto_ccs	media_pln_ccs
495	20.000	1	0	14.000	5.000	0.000	0.8201698320	0.7175748379
495	20.000	1	0	0.000	15.000	0.000	0.8201698320	0.7175748379
495	20.000	1	0	0.000	19.000	2.000	0.6786934689	0.5287970711
495	20.000	1	0	0.000	18.000	0.000	0.6786934689	0.5287970711
495	20.000	1	0	0.000	4.000	0.000	0.6378570886	0.4743073987
495	20.000	1	0	0.000	4.000	0.000	0.6378570886	0.4743073987
495	20.000	1	0	0.000	5.000	0.000	0.7073339388	0.5072712658
495	20.000	1	0	0.000	13.000	0.000	0.8418029263	0.8469200283
495	20.000	1	0	0.000	8.000	0.000	0.6860326224	0.6491706140
495	20.000	1	0	0.000	2.000	0.000	0.5584636551	0.5798333194
495	20.000	1	0	12.000	31.000	1.000	0.2007847147	0.1944631470
495	20.000	1	0	0.000	11.000	0.000	0.7425177768	0.7939256421
495	20.000	0	1	11.000	20.000	1.000	0.7806274778	0.8552131155
495	20.000	0	1	32.000	8.000	0.000	0.8981124742	0.8848881743
495	20.000	0	1	2.000	18.000	3.000	0.9139540613	0.9675229613
495	20.000	0	1	11.000	29.000	2.000	0.6793860062	0.6393714846
495	20.000	0	1	22.000	14.000	0.000	0.9139540414	0.9675229369
495	20.000	0	1	12.000	19.000	0.000	0.9139527838	0.9675213919
495	20.000	0	1	50.000	1.000	3.000	0.8981124742	0.8848881743
495	20.000	0	1	46.000	2.000	0.000	0.9239662405	0.9540489024

Fonte: Elaborado pelo autor.

A Figura 23 mostra um fragmento de registros de amostras para uma janela de 20 segundos. Os registros da figura mostram que cada amostra é atrelada a um rótulo de emoção (“confuso” ou “não confuso”). Os rótulos emocionais são atributos de dados categóricos representados nas amostras com a abordagem *one-hot-encoding*, comumente usada para a conversão de atributos de dados categóricos para um formato adequado para modelos de aprendizado de máquina que funcionam com vetores de atributos de dados numéricos

(BURKOV, 2019; SEGER, 2018). Nas amostras, estes atributos (“confuso” e “não\_confuso”) recebem os valores 0 ou 1 para cada um dos possíveis rótulos emocionais. O valor 0 indica a ausência do rótulo e o valor 1 indica a presença do rótulo. Para cada registro de amostra, apenas um destes atributos de dados recebe o valor 1. O outro recebe o valor 0.

A Tabela 22 apresenta a quantidade de amostras geradas para cada janela de dados de tempo fixo e variável, em cada nível de ensino e para cada rótulo de emoção (“confusão” e “não confusão”). A tabela mostra o balanceamento na quantidade de amostras para as classes de interesse (“confusão” e “não confusão”), uma característica importante para o funcionamento dos modelos de aprendizado de máquina da Tese e *baseline*.

Tabela 22 – Quantidade de amostras geradas para janelas de tamanho fixo e variável

Tamanho da janela	Nível de ensino	Quantidade “confusão”	Quantidade “não confusão”	Total Janela X Nível	Total Janela
5	Superior	466	466	932	1926
	Técnico	497	497	994	
10	Superior	453	453	906	1898
	Técnico	496	496	992	
20	Superior	450	450	900	1888
	Técnico	494	494	988	
40	Superior	445	445	890	1866
	Técnico	488	488	976	
60	Superior	440	440	880	1848
	Técnico	484	484	968	
90	Superior	430	430	860	1804
	Técnico	472	472	944	
120	Superior	418	418	836	1766
	Técnico	465	465	930	
180	Superior	392	392	784	1656
	Técnico	436	436	872	
240	Superior	357	357	714	1542
	Técnico	414	414	828	
360	Superior	304	304	608	1378
	Técnico	385	385	770	
Variável	Superior	466	466	932	1926
	Técnico	497	497	994	
<b>Total</b>		<b>9749</b>	<b>9749</b>	<b>19498</b>	

Fonte: Elaborado pelo autor.

## 9 EFEITOS DE ESTIMATIVAS DE CONHECIMENTO DO ALUNO EM DIFERENTES TAMANHOS DE JANELAS DE OBSERVAÇÃO

Este capítulo descreve os métodos utilizados pela Tese para alcançar os objetivos específicos OE1 e OE2, e apresenta os resultados obtidos. Alcançar estes objetivos específicos possibilita encontrar respostas para as subquestões de pesquisa SQP1 e SQP2. Os objetivos OE1 e OE2 são sobre verificar se modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente (abordagem proposta pela Tese) possuem desempenho superior a modelos que utilizam apenas dados sobre interação com o ambiente (abordagem *baseline*, dos trabalhos relacionados). O objetivo é comparar o desempenho dos modelos da Tese e o desempenho dos modelos *baseline* em diferentes configurações de janelas de dados, de tempo fixo e variável. Uma janela de dados é o tempo que o modelo observa o comportamento do aluno no ambiente para tentar inferir sua emoção. A Tese já apresentou uma discussão sobre a abordagem computacional para este tipo de problema: o desenvolvimento de modelos de aprendizado de máquina supervisionada que modelam classes de saída relacionadas aos estados emocionais dos alunos nas amostras, os estados de “confusão” e de “não confusão”. A Tese implementou dois métodos para alcançar os objetivos OE1 e OE2:

**Método 1: geração de modelos de aprendizado de máquina que refletem a abordagem da Tese e a abordagem *baseline*, e a comparação dos seus resultados.**

**Método 2: identificação dos atributos de dados mais relevantes para os modelos.**

A Seção 9.1 apresenta detalhes sobre o primeiro método. O segundo método é detalhado na Seção 9.2. Os resultados são apresentados na Seção 9.3.

### 9.1 Método 1: geração e comparação dos modelos

O **primeiro método** é sobre a geração de modelos de aprendizado de máquina da Tese e *baseline* e a comparação dos seus resultados. Os **modelos da Tese** são modelos de aprendizado de máquina supervisionada treinados com amostras da Tese, compostas por **atributos de dados sobre estimativas de conhecimento do aluno + dados sobre a interação do aluno com o ambiente de programação**. Os modelos da Tese representam a abordagem

apresentada na hipótese da Tese e usam uma abordagem mista dirigida a dados e guiada à teoria. Os **modelos *baseline*** são modelos de aprendizagem de máquina supervisionada treinados com amostras *baseline*, compostas apenas por atributos de dados sobre a interação do aluno com o ambiente de programação. Os modelos *baseline* representam os trabalhos relacionados, que usam uma abordagem dirigida a dados.

A Tese assume que um “modelo de aprendizado de máquina supervisionada” é composto por um algoritmo de aprendizado de máquina supervisionada e seus parâmetros. Os parâmetros de cada modelo são adequados em uma etapa chamada de **treinamento**, e o desempenho do modelo é obtido em uma etapa chamada de **teste**. Os modelos da Tese são treinados e testados com as amostras da Tese, enquanto os modelos *baseline* são treinados com as amostras *baseline*. O Capítulo 8 descreveu a geração das amostras da Tese e *baseline*.

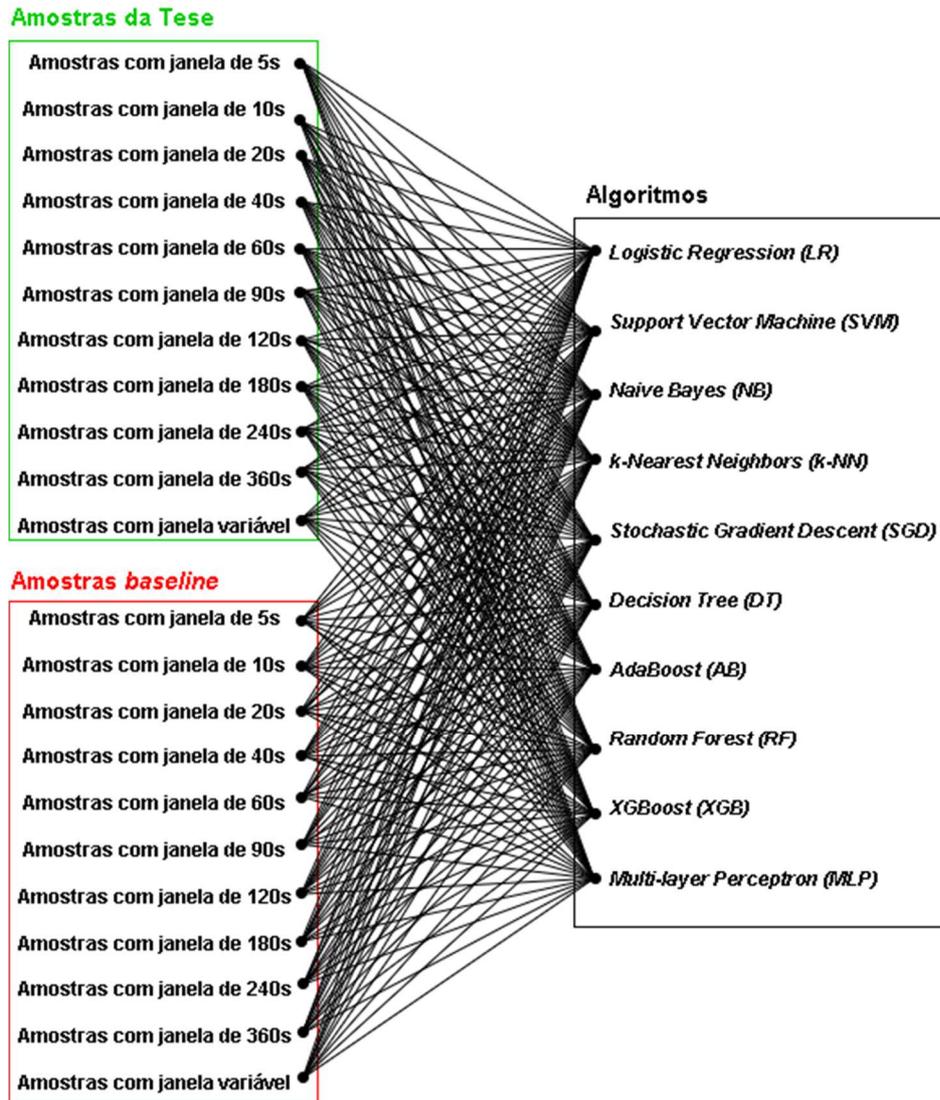
O primeiro método foi organizado em duas etapas: 1) geração e validação dos modelos de aprendizado de máquina supervisionada de classificação da confusão (“confusão” e “não confuso”) para amostras da Tese de diferentes tamanhos de janelas (5, 10, 20, 40, 60, 90, 120, 180, 240, 360 e variável) e para as amostras *baseline* de diferentes tamanhos de janelas (5, 10, 20, 40, 60, 90, 120, 180, 240, 360 e janela variável); e 2) comparação do desempenho dos modelos da Tese e *baseline*, nas diferentes janelas.

A Figura 24 mostra o esquema de geração dos modelos. A Tese utiliza 10 diferentes algoritmos de aprendizado de máquina supervisionada. Cada modelo é gerado para um algoritmo de aprendizado de máquina e um determinado conjunto de amostras. A figura indica que são gerados 110 modelos da Tese e 110 modelos *baseline*. Como será visto mais adiante, essa quantidade de modelos gerada é duplicada, pois cada modelo é gerado duas vezes, uma vez sem utilizar a técnica de seleção de atributos de dados e uma segunda vez utilizando a técnica de seleção de atributos de dados. Assim, a quantidade total de modelos gerados para os objetivos OE1 e OE2 chega a 220 para os modelos da Tese e 220 para os modelos *baseline*.

Uma etapa importante do método é a comparação dos resultados dos modelos da Tese e *baseline*. A única diferença entre os modelos da Tese e os modelos *baseline* são os atributos de dados sobre estimativas de conhecimento do aluno, que estão presentes apenas nos modelos da Tese. Se o desempenho dos modelos da Tese for superior aos modelos *baseline*, a Tese estará apresentando evidências para aceitar a sua hipótese, de que a utilização de dados sobre estimativas de conhecimento do aluno no domínio de programação de computadores, juntamente de dados sobre a interação do estudante com o ambiente computacional de aprendizagem de programação, pode melhorar o desempenho dos modelos livres de sensores na detecção da confusão do aluno nas tarefas de aprendizagem de programação de

computadores, quando comparado a utilizar apenas os dados de interação do estudante com o ambiente de aprendizagem. A comparação dos modelos será realizada para cada configuração de janelas de dados, de forma que será possível verificar como o desempenho dos modelos da Tese e *baseline* difere nas diferentes configurações de janelas de observação.

Figura 24 – Esquema de geração dos modelos da Tese e *baseline*



Fonte: Elaborado pelo autor.

A geração de cada modelo pode ser descrita como uma sequência de componentes de processamento de dados, chamada de *pipeline*. A próxima Seção descreve a *pipeline* utilizada para cada modelo. Nas seções seguintes, são descritos os algoritmos considerados na geração dos modelos, os componentes de processamento do *pipeline*, o método de validação dos

modelos, as métricas de avaliação de desempenho dos modelos, o ambiente de implementação e execução dos modelos e como foi realizada a comparação dos modelos da Tese e *baseline*.

### 9.1.1 Pipeline de geração dos modelos

Um *pipeline* é uma sequência de componentes de processamento de dados, comumente usada em tarefas de aprendizado de máquina (PAPER, 2021). A Figura 25 exibe o *pipeline* definido para a geração e validação de um modelo. A entrada do *pipeline* é composta por **amostras de dados não normalizadas** e **um algoritmo não otimizado**. Uma amostra não normalizada é uma amostra que não passou por nenhum tipo de ajuste de seus valores, como um alinhamento dos valores para uma distribuição normal. Um algoritmo de aprendizado de máquina pode se comportar mal se, individualmente, os atributos de dados das amostras não se parecerem com uma distribuição normal (KUBAT, 2017). Um algoritmo não otimizado, por sua vez, é um algoritmo que não teve seus hiperparâmetros otimizados.

Figura 25 – Pipeline de geração e validação de um modelo de aprendizado de máquina



Fonte: Elaborado pelo autor.

Na primeira atividade do *pipeline*, as amostras passam por um processamento de normalização (a Seção 9.1.3 descreve como foi feita a normalização). A segunda atividade do *pipeline* recebe como entrada as amostras normalizadas e o algoritmo não otimizado. Nesta atividade, ocorre um processamento de otimização dos hiperparâmetros do algoritmo (a Seção

9.1.4 descreve como foi feita esta otimização). Na terceira e última atividade do *pipeline* ocorre a validação do modelo, ou seja, o modelo é treinado e testado e seu desempenho é verificado. Esta etapa recebe como entrada as amostras normalizadas e o algoritmo otimizado, e tem como saída o modelo validado. A Seção 9.1.7 mostra como foi feita a validação dos modelos e a Seção 9.1.5 descreve as métricas de desempenho consideradas.

Figura 26 – Pseudocódigo de geração e validação dos modelos

```

1. abordagens <- [tese, baseline]
2. janelas <- [5,10,20,40,60,90,120,180,240,360,variável]
3. algoritmos <- [LR,SVM,NB,kNN,SGD,DT,AB,RF,XGB,MLP]
4.
5. para cada abordagem em abordagens faça
6.   para cada janela em janelas faça
7.     amostras <- obterAmostras(abordagem, janela)
8.     amostras <- normaliza(amostras)
9.     para cada algoritmo em algoritmos faça
10.      hiperParams <- obterMelhoresHiperParams(algoritmo, amostras)
11.      resultados[abordagem, janela, algoritmo] <- valida(algoritmo, hiperParams, amostras)
12.      melhorModelo[abordagem, janela] <- obterMelhorModelo(resultados[abordagem, janela])

```

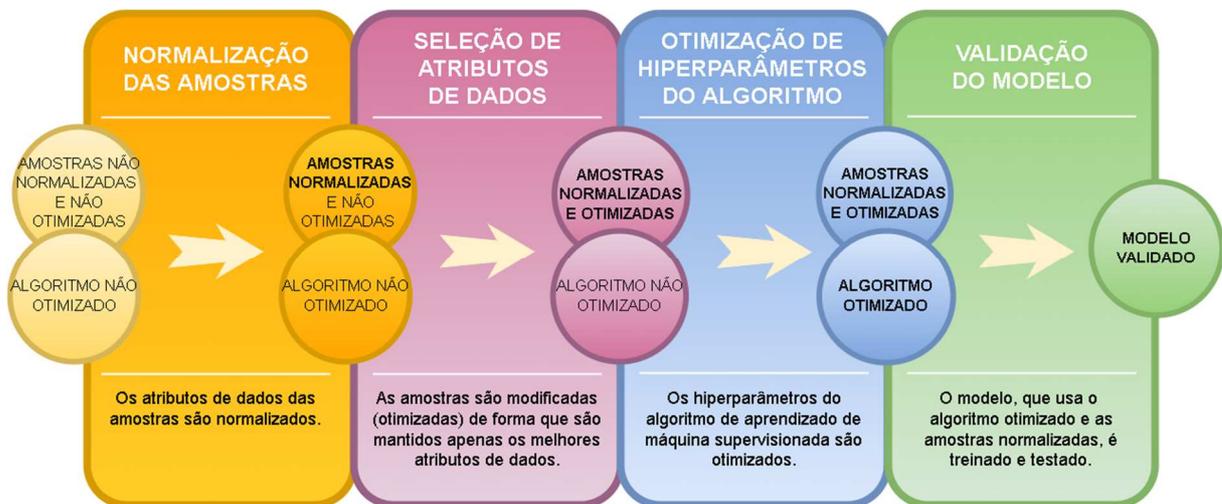
Fonte: Elaborado pelo autor.

A Figura 26 apresenta um pseudocódigo que descreve o fluxo de geração dos modelos. O pseudocódigo explicita o funcionamento da *pipeline* para cada modelo. Primeiramente, nas linhas 1, 2 e 3, são carregados em memória as abordagens das amostras (Tese e *baseline*), as janelas (5, 10, 20, 40, 60, 90, 120, 180, 240, 360 e variável), e os algoritmos de aprendizado de máquina. Para cada janela de cada abordagem (linhas 5 e 6), são executadas as operações descritas a seguir. São carregadas em memória as amostras relativas à abordagem e à janela (linha 7). Na linha 8, as amostras carregadas são atualizadas por uma operação de normalização. Para cada um dos algoritmos (linha 9), são executadas as seguintes operações. São obtidos os melhores hiperparâmetros para o algoritmo e as amostras carregadas (linha 10). Na linha 11, é realizada a validação do modelo composto pelo algoritmo, seus hiperparâmetros e as amostras carregadas. Os resultados de desempenho na validação são guardados em memória (linha 11). A iteração mais interna (linha 9) é executada 220 vezes, uma para cada modelo gerado. Além disso, para cada janela, em cada abordagem, é guardado em memória o modelo que obteve o melhor desempenho na métrica de acurácia (linha 12).

A Tese também realizou uma segunda rodada de geração e validação de todos os modelos, porém acrescentando ao *pipeline* de cada modelo uma atividade de **seleção dos melhores atributos de dados**, do inglês *feature selection*. Esta etapa é detalhada na Seção

9.1.6. O método de seleção de atributos de dados tem sido utilizado na literatura para reduzir a quantidade de atributos de dados das amostras, reduzir o custo computacional e melhorar o desempenho preditivo dos modelos (RASCHKA, 2018). O objetivo de executar novamente a geração dos modelos, acrescentando a atividade de seleção de atributos, foi verificar se esta abordagem proporciona uma melhora no desempenho preditivo dos modelos, ao identificar e remover atributos de dados que poderiam atrapalhar os modelos. Também tem interesse em verificar se, mesmo removendo atributos de dados que não contribuem para os modelos, os modelos da Tese apresentam desempenho superior aos modelos *baseline*. A Figura 27 mostra o mesmo *pipeline* apresentado na Figura 25, porém acrescenta ao *pipeline* uma atividade de seleção dos melhores atributos de dados, entre as atividades de normalização e otimização de hiperparâmetros. Esta atividade recebe de entrada um conjunto de amostras normalizadas, mas não otimizadas, e um algoritmo não otimizado. Por amostra não otimizada, entende-se uma amostra não formada por atributos de dados selecionados. A saída desta atividade é um conjunto de amostras normalizadas e otimizadas, juntamente do algoritmo não otimizado.

Figura 27 – *Pipeline* de geração e validação de um modelo de aprendizado de máquina com seleção de atributos



Fonte: Elaborado pelo autor.

A Figura 28 apresenta um pseudocódigo que descreve o fluxo de geração dos modelos, e inclui a atividade de seleção de atributos de dados. Primeiramente, nas linhas 1, 2 e 3, são carregados em memória as abordagens das amostras (Tese e *baseline*), as janelas (5, 10, 20, 40, 60, 90, 120, 180, 240, 360 e variável), e os algoritmos de aprendizado de máquina. Para cada janela de cada abordagem (linhas 5 e 6), são executadas as operações descritas a seguir. São

carregadas em memória as amostras relativas à abordagem e à janela (linha 7). Na linha 8, as amostras carregadas são atualizadas por uma operação de normalização.

Figura 28 – Pseudocódigo de geração e validação dos modelos com seleção de atributos

```

1. abordagens <- [tese, baseline]
2. janelas <- [5,10,20,40,60,90,120,180,240,360,variável]
3. algoritmos <- [LR,SVM,NE,kNN,SGD,DT,AB,RF,XGB,MLP]
4.
5. para cada abordagem em abordagens faça
6.   para cada janela em janelas faça
7.     amostras <- obterAmostras(abordagem, janela)
8.     amostras <- normaliza(amostras)
9.     melhorModelo <- obterMelhorModelo(abordagem, janela)
10.    melhoresAtributos <- obterMelhoresAtributos(melhorModelo, amostras)
11.    amostras <- filtraMelhoresAtributos(melhoresAtributos, amostras)
12.    para cada algoritmo em algoritmos faça
13.      hiperParams <- obterMelhoresHiperParams(algoritmo, amostras)
14.      resultados[abordagem, janela, algoritmo] <- valida(algoritmo, hiperParams, amostras)

```

Fonte: Elaborado pelo autor.

A partir da linha 9, o fluxo se difere do pseudocódigo da Figura 26. Nesta linha, o melhor modelo obtido para a janela e para a abordagem, na execução do pseudocódigo da Figura 26, é carregado em memória (linha 9). Isso significa que a geração e validação dos modelos sem seleção de atributos de dados é executada antes da geração e validação dos modelos com seleção de atributos de dados. Na linha 10 ocorre a obtenção dos melhores atributos de dados. Para isso, é utilizado o melhor modelo treinado para a abordagem e janela, e as amostras carregadas. Após, as amostras são atualizadas com uma operação de filtro (linha 11), na qual são mantidos apenas os atributos de dados selecionados. Na sequência do fluxo, na linha 12, para cada algoritmo, são realizadas as operações que se seguem. São obtidos os melhores hiperparâmetros para o algoritmo e as amostras (linha 13). Por fim, ocorre a operação de validação para o algoritmo com os hiperparâmetros e as amostras. Os resultados de desempenho são guardados em memória (linha 14).

Os métodos utilizados nos *pipelines* e os processos descritos nos pseudocódigos seguem boas práticas descritas na literatura (KUBAT, 2017; RASCHKA, 2018) para o desenvolvimento e validação de modelos de aprendizado de máquina supervisionada. As próximas Seções detalham as operações apresentadas nestes processos.

### 9.1.2 Algoritmos

Um algoritmo de aprendizagem de máquina é um conjunto de instruções que tentam modelar uma função alvo com base em uma amostra de entrada. No contexto da Tese, a função alvo possui como saída um de dois valores categóricos possíveis (“confusão” e “não confusão”) para cada entrada, ou seja, os algoritmos usados na Tese realizam uma tarefa de classificação binária. O tipo de problema que a Tese tenta resolver envolve a aplicação de algoritmos de aprendizagem de máquina do tipo supervisionada, no qual buscam aprender uma relação entre as entradas e as saídas através de amostras rotuladas de dados, ou seja, amostras vinculadas aos rótulos *ground-truth* de “confusão” e “não confusão”, obtidos via autorrelato dos alunos.

A Tese selecionou diferentes tipos de algoritmos de aprendizado de máquina supervisionada para serem utilizados nos seus modelos. Foram selecionados algoritmos utilizados nos trabalhos relacionados (ver Quadro 2 da Seção 3.4.2), como *Support Vector Machine* (SVM), *Naive Bayes* (NB), *Multi-layer Perceptron* (MLP), *Decision Tree* (DT) e *Random Forest* (RF). Também foram adicionados outros algoritmos tradicionais de aprendizado de máquina (KUBAT, 2017), como *Logistic Regression* (LR), *k-Nearest Neighbors* (k-NN), *Stochastic Gradient Descent* (SGD) e *AdaBoost* (AB). O algoritmo *XGBoost* (XGB) foi selecionado por usar conjuntos de árvores, recomendado para dados tabulares (SHWARTZ-ZIV; ARMON, 2022), como os usados nesta pesquisa.

### 9.1.3 Normalização das amostras

A normalização de amostras é o processo de ajustar os valores individuais dos atributos de dados das amostras, de forma a favorecer o comportamento dos algoritmos de aprendizado de máquina e melhorar o desempenho deles (KUBAT, 2017). Nos *pipelines* apresentados (Figura 25 e Figura 27), a etapa de normalização consistiu em redimensionar os valores dos atributos de dados, individualmente, de forma que eles passassem a ter a propriedade de uma distribuição normal padrão com  $\mu$  (valor médio do atributo de dados sobre todas as amostras) = 0 e  $\sigma$  (desvio padrão) = 1.

### 9.1.4 Otimização de hiperparâmetros

Os **hiperparâmetros dos modelos** de aprendizado de máquina são todos os parâmetros que são ajustados antes do treinamento, e diferem dos **parâmetros dos modelos** que são

ajustados durante o treinamento (RASCHKA, 2018). O ajuste destes hiperparâmetros controla o comportamento dos algoritmos e o melhor ajuste busca otimizar o desempenho. O ajuste dos hiperparâmetros é chamado de otimização de hiperparâmetros. A otimização de hiperparâmetros está presente em todos os *pipelines* apresentados (Figura 25 e Figura 27). Todos os algoritmos de aprendizagem de máquina considerados na Tese possuem um conjunto de hiperparâmetros candidatos. Um subconjunto destes hiperparâmetros, que leva o algoritmo a seu melhor desempenho, precisa ser selecionado antes do algoritmo ser utilizado no treinamento.

Quadro 9 – Valores candidatos para os hiperparâmetros do algoritmo *k-Nearest Neighbor*

Hiperparâmetro	Valores candidatos
n_neighbors	3, 4, 5, 6, 7, 8, 9, 10
weights	'uniform', 'distance'
algorithm	'ball_tree', 'kd_tree', 'brute'
leaf_size	1, 5, 10, 20, 30
p	1, 2

Fonte: Elaborado pelo autor.

O método utilizado pela Tese para selecionar o subconjunto de hiperparâmetros foi o de busca exaustiva (*grid search*), com validação cruzada *k-fold* ( $k = 10$ ) com *folds* estratificados. Neste método, o pesquisador define previamente, de forma arbitrária, conjuntos de valores candidatos para os diferentes hiperparâmetros de cada algoritmo. Na busca exaustiva, todas as combinações possíveis de valores dos hiperparâmetros são usadas no treinamento e validação de determinado algoritmo. O método de validação utilizado foi validação cruzada *k-fold* ( $k = 10$ ) com *folds* estratificados (a Seção 9.1.7 explica este método de validação cruzada). O Quadro 9 mostra os valores candidatos para os hiperparâmetros do algoritmo *k-Nearest Neighbor*. Os diferentes valores desses hiperparâmetros possibilitam 480 combinações possíveis ( $8 \times 2 \times 3 \times 5 \times 2$ ) de hiperparâmetros a serem testados. Considerando que foi aplicada uma validação cruzada *k-fold* de  $k = 10$ , a quantidade total de modelos gerados (treinados e validados) na aplicação da busca exaustiva foi de 4800 (480 combinações  $\times$  10) para o algoritmo *k-Nearest Neighbor*. O conjunto de hiperparâmetros do algoritmo que apresentar o melhor desempenho na métrica de acurácia, é selecionado. O Apêndice O apresenta todos os hiperparâmetros candidatos, escolhidos arbitrariamente pelo pesquisador, para cada algoritmo.

### 9.1.5 Métricas de avaliação

Os trabalhos relacionados utilizaram as seguintes métricas nas avaliações de desempenho dos seus modelos: *Kappa* de Cohen (BOSCH; CHEN; D’MELLO, 2014; GRAFSGAARD; BOYER; LESTER, 2011; LEE et al., 2011; TIAM-LEE; SUMI, 2018, 2019; VEA; RODRIGO, 2017c); acurácia (GRAFSGAARD; BOYER; LESTER, 2011; TIAM-LEE; SUMI, 2018, 2019; VEA; RODRIGO, 2017c); *TP Rate*, *FP Rate*, precisão, *recall*, *F-measure*, área *ROC* e matriz de confusão (FELIPE et al., 2012). Nas avaliações dos modelos da Tese e *baseline*, além das duas métricas mais utilizadas nos trabalhos relacionados (acurácia e *Kappa* de *Cohen*), também foram gerados valores de desempenho para as métricas AUC *ROC*, precisão (do inglês, *precision*), *recall* e F1.

Quando uma amostra de exemplo é testada contra um modelo de aprendizado de máquina treinado, obtém-se um de quatro resultados possíveis: 1) o exemplo é positivo (por exemplo, relacionado à confusão) e o classificador reconhece como exemplo positivo (VP = verdadeiro positivo); 2) o exemplo é negativo (por exemplo, relacionado à não confusão) e o classificador reconhece como exemplo negativo (VN = verdadeiro negativo); 3) o exemplo é positivo, mas o classificador reconhece como exemplo negativo (FN = falso negativo); 4) o exemplo é negativo, mas o classificador reconhece como exemplo positivo (FP = falso positivo). Quando o modelo de aprendizado de máquina é aplicado para um conjunto de amostras de exemplo, todos os quatro resultados possíveis ocorrerão em diferentes quantidades, de forma que diferentes critérios (métricas) de desempenho podem ser definidos, como a acurácia, precisão, *recall* e F1. Cada métrica apresenta um tipo diferente de informação sobre os resultados do modelo. Os próximos parágrafos descrevem cada uma das métricas utilizadas na Tese e como elas podem ser interpretadas.

A métrica de **acurácia** mede o desempenho geral do modelo, ou seja, o percentual de classificações corretas em relação à quantidade total de amostras de exemplo apresentadas ao modelo de aprendizado de máquina. A acurácia é calculada através da divisão da quantidade de classificações corretas pelo total de amostras de exemplo de entrada:  $(VP + VN) / (VP + VN + FN + FP)$ . Quando há um desequilíbrio na quantidade de amostras de exemplo para cada classe de interesse (no caso da Tese, as classes “confusão” e “não confusão”), a acurácia não apresenta uma boa resposta sobre o modelo. Nestes casos, métricas como precisão e *recall* são mais interessantes para explicar a utilidade dos modelos (KUBAT, 2017). No entanto, como foi possível gerar amostras balanceadas para o trabalho da Tese (ver Seção 8.3), a acurácia é uma métrica de desempenho geral interessante para a avaliação de desempenho dos modelos.

A métrica **AUC ROC** (do inglês *Area under curve of receiver operating characteristic*) mede o desempenho do modelo em classificar corretamente tanto as classes positivas (no caso da Tese, a classe “confusão”) nas amostras de exemplos positivos, quanto as classes negativas (no caso da Tese, a classe “não confusão”) nas amostras de exemplos negativos. O valor de AUC ROC é um número real entre 0,0 e 1,0. Quanto maior o valor (próximo de 1,0), melhor o desempenho do modelo em distinguir as classes positivas e negativas. Um valor de AUC ROC igual a 1,0 indica que o modelo é capaz de distinguir perfeitamente entre as classes positivas e negativas. Um valor de AUC ROC igual a 0,0 indica que o classificador estaria prevendo todos os exemplos positivos como negativos e todos os exemplos negativos como positivos. A métrica é utilizada em problemas de classificação binária, como é o caso da Tese, mas pode ser estendido para problemas com mais de duas classes (BURKOV, 2019).

A métrica *Kappa de Cohen* foi originalmente usada para medir a concordância entre dois juízes humanos. No entanto, ela tem sido utilizada nas avaliações de modelos de aprendizado de máquina, como nos trabalhos relacionados, para medir a concordância entre as saídas dos modelos e as classes de interesse. O *Kappa* é uma medida baseada na acurácia (mede o percentual de classificações corretas sobre o total de amostras), mas considera a possibilidade de a concordância ter ocorrido por acaso para fazer um ajuste na medida (COHEN, 1960). Na Tese, os dois juízes são os seguintes: o modelo de classificação; e os alunos participantes, que relataram sua confusão durante as tarefas de programação. Da mesma forma como a acurácia, o *Kappa* pode apresentar problemas para explicar os resultados dos modelos quando há um desequilíbrio de amostras para cada classe de interesse (DELGADO; TIBAU, 2019), o que não é um problema na Tese. O valor de *Kappa* é um valor real entre -1,0 e 1,0, sendo que valores iguais a 1,0 indicam a concordância perfeita entre os juízes. Valores iguais a 0,0 (zero) indicam que a concordância é a mesma que a esperada por acaso. Valores iguais a -1,0 indicam nenhuma concordância (COHEN, 1960). Em estatística, a faixa de valores de *Kappa* entre 0,4 e 0,6 é considerada uma concordância justa (apertada) e moderada, enquanto a faixa de valores entre 0,6 e 0,75 é considerada boa. A faixa com valores maiores que 0,75 é considerada excelente (KNOTTNERUS; TUGWELL, 2010). Para o melhor modelo da Tese, espera-se obter um valor de concordância superior a 0,6, considerado um bom objetivo em detectores computacionais de emoções, uma vez que esse é um valor de concordância aceito para juízes humanos (D’MELLO; GRAESSER, 2009; OCUMPAUGH; BAKER; OCUMPAUGH, J., & BAKER, 2014; STEIDL et al., 2005).

A **precisão** mede o percentual de verdadeiros positivos (VP), entre todos os exemplos que o classificador rotulou como positivo. É calculado através da divisão da quantidade de

verdadeiros positivos, pela quantidade de exemplos rotulados como positivos pelo classificador:  $VP / (VP + FP)$ . A métrica de precisão mede a probabilidade de o modelo estar correto quando classifica um exemplo como positivo (KUBAT, 2017). No contexto do problema enfrentado pela Tese, esta métrica mede a precisão dos modelos quando indica que o aluno está confuso, ou seja, o percentual de acerto dos modelos quando indicam que o aluno está confuso. Se num contexto educacional for um problema o modelo indicar que o aluno está confuso quando não está, a precisão é uma métrica que deve ser levada em consideração.

O *recall* mede o percentual de exemplos positivos que são corretamente identificados como positivos pelo classificador. É calculado através da divisão da quantidade de verdadeiros positivos, pela quantidade de amostras de exemplos positivos apresentados ao modelo:  $VP / (VP + FN)$  (KUBAT, 2017). Se num contexto educacional for fundamental o modelo identificar a confusão quando o aluno está confuso, o *recall* é uma métrica que deve ser levada em consideração.

A métrica **F1** mede uma média harmônica de precisão e *recall*. É útil quando não se sabe ao certo qual das duas métricas é mais importante no contexto. Um valor de F1 baixo indica que, ou a precisão, ou o recall, está baixo. A fórmula para o cálculo de F1 é o seguinte:  $(2 * \text{precisão} * \text{recall}) / (\text{precisão} + \text{recall})$  (KUBAT, 2017).

#### 9.1.6 Seleção dos melhores atributos de dados

Métodos de seleção de atributos de dados buscam selecionar um subconjunto de atributos de dados que melhor contribuem para um modelo de aprendizado de máquina modelar sua função de saída. O objetivo é remover atributos de dados irrelevantes ou redundantes, que reduzem o desempenho dos classificadores (DASH; LIU, 1997). Os trabalhos relacionados que indicaram usar algum método de seleção de atributos de dados (BOSCH; CHEN; D'MELLO, 2014; TIAM-LEE; SUMI, 2019; VEA; RODRIGO, 2017b) usaram o método *RELIEF-F*. O trabalho de Vea e Rodrigo (2017) removeu manualmente alguns atributos de dados que eram fortemente correlacionados com outros atributos de dados.

Para a seleção de atributos de dados, uma atividade do *pipeline* apresentado na Figura 27, a Tese utilizou o método *Recursive Feature Elimination* (RFE) (LE THI; NGUYEN; OUCHANI, 2008), também utilizado no trabalho sobre detecção de confusão baseado em sensores no contexto de MOOCs (*Massive Open Online Course*), de Erwianda et al. (2019).

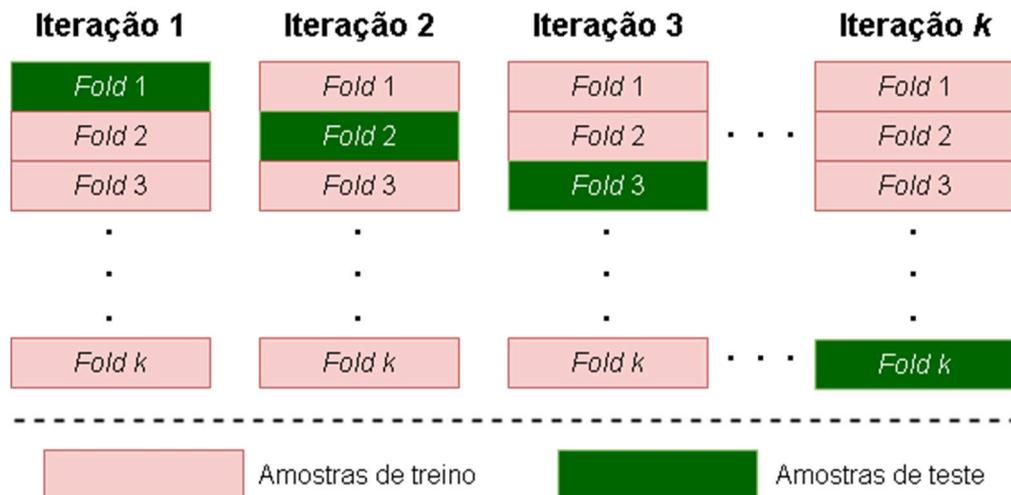
O RFE é um método iterativo que utiliza um modelo classificador. A Tese utiliza o melhor modelo classificador de confusão da respectiva janela. Antes de iniciar o processo

iterativo, o modelo classificador é treinado e validado (validação cruzada *k-fold* com  $k = 10$  e seleção estratificada das amostras dos *folds*) sobre amostras contendo todos os atributos de dados. Após, é iniciado o processo iterativo. A cada iteração, o modelo RFE gera pesos de importância para cada atributo de dados (utiliza algum critério específico do algoritmo utilizado) e ranqueia os atributos, dos mais importantes aos menos importantes. Ao final de cada iteração, o método elimina das amostras os atributos de dados menos importantes. A quantidade de atributos de dados que são eliminados a cada iteração é definida por um valor de *step*, definido previamente. Numa próxima iteração, o modelo é novamente treinado e validado, um novo ranqueamento é gerado, e a quantidade de atributos definida pelo *step* com pior ranqueamento é eliminada. Estas etapas são executadas recursivamente, removendo sempre uma quantidade de atributos de dados menos importantes, até que o modelo seja treinado e validado com a quantidade mínima de atributos de dados definido para a execução do método (LE THI; NGUYEN; OUCHANI, 2008). A Tese usou o valor 1 como quantidade mínima de atributos de dados. O valor de *step* utilizado também foi 1, de forma que a cada recursão, foi eliminado das amostras um atributo por vez. No final do processo, o RFE seleciona o conjunto de atributos de dados que obteve o melhor desempenho na métrica de acurácia.

#### 9.1.7 Validação dos modelos

O objetivo de validar um modelo de aprendizado de máquina é verificar se, após ser treinado com determinado conjunto de amostras (amostras de treino), possui bom desempenho de predição em um outro conjunto de amostras (amostras de teste) que não foi usado no treinamento. Em outras palavras, o objetivo de validar um modelo de aprendizado de máquina é verificar se ele generaliza bem para dados não vistos no treinamento (RASCHKA, 2018).

Figura 29 – Divisão dos *folds* no método de validação cruzada *k-fold*



Fonte: Elaborado pelo autor.

Os *pipelines* apresentados na Figura 25 e na Figura 27, incluem a atividade de validação de modelo. A Tese utilizou o método de validação cruzada  $k$ -fold com  $k = 10$  e com amostras estratificadas nos *folds*. O método divide aleatoriamente as amostras de dados em  $k$  conjuntos (*folds*) com tamanho igual de amostras em cada conjunto. Então são realizadas  $k$  iterações de treinamento e teste. Em cada iteração, são utilizados  $k - 1$  (9) conjuntos de amostras no treinamento do modelo, enquanto o conjunto restante de amostras é usado para o teste do modelo. A Figura 29 ilustra a divisão dos *folds* em cada iteração. Ao final das iterações, é calculada a média dos desempenhos (acurácia) nos testes, nas  $k$  iterações (WONG, 2015). Esta média indica o desempenho do modelo validado.

A Tese utilizou o método de estratificação na etapa em que os *folds* são formados. A estratificação garante que, na distribuição aleatória das amostras para os *folds*, cada *fold* contenha um conjunto balanceado de amostras para cada classe de saída (RASCHKA, 2018). No contexto da Tese, a estratificação garante que cada *fold* seja formado por um conjunto balanceado de amostras vinculadas aos rótulos de “confusão” e “não confusão”.

Os trabalhos relacionados utilizaram os seguintes métodos para validação de modelos de classificação: a) *leave-one-out cross fold validation (LOOCV)* (TIAM-LEE; SUMI, 2018); b) *Leave-p-out student-level cross-validation (LpOCV)* (BOSCH; CHEN; D’MELLO, 2014); *10-fold cross-validation* (TIAM-LEE; SUMI, 2019); *student-level 10-fold cross-validation* (LEE et al., 2011); *student-level cross-validation* (VEA; RODRIGO, 2017c); e 70% de treino e 30% de teste (FELIPE et al., 2012). Os trabalhos relacionados não citam o uso do método de estratificação.

Os métodos *LOOCV* e *LpOCV* utilizados por alguns trabalhos relacionados são métodos robustos de validação cruzada, mas possuem custo computacional elevado. No *LOOCV*, ocorre que  $k = n$  e  $n$  é igual a quantidade de amostras, de forma que a quantidade de iterações é a quantidade de amostras. No *LpOCV*, ocorre que  $k = a$  e  $a$  é igual a quantidade de estudantes, de forma que a quantidade de iterações é a quantidade de estudantes. Foi avaliado como inviável a aplicação destes métodos na Tese, considerando a quantidade de amostras da Tese e a quantidade de diferentes conjuntos de amostras usada na Tese (ver Tabela 22).

Outros trabalhos relacionados utilizaram validação cruzada *k-fold* com  $k = 10$ , à nível de estudante. Este método é robusto para demonstrar a generalização dos modelos para diferentes estudantes. Neste método, quando as amostras de determinado estudante são usadas nos *folds* de treino, nenhuma amostra do mesmo estudante é usada no *fold* de teste. Este método não foi utilizado na Tese, em função de ter sido identificado um desbalanceamento na quantidade de amostras de cada estudante, não permitindo o balanceamento adequado das classes de saída nos *k folds*.

O método de seleção de 70% de amostras para treino e 30% de amostras para teste, usado em um dos trabalhos relacionado, não foi utilizado na Tese, por ser considerado um método menos robusto para validação de modelos (RASCHKA, 2018).

O método usado na Tese foi o de validação cruzada *k-fold* com  $k = 10$  e amostras estratificadas. Este método faz um bom balanço de robustez e custo computacional e é bem aceito para a validação de modelos de aprendizado de máquina (RASCHKA, 2018).

#### 9.1.8 Comparação da acurácia preditiva dos modelos da Tese e *baseline*

Foram aplicados testes de hipótese estatística para comparar a acurácia preditiva dos modelos da Tese e *baseline* em cada configuração de janela de observação de dados. O teste estatístico utilizado foi o teste McNemar (1947), recomendado quando dois modelos de aprendizado de máquina são comparados (DIETTERICH, 1997; RASCHKA, 2018). O teste McNemar é um teste estatístico não paramétrico, também referenciado na literatura como um “teste chi-quadrado inter-sujeitos”. A hipótese nula é que a acurácia preditiva dos dois modelos comparados é igual, e a hipótese alternativa é que a acurácia preditiva dos dois modelos não é igual. Considerando um nível de significância estatística de 0,05 ( $\alpha = 0,05$ ), caso seja obtido um valor de  $p$  (*p-value*) inferior a  $\alpha$ , é possível rejeitar a hipótese nula e aceitar a hipótese alternativa (RASCHKA, 2018).

Figura 30 – Tabela de contingência 2 x 2 para dois cenários de exemplo

		Cenário 1		Cenário 2	
		Modelo 2 correto	Modelo 2 errado	Modelo 2 correto	Modelo 2 errado
Modelo 1	correto	A 9959	B 11	A 9945	B 25
	errado	C 1	D 29	C 15	D 15

Fonte: Adaptado de Raschka (2018).

