



Programa Interdisciplinar de Pós-Graduação em

**Computação Aplicada**

**Mestrado Acadêmico**

Marinês Refosco Severo Bertola

Cardiac-M: Sistema de Apoio para Identificação de Riscos em  
Pacientes Crônicos Cardíacos

São Leopoldo, 2011

Marinês Refosco Severo Bertola

CARDIAC-M: Sistema de Apoio para Identificação de Riscos em Pacientes Crônicos Cardíacos

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre, pelo Programa de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos – UNISINOS.

Orientador: Dr. João Gluz

São Leopoldo

2011

Dados Internacionais de Catalogação na Publicação (CIP)

B546c Bertola, Marinês Refosco Severo

Cardiac-M: sistema de apoio para identificação de riscos em pacientes crônicos cardíacos / Marinês Refosco Severo Bertola. - 2011.

79 f. : il. ; 30 cm.

Apresenta bibliografia.

Dissertação (Mestrado) – Universidade do Vale do Rio dos Sinos , Programa de Pós-Graduação em Computação Aplicada , 2011.

“Orientação: Prof. Dr. João Gluz.”

1. Cardiologia – Doenças crônicas. 2. Cardiologia – Tratamento preventivo. 3. Mineração de dados - Medicina. I. Título.

CDU: 616.12

Índice para o catálogo sistemático:

- |  |              |
|--|--------------|
| 1. Cardiologia – Doenças crônicas      | 616.12       |
| 2. Cardiologia – Tratamento preventivo | 616.12-005.8 |
| 3. Mineração de dados – Medicina       | 004.6:616    |

Catalogação na fonte elaborada pelo bibliotecário  
Marcos Leandro Freitas Hübner – CRB 10/ 1253

Marinês Refosco Severo Bertola

CARDIAC-M: Sistema de Apoio à Identificação de Riscos em Pacientes Crônicos Cardíacos

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

Aprovado em \_\_\_ / \_\_\_ / \_\_\_\_

BANCA EXAMINADORA

---

---

---

## **AGRADECIMENTOS**

Agradeço a todas as pessoas que de uma forma ou outra me ajudaram no desenvolvimento deste trabalho, seja com informações ou mesmo pela força, carinho e amizade.

À minha família agradeço pelo carinho, apoio e compreensão que me deram no decorrer desta jornada e, sobretudo, por acreditarem no sucesso desta. A eles, principalmente ao meu marido e filhos, que tiveram que abdicar de minha companhia e atenção, dedico este trabalho.

Um agradecimento especial ao meu orientador, prof. Dr. João Gluz, pelo apoio, atenção e o conhecimento a mim fornecido para desenvolvimento deste trabalho.

Agradeço também aos meus amigos e colegas da instituição de saúde onde trabalho, pelo apoio na disponibilização das informações e avaliação dos resultados, com agradecimento especial ao Dr. Raul cuja participação foi de vital importância neste trabalho.

E, sobretudo, agradeço a Deus pois tenho certeza que, sem a sua presença, não teria chego onde cheguei.

## RESUMO

As mudanças sociais, econômicas e a alteração nos hábitos alimentares são algumas das razões para o crescimento do número de pacientes portadores de doenças crônicas, principalmente doenças cardíacas. Com isso, é iminente a necessidade de um acompanhamento mais severo de forma a proporcionar a estes pacientes qualidade de vida, reduzindo, se possível, o nível de mortalidade de doentes cardíacos. O trabalho desenvolvido prevê atender essa necessidade através do projeto de um sistema multiagente proativo, envolvendo aplicação de técnicas de mineração de dados para recuperação e classificação de dados do prontuário médico do paciente de forma a identificar uma possibilidade de agravamento da doença.

Palavras-chave: mineração de dados, agentes, doenças crônicas

## **ABSTRACT**

*The social, economic and dietary habit changes are some reasons for the growing number of patients with chronic diseases, especially heart disease. Thus, it is the imminent need for stricter monitoring to assure patients' quality of life by reducing, if possible, the level of mortality of cardiac patients. This work tries to meet this need by designing a proactive multi-agent system and applying data mining techniques for data classification and retrieval of patient medical records to identify a possible worsening of the disease.*

*Keywords: data mining, agents, chronic diseases*

## LISTA DE FIGURAS

Figura 1 - Percentual de mortes por doenças crônicas na Espanha em 2002 .....	12
Figura 2 - Percentual de mortes por doenças crônicas nos EUA em 2002.....	12
Figura 3 - Percentual de mortes por doenças crônicas no Brasil em 2007 .....	13
Figura 4 - O Processo de Descoberta de Conhecimento em Bancos de Dados (KDD) .....	16
Figura 5 - Uma rede de crenças Bayesianas para detectar problemas cardíacos e azia em pacientes .....	20
Figura 6 – Exemplo de Rede Bayesiana em um Classificador <i>Naïve Bayes</i> .....	21
Figura 7 - Conceito de Agente.....	26
Figura 8 - Diagrama de Objetivos .....	32
Figura 9 - Diagrama de Agrupamento de Dados .....	33
Figura 10 - Diagrama de Interação de Agentes .....	33
Figura 11 - Diagrama Geral do Sistema .....	34
Figura 12 - Diagrama Geral do Agente <i>Reuniões</i> .....	34
Figura 13 - Árvore de Decisão obtida do algoritmo C4.5 aplicado sobre a base de dados do SINAN.....	39
Figura 14 – Melhores associações entre fatores de risco cardiovasculares dos pacientes portadores de DAC .....	43
Figura 15 - Modelo GRPC utilizando os conceitos de CRM .....	44
Figura 16 - Central de Monitoramento de Pacientes Crônicos.....	45
Figura 17 - Estrutura do Cenário <i>Admission</i> , onde os nodos representam o protocolo e as setas as transições.....	46
Figura 18 – Diagrama de Caso de Uso do Cardiac-M no Atendimento do Paciente .....	49
Figura 19 - Diagrama de Caso de Uso da Interação Pró-ativa com Sistemas Legados.....	50
Figura 20 – Diagrama de Relacionamento entre Agentes .....	51
Figura 21 - Arquitetura Geral do Sistema CARDIAC-M.....	52
Figura 22 - Diagrama de Interação CARDIAC-M .....	53
Figura 23 - Tela de Login da Aplicação Cardiac-M.....	54
Figura 24 - Tela de Registro de Atendimento da Aplicação Cardiac-M.....	54
Figura 25 – Resumo do Método <i>consultaProntuario</i> da Aplicação Cardiac-M.....	55
Figura 26 - Método <i>classificar</i> da Aplicação Cardiac-M.....	56
Figura 27 - Processo de Classificação Modelado no <i>RapidMiner</i> .....	57
Figura 28 - Método <i>solicitaAvaliacaoRisco</i> da Aplicação Cardiac-M.....	58
Figura 29 - Tela para Seleção de Paciente da Protótipo da Aplicação Cardiac-M.....	59
Figura 30 - Alerta exibido ao médico .....	59