A Figura 30 apresenta dois cenários de exemplo que ajudam a descrever o funcionamento do teste *McNemar*. Cada cenário possui uma tabela de contingência 2 X 2. As células A e D mostram quantas previsões foram, respectivamente, corretas e erradas em ambos os modelos de comparação do cenário, ou seja, em quantas amostras ambos os modelos tomaram as mesmas decisões. No entanto, as células B e C são mais interessantes para o teste *McNemar*, pois elas indicam quantas decisões dos modelos foram divergentes. No cenário 1, a acurácia do modelo 1 é de 0,997  $((9959 + 11) / (9959 + 11 + 1 + 29))$  e do modelo 2 de 0,996  $((9959 + 1) / (9959 + 11 + 1 + 29))$ . Apesar da diferença de acurácia de apenas 0,001 entre os modelos, o modelo 1 acertou 11 previsões que o modelo 2 errou, o que poderia dar uma intuição de que o modelo 1 obteve um desempenho superior ao modelo 2. No cenário 2, as acurácias dos modelos são as mesmas dos modelos do cenário 1, com o modelo 1 apresentando acurácia de 0,997  $((9945 + 25) / (9945 + 25 + 15 + 15))$  e o modelo 2 apresentando acurácia de 0,996  $((9945 + 15) / (9945 + 25 + 15 + 15))$ . No entanto, visualizando as células B e C, que representam a quantidade de amostras onde os modelos divergiram, não fica tão clara a diferença dos modelos 1 e 2. Apesar de a diferença de acurácia dos modelos nos cenários 1 e 2 seja a mesma (0,001), a valor de  $p$  para a comparação no cenário 1 é de 0,0039 e no cenário 2 de 0,1138. Com estes valores de  $p$ , aceitamos a hipótese alternativa no cenário 1, mas rejeitamos no cenário 2 (RASCHKA, 2018).

O modelo *McNemar* testa a diferença dos modelos em relação às decisões divergentes. Descrevendo de outra forma, a hipótese nula do teste *McNemar* é que as probabilidades  $p(B)$  e  $p(C)$  sejam as mesmas em ambos os modelos (RASCHKA, 2018).

### 9.1.9 Ambiente de implementação e execução dos modelos

Todos os modelos de aprendizado de máquina descritos na Tese foram implementados na linguagem de programação *Python*<sup>1</sup>, em ambientes *notebooks* do *Google Colaboratory*<sup>2</sup>. *Notebooks* são ambientes de edição de código que permitem combinar em um só documento códigos executáveis na linguagem *Python*, imagens e outras linguagens, como *HTML* e *LaTeX*. Nos *notebooks*, os códigos podem ser fragmentados em diferentes células de códigos, possibilitando que sejam executados de forma fragmentada em cada célula, e compartilhando o estado da memória em todo o documento *notebook*. O *Google Colaboratory* é um produto do *Google*, que utiliza *notebooks* para possibilitar ao usuário escrever códigos *Python* em um navegador *web* e guardar na “nuvem” o estado do documento. O *Google Colaboratory* também disponibiliza bibliotecas *Python* para tarefas de ciência de dados e aprendizado de máquina, além de recursos de *hardware* para a execução dos códigos, como GPUs (unidades de processamento gráfico), TPUs (unidades de processamento Tensor) e memória RAM. No entanto, a Tese não utilizou os recursos computacionais do *Google Colaboratory*. Todos os modelos foram executados em computadores pessoais, conectados aos *notebooks*. Os modelos foram executados no ambiente *Anaconda*<sup>3</sup>, uma distribuição *open-source* da linguagem de programação *Python*, que acompanha pacotes de implementações de algoritmos de aprendizado de máquina. Foram utilizadas as seguintes bibliotecas *Python*, disponibilizadas pelo *Anaconda*: *scikit-learn*<sup>4</sup>, que disponibilizou a maior parte dos algoritmos de aprendizado de máquina, além de funções para obtenção de métricas, normalização de dados, estratificação de amostras, seleção de atributos de dados, seleção de hiperparâmetros e validação cruzada; *NumPy*<sup>5</sup> e *pandas*<sup>6</sup>, na manipulação e análises de dados; *matplotlib*<sup>7</sup>, na geração de gráficos; e *XGBoost*<sup>8</sup>, na utilização do algoritmo *XGBoost*, não incluído no *scikit-learn*.

## 9.2 Método 2: identificação dos atributos de dados mais relevantes

O **segundo método** utilizado para os objetivos OE1 e OE2 busca identificar os atributos de dados mais relevantes para os modelos. O objetivo é verificar se os atributos de dados sobre

---

<sup>1</sup> Disponível em <https://www.python.org/>.

<sup>2</sup> Disponível em <https://colab.research.google.com/>

<sup>3</sup> Disponível em <https://www.anaconda.com/>.

<sup>4</sup> Disponível em <https://scikit-learn.org/>.

<sup>5</sup> Disponível em <https://numpy.org/>

<sup>6</sup> Disponível em <https://pandas.pydata.org/>

<sup>7</sup> Disponível em <https://matplotlib.org/>

<sup>8</sup> Disponível em <https://xgboost.ai/>

estimativas de conhecimento do aluno aparecem entre os mais relevantes nos melhores modelos da Tese, em cada uma das configurações de janelas de observação de dados. Também foram verificados os atributos de dados mais relevantes nos melhores modelos *baseline*.

O método de verificação dos atributos mais relevantes foi dividido em duas etapas: 1) ranqueamento dos atributos de dados que mais foram selecionados na etapa de seleção de atributos de dados (método RFE) dos modelos da Tese e *baseline*, nas diferentes configurações de janelas; 2) aplicação do método de importância de permutação de atributos de dados (do inglês *permutation feature importance*) nos melhores modelos da Tese e *baseline* de cada configuração de janelas de dados, e ranqueamento dos dez atributos mais relevantes em cada configuração de janelas.

No método de **importância de permutação de atributos de dados**, a relevância de um atributo de dado para um modelo particular é definida como a diminuição no score de desempenho do modelo quando os valores do atributo são embaralhados (do inglês *shuffle*) aleatoriamente. Quanto maior for a diminuição do score de desempenho do modelo particular, quando os valores de determinado atributo de dado são embaralhados aleatoriamente, maior é a relevância do atributo para o modelo (BREIMAN, 2001; MOLNAR et al., 2021). Na aplicação do método pela Tese, cada atributo foi embaralhado 50 vezes em cada modelo. O score de desempenho considerado foi a acurácia. O método foi aplicado apenas nos melhores modelos da Tese e *baseline*, em cada configuração de janela.

### 9.3 Resultados

Esta Seção apresenta os resultados encontrados na aplicação dos métodos descritos nas Seções anteriores, para alcançar os objetivos específicos OE1 e OE2. **Primeiramente são apresentados os resultados para o primeiro método, descrito na Seção 9.1.**

A Tabela 23 apresenta os valores de acurácia, *AUC ROC*, *Kappa*, precisão, *recall* e F1 dos modelos da Tese e dos modelos *baseline* que apresentaram melhor acurácia em cada configuração de janela de observação de tempo fixo. Os resultados de todos os modelos gerados foram disponibilizados no Apêndice A. A Tabela 23 também apresenta os resultados dos testes estatísticos (*McNemar*) que comparam a acurácia preditiva dos modelos da Tese e *baseline*, em cada configuração de janela de tempo fixo. Também são apresentados os algoritmos destes melhores modelos.

Tabela 23 – Desempenho e comparação dos melhores modelos com janelas fixas e sem seleção de atributos de dados

<i>Jan.</i>		<i>Algor.</i>	<i>Acur.</i>	<i>AUC ROC</i>	<i>Kappa</i>	<i>Prec.</i>	<i>Recall</i>	<i>F1</i>	<i>Teste McNemar</i>	
									<i>chi<sup>2</sup></i>	<i>p</i>
5	Tese	XGB	,89197	,94857	,78394	,87273	,92523	,89495	396,19565	<,00001
	Base	XGB	,76480	,81620	,52951	,70976	<b>,93039</b>	,79286		
10	Tese	RF	<b>,89725</b>	<b>,95082</b>	<b>,79452</b>	<b>,87981</b>	,92308	<b>,89992</b>	280,13880	<,00001
	Base	XGB	,76607	,83293	,53217	,73102	,91043	,78888		
20	Tese	XGB	,87710	,93528	,75419	,86296	,91425	,87957	284,08503	<,00001
	Base	XGB	,77965	,84468	,55930	,74315	,89407	,79502		
40	Tese	XGB	,84831	,91918	,69661	,83298	,87788	,85218	42,22340	<,00001
	Base	RF	,78563	,84947	,57123	,75542	,86611	,79769		
60	Tese	XGB	,84303	,91161	,68611	,82421	,87442	,84813	175,00564	<,00001
	Base	RF	,78784	,84893	,57573	,75776	,85393	,80013		
90	Tese	XGB	,82151	,90439	,64303	,80924	,85262	,82502	368,00270	<,00001
	Base	MLP	,77549	,83442	,55102	,74870	,86255	,78885		
120	Tese	RF	,82274	,90144	,64546	,81377	,83801	,82543	31,03030	<,00001
	Base	RF	,76612	,84344	,53219	,74800	,83921	,77915		
180	Tese	XGB	,80798	,88887	,61597	,80004	,82252	,81071	323,14540	<,00001
	Base	MLP	,75962	,82064	,51923	,74447	,80427	,76582		
240	Tese	XGB	,80221	,86712	,60440	,79828	,80930	,80351	402,00247	<,00001
	Base	LG	,74971	,82589	,49941	,74160	,78081	,75753		
360	Tese	RF	,79610	,86917	,59220	,80007	,79386	,79504	287,08417	<,00001
	Base	MLP	,73951	,81153	,47908	,73331	,77510	,74770		

Fonte: Elaborado pelo autor.

A primeira observação é que para todas as métricas de desempenho geral (acurácia, AUC ROC e Kappa), em todas as configurações de janelas fixas, os valores dos melhores modelos da Tese foram superiores aos valores dos melhores modelos *baseline*. Os testes estatísticos (*McNemar*) também mostraram que em todas as janelas de dados, a diferença de acurácia preditiva entre o melhor modelo da Tese e o melhor modelo *baseline* foi significativa, sempre com valor de  $p < 0,00001$ . É possível concluir que a acurácia preditiva dos melhores modelos da Tese foi significativamente superior a acurácia dos melhores modelos *baseline*.

Todos os melhores modelos da Tese apresentaram valores superiores na métrica de precisão, em relação aos melhores modelos *baseline*. No contexto da Tese, a precisão mede o percentual de acerto quando o modelo indica que o aluno está confuso. A precisão de um modelo é particularmente interessante em ambientes educacionais que entregam ações imediatas após a detecção de uma emoção. Se o ambiente age em função de uma detecção de confusão, é importante que esta detecção seja correta.

Aparentemente, a diferença de valores entre os modelos da Tese e *baseline* foi mais apertada na métrica de *recall*. Em uma das configurações de janela (5 segundos), o valor de *recall* foi superior no melhor modelo *baseline*. Em todas as demais configurações, os melhores modelos da Tese superaram os melhores modelos *baseline*. O *recall* mede o percentual de exemplos positivos que são corretamente identificados como positivos pelo classificador, ou

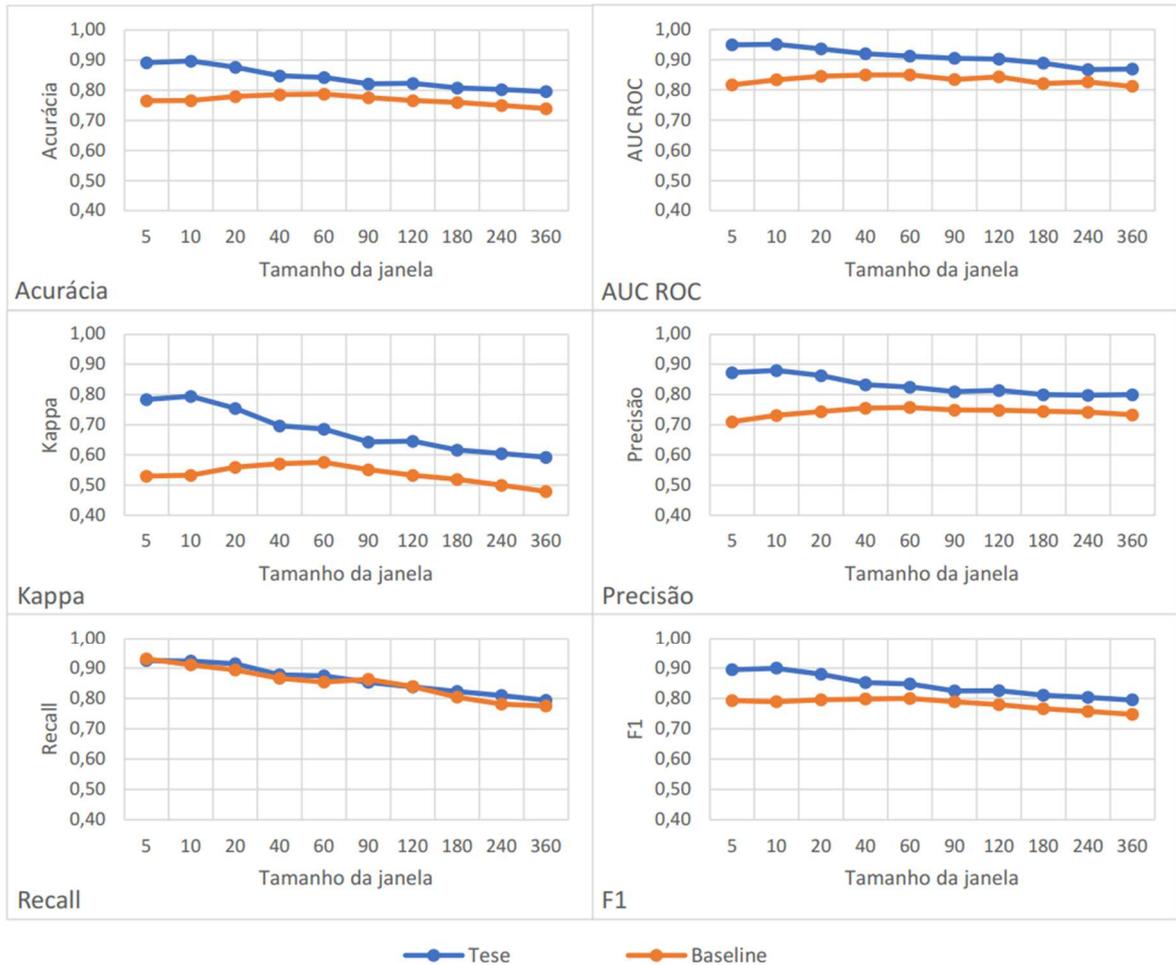
seja, se o aluno está confuso, o modelo identifica a confusão? Em um contexto educacional em que é rigorosamente necessário que o ambiente reaja a uma confusão do aluno, sempre que o aluno estiver confuso, os modelos da Tese não apresentam um benefício tão evidente, como nas demais métricas.

Em relação à métrica F1, todos os melhores modelos da Tese apresentaram valores superiores aos modelos *baseline*, em todas as configurações de janelas. A métrica F1 representa uma média harmônica de precisão e *recall*, útil quando não se sabe ao certo qual das duas métricas é mais importante no contexto de ensino.

Outra informação interessante apresentada na Tabela 23 é sobre os algoritmos dos melhores modelos da Tese e dos melhores modelos *baseline*. O algoritmo com mais ocorrência entre os melhores modelos é o *XGBoost* (10 ocorrências), seguido pelo algoritmo *Random Forest* (6 ocorrências), *Multilayer Perceptron* (3 ocorrências) e *Logistic Regression* (1 ocorrência). Considerando apenas os modelos da Tese, o *XGBoost* também é o algoritmo com mais ocorrências (7). Nos modelos *baseline*, três algoritmos empatam com três ocorrências cada, entre os melhores: *XGBoost*, *Random Forest* e *Multilayer Perceptron*. Considera-se importante reforçar que os modelos dos algoritmos apresentados na tabela são aqueles que obtiveram melhores valores de acurácia, uma métrica de desempenho geral. Outros algoritmos podem ter obtido melhores desempenhos em outras métricas.

O Gráfico 1 mostra o desempenho dos modelos da Tese e dos modelos *baseline*, em cada métrica, nas diferentes configurações de janelas de tempo fixo. O objetivo do gráfico é evidenciar a diferença de desempenho quando o tempo de observação aumenta. Uma primeira observação é que o desempenho superior dos modelos da Tese é mais aparente nas janelas de observação de 5, 10 e 20 segundos, com exceção na métrica de *recall*. A Tese levanta uma suposição para este efeito: os atributos de dados dos modelos *baseline*, como a “quantidade de teclas digitadas” e “tempo de movimento de mouse” não são tão representativos com tempos de observação menores. Em outras palavras, a suposição é que os atributos de dados dos trabalhos relacionados poderiam necessitar de um tempo maior de observação para serem mais representativos. Esta suposição não foi verificada pela Tese. Por outro lado, os atributos de dados sobre estimativas de conhecimento do aluno não seriam prejudicados pelo tempo menor de observação, pois as estimativas geradas por modelos BKT refletem um histórico de evidências de aplicação de conhecimento anterior (PELÁNEK, 2017a).

Gráfico 1 - Desempenho dos melhores modelos com janelas fixas e sem seleção de atributos



Fonte: Elaborado pelo autor.

A Tabela 24 apresenta o mesmo tipo de resultado apresentado na Tabela 23, mas relacionado à geração dos modelos com a etapa de seleção de atributos de dados. Os resultados de todos os modelos gerados foram disponibilizados no apêndice B. O objetivo de apresentar estes resultados foi verificar se, adicionando uma etapa de seleção de atributos de dados na geração dos modelos, poderiam ser encontrados resultados diferentes dos apresentados até aqui. No entanto, a aplicação de seleção de atributos de dados não apresentou, aparentemente, diferenças significativas. Nas métricas de acurácia, AUC ROC, *kappa*, precisão e F1, as diferenças dos resultados foram de, no máximo, três casas decimais. Na métrica de *recall*, a diferença foi de, no máximo, duas casas decimais, e em algumas janelas (5, 90 e 180), o *recall* foi maior nos modelos *baseline*. Não foram aplicados testes estatísticos comparando os modelos gerados sem seleção de atributos de dados e com seleção de atributos de dados.

Tabela 24 – Desempenho e comparação dos melhores modelos com janelas fixas e com seleção de atributos de dados

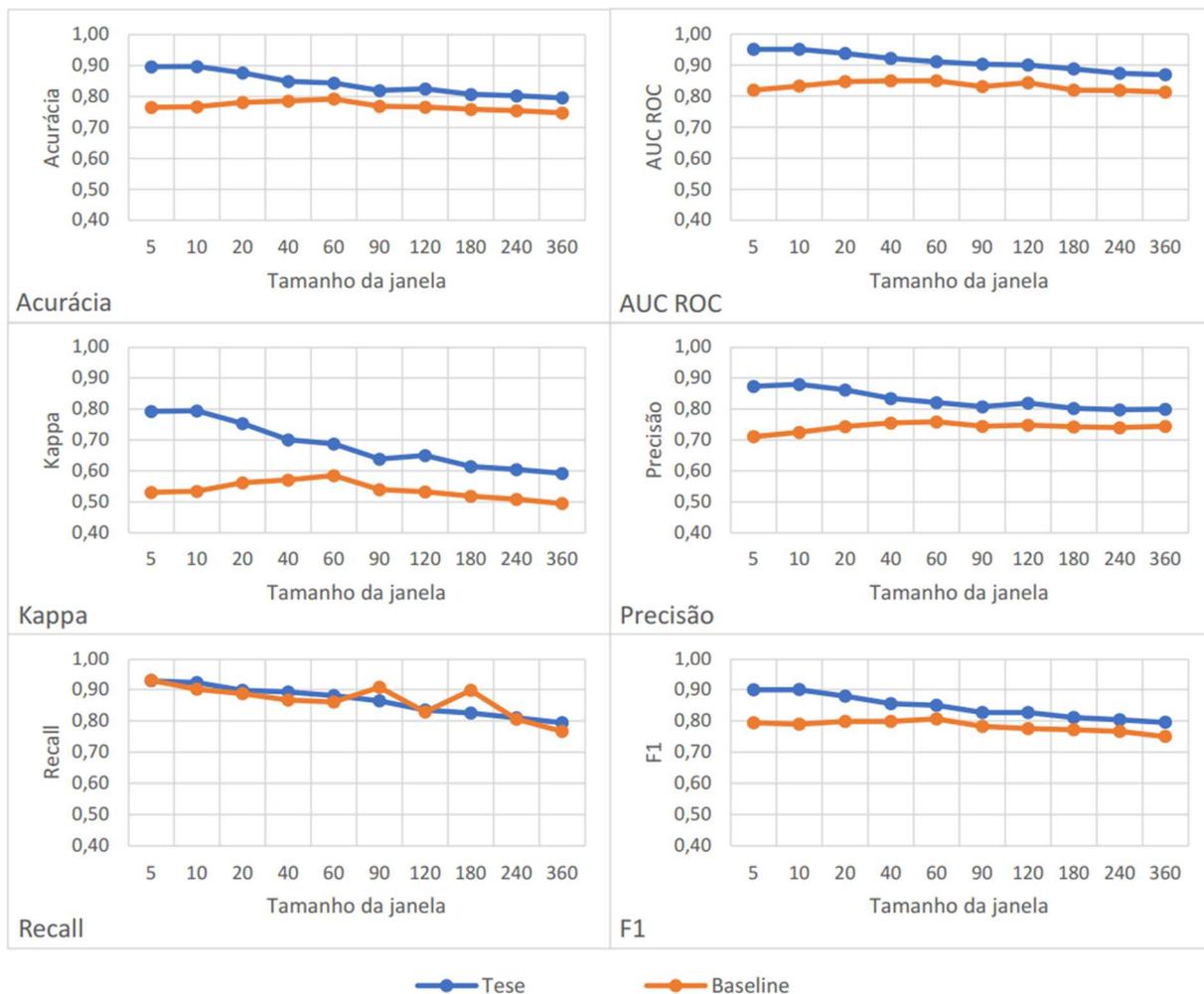
Jan.	Algor.	Acur.	AUC ROC	Kappa	Prec.	Recall	F1	Teste McNemar		
								chi <sup>2</sup>	p	
5	Tese	XGB	,89614	,95054	,79227	,87339	,92832	,89938	281,52011	<,00001
	Base	XGB	,76532	,81884	,53056	,71107	<b>,92938</b>	,79362		
10	Tese	RF	<b>,89725</b>	<b>,95082</b>	<b>,79452</b>	<b>,87981</b>	,92202	<b>,89992</b>	278,14603	<,00001
	Base	RF	,76712	,83245	,53427	,72463	,90095	,78909		
20	Tese	XGB	,87657	,93752	,75313	,86215	,89725	,87915	367,02412	<,00001
	Base	XGB	,78073	,84636	,56145	,74368	,88666	,79802		
40	Tese	XGB	,84938	,92118	,70088	,83414	,89179	,85449	42,22340	<,00001
	Base	XGB	,78563	,84947	,57123	,75542	,86611	,79769		
60	Tese	XGB	,84358	,91113	,68721	,82066	,88094	,84939	152,94943	<,00001
	Base	XGB	,79272	,84952	,58549	,75901	,86044	,80603		
90	Tese	SVM	,81930	,90307	,63864	,80755	,86368	,82637	12,56281	<b>,00039</b>
	Base	SVM	,76884	,83076	,53998	,74457	,90687	,78174		
120	Tese	RF	,82501	,90040	,64998	,81965	,83461	,82666	31,03030	<,00001
	Base	RF	,76612	,84344	,53219	,74800	,82779	,77538		
180	Tese	XGB	,80734	,88776	,61471	,80253	,82496	,81061	288,34042	<,00001
	Base	XGB	,75904	,81888	,51809	,74282	,89732	,77163		
240	Tese	XGB	,80221	,87326	,60440	,79828	,80930	,80351	313,00317	<,00001
	Base	XGB	,75421	,81799	,50842	,73956	,80542	,76577		
360	Tese	RF	,79610	,86917	,59220	,80007	,79386	,79504	259,09293	<,00001
	Base	RF	,74747	,81298	,49500	,74459	,76639	,75010		

Fonte: Elaborado pelo autor.

Outra diferença entre os resultados sem seleção de atributos e com seleção de atributos é que na janela de 90 segundos, a diferença estatística entre o melhor modelo da Tese e o melhor modelo *baseline* foi menos significativa, embora ainda abaixo do nível de significância estatística ( $\alpha = 0,05$ ). Ao invés de  $p < 0,00001$ , foi encontrado  $p = 0,00039$ . Além disso, o *XGBoost* continuou sendo o algoritmo com mais ocorrências (12), seguido por *Random Forest* (6 ocorrências) e *Support Vector Machines* (2 ocorrências).

O Gráfico 2 mostra o desempenho dos modelos da Tese e dos modelos *baseline*, em cada métrica, nas janelas de dados de tempo fixo, com seleção de atributos de dados. O gráfico mostra apenas uma alteração visível na métrica de *recall*, uma diferença já descrita em parágrafos anteriores. Uma conclusão é que os modelos da Tese e os modelos *baseline* não se beneficiaram tanto de um procedimento de seleção dos melhores atributos de dados.

Gráfico 2 - Desempenho dos melhores modelos com janelas fixas e com seleção de atributos



Fonte: Elaborado pelo autor.

Os gráficos 1 e 2 mostram uma aparente perda de desempenho de todos os modelos, à medida que o tempo de observação aumenta. Esse efeito parece ser mais evidente na métrica *kappa*, e nos modelos da Tese, em geral. Este efeito já era esperado pelo pesquisador. Uma dúvida levantada pelo pesquisador, antes da obtenção dos resultados, foi a seguinte: qual o efeito da subnotificação de relato de confusão dos alunos, no desempenho dos modelos? É conhecido na literatura que uma das desvantagens dos métodos de autorrelato de emoções é que o aumento da carga cognitiva do aluno, que precisa desenvolver a tarefa de programação juntamente da tarefa de relato de confusão, pode fazer o aluno não se engajar na tarefa de autorrelato (PORAYSKA-POMSTA et al., 2013), ocasionando na subnotificação de relatos de confusão. Isso poderia ocasionar a geração de amostras de dados que refletem padrões de confusão, mas vinculados a rótulos de “não confusão”. No entanto, uma suposição da Tese, baseada em observações em testes piloto, é que quando o aluno enfrenta uma confusão que o deixa travado na tarefa, o aluno se desengaja da tarefa e relata sua confusão. Esta é apenas uma suposição não verificada pela Tese, que poderia explicar a redução de desempenho dos modelos

à medida que as janelas crescem. Outra suposição, também não verificada, é que a redução da quantidade de amostras, à medida que o tamanho das janelas aumenta (ver Tabela 22), poderia ter contribuído para a queda de desempenho dos modelos.

A Tabela 25 apresenta os valores de acurácia, AUC ROC, *Kappa*, precisão, *recall* e F1 dos modelos da Tese e dos modelos *baseline* que apresentaram melhor acurácia na janela de observação de tempo variável, sem e com a seleção dos atributos de dados. Os resultados de todos os modelos gerados, foram disponibilizados no apêndice C. A Tabela 25 também apresenta os resultados de testes estatísticos (*McNemar*) que comparam a acurácia preditiva dos modelos da Tese e *baseline*.

Tabela 25 - Desempenho e comparação dos melhores modelos com janela variável

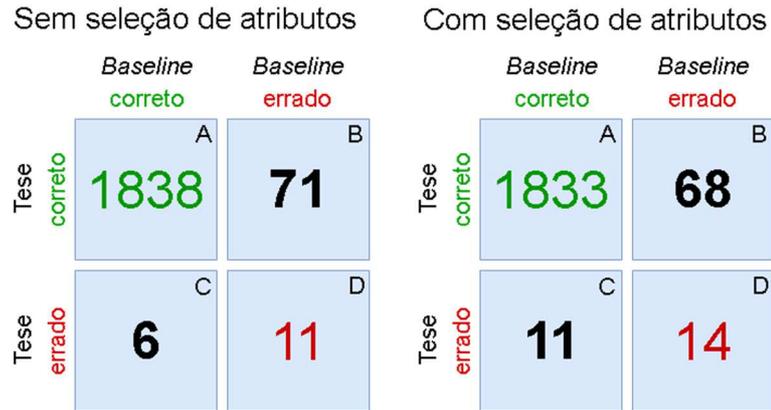
SA	Algor.	Acur.	AUC ROC	Kappa	Prec.	Recall	F1	Teste McNemar		
								chi <sup>2</sup>	p	
Não	Tese	XGB	<b>,87850</b>	<b>,94433</b>	<b>,75703</b>	,86478	,89827	<b>,88093</b>	53,19480	<,00001
	Base	XGB	,86762	,93583	,73526	<b>,90226</b>	<b>,90245</b>	,87169		
Sim	Tese	XGB	<b>,88161</b>	<b>,94435</b>	<b>,76324</b>	,86710	<b>,90346</b>	<b>,88414</b>	39,69620	<,00001
	Base	XGB	,86762	,93583	,73526	<b>,90226</b>	,90245	,87169		

Fonte: Elaborado pelo autor.

A primeira observação, é que em todos os resultados das métricas de desempenho geral (acurácia, AUC ROC e *Kappa*), nos cenários com e sem seleção de atributos de dados (SA), os modelos da Tese obtiveram valores superiores. Embora a diferença de valores na métrica de acurácia tenha sido aparentemente estreita de, no máximo, 0,01399 (com seleção de atributos de dados), a diferença de acurácia preditiva foi significativa ( $p < ,00001$ ) em ambos os cenários. A Figura 31 mostra a tabela de contingência usada pelos testes estatísticos McNemar. Os testes encontraram diferenças significativas nas predições que os modelos discordam (células B e C), nos cenários sem seleção de atributos e com a seleção.

Nos modelos sem seleção de atributos, os modelos *da Tese* apresentaram desempenho inferior nas métricas de *precisão* e *recall*, mas superior na F1. No entanto, o modelo da Tese se beneficiou da seleção de atributos de dados, obtendo um valor superior no *recall*. Outra observação é que o algoritmo *XGBoost* apresentou o melhor desempenho de acurácia nos melhores modelos da Tese e *baseline* com janela variável, com e sem seleção dos melhores atributos de dados. Estes resultados confirmam o *XGBoost* como o algoritmo que resultou nos melhores valores de acurácia nas janelas de tempo fixo e variável. O Gráfico 3 mostra o desempenho dos melhores modelos nas diferentes métricas.

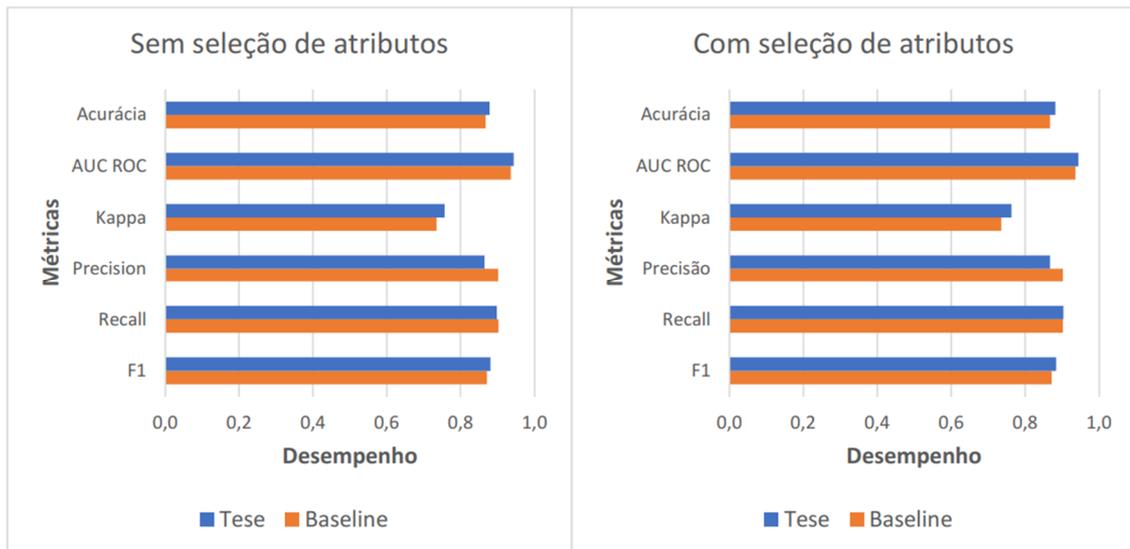
Figura 31 – Tabelas de contingência usadas nas comparações dos modelos de janela variável



Fonte: Adaptado e traduzido de (RASCHKA, 2018).

Uma conclusão geral é que, aparentemente, os modelos da Tese e *baseline* se beneficiaram da abordagem de janelas de observação de tempo variável, principalmente os modelos *baseline*. Nas janelas fixas, o melhor modelo *baseline* obteve a melhor acurácia (0,79272) na janela de 60 segundos, com seleção de atributos de dados. Com janela variável, o melhor modelo *baseline* obteve acurácia de 0,86762, mais de 7 pontos percentuais superior. Já o melhor modelo da Tese com janela variável obteve a terceira melhor acurácia (0,88161) de todos os resultados, apenas inferior ao desempenho nas janelas de 5 e 10 segundos.

Gráfico 3 - Desempenhos dos melhores modelos com janela variável



Fonte: Elaborado pelo autor.

Os próximos parágrafos apresentam os resultados do segundo método implementado para os objetivos específicos OE1 e OE2, descrito na Seção 9.2. O método busca identificar os atributos de dados mais relevantes para os modelos da Tese e *baseline*. O

método foi organizado em duas etapas. Primeiramente são apresentados os resultados da primeira etapa: ranqueamento dos atributos de dados que mais foram selecionados na etapa de seleção de atributos (método RFE) dos modelos da Tese e *baseline*, nas diferentes janelas.

A Tabela 26 mostra um panorama sobre a quantidade de atributos de dados selecionados pelo método RFE nos modelos da Tese e *baseline* que apresentaram melhor acurácia, em cada configuração de janela (fixa e variável). Também é apresentado o percentual de atributos selecionados, em relação ao total de atributos disponíveis (65 nos modelos da Tese e 23 nos modelos *baseline*). A tabela mostra que, entre os melhores modelos da Tese, na janela de 360 segundos, foi necessário selecionar todos os atributos para obter o melhor desempenho. Nas demais janelas, remover atributos resultou em um melhor desempenho do modelo. Nos modelos *baseline*, nas janelas de 40, 60, 120 e variável, todos os atributos foram selecionados para obter um melhor desempenho. No entanto, como apresentado em discussões anteriores, a melhoria de desempenho de acurácia foi obtida, na maioria das janelas e métricas, apenas a partir de três casas decimais. Aparentemente, seleção de atributos de dados não teve efeito tão evidente na geração dos modelos com janelas fixas e variável, embora não tenha sido aplicado testes estatísticos comparando os resultados dos cenários sem e com seleção de atributos de dados.

Tabela 26 – Atributos de dados selecionados nos melhores modelos

<i>Janela</i>	<i>Modelos da Tese</i>		<i>Modelos baseline</i>	
	<i>Quantidade de atributos selecionados</i>	<i>Percentual selecionado</i>	<i>Quantidade de atributos selecionados</i>	<i>Percentual selecionado</i>
5	55	,84615	18	,78260
10	55	,84615	19	,82608
20	53	,81538	21	,91304
40	60	,92307	23	1
60	64	,98461	23	1
90	63	,96923	16	,69565
120	60	,92307	23	1
180	63	,96923	14	,60869
240	64	,98461	16	,69565
360	65	1	17	,73913
variável	60	,92307	23	1

Fonte: Elaborado pelo autor.

A Tese também buscou verificar quais atributos de dados foram selecionados pelo método RFE nos modelos de melhor acurácia da Tese e *baseline*, em cada janela de dados (fixa e variável). Nos modelos da Tese, do total de 42 atributos de dados sobre estimativas de conhecimento do aluno, 33 atributos (0,78571 do total) foram selecionados pelo método RFE em todas as configurações de janelas, incluindo os atributos sobre a **média de  $p(L)$  nos**

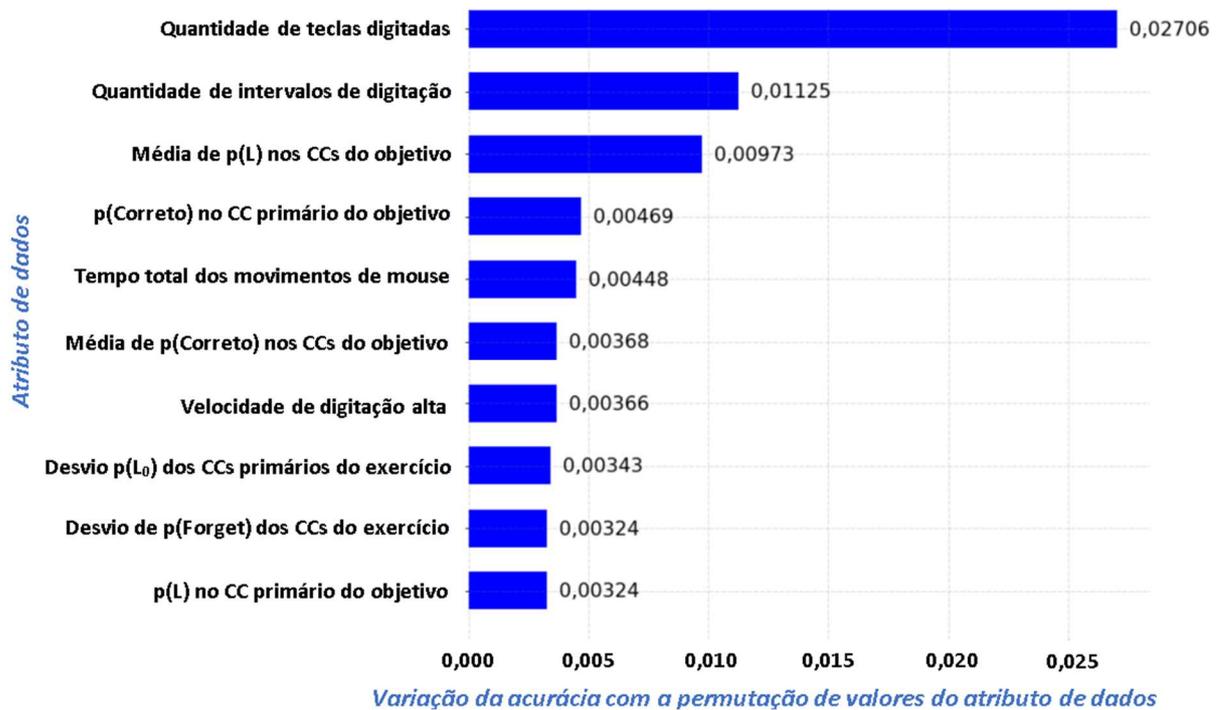
**componentes de conhecimento do objetivo selecionado** e a **média de  $p(\text{Correto})$  nos componentes de conhecimento do objetivo selecionado**. Os demais atributos que também foram selecionados nos melhores modelos da Tese, em todas as janelas de dados, sobre a interação do aluno com o ambiente, estão a **quantidade de erro de sintaxe no código**, a **quantidade de cliques do mouse no código**, a **quantidade de movimentos de mouse**, o **tempo total de movimentos de mouse**, a **quantidade de teclas digitadas no código**, a **quantidade de teclas de *backspace* no código**, a **quantidade de teclas de controle (setas de movimentação) no código**, a **quantidade de intervalos de digitação** e o **tempo ocioso**.

Nos modelos *baseline*, em que as amostras são compostas apenas pelos atributos de interação com o ambiente (mouse, teclado, entre outros), e representam os trabalhos relacionados, os atributos que foram selecionados em todas as configurações de janelas foram os seguintes: **quantidade de teclas digitadas no código**, **quantidade de teclas de *backspace* no código**, **quantidade de intervalos de digitação**, **quantidade de cliques do mouse no código**, **quantidade de erro de sintaxe no código**, e **velocidade de movimentos de mouse**. Apenas um destes atributos (velocidade de movimentos de mouse) não aparece também entre os de maior ocorrência nos modelos da Tese, o que reforça a relevância destes atributos de interação com o ambiente. Os apêndices D e E apresentam as tabelas completas com a quantidade de ocorrências de cada atributo nos melhores modelos da Tese e *baseline*, nas 11 configurações de janelas.

Os próximos parágrafos apresentam os resultados da segunda etapa do método 2: aplicação do método de importância de permutação de atributos de dados nos melhores modelos da Tese e *baseline* de cada configuração de janela de dados, e ranqueamento dos dez atributos mais relevantes em cada configuração de janela.

O Gráfico 4 mostra os dez atributos de dados mais relevantes no melhor modelo da Tese, que também é o melhor modelo de todo o trabalho. A Tese considera como melhor modelo, aquele com maior valor para a métrica de acurácia. O melhor modelo foi encontrado para uma janela de 10 segundos, sem seleção de atributos, e obteve acurácia de 0,89725 (ver Tabela 23). Este modelo também obteve os valores mais altos nas métricas AUC ROC (0,95082), *kappa* (0,79452), precisão (0,87981) e F1 (0,89992). No Gráfico 4, a relevância dos atributos (eixo horizontal) foi obtida pelo método de importância de permutação de atributos, descrito na Seção 9.2. O valor da relevância de um atributo reflete a diminuição do escore (acurácia) do modelo, quando os valores do atributo são embaralhados aleatoriamente. Observa-se que entre os dez atributos mais relevantes no melhor modelo da Tese, seis são atributos sobre estimativas do conhecimento do aluno, e quatro são atributos que representam os trabalhos relacionados.

Gráfico 4 – Os dez atributos de dados mais relevantes no melhor modelo da Tese



Fonte: Elaborado pelo autor.

Ainda sobre o Gráfico 4, observa-se que os dois atributos mais relevantes são sobre a interação do aluno com o ambiente (quantidade de teclas digitadas e quantidade de intervalos de digitação), seguido por dois atributos sobre estimativas de conhecimento do aluno (média da probabilidade de dominar os componentes do objetivo em codificação e a probabilidade de aplicar corretamente o componente de conhecimento primário do objetivo em codificação).

A Tabela 27 apresenta informações estatísticas de média e desvio padrão encontrados para os atributos de dados mais relevantes no melhor modelo da Tese (janela de observação de 10 segundos), nas amostras vinculadas aos rótulos de confusão e de não confusão. O objetivo é observar o comportamento desses dados, quando eles foram relevantes para modelar as fronteiras de classificação de confusão e de não confusão, nos modelos de aprendizado de máquina. A aplicação de testes estatísticos não paramétricos *MannWhitney U* mostrou que a diferença dos valores médios para cada atributo de dados entre as amostras de confusão e de não confusão são todas significativas, com valor de  $p$  abaixo do nível de significância ( $\alpha = 0,05$ ) (valores em negrito na Tabela 27).

Tabela 27 – Informações sobre os atributos mais relevantes no melhor modelo da Tese

#	Atributo	Confusão		Não confusão		MW U $p$
		Média	Desvio	Média	Desvio	

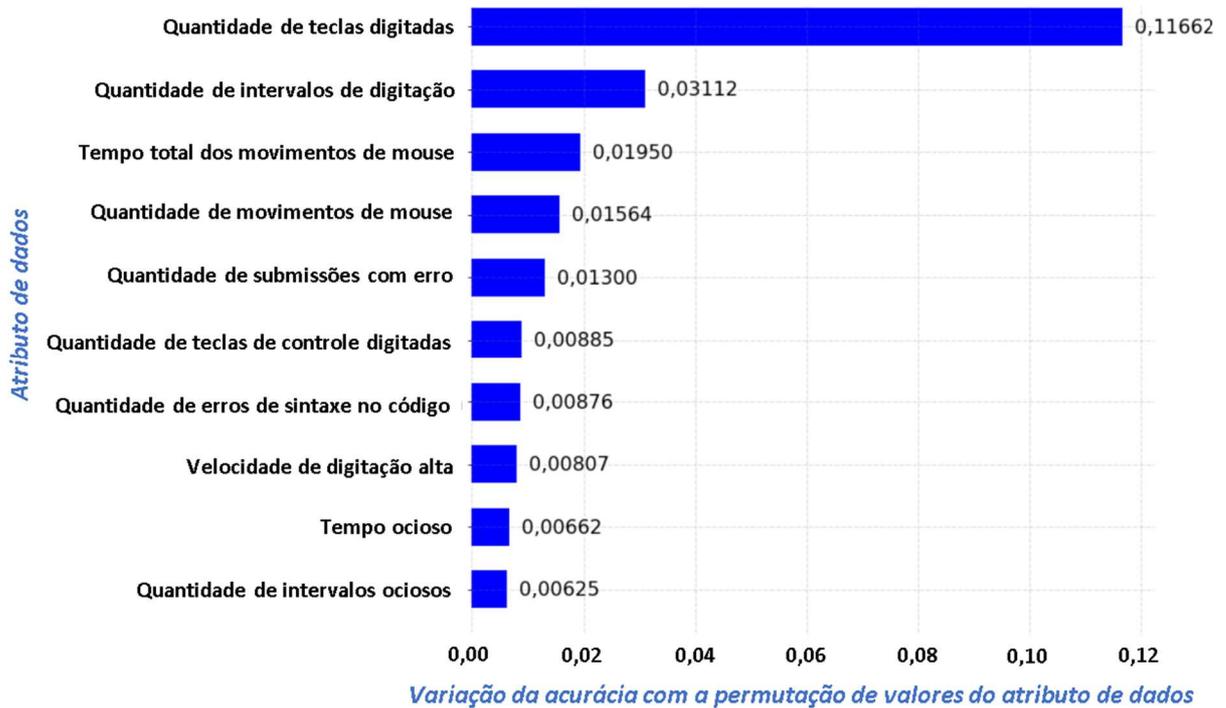
1	Quantidade de teclas digitadas	1,40569	3,71375	10,39409	12,57671	<,00001
2	Quantidade de intervalos de digitação	0,37091	0,80310	1,12118	1,33717	<,00001
3	Média de p(L) nos CCs do objetivo	0,68997	0,16661	0,89796	0,12073	<,00001
4	p(Correto) no CC primário do objetivo	0,45546	0,26097	0,77795	0,22304	<,00001
5	Tempo total de movimentos de mouse	2,21374	1,48403	1,56990	1,46963	<,00001
6	Média p(Correto) nos CCs do objetivo	0,71391	0,13421	0,87345	0,10695	<,00001
7	Velocidade de digitação alta	0,12539	0,33134	0,44784	0,49753	<,00001
8	Desvio p(L <sub>0</sub> ) nos CCs primários do exercício	0,10121	0,07449	0,09149	0,06507	,00459
9	Desvio p(Forget) nos CCs do exercício	0,09461	0,06068	0,08004	0,04855	<,00001
10	p(L) no CC primário do objetivo	0,48652	0,30520	0,83165	0,22230	<,00001

Fonte: Elaborado pelo autor.

Ainda na Tabela 27, observando os valores de média, verifica-se que, apesar de nos momentos de confusão os alunos terem apresentado menos atividade de digitação, eles passaram mais tempo movimentando o mouse. Além disso, quando não estavam confusos, os alunos que estavam digitando, digitaram num ritmo mais rápido. Em relação aos dados sobre estimativas de conhecimento do estudante, quando não confusos, os alunos apresentaram maior probabilidade de dominarem os componentes do objetivo do exercício em codificação, e maior probabilidade de aplicarem corretamente os componentes do objetivo do exercício em codificação. Também é possível observar que nos momentos de confusão, havia maior variação entre as probabilidades de conhecimento inicial nos componentes primários e maior variação entre as probabilidades de esquecimento dos componentes do exercício.

O Gráfico 5 mostra os dez atributos de dados mais relevantes no melhor modelo *baseline*, onde não são considerados os atributos sobre estimativas do conhecimento do aluno. O melhor modelo *baseline* foi encontrado para uma janela variável, sem seleção de atributos de dados, e obteve acurácia de 0,86762 (ver Tabela 25). Este modelo também obteve os valores mais altos nas métricas AUC ROC (0,93583), *kappa* (0,73526), precisão (0,90226) e F1 (0,87169). Observa-se que os dois atributos mais relevantes no melhor modelo *baseline* também são os dois atributos mais relevantes no melhor modelo da Tese. Estes resultados evidenciam que a quantidade de teclas digitadas e a quantidade de intervalos de digitação, são atributos interessantes para a detecção da confusão no contexto da Tese.

Gráfico 5 – Os dez atributos de dados mais relevantes no melhor modelo *baseline*



Fonte: Elaborado pelo autor.

Os apêndices F e G mostram os dez atributos de dados mais relevantes em cada configuração de janela de observação (fixas e variável) dos modelos da Tese e *baseline*.

Com o objetivo de resumir os resultados apresentados nestes apêndices, a Tabela 28 mostra uma relação dos dez atributos de dados que mais ocorrem entre os dez atributos de dados mais relevantes em cada configuração de janela dos melhores modelos da Tese e *baseline*. Observa-se que entre os atributos mais relevantes, cinco são atributos de dados sobre estimativas do conhecimento dos estudantes e outros cinco sobre a interação com o ambiente.

Tabela 28 – Os 10 atributos mais frequentes entre os 10 atributos mais relevantes nas janelas

<i>Modelos da Tese</i>			<i>Modelos baseline</i>		
#	Atributo	Oc.	#	Atributo	Oc.
1	Quantidade de teclas digitadas	10	1	Quantidade de teclas digitadas	11
2	<b>Média de p(L) nos CCs do objetivo</b>	<b>9</b>	2	Quantidade de intervalos de digitação	10
3	Quantidade de intervalos de digitação	8	3	Quantidade de submissões com erro	9
4	Tempo total de movimentos de mouse	8	4	Velocidade de digitação alta	8
5	<b>p(L) no CC primário do objetivo</b>	<b>7</b>	5	Quantidade de movimentos de mouse	7
6	<b>Média p(Correto) nos CCs do exercício</b>	<b>6</b>	6	Tempo total de movimentos de mouse	7
7	Quantidade de movimentos de mouse	6	7	Velocidade movimentos mouse alta	6
8	Quantidade de erros de sintaxe no código	5	8	Velocidade movimentos mouse média	6
9	<b>Média de p(L) nos CCs do exercício</b>	<b>4</b>	9	Quantidade de cliques no código	5
10	<b>p(Correto) no CC primário do objetivo</b>	<b>4</b>	10	Quantidade de intervalos ociosos	5

Fonte: Elaborado pelo autor.

Para finalizar a apresentação dos resultados obtidos na aplicação dos métodos para os objetos específicos OE1 e OE2, a Tabela 29 apresenta um resumo sobre os modelos da Tese e *baseline* que obtiveram os maiores valores para cada uma das métricas de desempenho. Observa-se que um único modelo obteve os melhores valores de acurácia, AUC ROC, *kappa*, precisão e F1, na abordagem da Tese, para uma janela de observação de 10 segundos, com e sem a etapa de seleção de atributos de dados (SA). Na abordagem *baseline*, também foi encontrado um único modelo que obteve os melhores valores de acurácia, AUC ROC, *kappa*, precisão e F1, para uma janela variável de observação, com e sem a etapa de seleção de atributos de dados (SA). Os algoritmos dos modelos de melhor desempenho foram o *Random Forest* e o *XGBoost*.

Tabela 29 – Os melhores valores obtidos em cada métrica nos modelos da Tese e *baseline*

<i>Métrica</i>	<i>Tese</i>				<i>Baseline</i>			
	<i>Valor</i>	<i>Algoritmo</i>	<i>Janela</i>	<i>SA</i>	<i>Valor</i>	<i>Algoritmo</i>	<i>Janela</i>	<i>SA</i>
Acurácia	<b>,89725</b>	RF	10	Ambos	,86762	XGB	Variável	Ambos
<i>AUC ROC</i>	<b>,95082</b>	RF	10	Ambos	,93583	XGB	Variável	Ambos
<i>Kappa</i>	<b>,79452</b>	RF	10	Ambos	,73526	XGB	Variável	Ambos
Precisão	,87981	RF	10	Ambos	<b>,90226</b>	XGB	Variável	Ambos
<i>Recall</i>	,92832	XGB	5	Sim	<b>,92938</b>	XGB	5	Sim
F1	<b>,89992</b>	RF	10	Ambos	,87169	XGB	Variável	Ambos

Fonte: Elaborado pelo autor.



## 10 EFEITOS DE ESTIMATIVAS DE CONHECIMENTO DO ALUNO NA GENERALIZAÇÃO DOS MODELOS PARA ALUNOS HETEROGÊNEOS

Este capítulo descreve o método utilizado pela Tese para alcançar o objetivo específico OE3, e apresenta os resultados obtidos. Alcançar este objetivo específico possibilita encontrar respostas para a subquestão de pesquisa SQP3. A Seção 10.1 descreve o método utilizado para alcançar o objetivo. Os resultados são apresentados na Seção 10.2.

### 10.1 Método

Para alcançar o objetivo específico OE3, o trabalho da Tese verifica o poder de generalização dos modelos para estudantes com características heterogêneas. A Tese assume que os alunos participantes das turmas de ensino técnico e os alunos das turmas de ensino superior constituem dois grupos com características heterogêneas. O Capítulo 5 apresentou as características emocionais, de idade e de experiências anteriores em programação desses dois grupos de estudantes. Não foram encontradas diferenças significativas entre os grupos em relação as características emocionais, mas foram encontradas diferenças significativas em relação as idades e as experiências com programação.

O método utilizado para alcançar o objetivo específico OE3 é similar ao primeiro método utilizado para alcançar os objetivos OE1 e OE2, descrito na Seção 9.1: **geração de modelos de aprendizado de máquina que refletem a abordagem da Tese e a abordagem baseline, e a comparação dos seus resultados**. O método utilizado aqui utiliza as mesmas *pipelines*, algoritmos, métricas e testes estatísticos do método apresentado na Seção 9.1. A diferença está apenas na etapa de validação dos modelos. No presente método, ao invés de validar os modelos com validação cruzada *k-fold* ( $k = 10$ ), os modelos foram treinados com as amostras de alunos de um determinado nível de ensino e foram testados com as amostras de alunos do outro nível de ensino. O método é similar ao de uma validação cruzada *k-fold*, com  $k = 2$ , no qual  $k$  corresponde ao nível de ensino dos participantes da pesquisa (ensino técnico e ensino superior). Para cada modelo gerado, são realizadas  $k$  iterações (2 iterações): na primeira iteração, aplica-se o cenário 1, no qual o modelo é treinado com as amostras dos alunos do ensino técnico, testado com as amostras dos alunos do ensino superior e são gerados valores para as métricas de desempenho. Na segunda iteração, aplica-se o cenário 2, no qual o modelo é treinado com as amostras dos alunos do ensino superior, testado com as amostras dos alunos do ensino técnico e são gerados valores para as métricas. Ao final das duas iterações, para cada

métrica é calculada uma média simples dos valores encontrados nas duas iterações. Em cada iteração, é realizada a otimização dos hiperparâmetros do modelo.

O método é executado para cada uma das 11 configurações de janelas de observação de dados (5, 10, 20, 40, 60, 90, 120, 180, 240, 360 e variável) dos modelos da Tese e *baseline* e para cada um dos 10 algoritmos utilizados na Tese. Também são realizados testes estatísticos (McNemar) que comparam o desempenho do melhor modelo da Tese e o melhor modelo *baseline* em determinado cenário, como a comparação da acurácia preditiva do melhor modelo da Tese com o melhor modelo *baseline* quando as amostras dos alunos do ensino técnico são usadas no treino e as amostras dos alunos do ensino superior são usadas no teste, para uma determinada janela de observação.

Figura 32 – Pseudocódigo de geração dos modelos para verificação de generalização

```

1.  abordagens <- [tese, baseline]
2.  janelas <- [5,10,20,40,60,90,120,180,240,360,variável]
3.  algoritmos <- [LR,SVM,NB,kNN,SGD,DT,AB,RF,XGB,MLP]
4.  níveis <- [técnico, graduação]
5.
6.  para cada abordagem em abordagens faça
7.    para cada janela em janelas faça
8.      para cada nível em níveis faça
9.        amostrasTreino, amostrasTeste <- obterAmostras(abordagem,
                                                         janela,
                                                         nível)
10.       amostrasTreino, amostrasTeste <- normaliza(amostrasTreino,
                                                         amostrasTeste)
11.      para cada algoritmo em algoritmos faça
12.        hiperParams <- obterMelhoresHiperParams(algoritmo,
                                                         amostrasTreino,
                                                         amostrasTeste)
13.       resultados[abordagem, janela, nível, algoritmo] <- valida(algoritmo,
                                                                    hiperParams,
                                                                    amostrasTreino,
                                                                    amostrasTeste)
14.       melhorModelo[abordagem, janela, nível] <- obterMelhorModelo(resultados[abordagem,
                                                                               janela,
                                                                               nível])

```

Fonte: Elaborado pelo autor.

A Figura 32 apresenta um pseudocódigo que descreve o fluxo de geração dos modelos. Primeiramente, nas linhas 1, 2, 3 e 4, são carregados em memória as abordagens das amostras (Tese e *baseline*), as janelas (5, 10, 20, 40, 60, 90, 120, 180, 240, 360 e variável), os algoritmos de aprendizado de máquina, e os níveis de ensino (técnico e graduação). Para cada nível de ensino, em cada janela de cada abordagem (linhas 6, 7 e 8), são executadas as operações

descritas a seguir. São carregadas em memória as amostras de treino e testes relativas à abordagem, à janela e ao nível de ensino (linha 9). Para determinado nível de ensino, as amostras de treino são as amostras relativas a este nível de ensino, enquanto as amostras de teste são as amostras do outro nível de ensino. Na linha 10, as amostras carregadas são atualizadas por uma operação de normalização. Para cada um dos algoritmos (linha 11), são executadas as seguintes operações. São obtidos os melhores hiperparâmetros para o algoritmo da iteração, de acordo com as características das amostras carregadas (linha 12). Na linha 13, é realizada a validação (treino com as amostras de um nível de ensino e teste com as amostras do outro nível de ensino) do modelo composto pelo algoritmo, seus hiperparâmetros e as amostras carregadas. O resultado da validação é guardado em memória. Observe que a iteração mais interna (linha 11) é executada 440 vezes. Além disso, para cada janela, em cada abordagem, e para cada nível de ensino, é guardado em memória o modelo que obteve o melhor desempenho de acurácia (linha 14).

Figura 33 – Pseudocódigo de geração dos modelos para verificação de generalização com seleção de atributos de dados

```

1.  abordagens <- [tese, baseline]
2.  janelas <- [5,10,20,40,60,90,120,180,240,360,variável]
3.  algoritmos <- [LR,SVM,NB,kNN,SGD,DT,AB,RF,XGB,MLP]
4.  níveis <- [técnico, graduação]
5.
6.  para cada abordagem em abordagens faça
7.    para cada janela em janelas faça
8.      para cada nível em níveis faça
9.        amostrasTreino, amostrasTeste <- obterAmostras(abordagem,
                                                         janela,
                                                         nível)
10.       amostrasTreino, amostrasTeste <- normaliza(amostrasTreino,
                                                    amostrasTeste)
11.       melhorModelo <- obterMelhorModelo(abordagem,
                                             janela,
                                             nível)
12.       melhoresAtributos <- obterMelhoresAtributos(melhorModelo,
                                                    amostrasTreino,
                                                    amostrasTeste)
13.       amostrasTreino, amostrasTeste <- filtraMelhoresAtributos(melhoresAtributos,
                                                                    amostrasTreino,
                                                                    amostrasTeste)
14.     para cada algoritmo em algoritmos faça
15.       hiperParams <- obterMelhoresHiperParams(algoritmo,
                                                  amostrasTreino,
                                                  amostrasTeste)
16.       resultados[abordagem, janela, algoritmo, nível] <- valida(algoritmo,
                                                                    hiperParams,
                                                                    amostrasTreino,
                                                                    amostrasTeste)

```

Fonte: Elaborado pelo autor.

Com o objetivo de verificar se podem ser encontrados subconjuntos de atributos de dados que melhorem o desempenho de generalização dos modelos para estudantes heterogêneos, todos os modelos foram gerados uma segunda vez, com a inclusão da etapa de seleção dos melhores atributos de dados. A Figura 33 apresenta um pseudocódigo que descreve o fluxo de geração dos modelos, acrescentando a atividade de seleção de atributos. Primeiramente, nas linhas 1, 2, 3 e 4, são carregados em memória as abordagens das amostras (Tese e *baseline*), as janelas (5, 10, 20, 40, 60, 90, 120, 180, 240, 360 e variável), os algoritmos de aprendizado de máquina, e os níveis de ensino (técnico e graduação). Para cada uma das abordagens (linha 6), são executadas as operações descritas a seguir. Para cada uma das janelas (linha 7), são carregadas as operações descritas a seguir. Para cada um dos níveis de ensino (linha 8), são carregadas as operações descritas a seguir. São carregadas em memória as amostras de treino e testes relativas à abordagem, à janela e ao nível de ensino (linha 9). Para determinado nível de ensino, as amostras de treino são as amostras relativas a este nível de ensino, enquanto as amostras de teste são as amostras do outro nível de ensino.

Na linha 10 do pseudocódigo da Figura 33, as amostras carregadas são atualizadas por uma operação de normalização. Na linha 11, o melhor modelo obtido para a abordagem, janela e nível de ensino, é carregado em memória. Este melhor modelo é obtido na execução do pseudocódigo da Figura 32 (linha 14). Isso significa que a geração dos modelos com seleção de atributos de dados é executada depois da geração dos modelos sem seleção de atributos de dados (pseudocódigo da Figura 32). Na linha 12, ocorre a obtenção dos melhores atributos de dados. Para isso, é utilizado o melhor modelo treinado para a abordagem, janela e nível, e as amostras carregadas. Após, as amostras são atualizadas com uma operação de filtro (linha 13), no qual são mantidos apenas os atributos de dados selecionados. Na sequência do fluxo, na linha 14, para cada algoritmo, são realizadas as operações descritas a seguir. São obtidos os melhores hiperparâmetros para o algoritmo e as amostras (linha 15). Por fim, é realizada a validação (treino com as amostras de um nível de ensino e teste com as amostras do outro nível de ensino) do modelo composto pelo algoritmo, seus hiperparâmetros e as amostras carregadas. O resultado da validação é guardado em memória (linha 16).

## 10.2 Resultados

Esta Seção apresenta os resultados encontrados na aplicação do método descrito na Seção anterior. A Tabela 30 apresenta o desempenho dos melhores modelos da Tese e dos melhores modelos *baseline* nas diferentes métricas (acurácia, *AUC ROC*, *kappa*, precisão,

recall e F1) e diferentes configurações de janelas de observação de tempo fixo. São considerados “melhores modelos” aqueles com valores superiores na métrica de acurácia.

Tabela 30 – Desempenho dos modelos de janela fixa quando treinados com amostras de cada grupo de estudantes

Jan.	Algor.	Acur.	AUC ROC	Kappa	Prec.	Recall	F1	Teste McNemar		
								chi <sup>2</sup>	p	
<i>Treinamentos com o ensino técnico e testes com o ensino superior</i>										
5	Tese	SGD	<b>,84104</b>	<b>,89730</b>	<b>,68209</b>	<b>,89639</b>	<b>,92354</b>	<b>,84894</b>	95,24497	<,00001
	Base	ADB	,68511	,73062	,37022	,63068	<b>,92354</b>	,73938		
10	Tese	SGD	,83165	,89536	,66330	,83766	,85685	,83416	79,24163	<,00001
	Base	ADB	,68346	,72818	,36693	,63343	,91330	,74262		
20	Tese	SGD	,79149	,85317	,58299	,77400	,85627	,79961	29,67804	<,00001
	Base	LR	,71153	,76608	,42307	,65761	,90890	,75513		
40	Tese	RF	,73565	,81579	,47131	,82587	,74590	,72017	0,0	1,0
	Base	LR	,73463	,78354	,46926	,68566	,90368	,77220		
60	Tese	SGD	,73966	,80102	,47933	,79329	,79132	,75147	1,01554	0,31357
	Base	RF	,72417	,78097	,44834	,66979	,89256	,76224		
90	Tese	SGD	,71928	,76350	,43855	,74346	1,0000	,72708	0,75625	0,38450
	Base	LR	,73199	,78220	,46398	,68159	,87711	,76465		
120	Tese	ADB	,71075	,78547	,42150	,81818	,78064	,69213	0,98299	0,32146
	Base	RF	,73010	,79596	,46021	,68706	,89892	,76199		
180	Tese	SGD	,71559	,78614	,43119	,80078	,77752	,73218	0,18652	0,66582
	Base	RF	,72362	,79833	,44724	,69306	,91513	,75125		
240	Tese	SGD	,69686	,78550	,39372	,78461	,77777	,71182	0,40251	0,52579
	Base	ADB	,70772	,78245	,41545	,67252	,87922	,74364		
360	Tese	SGD	,67272	,73214	,34545	,70129	,98181	,72727	7,85256	<b>0,00507</b>
	Base	MLP	,71948	,78919	,43896	,68903	,89610	,74537		
<i>Treinamentos com o ensino superior e testes com o ensino técnico</i>										
5	Tese	LR	,86695	,93536	,73390	,89419	,99141	,87843	13,14935	<b>0,00028</b>
	Base	NB	,81759	,87673	,63519	,81075	,92489	,83527		
10	Tese	SGD	<b>,87638</b>	<b>,94007</b>	<b>,75275</b>	<b>,94095</b>	<b>,99779</b>	<b>,88822</b>	34,75177	<,00001
	Base	NB	,79801	,86757	,59602	,80045	,91170	,81863		
20	Tese	SGD	,84000	,91446	,68000	,90322	,98666	,84905	4,81283	<b>0,02824</b>
	Base	LR	,80555	,87184	,61111	,82051	,80222	,80091		
40	Tese	MLP	,79662	,87326	,59325	,83612	,96853	,81881	0,54644	0,45977
	Base	MLP	,78427	,85357	,56853	,78048	,83595	,78807		
60	Tese	LR	,78295	,86693	,56590	,85317	,96818	,80083	0,0	1,0
	Base	MLP	,78181	,86191	,56363	,80845	,81363	,77197		
90	Tese	MLP	,75232	,84171	,50465	,80392	,98372	,74812	0,85207	0,35596
	Base	SGD	,73720	,82113	,47441	,75668	,78139	,74749		
120	Tese	MLP	,74760	,82356	,49521	,77591	,99521	,75854	1,05140	0,30518
	Base	SVM	,72846	,80439	,45693	,75202	,71531	,72147		
180	Tese	MLP	,72193	,80301	,44387	,75390	,98979	,73985	2,54326	0,11076
	Base	SGD	,69132	,77762	,38265	,73734	,64030	,66314		
240	Tese	SGD	,71008	,78355	,42016	,74008	,97479	,73764	1,44632	0,22911
	Base	SGD	,68627	,76021	,37254	,74007	,61904	,65578		
360	Tese	SGD	,71710	,77750	,43421	,73366	,97368	,72168	6,62727	<b>0,010042</b>
	Base	NB	,67105	,74992	,34210	,71495	,62500	,65517		

Fonte: Elaborado pelo autor.

Os resultados da Tabela 30 são apresentados para **dois cenários: no primeiro cenário**, os modelos são treinados com as amostras do ensino técnico e testados com as amostras do ensino superior; no **segundo cenário**, os modelos são treinados com as amostras do ensino superior e testados com as amostras do ensino técnico. Também são apresentados os resultados de testes estatísticos (McNemar) que comparam a acurácia dos melhores modelos da Tese e *baseline*. Todos os resultados (não apenas os melhores modelos), estão disponíveis nos apêndices H e J.

A primeira observação é que nos dois cenários, os melhores modelos da Tese possuem acurácia preditiva significativamente superior ( $p < 0,05$ ) aos melhores modelos *baseline*, nas janelas de 5, 10 e 20 segundos. Nestas janelas, em ambos os cenários, os melhores modelos da Tese apresentam valores superiores ou iguais aos melhores modelos *baseline*, para todas as métricas, com exceção da métrica de *recall*. Nas janelas de 40, 60, 90, 120, 180 e 240 segundos, em ambos os cenários, não foi encontrada diferença estatística significativa na acurácia preditiva das abordagens. Na janela de 360 segundos, uma diferença significativa é novamente encontrada, em ambos os cenários, mas com uma variação nos cenários. No cenário 1, a acurácia do melhor modelo *baseline* é superior, enquanto no cenário 2, a acurácia do melhor modelo da Tese é superior. Os melhores resultados de cada métrica são destacados em negrito.

Os resultados da Tabela 30 apresentam evidências de que a abordagem da Tese supera a abordagem *baseline* de forma significativa e consistente (em ambos os cenários), na generalização para estudantes com características heterogêneas, apenas nas janelas de 5, 10 e 20 segundos. Nas demais janelas, nenhuma abordagem sobressaiu sobre a outra de forma significativa e em ambos os cenários.

A Tabela 31 apresenta a média dos valores obtidos nos dois cenários de treino e teste (treino com amostras do ensino técnico e teste com amostras do ensino superior, e treino com amostras do ensino superior e teste com amostras do ensino técnico) para cada métrica, cada janela de tempo fixo e cada abordagem (Tese e *baseline*). Os resultados são mostrados apenas para os modelos que apresentaram melhor média de acurácia em cada janela de tempo fixo e cada abordagem. Os resultados completos (não apenas dos melhores modelos), são disponibilizados no apêndice L. Os valores mais altos, em todas as métricas, foram encontrados nos modelos da Tese, nas janelas de 5 e 10 segundos (destacados em negrito). Nas janelas de 5, 10 e 20 segundos, todos os valores encontrados para os melhores modelos da Tese são superiores aos valores dos melhores modelos *baseline*. A partir da janela de 40 segundos, há uma variação no predomínio dos modelos de cada abordagem, com algumas métricas obtendo valores superiores nos melhores modelos da Tese, e em outras, obtendo valores superiores nos

modelos *baseline*. Estes resultados reforçam que os desempenhos de generalização mais consistentes são encontrados para os modelos da Tese, em janelas de observação de tempos menores, de 5, 10 e 20 segundos. Os algoritmos com mais ocorrências nos melhores modelos foram, respectivamente, *Logistic Regression* (6 ocorrências), *Stochastic Gradient Descent* (5 ocorrências) e *Multilayer Perceptron* (4 ocorrências), *XGBoost* (2 ocorrências), *AdaBoost* (2 ocorrências) e *Random Forest* (1 ocorrência).

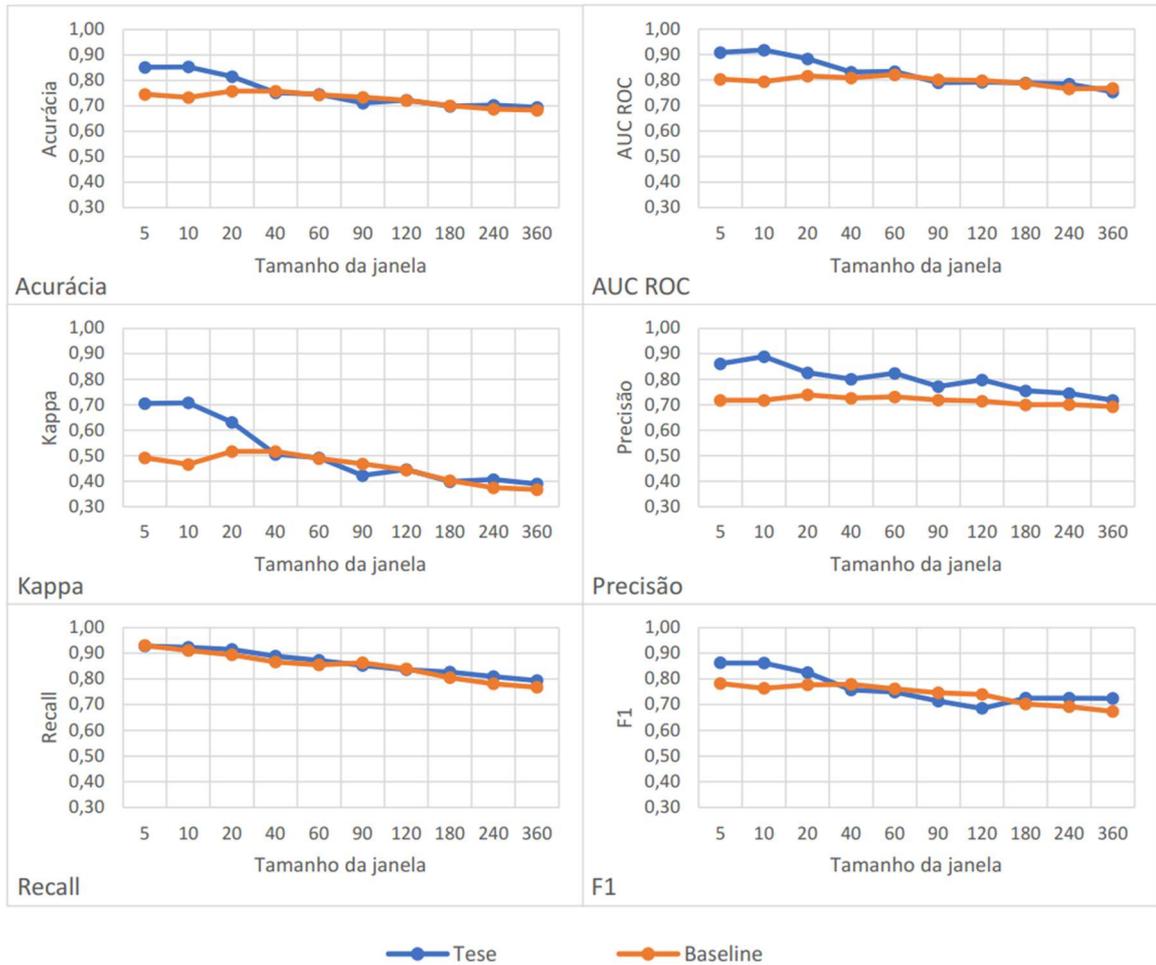
Tabela 31 – Desempenho dos modelos de janela fixa que obtiveram melhor média de acurácia nos dois cenários

<i>Jan.</i>		<i>Algor.</i>	<i>Acur.</i>	<i>AUC ROC</i>	<i>Kappa</i>	<i>Prec.</i>	<i>Recall</i>	<i>F1</i>
5	Tese	SGD	,85239	,90843	,70478	,86124	<b>,92522</b>	<b>,86215</b>
	Base	LR	,74582	,80367	,49164	,71778	,91516	,78198
10	Tese	SGD	<b>,85401</b>	<b>,91771</b>	<b>,70803</b>	<b>,88931</b>	,91455	,86119
	Base	ADB	,73301	,79394	,46602	,71694	,88629	,76367
20	Tese	SGD	,81574	,88382	,63149	,82594	,86599	,82433
	Base	LR	,75854	,81602	,51709	,73906	,85556	,77729
40	Tese	XGB	,75257	,83111	,50515	,80054	,79459	,75728
	Base	LR	,75832	,80918	,51665	,72583	,86869	,77880
60	Tese	LR	,74581	,83334	,49163	,82323	,83336	,74880
	Base	MLP	,74421	,82144	,48843	,73068	,83450	,76120
90	Tese	MLP	,71143	,79045	,42287	,77102	,90581	,71363
	Base	LR	,73401	,80148	,46803	,71798	,82925	,74541
120	Tese	ADB	,72319	,79170	,44639	,79704	,82879	,68538
	Base	XGB	,72205	,79806	,44411	,71383	,80711	,73906
180	Tese	MLP	,69927	,78874	,39854	,75437	,84008	,72555
	Base	LR	,70148	,78732	,40296	,69960	,73642	,70239
240	Tese	SGD	,70347	,78452	,40694	,74422	,78195	,72473
	Base	RF	,68717	,76584	,37434	,70107	,72363	,69212
360	Tese	SGD	,69491	,75393	,38983	,71748	,80314	,72447
	Base	MLP	,68375	,76792	,36750	,69267	,76055	,67350

Fonte: Elaborado pelo autor.

O Gráfico 6 mostra o desempenho dos modelos na generalização dos resultados. São mostrados os modelos com melhor média de acurácia nos cenários, nas diferentes configurações de janelas fixas. Observa-se que nas métricas de acurácia, AUC ROC, *kappa*, precisão e F1, os modelos da Tese sobressaem aos modelos *baseline* nas janelas até 20 segundos. O gráfico também mostra que não há diferença evidente entre as abordagens na métrica de *recall*.

Gráfico 6 – Desempenho dos modelos de janela fixa que obtiveram melhor média de acurácia nos cenários



Fonte: Elaborado pelo autor.

A Tabela 32 apresenta os resultados dos melhores modelos nos dois cenários de treino e teste, quando é incluída a etapa de seleção de atributos de dados. Os resultados de todos os modelos gerados (não apenas os melhores), são disponibilizados nos apêndices I e K. Os resultados mostram desempenhos similares aos apresentados quando os modelos foram gerados sem a etapa de seleção de atributos. A abordagem da Tese apresentou acurácia significativamente superior e mais consistente (em ambos os cenários), em relação à abordagem *baseline*, nas janelas de 5, 10 e 20 segundos. Nas demais janelas, nenhuma abordagem sobressaiu sobre a outra de forma significativa e em ambos os cenários. Os melhores resultados para cada métrica são destacados em negrito. Não foram aplicados testes estatísticos para comparação da acurácia preditiva dos modelos sem e com seleção de atributos. Os ganhos com a aplicação da técnica de seleção de atributos de dados foi de, no máximo, 2 casas decimais. As métricas que mais se beneficiaram foram a *precisão* e o *recall*.

Tabela 32 – Modelos de janela fixa com melhor generalização com seleção de atributos

Jan.	Algor.	Acur.	AUC ROC	Kappa	Prec.	Recall	F1	Teste McNemar		
								chi <sup>2</sup>	p	
<i>Treinamentos com o ensino técnico e testes com o ensino superior</i>										
5	Tese	SGD	<b>,84104</b>	,89747	<b>,68209</b>	,86186	,93159	<b>,84923</b>	108,01606	<b>&lt;,00001</b>
	Base	ADB	,67505	,72310	,35010	,62146	<b>,96177</b>	,73632		
10	Tese	SGD	,83467	<b>,90030</b>	,66935	<b>,87469</b>	,88104	,83632	82,30115	<b>&lt;,00001</b>
	Base	ADB	,68649	,72807	,37298	,63135	,91733	,74104		
20	Tese	MLP	,77530	,84718	,55060	,82178	,86639	,79406	19,74299	<b>&lt;,00001</b>
	Base	LR	,70850	,77071	,41700	,65594	,90890	,75257		
40	Tese	SGD	,75102	,81533	,50204	,80327	,75409	,74213	0,62592	0,42885
	Base	kNN	,73668	,78487	,47336	,69411	,90778	,77208		
60	Tese	SVM	,75206	,80213	,50413	,77915	,80991	,76562	1,94011	0,16365
	Base	RF	,73243	,77928	,46487	,67550	,89462	,76977		
90	Tese	LR	,75317	,81277	,50635	,78163	,78813	,74756	1,25	0,26355
	Base	MLP	,73622	,78546	,47245	,68614	,90042	,76750		
120	Tese	RF	,70752	,78352	,41505	,80069	,81290	,72762	1,53554	0,21528
	Base	RF	,72795	,79633	,45591	,68531	,89032	,76296		
180	Tese	SGD	,72477	,81082	,44954	,75000	,87156	,75757	0,0	1,0
	Base	RF	,72591	,79931	,45183	,69505	,85779	,75100		
240	Tese	NB	,70410	,80480	,40821	,72295	,98309	,74666	1,27058	0,25965
	Base	MLP	,72705	,79048	,45410	,70434	,86956	,74655		
360	Tese	SGD	,67272	,77491	,34545	,71098	,94026	,73586	8,47058	<b>0,00360</b>
	Base	SGD	,72077	,78764	,44155	,70926	,96883	,75428		
<i>Treinamentos com o ensino superior e testes com o ensino técnico</i>										
5	Tese	ADB	,87446	,92973	,74892	,88850	,97424	,88335	13,06666	<b>0,00030</b>
	Base	SGD	,82832	,87170	,65665	,79772	,94420	,84534		
10	Tese	SGD	<b>,88300</b>	<b>,94401</b>	<b>,76600</b>	<b>,91503</b>	<b>,99337</b>	<b>,89292</b>	36,62307	<b>&lt;,00001</b>
	Base	NB	,80574	,87355	,61147	,79723	,92494	,82643		
20	Tese	SGD	,86777	,93118	,73555	,92951	,98888	,86994	16,33986	<b>0,00005</b>
	Base	LR	,81111	,87609	,62222	,82558	,85333	,81098		
40	Tese	LR	,81685	,88871	,63370	,84561	,99550	,83282	3,819209	0,050668
	Base	MLP	,78651	,84525	,57303	,76989	,83595	,79302		
60	Tese	ADB	,78068	,86949	,56136	,85099	,96818	,79707	0,0	1,0
	Base	MLP	,78181	,86191	,56363	,80845	,81363	,77197		
90	Tese	SGD	,75930	,83876	,51860	,81318	,98372	,79196	0,22085	0,63838
	Base	NB	,75116	,82101	,50232	,75990	,82325	,76789		
120	Tese	SGD	,75598	,84219	,51196	,83259	,98564	,77874	2,02808	0,15441
	Base	MLP	,73205	,80425	,46411	,74525	,72009	,72881		
180	Tese	MLP	,74107	,79686	,48214	,75100	,95153	,75334	5,98963	0,01439
	Base	NB	,69642	,77709	,39285	,73333	,75510	,71325		
240	Tese	SGD	,71288	,80454	,42577	,74902	,98039	,75211	0,87111	0,35064
	Base	SGD	,69187	,76281	,38375	,74552	,68347	,67217		
360	Tese	SGD	,71710	,77750	,43421	,73366	,97368	,72168	4,72321	<b>0,02975</b>
	Base	SGD	,67763	,75457	,35526	,73478	,61184	,65148		

Fonte: Elaborado pelo autor.

A Tabela 33 mostra os valores médios obtidos nos dois cenários de treino e teste, quando aplicada a etapa de seleção de atributos de dados. Os resultados são mostrados apenas para os modelos que apresentaram melhor média de acurácia. Os resultados de todos os modelos gerados (não apenas os melhores), foram disponibilizados no apêndice M. Novamente, os melhores resultados foram obtidos nas janelas de 5 e 10 segundos (destacados em negrito). Nas

janelas de 5, 10 e 20 segundos, a abordagem da Tese apresenta valores superiores nas métricas de acurácia, AUC ROC, *kappa*, precisão e F1.

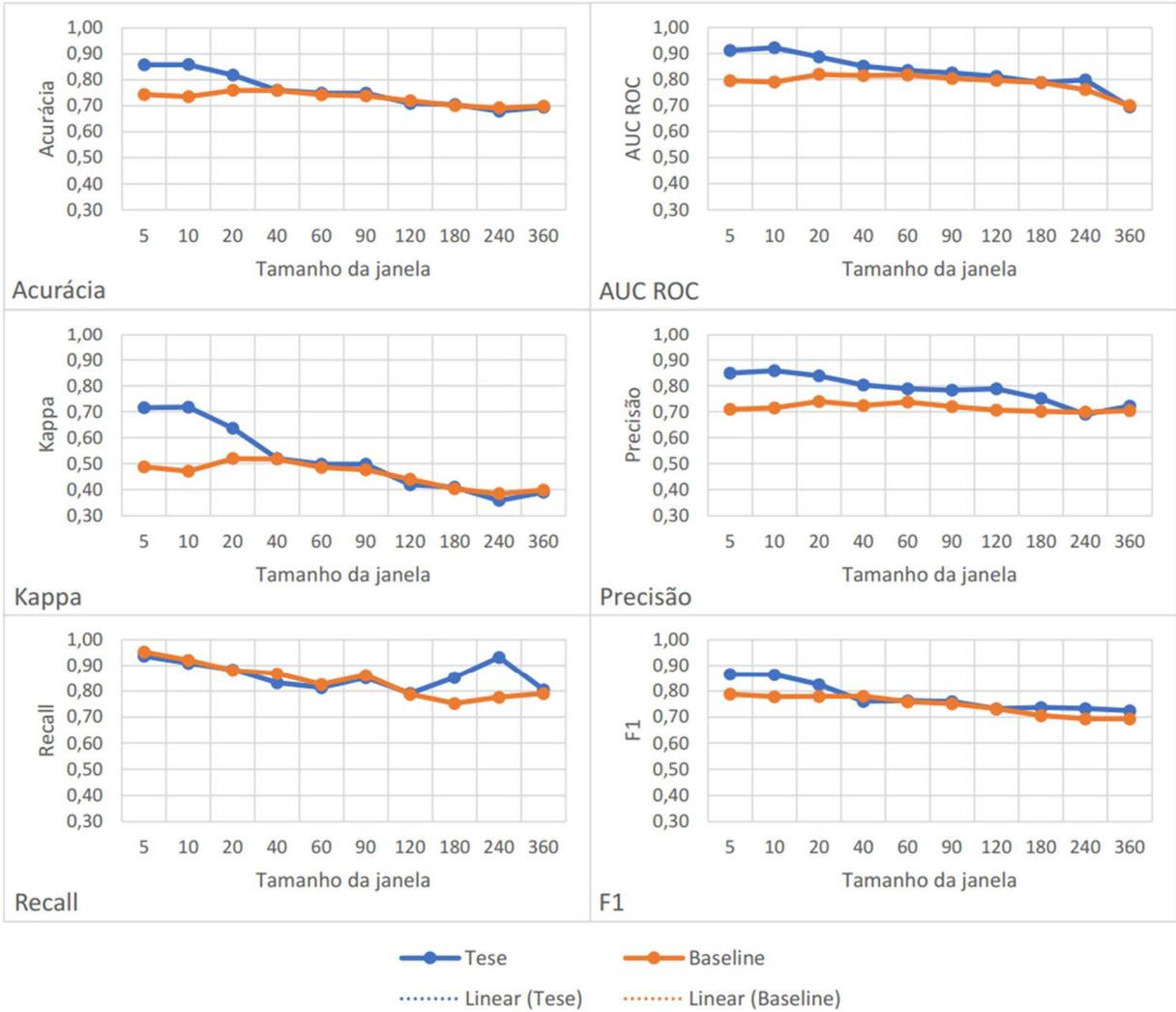
Tabela 33 – Desempenho dos modelos de janela fixa que obtiveram melhor média de acurácia nos cenários com seleção de atributos de dados

<i>Jan.</i>		<i>Algor.</i>	<i>Acur.</i>	<i>AUC ROC</i>	<i>Kappa</i>	<i>Prec.</i>	<i>Recall</i>	<i>F1</i>
5	Tese	SGD	,85775	,91061	,71551	,84906	,93742	<b>,86629</b>
	Base	SGD	,74414	,79456	,48828	,70923	<b>,95298</b>	,78839
10	Tese	SGD	<b>,85884</b>	<b>,92155</b>	<b>,71768</b>	<b>,85829</b>	,91023	,86462
	Base	ADB	,73563	,78940	,47126	,71422	,92114	,77838
20	Tese	SGD	,81850	,88647	,63700	,83798	,88430	,82493
	Base	LR	,75980	,81943	,51961	,73965	,88112	,77969
40	Tese	XGB	,76014	,85106	,52028	,80263	,83259	,76102
	Base	LR	,75883	,81426	,51767	,72423	,86850	,78000
60	Tese	ADB	,74932	,83452	,49865	,78849	,81312	,76289
	Base	MLP	,74266	,81686	,48533	,73728	,82624	,75842
90	Tese	SGD	,74935	,82473	,49870	,78323	,85203	,75961
	Base	LR	,73808	,80313	,47617	,71985	,86184	,75108
120	Tese	MLP	,70969	,81210	,41938	,78806	,79161	,73181
	Base	LR	,72002	,79629	,44004	,70670	,78752	,73205
180	Tese	LR	,70505	,78785	,41011	,75050	,85305	,73732
	Base	LR	,70154	,78820	,40309	,70135	,75255	,70605
240	Tese	ADB	,67929	,79839	,35859	,69029	,93286	,73338
	Base	RF	,69241	,76109	,38482	,69884	,77651	,69393
360	Tese	SGD	,69491	,69491	,38983	,72232	,80573	,72506
	Base	SGD	,69920	,69920	,39841	,70412	,79033	,69362

Fonte: Elaborado pelo autor.

O Gráfico 7 reforça o desempenho superior dos melhores modelos da Tese na generalização de seus resultados para estudantes com características heterogêneas, nas janelas de observação com tempos menores, de 5, 10 e 20 segundos, quando é aplicado o método de seleção dos melhores atributos de dados. Nas demais janelas, o desempenho de generalização de ambas as abordagens (Tese e *baseline*) é equiparado. O gráfico também mostra que na métrica de *recall*, nas janelas de 180, 240 e 360 segundos, os melhores modelos da Tese aparentemente se beneficiaram da etapa de seleção de atributos de dados.

Gráfico 7 - Desempenho dos modelos de janela fixa que obtiveram melhor média de acurácia nos cenários com seleção de atributos de dados



Fonte: Elaborado pelo autor.

Por fim, os próximos parágrafos descrevem os resultados sobre o poder de generalização dos modelos para alunos com características heterogêneas, quando as amostras refletem janelas de observação de tempo variável. A Tabela 34 mostra os resultados dos melhores modelos da Tese e *baseline* nos dois cenários de treino e teste, sem e com seleção de atributos (SA). Também são apresentados os resultados de testes estatísticos que comparam as abordagens. A tabela também apresenta os modelos que obtiveram melhor média de acurácia nos dois cenários de treino e teste. Os resultados de todos os modelos gerados (não apenas os melhores) foram disponibilizados no apêndice N. Os melhores resultados em cada cenário são destacados em negrito.

Tabela 34 - Modelos de janela variável com melhor desempenho de generalização

<i>SA</i>	<i>Algor.</i>	<i>Acur.</i>	<i>Kappa</i>	<i>Prec.</i>	<i>Recall</i>	<i>F1</i>	<i>Teste McNemar</i>
-----------	---------------	--------------	--------------	--------------	---------------	-----------	----------------------

				<i>AUC</i>				<i>chi</i> <sup>2</sup>	<i>p</i>	
				<i>ROC</i>						
<b><i>Treinamentos com o ensino técnico e testes com o ensino superior</i></b>										
Não	Tese	DT	,78269	,88483	,56539	<b>,96734</b>	,89134	,80399	8,62676	<b>0,00331</b>
	Base	SGD	,81891	,90253	,63782	,78253	<b>,94969</b>	,82986		
Sim	Tese	RF	,82293	<b>,90848</b>	,64587	,81910	,93360	,84000	1,26041	0,26157
	Base	MLP	<b>,83501</b>	,90591	<b>,67002</b>	,79639	,91549	<b>,84730</b>		
<b><i>Treinamentos com o ensino superior e testes com o ensino técnico</i></b>										
Não	Tese	ADB	,78970	,90072	,57939	,89795	,98283	,77367	2,08333	0,14891
	Base	RF	,77253	,89928	,54506	,89772	,62017	,73164		
Sim	Tese	ADB	,75321	<b>,90384</b>	,50643	,88659	<b>,98712</b>	,77924	30,69473	3,01983
	Base	DT	<b>,81223</b>	,89909	<b>,62446</b>	<b>,90000</b>	,71244	<b>,79141</b>		
<b><i>Modelos que obtiveram melhor média de acurácia nos dois cenários</i></b>										
Não	Tese	ADB	,76205	,85562	,52410	<b>,89245</b>	,90489	,74876	-	-
	Base	LR	,77986	,88889	,55972	,83785	,76669	,76134		
Sim	Tese	ADB	,78405	,89013	,56810	,82638	<b>,91106</b>	,79122	-	-
	Base	DT	<b>,79997</b>	<b>,89874</b>	<b>,59995</b>	,83939	,81296	<b>,80142</b>		

Fonte: Elaborado pelo autor.

Os resultados apresentados na Tabela 34 mostram que o único teste estatístico que apresentou diferença significativa de acurácia preditiva entre as abordagens da Tese e *baseline* foi observado no cenário 1 (treinamento com o ensino técnico e teste com o ensino superior), sem seleção de atributos (SA). Neste cenário, os valores das métricas de acurácia, AUC ROC, *kappa*, *recall* e F1 são superiores no melhor modelo *baseline*. Quando observamos os modelos que obtiveram melhor média de acurácia em ambos os cenários, os melhores modelos *baseline* também são superiores nas métricas de acurácia, AUC ROC, *kappa* e F1. No entanto, os melhores modelos da Tese possuem valores médios superiores aos melhores modelos *baseline* na métrica de *recall* e precisão. Os resultados apresentados na Tabela 34 mostram que não é possível identificar uma clara vantagem de nenhuma das abordagens (Tese e *baseline*) na generalização dos modelos para estudantes com características heterogêneas, quando são usadas amostras que refletem janelas de observação de tempo variável.