Figura 31 - Fatores de Risco Cardiovascular .....	62
Figura 32 – Critérios para Cálculo do Escore de Framingham .....	62
Figura 33 - Interface Gráfica da Ferramenta RapidMiner.....	64
Figura 34 – Diagrama do Processo de Construção do Modelos de Classificação.....	65
Figura 35 – Estrutura do Processo de Treinamento e Avaliação do Modelo NaiveBayes.....	65
Figura 36 – Estrutura do Processo de Treinamento e Avaliação do Modelo KNN .....	66
Figura 37 - Seleção do Paciente para Atendimento no Protótipo do Sistema Cardiac-M.....	69
Figura 38 - Interface para Registro de Atendimento Médico do Protótipo do Sistema Cardiac-M.....	70
Figura 39 - Alerta emitido ao médico quando o paciente é avaliado com risco por ambos classificadores.....	70
Figura 40 - Interface do Registro de Atendimento do Paciente <i>H123</i> .....	71
Figura 41 - Alerta emitido ao médico quando o paciente é avaliado com risco por apenas um dos classificadores .....	71

## LISTA DE TABELAS

Tabela 1 - Exemplo de base de dados de treinamento para aplicação do classificador <i>Naïve Bayes</i> .....	22
Tabela 2 - Cálculo das Medidas de Ponderação do Atributo Idade.....	25
Tabela 3 - Classificação da Pressão Arterial .....	36
Tabela 4 – Regras de Classificação obtidas do algoritmo C4.5 aplicado sobre a base de dados do SINAN.....	39
Tabela 5 - Análise do estilo de vida da população da incidência de DAC no Hospital Monte Sinai .....	42
Tabela 6 - Análise do estilo de vida da população da incidência de DAC no Hospital Monte Sinai .....	42
Tabela 7 - Análise antropométrica da população da incidência de DAC no Hospital Monte Sinai.....	43
Tabela 8 - Análise Comparativa dos Trabalhos Relacionados que utilizam Técnicas de Mineração de Dados .....	47
Tabela 9 - Análise Comparativa dos Trabalhos Relacionados que utilizam Tecnologia de Agentes .....	47
Tabela 10 – Tabela com os atributos extraídos do Prontuário Médico para treino do algoritmo de classificação. ....	60
Tabela 11 – Matriz de Confusão .....	66
Tabela 12 – Métricas dos Modelos de Classificação.....	67
Tabela 13 – Avaliação da Previsão dos Modelos de Classificação.....	68
Tabela 14 - Resultados dos Casos de Teste do Protótipo do Sistema Cardiac-M.....	71

## LISTA DE SIGLAS

AI – Angina Instável  
ANS – Agência Nacional de Saúde  
BBN – Rede de Crenças Bayesiana  
CIA – Comunicação Interatrial  
CID – Classificação Internacional de Doenças  
CIV – Comunicação Interventricular  
CRM – Conselho Regional de Medicina  
DAC – Doença da Artéria Coronariana  
DAG – Grafo Acíclico Direcionado  
DC – Doenças Cardiovasculares  
EF – Escore de Framingham  
EHR – Registro de Saúde Eletrônico  
ES – Engenharia de Software  
HAP – Hipertensão Arterial Pulmonar  
HFDC – Histórico Familiar de Doenças Cardiovasculares  
IAD – Inteligência Artificial Distribuída  
IAM – Infarto Agudo do Miocárdio  
IBGE – Instituto Brasileiro de Geografia e Estatística  
IMC – Índice de Massa Corpórea  
KDD – Descoberta de Conhecimento em Bancos de Dados  
OMS – Organização Mundial de Saúde  
PA – Pressão Arterial  
SBC – Sociedade Brasileira de Cardiologia  
SINAN – Sistema de Informações de Agravos de Notificação  
SQL – Linguagem de Consultas Estruturada  
SR – Sem Resposta  
TIC – Tecnologia de Informação e Comunicação  
TISS – Troca de Informações em Saúde Suplementar

## LISTA DE EQUAÇÕES

$$P(A|B) = \frac{P(A \text{ e } B)}{P(B)} \quad (1)$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (2)$$

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (3)$$

$$X^* = \frac{X - \min(X)}{\text{range}(X)} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (4)$$

$$X^* = \frac{X - \text{mean}(X)}{SD(X)} \quad (5)$$

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (6)$$

$$\text{Recall} = \frac{VP}{VP + FN} \quad (7)$$

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (8)$$

## SUMÁRIO

1 INTRODUÇÃO.....	11
1.1 Motivação e Contexto do Trabalho .....	11
1.2 Problema.....	14
1.3 Objetivo .....	14
1.4 Organização do Trabalho.....	15
2 REFERENCIAL TEÓRICO.....	16
2.1 Mineração de Dados .....	16
2.1.1 Classificadores Bayesianos.....	18
2.1.2 Classificador KNN .....	23
2.2 Agentes e Sistemas Multiagente.....	25
2.2.1 Agentes de Software.....	26
2.2.2 Sistemas Multiagente.....	30
2.2.3 Engenharia de Software Orientada a Agentes .....	31
2.3 Doenças Crônicas Cardíacas .....	35
3 TRABALHOS RELACIONADOS .....	38
3.1 Utilizando Técnicas de Data Mining para Descoberta de Padrões de Comportamento das Hepatites Virais .....	38
3.2 Identificação dos Fatores de Risco em Doenças Cardíacas utilizando Técnicas de Mineração de Dados .....	40
3.3 Modelo de Gestão do Relacionamento do Paciente Crônico (GRPC) .....	44
3.4 Sistema HCDMP .....	46
3.5 Análise Comparativa dos Trabalhos.....	47
4 ARQUITETURA DO SISTEMA CARDIAC-M.....	49
4.1 Casos de Uso .....	49
4.2 Modelo de Arquitetura.....	50
4.3 Protótipo do Cardiac-M.....	53
5 MODELOS DE ESTIMATIVA DE RISCO .....	60
5.1 Bases de Dados.....	60
5.2 Critérios de Análise dos Dados .....	61
5.3 Ferramenta de Mineração de Dados .....	63
5.4 Construção dos Modelos .....	65
6 EXPERIMENTOS E AVALIAÇÃO .....	68

6.1 Experimentos de Avaliação do Modelo de Risco.....	68
6.2 Avaliação do Protótipo do Sistema Cardiac-M.....	69
7 CONSIDERAÇÕES FINAIS .....	74
7.1 Conclusões.....	74
7.2 Trabalhos Futuros .....	75
REFERÊNCIAS BIBLIOGRÁFICAS .....	76

# 1 INTRODUÇÃO

A saúde sempre foi um fator preocupante e alvo de discussões entre órgãos governamentais e sociais. Há algum tempo, o foco era a busca do tratamento eficaz para a doença e, hoje, a promoção de saúde. As transformações sociais e econômicas das últimas décadas e as conseqüentes mudanças dos hábitos alimentares, aumento do sedentarismo e estresse, além do aumento da expectativa de vida da população colaboram para o aumento da incidência de doenças crônicas (BRASIL, 2008).

Doenças crônicas são aquelas cujo tratamento não é realizado em um tempo curto, não sendo consideradas emergências médicas. O tratamento é extenso, mais de seis meses, e apresentam efeitos de longo prazo, difíceis de prever (WHO, 2011). O diagnóstico de uma doença crônica altera significativamente a qualidade de vida e atividade das pessoas.