## 11 CONCLUSÃO

O trabalho desenvolvido pela Tese buscou verificar os efeitos de utilizar dados sobre estimativas do conhecimento do aluno em modelos de aprendizado de máquina livres de sensores, desenvolvidos para detectar a confusão do estudante em amostras de dados que refletem o comportamento do aluno em tarefas de programação de computadores. A hipótese levantada é que integrar estes dados sobre estimativas do conhecimento do aluno a dados de interação do estudante com o ambiente (mouse, teclado, erros de codificação, entre outros) pode melhorar o desempenho destes detectores de confusão, em relação aos modelos que utilizam apenas os dados de interação com o ambiente. A hipótese é justificada por teorias de emoções, que relacionam a confusão a um componente cognitivo de avaliação (*appraisal*). Para estas teorias, a confusão surge quando há uma avaliação sobre a falta de correspondência entre a informação que chega, como o enunciado de um exercício de programação ou uma situação enfrentada pelo aluno durante a codificação, e o modelo mental de conhecimento e experiências anteriores do estudante (D'MELLO; GRAESSER, 2014b; SILVIA, 2010). A abordagem apresentada na hipótese implementa uma sugestão discutida na literatura (D'MELLO; GRAESSER, 2014b) sobre a geração de modelos de detecção de confusão livres de sensores que usam uma abordagem mista guiada à teoria e dirigida a dados, ou seja, que utilizam amostras compostas por atributos de dados justificados por alguma teoria (abordagem guiada à teoria) e outros atributos não justificados por qualquer teoria (abordagem dos modelos dirigidos a dados).

Uma revisão da literatura encontrou apenas três trabalhos relacionados ao tema da Tese que avaliaram modelos livres de sensores na aprendizagem de programação de computadores (FELIPE et al., 2012; LEE et al., 2011; VEA; RODRIGO, 2017a). Estes trabalhos usaram a abordagem dirigida a dados, na qual usaram dados de interação (mouse, teclado, erros de código, compilação de código, entre outros) para gerar modelos de detecção de confusão. Um dos trabalhos não encontrou resultados tão fortes de desempenho (*kappa* de 0,572 e precisão de 0,74) (VEA; RODRIGO, 2017a). Os outros dois trabalhos (FELIPE et al., 2012; LEE et al., 2011) encontraram bons resultados para a métrica de *kappa* (0,856 e 0,860), mas apresentaram limitações importantes: tempo bastante reduzido de observação dos estudantes para formação das amostras (FELIPE et al., 2012); poucos estudantes participantes (FELIPE et al., 2012); falta de verificação do poder de generalização dos seus resultados para alunos com características heterogêneas (FELIPE et al., 2012; LEE et al., 2011); e utilização de uma modalidade de dados que não permite a identificação da confusão durante a codificação (dados sobre compilação,

que são gerados apenas após a codificação) (LEE et al., 2011). A quantidade de trabalhos relacionados no contexto de aprendizagem de programação de computadores sugere que a detecção de confusão livre de sensores neste contexto tem sido pouco explorada na literatura.

Para verificar a sua hipótese, a Tese comparou os resultados da abordagem apresentada na sua hipótese com a abordagem dos trabalhos relacionados, dirigidos a dados (*baseline*). O trabalho também buscou verificar se a abordagem da Tese possibilita aos modelos generalizar bem seus resultados para estudantes com características heterogêneas, em comparação com a abordagem *baseline*. Além disso, a hipótese da Tese foi verificada quando foram utilizadas diferentes configurações de janelas de dados de observação do comportamento dos alunos. Assim, a Tese apresenta evidências sobre os efeitos da abordagem da hipótese e da abordagem *baseline* em diferentes janelas de observação do comportamento dos alunos.

O primeiro objetivo específico da Tese foi verificar se modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente possuem desempenho superior a modelos que utilizam apenas dados sobre interação com o ambiente, quando são consideradas diferentes janelas de observação de tamanho fixo. Este objetivo também apresenta respostas para a seguinte subquestão de pesquisa: qual o desempenho dos modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente e dos modelos que utilizam apenas dados sobre interação com o ambiente, em diferentes janelas de observação de tamanho fixo?

Os resultados mostraram que para todas as janelas de observação de tempo fixo (5, 10, 20, 40, 60, 90, 120, 180, 240 e 360 segundos), a diferença de acurácia preditiva entre melhores modelos da Tese e os melhores modelos *baseline* foi significativa. Em todas estas janelas, os valores de desempenho geral (acurácia, AUC ROC e *kappa*) dos melhores modelos da Tese foram superiores aos melhores modelos *baseline*. Os melhores modelos da Tese também obtiveram valores superiores nas métricas de precisão e F1. Na métrica de *recall*, aparentemente nenhuma das abordagens sobressaiu sobre a outra, com algumas janelas apresentando valores superiores para uma abordagem, e outras janelas apresentando valores superiores para a outra abordagem. Aparentemente, os modelos de ambas as abordagens não se beneficiaram de forma significativa com a aplicação do método de seleção de atributos de dados.

Outra observação é que, com exceção do *recall*, em todas as demais métricas, o desempenho superior dos melhores modelos da Tese foi mais evidente em janelas de observação de 5, 10 e 20 segundos. Nestas janelas, foram obtidos os melhores resultados apresentados pelo trabalho da Tese nas métricas de acurácia (**0,89725**, na janela de 10

segundos), AUC ROC (**0,95082**, na janela de 10 segundos), *kappa* (**0,79452**, na janela de 10 segundos) e F1 (**0,89992**, na janela de 10 segundos), ambos resultados de melhores modelos na abordagem da Tese. Para a métrica de *recall*, o melhor resultado foi encontrado para um melhor modelo da abordagem *baseline* (**0,92938** na janela de 5 segundos), com seleção de atributos de dados. Outro resultado interessante é que os algoritmos que compuseram os melhores modelos com janelas de observação de tempo fixo foram os algoritmos *XGBoost* e o *Random Forest*.

O segundo objetivo específico da Tese foi verificar se modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente possuem desempenho superior a modelos que utilizam apenas dados sobre interação com o ambiente, quando é considerada uma janela de observação de tamanho variável. Este objetivo também apresenta respostas para a seguinte subquestão de pesquisa: qual o desempenho dos modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente e dos modelos que utilizam apenas dados sobre interação com o ambiente, em uma janela de observação de tamanho variável?

Os resultados mostraram que a diferença de acurácia preditiva entre o melhor modelo da Tese e o melhor modelo *baseline* foi significativa quando as amostras foram compostas por atributos de dados que refletiam observações de tempo variável. Como os valores de desempenho geral (acurácia, AUC ROC e *kappa*) foram superiores nos melhores modelos da Tese, é possível sugerir que, em relação ao desempenho geral, os melhores modelos da Tese superaram o desempenho dos melhores modelos *baseline*. No entanto, observou-se que, com janelas de tempo variável, os melhores modelos da Tese não se beneficiaram tanto em relação ao uso de janelas de tempo fixo, pois apresentaram valores inferiores aos melhores modelos com janelas fixas. Já os modelos com a abordagem *baseline* parecem ter se beneficiado mais do cenário de janelas variáveis. Os melhores resultados encontrados na Tese para a abordagem *baseline* foram encontrados no cenário de janelas variáveis, para as métricas de acurácia (**0,86762**), AUC ROC (**0,93583**), *kappa* (**0,73526**), precisão (**0,90226**) e F1 (**0,87169**). O melhor valor de precisão encontrado no trabalho da Tese foi obtido para o melhor modelo *baseline* com janela de tempo variável. Em todos os melhores modelos com janelas variáveis (Tese e *baseline*, com e sem seleção de atributos), o algoritmo utilizado foi o *XGBoost*, o que reforça o bom desempenho deste algoritmo no contexto da Tese.

O trabalho da Tese também verificou a relevância dos atributos de dados nas diferentes configurações de janela de tempo fixo e tempo variável, nas abordagens da Tese e *baseline*. Foi encontrado que, entre os dez atributos de dados mais relevantes na abordagem da Tese, cinco

foram atributos sobre estimativas de conhecimento do estudante e os outros cinco foram atributos sobre a interação do aluno com o ambiente. Os dez atributos mais relevantes encontrados na abordagem da Tese, na ordem de relevância, foram os seguintes: 1) **quantidade de teclas digitadas**; 2) **média das probabilidades de domínio do aluno nos componentes de conhecimento do objetivo em codificação**; 3) **quantidade de intervalos de digitação**; 4) **tempo total de movimentos de mouse**; 5) **probabilidade de domínio do aluno no componente de conhecimento primário do objetivo em codificação**; 6) **média das probabilidades de o aluno aplicar corretamente os componentes de conhecimento do exercício em codificação**; 7) **quantidade de movimentos de mouse**; 8) **quantidade de erros de sintaxe na codificação**; 9) **média das probabilidades de domínio do aluno nos componentes de conhecimento do exercício em codificação**; e 10) **a probabilidade de o aluno aplicar corretamente o componente de conhecimento primário do objetivo em codificação**. Outro resultado interessante é que o atributo sobre quantidade de teclas digitadas foi o mais relevante tanto nos melhores modelos da Tese quanto nos melhores modelos *baseline*.

O terceiro objetivo específico da Tese foi verificar se modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente generalizam bem seus resultados para estudantes com características heterogêneas, comparado a modelos que utilizam apenas dados sobre interação com o ambiente. Este objetivo também apresenta respostas para a seguinte subquestão de pesquisa: qual o desempenho dos modelos de detecção de confusão livres de sensores que utilizam dados sobre estimativas de conhecimento do aluno + dados sobre interação com o ambiente e dos modelos que utilizam apenas dados sobre interação com ambiente na generalização de seus resultados para estudantes com características heterogêneas?

Quando os modelos foram treinados usando apenas as amostras dos alunos de um nível de ensino e testados usando as amostras dos alunos de outro nível de ensino, foi encontrada diferença significativa de acurácia preditiva entre os modelos da Tese e *baseline* nas janelas de 5, 10 e 20 segundos, no cenário 1 (treino com as amostras do ensino técnico e teste com as amostras do ensino superior) e também no cenário 2 (treino com as amostras do ensino superior e teste com as amostras do ensino técnico). Nestas janelas, os valores para as métricas de desempenho geral acurácia, AUC ROC e *kappa*, e as métricas de precisão e F1, foram superiores nas amostras da Tese, em ambos os cenários. Também foi encontrada uma diferença significativa na janela de 360 segundos, mas no cenário 1, o desempenho geral foi superior no melhor modelo *baseline*, enquanto no cenário 2, o desempenho geral foi superior no melhor

modelo da Tese. Nas demais janelas (40, 60, 90, 120, 180 e 240), não houve diferença significativa entre os melhores modelos. Dessa forma, os resultados foram mais consistentes nas janelas iniciais, de 5, 10 e 20 segundos, com desempenho geral superior nos melhores modelos da Tese.

Quando os modelos foram treinados com amostras compostas por atributos que refletem janelas de observação de tempo variável, foi encontrada uma diferença significativa de acurácia preditiva apenas quando os modelos foram treinados com as amostras do ensino técnico. Neste cenário, o desempenho geral foi superior nos melhores modelos *baseline*, com exceção da métrica de precisão, que foi superior na abordagem da Tese. Nos demais cenários de verificação dos resultados de generalização para estudantes com características heterogêneas, com janelas variáveis, não fica claro qual abordagem (Tese ou *baseline*) sobressai sobre a outra, em função de haver muita variação nos resultados das métricas.

Não é intenção da Tese fazer comparações estatísticas diretas de seus resultados com os resultados dos trabalhos relacionados, pois as amostras utilizadas nos trabalhos relacionados foram diferentes das utilizadas na Tese. A Tese gerou modelos que representam a abordagem conceitual dos trabalhos relacionados e a abordagem apresentada em sua hipótese e comparou os resultados das abordagens. No entanto, é possível fazer uma comparação indireta dos resultados apresentados pela Tese e pelos trabalhos relacionados. O Quadro 2 (Seção 3.4.2) descreveu os sete trabalhos relacionados que apresentaram modelos de detecção de confusão em atividades de programação e computadores, sendo três deles mais alinhados com a Tese (modelos livres de sensores). A comparação aqui é feita com os três trabalhos livres de sensores. Estes trabalhos limitaram seus resultados às métricas de acurácia, precisão e *kappa*. O melhor resultado de *kappa* dos trabalhos relacionados foi 0,860, um valor superior ao melhor valor encontrado na Tese (0,79452 para um modelo com a abordagem da hipótese da Tese). O melhor valor de precisão dos trabalhos relacionados foi de 0,74 (um único trabalho que apresentou resultado nesta métrica). Na Tese, este resultado foi superado. O melhor resultado de precisão encontrado na Tese foi para um modelo *baseline* (0,90226), enquanto o melhor resultado de precisão para um modelo com a abordagem da Tese foi de 0,87981. O melhor resultado de acurácia dos trabalhos relacionados foi 0,77 (um único trabalho apresentou resultado nesta métrica), um valor inferior ao encontrado na Tese (0,89725 para um modelo com a abordagem da hipótese da Tese).

Observando todos os resultados apresentados na Tese, os resultados dos melhores modelos *baseline* superaram os resultados dos melhores modelos da Tese nas métricas de precisão (**0,90226** para o melhor modelo *baseline* e **0,87981** para o melhor modelo da Tese) e

*recall* (**0,92938** para o melhor modelo *baseline* e **0,92832** para o melhor modelo da Tese). No entanto, os resultados dos melhores modelos com a abordagem da Tese superaram os resultados dos melhores modelos da abordagem *baseline* nas métricas de acurácia (**0,89725** para o melhor modelo da Tese e **0,86762** para o melhor modelo *baseline*), AUC ROC (**0,95082** para o melhor modelo da Tese e **0,93583** para o melhor modelo *baseline*), *kappa* (**0,79452** para o melhor modelo da Tese e **0,73526** para o melhor modelo *baseline*) e F1 (**0,89992** para o melhor modelo da Tese e **0,87169** para o melhor modelo *baseline*). Além disso, para todas as janelas de tempo fixo (5, 10, 20, 40, 60, 90, 120, 180, 240 e 360) e variável, a diferença de acurácia preditiva foi significativa, segundo testes estatísticos. Em todos estes cenários de janelas fixas e variáveis com diferença significativa entre as abordagens, os valores para as métricas de desempenho geral (acurácia, AUC ROC e *kappa*) e F1, foram superiores para os melhores modelos da Tese. Na verificação de generalização dos resultados dos modelos para alunos com características heterogêneas, a diferença entre as abordagens só foi significativa e consistente (em ambos os cenários de treinos e testes dos modelos) nas configurações de janelas de 5, 10 e 20 segundos, quando os valores de desempenho geral e F1 observados também foram superiores para os modelos da abordagem da Tese. Apenas nas métricas de precisão e *recall*, houve uma variação de configurações de janelas onde uma e outra abordagem sobressaiu sobre a outra.

O trabalho da Tese apresenta evidências significativas para aceitar a hipótese de que a utilização de dados sobre estimativas de conhecimento do aluno no domínio de programação de computadores, juntamente de dados sobre a interação do estudante com o ambiente computacional de aprendizagem de programação, pode melhorar o desempenho dos modelos livres de sensores na detecção da confusão do aluno nas tarefas de aprendizagem de programação de computadores, quando comparado a utilizar apenas os dados de interação do estudante com o ambiente de aprendizagem. Os resultados também apresentam uma resposta para a principal questão de pesquisa levantada pela Tese, de que modelos livres de sensores que utilizam uma abordagem mista guiada à teoria e dirigida a dados, podem apresentar um desempenho superior a modelos livres de sensores que utilizam uma abordagem dirigida a dados, na detecção da confusão do aluno durante tarefas de aprendizagem de programação.

As evidências apresentadas pela Tese sugerem que ambientes computacionais adaptativos de aprendizagem de programação, como sistemas tutores inteligentes no domínio de programação, que utilizam informações inferidas pelo sistema para adaptar suas decisões instrucionais e estratégias pedagógicas, e que mantém um modelo de estimativas de conhecimento do aluno, como o modelo BKT, podem se beneficiar da abordagem da Tese para a detecção da confusão do aluno. Os resultados da Tese ainda indicam que os cenários onde a

abordagem da Tese é mais vantajosa é quando os dados dos alunos usados no modelo de detecção de confusão representam observações de comportamento sobre intervalos de tempo de 5, 10 e 20 segundos.

Os resultados apresentados pela Tese para os modelos livres de sensores que utilizam somente dados de interação com o ambiente também são uma contribuição para as pesquisas em computação afetiva. Os resultados da Tese para os melhores modelos *baseline* não podem ser considerados ruins, como uma acurácia de 0,86762, superior ao resultado do único trabalho relacionado que apresentou resultado para esta métrica. O melhor modelo *baseline* também apresentou um valor de *kappa* de 0,73526, uma concordância superior à esperada para juízes humanos (0,6) (OCUMPAUGH; BAKER; OCUMPAUGH, J., & BAKER, 2014). Ambientes computacionais de aprendizagem que não mantenham um modelo probabilístico de conhecimento do aluno ou quando não há dados suficientes para o modelo inferir estimativas sobre o conhecimento do estudante (problema do início frio<sup>1</sup>), podem ainda se beneficiar de dados de interação na detecção da confusão, quando usada uma janela de observação de tamanho variável (cenário dos melhores resultados da abordagem *baseline*).

Outro resultado interessante é que o algoritmo de aprendizado de máquina que mais se destacou, ou seja, o que mais ocorreu entre os melhores modelos da Tese e *baseline*, nas configurações de janelas fixas e janela variável, foi o algoritmo *XGBoost* (*eXtreme Gradient Boosting*) (CHEN; GUESTRIN, 2016). Este algoritmo tem apresentado resultados promissores nos últimos anos, tendo seu desempenho destacado em competições de aprendizado de máquina (ROUSSEAU; USTYUZHANIN, 2022; ZHANG; ZHAO; LI, 2019). O *XGBoost* é baseado em técnicas de *Gradient Tree Boosting* e adiciona um mecanismo de regularização na sua função objetivo que controla a complexidade do modelo e reduz a probabilidade de sobreajuste (CHEN; GUESTRIN, 2016). O segundo algoritmo que mais se destacou (mais ocorrências) entre os melhores modelos foi o *Random Forest* (BREIMAN, 2001). O algoritmo gera diversas árvores de decisão que tomam decisões individuais sobre a classe de saída. A decisão final do algoritmo será a classe de saída que mais ocorreu (a mais “votada”) nas diversas árvores de decisão (BREIMAN, 2001). Algoritmos como o *XGBoost* e o *Random Forest*, que utilizam conjuntos de árvores de decisão, além de outras características individuais, são os algoritmos mais recomendados em problemas que envolvem dados tabulares (SHWARTZ-ZIV; ARMON, 2022), como o problema abordado na Tese. Os resultados da Tese sugerem a implementadores

---

<sup>1</sup> O problema de início frio em modelos probabilísticos de conhecimento do aluno, como os modelos BKT, ocorre quando o modelo não possui dados suficientes para gerar estimativas sobre o conhecimento do estudante, como quando a interação do aluno com o ambiente é ainda incipiente (DAS et al., 2021).

de modelos de detecção de confusão no contexto da Tese que considerem a utilização destes algoritmos.

Além das evidências apresentadas pela Tese para sua hipótese e para suas questões de pesquisa, a Tese apresentou outras contribuições interessantes: o ambiente de programação de computadores adaptado pela Tese (descrito no Capítulo 6) e a descrição do método usado pela Tese para a coleta de dados e geração de amostras para modelos de aprendizado de máquina (Capítulo 8) podem ser úteis para outros pesquisadores e implementadores de modelos de detecção de confusão no contexto da Tese; o modelo de inferência do conhecimento do aluno implementado pela Tese (descrito no Capítulo 7), apesar de não apresentar uma contribuição original para a literatura, apresentou resultados similares aos obtidos na literatura para modelos BKT, e reforça as evidências já existentes para este tipo de modelo, no domínio de programação de computadores.

O trabalho desenvolvido na Tese também resultou em contribuições através da publicação de dois artigos. O primeiro foi publicado na conferência EC-TEL 2019 (*European Conference on Technology-Enhanced Learning*) (MORAIS et al., 2019). O artigo, desenvolvido em conjunto com um colega de pesquisa, Felipe de Moraes, descreve o *EmAP-ML (Emotions Annotation Protocol for Machine Learning)*, um protocolo para a obtenção de rótulos de emoções para modelos de aprendizado de máquina (descrito na Seção 8.2.1). Como descrito na Seção 8.2.2, este método seria utilizado para a obtenção dos rótulos de confusão dos modelos da Tese e *baseline*, mas acabou não sendo utilizado, em função da realização da coleta de dados no regime de aulas remotas, decorrente da declaração de pandemia de COVID-19. O segundo artigo, publicado no evento SBIE 2020 (Simpósio Brasileiro de Informática na Educação) (KAUTZMANN; JAQUES, 2020) apresentou um modelo de identificação de componentes de conhecimento de programação em processo de aplicação durante a codificação. Como descrito na Seção 6.3, o mecanismo deste modelo também acabou não sendo utilizado no trabalho da Tese, em função da realização da coleta de dados no regime de aulas remotas. Os dois trabalhos publicados são contribuições interessantes da Tese. A expectativa do autor é que novos artigos sejam publicados com os resultados da Tese.

### **11.1 Limitações e trabalhos futuros**

Os resultados dos modelos da Tese poderiam ter sido ainda melhores em detectar a confusão dos alunos, com um melhor ajuste do modelo de conhecimento do aluno para alguns componentes de conhecimento. Alguns componentes como “Repetição” e “Escrita de array”,

apresentaram acurácia inferior a 0,7 e erro (RMSE) superior a 0,4 (ver Seção 7.5.1). Para estes componentes com desempenho não tão satisfatório na inferência das suas estimativas, os modelos de detecção da Tese podem ter tido mais dificuldade em diferenciar as classes de “confusão” e “não confusão”. Uma possibilidade seria aumentar a granularidade desses componentes, ou seja, derivar outros componentes de conhecimento mais específicos, como “Repetição com quantidade pré-definida de iterações” e “Repetição com quantidade indefinida de iterações”. No entanto, é conhecida a dificuldade de gerar modelos de conhecimento mais precisos em domínios de aprendizagem complexa, como a de programação de computadores. Um exemplo de dificuldade encontrada é que quanto menor for a granularidade dos componentes, maior é a demanda por oportunidades de aplicação suficientes desses conhecimentos nas tarefas, para gerar suas estimativas. Isso nem sempre é viável de ser aplicado na prática (PELÁNEK, 2017b).

Outra possível limitação do trabalho foi identificada no método de verificação sobre a generalização dos resultados para estudantes com características heterogêneas (Seção 10.1). O trabalho criou dois grupos de estudantes participantes (ensino superior e ensino técnico) e assumiu que os grupos possuíam características heterogêneas. Apesar de haver uma diferença significativa de idade e experiências anteriores de programação de computadores entre os grupos, não houve diferença significativa em relação à percepção emocional dos estudantes (ver Capítulo 5). Mesmo que, em média, os alunos participantes do ensino superior possuem mais experiência anterior que os alunos do ensino técnico, foi identificado que alguns alunos do ensino superior possuíam a mesma experiência média em programação dos alunos do ensino técnico. Talvez fosse uma abordagem mais interessante agrupar os estudantes pela experiência anterior de programação, do que pelo nível de ensino (técnico e superior), utilizando algum método de agrupamento.

Apesar dos resultados interessantes encontrados pela Tese tanto nos modelos que refletem a hipótese da Tese, quanto nos modelos *baseline*, com algoritmos de aprendizado de máquina, uma sugestão para trabalhos futuros é verificar o desempenho de algoritmos de *deep learning* (aprendizagem profunda) no contexto da Tese. O trabalho de Botelho, Baker e Heffernan (BOTELHO; BAKER; HEFFERNAN, 2017) usou diferentes redes neurais recorrentes (*Recurrent Neural Network* (RNN), *Gated Recurrent Unit Neural Network* (GRU) e *Long-Short Term Memory Network* (LSTM)) e dados de interação dos alunos com o tutor de matemática *ASSISTments* para modelar um detector livre de sensores de quatro estados afetivos (concentração engajada, tédio, confusão e frustração). O melhor resultado encontrado foi um valor de 0,78 para a métrica AUC ROC, com uma rede do tipo LSTM. Este resultado foi

superior ao encontrado por dois outros trabalhos de referência que usaram a mesma base de dados do sistema ASSISTments (WANG; HEFFERNAN; HEFFERNAN, 2015). Redes neurais recorrentes possuem estruturas que capturam informações recorrentes em dados de séries de tempo (WILLIAMS; ZIPSER, 1989). A suposição dos autores para o desempenho superior encontrado com estes modelos, no domínio de matemática, é que a detecção de estados afetivos poderia se beneficiar de modelos que observam a estrutura temporal dos dados de entrada (BOTELHO; BAKER; HEFFERNAN, 2017). No entanto, é possível fazer a seguinte reflexão sobre os resultados deste trabalho. O melhor resultado do trabalho de Botelho, Baker e Heffernan (BOTELHO; BAKER; HEFFERNAN, 2017), no domínio de matemática, considerando apenas dados de interação, foi um valor de 0,78 para AUC ROC. No trabalho da Tese, no domínio de programação de computadores, o melhor resultado para a abordagem *baseline* (apenas dados de interação) para AUC ROC foi de 0,93583 (janela variável), e para a abordagem da Tese, foi de 0,95082 (janela de 10 segundos). Apesar de parecer não haver tanto espaço de evolução, poderia ser interessante verificar o desempenho de redes neurais recorrentes no contexto da Tese.

Em ambientes de aprendizagem complexa, como a de programação de computadores, o aluno precisar ler enunciados de problemas, diagnosticar os problemas, acessar conhecimentos prévios no domínio, estabelecer relações de conceitos, fazer inferências e gerar explicações. Nestes cenários de aprendizagem, a confusão é mais propensa de ocorrer e talvez seja mais norma do que exceção (D’MELLO et al., 2014; LEHMANN; GRAESSER, 2015). As emoções possuem um papel significativo nas experiências de aprendizagem do aluno, mas o impacto delas na aprendizagem nem sempre é tão óbvio. A confusão é um exemplo de emoção que pode ter impacto positivo para a aprendizagem, mas quando não regulada, pode ter impacto negativo e desengajar o aluno (D’MELLO; GRAESSER, 2012a). Evidências sugerem que a confusão pode proporcionar oportunidades de aprendizagem sob certas condições: a confusão deve ser suficiente para manter o aluno engajado na tarefa de aprendizagem; o aprendiz deve ter conhecimento prévio suficiente e/ou habilidades para resolver a confusão, como habilidades de regulação de emoções e habilidades metacognitivas; a confusão ser resolvida pelo próprio estudante (ARGUEL et al., 2019b; D’MELLO; GRAESSER, 2014a). Nesse sentido, outra possibilidade de trabalho futuro é integrar o modelo da Tese a um ambiente computacional adaptativo de aprendizagem de programação de computadores. O ambiente poderia utilizar as saídas do modelo (detecção de confusão e não confusão) nas suas decisões instrucionais durante as tarefas de programação. O ambiente poderia apresentar *feedbacks* sobre o conteúdo e diminuir as lacunas de conhecimento do estudante, além de ajudar o aluno a gerenciar a sua

confusão, ensinando estratégias de regulação emocional e estratégias autorregulatórias (metacognitivas) (ARGUEL et al., 2019b).

É uma tarefa desafiadora desenvolver mecanismos que detectam precisamente as emoções do aluno, como a confusão, para que ambientes computacionais de aprendizagem possam tomar decisões instrucionais importantes e adaptar o ambiente, de forma a usar as emoções a favor da aprendizagem. O autor da Tese espera que os resultados positivos apresentados na Tese possam inspirar outros pesquisadores na implementação e avaliação de modelos de detecção de confusão em contextos similares ao da Tese.

## REFERÊNCIAS

- AGARWAL, D.; BAKER, R. Dynamic knowledge tracing through data driven recency weights. **27th International Conference on Computers in Education**, n. August, 2019.
- ALEVEN, V. et al. Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. **International Journal of ...**, v. 16, p. 101–130, 2006.
- ALVES, F. P.; JAQUES, P. **Um Ambiente Virtual com Feedback Personalizado para Apoio a Disciplinas de Programação**. 3 nov. 2014.
- ALZOUBI, O.; D'MELLO, S. K.; CALVO, R. A. Detecting Naturalistic Expressions of Nonbasic Affect Using Physiological Signals. **IEEE Transactions on Affective Computing**, v. 3, n. 3, p. 298–310, jul. 2012.
- ARGUEL, A. et al. Puzzle-solving activity as an indicator of epistemic confusion. **Frontiers in Psychology**, v. 10, n. JAN, p. 1–11, 2019a.
- ARGUEL, A. et al. Seeking optimal confusion: a review on epistemic emotion management in interactive digital learning environments. **Interactive Learning Environments**, v. 27, n. 2, p. 200–210, 2019b.
- ARNOLD, M. B. **Emotion and personality**. New York, NY, US: Columbia University Press, 1960.
- ARROYO, I. et al. **Emotion Sensors Go To School**. INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE IN EDUCATION. **Anais...**2009.
- AZEVEDO, R.; TAUB, M.; MUDRICK, N. V. Understanding and Reasoning about Real-Time Cognitive, Affective, and Metacognitive Processes to Foster Self-Regulation with Advanced Learning Technologies. **Handbook of Self-Regulation of Learning and Performance**, n. 11237, p. 254–270, 2019.
- BADRINATH, A.; WANG, F.; PARDOS, Z. pyBKT: An Accessible Python Library of Bayesian Knowledge Tracing Models. 1 maio 2021.
- BAKER, R. S. J. D. Data mining for education. **International Encyclopedia of Education**, v. 7, p. 112–118, 2010.
- BAKER, R. S. J. D. et al. Contextual slip and prediction of student performance after use of an intelligent tutor. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6075 LNCS, n. June, p. 52–63, 2010.
- BAKER, R. S. J. D. et al. Ensembling predictions of student knowledge within intelligent tutoring systems. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6787 LNCS, p. 13–24, 2011.

BAKER, R. S. J. D. et al. Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. **Proceedings of the 5th International Conference on Educational Data Mining**, p. 126–133, 2012.

BAKER, R. S. J. D.; CORBETT, A. T.; ALEVEN, V. More Accurate Student Modeling through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. **Intelligent Tutoring Systems**, p. 406–415, 2008.

BAKER, R. S. J. D.; OCUMPAUGH, J. Interaction-Based Affect Detection in Educational Software. In: **The Oxford Handbook of Affective Computing**. [s.l.] Oxford University Press, 2015. p. 233–245.

BARRETT, L. F. Solving the Emotion Paradox: Categorization and the Experience of Emotion. **Personality and Social Psychology Review**, v. 10, n. 1, p. 20–46, 21 fev. 2006.

BARRICK, M. R.; MOUNT, M. K. THE BIG FIVE PERSONALITY DIMENSIONS AND JOB PERFORMANCE: A META-ANALYSIS. **Personnel Psychology**, v. 44, n. 1, p. 1–26, mar. 1991.

BECK, J. E.; CHANG, K. M. Identifiability: A fundamental problem of student modeling. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 4511 LNCS, p. 137–146, 2007.

BIXLER, R.; D’MELLO, S. Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. **Proceedings of the 2013 international conference on Intelligent user interfaces - IUI ’13**, p. 225, 2013.

BLOOM, B. S. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. **Educational Researcher**, v. 13, n. 6, p. 4–16, 1984.

BOSCH, N. et al. Temporal generalizability of face-based affect detection in noisy classroom environments. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 9112, p. 44–53, 2015.

BOSCH, N. et al. Detecting student emotions in computer-enabled classrooms. **IJCAI International Joint Conference on Artificial Intelligence**, v. 2016- Janua, p. 4125–4129, 2016.

BOSCH, N. et al. **Students’ Verbalized Metacognition During Computerized Learning**. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. **Anais...**New York, NY, USA: ACM, 6 maio 2021.

BOSCH, N.; CHEN, Y.; D’MELLO, S. It’s written on your face: Detecting affective states from facial expressions while learning computer programming. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 8474 LNCS, p. 39–44, 2014.

BOSCH, N.; D’MELLO, S. The Affective Experience of Novice Computer Programmers. **International Journal of Artificial Intelligence in Education**, v. 27, n. 1, p. 181–206, 2017.

BOSCH, N.; D’MELLO, S.; MILLS, C. What emotions do novices experience during their first computer programming learning session? **Lecture Notes in Computer Science**

(including subseries **Lecture Notes in Artificial Intelligence** and **Lecture Notes in Bioinformatics**), v. 7926 LNAI, p. 11–20, 2013.

BOTELHO, A. F. et al. Studying affect dynamics and chronometry using sensor-free detectors. **Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018**, 2018.

BOTELHO, A. F.; BAKER, R. S.; HEFFERNAN, N. T. Improving Sensor-Free Affect Detection Using Deep Learning. In: [s.l: s.n.]. p. 40–51.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

BURKOV, A. **The Hundred-Page Machine Learning Book**. [s.l.] Andriy Burkov, 2019.

CALVO, R. A.; D’MELLO, S. Affect detection: An interdisciplinary review of models, methods, and their applications. **IEEE Transactions on Affective Computing**, v. 1, n. 1, p. 18–37, 2010.

CHEN, T.; GUESTRIN, C. **XGBoost**. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. **Anais...**New York, NY, USA: ACM, 13 ago. 2016.

CHI, M. T. H.; OHLSSON, S. Complex Declarative Learning. In: **The Cambridge handbook of thinking and reasoning**. New York, NY, US: Cambridge University Press, 2005. p. 371–399.

COHEN, J. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37–46, 1960.

CONATI, C.; MACLAREN, H. Empirically building and evaluating a probabilistic model of user affect. **User Modeling and User-Adapted Interaction**, v. 19, n. 3, p. 267–303, 2009.

CORBETT, A. et al. A Cognitive Tutor for Genetics Problem Solving: Learning Gains and Student Modeling. **Journal of Educational Computing Research**, v. 42, n. 2, p. 219–239, 9 mar. 2010.

CORBETT, A. T.; ANDERSON, J. R. Knowledge Tracing : Modeling the Acquisition of Procedural Knowledge. **User Modeling and User-Adapted Interaction**, v. 4, n. 4, p. 253–278, 1995.

COTO, M. et al. Emotions and programming learning: systematic mapping. **Computer Science Education**, v. 00, n. 00, p. 1–36, 2021.

CRAIG, S. et al. Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. **Journal of Educational Media**, v. 29, n. 3, p. 241–250, 2004.

CRAIG, S. D. et al. Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive - Affective states during learning. **Cognition and Emotion**, v. 22, n. 5, p. 777–788, 2008a.

CRAIG, S. D. et al. Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive - Affective states during learning. **Cognition and Emotion**, v. 22, n. 5, p. 777–788, 2008b.

CRAIG, S. D. et al. Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive - Affective states during learning. **Cognition and Emotion**, v. 22, n. 5, p. 777–788, 2008c.

CSIKSZENTMIHALYI, M.; CSIKZENTMIHALY, M. **Flow: The psychology of optimal experience**. [s.l.] Harper & Row New York, 1990. v. 1990

DARWIN, C. **The expression of the emotions in man and animals**. [s.l.] John Murray, 1872.

DAS, R. et al. A New Interpretation of Knowledge Tracing Models' Predictive Performance in Terms of the Cold Start Problem. **CEUR Workshop Proceedings**, v. 3051, 2021.

DASH, M.; LIU, H. Feature selection for classification. **Intelligent Data Analysis**, v. 1, n. 3, p. 131–156, 1997.

DECATUR, S. E.; GOLDREICH, O.; RON, D. Computational Sample Complexity. In: **Human Evolution**. [s.l.] Oxford University Press, 1997. p. 1–6.

DELGADO, R.; TIBAU, X. A. Why Cohen's Kappa should be avoided as performance measure in classification. **PLoS ONE**, v. 14, n. 9, 2019.

DI LEO, I.; MUIS, K. R. Confused, now what? A Cognitive-Emotional Strategy Training (CEST) intervention for elementary students during mathematics problem solving. **Contemporary Educational Psychology**, v. 62, n. May, p. 101879, 2020.

DIETTERICH, T. G. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms 1 Introduction. p. 1–24, 1997.

D'MELLO, S. A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. **Journal of Educational Psychology**, v. 105, n. 4, p. 1082–1099, 2013.

D'MELLO, S. et al. Confusion can be beneficial for learning. **Learning and Instruction**, v. 29, p. 153–170, 2014.

D'MELLO, S.; CALVO, R. A. Beyond the basic emotions. **CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13**, p. 2287, 2013.

D'MELLO, S.; DALE, R.; GRAESSER, A. Disequilibrium in the mind, disharmony in the body. **Cognition and Emotion**, v. 26, n. 2, p. 362–374, 2012.

D'MELLO, S.; GRAESSER, A. Automatic detection of learner's affect from gross body language. **Applied Artificial Intelligence**, v. 23, n. 2, p. 123–150, 2009.

D'MELLO, S.; GRAESSER, A. Dynamics of affective states during complex learning. **Learning and Instruction**, v. 22, n. 2, p. 145–157, 2012a.

- D'MELLO, S.; GRAESSER, A. AutoTutor and affective autotutor. **ACM Transactions on Interactive Intelligent Systems**, v. 2, n. 4, p. 1–39, 1 dez. 2012b.
- D'MELLO, S.; GRAESSER, A. Confusion and its dynamics during device comprehension with breakdown scenarios. **Acta Psychologica**, v. 151, p. 106–116, 2014a.
- D'MELLO, S.; GRAESSER, A.; PICARD, R. W. Toward an affect-sensitive autotutor. **IEEE Intelligent Systems**, v. 22, n. 4, p. 53–61, 2007.
- D'MELLO, S. K. et al. Automatic detection of learner's affect from conversational cues. **User Modeling and User-Adapted Interaction**, v. 18, n. 1–2, p. 45–80, 2008.
- D'MELLO, S. K. Big data in the science of learning. **Big data in psychological research.**, p. 203–225, 2020.
- D'MELLO, S. K.; GRAESSER, A. Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. **User Modeling and User-Adapted Interaction**, v. 20, n. 2, p. 147–187, 2010.
- D'MELLO, S. K.; GRAESSER, A. C. Confusion. In: **International handbook of emotions in education**. [s.l.] Routledge, 2014b. p. 299–320.
- D'MELLO, S. K.; KORY, J. A review and meta-analysis of multimodal affect detection systems. **ACM Computing Surveys**, v. 47, n. 3, 2015.
- D'MELLO, S. K.; LEHMAN, B.; PERSON, N. Monitoring affect states during effortful problem solving activities. **International Journal of Artificial Intelligence in Education**, v. 20, n. 4, p. 361–389, 2010.
- D'MELLO, S.; KORY, J. Consistent but Modest: A Meta-Analysis on Unimodal and Multimodal Affect Detection Accuracies from 30 Studies. **Proceedings of the 14th ACM international conference on Multimodal interaction - ICMI '12**, p. 31–38, 2012.
- D'MELLO, S.; PERSON, N.; LEHMAN, B. Antecedent-consequent relationships and cyclical patterns between affective states and problem solving outcomes. **Frontiers in Artificial Intelligence and Applications**, v. 200, n. 1, p. 57–64, 2009.
- DRAGON, T. et al. Viewing Student Affect and Learning through Classroom Observation and Physical Sensors. In: **International Handbook of the Learning Sciences**. [s.l.: s.n.]. p. 29–39.
- EAGLE, M. et al. Exploring Learner Model Differences Between Students. In: ANDRÉ, E. et al. (Eds.). **Lecture Notes in Computer Science**. Cham: Springer International Publishing, 2017. v. 10331p. 494–497.
- EBERT, F. et al. Confusion detection in code reviews. **Proceedings - 2017 IEEE International Conference on Software Maintenance and Evolution, ICSME 2017**, n. September, p. 549–553, 2017.
- EKMAN, P.; FRIESEN, W. V. **Universals and cultural differences in the judgments of facial expressions of emotion**. **Journal of Personality and Social Psychology**, 1987.

ELFENBEIN, H. A.; AMBADY, N. On the universality and cultural specificity of emotion recognition: A meta-analysis. **Psychological Bulletin**, v. 128, n. 2, p. 203–235, 2002.

ELFENBEIN, H. A.; AMBADY, N. Universals and Cultural Differences in Recognizing Emotions. **Current Directions in Psychological Science**, v. 12, n. 5, p. 159–164, 2003.

ERWIANDA, M. S. F. et al. Improving Confusion-State Classifier Model Using XGBoost and Tree-Structured Parzen Estimator. **2019 2nd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2019**, p. 309–313, 2019.

FANCSALI, S. E. et al. Optimal and Worst-Case Performance of Mastery Learning Assessment with Bayesian Knowledge Tracing. **Educational data mining 2013**, 2013.

FARIA, L.; SANTOS, N. L. Questionário de Competência Emocional revisitado (QCE). In: **Adaptação, desenvolvimento e sucesso acadêmico dos estudantes do ensino superior: Instrumentos de avaliação**. [s.l.] Associação para o Desenvolvimento da Investigação em Psicologia da Educação, 2017. p. 70–79.

FELIPE, D. A. M. et al. **Towards the development of intelligent agent for novice C/C++ programmers through affective analysis of event logs**. Proceedings of the International Multiconference of Engineers and Computers Scientists. **Anais...Hong Kong**: 2012.

FREDRICKSON, B. L. What good are positive emotions? **Review of General Psychology**, v. 2, n. 3, p. 300–319, 1998.

FRIJDA, N. H. The Laws of Emotion. **American Psychologist**, v. 43, n. 5, p. 349–358, 1988.

GONG, Y.; BECK, J. E.; HEFFERNAN, N. T. Comparing knowledge tracing and performance factor analysis by using multiple model fitting procedures. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6094 LNCS, n. PART 1, p. 35–44, 2010.

GOWDA, S. M. et al. Improving models of slipping, guessing, and moment-by-moment learning with estimates of skill difficulty. **EDM 2011 - Proceedings of the 4th International Conference on Educational Data Mining**, p. 199–208, 2011.

GRAESSER, A. et al. Emotions and Learning with AutoTutor. **Proceedings of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work**, p. 569–571, 2007.

GRAESSER, A.; D’MELLO, S. K. Theoretical Perspectives on Affect and Deep Learning. In: **New Perspectives on Affect and Learning Technologies**. New York, NY: Springer New York, 2011. p. 11–21.

GRAFSGAARD, J. F. et al. Multimodal Analysis of the Implicit Affective Channel in Computer-Mediated Textual Communication. **In Proceedings of the 14th ACM International Conference on Multimodal Interaction**, p. 145–152, 2012.

GRAFSGAARD, J. F.; BOYER, K. E.; LESTER, J. C. Predicting facial indicators of confusion with hidden markov models. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6974 LNCS, n. PART 1, p. 97–106, 2011.

HALGREN, E. et al. N400-like Magnetoencephalography Responses Modulated by Semantic Context, Word Frequency, and Lexical Class in Sentences. **NeuroImage**, v. 17, n. 3, p. 1101–1116, nov. 2002.

HEFFERNAN, N. T.; HEFFERNAN, C. L. The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. **International Journal of Artificial Intelligence in Education**, v. 24, n. 4, p. 470–497, 2014.

HESS, U. Now you see it, now you don't--the confusing case of confusion as an emotion: Commentary on Rozin and Cohen (2003). **Emotion**, v. 3, n. 1, p. 76–80, 2003.

IZARD, C. E. The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. **Emotion Review**, v. 2, n. 4, p. 363–370, 2010.

JAQUES, P. A.; VICARI, R. M. Estado da Arte em Ambientes Inteligentes de Aprendizagem que Consideram a Afetividade do Aluno. **Informática na Educação**, v. 8, n. 1, p. 15–38, 2005.

K. D'MELLO, S.; GRAESSER, A. **Mining Bodily Patterns of Affective Experience during Learning**. 2010.

KASURINEN, J.; NIKULA, U. Estimating programming knowledge with Bayesian knowledge tracing. **ACM SIGCSE Bulletin**, v. 41, n. 3, p. 313, 2009.

KAUTZMANN, T. R.; JAQUES, P. A. **Modelo de identificação de unidades de conhecimento de programação em processo de aplicação durante a codificação**. Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020). **Anais...**Sociedade Brasileira de Computação, 24 nov. 2020.

KELTNER, D.; SHIOTA, M. N. New displays and new emotions: A commentary on Rozin and Cohen (2003). **Emotion**, v. 3, n. 1, p. 86–91, 2003.

KHAJAH, M.; LINDSEY, R. V; MOZER, M. C. **How deep is knowledge tracing?** Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016. **Anais...**2016.

KINNUNEN, P.; SIMON, B. Experiencing programming assignments in CS1: The emotional toll. **ICER'10 - Proceedings of the International Computing Education Research Workshop**, p. 77–85, 2010.

KNOTTNERUS, J. A.; TUGWELL, P. Real world research. **Journal of Clinical Epidemiology**, v. 63, n. 10, p. 1051–1052, out. 2010.

KOEDINGER, K. R. et al. Data mining and education. **Wiley Interdisciplinary Reviews: Cognitive Science**, v. 6, n. 4, p. 333–353, 2015.

KOEDINGER, K. R.; CORBETT, A. Cognitive Tutors: Technology Brining Learning Science to the Classroom. In K. Sawyer (Ed.), **The Cambridge Handbook of Learning Sciences**, v. 21, n. 3, p. 61–77, 2006.

KOEDINGER, K. R.; CORBETT, A. T.; PERFETTI, C. The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. **Cognitive Science**, v. 36, n. 5, p. 757–798, 2012.

KÖLLING, M. et al. The BlueJ System and its Pedagogy. **Computer Science Education**, v. 13, n. 4, p. 249–268, 9 dez. 2003.

KUBAT, M. **An Introduction to Machine Learning**. Cham: Springer International Publishing, 2017.

KURNIA, A.; LIM, A.; CHEANG, B. Online Judge. **Computers & Education**, v. 36, n. 4, p. 299–315, 2001.

KUTAS, M.; HILLYARD, S. A. **Reading senseless sentences: brain potentials reflect semantic incongruity.** , 1980.

LAZARUS, R. S. **Psychological stress and the coping process**. New York, NY, US: McGraw-Hill, 1966.

LE THI, H. A.; NGUYEN, V. V.; OUCHANI, S. Gene selection for cancer classification using DCA. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 5139 LNAI, p. 62–72, 2008.

LEE, D. M. C. et al. Exploring the relationship between novice programmer confusion and achievement. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6974 LNCS, n. PART 1, p. 175–184, 2011.

LEE, J. I.; BRUNSKILL, E. The impact on individualizing student models on necessary practice opportunities. **Proceedings of the 5th International Conference on Educational Data Mining, EDM 2012**, p. 118–125, 2012.

LEHMAN, B. et al. Inducing and Tracking Confusion with Contradictions during Complex Learning. **International Journal of Artificial Intelligence in Education**, v. 22, n. 1–2, p. 85–105, 2013.

LEHMAN, B.; D’MELLO, S.; GRAESSER, A. Confusion and complex learning during interactions with computer learning environments. **Internet and Higher Education**, v. 15, n. 3, p. 184–194, 2012.

LEHMANN, B.; GRAESSER, A. To Resolve or Not to Resolve? That is the big question about confusion. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 9112, n. June, p. 883, 2015.

LEWIS, J. R. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. **International Journal of Human-Computer Interaction**, v. 7, n. 1, p. 57–78, 1995.

LEWIS, J. R. Measuring Perceived Usability: The CSUQ, SUS, and UMUX. **International Journal of Human-Computer Interaction**, v. 34, n. 12, p. 1148–1156, 2018.

LIN, C.; CHI, M. Intervention-BKT: Incorporating instructional interventions into Bayesian knowledge tracing. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 9684, p. 208–218, 2016.

LIN, C.; CHI, M. Deep Learning vs . Bayesian Knowledge Tracing : Student Models for Interventions. **Journal of Educational Data Mining**, v. 10, n. 2, p. 28–54, 2018.

LYNCH, C. et al. **Concepts, Structures, and Goals: Redefining Ill-Definedness** *International Journal of Artificial Intelligence in Education*. [s.l.] IOS Press, 2009.

MA, W. et al. Intelligent tutoring systems and learning outcomes: A meta-analysis. **Journal of Educational Psychology**, v. 106, n. 4, p. 901–918, 2014.

MACDOWELL, K. A.; MANDLER, G. Constructions of emotion: Discrepancy, arousal, and mood. **Motivation and Emotion**, v. 13, n. 2, p. 105–124, 1989.

MARCOT, B. G.; HANEA, A. M. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? **Computational Statistics**, v. 36, n. 3, p. 2009–2031, 2021.

MAYER, R. E. et al. Applying the segmenting principle to online geography slideshow lessons. **Educational Technology Research and Development**, v. 66, n. 3, p. 563–577, 2018.

MCDANIEL, B. et al. **Facial Features for Affective State Detection in Learning Environments**. Proceedings of the Annual Meeting of the Cognitive Science Society. *Anais...*2007.

MCNEMAR, Q. Note on the sampling error of the difference between correlated proportions or percentages. **Psychometrika**, v. 12, n. 2, p. 153–157, jun. 1947.

MELLO, S. D. et al. I Feel Your Pain : A Selective Review of Affect-Sensitive Instructional Strategies. **Design Recommendations for Intelligent Tutoring Systems**, v. 2, p. 35–48, 2014.

MELLO, S. K. D. et al. Integrating Affect Sensors in an Intelligent Tutoring System. **In Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International Conference on Intelligent User Interfaces**, p. 7–13, 2005.

MELLO, S. K. D.; SCIENCE, C.; HALL, D. Predicting Affective States expressed through an Emote-Aloud Procedure from AutoTutor ' s Mixed- Initiative Dialogue. **International Journal of Artificial Intelligence in Education**, v. 16, n. 1, p. 3–28, 2006.

MILLS, C.; D'MELLO, S. Emotions during Writing on Topics That Align or Misalign with Personal Beliefs. In: [s.l: s.n.]. p. 638–639.

MOLNAR, C. et al. Relating the Partial Dependence Plot and Permutation Feature Importance to the Data Generating Process. 3 set. 2021.

MOON, T. K. The expectation-maximization algorithm. **IEEE Signal Processing Magazine**, v. 13, n. 6, p. 47–60, 1996.

MOORS, A. et al. Appraisal theories of emotion: State of the art and future development. **Emotion Review**, v. 5, n. 2, p. 119–124, 2013.

MOORS, A.; SCHERER, K. R. The Role of Appraisal in Emotion. In: ROBINSON, M.; WATKINS, E.; HARMON-JONES, E. (Eds.). **Handbook of cognition and emotion**. New York: Guilford Press, 2013. p. 135–155.

MORAIS, F. DE et al. **EmAP-ML: A Protocol of Emotions and Behaviors Annotation for Machine Learning Labels**. (M. Scheffel et al., Eds.) 14th European Conference on Technology Enhanced Learning, EC-TEL 2019. **Anais...**Delft: Springer International Publishing, 2019.

MOTA, S.; PICARD, R. W. Automated Posture Analysis for Detecting Learner's Interest Level. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops**, v. 5, p. 1–6, 2003.

MYERS, G. J.; SANDLER, C.; BADGETT, T. **The Art of Software Testing**. [s.l.: s.n.].

NAGATANI, K. et al. Augmenting knowledge tracing by considering forgetting behavior. **The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019**, p. 3101–3107, 2019.

OCUMPAUGH, J.; BAKER, R. S.; OCUMPAUGH, J., & BAKER, R. S. (2014). **B. R. O. M. P. ( B. ) 2 . 0 T. AND T. MANUAL. Baker Rodrigo Ocumpaugh Monitoring Protocol ( BROMP ) 2 . 0 Technical and Training Manual. , 2014.**

OLIVEIRA, E. DE; JAQUES, P. A. Inferindo as emoções do usuário pela face através de um sistema psicológico de codificação facial. **Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems**, n. January 2008, p. 156–165, 2008.

ORTONY, A.; CLORE, G. L.; COLLINS, A. **The Cognitive Structure of Emotions**. New York: Cambridge University Press, 1988.

PANG, B.; LEE, L. **Opinion Mining and Sentiment Analysis: Foundations and Trends in Information Retrieval**. [s.l.: s.n.]. v. 2

PAPER, D. Introduction to Deep Learning. In: **TensorFlow 2.x in the Colaboratory Cloud: An Introduction to Deep Learning on Google's Cloud Service**. Berkeley, CA: Apress, 2021. p. 1–24.

PARDOS, Z. A.; HEFFERNAN, N. T. Modeling individualization in a Bayesian networks implementation of knowledge tracing. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6075 LNCS, p. 255–266, 2010a.

PARDOS, Z. A.; HEFFERNAN, N. T. Navigating the parameter space of Bayesian Knowledge Tracing models: Visualizations of the convergence of the expectation maximization algorithm. **Educational Data Mining 2010 - 3rd International Conference on Educational Data Mining**, p. 161–170, 2010b.

PARDOS, Z. A.; HEFFERNAN, N. T. KT-IDEM: Introducing item difficulty to the knowledge tracing model. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6787 LNCS, p. 243–254, 2011.

PEKRUN, R. Emotions as Drivers of Learning and Cognitive Development. In: **New Perspectives on Affect and Learning Technologies**. New York, NY: Springer New York, 2011. p. 23–39.

PEKRUN, R.; STEPHENS, E. J. Academic emotions. In: **APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors**. Washington: American Psychological Association, 2012. v. 2p. 3–31.

PELÁNEK, R. Metrics for Evaluation of Student Models. **JEDM - Journal of Educational Data Mining**, v. 7, n. 2, p. 1–19, 2015.

PELÁNEK, R. Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. **User Modeling and User-Adapted Interaction**, v. 27, n. 3–5, p. 313–350, 2017a.

PELÁNEK, R. Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. **User Modeling and User-Adapted Interaction**, v. 27, n. 3–5, p. 313–350, 2017b.

PENMETSA, P.; SHI, Y.; PRICE, T. Investigate Effectiveness of Code Features in Knowledge Tracing Task on Novice Programming Course. **CEUR Workshop Proceedings**, v. 3051, p. 0–5, 2021.

PICARD, R. W. Affective Computing. **MIT press**, n. 321, p. 1–16, 1995.

PICARD, R. W. Affective Computing: From laughter to IEEE. **IEEE Transactions on Affective Computing**, v. 1, n. 1, p. 11–17, 2010.

PORAYSKA-POMSTA, K. et al. Knowledge elicitation methods for affect modelling in education. **International Journal of Artificial Intelligence in Education**, v. 22, n. 3, p. 107–140, 2013.

QIU, Y. et al. Does time matter? Modeling the effect of time in bayesian knowledge tracing. **EDM 2011 - Proceedings of the 4th International Conference on Educational Data Mining**, n. Table I, p. 139–148, 2011.

RAPOSO, A. C.; MARANHÃO, D.; SOARES NETO, C. **Análise do Modelo BKT na Avaliação da Curva de Aprendizagem de Alunos de Algoritmos**. Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019). **Anais...Brazilian Computer Society (Sociedade Brasileira de Computação - SBC)**, 11 nov. 2019.

RASCHKA, S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. 13 nov. 2018.

REIS, H. et al. Analysis of Permanence Time in Emotional States: A Case Study Using Educational Software. **International Conference on Intelligent Tutoring Systems (ITS)**, p. 180–190, 2018.

- REYNOLDS, C.; PICARD, R. Affective sensors, privacy, and ethical contracts. **Conference on Human Factors in Computing Systems - Proceedings**, p. 1103–1106, 2004.
- RICHEY, J. E. et al. More confusion and frustration, better learning: The impact of erroneous examples. **Computers and Education**, v. 139, n. May, p. 173–190, 2019.
- RITTER, S. et al. How mastery learning works at scale. **L@S 2016 - Proceedings of the 3rd 2016 ACM Conference on Learning at Scale**, p. 71–79, 2016.
- RODRIGO, M. M. T.; BAKER, R. S. J.; NABOS, J. Q. The Relationships Between Sequences of Affective States and Learner Achievement. **Proceedings of the 18th International Conference on Computers in Education**, n. December 2016, p. 56–60, 2010.
- RODRIGO, MA. M. T. et al. Affective and behavioral predictors of novice programmer achievement. **ACM SIGCSE Bulletin**, v. 41, n. 3, p. 156, 2009.
- ROUSSEAU, D.; USTYUZHANIN, A. Machine Learning Scientific Competitions and Datasets. **Artificial Intelligence for High Energy Physics**, p. 765–812, 2022.
- ROZIN, P.; COHEN, A. B. High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. **Emotion**, v. 3, n. 1, p. 68–75, 2003.
- RUSSELL, J. A. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. **Psychological Bulletin**, v. 115, n. 1, p. 102–141, 1994.
- RUSSELL, J. A. Core affect and the psychological construction of emotion. **Psychological Review**, v. 110, n. 1, p. 145–172, 2003.
- RUSSELL, J. A.; BACHOROWSKI, J.; FERNÁNDEZ-DOLS, J.-M. Facial and Vocal Expressions of Emotion. **Annual Review of Psychology**, 2003.
- SAURO, J.; LEWIS, J. R. **Quantifying the User Experience**. [s.l.: s.n.]. v. 148
- SAYFAN, L.; LAGATTUTA, K. H. Grownups are not afraid of scary stuff, but kids are: Young children's and adults' reasoning about children's, infants', and adults' fears. **Child Development**, v. 79, n. 4, p. 821–835, 2008.
- SCHERER, K. R. Toward a dynamic theory of emotion. **Geneva Studies in Emotion**, v. 1, p. 1–96, 1987.
- SCHERER, K. R. Appraisal Theory. In: DALGLEISH, T.; POWER, M. J. (Eds.). **handbook of Cognition and Emotion**. [s.l.] John Wiley & Sons, 1999. p. 637–663.
- SCHERER, K. R. What are emotions? and how can they be measured? **Social Science Information**, v. 44, n. 4, p. 695–729, 2005.
- SCHERER, K. R. The dynamic architecture of emotion: Evidence for the component process model. **Cognition & Emotion**, v. 23, n. 7, p. 1307–1351, 2009.

- SCHNEIDER, G.; JAQUES, P. A. Combinando técnicas de análise estática e avaliação dinâmica para avaliação de código em ambientes de aprendizagem de programação. **Revista Brasileira de Computação Aplicada**, v. 8, n. 1, p. 114–129, 2016.
- SEGER, C. An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. **Degree Project Technology**, p. 41, 2018.
- SHU, L. et al. A review of emotion recognition using physiological signals. **Sensors (Switzerland)**, v. 18, n. 7, 2018.
- SHWARTZ-ZIV, R.; ARMON, A. Tabular data: Deep learning is not all you need. **Information Fusion**, v. 81, p. 84–90, 1 maio 2022.
- SILVIA, P. J. Confusion and interest: The role of knowledge emotions in aesthetic experience. **Psychology of Aesthetics, Creativity, and the Arts**, v. 4, n. 2, p. 75–80, 2010.
- SLATER, S.; BAKER, R. S. Degree of error in Bayesian knowledge tracing estimates from differences in sample sizes. **Behaviormetrika**, v. 45, n. 2, p. 475–493, 2018.
- SMIDERLE, R. **O efeito da gamificação no engajamento e aprendizagem de programação: Um estudo considerando a personalidade e a orientação motivacional dos estudantes**. [s.l.] Universidade do Vale do Rio dos Sinos, 2019.
- STEIDL, S. et al. “Of All Things the Measure Is Man”: Automatic Classification of Emotions and Inter-Labeler Consistency. Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. **Anais...IEEE**, 2005.
- TAKSIC, V. Convergent and divergent validity of the Emotional Skills and Competence Questionnaire. **Comunicación presentada en XII Days of Psychology, Zadar, Croacia**, 2000.
- TIAM-LEE, T. J.; SUMI, K. Adaptive Feedback Based on Student Emotion in a System for Programming Practice. **International Conference on Intelligent Tutoring Systems**, p. 243–255, 2018.
- TIAM-LEE, T. J.; SUMI, K. Analysis and Prediction of Student Emotions While Doing Programming Exercises. In: COY, A.; HAYASHI, Y.; CHANG, M. (Eds.). **International Conference on Intelligent Tutoring Systems**. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019. v. 11528p. 24–33.
- VANLEHN, K. et al. Why Do Only Some Events Cause Learning During Human Tutoring? **Cognition and Instruction**, v. 21, n. 3, p. 209–249, 2003.
- VANLEHN, K. The Behavior of tutoring systems. **International Journal of Artificial Intelligence in Education**, v. 16, n. 3, p. 227–265, 2006.
- VANLEHN, K. The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. **Educational Psychologist**, v. 46, n. 4, p. 197–221, out. 2011.
- VEA, L. A. et al. Development of an Intelligent Agent that Detects Student’s Negative Affect while Making a Computer Program. **Proceedings of the International Conference on Advances in Image Processing - ICAIP 2017**, p. 123–128, 2017.

VEA, L.; RODRIGO, M. M. Modeling negative affect detector of novice programming students using keyboard dynamics and mouse behavior. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 10004 LNAI, p. 127–138, 2017a.

VEA, L.; RODRIGO, M. M. Modeling negative affect detector of novice programming students using keyboard dynamics and mouse behavior. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 10004 LNAI, p. 127–138, 2017b.

VEA, L.; RODRIGO, M. M. Modeling negative affect detector of novice programming students using keyboard dynamics and mouse behavior. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 10004 LNAI, p. 127–138, 2017c.

VIERA, A. J.; GARRETT, J. M. Understanding interobserver agreement: the kappa statistic. **Family medicine**, v. 37, n. 5, p. 360–3, maio 2005.

WANG, C. et al. Knowledge tracing for complex problem solving: Granular rank-based tensor factorization. **UMAP 2021 - Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization**, p. 179–188, 2021.

WANG, S. et al. Modeling student learning outcomes in studying programming language course. **7th International Conference on Information Science and Technology, ICIST 2017 - Proceedings**, p. 263–270, 2017.

WANG, Y.; HEFFERNAN, N. T.; HEFFERNAN, C. Towards Better Affect Detectors: Effect of Missing Skills, Class Features and Common Wrong Answers. **Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15**, p. 31–35, 2015.

WILLIAMS, R. J.; ZIPSER, D. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. **Neural Computation**, v. 1, n. 2, p. 270–280, 1989.

WINNE, P. H.; BAKER, R. S. J. D. The Potentials of Educational Data Mining for Researching Metacognition, Motivation and Self-Regulated Learning. **JEDM - Journal of Educational Data Mining**, v. 5, n. 1, p. 1–8, 2013.

WONG, T. T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. **Pattern Recognition**, v. 48, n. 9, p. 2839–2846, 2015.

WOOLF, B. et al. Affect-aware tutors: recognising and responding to student affect. **International Journal of Learning Technology**, v. 4, n. 3/4, p. 129, 2009.

WOOLF, B. P. **Building Intelligent Interactive Tutors**. [s.l.] Elsevier, 2009. v. 11

YANG, T. Y. et al. Active learning for student affect detection. **EDM 2019 - Proceedings of the 12th International Conference on Educational Data Mining**, n. June, p. 208–217, 2019.

YUDELSON, M. V. Individualizing Bayesian Knowledge Tracing. Are Skill Parameters More Important Than Student Parameters? **Proceedings of the 9th International Conference on Educational Data Mining**, p. 556–561, 2016.

YUDELSON, M. V.; KOEDINGER, K. R.; GORDON, G. J. Individualized bayesian knowledge tracing models. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 7926 LNAI, p. 171–180, 2013.

ZENG, Z. et al. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 31, n. 1, p. 39–58, 2009.

ZHANG, W.; ZHAO, X.; LI, Z. A Comprehensive Study of Smartphone-Based Indoor Activity Recognition via Xgboost. **IEEE Access**, v. 7, p. 80027–80042, 2019.

ZHOU, Y. et al. Confusion State Induction and EEG-based Detection in Learning. **Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference**, v. 2018, n. Shaoqi Li, p. 3290–3293, 2018.

**APÊNDICE A – TODOS OS RESULTADOS DE VALIDAÇÃO CRUZADA DOS MODELOS NAS AMOSTRAS DE JANELAS FIXAS**

		Resultados de acurácia (ac), AUC ROC (auc), kappa (k), precisão (p), recall (r) e F1-score (f1) dos modelos da Tese (T) e <i>baseline</i> (B) nas amostras de janelas fixas (5s, 10s, 20s, 40s, 60s, 90s, 120s, 180s, 240s e 360s)																			
Modelo		T 5	B 5	T 10	B 10	T 20	B 20	T 40	B 40	T 60	B 60	T 90	B 90	T 120	B 120	T 180	B 180	T 240	B 240	T 360	B 360
<i>LR</i>	ac	,86912	,75131	,86564	,75235	,84162	,76906	,80815	,76313	,81058	,76891	,77830	,76996	,76841	,75254	,77777	,75364	,75357	,74971	,75109	,72061
	auc	,92833	,79479	,93376	,81901	,91499	,83159	,87679	,82130	,87054	,83134	,85422	,81420	,84668	,81367	,84603	,81226	,82497	,79886	,81389	,78231
	k	,73823	,50257	,73130	,50476	,68323	,53810	,61625	,52620	,62119	,53787	,55659	,52106	,53678	,50504	,55556	,49025	,50712	,45917	,50219	,44129
	p	,84631	,69170	,85441	,69927	,82729	,72670	,78972	,72686	,79829	,73903	,76701	,73368	,76991	,73495	,77808	,73291	,74967	,71622	,74418	,71763
	r	,90335	,90862	,88302	,88935	,86435	,86221	,84246	,84571	,83220	,83334	,80042	,81925	,76776	,78927	,77896	,77165	,76128	,75880	,76626	,73152
	f1	,87337	,78529	,86803	,78241	,84522	,78861	,81460	,78119	,81447	,78302	,78262	,77355	,76807	,76087	,77807	,75128	,75507	,73618	,75451	,72320
<i>SVM</i>	ac	,88107	,75493	,87777	,75764	,85595	,77437	,83760	,77544	,82627	,77379	,81985	,76941	,79953	,75989	,79588	,75299	,77299	,74969	,77361	,73441
	auc	,93534	,80292	,92997	,80253	,92116	,82509	,89286	,83137	,89290	,83241	,88577	,82451	,87577	,82331	,86130	,80907	,84039	,81142	,84053	,80123
	k	,76215	,50983	,75554	,51533	,71188	,54877	,67517	,55087	,65260	,54764	,63971	,53994	,59902	,51973	,59180	,50728	,54595	,49941	,54718	,46886
	p	,85162	,69333	,85110	,69963	,84403	,73880	,81419	,73769	,80095	,74309	,80105	,74102	,78785	,73365	,78236	,73219	,76930	,73784	,76768	,72221
	r	,92523	,91586	,91676	,90520	,87507	,84968	,87679	,85529	,87124	,84094	,85262	,83040	,82102	,81648	,82248	,80063	,78075	,77439	,78795	,76779
	f1	,88627	,78906	,88233	,78888	,85889	,78983	,84383	,79199	,83385	,78842	,82502	,78279	,80364	,77236	,80118	,76444	,77409	,75515	,77640	,74285
<i>NB</i>	ac	,83434	,72894	,81082	,73815	,77068	,74525	,72240	,71601	,74186	,73915	,71623	,76497	,72083	,71857	,70592	,75058	,69387	,74967	,68936	,69666
	auc	,87591	,77984	,86972	,77850	,83074	,77497	,79001	,77467	,80182	,78934	,77137	,77189	,77069	,77906	,75992	,76957	,74476	,76642	,72880	,75213
	k	,66869	,45785	,62164	,47630	,54129	,49046	,44481	,43196	,48377	,47837	,43246	,40254	,44165	,43707	,41190	,43589	,38773	,42031	,37874	,39341
	p	,80300	,66396	,78463	,67732	,73556	,68891	,67209	,66782	,76350	,71398	,72316	,65565	,72865	,69529	,72491	,72280	,69978	,72711	,68021	,70732
	r	,88782	,93039	,85983	,91043	,84963	,89407	,87139	,86185	,70244	,79986	,70065	,86255	,70437	,78584	,66675	,70762	,68085	,67450	,71402	,67495
	f1	,84301	,77463	,81980	,77658	,78791	,77808	,75863	,75221	,73096	,75387	,71096	,74293	,71548	,73415	,69340	,71395	,68875	,69688	,69608	,68967
<i>k-NN</i>	ac	,86393	,73418	,86722	,73604	,84588	,75425	,81674	,76098	,81220	,75594	,79047	,76333	,78993	,74861	,79289	,74574	,75741	,73736	,75976	,72496
	auc	,91914	,79561	,92946	,80227	,91387	,82146	,88615	,82711	,88442	,82446	,87316	,80864	,85733	,82205	,85592	,80635	,83400	,80748	,83113	,79539
	k	,72786	,46834	,73446	,47213	,69177	,50853	,63349	,52199	,62445	,51201	,58096	,50011	,57980	,49716	,58577	,50124	,51483	,49935	,51954	,45001
	p	,82825	,70976	,83221	,71722	,80475	,74158	,78382	,74132	,79094	,73114	,77268	,73183	,77004	,73669	,78976	,73462	,75568	,73654	,75313	,72115
	r	,92003	,79440	,91995	,78194	,91425	,78077	,87788	,80283	,85063	,81605	,82381	,79053	,82890	,77583	,80074	,78736	,76257	,77820	,77640	,73734
	f1	,87139	,74937	,87375	,74760	,85579	,76043	,82760	,77057	,81939	,77022	,79698	,75962	,79776	,75491	,79435	,75911	,75828	,75617	,76366	,72757
<i>SGD</i>	ac	,87223	,75390	,87354	,75604	,84535	,76906	,81672	,76690	,81004	,77053	,78494	,76054	,77971	,75367	,78619	,74510	,76196	,73281	,75615	,72061
	auc	,92801	,79509	,92867	,81524	,91117	,82775	,87164	,81653	,86904	,82734	,85053	,81220	,84805	,81345	,84573	,80869	,82032	,79871	,81138	,78157
	k	,74445	,50775	,74710	,51215	,69070	,53813	,63339	,53376	,62012	,54113	,56990	,52996	,55939	,50730	,57243	,50604	,52391	,46563	,51231	,44129

	p	.84440	.69362	.85691	.70190	.82946	.72557	.79399	.73510	.78907	.74642	.77155	.74305	.78405	.73557	.77975	.73042	.75902	.71590	.75159	.71502
	r	.91377	.91173	.89780	.89358	.87076	.86547	.85749	.83500	.84845	.82364	.81042	.81153	.77452	.79153	.79947	.80427	.76778	.77046	.76479	.73734
	fl	.87744	.78770	.87651	.78572	.84932	.78918	.82398	.78143	.81722	.78246	.79003	.77520	.77875	.76226	.78894	.76498	.76257	.74143	.75761	.72484
<i>DT</i>	ac	.87485	.75805	.86985	.74921	.83263	.75529	.79042	.75293	.78950	.75483	.77883	.75275	.75708	.72423	.75905	.74330	.74447	.72958	.74457	.68363
	auc	.91873	.77442	.92148	.81157	.87879	.78761	.83510	.79267	.84609	.80683	.82648	.76588	.82189	.78571	.81524	.76279	.79036	.75654	.75178	.71958
	k	.74970	.51604	.73971	.49845	.66524	.51061	.58087	.50584	.57906	.50965	.55765	.46455	.51412	.44847	.51821	.42504	.48891	.40974	.48919	.36724
	p	.84440	.69749	.84640	.71940	.82334	.71529	.76748	.70706	.78803	.72981	.76045	.70202	.75814	.72740	.76768	.70827	.73772	.70001	.72370	.66964
	r	.91898	.91279	.90412	.81877	.84745	.84966	.83606	.86611	.79436	.81491	.81597	.80931	.75649	.72022	.74650	.72217	.76258	.72364	.79104	.74307
	fl	.87999	.79054	.87404	.76514	.83462	.77612	.79958	.77745	.79047	.76838	.78637	.75100	.75670	.72195	.75558	.71440	.74891	.70999	.75543	.70239
<i>AB</i>	ac	.87849	.75962	.88040	.75765	.85381	.76375	.81028	.75615	.80028	.75971	.78329	.75004	.77458	.74743	.76628	.73184	.75418	.72633	.74234	.71481
	auc	.93686	.77601	.93533	.82734	.91986	.82106	.88000	.81440	.87043	.81241	.85626	.80572	.84552	.80233	.84109	.79870	.81904	.79228	.81994	.77452
	k	.75698	.51918	.76080	.51533	.70761	.52746	.62049	.51224	.60066	.51946	.56661	.50553	.54920	.49484	.53258	.46374	.50832	.45258	.48463	.42954
	p	.84927	.70559	.85371	.70169	.83584	.71969	.78840	.72007	.77327	.72808	.76842	.71950	.76324	.71287	.75636	.70726	.74974	.72113	.73453	.71634
	r	.92211	.89307	.91885	.89886	.88240	.86435	.85315	.83811	.85396	.83119	.81051	.82816	.79729	.82779	.78623	.79099	.76516	.74182	.76180	.71393
	fl	.88371	.78819	.88476	.78776	.85812	.78516	.81824	.77410	.81069	.77594	.78858	.76959	.77952	.76561	.77050	.74651	.75696	.73035	.74664	.71399
<i>RF</i>	ac	.89094	.75442	.89725	.75660	.87024	.77755	.84563	.78563	.83927	.78784	.81708	.73228	.82274	.76612	.80556	.71794	.79374	.71013	.79610	.73512
	auc	.94857	.81620	.95082	.82998	.93528	.84449	.91918	.84947	.91161	.84893	.90439	.83442	.90144	.84344	.88887	.82064	.86517	.82589	.86917	.81153
	k	.78190	.50876	.79452	.51323	.74050	.55511	.69126	.57123	.67859	.57573	.63419	.53888	.64546	.53219	.61118	.49153	.58745	.49938	.59220	.47031
	p	.87060	.70585	.87981	.73102	.85881	.74040	.82965	.75542	.82023	.75776	.80791	.74870	.81377	.74800	.79989	.73421	.79443	.73671	.80007	.73331
	r	.91899	.87431	.92202	.81456	.88673	.85492	.87142	.84565	.87017	.84852	.83161	.81162	.83801	.80172	.81650	.77164	.79370	.78081	.79096	.74456
	fl	.89393	.78086	.89992	.77012	.87234	.79349	.84965	.79769	.84408	.80013	.81918	.77820	.82543	.77343	.80764	.75188	.79353	.75753	.79504	.73690
<i>XGB</i>	ac	.89197	.76480	.89463	.76607	.87710	.77965	.84831	.78239	.84303	.78190	.82151	.70126	.81141	.76384	.80798	.71250	.80221	.70488	.79463	.73153
	auc	.94483	.81249	.94528	.83293	.93369	.84468	.90643	.83931	.90410	.83953	.89867	.81409	.88228	.82301	.87700	.81187	.86712	.80279	.86557	.79825
	k	.78394	.52951	.78927	.53217	.75419	.55930	.69661	.56477	.68611	.56385	.64303	.52669	.62281	.52763	.61597	.48664	.60440	.47476	.58926	.46311
	p	.87191	.70935	.87439	.72591	.86296	.74315	.83298	.75004	.82421	.74810	.80924	.73010	.79742	.73803	.80004	.73196	.79828	.72838	.79620	.72921
	r	.92004	.89924	.92308	.85565	.89723	.85490	.87357	.84774	.87442	.85393	.84151	.83592	.83567	.81751	.82252	.76804	.80930	.76009	.79386	.74311
	fl	.89495	.79286	.89769	.78519	.87957	.79502	.85218	.79534	.84813	.79688	.82460	.77899	.81584	.77538	.81071	.74899	.80351	.74335	.79461	.73456
<i>MLP</i>	ac	.88468	.76168	.88041	.75974	.86441	.77754	.83600	.77867	.82626	.78188	.80820	.77549	.80293	.76273	.79406	.75962	.78015	.74969	.78231	.73951
	auc	.92723	.81268	.93662	.83215	.92792	.83857	.89403	.83879	.89134	.84134	.87191	.81811	.86890	.82373	.85942	.81709	.83866	.81772	.83106	.80731
	k	.76936	.52327	.76083	.51951	.72879	.55511	.67200	.55733	.65254	.56384	.61643	.55102	.60584	.52546	.58816	.51923	.56031	.49939	.56466	.47908
	p	.87273	.70352	.86226	.70455	.84925	.73847	.81275	.74458	.80609	.75088	.79286	.74410	.78944	.72789	.78269	.74447	.77646	.74160	.78111	.72486
	r	.90234	.90647	.90622	.89676	.88659	.86017	.87568	.84886	.86034	.84744	.83484	.84037	.82779	.83921	.81517	.78971	.78984	.76661	.78953	.77510
	fl	.88692	.79181	.88328	.78875	.86722	.79455	.84236	.79258	.83208	.79568	.81296	.78885	.80771	.77915	.79819	.76582	.78227	.75326	.78360	.74770

**APÊNDICE B – TODOS OS RESULTADOS DE VALIDAÇÃO CRUZADA DOS MODELOS NAS AMOSTRAS DE JANELAS FIXAS  
COM ATRIBUTOS DE DADOS SELECIONADOS**

Modelo		Resultados de acurácia (ac), AUC ROC (auc), kappa (k), precisão (p), recall (r) e F1-score (f1) dos modelos da Tese (T) e <i>baseline</i> (B) nas amostras de janelas fixas (5s, 10s, 20s, 40s, 60s, 90s, 120s, 180s, 240s e 360s)																			
		T 5	B 5	T 10	B 10	T 20	B 20	T 40	B 40	T 60	B 60	T 90	B 90	T 120	B 120	T 180	B 180	T 240	B 240	T 360	B 360
<i>LR</i>	ac	,86756	,75338	,86617	,74920	,83474	,76853	,80762	,76313	,80408	,76891	,77884	,76165	,76671	,75254	,77595	,75115	,75357	,73735	,75109	,71698
	auc	,92599	,79535	,93244	,80687	,90736	,83249	,87601	,82130	,87009	,83134	,85235	,81436	,84726	,81367	,84550	,81123	,82512	,79996	,81389	,78083
	k	,73512	,50673	,73235	,49843	,66946	,53703	,61519	,52620	,60819	,53787	,55768	,52331	,53339	,50504	,55194	,50237	,50712	,47470	,50219	,43403
	p	,84310	,69291	,85795	,70316	,81751	,72643	,79507	,72686	,79269	,73903	,76667	,73328	,76882	,73495	,77874	,73684	,74967	,72066	,74418	,71765
	r	,90442	,91173	,87880	,86509	,86330	,86114	,83175	,84571	,82463	,83334	,80263	,82483	,76552	,78927	,77293	,78135	,76128	,77437	,76626	,71988
	f1	,87224	,78722	,86780	,77545	,83949	,78798	,81223	,78119	,80800	,78302	,78361	,77563	,76628	,76087	,77514	,75811	,75507	,74610	,75451	,71705
<i>SVM</i>	ac	,88159	,75492	,87882	,75289	,85754	,77755	,83546	,77544	,82356	,77379	,81930	,76221	,79898	,75989	,79649	,75904	,77235	,73607	,77361	,73876
	auc	,93148	,79997	,93686	,80172	,92148	,82739	,89880	,83137	,89289	,83241	,88947	,81762	,87845	,82331	,86397	,81841	,84059	,81256	,84053	,80737
	k	,76320	,50980	,75766	,50580	,71505	,55514	,67090	,55087	,64718	,54764	,63860	,52445	,59792	,51973	,59301	,51809	,54468	,47215	,54718	,47756
	p	,86007	,69482	,85975	,69563	,83620	,74367	,80808	,73769	,79589	,74309	,79364	,72535	,78394	,73365	,78392	,73303	,76701	,73112	,76768	,72929
	r	,91275	,91067	,90622	,90095	,88986	,84864	,88215	,85529	,87339	,84094	,86368	,84481	,82672	,81648	,82127	,81520	,78336	,74588	,78795	,76639
	f1	,88530	,78803	,88203	,78482	,86202	,79199	,84297	,79199	,83211	,78842	,82637	,78020	,80449	,77236	,80147	,77163	,77425	,73770	,77640	,74539
<i>NB</i>	ac	,78607	,73571	,76814	,74025	,77013	,75214	,72772	,71601	,74186	,73915	,71456	,68962	,70892	,71857	,70592	,65518	,69387	,71334	,68936	,67995
	auc	,87266	,78529	,86667	,77732	,83547	,78142	,79260	,77467	,80211	,78934	,77259	,77616	,76827	,77906	,76103	,77887	,74740	,77127	,72880	,74524
	k	,57217	,47140	,53632	,48052	,54022	,50423	,45547	,43196	,48376	,47837	,42912	,37920	,41782	,43707	,41189	,31026	,38773	,42670	,37874	,36001
	p	,81841	,67053	,81623	,68471	,73642	,69865	,75217	,66782	,76404	,71398	,72464	,63396	,76976	,69529	,72814	,60470	,69576	,72249	,68021	,71215
	r	,73728	,92938	,69338	,89148	,84433	,88666	,68278	,86185	,70135	,79986	,69177	,90687	,59905	,78584	,66069	,89732	,69124	,69781	,71402	,60963
	f1	,77525	,77883	,74906	,77435	,78619	,78141	,71509	,75221	,73064	,75387	,70700	,74527	,67277	,73415	,69145	,72235	,69218	,70701	,69608	,65510
<i>k-NN</i>	ac	,86291	,73625	,86986	,73288	,83953	,75266	,81939	,76098	,81166	,75594	,79047	,75999	,79673	,74861	,79229	,74574	,75806	,74056	,75976	,73150
	auc	,91914	,79983	,93141	,79923	,90348	,82351	,88825	,82711	,88498	,82446	,86302	,80940	,86914	,82205	,85449	,79456	,83101	,80916	,83113	,79967
	k	,72583	,47249	,73973	,46580	,67907	,50536	,63878	,52199	,62337	,51201	,58097	,52000	,59342	,49716	,58456	,49152	,51612	,48109	,51954	,46309
	p	,82500	,71107	,84047	,71188	,81040	,74129	,78247	,74132	,79207	,73114	,77091	,73638	,77699	,73669	,78673	,73475	,76083	,72946	,75313	,72529
	r	,92212	,79751	,91362	,78508	,88777	,77654	,88859	,80283	,84631	,81605	,82714	,81042	,83351	,77583	,80439	,77036	,75348	,76781	,77640	,75038
	f1	,87053	,75153	,87525	,74624	,84703	,75826	,83149	,77057	,81796	,77022	,79751	,77114	,80392	,75491	,79465	,75098	,75672	,74754	,76366	,73611
<i>SGD</i>	ac	,87378	,75650	,86934	,75288	,83899	,77065	,81565	,76690	,80624	,77053	,78548	,76877	,77916	,75367	,78439	,75480	,76131	,73930	,75615	,71843
	auc	,92207	,79576	,93317	,80716	,90953	,83051	,87172	,81653	,86807	,82734	,85377	,81245	,84914	,81345	,84676	,80977	,82257	,79985	,81138	,77993

	k	,74756	,51296	,73870	,50580	,67797	,54132	,63124	,53376	,61253	,54113	,57097	,53998	,55828	,50730	,56881	,50966	,52261	,47860	,51231	,43694
	p	,85419	,69599	,85496	,70636	,82540	,72647	,78941	,73510	,78560	,74642	,77431	,74439	,78346	,73557	,77806	,73285	,75976	,72386	,75159	,71794
	r	,90233	,91276	,89042	,86931	,86117	,86866	,86389	,83500	,84526	,82364	,80705	,82487	,77343	,79153	,79703	,80310	,76518	,77307	,76479	,72425
	fl	,87728	,78960	,87202	,77894	,84262	,79098	,82413	,78143	,81362	,78246	,78980	,78174	,77784	,76226	,78682	,76609	,76150	,74711	,75761	,71935
<i>DT</i>	ac	,88055	,75805	,86985	,74921	,82682	,75529	,79739	,75293	,78623	,75483	,77829	,73452	,76443	,72423	,75905	,70103	,74447	,70878	,74457	,68429
	auc	,91743	,77344	,92148	,81157	,89554	,78790	,83415	,79267	,84606	,80683	,83778	,76996	,76752	,78571	,81479	,76349	,79042	,75735	,75178	,73944
	k	,76112	,51604	,73971	,49845	,65360	,51061	,59479	,50584	,57252	,50965	,55660	,46901	,52888	,44847	,51821	,40217	,48891	,41758	,48919	,36861
	p	,85564	,69749	,84640	,71940	,83257	,71529	,78614	,70706	,78246	,72981	,75999	,70960	,76637	,72740	,76768	,69531	,73772	,70364	,72370	,67948
	r	,91690	,91279	,90412	,81877	,82091	,84966	,81892	,86611	,79762	,81491	,81490	,79604	,76215	,72022	,74650	,72108	,76258	,72625	,79104	,70098
	fl	,88478	,79054	,87404	,76514	,82571	,77612	,80169	,77745	,78815	,76838	,78588	,74963	,76398	,72195	,75558	,70680	,74891	,71441	,75543	,68921
<i>AB</i>	ac	,87953	,75961	,87936	,75975	,85645	,76375	,81136	,75615	,80028	,75971	,78329	,75055	,77175	,74743	,76326	,73425	,75418	,72631	,74234	,71044
	auc	,93606	,77836	,93565	,82299	,92190	,82106	,88597	,81440	,87043	,81241	,85626	,80912	,84501	,80233	,84026	,80011	,81904	,79475	,81994	,77152
	k	,75906	,51917	,75872	,51954	,71288	,52746	,62268	,51224	,60066	,51946	,56661	,50111	,54355	,49484	,52655	,46855	,50832	,45260	,48463	,42090
	p	,85142	,70527	,85299	,70443	,83632	,71969	,79435	,72007	,77327	,72808	,76842	,71574	,76084	,71287	,75487	,72004	,74974	,72028	,73453	,71762
	r	,92108	,89410	,91781	,89780	,88875	,86435	,84464	,83811	,85396	,83119	,81051	,83150	,79388	,82779	,78020	,76807	,76516	,74187	,76180	,69808
	fl	,88444	,78838	,88387	,78909	,86121	,78516	,81792	,77410	,81069	,77594	,78858	,76891	,77664	,76561	,76691	,74278	,75696	,73060	,74664	,70660
<i>RF</i>	ac	,89042	,75857	,89725	,75765	,87077	,77489	,84904	,78563	,83710	,79272	,81652	,76884	,82501	,76612	,80316	,75116	,79632	,75097	,79610	,73948
	auc	,95054	,81884	,95082	,83245	,93752	,84636	,92118	,84947	,91113	,84952	,90307	,83076	,90040	,84344	,88776	,81888	,87326	,81799	,86917	,80953
	k	,78086	,51707	,79452	,51534	,74153	,54980	,70088	,57123	,67426	,58549	,63308	,53774	,64998	,53219	,60637	,50237	,59262	,50192	,59220	,47902
	p	,87339	,70913	,87981	,72395	,85744	,74368	,83414	,75542	,81910	,75901	,80516	,74457	,81965	,74800	,80253	,73443	,79810	,73956	,80007	,73654
	r	,91377	,87845	,92202	,83564	,88988	,83901	,87679	,84565	,86692	,86044	,83492	,81934	,83460	,80172	,80564	,78733	,79500	,77687	,79096	,75036
	fl	,89288	,78452	,89992	,77513	,87318	,78835	,85449	,79769	,84187	,80603	,81939	,77941	,82666	,77343	,80352	,75943	,79616	,75741	,79504	,74187
<i>XGB</i>	ac	,89614	,76532	,89357	,76712	,87657	,77807	,84938	,78239	,84358	,78190	,81930	,76555	,81253	,76384	,80734	,75115	,80221	,74319	,79463	,73440
	auc	,94664	,81122	,94680	,83064	,93358	,84578	,90629	,83931	,90856	,83953	,89886	,82026	,88420	,82301	,87870	,81156	,86712	,80853	,86557	,79642
	k	,79227	,53056	,78715	,53427	,75313	,55614	,69875	,56477	,68721	,56385	,63864	,53114	,62505	,52763	,61471	,50237	,60440	,48638	,58926	,46886
	p	,87270	,70934	,87362	,72463	,86215	,73623	,83185	,75004	,82066	,74810	,80755	,73483	,79966	,73803	,79882	,73736	,79828	,72999	,79620	,73543
	r	,92832	,90134	,92202	,86301	,89725	,86655	,87787	,84774	,88094	,85393	,83932	,83151	,83461	,81751	,82365	,78015	,80930	,77302	,79386	,73729
	fl	,89938	,79362	,89674	,78747	,87915	,79595	,85365	,79534	,84939	,79688	,82250	,77987	,81664	,77538	,81061	,75784	,80351	,75055	,79461	,73470
<i>MLP</i>	ac	,88314	,76066	,88463	,76343	,85912	,78073	,83815	,77761	,82572	,78028	,80820	,76828	,80804	,76270	,79471	,75842	,76653	,75421	,78231	,74747
	auc	,93476	,81207	,93398	,82254	,92438	,84051	,89977	,83008	,89128	,83750	,87557	,81786	,86689	,82331	,85550	,81695	,82990	,81229	,83106	,81298
	k	,76628	,52124	,76927	,52690	,71823	,56145	,67625	,55520	,65148	,56065	,61641	,53659	,61605	,52539	,58944	,51684	,53308	,50842	,56466	,49500
	p	,86409	,70354	,86628	,71612	,84198	,74018	,80630	,73734	,81177	,75796	,79085	,73997	,79917	,73887	,77900	,74282	,75859	,73033	,78111	,74459
	r	,91069	,90234	,91043	,87777	,88450	,86760	,89179	,86284	,85057	,82795	,83930	,82928	,82446	,81427	,82496	,78975	,78341	,80542	,78953	,75767
	fl	,88635	,79042	,88743	,78794	,86256	,79802	,84655	,79486	,83017	,79067	,81391	,78123	,81088	,77411	,80029	,76490	,77038	,76577	,78360	,75010

**APÊNDICE C – TODOS OS RESULTADOS DE VALIDAÇÃO CRUZADA DOS MODELOS  
NAS AMOSTRAS DE JANELAS VARIÁVEIS**

Modelo		Sem seleção de atributos de dados		Com seleção de atributos de dados	
		Tese	Baseline	Tese	Baseline
<i>LR</i>	ac	,82085	,80946	,82655	,80946
	auc	,90772	,88846	,90744	,88846
	k	,64172	,61894	,65313	,61894
	p	,83129	,85713	,83353	,85713
	r	,80586	,74357	,81621	,74357
	f1	,81789	,79573	,82449	,79573
<i>SVM</i>	ac	,85150	,82971	,85201	,82971
	auc	,92020	,90755	,92131	,90755
	k	,70299	,65945	,70403	,65945
	p	,85373	,86812	,85305	,86812
	r	,84838	,77788	,85049	,77788
	f1	,85091	,82001	,85157	,82001
<i>NB</i>	ac	,71339	,66668	,71598	,66668
	auc	,85908	,85288	,86183	,85288
	k	,42676	,33327	,43197	,33327
	p	,84291	,90226	,85094	,90226
	r	,52538	,37483	,52539	,37483
	f1	,64622	,52749	,64831	,52749
<i>k-NN</i>	ac	,83694	,84115	,84006	,84115
	auc	,91527	,90209	,91600	,90209
	k	,67392	,68231	,68015	,68231
	p	,83268	,85202	,83463	,85202
	r	,84634	,82662	,85049	,82662
	f1	,83863	,83882	,84175	,83882
<i>SGD</i>	ac	,83436	,81310	,83489	,81310
	auc	,90714	,85837	,90869	,85837

	k	,66874	,62621	,66977	,62621
	p	,84730	,80960	,84986	,80960
	r	,81621	,81937	,81411	,81937
	f1	,83121	,81397	,83134	,81397
<i>DT</i>	ac	,85878	,83694	,85565	,83694
	auc	,89444	,88093	,89457	,88093
	k	,71762	,67395	,71136	,67395
	p	,84291	,81395	,83510	,81395
	r	,88480	,87651	,88895	,87651
	f1	,86258	,84319	,86054	,84319
<i>AB</i>	ac	,85670	,85152	,85669	,85152
	auc	,92216	,91543	,92204	,91543
	k	,71343	,70306	,71343	,70306
	p	,83179	,83209	,83246	,83209
	r	,89623	,88168	,89518	,88168
	f1	,86241	,85592	,86228	,85592
<i>RF</i>	ac	,87745	,86709	,87796	,86709
	auc	,94433	,93583	,94435	,93583
	k	,75491	,73422	,75596	,73422
	p	,86285	,84382	,86026	,84382
	r	,89825	,90245	,90346	,90245
	f1	,88002	,87169	,88109	,87169
<i>XGB</i>	ac	,87850	,86762	,88161	,86762
	auc	,93770	,92886	,93964	,92886
	k	,75703	,73526	,76324	,73526
	p	,86478	,85027	,86710	,85027
	r	,89827	,89309	,90243	,89309
	f1	,88093	,87084	,88414	,87084
<i>MLP</i>	ac	,86085	,85515	,86135	,85515
	auc	,92052	,91356	,92555	,91356
	k	,72168	,71032	,72270	,71032
	p	,85892	,85621	,85712	,85621
	r	,86396	,85469	,86810	,85469
	f1	,86076	,85495	,86223	,85495