A maioria das doenças crônicas não tem cura, como diabetes, asma, doença de Alzheimer e hipertensão. Entretanto, várias delas podem ser prevenidas ou controladas através da detecção precoce, adoção de dieta e hábitos saudáveis, prática de exercícios e acesso ao tratamento adequado recomendado pelo profissional de saúde (BRASIL, 2008).

## 1.1 *Motivação e Contexto do Trabalho*

Atualmente, as doenças crônicas constituem um sério problema de saúde pública. Doenças cardiovasculares, câncer, diabetes e doenças crônicas respiratórias são as maiores responsáveis pela mortalidade no mundo, representando 60% de todas as mortes (WHO, 2011).

Nos Estados Unidos, em 2002, cerca de 88% das mortes foram decorrentes de doenças crônicas (WHO US, 2011). No Brasil, entre 2002 e 2004, óbitos provocados por doenças crônicas representam 74% dos óbitos do país neste período (BRASIL, 2008), reduzindo para 67,3% em 2007, devido ao resultado da aplicação de ações preventivas das principais doenças (BRASIL, 2010).

Na pesquisa nacional realizada pelo IBGE em 2008, 31,3% da população afirmam ter pelo menos uma doença crônica, cuja incidência maior é predominante nas regiões Sul e Sudeste, com 35,8% e 34,2% respectivamente. Das doenças crônicas citadas na pesquisa, o maior índice é de doenças cardiovasculares, representando 18%, seguido de doenças osteomusculares com 13,5% (IBGE, 2008).

Segundo pesquisa realizada por Musavi (2008), em 2002 90% das mortes da Espanha foram de pacientes com doenças crônicas, dos quais 33% possuíam doenças cardíacas. Complementando esses resultados, a OMS identificou fator semelhante nos Estados Unidos, onde, no mesmo período, 38% das mortes foram de pacientes portadores de doenças cardíacas (WHO US, 2011). No Brasil, em 2007, as doenças cardiovasculares foram responsáveis por 29,4% dos óbitos do país (BRASIL, 2010). Estes fatos ilustram a afirmação da forte incidência de doenças cardiovasculares na população. As figuras 1, 2 e 3 apresentam esses resultados.

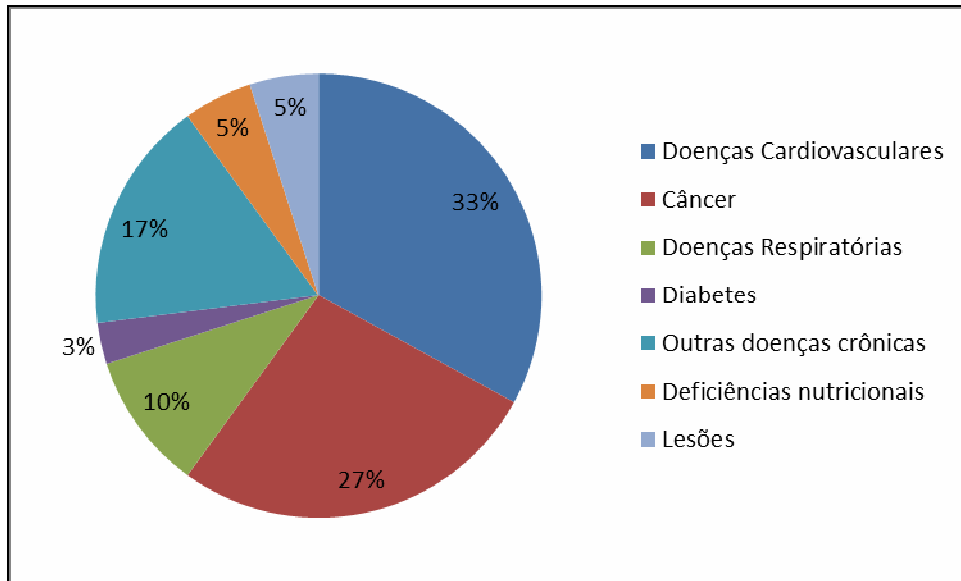


Figura 1 - Percentual de mortes por doenças crônicas na Espanha em 2002 (MUSAVI et al, 2008)

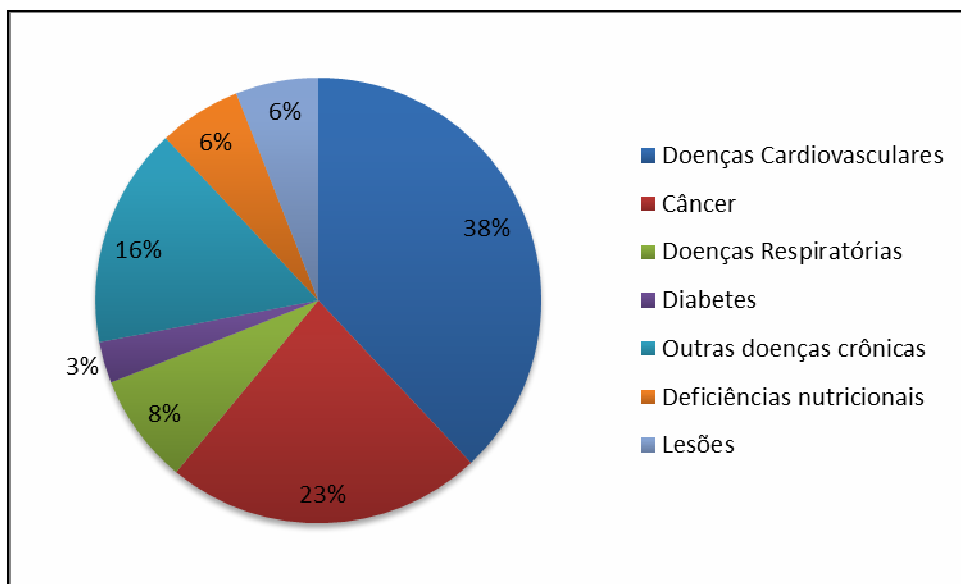


Figura 2 - Percentual de mortes por doenças crônicas nos EUA em 2002 (WHO US, 2011)