---

## APÊNDICE D – OCORRÊNCIAS DOS ATRIBUTOS DE DADOS SELECIONADOS NOS MODELOS DA TESE

Atributos de dados	Quantidade de janelas em que o atributo foi selecionado	Percentual de seleção nas janelas
Desvio p(Correto) nos CCs do exercício	11	1
Desvio p(Correto) nos CCs primários do exercício	11	1
Desvio p(F) nos CCs do exercício	11	1
Desvio p(G) nos CCs do exercício	11	1
Desvio p(G) nos CCs primários do exercício	11	1
Desvio p(L) nos CCs do exercício	11	1
Desvio p(L) nos CCs primários do exercício	11	1
Desvio p(L <sub>0</sub> ) nos CCs do exercício	11	1
Desvio p(T) nos CCs do exercício	11	1
Média p(Correto) nos CCs do exercício	11	1
Média p(Correto) nos CCs do objetivo	11	1
Média p(Correto) nos CCs primários do exercício	11	1
Média p(F) nos CCs do exercício	11	1
Média p(F) nos CCs do objetivo	11	1
Média p(G) nos CCs do objetivo	11	1
Média p(G) nos CCs primários do exercício	11	1
Média p(L) nos CCs do exercício	11	1
Média p(L) nos CCs do objetivo	11	1
Média p(L) nos CCs primários do exercício	11	1
Média p(L <sub>0</sub> ) nos CCs do exercício	11	1
Média p(L <sub>0</sub> ) nos CCs do objetivo	11	1
Média p(S) nos CCs do exercício	11	1
Média p(S) nos CCs do objetivo	11	1
Média p(T) nos CCs do exercício	11	1
Média p(T) nos CCs do objetivo	11	1
Média p(T) nos CCs primários do exercício	11	1
p(Correto) no CC primário do objetivo	11	1
p(F) no CC primário do objetivo	11	1
p(G) no CC primário do objetivo	11	1
p(L) no CC primário do objetivo	11	1
p(L <sub>0</sub> ) no CC primário do objetivo	11	1

p(S) no CC primário do objetivo	11	1
p(T) no CC primário do objetivo	11	1
Quantidade de cliques de mouse no código	11	1
Quantidade de erros de sintaxe no código	11	1
Quantidade de intervalos de digitação	11	1
Quantidade de movimentos de mouse	11	1
Quantidade de teclas <i>backspace</i>	11	1
Quantidade de teclas de controle	11	1
Quantidade de teclas digitadas	11	1
Tempo ocioso	11	1
Tempo total de movimentos de mouse	11	1
Desvio p(F) nos CCs primários do exercício	10	,90909
Desvio p(L <sub>0</sub> ) nos CCs primários do exercício	10	,90909
Desvio p(S) nos CCs do exercício	10	,90909
Desvio p(T) nos CCs primários do exercício	10	,90909
Média p(F) nos CCs primários do exercício	10	,90909
Média p(G) nos CCs do exercício	10	,90909
Média p(L <sub>0</sub> ) nos CCs primários do exercício	10	,90909
Quantidade de submissões	10	,90909
Quantidade de submissões com erro	10	,90909
Velocidade de digitação alta	10	,90909
Velocidade de digitação média	10	,90909
Velocidade de movimentos de mouse alta	10	,90909
Velocidade de movimentos de mouse média	10	,90909
Desvio p(S) nos CCs primários do exercício	9	,81818
Média p(S) nos CCs primários do exercício	9	,81818
Quantidade de aberturas de tela de enunciados de objetivos	9	,81818
Quantidade de submissões com erro de objetivo de exercício	9	,81818
Quantidade de intervalos ociosos	8	,72727
Quantidade de submissões com sucesso	8	,72727
Quantidade de submissões com erro de compilação	6	,54545
Quantidade de teclas <i>delete</i>	6	,54545
Velocidade de digitação alta	4	,36363
Velocidade de movimentos de mouse baixa	2	,18181

**APÊNDICE E – OCORRÊNCIAS DOS ATRIBUTOS DE DADOS SELECIONADOS NOS MODELOS *BASELINE***

<b>Atributos de dados</b>	<b>Quantidade de janelas em que o atributo foi selecionado</b>	<b>Percentual de seleção nas janelas</b>
Quantidade de cliques de mouse no código	11	1
Quantidade de erros de sintaxe no código	11	1
Quantidade de intervalos de digitação	11	1
Quantidade de teclas <i>backspace</i>	11	1
Quantidade de teclas digitadas	11	1
Velocidade de movimentos de mouse alta	11	1
Quantidade de submissões	10	,909091
Quantidade de submissões com erro	10	,909091
Quantidade de submissões com erro de objetivo de exercício	10	,909091
Quantidade de submissões com sucesso	10	,909091
Tempo total de movimentos de mouse	10	,909091
Velocidade de digitação alta	10	,909091
Velocidade de movimentos de mouse média	10	,909091
Quantidade de intervalos ociosos	9	,818182
Quantidade de movimentos de mouse	9	,818182
Tempo ocioso	9	,818182
Velocidade de digitação média	9	,818182
Quantidade de teclas de controle	8	,727273
Velocidade de movimentos de mouse baixa	8	,727273
Quantidade de abertura de tela de enunciados de objetivos	7	,636364
Quantidade de submissões com erro de compilação	6	,545455
Quantidade de teclas <i>delete</i>	6	,545455
Velocidade de digitação baixa	6	,545455
Desvio p(Correto) nos CCs do exercício	0	0
Desvio p(Correto) nos CCs primários do exercício	0	0
Desvio p(F) nos CCs do exercício	0	0
Desvio p(F) Nos CCs primários do exercício	0	0
Desvio p(G) nos CCs do exercício	0	0
Desvio p(G) nos CCs primários do exercício	0	0
Desvio p(L) nos CCs do exercício	0	0
Desvio p(L) nos CCs primários do exercício	0	0
Desvio p(L <sub>0</sub> ) nos CCs do exercício	0	0
Desvio p(L <sub>0</sub> ) nos CCs primários do exercício	0	0

Desvio p(S) nos CCs do exercício	0	0
Desvio p(S) nos CCs primários do exercício	0	0
Desvio p(T) nos CCs do exercício	0	0
Desvio p(T) nos CCs primários do exercício	0	0
Média p(Correto) nos CCs do exercício	0	0
Média p(Correto) nos CCs do objetivo	0	0
Média p(Correto) nos CCs primários do exercício	0	0
Média p(F) nos CCs do exercício	0	0
Média p(F) nos CCs do objetivo	0	0
Média p(F) nos CCs primários do exercício	0	0
Média p(G) nos CCs do exercício	0	0
Média p(G) nos CCs do objetivo	0	0
Média p(G) nos CCs primários do exercício	0	0
Média p(L) nos CCs do exercício	0	0
Média p(L) nos CCs do objetivo	0	0
Média p(L) nos CCs primários do exercício	0	0
Média p(L <sub>0</sub> ) nos CCs do exercício	0	0
Média p(L <sub>0</sub> ) nos CCs do objetivo	0	0
Média p(L <sub>0</sub> ) nos CCs primários do exercício	0	0
Média p(S) nos CCs do exercício	0	0
Média p(S) nos CCs do objetivo	0	0
Média p(S) nos CCs primários do exercício	0	0
Média p(T) nos CCs do exercício	0	0
Média p(T) nos CCs do objetivo	0	0
Média p(T) nos CCs primários do exercício	0	0
p(Correto) no CC primário do objetivo	0	0
p(F) no CC primário do objetivo	0	0
p(G) no CC primário do objetivo	0	0
p(L) no CC primário do objetivo	0	0
P(L <sub>0</sub> ) no CC primário do objetivo	0	0
p(S) no CC primário do objetivo	0	0
p(T) no CC primário do objetivo	0	0

**APÊNDICE F – ATRIBUTOS DE DADOS RELEVANTES NAS AMOSTRAS DA TESE**

<b>Atributo – Janela de 5 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 10 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 20 segundos</b>	<b>Escore</b>
Quantidade de teclas digitadas	,10430	Quantidade de teclas digitadas	,02706	Média de p(L) nos CCs do objetivo	,15068
Média p(L) nos CCS do objetivo	,06972	Quantidade de intervalos de digitação	,01125	Quantidade de teclas digitadas	,07981
Tempo total de movimentos de mouse	,00910	Média de p(L) nos CCs do objetivo	,00974	Quantidade de movimentos de mouse	,01299
Média de p(L) CCs primários exercício	,00417	p(Correto) no CC primário do objetivo	,00470	Média de p(L) nos CCs do exercício	,01229
Quantidade de cliques no código	,00411	Tempo total de movimentos de mouse	,00449	Quantidade de intervalos de digitação	,01201
p(Correto) no CC primário do objetivo	,00312	Média p(Correto) nos CCs do objetivo	,00369	Média p(Correto) CCs primários exercício	,01184
Desvio p(Correto) CCs primários exerc.	,00297	Velocidade de digitação alta	,00367	Quantidade de erros de sintaxe no código	,01159
Desvio p(Correto) nos CCs do exercício	,00222	Desvio p(L <sub>0</sub> ) CCs primários do exercício	,00344	p(T) no CC primário do objetivo	,00655
p(L) no CC primário do objetivo	,00218	Desvio p(Forget) nos CCs do exercício	,00325	Média p(Correto) nos CCs do exercício	,00625
Média p(Correto) nos CCs do exercício	,00195	p(L) no CC primário do objetivo	,00325	Média p(Forget) nos CCs do objetivo	,00555
<b>Atributo – Janela de 40 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 60 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 90 segundos</b>	<b>Escore</b>
Quantidade de teclas digitadas	,09661	Quantidade de teclas digitadas	,09758	Quantidade de teclas digitadas	,11776
Média de p(L) nos CCs do objetivo	,08815	Média de p(L) nos CCs do objetivo	,06413	Média de p(L) nos CCs do objetivo	,05235
Quantidade de intervalos de digitação	,02217	Tempo total de movimentos de mouse	,01299	Tempo ocioso	,01415
p(L) no CC primário do objetivo	,01899	Média p(Correto) nos CCs do exercício	,01045	Desvio p(L) nos CCs primários exercício	,01217
Tempo total de movimentos de mouse	,01567	Quantidade de intervalos de digitação	,00848	Desvio p(Correto) CCs primários exerc.	,01047
p(Correto) no CC primário do objetivo	,01293	Quantidade de submissões com erro	,00719	p(L) no CC primário do objetivo	,00973
Média p(Correto) nos CCs do exercício	,00860	Média de p(L) nos CCs do exercício	,00710	Quantidade de erros de sintaxe no código	,00965
Média de p(L) nos CCs do exercício	,00830	p(Correto) no CC primário do objetivo	,00690	Quantidade de intervalos de digitação	,00942
Média p(Correto) nos CCs do objetivo	,00557	Desvio p(Correto) CCs primários exerc.	,00654	Tempo total de movimentos de mouse	,00936
Quantidade de movimentos de mouse	,00523	p(L) no CC primário do objetivo	,00448	Quantidade de teclas <i>backspace</i>	,00902
<b>Atributo – Janela de 120 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 180 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 240 segundos</b>	<b>Escore</b>
Quantidade de teclas digitadas	,02147	Quantidade de teclas digitadas	,06191	Quantidade de teclas digitadas	,05380
Quantidade de movimentos de mouse	,00913	Média de p(L) nos CCs do objetivo	,04691	Média de p(L) nos CCs do objetivo	,01777
Tempo total de movimentos de mouse	,00673	Quantidade de movimentos de mouse	,02703	Quantidade de movimentos de mouse	,01741
Tempo ocioso	,00553	Quantidade de submissões com erro	,01802	Tempo total de movimentos de mouse	,01201
Quantidade de teclas <i>backspace</i>	,00442	Média de p(G) nos CCs do objetivo	,01169	Quantidade de erros de sintaxe no código	,01183
Média de p(L <sub>0</sub> ) nos CCs do exercício	,00367	Média p(Correto) nos CCs do exercício	,01164	Quantidade de submissões com erro	,01118
Média p(Correto) nos CCs do exercício	,00335	Quantidade de cliques no código	,00964	Quantidade de intervalos de digitação	,00934
Média p(S) CCs primários exercício	,00335	Média p(L) nos CCs primários exercício	,00932	Média de p(L) nos CCs do exercício	,00833
p(L <sub>0</sub> ) no CC primário do objetivo	,00322	Quantidade de intervalos de digitação	,00853	Média p(Correto) nos CCs do objetivo	,00695
Desvio p(L <sub>0</sub> ) CCs primários exercício	,00310	p(L) no CC primário do objetivo	,00703	Quantidade de intervalos ociosos	,00615

<b>Atributo – Janela de 360 segundos</b>	<b>Escore</b>	<b>Atributo – Janela variável</b>	<b>Escore</b>
Quantidade de teclas digitadas	,01643	Média de p(L) nos CCs do objetivo	,02042
Quantidade de intervalos de digitação	,01129	Quantidade de intervalos ociosos	,01493
Média p(Correto) CCs primários exercício	,00917	Tempo ocioso	,00918
Média p(L) nos CCs primários exercício	,00859	Quantidade de movimentos de mouse	,00831
Desvio p(L <sub>0</sub> ) nos CCs do exercício	,00731	Tempo total de movimentos de mouse	,00517
Quantidade de submissões com erro	,00717	Quantidade de erros de sintaxe no código	,00498
Quantidade de intervalos ociosos	,00671	p(L) no CC primário do objetivo	,00455
Média de p(L) nos CCs do objetivo	,00662	Quantidade de submissões com sucesso	,00239
Quantidade de erros de sintaxe no código	,00636	Média de p(G) nos CCs do objetivo	,00197
Tempo ocioso	,00633	Média de p(T) nos CCs do exercício	,00137

**APÊNDICE G – ATRIBUTOS DE DADOS RELEVANTES NAS AMOSTRAS BASELINE**

<b>Atributo – Janela de 5 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 10 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 20 segundos</b>	<b>Escore</b>
Velocidade movimentos mouse média	,07562	Quantidade de teclas digitadas	,11547	Quantidade de teclas digitadas	,12913
Quantidade de teclas digitadas	,07080	Tempo total de movimentos de mouse	,05028	Velocidade movimentos mouse alta	,08324
Velocidade movimentos mouse alta	,06706	Quantidade de intervalos de digitação	,02723	Velocidade movimentos mouse média	,07180
Velocidade movimentos mouse baixa	,02368	Quantidade de submissões com sucesso	,01581	Quantidade de intervalos de digitação	,03347
Quantidade de intervalos de digitação	,01844	Quantidade de subm. com erro de objet.	,01096	Quantidade de submissões com sucesso	,01691
Quantidade de submissões	,01477	Quantidade de aberturas de enunciados	,01005	Velocidade de digitação média	,01576
Tempo ocioso	,00976	Quantidade de erros de sintaxe no código	,00856	Quantidade de subm. com erro de objet.	,01390
Quantidade de aberturas de enunciados	,00968	Velocidade de digitação alta	,00301	Velocidade de digitação alta	,01324
Velocidade de digitação alta	,00879	Velocidade movimentos mouse média	,00259	Quantidade de submissões com erro	,01072
Quantidade de movimentos de mouse	,00752	Quantidade de submissões com erro	,00207	Quantidade de cliques no código	,01057
<b>Atributo – Janela de 40 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 60 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 90 segundos</b>	<b>Escore</b>
Quantidade de teclas digitadas	,09730	Quantidade de teclas digitadas	,11662	Quantidade de teclas digitadas	,08388
Quantidade de intervalos de digitação	,04004	Quantidade de intervalos de digitação	,03113	Quantidade de intervalos de digitação	,02902
Tempo total de movimentos de mouse	,01503	Tempo total de movimentos de mouse	,01950	Quantidade de cliques no código	,01313
Quantidade de submissões com sucesso	,01068	Quantidade de movimentos de mouse	,01565	Velocidade movimentos mouse alta	,00847
Quantidade de teclas backspace	,00643	Quantidade de submissões com erro	,01301	Velocidade de digitação alta	,00685
Velocidade de digitação alta	,00448	Quantidade de teclas de controle	,00885	Quantidade de teclas de controle	,00659
Quantidade de subm. com erro de objet.	,00405	Quantidade de erros de sintaxe no código	,00877	Quantidade de submissões com sucesso	,00401
Velocidade movimentos mouse média	,00274	Velocidade de digitação alta	,00807	Quantidade de aberturas de enunciados	,00275
Velocidade de digitação média	,00208	Tempo ocioso	,00662	Quantidade de teclas <i>delete</i>	,00184
Quantidade de submissões	,00176	Quantidade de intervalos ociosos	,00626	Quantidade de submissões com erro	,00171
<b>Atributo – Janela de 120 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 180 segundos</b>	<b>Escore</b>	<b>Atributo – Janela de 240 segundos</b>	<b>Escore</b>
Quantidade de teclas digitadas	,06677	Quantidade de teclas digitadas	,06708	Quantidade de teclas digitadas	,12903
Quantidade de intervalos de digitação	,03511	Quantidade de intervalos de digitação	,03205	Velocidade movimentos mouse alta	,05357
Tempo total de movimentos de mouse	,01450	Quantidade de submissões com erro	,02377	Quantidade de intervalos de digitação	,04402
Quantidade de cliques no código	,00795	Quantidade de teclas <i>backspace</i>	,02157	Velocidade movimentos mouse média	,02620
Quantidade de submissões com erro	,00779	Velocidade de digitação média	,01662	Quantidade de submissões com erro	,01888
Quantidade de movimentos de mouse	,00716	Quantidade de cliques no código	,01514	Quantidade de movimentos de mouse	,01738
Quantidade de intervalos ociosos	,00589	Velocidade de digitação alta	,01406	Quantidade de teclas <i>backspace</i>	,01606
Velocidade de digitação alta	,00458	Tempo total de movimentos de mouse	,01336	Tempo total de movimentos de mouse	,01219
Quantidade de submissões	,00211	Velocidade movimentos mouse alta	,01036	Quantidade de intervalos ociosos	,01211
Quantidade de subm. com erro de objet.	,00174	Quantidade de movimentos de mouse	,01017	Quantidade de erros de sintaxe no código	,01160

<b>Atributo – Janela de 360 segundos</b>	<b>Escore</b>	<b>Atributo – Janela variável</b>	<b>Escore</b>
Velocidade movimentos mouse alta	,05277	Tempo ocioso	,03470
Quantidade de teclas digitadas	,04761	Quantidade de intervalos ociosos	,01780
Quantidade de movimentos de mouse	,04731	Quantidade de submissões com erro	,01597
Quantidade de cliques no código	,04229	Quantidade de submissões com sucesso	,01090
Quantidade de intervalos ociosos	,03353	Quantidade de movimentos de mouse	,00881
Velocidade movimentos mouse média	,02758	Quantidade de teclas de controle	,00731
Quantidade de intervalos de digitação	,02639	Tempo total de movimentos de mouse	,00640
Quantidade de submissões com erro	,02017	Quantidade de teclas digitadas	,00509
Quantidade de submissões	,01878	Quantidade de erros de sintaxe no código	,00399
Quantidade de teclas <i>backspace</i>	,01808	Velocidade movimentos mouse baixa	,00307

**APÊNDICE H – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS - TREINAMENTO COM AMOSTRAS  
DO TÉCNICO E TESTES COM AMOSTRAS DO ENSINO SUPERIOR**

Modelo		Resultados de acurácia (ac), AUC ROC (auc), kappa (k), precisão (p), recall (r) e F1-score (f1) dos modelos da Tese (T) e baseline (B) nas amostras de janelas fixas (5s, 10s, 20s, 40s, 60s, 90s, 120s, 180s, 240s e 360s)																			
		T 5	B 5	T 10	B 10	T 20	B 20	T 40	B 40	T 60	B 60	T 90	B 90	T 120	B 120	T 180	B 180	T 240	B 240	T 360	B 360
<i>LR</i>	ac	,74849	,67404	,81350	,67540	,75506	,71153	,63422	,73463	,70867	,69214	,52754	,73199	,64086	,71505	,62844	,71674	,55797	,68478	,54935	,71428
	auc	,82638	,70761	,88858	,72031	,82096	,76021	,80715	,78125	,76576	,77487	,69291	,78220	,74045	,77919	,70320	,79833	,65962	,77340	,69151	,78919
	k	,49698	,34808	,62701	,35080	,51012	,42307	,26844	,46926	,41735	,38429	,05508	,46398	,28172	,43010	,25688	,43348	,11594	,36956	,09870	,42857
	p	,69510	,61736	,78849	,62393	,72992	,65761	,82587	,67642	,72645	,63798	,51416	,68159	,73144	,66286	,69580	,66965	,74489	,64140	,52646	,67223
	r	,88531	,91549	,85685	,88306	,80971	,88259	,34016	,89959	,66942	,88843	1,0000	,87076	,44516	,87526	,45642	,85550	,17632	,83816	,98181	,83636
	f1	,77876	,73743	,82125	,73121	,76775	,75367	,48185	,77220	,69677	,74266	,67913	,76465	,55347	,75440	,55124	,75125	,28515	,72670	,68540	,74537
<i>SVM</i>	ac	,72434	,65895	,69052	,66633	,75303	,69433	,69467	,71926	,68595	,69731	,62288	,70550	,61612	,71505	,39908	,68119	,61835	,68599	,38571	,68051
	auc	,72101	,71994	,69719	,70955	,81278	,73394	,77648	,76121	,76168	,77843	,73760	,77582	,68925	,77988	,36694	,73738	,69194	,73736	,34233	,72812
	k	,44869	,31790	,38104	,33266	,50607	,38866	,38934	,43852	,37190	,39462	,24576	,41101	,23225	,43010	-,2018	,36238	,23671	,37198	-,2285	,36103
	p	,71820	,60646	,67027	,61836	,76824	,65635	,73399	,67313	,66483	,64317	,67575	,67896	,62272	,66025	,41881	,67555	,62760	,65098	,40557	,66992
	r	,73843	,90543	,75000	,86895	,72469	,81578	,61065	,85245	,75000	,88636	,47245	,77966	,58924	,88602	,52064	,69724	,58212	,80193	,49090	,71168
	f1	,72817	,72639	,70789	,72254	,74583	,72743	,66666	,75226	,70485	,74543	,55611	,72583	,60552	,75665	,46421	,68623	,60401	,71861	,44418	,69017
<i>NB</i>	ac	,73843	,64084	,74395	,64616	,70141	,65283	,63627	,57786	,69628	,64876	,67055	,58898	,70537	,63655	,69151	,61009	,67874	,56401	,65064	,58051
	auc	,81280	,70504	,81634	,64758	,76235	,68152	,69077	,70352	,74373	,73348	,71316	,69294	,74319	,74145	,73605	,73951	,72401	,70485	,68706	,70087
	k	,47686	,28169	,48790	,29233	,40283	,30566	,27254	,15573	,39256	,29752	,34110	,17796	,41075	,27311	,38302	,22018	,35748	,12801	,30129	,16103
	p	,81432	,59090	,75527	,60225	,73747	,60107	,69387	,54714	,73750	,60526	,70277	,55645	,72470	,58956	,69195	,56837	,78461	,53925	,68709	,54936
	r	,61770	,91549	,72177	,86088	,62550	,90890	,48770	,90368	,60950	,85537	,59110	,87711	,66236	,89892	,69036	,91513	,49275	,87922	,55324	,89610
	f1	,70251	,71823	,73814	,70871	,67688	,72361	,57280	,68160	,66742	,70890	,64211	,68092	,69213	,71209	,69116	,70123	,60534	,66850	,61295	,68114
<i>k-NN</i>	ac	,81187	,66398	,79536	,66633	,73987	,67510	,69467	,69159	,67562	,68905	,67796	,68961	,64946	,69892	,62729	,66169	,65579	,65700	,57532	,62467
	auc	,87262	,70318	,85724	,69156	,79381	,72884	,74839	,73723	,71656	,72325	,69029	,74099	,67757	,74484	,68743	,73043	,69134	,70629	,60636	,66555
	k	,62374	,32796	,59072	,33266	,47975	,35020	,38934	,38319	,35124	,37809	,35593	,37923	,29892	,39784	,25458	,32339	,31159	,31401	,15064	,24935
	p	,75833	,61366	,76588	,61939	,71663	,63246	,69076	,66317	,66666	,65326	,67213	,66791	,65831	,68316	,63979	,66059	,68587	,64192	,58333	,62698
	r	,91549	,88531	,85080	,86290	,79352	,83603	,70491	,77868	,70247	,80578	,69491	,75423	,62150	,74193	,58256	,66513	,57487	,71014	,52727	,61558
	f1	,82953	,72487	,80611	,72114	,75312	,72013	,69776	,71630	,68410	,72155	,68333	,70845	,63938	,71134	,60984	,66285	,62549	,67431	,55388	,62123
<i>SGD</i>	ac	,84104	,67002	,83165	,66935	,79149	,71052	,72233	,72848	,73966	,69214	,71928	,72563	,58709	,71290	,71559	,70298	,69686	,68236	,67272	,71039
	auc	,89398	,70575	,89536	,71380	,85317	,76608	,81579	,77005	,80102	,77489	,76350	,78183	,69452	,77959	,78614	,78658	,78550	,77147	,73214	,78456

	k	.68209	.34004	.66330	.33871	.58299	.42105	.44467	.45696	.47933	.38429	.43855	.45127	.17419	.42580	.43119	.40596	.39372	.36473	.34545	.42077
	p	.80874	.61559	.82191	.62069	.76966	.65430	.80911	.67726	.71886	.63798	.70741	.67720	.69756	.66019	.69183	.65888	.67833	.63802	.62337	.67382
	r	.89336	.90543	.84677	.87096	.83198	.89271	.58196	.87295	.78719	.88843	.74788	.86228	.30752	.87741	.77752	.84174	.74879	.84299	.87272	.81558
	fl	.84894	.73289	.83416	.72483	.79961	.75513	.67699	.76275	.75147	.74266	.72708	.75862	.42686	.75346	.73218	.73917	.71182	.72632	.72727	.73795
<i>DT</i>	ac	.67605	.64889	.75403	.66129	.69736	.65789	.67315	.65471	.65909	.70351	.49364	.61652	.62365	.70107	.62156	.63761	.52415	.67874	.50779	.65324
	auc	.78403	.65752	.76866	.70163	.76764	.67842	.63100	.70644	.73761	.72290	.51076	.68680	.61774	.74845	.62156	.71936	.51573	.71555	.50954	.69822
	k	.35211	.29778	.50806	.32258	.39473	.31578	.34631	.30942	.31818	.40702	-.0127	.23305	.24731	.40215	.24311	.27522	.04830	.35748	.01558	.30649
	p	.61776	.59610	.71575	.61976	.70700	.62662	.65116	.63702	.62581	.65810	.48897	.58957	.59410	.67608	.66772	.64925	.52538	.63703	.50903	.64676
	r	.92354	.92354	.84274	.83467	.67408	.78137	.74590	.71926	.79132	.84710	.28178	.76694	.78064	.77204	.48394	.59862	.50000	.83091	.43896	.67532
<i>AB</i>	fl	.74032	.72454	.77407	.71134	.69015	.69549	.69532	.67565	.69890	.74074	.35752	.66666	.67472	.72088	.56117	.62291	.51237	.72117	.47140	.66073
	ac	.75754	.68511	.81754	.68346	.76012	.69939	.65778	.70389	.72830	.68905	.67478	.68644	.71075	.71290	.68004	.68233	.64734	.70772	.62077	.68441
	auc	.84913	.70633	.87966	.72073	.79519	.72483	.75432	.77159	.78755	.73538	.73816	.74186	.78547	.77822	.75179	.77813	.75338	.78245	.69691	.75245
	k	.51509	.37022	.63508	.36693	.52024	.39878	.31557	.40778	.45661	.37809	.34957	.37288	.42150	.42580	.36009	.36467	.29468	.41545	.24155	.36883
	p	.80331	.63068	.82340	.62569	.71816	.64506	.67500	.64828	.71373	.63436	.63865	.63836	.81818	.66071	.77351	.64120	.61685	.66226	.70129	.65368
	r	.68209	.89336	.80846	.91330	.85627	.88664	.60860	.89139	.76239	.89256	.80508	.86016	.54193	.87526	.50917	.82798	.77777	.84782	.42077	.78441
<i>RF</i>	fl	.73775	.73938	.81587	.74262	.78116	.74680	.64008	.75064	.73726	.74163	.71227	.73285	.65200	.75300	.61410	.72272	.68803	.74364	.52597	.71310
	ac	.82696	.68008	.82358	.67540	.75303	.69129	.73565	.73360	.71797	.72417	.71716	.71292	.66666	.73010	.68119	.72362	.68357	.69927	.62337	.69090
	auc	.89730	.73062	.88769	.71107	.82427	.73948	.78676	.78354	.76692	.76751	.75505	.76350	.74234	.79596	.72697	.78959	.72538	.76826	.68887	.75492
	k	.65392	.36016	.64717	.35080	.50607	.38259	.47131	.46721	.43595	.44834	.43432	.42584	.33333	.46021	.36238	.44724	.36715	.39855	.24675	.38181
	p	.82828	.62482	.82424	.63343	.74414	.65023	.76497	.68566	.71574	.66979	.74346	.66722	.75244	.68706	.75483	.69306	.74836	.66401	.66101	.66974
	r	.82495	.90140	.82258	.83266	.77125	.82793	.68032	.86270	.72314	.88429	.66313	.84957	.49677	.84516	.53669	.80275	.55314	.80676	.50649	.75324
<i>XGB</i>	fl	.82661	.73805	.82341	.71951	.75745	.72840	.72017	.76406	.71942	.76224	.70100	.74743	.59844	.75795	.62734	.74388	.63611	.72846	.57352	.70904
	ac	.78772	.67404	.77923	.68044	.75809	.69736	.71413	.71516	.72417	.70454	.70550	.69491	.66881	.72043	.68578	.71674	.65821	.69686	.65194	.68441
	auc	.88469	.70743	.84763	.72055	.80544	.73043	.76516	.75548	.78037	.76049	.76131	.74710	.73073	.79313	.74197	.79760	.73028	.74709	.70000	.75477
	k	.57545	.34808	.55846	.36088	.51619	.39473	.42827	.43032	.44834	.40909	.41101	.38983	.33763	.44086	.37156	.43348	.31642	.39372	.30389	.36883
	p	.81919	.61736	.74513	.63027	.73224	.64840	.70941	.66826	.70905	.65517	.68945	.64330	.65858	.66998	.67016	.66965	.63617	.65886	.64096	.65502
	r	.73843	.91549	.84879	.87298	.81376	.86234	.72541	.85450	.76033	.86363	.74788	.87500	.70107	.86881	.73165	.85550	.73913	.81642	.69090	.77922
<i>MLP</i>	fl	.77672	.73743	.79359	.73203	.77085	.74022	.71732	.75000	.73379	.74509	.71748	.74147	.67916	.75655	.69956	.75125	.68379	.72923	.66500	.71174
	ac	.67706	.66599	.81451	.67842	.77732	.69838	.70696	.70696	.71694	.70661	.67055	.71716	.61075	.72258	.67660	.70986	.54589	.68961	.63506	.71948
	auc	.85316	.72064	.88212	.72818	.84924	.73644	.77590	.74552	.79974	.78097	.71367	.74790	.72794	.78878	.77448	.76802	.69422	.74073	.68475	.77999
	k	.35412	.33199	.62903	.35685	.55465	.39676	.41393	.41393	.43388	.41322	.34110	.43432	.22150	.44516	.35321	.41972	.09178	.37922	.27013	.43896
	p	.89639	.61073	.83766	.62697	.77400	.65264	.72345	.66031	.79329	.65290	.66947	.66942	.75369	.66720	.80078	.67362	.73170	.67252	.65028	.68903
	r	.40040	.91549	.78024	.88104	.78340	.84817	.67008	.85245	.58677	.88223	.67372	.85805	.32903	.88817	.47018	.81422	.14492	.73913	.58441	.80000
	fl	.55354	.73268	.80793	.73260	.77867	.73767	.69574	.74418	.67458	.75043	.67159	.75208	.45808	.76199	.59248	.73727	.24193	.70425	.61559	.74038

**APÊNDICE I – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS - TREINAMENTO COM AMOSTRAS DO TÉCNICO E TESTES COM AMOSTRAS DO ENSINO SUPERIOR – COM ATRIBUTOS DE DADOS SELECIONADOS**

Modelo		Resultados de acurácia (ac), AUC ROC (auc), kappa (k), precisão (p), recall (r) e F1-score (f1) dos modelos da Tese (T) e baseline (B) nas amostras de janelas fixas (5s, 10s, 20s, 40s, 60s, 90s, 120s, 180s, 240s e 360s)																			
		T 5	B 5	T 10	B 10	T 20	B 20	T 40	B 40	T 60	B 60	T 90	B 90	T 120	B 120	T 180	B 180	T 240	B 240	T 360	B 360
<i>LR</i>	ac	,80382	,65593	,81350	,67943	,74493	,70850	,57582	,73565	,71694	,70867	,75317	,73199	,65268	,72473	,70986	,71559	,65458	,68478	,63376	,71948
	auc	,88648	,71489	,90030	,68851	,81228	,76030	,81453	,78487	,75911	,77928	,81277	,78382	,78352	,78682	,80832	,79931	,80237	,76060	,72429	,78764
	k	,60764	,31187	,62701	,35887	,48987	,41700	,15163	,47131	,43388	,41735	,50635	,46398	,30537	,44946	,41972	,43119	,30917	,36956	,26753	,43896
	p	,74917	,60292	,87469	,62191	,73818	,65373	,80327	,67857	,69962	,65586	,76496	,67980	,78861	,66991	,65857	,66785	,59552	,64970	,58293	,67209
	r	,91348	,91348	,73185	,91532	,75910	,88664	,20082	,89549	,76033	,87809	,73093	,87711	,41720	,88602	,87156	,85779	,96376	,80193	,94026	,85714
	f1	,82320	,72640	,79692	,74062	,74850	,75257	,32131	,77208	,72871	,75088	,74756	,76595	,54571	,76296	,75024	,75100	,73616	,71783	,71968	,75342
<i>SVM</i>	ac	,75855	,66499	,30342	,67338	,68623	,69635	,67725	,72233	,75206	,69731	,74046	,71822	,61828	,71612	,70527	,67316	,62560	,70652	,60389	,69610
	auc	,76372	,68689	,26725	,70032	,77348	,71495	,76177	,77864	,79954	,77769	,81148	,78546	,66422	,79212	,79311	,73881	,68971	,77550	,64014	,76549
	k	,51710	,32998	-,3931	,34677	,37247	,39271	,35450	,44467	,50413	,39462	,48093	,43644	,23655	,43225	,41055	,34633	,25120	,41304	,20779	,39220
	p	,74759	,61452	,32368	,62610	,76285	,65594	,72584	,66514	,72592	,64148	,78163	,68007	,59581	,66028	,74655	,66098	,57761	,68627	,58130	,66666
	r	,78068	,88531	,36088	,86088	,54048	,82591	,56967	,89549	,80991	,89462	,66737	,82415	,73548	,89032	,62156	,71100	,93478	,76087	,74285	,78441
	f1	,76378	,72547	,34127	,72495	,63270	,73118	,63834	,76331	,76562	,74719	,72000	,74521	,65832	,75824	,67834	,68508	,71402	,72164	,65222	,72076
<i>NB</i>	ac	,74245	,64688	,62701	,66129	,69736	,65688	,66393	,61782	,69421	,67148	,69491	,57309	,70000	,66451	,68692	,66743	,70410	,59299	,62987	,55324
	auc	,85555	,71636	,82331	,66631	,79020	,68180	,71119	,72967	,75279	,73739	,73872	,68869	,75143	,73479	,76935	,73665	,75545	,68970	,68180	,75161
	k	,48490	,29376	,25403	,32258	,39473	,31376	,32786	,23565	,38843	,34297	,38983	,14618	,40000	,32903	,37385	,33486	,40821	,18599	,25974	,10649
	p	,86186	,59012	,87058	,60666	,82178	,60430	,71164	,57457	,74102	,62848	,68548	,54417	,78882	,62085	,66464	,64370	,72295	,55987	,66891	,52907
	r	,57746	,96177	,29838	,91733	,50404	,90890	,55123	,90778	,59710	,83884	,72033	,90042	,54623	,84516	,75458	,75000	,66183	,86956	,51428	,96883
	f1	,69156	,73144	,44444	,73033	,62484	,72595	,62124	,70373	,66132	,71858	,70247	,67837	,64548	,71584	,70676	,69279	,69104	,68117	,58149	,68440
<i>k-NN</i>	ac	,83400	,65593	,81552	,66532	,73785	,65587	,66291	,73668	,67045	,68698	,70868	,70127	,62688	,69139	,71215	,68463	,68236	,67270	,56753	,66233
	auc	,87728	,72100	,86430	,69783	,80080	,70423	,71505	,77133	,72576	,73274	,77319	,73396	,66056	,74641	,76920	,73952	,75074	,73810	,60690	,72485
	k	,66800	,31187	,63104	,33064	,47570	,31174	,32582	,47336	,34090	,37396	,41737	,40254	,25376	,38279	,42431	,36926	,36473	,34541	,13506	,32467
	p	,80181	,60601	,77896	,62424	,76404	,62500	,64749	,69411	,65249	,65365	,75718	,67992	,61895	,68237	,71362	,68008	,65314	,64444	,57142	,64637
	r	,88732	,89134	,88104	,83064	,68825	,77935	,71516	,84631	,72933	,79545	,61440	,76059	,66021	,71612	,70871	,69724	,77777	,77053	,54026	,71688
	f1	,84240	,72149	,82686	,71280	,72417	,69369	,67964	,76269	,68878	,71761	,67836	,71800	,63891	,69884	,71116	,68856	,71003	,70187	,55540	,67980
<i>SGD</i>	ac	,84104	,65996	,83467	,67943	,76923	,70242	,75102	,72745	,62913	,70247	,73940	,73093	,61182	,71612	,72477	,69380	,60748	,68599	,67272	,72077
	auc	,89747	,71487	,89627	,66846	,84177	,77071	,81533	,78266	,77245	,77349	,81070	,78524	,78201	,78834	,81082	,74986	,79224	,75810	,73587	,78600

	k	.68209	.31992	.66935	.35887	.53846	.40485	.50204	.45491	.25826	.40495	.47881	.46186	.22365	.43225	.44954	.38761	.21497	.37198	.34545	.44155
	p	.80762	.60585	.82806	.62292	.74538	.64925	.79661	.66818	.72727	.64848	.76279	.67868	.79213	.66080	.67689	.67280	.56137	.65217	.62246	.67346
	r	.89537	.91549	.84475	.90927	.81781	.88056	.67418	.90368	.41322	.88429	.69491	.87711	.30322	.88817	.86009	.75458	.98309	.79710	.87792	.85714
	fl	.84923	.72916	.83632	.73934	.77992	.74742	.73030	.76829	.52700	.74825	.72727	.76525	.43856	.75779	.75757	.71135	.71466	.71739	.72844	.75428
<i>DT</i>	ac	.62374	.67303	.70362	.62802	.74696	.65890	.71516	.71004	.65909	.70557	.61440	.61440	.61612	.70107	.53211	.63761	.66908	.66183	.56233	.67013
	auc	.79857	.70289	.78685	.67710	.77309	.67710	.76572	.72908	.73348	.73709	.66819	.69825	.69698	.74845	.62376	.71904	.68862	.74188	.56410	.67671
	k	.24748	.34607	.40725	.25604	.49392	.31781	.43032	.42008	.31818	.41115	.22881	.22881	.23225	.40215	.06422	.27522	.33816	.32367	.12467	.34026
	p	.57658	.62146	.66031	.58956	.71554	.62284	.69962	.66040	.62581	.66500	.59677	.58766	.64210	.67608	.55785	.64925	.64227	.63453	.60084	.70926
	r	.93159	.88531	.83871	.84274	.81983	.80566	.75409	.86475	.79132	.82851	.70550	.76694	.52473	.77204	.30963	.59862	.76328	.76328	.37142	.57662
	fl	.71230	.73029	.73889	.69377	.76415	.70256	.72583	.74889	.69890	.73781	.64660	.66544	.57751	.72088	.39823	.62291	.69757	.69298	.45906	.63610
<i>AB</i>	ac	.78370	.67505	.76915	.68649	.71659	.69939	.65676	.70184	.71797	.69214	.65678	.67266	.68494	.71075	.70527	.67201	.67512	.71135	.59480	.70000
	auc	.83860	.70465	.87748	.72807	.78308	.72483	.73920	.76016	.78870	.73448	.72959	.73639	.74098	.78021	.76185	.76063	.76162	.79048	.71194	.75491
	k	.56740	.35010	.53830	.37298	.43319	.39878	.31352	.40368	.43595	.38429	.31355	.34533	.36989	.42150	.41055	.34403	.35024	.42270	.18961	.40000
	p	.80387	.61950	.83797	.63120	.77157	.64506	.67915	.64946	.70485	.63676	.64015	.62329	.80069	.66118	.74655	.63297	.63551	.66540	.71098	.67111
	r	.75050	.90744	.66733	.89717	.61538	.88664	.59426	.87704	.75000	.89462	.71610	.87288	.49247	.86451	.62156	.81880	.82125	.85024	.31948	.78441
	fl	.77627	.73632	.74298	.74104	.68468	.74680	.63388	.74629	.72672	.74398	.67600	.72727	.60985	.74930	.67834	.71400	.71654	.74655	.44086	.72335
<i>RF</i>	ac	.81086	.67102	.83064	.66028	.74595	.69736	.74795	.70184	.72624	.73243	.70444	.71186	.70752	.72795	.64908	.72591	.65700	.71256	.65584	.68051
	auc	.87321	.72310	.88858	.71055	.81678	.73515	.79177	.76757	.77184	.77442	.76706	.76070	.77638	.79633	.73158	.78869	.74473	.76993	.70050	.74573
	k	.62173	.34205	.66129	.32056	.49190	.39473	.49590	.40368	.45247	.46487	.40889	.42372	.41505	.45591	.29816	.45183	.31401	.42512	.31168	.36103
	p	.80961	.62073	.82411	.62519	.74645	.65451	.75965	.65463	.73548	.67550	.75328	.66393	.72705	.68531	.75000	.69505	.63157	.68565	.69867	.65410
	r	.81287	.87927	.84072	.80040	.74493	.83603	.72541	.85450	.70661	.89462	.60805	.85805	.66451	.84301	.44724	.80504	.75362	.78502	.54805	.76623
	fl	.81124	.72772	.83233	.70203	.74569	.73422	.74213	.74133	.72075	.76977	.67291	.74861	.69438	.75602	.56034	.74601	.68722	.73198	.61426	.70574
<i>XGB</i>	ac	.78772	.65191	.81149	.67943	.74797	.69433	.72028	.71209	.72210	.70971	.71928	.69491	.64838	.71612	.62156	.70183	.60869	.70893	.63636	.67792
	auc	.87900	.71512	.87663	.72149	.80034	.73445	.76903	.76568	.78181	.76211	.75441	.74769	.72818	.78684	.68216	.74772	.71145	.78319	.70797	.75689
	k	.57545	.30382	.62298	.35887	.49595	.38866	.44057	.42418	.44421	.41942	.43855	.38983	.29677	.43225	.24311	.40367	.21739	.41787	.27272	.35584
	p	.80819	.59973	.81212	.62973	.76923	.64860	.72164	.66046	.70873	.65736	.70495	.64330	.69491	.66947	.61087	.66666	.57475	.66290	.65043	.64989
	r	.75452	.91348	.81048	.87096	.70850	.84817	.71721	.87295	.75413	.87603	.75423	.87500	.52903	.85376	.66972	.80733	.83574	.85024	.58961	.77142
	fl	.78043	.72408	.81130	.73096	.73761	.73508	.71942	.75198	.73073	.75110	.72876	.74147	.60073	.75047	.63895	.73029	.68110	.74497	.61852	.70546
<i>MLP</i>	ac	.67002	.65895	.82358	.68246	.77530	.68724	.65573	.70389	.73243	.70351	.72245	.73622	.69569	.69677	.59633	.71215	.70169	.72705	.66623	.70909
	auc	.84826	.71620	.88454	.70215	.84718	.74251	.75619	.74345	.80213	.75873	.77580	.78397	.77006	.75085	.64870	.78352	.80480	.77611	.77491	.78614
	k	.34004	.31790	.64717	.36491	.55060	.37449	.31147	.40778	.46487	.40702	.44491	.47245	.39139	.39354	.19266	.42431	.40338	.45410	.33246	.41818
	p	.82625	.60941	.81044	.63135	.73287	.64521	.75850	.65331	.77915	.66610	.69662	.68614	.65853	.66727	.62209	.66606	.64884	.70434	.60884	.68171
	r	.43058	.88531	.84475	.87701	.86639	.83198	.45696	.86885	.64876	.81611	.78813	.87076	.81290	.78494	.49082	.85091	.87922	.78260	.92987	.78441
	fl	.56613	.72190	.82724	.73417	.79406	.72679	.57033	.74582	.70800	.73351	.73956	.76750	.72762	.72134	.54871	.74723	.74666	.74141	.73586	.72946