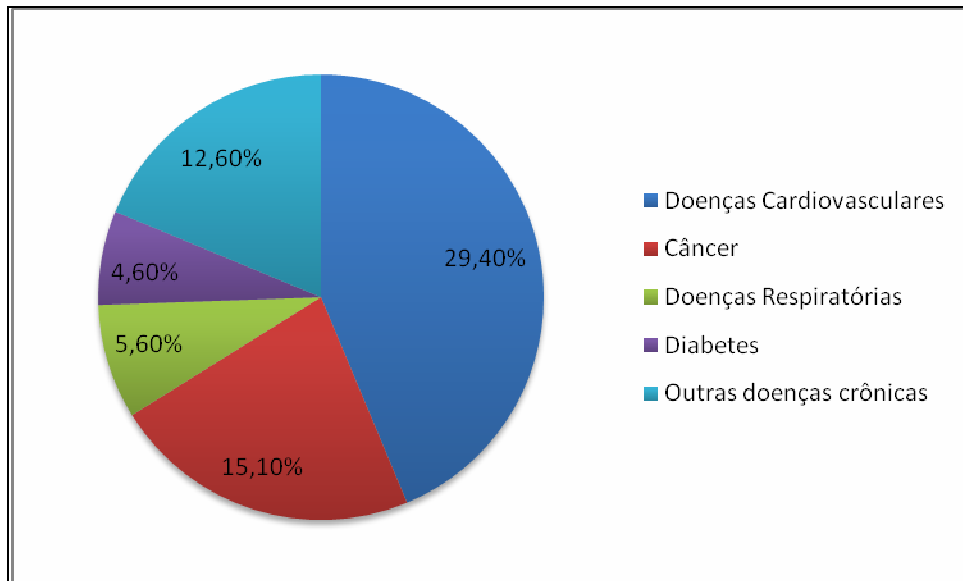


Figura 3 - Percentual de mortes por doenças crônicas no Brasil em 2007 (BRASIL, 2010)

No entanto, mais importante do que identificar um paciente crônico é identificar a gravidade da doença de forma a prevenir uma morte precoce. Dentro do universo de doentes crônicos, existe uma parcela destes doentes que podem ser classificados como *doentes crônicos de risco*, que apresentam uma probabilidade alta de sofrer um forte agravamento da doença, podendo levar a situações de risco de vida. Isto é particularmente relevante no caso das doenças cardíacas.

Assim, a identificação precoce dos doentes crônicos de risco, quando possível, oferece condições para uma terapia mais efetiva. Esta terapia irá requerer um acompanhamento efetivo do paciente, avaliando continuamente seus hábitos de vida, dietas e interação de medicamentos, conforme tratamento estabelecido.

Este acompanhamento não é trivial, uma vez que depende muito do paciente que, por várias razões, pode não seguir o tratamento adequadamente, tornando-se um potencial paciente de risco. Assim, para um processo terapêutico mais efetivo também são necessárias soluções que possam apoiar um acompanhamento *proativo* da doença crônica, auxiliando a equipe médica e o próprio paciente na gestão da doença.

Conforme Dunbar-Jacob (2001), “a falta de adesão ao tratamento é um problema dos pacientes com doenças crônicas. Aproximadamente metade dos pacientes crônicos têm dificuldades em seguir o tratamento corretamente, tornando-os incapazes de obter os resultados esperados”.

A adesão ao tratamento adequado é fundamental para a melhor gestão da doença crônica e esses benefícios atingem não somente o próprio paciente, como seus familiares, sistemas de saúde e economia do país. O sistema de saúde economiza com a redução de internações emergenciais e intervenções cirúrgicas e a economia ganha com o aumento de produtividade. Em contrapartida, a baixa adesão afeta negativamente a evolução clínica do paciente, trazendo consequências pessoais, sociais e econômicas (DEWULF et.al, 2006).

Esta situação oferece um claro incentivo para que sejam projetadas e desenvolvidos sistemas e ferramentas de Tecnologia de Informação e Comunicação (TIC) que possam apoiar a detecção e acompanhamento de doenças crônicas.

Dessa forma, a pesquisa desenvolvida na presente dissertação se insere neste contexto, propondo uma solução tecnológica para acompanhamento de pacientes cardíacos crônicos para prevenir agravamento dessas doenças e a redução dos óbitos.

Há uma preocupação em particular de se poder prever situações onde a doença cardíaca crônica passe a ser crítica, se tornando, então, um risco muito maior ao paciente. O sistema de acompanhamento deve ser capaz de avisar de forma proativa quando esta situação ocorre.

Para que isso seja possível serão empregadas técnicas avançadas de mineração de dados, modelagem probabilística, e projeto de sistemas multiagente, no processo de projeto e construção deste sistema. Essa é a maior contribuição do presente trabalho, mostrando que é viável utilizar as técnicas e modelos citados acima no projeto e desenvolvimento de um sistema de acompanhamento de doentes cardíacos crônicos.

## **1.2 Problema**

A questão tecnológica a ser pesquisada no presente trabalho busca descobrir se a utilização de um modelo de decisão probabilístico, construído através de técnicas de mineração de dados aplicadas sobre bases de dados de informações médicas, permite identificar previamente características de pacientes cardíacos de risco. Também será analisada a aplicação de tecnologias de projeto e desenvolvimento de sistemas multiagente, de forma a garantir um comportamento proativo na detecção e aviso de ocorrência de situações de risco.

## **1.3 Objetivo**

O principal objetivo deste trabalho é criar recursos computacionais que possam fornecer auxílio à prevenção de agravamento de pacientes portadores de doenças crônicas cardíacas. Na prática este objetivo se traduz na criação de uma ferramenta de apoio ao diagnóstico, capaz de identificar, com base em um modelo computacional de identificação de pacientes de risco, qual a possibilidade de um doente cardíaco crônico de se transformar em um doente crônico de risco.

Para tanto será necessário atingir toda uma série de objetivos secundários, listados a seguir:

1. Estudo e análise das áreas de mineração de dados, com ênfase nos modelos probabilísticos de decisão/classificação, e de projeto e desenvolvimento de agentes e sistemas multiagente. Estas áreas fornecerão as principais tecnologias para construir o modelo computacional e a ferramenta de auxílio ao diagnóstico definidos no presente trabalho.
2. Obtenção de uma base de dados de saúde contendo informações sobre o perfil de pacientes crônicos.
3. Análise e extração das informações dos perfis de pacientes crônicos cardíacos, com a identificação de possíveis casos de doentes crônicos com risco de agravamento através de auxílio de médico especialista.

4. Elaboração do modelo computacional probabilístico de identificação de pacientes de risco através de técnicas de mineração de dados.
5. Projeto e desenvolvimento do protótipo da ferramenta de apoio ao diagnóstico.
6. Execução de testes e experimentos de validação da ferramenta e do modelo de identificação de pacientes de risco.

#### **1.4 Organização do Trabalho**

Este trabalho está dividido em sete capítulos. Além do capítulo inicial de introdução e de um capítulo final, com as conclusões do trabalho e sugestões de trabalhos futuros, os demais capítulos são os seguintes:

- Referencial Teórico: neste capítulo serão citadas as tecnologias e estudos que serão utilizados no desenvolvimento do trabalho. São elas: técnicas de mineração de dados, tecnologia de agentes e doenças crônicas cardíacas.
- Trabalhos Relacionados: neste capítulo serão apresentados alguns trabalhos relacionados ao tratamento e acompanhamento de pacientes crônicos, utilizando as tecnologias que são utilizadas neste trabalho.
- Arquitetura do Sistema: neste capítulo são descritos os casos de uso e as principais características do projeto e arquitetura do sistema desenvolvido utilizando tecnologia de agentes e mineração de dados, além da apresentação das funcionalidades do protótipo desenvolvido.
- Modelos de Estimativa de Risco: neste capítulo são apresentados os modelos de estimativas de risco a serem utilizados no trabalho, origem dos dados e critérios utilizados para análise e construção desses modelos a partir da ferramenta de mineração *RapidMiner*.
- Experimentação e Avaliação da Ferramenta: neste capítulo serão relatados os experimentos utilizados para avaliação dos modelos e da ferramenta. Completando com a análise, do ponto de vista médico, de sua usabilidade e aplicabilidade no contexto do problema de pesquisa.