**APÊNDICE J – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS - TREINAMENTO COM AMOSTRAS  
DO ENSINO SUPERIOR E TESTES COM AMOSTRAS DO TÉCNICO**

Resultados de acurácia (ac), AUC ROC (auc), kappa (k), precisão (p), recall (r) e F1-score (f1) dos modelos da Tese (T) e baseline (B) nas amostras de janelas fixas (5s, 10s, 20s, 40s, 60s, 90s, 120s, 180s, 240s e 360s)																					
Modelo		T 5	B 5	T 10	B 10	T 20	B 20	T 40	B 40	T 60	B 60	T 90	B 90	T 120	B 120	T 180	B 180	T 240	B 240	T 360	B 360
<i>LR</i>	ac	,86695	,81759	,85099	,78918	,83444	,80555	,78314	,78202	,78295	,77500	,72674	,73604	,62918	,71890	,70535	,68622	,70448	,66946	,46875	,63815
	auc	,93131	,87224	,92662	,86757	,90306	,87184	,85506	,83711	,85714	,85443	,82199	,81895	,71963	,80353	,77866	,77632	,77972	,75703	,52198	,74665
	k	,73390	,63519	,70198	,57836	,66888	,61111	,56629	,56404	,56590	,55000	,45348	,47209	,25837	,43779	,41071	,37244	,40896	,33893	-,0625	,27631
	p	,80866	,78793	,78291	,79241	,79452	,82051	,74513	,77342	,73988	,79803	,69383	,75438	,60931	,74663	,68337	,72956	,69945	,72161	,47197	,70792
	r	,96137	,86909	,97130	,78366	,90222	,78222	,86067	,79775	,87272	,73636	,81162	,70000	,72009	,66267	,76530	,59183	,71708	,55182	,52631	,47039
	f1	,87843	,82653	,86699	,78801	,84495	,80091	,79874	,78539	,80083	,76595	,74812	,72617	,66008	,70215	,72202	,65352	,70816	,62539	,49766	,56521
<i>SVM</i>	ac	,71137	,81759	,78366	,76379	,80777	,78888	,77752	,74719	,57840	,73977	,71627	,71860	,62201	,72846	,71301	,65561	,49579	,62885	,53289	,65131
	auc	,90621	,86050	,88046	,83195	,91165	,85521	,85857	,80036	,60784	,81741	,83035	,79318	,68853	,79502	,74859	,72491	,49811	,67518	,54133	,74255
	k	,42274	,63519	,56732	,52759	,61555	,57777	,55505	,49438	,15681	,47954	,43255	,43720	,24401	,45693	,42602	,31122	-,0084	,25770	,06578	,30263
	p	,63548	,76148	,71169	,72936	,88365	,80232	,81108	,73206	,60883	,77402	,80392	,73618	,58095	,74055	,74925	,67836	,49350	,65862	,54672	,71495
	r	,99141	,92489	,95364	,83885	,70888	,76666	,72359	,77977	,43863	,67727	,57209	,68139	,87559	,70334	,64030	,59183	,31932	,53501	,38486	,50328
	f1	,77451	,83527	,81509	,78028	,78668	,78409	,76484	,75516	,50990	,72242	,66847	,70772	,69847	,72147	,69050	,63215	,38775	,59041	,45173	,59073
<i>NB</i>	ac	,68025	,81759	,54966	,79801	,58111	,78555	,56629	,74831	,55227	,75795	,55000	,73604	,52870	,70813	,51913	,61224	,52381	,63865	,52467	,67105
	auc	,76102	,86160	,59510	,84946	,62218	,82354	,68550	,78338	,58876	,79957	,57379	,77918	,53452	,76185	,51986	,68283	,53033	,70136	,54053	,71194
	k	,36051	,63519	,09933	,59602	,16222	,57111	,13258	,49662	,10454	,51590	,10000	,47209	,05741	,41626	,03826	,22449	,04761	,27731	,04934	,34210
	p	,61111	,76909	,52619	,74280	,54478	,77634	,53673	,71209	,52853	,73210	,52677	,71641	,51485	,70518	,50985	,64569	,51251	,69411	,51299	,68840
	r	,99141	,90772	,99779	,91170	,98666	,80222	,96853	,83370	,96818	,81363	,98372	,78139	,99521	,71531	,98979	,49744	,97479	,49579	,97368	,62500
	f1	,75613	,83267	,68902	,81863	,70197	,78907	,69070	,76811	,68378	,77072	,68613	,74749	,67863	,71021	,67302	,56196	,67181	,57843	,67196	,65517
<i>k-NN</i>	ac	,84227	,76502	,78697	,71192	,71333	,71777	,74269	,72022	,71818	,73750	,64651	,69767	,65789	,68779	,64795	,64795	,64986	,60924	,57730	,60855
	auc	,88604	,84948	,85303	,81430	,76666	,80220	,81061	,78138	,79181	,80262	,69369	,73948	,71951	,75187	,67239	,68630	,68126	,65605	,60433	,65882
	k	,68454	,53004	,57395	,42384	,42666	,43555	,48539	,44044	,43636	,47500	,29302	,39534	,31578	,37559	,29591	,29591	,29972	,21848	,15460	,21710
	p	,82092	,78004	,79816	,73762	,73414	,77071	,75116	,71491	,76519	,74940	,65517	,69767	,68857	,70712	,67682	,66111	,67893	,62662	,59362	,62500
	r	,87553	,73819	,76821	,65783	,66888	,62000	,72584	,73258	,62954	,71363	,61860	,69767	,57655	,64114	,56632	,60714	,56862	,54061	,49013	,54276
	f1	,84735	,75854	,78290	,69544	,70000	,68719	,73828	,72364	,69077	,73108	,63636	,69767	,62760	,67252	,61666	,63297	,61890	,58045	,53693	,58098
<i>SGD</i>	ac	,86373	,81652	,87638	,78697	,84000	,78000	,75955	,78427	,68068	,77613	,64767	,73720	,72966	,72009	,67984	,69132	,71008	,68627	,71710	,64473
	auc	,92289	,87197	,94007	,86483	,91446	,85170	,79943	,83513	,79581	,85385	,69660	,82113	,80923	,80439	,75939	,77762	,78355	,76021	,77573	,74992
	k	,72746	,63304	,75275	,57395	,68000	,56000	,51910	,56853	,36136	,55227	,29534	,47441	,45933	,44019	,35969	,38265	,42016	,37254	,43421	,28947

	p	,80651	,78865	,81056	,79017	,80357	,79166	,69609	,77440	,62926	,80000	,61278	,75373	,68532	,74731	,64071	,73734	,67361	,74007	,71019	,71153
	r	,95708	,86480	,98234	,78145	,90000	,76000	,92134	,80224	,87954	,73636	,80232	,70465	,84928	,66507	,81887	,59438	,81512	,57423	,73355	,48684
	f1	,87536	,82497	,88822	,78579	,84905	,77551	,79303	,78807	,73364	,76686	,69486	,72836	,75854	,70379	,71892	,65819	,73764	,64668	,72168	,57812
<i>DT</i>	ac	,80686	,76824	,78145	,75607	,75555	,72333	,72921	,75618	,71136	,70113	,63488	,63720	,57894	,68899	,64030	,62500	,55042	,59523	,56743	,63486
	auc	,91594	,78860	,77416	,83583	,74915	,79942	,72594	,76307	,70619	,76802	,63488	,68823	,57674	,71253	,60626	,69382	,57088	,59523	,56743	,66095
	k	,61373	,53648	,56291	,51214	,51111	,44666	,45842	,51236	,42272	,40227	,26976	,27441	,15789	,37799	,28061	,25000	,10084	,19047	,13486	,26973
	p	,88648	,76595	,89968	,76728	,78607	,74216	,77419	,72093	,72463	,70344	,71969	,68209	,60000	,75159	,69503	,62250	,56122	,59770	,59903	,64539
	r	,70386	,77253	,63355	,73509	,70222	,68444	,64719	,83595	,68181	,69545	,44186	,51395	,47368	,56459	,50000	,63520	,46218	,58263	,40789	,59868
	f1	,78468	,76923	,74352	,75084	,74178	,71213	,70501	,77419	,70257	,69942	,54755	,58620	,52941	,64480	,58160	,62878	,50691	,59007	,48532	,62116
<i>AB</i>	ac	,79291	,79721	,87196	,78256	,82555	,78666	,78089	,76404	,75000	,77613	,72209	,71162	,73564	,71650	,67729	,66071	,64705	,64845	,65296	,64967
	auc	,92720	,84583	,92238	,84244	,90856	,83539	,86237	,81677	,83658	,83568	,81698	,77800	,79793	,76565	,77046	,72025	,70428	,71203	,72193	,70697
	k	,58583	,59442	,74392	,56512	,65111	,57333	,56179	,52809	,50000	,55227	,44418	,42325	,47129	,43301	,35459	,32142	,29411	,29691	,30592	,29934
	p	,88450	,75887	,82466	,79906	,85819	,79452	,80487	,77262	,80219	,79562	,78171	,71770	,77591	,71912	,71924	,68103	,68041	,66459	,73366	,70222
	r	,67382	,87124	,94481	,75496	,78000	,77333	,74157	,74831	,66363	,74318	,61627	,69767	,66267	,71052	,58163	,60459	,55462	,59944	,48026	,51973
	f1	,76492	,81118	,88065	,77639	,81722	,78378	,77193	,76027	,72636	,76850	,68920	,70754	,71483	,71480	,64315	,64054	,61111	,63033	,58051	,59735
<i>RF</i>	ac	,74785	,81008	,74392	,78918	,72222	,73000	,72584	,75730	,70227	,72386	,69651	,71976	,65191	,70215	,66581	,65943	,65266	,67507	,62006	,63157
	auc	,91406	,87673	,91834	,86139	,88216	,82682	,86530	,81753	,85242	,81425	,82586	,78021	,78542	,78019	,74830	,72516	,73551	,72367	,70806	,69773
	k	,49570	,62017	,48785	,57836	,44444	,46000	,45168	,51460	,40454	,44772	,39302	,43953	,30382	,40430	,33163	,31887	,30532	,35014	,24013	,26315
	p	,89419	,81075	,89045	,80045	,90322	,77023	,83612	,75963	,84496	,74686	,79858	,74934	,77253	,71177	,75390	,67705	,74008	,69716	,71345	,66000
	r	,56223	,80901	,55629	,77041	,49777	,65555	,56179	,75280	,49545	,67727	,52558	,66046	,43062	,67942	,49234	,60969	,47058	,61904	,40131	,54276
	f1	,69038	,80988	,68478	,78515	,64183	,70828	,67204	,75620	,62464	,71036	,63394	,70210	,55299	,69522	,59567	,64161	,57534	,65578	,51368	,59566
<i>XGB</i>	ac	,82081	,78648	,81015	,78145	,78333	,70555	,79101	,72584	,71022	,76818	,70000	,71860	,69258	,72368	,68494	,66581	,66106	,64705	,65131	,60855
	auc	,91703	,86871	,89523	,85673	,87972	,80208	,87326	,78948	,84118	,83315	,80125	,77745	,79246	,80299	,77700	,71875	,71115	,70572	,71403	,69501
	k	,64163	,57296	,62030	,56291	,56666	,41111	,58202	,45168	,42045	,53636	,40000	,43720	,38516	,44736	,36989	,33163	,32212	,29411	,30263	,21710
	p	,88040	,78958	,83532	,79043	,82776	,72952	,81662	,72483	,80528	,77441	,78666	,75133	,77104	,75202	,73311	,68361	,70034	,67101	,70720	,65420
	r	,74248	,78111	,77262	,76600	,71555	,65333	,75056	,72809	,55454	,75681	,54883	,65348	,54784	,66746	,58163	,61734	,56302	,57703	,51644	,46052
	f1	,80558	,78532	,80275	,77802	,76758	,68933	,78220	,72645	,65679	,76551	,64657	,69900	,64055	,70722	,64864	,64879	,62422	,62048	,59695	,54054
<i>MLP</i>	ac	,83905	,79613	,76379	,77262	,83444	,71333	,79662	,78427	,70227	,78181	,75232	,72325	,74760	,69378	,72193	,67474	,59383	,67366	,70559	,64802
	auc	,93536	,85733	,93114	,84447	,90909	,79071	,86079	,85357	,86693	,86191	,84171	,81091	,82356	,77225	,80301	,72876	,66273	,75697	,77750	,73634
	k	,67811	,59227	,52759	,54525	,66888	,42666	,59325	,56853	,40454	,56363	,50465	,44651	,49521	,38756	,44387	,34949	,18767	,34733	,41118	,29605
	p	,77335	,78163	,94095	,77024	,81953	,74489	,73826	,78048	,85317	,80845	,78628	,75668	,74238	,72752	,69506	,68767	,64888	,72963	,71929	,68750
	r	,95922	,82188	,56291	,77704	,85777	,64888	,91910	,79101	,48863	,73863	,69302	,65814	,75837	,61961	,79081	,64030	,40896	,55182	,67434	,54276
	f1	,85632	,80125	,70442	,77362	,83821	,69358	,81881	,78571	,62138	,77197	,73671	,70398	,75029	,66925	,73985	,66314	,50171	,62838	,69609	,60661

**APÊNDICE K – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS - TREINAMENTO COM AMOSTRAS  
DO ENSINO SUPERIOR E TESTES COM AMOSTRAS DO TÉCNICO – COM ATRIBUTOS DE DADOS SELECIONADOS**

Resultados de acurácia (ac), AUC ROC (auc), kappa (k), precisão (p), recall (r) e F1-score (f1) dos modelos da Tese (T) e baseline (B) nas amostras de janelas fixas (5s, 10s, 20s, 40s, 60s, 90s, 120s, 180s, 240s e 360s)																					
Modelo		T 5	B 5	T 10	B 10	T 20	B 20	T 40	B 40	T 60	B 60	T 90	B 90	T 120	B 120	T 180	B 180	T 240	B 240	T 360	B 360
<i>LR</i>	ac	,85836	,82510	,86423	,78697	,85444	,81111	,81685	,78202	,77954	,77500	,74069	,74418	,44378	,71531	,70025	,68750	,47479	,68907	,46875	,65953
	auc	,92204	,87039	,94280	,87072	,92537	,87609	,88758	,84364	,86289	,85443	,83464	,81934	,41670	,80389	,76738	,77709	,40819	,76157	,52198	,75457
	k	,71673	,65021	,72847	,57395	,70888	,62222	,63370	,56404	,55909	,55000	,48139	,48837	-,1124	,43062	,40051	,37500	-,0504	,37815	-,0625	,31907
	p	,79715	,77900	,79676	,79148	,81709	,82558	,76603	,76989	,73837	,79803	,77748	,75990	,46141	,74064	,69287	,73333	,48640	,74193	,47197	,72350
	r	,96137	,90772	,97792	,77924	,91333	,78888	,91236	,80449	,86590	,73636	,67441	,71395	,67224	,66267	,71938	,58928	,90196	,57983	,52631	,51644
	f1	,87159	,83845	,87809	,78531	,86253	,80681	,83282	,78681	,79707	,76595	,72229	,73621	,54722	,69949	,70588	,65346	,63199	,65094	,49766	,60268
<i>SVM</i>	ac	,62875	,81545	,78035	,78035	,68444	,79000	,76629	,74719	,74090	,73977	,48488	,69418	,75239	,71052	,31760	,66581	,70588	,62465	,53289	,64967
	auc	,64702	,84277	,91522	,84357	,87257	,84802	,86203	,78445	,86949	,81741	,51349	,76506	,82424	,75484	,23801	,73875	,76418	,66956	,54133	,71780
	k	,25751	,63090	,56070	,56070	,36888	,58000	,53258	,49438	,48181	,47954	-,0302	,38837	,50478	,42105	-,3647	,33163	,41176	,24930	,06578	,29934
	p	,63100	,76534	,91503	,78222	,86725	,79862	,81769	,72357	,85099	,77402	,47186	,73130	,75179	,72916	,31989	,68465	,73333	,65397	,54672	,68571
	r	,62017	,90987	,61810	,77704	,43555	,77555	,68539	,80000	,58409	,67727	,25348	,61395	,75358	,66985	,32398	,61479	,64705	,52941	,38486	,55263
	f1	,62554	,83137	,73781	,77962	,57988	,78692	,74572	,75987	,69272	,72242	,32980	,66750	,75268	,69825	,32192	,64784	,68750	,58513	,45173	,61202
<i>NB</i>	ac	,80150	,82725	,58388	,80574	,66666	,80111	,58988	,76629	,54318	,75795	,54767	,75116	,53708	,70095	,56505	,69642	,53081	,66666	,52467	,67269
	auc	,86967	,86754	,68342	,87355	,75526	,84008	,78289	,80352	,57211	,79957	,57104	,80878	,59508	,76698	,68833	,74381	,54364	,70571	,54053	,73116
	k	,60300	,65450	,16777	,61147	,33333	,60222	,17977	,53258	,08636	,51590	,09534	,50232	,07416	,40191	,13010	,39285	,06162	,33333	,04934	,34539
	p	,72408	,76521	,54611	,74688	,60135	,77263	,54962	,73652	,52334	,73210	,52546	,71951	,51954	,71649	,53669	,67579	,51622	,66124	,51299	,69662
	r	,97424	,94420	,99337	,92494	,98888	,85333	,99550	,82921	,96818	,81363	,98372	,82325	,98564	,66507	,95153	,75510	,98039	,68347	,97368	,61184
	f1	,83074	,84534	,70477	,82643	,74789	,81098	,70823	,78012	,67942	,77072	,68502	,76789	,68042	,68982	,68629	,71325	,67632	,67217	,67196	,65148
<i>k-NN</i>	ac	,81008	,81330	,76490	,79249	,72333	,72888	,72809	,75168	,70909	,73750	,66162	,70116	,65789	,68660	,63010	,64668	,61204	,63865	,57730	,64473
	auc	,88004	,84825	,85085	,84388	,80276	,80788	,75891	,80271	,79214	,80262	,71279	,74755	,73194	,75294	,67853	,69940	,65661	,68385	,60433	,70663
	k	,62017	,62660	,52980	,58498	,44666	,45777	,45618	,50337	,41818	,47500	,32325	,40232	,31578	,37320	,26020	,29336	,22409	,27731	,15460	,28947
	p	,82039	,76259	,76200	,74952	,77686	,76822	,75061	,73931	,75000	,74940	,67331	,71149	,67187	,69797	,65361	,66381	,61428	,64102	,59362	,66176
	r	,79399	,90987	,77041	,87858	,62666	,65555	,68314	,77752	,62727	,71363	,62790	,67674	,61722	,65789	,55357	,59438	,60224	,63025	,49013	,59210
	f1	,80697	,82974	,76619	,80894	,69372	,70743	,71529	,75794	,68316	,73108	,64981	,69368	,64339	,67734	,59944	,62718	,60820	,63559	,53693	,62500
<i>SGD</i>	ac	,87446	,82832	,88300	,78366	,86777	,80333	,73932	,78539	,72159	,77613	,75930	,74069	,75598	,71650	,68494	,68239	,71288	,69187	,71710	,67763
	auc	,92376	,87170	,94401	,86986	,93118	,86815	,88059	,84525	,78636	,85385	,83876	,82101	,84219	,80425	,76157	,77445	,80454	,76281	,77573	,75162
	k	,74892	,65665	,76600	,56732	,73555	,60666	,47865	,57078	,44318	,55227	,51860	,48139	,51196	,43301	,36989	,36479	,42577	,38375	,43421	,35526

	p	,82495	,78125	,82309	,79138	,85591	,81670	,65919	,76906	,66386	,80000	,69734	,75682	,71230	,74525	,65072	,72990	,66170	,74552	,71019	,73478
	r	,95064	,91201	,97571	,77041	,88444	,78222	,99101	,81573	,89772	,73636	,91627	,70930	,85885	,65789	,79846	,57908	,87114	,58263	,73355	,55592
	fl	,88335	,84158	,89292	,78076	,86994	,79909	,79174	,79171	,76328	,76686	,79196	,73229	,77874	,69885	,71706	,64580	,75211	,65408	,72168	,63295
<i>DT</i>	ac	,78004	,75000	,66666	,73509	,54222	,72888	,72247	,76516	,70568	,70113	,59651	,65232	,64712	,69617	,69642	,64413	,60084	,53781	,56743	,59703
	auc	,74700	,79531	,71082	,81232	,54029	,80212	,71182	,78920	,72386	,76802	,59651	,70326	,66667	,72901	,71385	,69034	,62913	,54282	,56743	,62119
	k	,56008	,50000	,33333	,47019	,08444	,45777	,44494	,53033	,41136	,40227	,19302	,30465	,29425	,39234	,39285	,28826	,20168	,07563	,13486	,19407
	p	,85175	,75717	,70572	,75912	,55937	,75245	,74146	,73228	,75931	,70344	,60000	,67466	,67226	,70098	,73475	,65229	,60465	,54205	,59903	,60000
	r	,67811	,73605	,57174	,68874	,39777	,68222	,68314	,83595	,60227	,69545	,57907	,58837	,57416	,68421	,61479	,61734	,58263	,48739	,40789	,58223
	fl	,75507	,74646	,63170	,72222	,46493	,71561	,71111	,78069	,67173	,69942	,58934	,62857	,61935	,69249	,66944	,63433	,59343	,51327	,48532	,59098
<i>AB</i>	ac	,87446	,80901	,87969	,78476	,81666	,79000	,80337	,75730	,78068	,77613	,73372	,71279	,68779	,71650	,69515	,68622	,68347	,64845	,65296	,64802
	auc	,92973	,84741	,91828	,85074	,89106	,83190	,88871	,81688	,83947	,83568	,81289	,78110	,80841	,76565	,75820	,75341	,75105	,71203	,72193	,70538
	k	,74892	,61802	,75938	,56953	,63333	,58000	,60674	,51460	,56136	,55227	,46744	,42558	,37559	,43301	,39030	,37244	,36694	,29691	,30592	,29605
	p	,87366	,75441	,83858	,79723	,85185	,79194	,82451	,76690	,80493	,79562	,76240	,72932	,77543	,71912	,70619	,70505	,73818	,66459	,73366	,70833
	r	,87553	,91630	,94039	,76379	,76666	,78666	,77078	,73932	,74090	,74318	,67907	,67674	,52870	,71052	,66836	,64030	,56862	,59944	,48026	,50328
	fl	,87459	,82751	,88657	,78015	,80701	,78929	,79674	,75286	,77159	,76850	,71832	,70205	,62873	,71480	,68676	,67112	,64240	,63033	,58051	,58846
<i>RF</i>	ac	,73927	,78111	,74172	,76710	,71666	,73666	,72134	,74494	,70568	,72386	,69883	,71860	,68062	,70813	,65943	,65306	,67787	,67226	,62006	,64638
	auc	,89375	,86602	,89249	,86085	,86529	,82719	,86166	,81151	,85796	,81425	,82151	,77688	,79681	,78352	,75830	,71331	,75482	,72035	,70806	,70202
	k	,47854	,56223	,48344	,53421	,43333	,47333	,44269	,48988	,41136	,44772	,39767	,43720	,36124	,41626	,31887	,30612	,35574	,34453	,24013	,29276
	p	,88850	,79772	,89247	,78673	,92951	,77099	,84561	,74885	,84150	,74686	,81318	,75683	,83259	,71534	,75100	,67751	,74902	,69040	,71345	,68312
	r	,54721	,75321	,54966	,73289	,46888	,67333	,54157	,73707	,50681	,67727	,51627	,64418	,45215	,69138	,47704	,58418	,53501	,62465	,40131	,54605
	fl	,67729	,77483	,68032	,75885	,62333	,71886	,66027	,74292	,63262	,71036	,63157	,69598	,58604	,70316	,58346	,62739	,62418	,65588	,51368	,60694
<i>XGB</i>	ac	,82725	,77897	,81457	,76931	,76444	,71777	,80000	,72809	,71136	,76818	,70814	,71627	,67344	,71650	,67984	,66326	,66386	,65406	,65131	,63815
	auc	,91073	,86688	,90244	,85360	,84978	,81305	,87316	,79393	,84967	,83315	,80934	,78147	,77313	,80311	,76769	,71311	,73966	,69936	,71403	,70098
	k	,65450	,55794	,62913	,53863	,52888	,43555	,60000	,45618	,42272	,53636	,41627	,43255	,34689	,43301	,35969	,32653	,32773	,30812	,30263	,27631
	p	,86052	,78260	,84337	,78372	,81648	,74019	,83291	,72605	,80592	,77441	,76557	,74603	,78431	,74393	,71428	,68713	,73877	,67973	,70720	,67213
	r	,78111	,77253	,77262	,74392	,68222	,67111	,75056	,73258	,55681	,75681	,60000	,65581	,47846	,66028	,59949	,59949	,50700	,58263	,51644	,53947
	fl	,81889	,77753	,80645	,76330	,74334	,70396	,78959	,72930	,65860	,76551	,67275	,69802	,59435	,69962	,65187	,64032	,60132	,62745	,59695	,59854
<i>MLP</i>	ac	,76716	,72210	,82560	,76379	,80444	,74000	,77078	,78651	,75909	,78181	,73255	,69302	,72368	,73205	,74107	,68239	,60644	,64425	,70559	,67269
	auc	,88507	,84781	,89628	,83946	,87153	,83548	,88194	,84277	,84650	,86191	,80246	,78752	,80413	,80202	,79686	,76679	,65889	,71567	,77750	,74169
	k	,53433	,44420	,65121	,52759	,60888	,48000	,54157	,57303	,51818	,56363	,46511	,38604	,44736	,46411	,48214	,36479	,21288	,28851	,41118	,34539
	p	,85470	,77308	,82997	,77985	,75464	,78723	,68798	,76955	,71111	,80845	,73148	,71842	,70459	,73774	,71925	,70370	,61377	,67944	,71929	,71255
	r	,64377	,62875	,81898	,73509	,90222	,65777	,99101	,81797	,87272	,73863	,73488	,63488	,77033	,72009	,79081	,63010	,57423	,54621	,67434	,57894
	fl	,73439	,69349	,82444	,75681	,82186	,71670	,81215	,79302	,78367	,77197	,73317	,67407	,73600	,72881	,75334	,66487	,59334	,60559	,69609	,63883

**APÊNDICE L – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS – MÉDIA DOS RESULTADOS**

Resultados de acurácia (ac), AUC ROC (auc), kappa (k), precisão (p), recall (r) e F1-score (f1) dos modelos da Tese (T) e baseline (B) nas amostras de janelas fixas (5s, 10s, 20s, 40s, 60s, 90s, 120s, 180s, 240s e 360s)																					
Modelo		T	B	T	B	T	B	T	B	T	B	T	B	T	B	T	B	T	B	T	B
		5	5	10	10	20	20	40	40	60	60	90	90	120	120	180	180	240	240	360	360
<i>LR</i>	ac	,80772	,74582	,83225	,73229	,79475	,75854	,70868	,75832	,74581	,73357	,62714	,73401	,63502	,71697	,66689	,70148	,63122	,67712	,50905	,67622
	auc	,87885	,78993	,90760	,79394	,86201	,81602	,83111	,80918	,81145	,81465	,75745	,80057	,73004	,79136	,74093	,78732	,71967	,76522	,60675	,76792
	k	,61544	,49164	,66450	,46458	,58950	,51709	,41736	,51665	,49163	,46714	,25428	,46803	,27004	,43395	,33379	,40296	,26245	,35425	,01810	,35244
	p	,75188	,70265	,78570	,70817	,76222	,73906	,78550	,72492	,73317	,71800	,60399	,71798	,67038	,70474	,68958	,69960	,72217	,68150	,49921	,69007
	r	,92334	,89229	,91407	,83336	,85596	,83240	,60041	,84867	,77107	,81239	,90581	,78538	,58262	,76897	,61086	,72367	,44670	,69499	,75406	,65337
	f1	,82859	,78198	,84412	,75961	,80635	,77729	,64030	,77880	,74880	,75430	,71363	,74541	,60678	,72827	,63663	,70239	,49665	,67604	,59153	,65529
<i>SVM</i>	ac	,71786	,73827	,73709	,71506	,78040	,74161	,73610	,73322	,63218	,71854	,66958	,71205	,61906	,72176	,55604	,66840	,55707	,65742	,45930	,66591
	auc	,81361	,79022	,78883	,77075	,86221	,79458	,81753	,78079	,68476	,79792	,78397	,78450	,68889	,78745	,55776	,73115	,59502	,70627	,44183	,73533
	k	,43571	,47655	,47418	,43012	,56081	,48322	,47220	,46645	,26436	,43708	,33916	,42411	,23813	,44352	,11209	,33680	,11415	,31484	-,0813	,33183
	p	,67684	,68397	,69098	,67386	,82594	,72933	,77253	,70260	,63683	,70860	,73984	,70757	,60184	,70040	,58403	,67695	,56055	,65480	,47615	,69244
	r	,86492	,91516	,85182	,85390	,71679	,79122	,66712	,81611	,59431	,78181	,52227	,73052	,73242	,79468	,58047	,64454	,45072	,66847	,43788	,60748
	f1	,75134	,78083	,76149	,75141	,76625	,75576	,71575	,75371	,60738	,73393	,61229	,71678	,65199	,73906	,57736	,65919	,49588	,65451	,44796	,64045
<i>NB</i>	ac	,70934	,72922	,64681	,72209	,64126	,71919	,60128	,66309	,62427	,70335	,61027	,66251	,61704	,67234	,60532	,61116	,60127	,60133	,58766	,62578
	auc	,78691	,78332	,70572	,74852	,69227	,75253	,68814	,74345	,66624	,76653	,64348	,73606	,63885	,75165	,62796	,71117	,62717	,70310	,61379	,70640
	k	,41868	,45844	,29362	,44418	,28252	,43839	,20256	,32618	,24855	,40671	,22055	,32503	,23408	,34469	,21064	,22233	,20255	,20266	,17532	,25157
	p	,71271	,68000	,64073	,67253	,64112	,68870	,61530	,62961	,63301	,66868	,61477	,63643	,61977	,64737	,60090	,60703	,64856	,61668	,60004	,61888
	r	,80456	,91160	,85978	,88629	,80608	,85556	,72812	,86869	,78884	,83450	,78741	,82925	,82879	,80711	,84008	,70629	,73377	,68751	,76346	,76055
	f1	,72932	,77545	,71358	,76367	,68943	,75634	,63175	,72486	,67560	,73981	,66412	,71420	,68538	,71115	,68209	,63159	,63857	,62346	,64245	,66815
<i>k-NN</i>	ac	,82707	,71450	,79116	,68912	,72660	,69643	,71868	,70591	,69690	,71327	,66223	,69364	,65367	,69336	,63762	,65482	,65282	,63312	,57631	,61661
	auc	,87933	,77633	,85514	,75293	,78024	,76552	,77950	,75931	,75418	,76294	,69199	,74023	,69854	,74836	,67991	,70836	,68630	,68117	,60534	,66219
	k	,65414	,42900	,58233	,37825	,45321	,39287	,43736	,41182	,39380	,42655	,32447	,38729	,30735	,38672	,27525	,30965	,30565	,26624	,15262	,23322
	p	,78962	,69685	,78202	,67850	,72539	,70159	,72096	,68904	,71593	,70133	,66365	,68279	,67344	,69514	,65831	,66085	,68240	,63427	,58847	,62599
	r	,89551	,81175	,80950	,76037	,73120	,72801	,71538	,75563	,66601	,75971	,65676	,72595	,59903	,69154	,57444	,63614	,57175	,62538	,50870	,57917
	f1	,83844	,74171	,79450	,70829	,72656	,70366	,71802	,71997	,68743	,72631	,65984	,70306	,63349	,69193	,61325	,64791	,62219	,62738	,54541	,60110
<i>SGD</i>	ac	,85239	,74327	,85401	,72816	,81574	,74526	,74094	,75637	,71017	,73414	,68347	,73142	,65838	,71649	,69772	,69715	,70347	,68432	,69491	,67756
	auc	,90843	,78886	,91771	,78931	,88382	,80889	,80761	,80259	,79841	,81437	,73005	,80148	,75187	,79199	,77277	,78210	,78452	,76584	,75393	,76724
	k	,70478	,48654	,70803	,45633	,63149	,49052	,48188	,51275	,42035	,46828	,36695	,46284	,31676	,43299	,39544	,39430	,40694	,36864	,38983	,35512
	p	,80762	,70212	,81624	,70543	,78661	,72298	,75260	,72583	,67406	,71899	,66010	,71546	,69144	,70375	,66627	,69811	,67597	,68904	,66678	,69267

	r	,92522	,88512	,91455	,82621	,86599	,82635	,75165	,83759	,83336	,81239	,77510	,78347	,57840	,77124	,79820	,71806	,78195	,70861	,80314	,65121
	fl	,86215	,77893	,86119	,75531	,82433	,76532	,73501	,77541	,74256	,75476	,71097	,74349	,59270	,72863	,72555	,69868	,72473	,68650	,72447	,65804
<i>DT</i>	ac	,74146	,70856	,76774	,70868	,72646	,69061	,70118	,70544	,68522	,70232	,56426	,62686	,60130	,69503	,63093	,63130	,53728	,63699	,53761	,64405
	auc	,84998	,72306	,77141	,76873	,75840	,73892	,67847	,73475	,72190	,74546	,57282	,68752	,59724	,73049	,61391	,70659	,54330	,65539	,53848	,67959
	k	,48292	,41713	,53548	,41736	,45292	,38122	,40236	,41089	,37045	,40464	,12852	,25373	,20260	,39007	,26186	,26261	,07457	,27398	,07522	,28811
	p	,75212	,68103	,80772	,69352	,74653	,68439	,71267	,67897	,67522	,68077	,60433	,63583	,59705	,71383	,68137	,63587	,54330	,61736	,55403	,64607
	r	,81370	,84803	,73814	,78488	,68815	,73291	,69654	,77760	,73657	,77128	,36182	,64045	,62716	,66831	,49197	,61691	,48109	,70677	,42342	,63700
	fl	,76250	,74688	,75879	,73109	,71597	,70381	,70016	,72492	,70074	,72008	,45253	,62643	,60206	,68284	,57138	,62585	,50964	,65562	,47836	,64094
<i>AB</i>	ac	,77523	,74116	,84475	,73301	,79283	,74303	,71934	,73396	,73915	,73259	,69844	,69903	,72319	,71470	,67867	,67152	,64720	,67809	,63687	,66704
	auc	,88817	,77608	,90102	,78158	,85188	,78011	,80834	,79418	,81206	,78553	,77757	,75993	,79170	,77194	,76112	,74919	,72883	,74724	,70942	,72971
	k	,55046	,48232	,68950	,46602	,58567	,48605	,43868	,46793	,47830	,46518	,39688	,39806	,44639	,42941	,35734	,34305	,29440	,35618	,27374	,33408
	p	,84391	,69478	,82403	,71237	,78817	,71979	,73993	,71045	,75796	,71499	,71018	,67803	,79704	,68992	,74638	,66112	,64863	,66343	,71748	,67795
	r	,67795	,88230	,87664	,83413	,81813	,82998	,67509	,81985	,71301	,81787	,71068	,77892	,60230	,79289	,54540	,71628	,66620	,72363	,45052	,65207
	fl	,75134	,77528	,84826	,75950	,79919	,76529	,70600	,75546	,73181	,75506	,70074	,72020	,68342	,73390	,62863	,68163	,64957	,68699	,55324	,65522
<i>RF</i>	ac	,78740	,74508	,78375	,73229	,73762	,71064	,73074	,74545	,71012	,72401	,70683	,71634	,65929	,71613	,67350	,69153	,66811	,68717	,62172	,66124
	auc	,90568	,80367	,90302	,78623	,85321	,78315	,82603	,80053	,80967	,79088	,79045	,77186	,76388	,78808	,73763	,75737	,73045	,74597	,69847	,72633
	k	,57481	,49016	,56751	,46458	,47525	,42129	,46149	,49091	,42024	,44803	,41367	,43269	,31858	,43226	,34700	,38306	,33623	,37434	,24344	,32248
	p	,86124	,71778	,85735	,71694	,82368	,71023	,80054	,72265	,78035	,70833	,77102	,70828	,76248	,69942	,75437	,68506	,74422	,68058	,68723	,66487
	r	,69359	,85521	,68943	,80154	,63451	,74174	,62106	,80775	,60929	,78078	,59435	,75502	,46369	,76229	,51452	,70622	,51186	,71290	,45390	,64800
	fl	,75849	,77396	,75409	,75233	,69964	,71834	,69610	,76013	,67203	,73630	,66747	,72476	,57572	,72659	,61151	,69275	,60572	,69212	,54360	,65235
<i>XGB</i>	ac	,80427	,73026	,79469	,73095	,77071	,70146	,75257	,72050	,71720	,73636	,70275	,70676	,68070	,72205	,68536	,69128	,65963	,67195	,65163	,64648
	auc	,90086	,78807	,87143	,78864	,84258	,76626	,81921	,77248	,81078	,79682	,78128	,76228	,76159	,79806	,75949	,75817	,72071	,72640	,70701	,72489
	k	,60854	,46052	,58938	,46190	,54143	,40292	,50515	,44100	,43440	,47272	,40550	,41352	,36140	,44411	,37072	,38255	,31927	,34391	,30326	,29296
	p	,84980	,70347	,79022	,71035	,78000	,68896	,76302	,69655	,75716	,71479	,73806	,69732	,71481	,71100	,70164	,67663	,66826	,66494	,67408	,65461
	r	,74046	,84830	,81070	,81949	,76466	,75784	,73798	,79129	,65743	,81022	,64835	,76424	,62446	,76814	,65664	,73642	,65107	,69672	,60367	,61987
	fl	,79115	,76138	,79817	,75503	,76921	,71477	,74976	,73822	,69529	,75530	,68202	,72023	,65986	,73188	,67410	,70002	,65401	,67485	,63097	,62614
<i>MLP</i>	ac	,75805	,73106	,78915	,72552	,80588	,70585	,75179	,74561	,70960	,74421	,71143	,72020	,67918	,70818	,69927	,69230	,56986	,68164	,67032	,68375
	auc	,89426	,78899	,90663	,78633	,87917	,76357	,81834	,79954	,83334	,82144	,77769	,77941	,77575	,78052	,78874	,74839	,67847	,74885	,73113	,75817
	k	,51611	,46213	,57831	,45105	,61177	,41171	,50359	,49123	,41921	,48843	,42287	,44041	,35836	,41636	,39854	,38460	,13973	,36328	,34065	,36750
	p	,83487	,69618	,88931	,69860	,79676	,69877	,73085	,72040	,82323	,73068	,72787	,71305	,74804	,69736	,74792	,68064	,69029	,70107	,68479	,68826
	r	,67981	,86869	,67157	,82904	,82058	,74853	,79459	,82173	,53770	,81043	,68337	,75809	,54370	,75389	,63050	,72726	,27694	,64547	,62937	,67138
	fl	,70493	,76697	,75617	,75311	,80844	,71563	,75728	,76495	,64798	,76120	,70415	,72803	,60419	,71562	,66617	,70021	,37182	,66632	,65584	,67350

**APÊNDICE M – TODOS OS RESULTADOS SOBRE A GENERALIZAÇÃO DOS MODELOS – MÉDIA DOS RESULTADOS – COM  
ATRIBUTOS SELECIONADOS**

Resultados de acurácia (ac), AUC ROC (auc), kappa (k), precisão (p), recall (r) e F1-score (f1) dos modelos da Tese (T) e baseline (B) nas amostras de janelas fixas (5s, 10s, 20s, 40s, 60s, 90s, 120s, 180s, 240s e 360s)																					
Modelo		T 5	B 5	T 10	B 10	T 20	B 20	T 40	B 40	T 60	B 60	T 90	B 90	T 120	B 120	T 180	B 180	T 240	B 240	T 360	B 360
<i>LR</i>	ac	,83109	,74052	,83887	,73320	,79969	,75980	,69633	,75883	,74824	,74183	,74693	,73808	,54823	,72002	,70505	,70154	,56469	,68692	,55125	,68951
	auc	,90426	,79264	,92155	,77961	,86883	,81819	,85106	,81426	,81100	,81686	,82370	,80158	,60011	,79536	,78785	,78820	,60528	,76109	,55125	,68951
	k	,66219	,48104	,67774	,46641	,59938	,51961	,39267	,51767	,49648	,48367	,49387	,47617	,09646	,44004	,41011	,40309	,12937	,37385	,10251	,37902
	p	,77316	,69096	,83573	,70669	,77764	,73965	,78465	,72423	,71899	,72694	,77122	,71985	,62501	,70528	,67572	,70059	,54096	,69582	,52745	,69780
	r	,93742	,91060	,85489	,84728	,83622	,83776	,55659	,84999	,81312	,80723	,70267	,79553	,54472	,77435	,79547	,72354	,93286	,69088	,73328	,68679
	f1	,84740	,78242	,83751	,76296	,80552	,77969	,57706	,77944	,76289	,75842	,73492	,75108	,54646	,73122	,72806	,70223	,68407	,68439	,60867	,67805
<i>SVM</i>	ac	,69365	,74022	,54189	,72687	,68534	,74317	,72177	,73476	,74648	,71854	,61267	,70620	,68533	,71332	,51143	,66949	,66574	,66558	,56839	,67288
	auc	,70537	,76483	,59124	,77194	,82303	,78149	,81190	,78155	,83452	,79755	,66249	,77526	,74423	,77348	,51556	,73878	,72694	,72253	,56839	,67288
	k	,38730	,48044	,08378	,45374	,37067	,48635	,44354	,46952	,49297	,43708	,22535	,41240	,37067	,42665	,02287	,33898	,33148	,33117	,13679	,34577
	p	,68929	,68993	,61936	,70416	,81505	,72728	,77177	,69436	,78846	,70775	,62675	,70568	,67380	,69472	,53322	,67282	,65547	,67012	,56401	,67619
	r	,70042	,89759	,48949	,81896	,48802	,80073	,62753	,84774	,69700	,78595	,46043	,71905	,74453	,78009	,47277	,66290	,79092	,64514	,56386	,66852
	f1	,69466	,77842	,53954	,75229	,60629	,75905	,69203	,76159	,72917	,73481	,52490	,70636	,70550	,72824	,50013	,66646	,70076	,65339	,55198	,66639
<i>NB</i>	ac	,77197	,73706	,60545	,73351	,68201	,72899	,62691	,69206	,61869	,71472	,62129	,66212	,61854	,68273	,62598	,68193	,61745	,62983	,57727	,61297
	auc	,86261	,79195	,75336	,76993	,77273	,76094	,74704	,76659	,66245	,76848	,65488	,74874	,67325	,75088	,72884	,74023	,64955	,69771	,57727	,61297
	k	,54395	,47413	,21090	,46703	,36403	,45799	,25382	,38412	,23739	,42944	,24259	,32425	,23708	,36547	,25197	,36386	,23491	,25966	,15454	,22594
	p	,79297	,67767	,70835	,67677	,71156	,68847	,63063	,65555	,63218	,68029	,60547	,63184	,65418	,66867	,60066	,65975	,61959	,61056	,59095	,61285
	r	,77585	,95298	,64588	,92114	,74646	,88112	,77336	,86850	,78264	,82624	,85203	,86184	,76594	,75511	,85305	,75255	,82111	,77651	,74398	,79033
	f1	,76115	,78839	,57461	,77838	,68637	,76846	,66474	,74193	,67037	,74465	,69375	,72313	,66295	,70283	,69653	,70302	,68368	,67667	,62673	,66794
<i>k-NN</i>	ac	,82204	,73462	,79021	,72890	,73059	,69238	,69550	,74418	,68977	,71224	,68515	,70121	,64238	,68900	,67112	,66565	,64720	,65568	,57241	,65353
	auc	,87866	,78463	,85757	,77085	,80178	,75605	,73698	,78702	,75895	,76768	,74299	,74075	,69625	,74967	,72387	,71946	,70368	,71097	,57241	,65353
	k	,64409	,46924	,58042	,45781	,46118	,38475	,39100	,48836	,37954	,42448	,37031	,40243	,28477	,37800	,34225	,33131	,29441	,31136	,14483	,30707
	p	,81110	,68430	,77048	,68688	,77045	,69661	,69905	,71671	,70124	,70152	,71524	,69570	,64541	,69017	,68362	,67195	,63371	,64273	,58252	,65406
	r	,84065	,90061	,82573	,85461	,65746	,71745	,69915	,81192	,67830	,75454	,62115	,71866	,63872	,68701	,63114	,64581	,69000	,70039	,51519	,65449
	f1	,82469	,77562	,79653	,76087	,70895	,70056	,69747	,76031	,68597	,72434	,66409	,70584	,64115	,68809	,65530	,65787	,65911	,66873	,54617	,65240
<i>SGD</i>	ac	,85775	,74414	,85884	,73155	,81850	,75288	,74517	,75642	,67536	,73930	,74935	,73581	,68390	,71631	,70486	,68810	,66018	,68893	,69491	,69920
	auc	,91061	,79328	,92014	,76916	,88647	,81943	,84796	,81395	,77941	,81367	,82473	,80313	,81210	,79629	,78620	,76216	,79839	,76046	,69491	,69920
	k	,71551	,48828	,71768	,46310	,63700	,50576	,49035	,51285	,35072	,47861	,49870	,47163	,36780	,43263	,40972	,37620	,32037	,37786	,38983	,39841

	p	,81628	,69355	,82557	,70715	,80065	,73298	,72790	,71862	,69556	,72424	,73006	,71775	,75221	,70302	,66381	,70135	,61154	,69884	,66632	,70412
	r	,92300	,91375	,91023	,83984	,85112	,83139	,83259	,85970	,65547	,81033	,80559	,79321	,58103	,77303	,82928	,66683	,92712	,68986	,80573	,70653
	fl	,86629	,78537	,86462	,76005	,82493	,77325	,76102	,78000	,64514	,75755	,75961	,74877	,60865	,72832	,73732	,67857	,73338	,68574	,72506	,69362
<i>DT</i>	ac	,70189	,71151	,68514	,68156	,64459	,69389	,71881	,73760	,68238	,70335	,60545	,63336	,63162	,69862	,61426	,64087	,63496	,59982	,56488	,63358
	auc	,77278	,74910	,74884	,74471	,65669	,73961	,73877	,75914	,72867	,75256	,63235	,70075	,68182	,73873	,66881	,70469	,65887	,64235	,56488	,63358
	k	,40378	,42303	,37029	,36312	,28918	,38779	,43763	,47521	,36477	,40671	,21091	,26673	,26325	,39724	,22853	,28174	,26992	,19965	,12977	,26716
	p	,71417	,68932	,68302	,67434	,63746	,68765	,72054	,69634	,69256	,68422	,59838	,63116	,65718	,68853	,64630	,65077	,62346	,58829	,59993	,65463
	r	,80485	,81068	,70522	,76574	,60880	,74394	,71862	,85035	,69679	,76198	,64228	,67766	,54944	,72812	,46221	,60798	,67295	,62534	,38966	,57943
	fl	,73369	,73837	,68530	,70799	,61454	,70908	,71847	,76479	,68532	,71862	,61797	,64700	,59843	,70668	,53383	,62862	,64550	,60312	,47219	,61354
<i>AB</i>	ac	,82908	,74203	,82442	,73563	,76663	,74469	,73006	,72957	,74932	,73414	,69525	,69273	,68637	,71363	,70021	,67912	,67929	,67990	,62388	,67401
	auc	,88416	,77603	,89788	,78940	,83707	,77836	,81396	,78852	,81408	,78508	,77124	,75875	,77469	,77293	,76003	,75702	,75633	,75125	,62388	,67401
	k	,65816	,48406	,64884	,47126	,53326	,48939	,46013	,45914	,49865	,46828	,39050	,38546	,37274	,42726	,40042	,35824	,35859	,35981	,24776	,34802
	p	,83877	,68696	,83827	,71422	,81171	,71850	,75183	,70818	,75489	,71619	,70127	,67631	,78806	,69015	,72637	,66901	,68684	,66500	,72232	,68972
	r	,81302	,91187	,80386	,83048	,69102	,83665	,68252	,80818	,74545	,81890	,69758	,77481	,51059	,78752	,64496	,72955	,69494	,72484	,39987	,64385
	fl	,82543	,78192	,81478	,76060	,74585	,76805	,71531	,74957	,74916	,75624	,69716	,71466	,61929	,73205	,68255	,69256	,67947	,68844	,51068	,65590
<i>RF</i>	ac	,77506	,72607	,78618	,71369	,73130	,71701	,73465	,72339	,71596	,72815	,70164	,71523	,69407	,71804	,65426	,68948	,66743	,69241	,63795	,66345
	auc	,88348	,79456	,89053	,78570	,84104	,78117	,82672	,78954	,81490	,79434	,79428	,76879	,78659	,78993	,74494	,75100	,74977	,74514	,63795	,66345
	k	,55013	,45214	,57236	,42739	,46261	,43403	,46929	,44678	,43192	,45630	,40328	,43046	,38814	,43609	,30852	,37897	,33487	,38482	,27591	,32690
	p	,84906	,70923	,85829	,70596	,83798	,71275	,80263	,70174	,78849	,71118	,78323	,71038	,77982	,70033	,75050	,68628	,69029	,68802	,70606	,66861
	r	,68004	,81624	,69519	,76664	,60691	,75468	,63349	,79579	,60671	,78595	,56216	,75111	,55833	,76719	,46214	,69461	,64431	,70483	,47468	,65614
	fl	,74426	,75128	,75633	,73044	,68451	,72654	,70120	,74212	,67669	,74007	,65224	,72229	,64021	,72959	,57190	,68670	,65570	,69393	,56397	,65634
<i>XGB</i>	ac	,80749	,71544	,81303	,72437	,75621	,70605	,76014	,72009	,71673	,73894	,71371	,70559	,66091	,71631	,65070	,68255	,63628	,68149	,64384	,65804
	auc	,89486	,79100	,88954	,78754	,82506	,77375	,82109	,77981	,81574	,79763	,78188	,76458	,75066	,79497	,72493	,73042	,72555	,74128	,64384	,65804
	k	,61498	,43088	,62606	,44875	,51242	,41211	,52028	,44018	,43347	,47789	,42741	,41119	,32183	,43263	,30140	,36510	,27256	,36299	,28767	,31608
	p	,83435	,69117	,82774	,70672	,79286	,69440	,77728	,69326	,75732	,71589	,73526	,69466	,73961	,70670	,66258	,67690	,65676	,67131	,67881	,66101
	r	,76782	,84300	,79155	,80744	,69536	,75964	,73388	,80276	,65547	,81642	,67711	,76540	,50375	,75702	,63460	,70341	,67137	,71643	,55302	,65545
	fl	,79966	,75081	,80887	,74713	,74048	,71952	,75451	,74064	,69466	,75831	,70075	,71974	,59754	,72504	,64541	,68530	,64121	,68621	,60774	,65200
<i>MLP</i>	ac	,71859	,69052	,82459	,72312	,78987	,71362	,71326	,74520	,74576	,74266	,72750	,71462	,70969	,71441	,66870	,69727	,65406	,68565	,68591	,69089
	auc	,86666	,78201	,89041	,77080	,85935	,78899	,81907	,79311	,82432	,81032	,78913	,78575	,78710	,77644	,72278	,77515	,73185	,74589	,68591	,69089
	k	,43718	,38105	,64919	,44625	,57974	,42724	,42652	,49041	,49152	,48533	,45501	,42925	,41938	,42883	,33740	,39455	,30813	,37131	,37182	,38178
	p	,84047	,69125	,82021	,70560	,74376	,71622	,72324	,71143	,74513	,73728	,71405	,70228	,68156	,70251	,67067	,68488	,63130	,69189	,66407	,69713
	r	,53718	,75703	,83187	,80605	,88430	,74488	,72398	,84341	,76074	,77737	,76151	,75282	,79161	,75252	,64082	,74051	,72672	,66441	,80210	,68168
	fl	,65026	,70769	,82584	,74549	,80796	,72174	,69124	,76942	,74583	,75274	,73637	,72079	,73181	,72507	,65103	,70605	,67000	,67350	,71598	,68415