## 2 REFERENCIAL TEÓRICO

Neste capítulo é feita uma revisão bibliográfica das principais técnicas e tecnologias empregadas neste trabalho. Assim, é apresentado de forma resumida o estado da arte das áreas de mineração de dados, computação baseada em agentes e doenças crônicas cardíacas.

### 2.1 Mineração de Dados

Mineração de dados (*Data Mining*) é a análise de um grande volume de dados com o objetivo de encontrar novas relações não explícitas, padrões de informações ou mesmo prever resultados de registros futuros. Essa análise é feita utilizando tecnologias de reconhecimento de padrões bem como técnicas matemáticas e estatísticas (LAROSE, 2004).

O uso de técnicas de mineração de dados vem recebendo grandes adeptos, uma vez que as bases de dados vêm crescendo muito nos últimos anos e técnicas tradicionais de análise de dados, como consultas em SQL, aplicadas em um grande volume de dados nem sempre retornam todas as informações necessárias. Esse crescimento se deve, principalmente, ao custo reduzido para armazenamento de dados e à heterogeneidade de fontes de informações, principalmente a WEB.

Freqüentemente, o termo mineração de dados é confundido com “descoberta de conhecimento”, KDD (*Knowledge Discovery in Databases*). Segundo (TAN, STEINBACH e KUMAR, 2009), mineração de dados é o processo de descoberta automática de informações úteis em grandes depósitos de dados. Técnicas de mineração de dados são aplicadas sobre grandes bancos de dados a fim de descobrir padrões de informação que poderiam, de outra forma, permanecer ignorados. Dessa forma, a mineração de dados é uma parte integral da KDD, que é o processo geral de conversão de dados brutos em informações úteis (TAN, STEINBACH e KUMAR, 2009). Este processo consiste de uma série de passos de transformação, indo do pré-processamento dos dados até o pós-processamento dos resultados da mineração de dados, conforme visto na Figura 4.

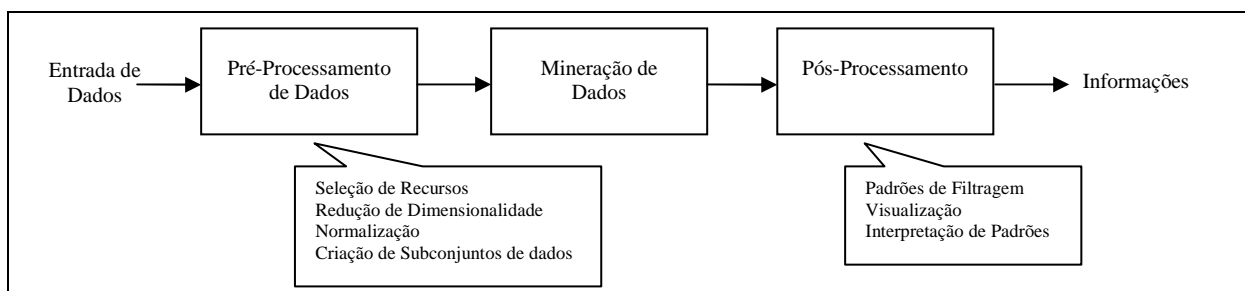


Figura 4 - O Processo de Descoberta de Conhecimento em Bancos de Dados (KDD) (TAN et al, 2009)

Todas as etapas ilustradas são importantes no processo de descoberta de conhecimento. Os passos envolvidos em cada etapa são os seguintes (TAN et al, 2009):

- **Pré-Processamento:** É considerado o processo mais trabalhoso, pois é nesta etapa que os dados devem ser preparados para análise e para aplicação das técnicas de mineração de dados. Para isso, deverão ser executadas rotinas de junção de dados (*data integration*), limpeza de dados para remoção de ruídos e dados duplicados (*data cleaning*) e a transformação dos dados no formato apropriado para executar técnicas de mineração de dados, executando agrupamentos e associações, se necessário (*data transformation*).
- **Mineração de Dados:** processo onde métodos inteligentes são aplicados para extrair padrões de informações. Neste processo, são especialmente importantes os algoritmos de aprendizagem de máquina (MITCHEL, 1997; RUSSEL e NORVIG, 2003) capazes de criar modelos computacionais que identificam estes padrões.
- **Pós-Processamento:** Nesta etapa do processo são identificados os padrões que representam conhecimento de interesse (*pattern evolution*), bem como o uso de técnicas para representação e visualização do conhecimento (*knowledge presentation*), permitindo ao analista a exploração dos dados e aplicação de métodos de teste de hipóteses para eliminar resultados incorretos do processo de *data mining*.

Por sua vez, o processo de mineração de dados pode ser detalhado nas seguintes atividades (TAN et al, 2009):

- **Modelagem Previsora:** Tem como principal objetivo prever o valor de um determinado atributo mediante análise das informações de outros atributos. Os seguintes mecanismos ou modelos de previsão são usados nesta atividade:
  - **Classificação:** Dado um conjunto de classes pré-definidas (variáveis discretas), é determinar a qual destas classes um novo item pertence. Esse método se utiliza de uma coleção de registros (conjunto de treinamento), onde um dos atributos é a classe. Aplicando o algoritmo de classificação, outro conjunto de registros (conjunto de teste) é usado para validar o modelo aplicado.
  - **Regressão:** Dado um conjunto de itens, é a análise da dependência entre os valores de atributos e, automaticamente, produzir um modelo que possa prever valores de atributos para novos itens. É utilizada para prever valores de variáveis contínuas.
- **Análise de Agrupamentos (*Clustering*):** Dado um conjunto de itens, determina-se um conjunto de classes, nos quais os itens são agrupados de acordo com suas características.
- **Análise de Associação:** Dado um conjunto de itens, é a identificação dos relacionamentos existentes entre os atributos destes itens.
- **Detecção de Anomalias:** Tem como principal objetivo identificar dados com características muito diferentes dos demais, bem como identificar uma classificação incorreta de objeto anômalo.

Neste trabalho, os modelos de classificação empregados na modelagem previsora do processo de classificação e na análise de agrupamentos serão relevantes para a construção do modelo de decisão de fatores de risco.

Dentre os diversos métodos de classificação que podem ser usados em processos de mineração de dados, neste trabalho foram avaliadas principalmente alternativas capazes de trabalhar com casos de incerteza, ou seja, foram considerados apenas métodos de classificação probabilística.

Assim as alternativas principais empregadas para a construção do modelo de decisão de doentes crônicos cardíacos de risco foram:

- Classificador de vizinho mais próximo (denominado usualmente de classificador KNN – *K-Nearest Neighbour*).
- Classificadores bayesianos, incluindo o classificador bayesiano simples (*Näive Bayes*), que trabalha com redes bayesianas simplificadas, e o classificador bayesiano completo (*BayesNet*), que trabalha com redes bayesianas com estrutura arbitrária.

Ambos tipos de classificadores são detalhados a seguir.

### **2.1.1 Classificadores Bayesianos**

As várias técnicas de classificação bayesiana são baseadas no Teorema de Bayes. Classificadores bayesianos são facilmente aplicáveis em grandes volumes de dados. Com relação ao seu desempenho, em diversos casos, como, por exemplo, classificação de textos, os classificadores bayesianos podem ser mais eficazes que árvores de decisão e redes neurais (HAN e KAMBER, 2000).

A base do processo de classificação bayesiana é a aplicação do Teorema de Bayes na descoberta da probabilidade condicional *a posteriori*  $P(B|A)$  de um determinado evento  $B$  ocorrer dado a ocorrência do evento  $A$ . O Teorema de Bayes permite obter essa probabilidade *a posteriori* a partir do conhecimento de estimativas das probabilidades de ocorrência de  $A$  e  $B$  (probabilidade prévias  $P(A)$  e  $P(B)$ ) e do conhecimento da probabilidade condicional  $P(A | B)$ .

#### Probabilidade Condicional

A noção de *probabilidade condicional* de um evento  $A$  em relação a outro evento  $B$  é definida informalmente como a chance de ocorrência do evento  $A$  dado que o evento  $B$  ocorreu. Na prática, somente faz sentido falar em probabilidade condicional nas situações em que ambos eventos  $A$  e  $B$  ocorrem. Por outro lado, dado que se assume que o evento  $B$  tenha realmente ocorrido, o que se obtém como probabilidade condicional destes eventos é o percentual que a probabilidade atribuída para a ocorrência de ambos eventos  $A$  e  $B$  tem sobre a probabilidade de ocorrência do evento  $B$ , conforme equação (1).

### Teorema de Bayes

O Teorema de *Bayes* é base de alguns dos modernos sistemas de Inteligência Artificial para inferência probabilística e sua equação é apresentada em (2).

Considere o exemplo (TAN et al. 2009) de um jogo de futebol entre duas equipes: a equipe 0 e a equipe 1, onde supõe-se que a equipe 0 vença 65% do tempo e a equipe 1 vença os jogos restantes. Entre os jogos vencidos pela equipe 0, apenas 30% deles vêm de jogos no campo da equipe 1. Por outro lado, 75% das vitórias da equipe 1 são obtidas jogando em casa. O problema do exemplo é identificar a equipe com mais chance de vencer, considerando que a equipe 1 receba o próximo jogo.

As informações do problema podem ser representadas da seguinte forma:

- A é uma variável aleatória que representa a equipe local da partida.
- B é uma variável aleatória que representa o vencedor da partida.
- A probabilidade da equipe 0 vencer é  $P(B = 0) = 0,65$ .
- A probabilidade da equipe 1 vencer é  $P(B = 1) = 1 - P(B=0) = 0,35$ .
- A probabilidade da equipe 1 ser o local e vencer é  $P(A=1|B=1) = 0,75$ .
- A probabilidade da equipe 1 ser o local e a vencedora ser a Equipe 0 é  $P(A=1|B=0) = 0,3$ .

O objetivo do problema é calcular  $P(B=1|A=1)$ , que representa a probabilidade condicional da Equipe 1 vencer a próxima partida que jogará em casa e comparar isso a  $P(B=0|A=1)$ . Utilizado o teorema de *Bayes* obtemos:

$$\begin{aligned}
 P(B=1|A=1) &= \frac{P(A=1 / B = 1) \times P(B=1)}{P(A=1)} \\
 &= \frac{P(A=1 / B=1) \times P(B=1)}{P(A=1, B=1) + P(A=1, B=0)} \\
 &= \frac{P(A=1 / B=1) \times P(B=1)}{P(A=1 / B=1) P(B=1) + P(A=1 / B=0) P(B=0)} \\
 &= \frac{0,75 \times 0,35}{0,75 \times 0,35 + 0,30 \times 0,65} \\
 &= 0,5738
 \end{aligned}$$

Portanto, a probabilidade da equipe 1 vencer a partida é 57,38% e da Equipe 0 ( $1 - P(B=1 | A=1)$ ) é de 42,62%.

## Redes Bayesianas

Uma rede de crenças *bayesiana* (BBN – *Bayesian Belief Network*), também chamada de rede *bayesiana*, fornece uma representação gráfica dos relacionamentos probabilísticos entre um conjunto aleatório de variáveis. Há dois elementos chaves de uma rede *bayesiana* (TAN et al, 2009):

- Um grafo acíclico direcionado (DAG – *Directed Acyclic Graph*) codificando os relacionamentos de dependência entre um conjunto de variáveis;
- Uma tabela de probabilidades condicionais associando cada nodo a seus nodos pais imediatos.

Uma rede bayesiana é representada por um conjunto de nós, onde cada nó representa uma variável aleatória e a ligação entre estes nós é representada através de setas que são dispostas de acordo com as dependências de probabilidade condicional entre essas variáveis. A Figura 5 ilustra um exemplo de rede de crenças Bayesianas aplicada na área de saúde:

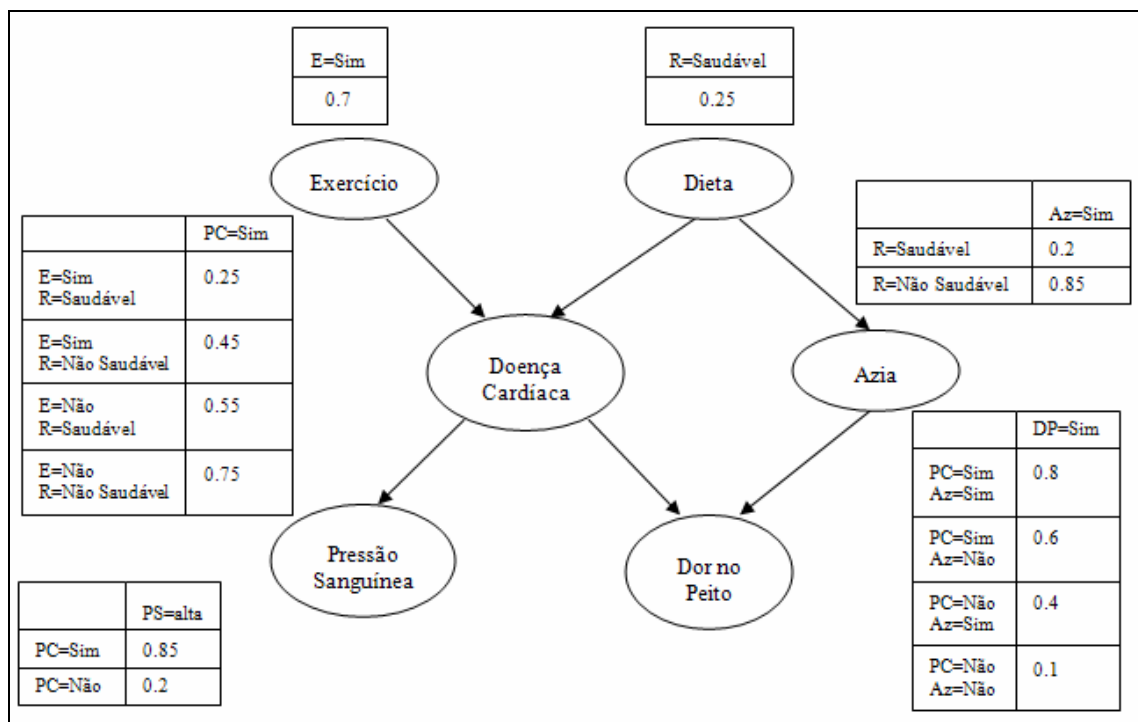


Figura 5 - Uma rede de crenças Bayesianas para detectar problemas cardíacos e azia em pacientes (TAN et al, 2009)

As redes bayesianas se propõem a estimar probabilidades futuras (a *posteriori*) de qualquer atributo ou variável aleatória usando uma definição recursiva do Teorema de Bayes (inferência bayesiana). A partir do conhecimento dos dados a *priori* são previstas as probabilidades a *posteriori*, analisando as relações de dependência entre os atributos.