**APÊNDICE N – TODOS OS RESULTADOS SOBRE GENERALIZAÇÃO NOS MODELOS DE JANELAS VARIÁVEIS**

Modelo		Cenário 1 (Treino com o ensino técnico e teste com o ensino superior)				Cenário 2 (Treino com o ensino superior e teste com o ensino técnico)				Média dos cenários			
		Sem Sel. Atr.		Com Sel. Atr.		Sem Sel. Atr.		Com Sel. Atr.		Sem Sel. Atr.		Com Sel. Atr.	
		Tese	Baseline	Tese	Baseline	Tese	Baseline	Tese	Baseline	Tese	Baseline	Tese	Baseline
<i>LR</i>	ac	,76660	,81187	,80281	,81690	,49570	,74785	,73068	,70279	,63115	,77986	,76675	,75984
	auc	,85905	,90253	,89165	,90591	,59867	,84198	,86344	,86760	,72886	,87225	,87755	,88675
	k	,53319	,62374	,60563	,63380	-,00858	,49570	,46137	,40557	,26230	,55972	,53350	,51969
	p	,72572	,77777	,80160	,78584	,49615	,88372	,66020	,88571	,61093	,83074	,73090	,83577
	r	,85714	,87323	,80482	,87122	,55364	,57081	,95064	,46566	,70539	,72202	,87773	,66844
<i>SVM</i>	f1	,78597	,82274	,80321	,82633	,52332	,69361	,77924	,61040	,65465	,75818	,79122	,71837
	ac	,71428	,81086	,73641	,83098	,76073	,72532	,66845	,73605	,73750	,76809	,70243	,78351
	auc	,81998	,90098	,84770	,90147	,82901	,87680	,82896	,89601	,82449	,88889	,83833	,89874
	k	,42857	,62173	,47283	,66197	,52145	,45064	,33691	,47210	,47501	,53618	,40487	,56703
	p	,77099	,77938	,77011	,79639	,77676	,85958	,83404	,87162	,77387	,81948	,80207	,83400
<i>NB</i>	r	,60965	,86720	,67404	,88933	,73176	,53862	,42060	,55364	,67070	,70291	,54732	,72149
	f1	,68089	,82095	,71888	,84030	,75359	,66226	,55920	,67716	,71724	,74161	,63904	,75873
	ac	,76156	,80985	,81589	,78873	,64163	,57510	,56223	,56008	,70160	,69248	,68906	,67440
	auc	,82301	,84532	,90476	,84639	,75059	,80699	,72223	,85302	,78680	,82615	,81350	,84970
	k	,52313	,61971	,63179	,57746	,28326	,15021	,12446	,12017	,40320	,38496	,37812	,34881
<i>k-NN</i>	p	,79545	,77797	,80426	,75488	,58571	,89772	,53364	,90000	,69058	,83785	,66895	,82744
	r	,70422	,86720	,83501	,85513	,96781	,16952	,98712	,13519	,83601	,51836	,91106	,49516
	f1	,74706	,82017	,81934	,80188	,72977	,28519	,69277	,23507	,73841	,55268	,75606	,51848
	ac	,70925	,80784	,78973	,81690	,65665	,72854	,64377	,77145	,68295	,76819	,71675	,79418
	auc	,75269	,86797	,85990	,87029	,73894	,83220	,78314	,87480	,74582	,85009	,82152	,87255
<i>SGD</i>	k	,41851	,61569	,57947	,63380	,31330	,45708	,28755	,54291	,36590	,53638	,43351	,58836
	p	,72707	,76842	,78235	,77015	,73548	,85618	,81308	,87315	,73127	,81230	,79771	,82165
	r	,67002	,88128	,80281	,90342	,48927	,54935	,37339	,63519	,57964	,71532	,58810	,76930
	f1	,69738	,82099	,79245	,83148	,58762	,66928	,51176	,73540	,64250	,74513	,65210	,78344
	ac	,76861	,81891	,81589	,81790	,58369	,73068	,63090	,76287	,67615	,77480	,72339	,79039
<i>SGD</i>	auc	,84293	,86972	,89947	,89369	,65025	,83051	,71646	,87626	,74659	,85012	,80797	,88498
	k	,53722	,63782	,63179	,63581	,16738	,46137	,26180	,52575	,35230	,54960	,44679	,58078

	p	,74054	,78253	,81910	,79477	,54653	,86941	,57644	,88401	,64354	,82597	,69777	,83939
	r	,82696	,88330	,81086	,85714	,98283	,54291	,98712	,60515	,90489	,71310	,89899	,73114
	fl	,78136	,82986	,81496	,82478	,70245	,66842	,72784	,71847	,74191	,74914	,77140	,77162
<i>DT</i>	ac	,78269	,80080	,78772	,78772	,60944	,72639	,62339	,81223	,69606	,76360	,70555	,79997
	auc	,83916	,86362	,84071	,86390	,68442	,82255	,67612	,88013	,76179	,84309	,75841	,87201
	k	,56539	,60161	,57545	,57545	,21888	,45279	,24678	,62446	,39213	,52720	,41111	,59995
	p	,73223	,73178	,74155	,72990	,68214	,81681	,68025	,89008	,70718	,77430	,71090	,80999
	r	,89134	,94969	,88330	,91348	,40987	,58369	,46566	,71244	,65061	,76669	,67448	,81296
	fl	,80399	,82662	,80624	,81143	,51206	,68085	,55286	,79141	,65802	,75373	,67955	,80142
<i>AB</i>	ac	,73440	,80382	,81488	,81086	,78970	,75321	,75321	,75751	,76205	,77852	,78405	,78418
	auc	,80666	,86353	,87794	,87429	,89078	,88417	,90231	,88279	,84872	,87385	,89013	,87854
	k	,46881	,60764	,62977	,62173	,57939	,50643	,50643	,51502	,52410	,55704	,56810	,56837
	p	,75381	,75593	,75867	,75966	,83750	,88562	,88562	,88216	,79565	,82077	,82214	,82091
	r	,69617	,89738	,92354	,90945	,71888	,58154	,58154	,59442	,70753	,73946	,75254	,75193
	fl	,72384	,82060	,83303	,82783	,77367	,70207	,70207	,71025	,74876	,76134	,76755	,76904
<i>RF</i>	ac	,70925	,75855	,82293	,74949	,75107	,77253	,74141	,76394	,73016	,76554	,78217	,75672
	auc	,80823	,84436	,86447	,85779	,90072	,89928	,90384	,89909	,85447	,87182	,88415	,87844
	k	,41851	,51710	,64587	,49899	,50214	,54506	,48283	,52789	,46032	,53108	,56435	,51344
	p	,74299	,73236	,76616	,71831	,89795	,89197	,88659	,88679	,82047	,81217	,82638	,80255
	r	,63983	,81488	,92957	,82092	,56652	,62017	,55364	,60515	,60318	,71753	,74161	,71303
	fl	,68756	,77142	,84000	,76619	,69473	,73164	,68163	,71938	,69115	,75153	,76081	,74279
<i>XGB</i>	ac	,76056	,76257	,80885	,76559	,74570	,76180	,69420	,77360	,75313	,76218	,75153	,76959
	auc	,82261	,83274	,84973	,85158	,88863	,87839	,86824	,88895	,85562	,85557	,85898	,87026
	k	,52112	,52515	,61770	,53118	,49141	,52360	,38841	,54721	,50627	,52437	,50305	,53919
	p	,75542	,72935	,74878	,72837	,88552	,86526	,87866	,86743	,82047	,79731	,81372	,79790
	r	,77062	,83501	,92957	,84708	,56437	,62017	,45064	,64592	,66750	,72759	,69011	,74650
	fl	,76294	,77861	,82944	,78325	,68938	,72250	,59574	,74046	,72616	,75055	,71259	,76186
<i>MLP</i>	ac	,73038	,81589	,80181	,83501	,60085	,71781	,60515	,76180	,66562	,76685	,70348	,79840
	auc	,88483	,89344	,90848	,89758	,74052	,78154	,65794	,89706	,81268	,83749	,78321	,89732
	k	,46076	,63179	,60362	,67002	,20171	,43562	,21030	,52360	,33124	,53370	,40696	,59681
	p	,96734	,78035	,73885	,78856	,81756	,88593	,58361	,86747	,89245	,83314	,66123	,82801
	r	,47686	,87927	,93360	,91549	,25965	,50000	,73390	,61802	,36825	,68963	,83375	,76675
	fl	,63881	,82686	,82488	,84730	,39413	,63923	,65019	,72180	,51647	,73305	,73754	,78455

## APÊNDICE O – HIPERPARÂMETROS CANDIDATOS

### Logistic Regression (LR)

```
[{'solver': ['liblinear'],
  'penalty': ['l1','l2'],
  'C': [1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5]},
 {'solver': ['lbfgs'],
  'penalty': ['l2'],
  'C': [1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5]},
 {'solver': ['lbfgs'],
  'penalty': ['none']}]
```

### Support Vector Machine (SVM)

```
[{'kernel': ['linear'],
  'C': [1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5]},
 {'kernel': ['rbf', 'sigmoid'],
  'C': [1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5],
  'coef0': [0.0, 0.001, 0.01, 0.1],
  'gamma': ['scale', 'auto'],
  'coef0': [0.0, 0.001, 0.01, 0.1]},
 {'kernel': ['poly'],
  'C': [1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5],
  'degree': [1, 2, 3, 4, 5],
  'gamma': ['scale', 'auto'],
  'coef0': [0.0, 0.001, 0.01, 0.1]}]
```

### Naive Bayes (NB)

```
[{'var_smoothing': np.logspace(0, -9, num=100)}]
```

### k-Nearest Neighbors (k-NN)

```
[{'n_neighbors': list(range(3, 10)),
  'weights': ['uniform', 'distance'],
  'algorithm': ['ball_tree', 'kd_tree', 'brute'],
  'leaf_size': [1, 5, 10, 20, 30],
  'p': [1, 2]}]
```

### Stochastic Gradient Descent (SGD)

```
[{'loss': ['hinge', 'log', 'modified_huber',
  'squared_hinge', 'perceptron'],
  'penalty': ['l1','l2'],
  'alpha': [0.00001, 0.0001, 0.001],
  'learning_rate': ['constant', 'optimal',
  'invscaling', 'adaptive'],
  'eta0': [0.001, 0.01, 0.05, 0.1, 0.15]}]
```

### Decision Tree (DT)

```
[{'criterion': ['gini', 'entropy'],
  'max_depth': [None, 2, 3, 5, 10, 15],
  'min_samples_leaf': [1, 3, 5, 10, 20, 30],
  'min_samples_split': [2, 3, 5, 10, 20, 30, 50],
  'max_features': [None, 'auto', 'sqrt', 'log2'],
  'min_weight_fraction_leaf': [0.0001, 0.001, 0.01],
  'max_leaf_nodes': [None, 2, 5]}]
```

### AdaBoost (AB)

```
[{'n_estimators': [10, 12, 15, 17, 20, 22,
  25, 27, 30, 32, 35],
  'learning_rate': [0.1, 0.2, 0.3, 0.4, 0.5,
  0.6, 0.7, 0.8, 0.9, 1.0],
  'algorithm': ['SAMME', 'SAMME.R']}]
```

### Random Forrest (RF)

```
[{'criterion': ['gini', 'entropy'],
  'n_estimators': [400, 500, 600, 700,
  800, 900, 1000],
  'min_samples_leaf': [1, 3, 5],
  'min_samples_split': [2, 5, 10, 15],
  'min_weight_fraction_leaf': [0.0001,
  0.001]}]
```

### XGBoost (XGB)

```
[{'booster': ['gblinear', 'dart']},
 {'booster': ['gbtree'],
  'eta': [0.1, 0.2, 0.3],
  'gamma': [0, 0.1, 0.3],
  'max_depth': [1, 2, 3, 4, 5],
  'lambda': [1, 2, 3, 4, 5],
  'alpha': [0, 1, 2]}]
```

### Multi-layer Perceptron (MLP)

```
[{'solver': ['sgd'],
  'hidden_layer_sizes': [(10),(20),(30),
  (5,10),(10,15)],
  'activation': ['logistic', 'tanh', 'relu'],
  'alpha': [0.0001, 0.01, 0.1],
  'learning_rate_init': [0.001, 0.025],
  'momentum': [0.6, 0.7, 0.8, 0.9],
  'max_iter': [100000],
  "batch_size": [50, 150, 250]},
 {'solver': ['adam'],
  'hidden_layer_sizes': [(10),(20),(30),
  (5,10),(10,15)],
  'activation': ['logistic', 'tanh', 'relu'],
  'alpha': [0.0001, 0.01, 0.1],
  'learning_rate_init': [0.001, 0.025],
  'max_iter': [100000],
  "batch_size": [50, 150, 250]}]
```

## APÊNDICE P – TERMO DE ASSENTIMENTO (TA)

### TERMO DE ASSENTIMENTO – TA

Você está sendo convidado(a) a participar, como voluntário(a), no projeto de pesquisa para o desenvolvimento de uma ferramenta computacional que detecta estados afetivos (como emoções, humor e atitudes) de estudantes enquanto estão realizando tarefas de programação de computadores em um ambiente computacional (software) de programação. O projeto é coordenado por Tiago Roberto Kautzmann, Mestre em Computação Aplicada, pesquisador em Inteligência Artificial e doutorando pelo Programa de Pós-Graduação em Computação Aplicada (PPGCA) da Universidade do Vale do Rio dos Sinos (UNISINOS).

A sua participação nesta pesquisa consistirá em ser um dos utilizadores de um ambiente computacional de programação de computadores, que coletará dados da sua interação com o ambiente enquanto realiza tarefas de programação. Para tanto, você irá usar o sistema e realizar as atividades determinadas no período especificado pelo(s) pesquisador(res) do projeto. Você também poderá ser convidado(a) a realizar testes para verificar seu conhecimento em programação, assim como responder a questionários. Além disso, você também poderá ser convidado(a) a participar de sessões individuais para narrar suas ações enquanto estiver usando o sistema.

Durante a sua participação no sistema, suas ações (teclado e mouse) poderão ser gravadas no banco de dados do sistema. Você também poderá ser gravado(a) com uma webcam ligada ao computador, para que possamos coletar dados sobre suas expressões faciais e corporais, enquanto usa o sistema. A tela do seu computador também poderá ser gravada para coletar dados da sua interação com o sistema. No caso de necessitarmos gravar a sua tela e/ou gravar você com a webcam, você será previamente convidado, podendo recusar a gravação. Esses dados coletados do sistema, webcam e tela do computador serão utilizados **apenas** para fins de pesquisa, de forma anônima. Apenas alguns poucos pesquisadores integrantes do projeto terão acesso às gravações. Esses dados serão guardados pelo pesquisador responsável em local seguro e por um período de 15 anos. Cabe aqui salientar que **a sua identidade será preservada, pois não serão divulgados nomes, informações ou imagens que possam identificá-lo(a).**

É importante observar que a sua participação neste projeto não terá nenhum impacto nas notas nas disciplinas da instituição de ensino. Por se tratar de utilização de um software educacional, acreditamos que os riscos a você sejam mínimos. Entre eles, podemos citar a possibilidade de constrangimento ao responder questionários e/ou testes de conhecimento; desconforto; estresse; cansaço ao responder às perguntas ou fazer exercícios; e quebra de anonimato. Os pesquisadores farão o possível para minimizar esses inconvenientes, mas você pode sempre optar, a qualquer momento, a

retirar seu consentimento se perceber qualquer um desses inconvenientes. O(s) pesquisador(res) tentará(ão) afetar minimamente o andamento normal das aulas. Como benefícios da participação desse experimento, você desenvolverá suas habilidades de programação de computadores. Outros benefícios potenciais são o desenvolvimento da habilidade de resolução de problemas, aumento da motivação para aprender programação e maior engajamento para resolver exercícios de programação.

Após ser esclarecido(a) sobre as informações acima, no caso de aceitar fazer parte desta pesquisa, por favor assinale o campo abaixo “Declaro que entendi os objetivos, riscos e benefícios da minha participação na pesquisa e concordo em participar”. Caso não deseje participar da pesquisa, assinale a opção “Não desejo participar dessa pesquisa”.

A qualquer momento, você pode retirar o consentimento. Sua recusa não trará nenhum prejuízo na sua relação com o pesquisador, com a UNISINOS e com a instituição de ensino.

Em caso de dúvida, você pode procurar o pesquisador Tiago Roberto Kautzmann no telefone (51) 99282-2114 ou pelo endereço de e-mail [tkautzmann@edu.unisinos.br](mailto:tkautzmann@edu.unisinos.br).

(  ) Declaro que entendi os objetivos, riscos e benefícios da minha participação na pesquisa e concordo em participar.

(  ) Não desejo participar dessa pesquisa.

\_\_\_\_ / \_\_\_\_ / \_\_\_\_\_  
Data

\_\_\_\_\_  
Me. Tiago Roberto Kautzmann – Coordenador da pesquisa

\_\_\_\_\_  
Nome completo do(a) aluno(a)

\_\_\_\_\_  
Assinatura do do(a) aluno(a)

## APÊNDICE Q – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE) PARA ALUNOS MENORES DE IDADE

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – TCLE

Senhores pais ou responsáveis, seu(ua) filho(a)/afilhado(a), aluno(a) de uma disciplina de programação de computadores na instituição de ensino \_\_\_\_\_, está sendo convidado(a) a participar, como voluntário(a), no projeto de pesquisa para o desenvolvimento de uma ferramenta computacional que detecta estados afetivos de estudantes enquanto estão realizando tarefas de programação de computadores em um ambiente (software) de programação. O projeto é coordenado por Tiago Roberto Kautzmann, Mestre em Computação Aplicada, pesquisador em Inteligência Artificial e doutorando pelo Programa de Pós-Graduação em Computação Aplicada (PPGCA) da Universidade do Vale do Rio dos Sinos (UNISINOS).

A participação do(a) aluno(a) nesta pesquisa consistirá em ser um dos utilizadores de um ambiente computacional de programação de computadores, que coletará dados da sua interação com o ambiente enquanto realiza tarefas de programação. Para tanto, o(a) aluno(a) deverá usar o sistema e realizar as atividades determinadas no período especificado pelo(os) pesquisador(res) do projeto. O(a) aluno(a) poderá ser convidado(a) a realizar testes para verificar seu conhecimento em programação, assim como responder a questionários. O(a) aluno(a) também poderá ser convidado(a) a participar de sessões individuais para narrar suas ações enquanto estiver usando o sistema.

Durante toda a participação do(a) aluno(a) no sistema, suas ações (teclado e mouse) poderão ser gravadas no banco de dados do sistema. O(a) estudante também poderá ser gravado(a) com uma webcam ligada ao computador, para que possamos coletar dados sobre suas expressões faciais e corporais, enquanto usa o sistema. A tela do computador também poderá ser gravada para coletar dados da interação do(a) aluno(a) com o sistema. No caso de necessitarmos gravar a tela e/ou gravar com a webcam, o(a) aluno(a) será previamente convidado, podendo recusar a gravação. Esses dados e gravações serão utilizados **apenas** para fins de pesquisa, de forma anônima. Apenas alguns poucos pesquisadores integrantes do projeto terão acesso às gravações. Esses dados serão guardados pelo pesquisador responsável em local seguro e por um período de 15 anos. Cabe aqui salientar que **a identidade do(a) aluno(a) será preservada, pois não serão divulgados nomes, informações ou imagens que possam identificá-lo(a).**

É importante observar que a participação do seu(ua) filho(a)/afilhado(a) neste projeto não terá nenhum impacto nas notas nas disciplinas da instituição de ensino. Por se tratar de utilização de um software educacional, acreditamos que os riscos ao estudante sejam mínimos. Entre eles, podemos citar a possibilidade de constrangimento ao responder questionários e/ou testes de conhecimento; desconforto; estresse; cansaço ao responder às perguntas ou fazer exercícios; e quebra de anonimato. Os pesquisadores farão o

possível para minimizar esses inconvenientes, mas você pode sempre optar, a qualquer momento, a retirar seu consentimento se perceber qualquer um desses inconvenientes. O(s) pesquisador(res) tentará(ão) afetar minimamente o andamento normal das aulas. Como benefícios da participação desse experimento, o(a) aluno(a) desenvolverá suas habilidades de programação de computadores. Outros benefícios potenciais são o desenvolvimento da habilidade de resolução de problemas, aumento da motivação para aprender programação e maior engajamento para resolver exercícios de programação.

Após ser esclarecido(a) sobre as informações acima, no caso de aceitar que seu(ua) filho(a) faça parte desta pesquisa, por favor assinale o campo abaixo “Declaro que entendi os objetivos, riscos e benefícios da participação do meu(inha) filho(a)/afilhado(a) na pesquisa e concordo com a participação dele(ela) nessa pesquisa”. Caso não deseje que seu(ua) filho(a)/afilhado(a) participe da pesquisa, assinale a opção “Não desejo que meu(inha) filho(a)/afilhado(a) participe dessa pesquisa”.

A qualquer momento, você pode retirar o consentimento. Sua recusa não trará nenhum prejuízo para seu(ua) filho(a)/afilhado(a) na relação com o pesquisador, com a UNISINOS e com a instituição de ensino. Em caso de dúvida, você pode procurar o pesquisador Tiago Roberto Kautzmann no telefone (51) 99282-2114 ou pelo endereço de e-mail [tkautzmann@edu.unisinos.br](mailto:tkautzmann@edu.unisinos.br).

(    ) Declaro que entendi os objetivos, riscos e benefícios da participação do meu(inha) filho(a)/afilhado(a) na pesquisa e concordo com a participação dele(ela) nessa pesquisa.

(    ) Não desejo que meu(inha) filho(a)/afilhado(a) participe dessa pesquisa.

\_\_\_\_ / \_\_\_\_ / \_\_\_\_\_  
Data

\_\_\_\_\_  
Me. Tiago Roberto Kautzmann – Coordenador da pesquisa

\_\_\_\_\_  
Nome completo do(a) aluno(a)

\_\_\_\_\_  
Nome completo do responsável pelo(a) aluno(a)

\_\_\_\_\_  
Assinatura do responsável pelo aluno(a)

## APÊNDICE R – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE) PARA ALUNOS MAIORES DE IDADE

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – TCLE

Você está sendo convidado(a) a participar, como voluntário(a), no projeto de pesquisa para o desenvolvimento de uma ferramenta computacional que detecta estados afetivos de estudantes enquanto estão realizando tarefas de programação de computadores em um ambiente (software) de programação. O projeto é coordenado por Tiago Roberto Kautzmann, Mestre em Computação Aplicada, pesquisador em Inteligência Artificial e doutorando pelo Programa de Pós-Graduação em Computação Aplicada (PPGCA) da Universidade do Vale do Rio dos Sinos (UNISINOS).

A sua participação nesta pesquisa consistirá em ser um dos utilizadores de um ambiente computacional de programação de computadores, que coletará dados da sua interação com o ambiente enquanto realiza tarefas de programação. Para tanto, você deverá usar o sistema e realizar as atividades determinadas no período especificado pelo(os) pesquisador(res) do projeto. Você poderá ser convidado(a) a realizar testes para verificar seu conhecimento em programação, assim como responder a questionários. Você também poderá ser convidado(a) a participar de sessões individuais para narrar suas ações enquanto estiver usando o sistema.

Durante toda a sua participação no sistema, suas ações (teclado e mouse) poderão ser gravadas no banco de dados do sistema. Você também poderá ser gravado(a) com uma webcam ligada ao computador, para que possamos coletar dados sobre suas expressões faciais e corporais, enquanto usa o sistema. A tela do computador também poderá ser gravada para coletar dados da sua interação com o sistema. No caso de necessitarmos gravar a tela e/ou gravar com a webcam, você será previamente convidado, podendo recusar a gravação. Esses dados e gravações serão utilizados **apenas** para fins de pesquisa, de forma anônima. Apenas alguns poucos pesquisadores integrantes do projeto terão acesso às gravações. Esses dados serão guardados pelo pesquisador responsável em local seguro e por um período de 15 anos. Cabe aqui salientar que **a sua identidade será preservada, pois não serão divulgados nomes, informações ou imagens que possam identificá-lo.**

É importante observar que a sua participação neste projeto não terá nenhum impacto nas notas nas disciplinas da instituição de ensino. Por se tratar de utilização de um software educacional, acreditamos que os riscos para você sejam mínimos. Entre eles, podemos citar a possibilidade de constrangimento ao responder questionários e/ou testes de conhecimento; desconforto; estresse; cansaço ao responder às perguntas ou fazer exercícios; e quebra de anonimato. Os pesquisadores farão o possível para minimizar esses inconvenientes, mas você pode sempre optar, a qualquer momento, a retirar seu consentimento se perceber qualquer um desses inconvenientes. O(s)

pesquisador(res) tentará(ão) afetar minimamente o andamento normal das aulas. Como benefícios da participação desse experimento, você desenvolverá suas habilidades de programação de computadores. Outros benefícios potenciais são o desenvolvimento da habilidade de resolução de problemas, aumento da motivação para aprender programação e maior engajamento para resolver exercícios de programação.

Após ser esclarecido(a) sobre as informações acima, no caso de você aceitar fazer parte desta pesquisa, por favor assinale o campo abaixo “Declaro que entendi os objetivos, riscos e benefícios da minha participação na pesquisa e concordo em participar”. Caso não deseje participar da pesquisa, assinale a opção “Não desejo participar dessa pesquisa”.

A qualquer momento, você pode retirar o consentimento. Sua recusa não trará nenhum prejuízo na sua relação com o pesquisador, com a UNISINOS e com a instituição de ensino.

Em caso de dúvida, você pode procurar o pesquisador Tiago Roberto Kautzmann no telefone (51) 99282-2114 ou pelo endereço de e-mail [tkautzmann@edu.unisinos.br](mailto:tkautzmann@edu.unisinos.br).

(  ) Declaro que entendi os objetivos, riscos e benefícios da minha participação na pesquisa e concordo em participar.

(  ) Não desejo participar dessa pesquisa.

\_\_\_\_ / \_\_\_\_ / \_\_\_\_\_

Data

\_\_\_\_\_  
Me. Tiago Roberto Kautzmann – Coordenador da pesquisa

\_\_\_\_\_  
Nome completo do aluno(a)

\_\_\_\_\_  
Assinatura do aluno(a)

## APÊNDICE S – PARECER DO CEP DA UNISINOS SOBRE O PROJETO DE PESQUISA RELACIONADO A ESTA TESE

UNIVERSIDADE DO VALE DO  
RIO DOS SINOS - UNISINOS



### PARECER CONSUBSTANCIADO DO CEP

#### DADOS DO PROJETO DE PESQUISA

**Título da Pesquisa:** Detecção de confusão de alunos em ambientes computacionais de aprendizagem de programação da computadores.

**Pesquisador:** TIAGO ROBERTO KAUTZMANN

**Área Temática:**

**Versão:** 1

**CAAE:** 24435519.2.0000.5344

**Instituição Proponente:** Universidade do Vale do Rio dos Sinos - UNISINOS

**Patrocinador Principal:** Financiamento Próprio

#### DADOS DO PARECER

**Número do Parecer:** 3.707.158

#### **Apresentação do Projeto:**

Os processos cognitivos desenvolvidos durante a aprendizagem, como a memorização e a tomada de decisão, podem ser prejudicados pelas emoções do estudante. As emoções negativas, como a frustração, a confusão e o tédio podem ser prejudiciais para a aprendizagem quando o aluno mantém essas emoções por um longo período de tempo. Elas são mais frequentes em tarefas complexas de aprendizagem, como é a programação de computadores. O presente estudo está interessado em uma das emoções não-básicas mais frequentes em ambientes computacionais de aprendizagem de programação: a confusão. Quando o aluno está confuso durante uma tarefa de aprendizagem, ele está num estado de desequilíbrio cognitivo resultante de contradições, conflitos e informações errôneas apresentadas a ele. A confusão é um estado afetivo que ocorre quando a nova informação que chega não corresponde ao conhecimento anterior do indivíduo ou quando ela não pode ser integrada ao modelo mental existente. Quando isto ocorre, além de entrar em confusão, o indivíduo poderá experimentar um conjunto de outras emoções negativas, tais como a frustração e o tédio. O objetivo desta pesquisa é implementar um algoritmo classificador de confusão de alunos enquanto eles estão realizando tarefas de programação de computadores em ambientes computacionais (software) de programação. Um detector de confusão poderia gerar alertas ao tutor humano e possibilitar uma adaptação automática do ambiente computacional de aprendizagem a partir da predição da confusão, como atribuir tarefas mais adequadas ao conhecimento do aluno. Para alcançar esse objetivo, a pesquisa realizará

**Endereço:** Av. Unisinos, 950

**Bairro:** Cristo Rei

**CEP:** 93.022-000

**UF:** RS

**Município:** SAO LEOPOLDO

**Telefone:** (51)3591-1198

**Fax:** (51)3590-8118

**E-mail:** cep@unisinos.br

Continuação do Parecer: 3.707.158

diversos estudos exploratórios e experimentais, com o objetivo de identificar dados dos estudantes que possuam capacidade preditiva de confusão em tarefas de programação. Serão consideradas as seguintes modalidades de dados sobre os alunos: interação com o sistema, conhecimento no domínio de programação e habilidades metacognitivas. Além de implementar um algoritmo de classificação de confusão em ambientes de programação, a pesquisa pretende encontrar correlações entre as variáveis de interesses, além de relações causais entre elas.

#### **Objetivo da Pesquisa:**

O objetivo primário do trabalho é treinar algoritmos computacionais para realizar a detecção da confusão de alunos durante tarefas de aprendizagem de programação, em ambientes computacionais de programação, identificando quando esta confusão está sendo benéfica para a aprendizagem do aluno. Este objetivo é desdobrado nos seguintes objetivos secundários: a) verificar correlações entre dados sobre o conhecimento dos estudantes e o estado afetivo de confusão; b) identificar correlações entre dados sobre habilidades metacognitivas dos estudantes e o estado afetivo de confusão; c) identificar correlações entre dados sobre as interações dos estudantes com o sistema e o estado afetivo de confusão; d) identificar a capacidade preditiva de diferentes conjuntos de dados envolvendo conhecimento do aluno, habilidades metacognitivas e interações com o sistema; e) identificar as configurações (parâmetros, estruturas e janelas de dados) dos algoritmos de classificação que possuam as melhores capacidades preditivas do estado afetivo de confusão; f) identificar quando a confusão está levando a uma melhor aprendizagem e quando está levando a uma pior aprendizagem.

#### **Avaliação dos Riscos e Benefícios:**

Os riscos foram descritos no projeto e nos termos de consentimento e assentimento anexados ao projeto, destacando-se entre eles os seguintes: possibilidade de constrangimento ao responder questionários e/ou testes de conhecimento, desconforto, estresse, cansaço ao responder às perguntas ou fazer exercícios e quebra de anonimato. Os pesquisadores comprometeram-se a fazer o possível para minimizar esses inconvenientes, deixando claro aos participantes que eles podem retirar seu consentimento de participação na pesquisa a qualquer momento.

#### **Comentários e Considerações sobre a Pesquisa:**

O projeto apresenta uma descrição clara, com objetivos e metodologia bem definidos, destacando-se o uso de uma ampla variedade de fontes de dados (imagens, gravações de seqüências de teclas, observação por humanos, questionários, testes) para tornar possível a verificação do grau de correlação entre os algoritmos propostos e a observação humana.

**Endereço:** Av. Unisinos, 950

**Bairro:** Cristo Rei

**CEP:** 93.022-000

**UF:** RS

**Município:** SAO LEOPOLDO

**Telefone:** (51)3591-1198

**Fax:** (51)3590-8118

**E-mail:** cep@unisinos.br

UNIVERSIDADE DO VALE DO  
RIO DOS SINOS - UNISINOS



Continuação do Parecer: 3.707.158

**Considerações sobre os Termos de apresentação obrigatória:**

Os termos de apresentação obrigatória, que abrangem um modelo de termo de assentimento para os menores de idade e dois modelos de TCLE (um deles para os pais ou responsáveis pelos menores e o outro para os adultos participantes da pesquisa), descrevem os riscos possíveis e enfatizam o direito de retirada do consentimento de participação na pesquisa a qualquer momento.

**Recomendações:**

Não há.

**Conclusões ou Pendências e Lista de Inadequações:**

Não há.

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1362644.pdf	29/10/2019 13:42:34		Aceito
Outros	anuenciaunisinos.pdf	29/10/2019 13:41:16	TIAGO ROBERTO KAUTZMANN	Aceito
Folha de Rosto	folhaDeRosto.pdf	28/10/2019 18:32:26	TIAGO ROBERTO KAUTZMANN	Aceito
Projeto Detalhado / Brochura Investigador	projeto.doc	28/10/2019 18:27:29	TIAGO ROBERTO KAUTZMANN	Aceito
Outros	anuenciaconcordia.pdf	18/10/2019 13:36:06	TIAGO ROBERTO KAUTZMANN	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TATermodeassentimento.docx	06/09/2019 14:32:23	TIAGO ROBERTO KAUTZMANN	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLEmenoresdeidade.docx	06/09/2019 14:31:56	TIAGO ROBERTO KAUTZMANN	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLEmaioresdeidade.docx	06/09/2019 14:31:30	TIAGO ROBERTO KAUTZMANN	Aceito
TCLE / Termos de Assentimento / Justificativa de	TCLE_TiagoRobertoKautzmann.pdf	18/11/2019 09:32:24	José Roque Junges	Aceito

Endereço: Av. Unisinos, 950

Bairro: Cristo Rei

CEP: 93.022-000

UF: RS

Município: SAO LEOPOLDO

Telefone: (51)3591-1198

Fax: (51)3590-8118

E-mail: cep@unisinos.br

UNIVERSIDADE DO VALE DO  
RIO DOS SINOS - UNISINOS



Continuação do Parecer: 3.707.158

Ausência	TCLE_TiagoRobertoKautzmann.pdf	18/11/2019 09:32:24	José Roque Junges	Aceito
----------	--------------------------------	------------------------	-------------------	--------

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

SAO LEOPOLDO, 18 de Novembro de 2019

---

**Assinado por:**  
**José Roque Junges**  
**(Coordenador(a))**

**Endereço:** Av. Unisinos, 950

**Bairro:** Cristo Rei

**CEP:** 93.022-000

**UF:** RS

**Município:** SAO LEOPOLDO

**Telefone:** (51)3591-1198

**Fax:** (51)3590-8118

**E-mail:** cep@unisinos.br

## ANEXO A - QUESTIONÁRIO DE COMPETÊNCIA EMOCIONAL (QCE)

### QUESTIONÁRIO DE COMPETÊNCIA EMOCIONAL

Autor: Vladimir Taksic' (2000)  
Adaptação Portuguesa: N. Lima Santos & L. Faria (2005)

Nas páginas seguintes vai encontrar um conjunto de afirmações, às quais pedimos que responda espontaneamente, sem pensar muito em cada uma delas, pois o que nos interessa é o que habitualmente sente e pensa acerca das mesmas.

Isto não é um teste, logo, não há boas nem más respostas. O que conta é a sua opinião sincera!

**MUITO OBRIGADO** por nos ter disponibilizado o seu tempo para responder a este Questionário.

Por favor, leia cada uma das afirmações que a seguir se apresentam e responda pondo um círculo ou uma cruz na letra que corresponde à sua opinião, usando a escala que se segue:

<i>A</i> NUNCA	<i>B</i> RARAMENTE	<i>C</i> POUCAS VEZES	<i>D</i> ALGUMAS VEZES	<i>E</i> FREQUENTEMENTE	<i>F</i> SEMPRE
-------------------	-----------------------	--------------------------	---------------------------	----------------------------	--------------------

1	Consigo manter-me de bom humor, mesmo que alguma coisa má aconteça.	A	B	C	D	E	F
2	Consigo exprimir os meus sentimentos e emoções em palavras.	A	B	C	D	E	F
3	Quando encontro alguém conhecido, apercebo-me logo da sua disposição.	A	B	C	D	E	F
4	Consigo manter-me de bom humor, mesmo quando os que me rodeiam estão de mau humor.	A	B	C	D	E	F
5	Quando alguma coisa me desagrada, demonstro-o logo.	A	B	C	D	E	F
6	Quando vejo como alguém se sente, geralmente sei o que lhe aconteceu.	A	B	C	D	E	F
7	As experiências desagradáveis ensinam-me o que não devo fazer.	A	B	C	D	E	F
8	Consigo facilmente arranjar maneira de me aproximar das pessoas de quem gosto.	A	B	C	D	E	F
9	Consigo perceber a diferença, se os meus amigos estão tristes ou dececionados.	A	B	C	D	E	F
10	Quando alguém me elogia, trabalho com maior entusiasmo.	A	B	C	D	E	F
11	Consigo facilmente descrever as emoções que estou a sentir.	A	B	C	D	E	F
12	Consigo facilmente notar as mudanças de humor nos meus amigos.	A	B	C	D	E	F
13	Quando não gosto de uma pessoa, arranjo maneira de lho mostrar.	A	B	C	D	E	F
14	Exprimo bem as minhas emoções.	A	B	C	D	E	F
15	Consigo facilmente fazer feliz um amigo no dia do seu aniversário.	A	B	C	D	E	F
16	Quando estou de bom humor, é difícil ficar maldisposto(a).	A	B	C	D	E	F
17	Consigo exprimir a forma como me sinto.	A	B	C	D	E	F
18	Ao observar uma pessoa junto de outras, sou capaz de descrever bem as suas emoções.	A	B	C	D	E	F

19	Quando estou de bom humor, todos os problemas parecem ter solução.	A	B	C	D	E	F
20	Consigo descrever o meu estado emocional actual.	A	B	C	D	E	F
21	Consigo perceber quando alguém se sente desanimado.	A	B	C	D	E	F
22	Quando estou com alguém que me admira, tenho cuidado com a forma como me comporto.	A	B	C	D	E	F
23	Posso afirmar que conheço bem o meu estado emocional.	A	B	C	D	E	F
24	Consigo descrever os sentimentos de uma pessoa a partir da expressão da sua cara.	A	B	C	D	E	F
25	Quando estou feliz e bem-humorado(a), estudo e trabalho melhor.	A	B	C	D	E	F
26	O meu comportamento reflete os meus sentimentos mais profundos.	A	B	C	D	E	F
27	Consigo detetar a inveja disfarçada nos outros.	A	B	C	D	E	F
28	Se eu quiser mesmo, consigo resolver problemas que parecem sem solução.	A	B	C	D	E	F
29	As pessoas são sempre capazes de descrever o meu estado de humor.	A	B	C	D	E	F
30	Percebo quando alguém tenta esconder o seu mau humor.	A	B	C	D	E	F
31	Consigo facilmente persuadir um amigo de que não há razões para se preocupar.	A	B	C	D	E	F
32	Normalmente sei as razões porque me sinto mal.	A	B	C	D	E	F
33	Percebo quando alguém se sente culpado.	A	B	C	D	E	F
34	Procuro moderar as emoções desagradáveis e reforçar as positivas.	A	B	C	D	E	F
35	Considero fácil manifestar carinho a pessoas do sexo oposto.	A	B	C	D	E	F
36	Percebo quando alguém tenta esconder os seus verdadeiros sentimentos.	A	B	C	D	E	F
37	Não há nada de mal com o modo como habitualmente me sinto.	A	B	C	D	E	F
38	Consigo nomear e descrever a maioria dos meus sentimentos.	A	B	C	D	E	F
39	Percebo quando alguém está em baixo.	A	B	C	D	E	F
40	Cumpro os meus deveres e obrigações com prontidão, em vez de estar a pensar neles.	A	B	C	D	E	F
41	Consigo reconhecer a maioria dos meus sentimentos.	A	B	C	D	E	F
42	Percebo quando o comportamento de alguém varia em função do seu humor.	A	B	C	D	E	F
43	Tento manter o bom humor.	A	B	C	D	E	F
44	Sei como surpreender de forma agradável cada um dos meus amigos.	A	B	C	D	E	F
45	No que me diz respeito, é normal sentir o que sinto agora.	A	B	C	D	E	F

MUITO OBRIGADO PELA SUA COLABORAÇÃO!