Existem dois tipos de classificadores que utilizam redes *bayesianas*:

- Classificador *Bayesiano* Simples (também denominado de *Naïve Bayes*)
- Classificador baseado em Redes *Bayesianas* genéricas



### Classificador *Bayesiano* Simples

O classificador bayesiano simples tem por objetivo calcular a probabilidade que uma amostra desconhecida pertença a cada uma das classes possíveis, ou seja, prever a classe mais provável (classificação de probabilidades) (DOMINGOS, P.; PAZZANI, M., 1997).

Neste classificador os atributos da base de dados são independentes entre si, mas dependentes de um nó pai, ou seja, dependentes de um atributo da base de dados eleito como principal, da qual todos os demais possuem alguma dependência. A figura abaixo ilustra a representação deste classificador.

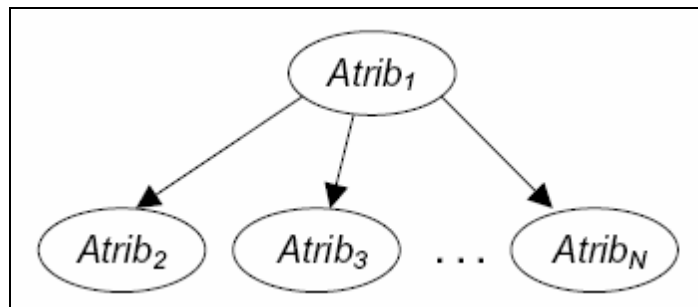


Figura 6 – Exemplo de Rede Bayesiana em um Classificador *Naïve Bayes*

Este algoritmo se utiliza de uma base de treinamento previamente classificada e possui as seguintes etapas:

1. Cálculo das probabilidades das classes:

Cada classe do conjunto de treinamento possui sua probabilidade calculada, dividindo-se o número de instâncias de determinada classe pelo número de instâncias do conjunto de treinamento.

2. Cálculo das probabilidades da amostra desconhecida:

Nesta etapa, cada valor de cada atributo da amostra possui sua probabilidade calculada para cada possível classe.

3. Cálculo das probabilidades:

As probabilidades calculadas na etapa 2 são multiplicadas e o resultado é multiplicado pela probabilidade calculada na etapa 1. Com isso, verifica-se qual a classe que possui maior probabilidade para a amostra e retorna esta classe.

De maneira geral o classificador *Naïve Bayes* é um dos métodos mais simples e eficientes em termos computacionais, apresentando robustez contra ruídos nos dados e atributos irrelevantes, de maneira que estes ruídos não venham a influenciar nas probabilidades de outros atributos (DOMINGOS e PAZZANI, 1997).

A Tabela 1 apresenta uma base de dados utilizada para treinamento do algoritmo. Um dos objetivos deste exemplo é aplicar o algoritmo para verificar se um determinado cliente, ainda não classificado, poderá ser inadimplente.

Tabela 1 - Exemplo de base de dados de treinamento para aplicação do classificador *Naïve Bayes*

Código Cliente	Sexo	Estado Civil	Escolaridade	Rendimentos	Inadimplente
1	Feminino	Solteiro	Ensino Médio Incompleto	Acima de Três Salários Mínimos	Não
2	Feminino	Solteiro	Ensino Médio Incompleto	Um Salário Mínimo	Não
3	Masculino	Solteiro	Ensino Médio Completo	Acima de Três Salários Mínimos	Sim
4	Feminino	Solteiro	Ensino Médio Completo	Um Salário Mínimo	Não
5	Feminino	Solteiro	Superior Incompleto	Dois Salários Mínimos	Não
6	Masculino	Casado	Ensino Médio Completo	Um Salário Mínimo	Não
7	Feminino	Casado	Ensino Médio Completo	Acima de Três Salários Mínimos	Não
8	Masculino	Casado	Ensino Médio Incompleto	Acima de Três Salários Mínimos	Não
9	Masculino	Casado	Ensino Médio Incompleto	Um Salário Mínimo	Não
10	Masculino	Casado	Superior Completo	Dois Salários Mínimos	Não
11	Feminino	Solteiro	Ensino Médio Completo	Acima de Três Salários Mínimos	Não
12	Feminino	Casado	Ensino Médio Incompleto	Dois Salários Mínimos	Sim
13	Feminino	Casado	Superior Incompleto	Um Salário Mínimo	Sim
14	Masculino	Casado	Superior Incompleto	Dois Salários Mínimos	Sim
15	Feminino	Solteiro	Superior Completo	Dois Salários Mínimos	Não

Fonte: Pichiliani<sup>1</sup>

A operação deste algoritmo é apresentada através de um exemplo que considera um cliente com os seguintes atributos:

Sexo: Masculino  
 Estado Civil: Solteiro  
 Escolaridade: Ensino Médio Incompleto  
 Rendimentos: Um salário mínimo

Aplicando os passos do algoritmo:

1. Cálculo das probabilidades prévias das classes:  
 $P(\text{Classe Inadimplente} = \text{Sim}) = 4/15 = 0,2667$   
 $P(\text{Classe Inadimplente} = \text{Não}) = 11/15 = 0,7334$
2. Cálculo das probabilidades da amostra desconhecida:  
 Atributo Sexo  
 $P(\text{Sexo} = \text{"Masculino"} \text{ e Inadimplente} = \text{"Sim"}) = 2/4 = 0,5$   
 $P(\text{Sexo} = \text{"Masculino"} \text{ e Inadimplente} = \text{"Não"}) = 4/11 = 0,3636$   
 Atributo Estado Civil  
 $P(\text{Estado Civil} = \text{"Solteiro"} \text{ e Inadimplente} = \text{"Sim"}) = 1/4 = 0,25$   
 $P(\text{Estado Civil} = \text{"Solteiro"} \text{ e Inadimplente} = \text{"Não"}) = 6/11 = 0,5455$   
 Atributo Escolaridade  
 $P(\text{Escolaridade} = \text{"Ensino Médio Incompleto"} \text{ e Inadimplente} = \text{"Sim"}) = 1/4 = 0,25$   
 $P(\text{Escolaridade} = \text{"Ensino Médio Incompleto"} \text{ e Inadimplente} = \text{"Não"}) = 4/11 = 0,3636$   
 Atributo Rendimentos  
 $P(\text{Rendimentos} = \text{"Um salário mínimo"} \text{ e Inadimplente} = \text{"Sim"}) = 1/4 = 0,25$   
 $P(\text{Rendimentos} = \text{"Um salário mínimo"} \text{ e Inadimplente} = \text{"Não"}) = 4/11 = 0,3636$

<sup>1</sup> Exemplo extraído do site *iMasters*, site com artigos e publicações de profissionais, estudantes e mestres em tecnologias e ferramentas voltadas para desenvolvimento WEB. Disponível no endereço [http://imasters.uol.com.br/artigo/4926/sql\\_server/data\\_mining\\_na\\_pratica\\_classificacao\\_bayesiana/](http://imasters.uol.com.br/artigo/4926/sql_server/data_mining_na_pratica_classificacao_bayesiana/). Acesso em 10 de maio de 2010.

3. Cálculo das probabilidades *a posteriori*:

$$P(\text{Classe Inadimplente} = \text{"Sim"}) = \underbrace{0,5 \times 0,25 \times 0,25 \times 0,25}_{\text{Probabilidades calculadas no passo 2}} \times \underbrace{0,2667}_{\text{Probabilidade calculada no passo 1}} = 0,0021$$

$$P(\text{Classe Inadimplente} = \text{"Não"}) = \underbrace{0,3636 \times 0,5455 \times 0,3636 \times 0,3636}_{\text{Probabilidades calculadas no passo 2}} \times \underbrace{0,7334}_{\text{Probabilidade calculada no passo 1}} = 0,0192$$

Como  $P(\text{Classe Inadimplente} = \text{"Não"}) = 0,0192 > P(\text{Classe Inadimplente} = \text{"Sim"}) = 0,0021$ , a classificação resultante do exemplo é Inadimplente = Não.

### Classificador baseado em Redes Bayesianas genéricas

Trata-se de uma abordagem mais flexível que os classificadores *bayesianos* simples, uma vez que não requer que todos os atributos sejam condicionalmente independentes. A diferença principal em relação aos classificadores *Naive Bayes* é a possibilidade de representação de dependência arbitrárias entre os subconjuntos de atributos. As Redes Bayesianas podem ser também aplicadas em tarefas de classificação, apesar da maior complexidade de computação (SAHAMI, 1998).

Quando a estrutura (topologia) da Rede Bayesiana é conhecida, incluindo os possíveis valores que as variáveis aleatórias podem assumir, então o processo de aprendizagem da rede bayesiana é relativamente simples, se restringindo a obtenção das tabelas de probabilidades prévias e condicionais para os diversos nós da rede, a partir da distribuição estatística dos valores dos dados de treinamento.

Porém quando a informação da topologia não está presente então é necessária a utilização de algoritmos mais sofisticados. Existem vários algoritmos capazes de executar este processo de aprendizagem, incluindo o algoritmo K2, que é um algoritmo de aprendizagem relativamente rápido e simples (WITTEN e FRANK, 2005), capaz de inferir uma rede bayesiana genérica a partir de uma dada classificação dos possíveis nós da rede (usualmente atributos das tabelas de treinamento). Outro algoritmo de aprendizagem eficiente de rede bayesiana é o algoritmo TAN (*Tree-Augmented Naive Bayes*) que é uma extensão do algoritmo de aprendizagem do classificador bayesiano simples, apto a identificar dependências (arcos) adicionais na rede correspondente ao classificador simples (WITTEN e FRANK, 2005).

#### **2.1.2 Classificador KNN**

O classificador KNN é um tipo de classificador baseado em exemplos que busca identificar os *k*-vizinhos mais próximos (exemplos mais próximos) do caso sendo classificado. O parâmetro *k* é usado para delimitar o número da vizinhança significativa de exemplos. Para que o algoritmo possa operar também é necessário definir uma métrica de similaridade (uma métrica de distância) entre os exemplos que permita definir a relação de proximidade (distância) entre os exemplos, necessária para estabelecer a vizinhança.

Portanto, para execução do algoritmo, faz-se necessário conhecimento de duas medidas muito importantes: número de vizinhos a considerar e distância entre os mesmos.

A escolha do número de vizinhos é muito importante, pois define a acuracidade do classificador. Um número muito pequeno, pode tornar o classificador suscetível a *overfitting*<sup>2</sup> devido ao ruído dos dados de treinamento e, um número muito grande, pode aumentar a possibilidade da informação ser classificada incorretamente, uma vez que a lista de vizinhos mais próximos pode incluir dados localizados longe de sua vizinhança (TAN et al. 2009).

A similaridade é medida pela distância  $d$ , considerando as coordenadas  $x$ ,  $y$  e  $z$ , onde (LAROSE, 2004):

1.  $d(x,y) \geq 0$ , e  $d(x,y)=0$  se  $x=y$
2.  $d(x,y) = d(y,x)$
3.  $d(x,z) \leq d(x,y) + d(y,z)$

Definidas as propriedades acima, é calculada a distância euclidiana, conforme exibido na equação (3).

Considerando o exemplo ilustrado por Larose (2004), suponha um paciente A com 20 anos de idade ( $x_1=20$ ) e medida de sódio 12 ( $x_2=12$ ) e outro paciente B com 30 anos de idade ( $y_1=30$ ) e medida de sódio 8 ( $y_2=8$ ). A distância euclidiana entre estes pontos é calculada da seguinte forma:

$$d(x,y) = \sqrt{(20 - 30)^2 + (12 - 8)^2} = \sqrt{100 + 16} = 10,77$$

O cálculo da distância considera todos os atributos dos dados de treinamento. Em situações onde somente parte dos atributos é necessário para classificar o registro ou quando o valor de um atributo influi mais do que outro no resultado da classificação, faz-se necessário efetuar um cálculo de ponderação dos atributos.

Larose (2004) cita um exemplo para ilustrar o problema: se temos um paciente do sexo masculino com 50 anos de idade, ele é similar a um paciente do sexo masculino com 20 anos de idade ou a um paciente do sexo feminino com 50 anos de idade? Qual atributo é mais relevante, a idade ou o sexo? O atributo sexo é discreto, pois assume dois valores apenas (masculino e feminino), enquanto a idade é considerada um atributo contínuo, uma vez que seu conteúdo possui uma larga escala de valores.

O cálculo para ponderação de atributos contínuos dá-se através das medidas *MinMax* e *Z-score*, cujas fórmulas são apresentadas nas equações (4) e (5). A escolha da medida a ser utilizada dá-se dos tipos de atributos considerados. Quando existem atributos contínuos e discretos, a medida *MinMax* gera o melhor resultado (LAROSE, 2004).

Considerando as informações do exemplo, inicialmente calcula-se a distância euclidiana entre o paciente A e B e a distância entre A e C:

$$d(A,B) = \sqrt{(50 - 20)^2 + 0^2} = 30$$

$$d(A,C) = \sqrt{(20 - 20)^2 + 1^2} = 1$$

---

<sup>2</sup> Termo utilizado em mineração de dados para indicar erros de classificação no conjunto de treinamento ou erros esperados no modelo em registros não vistos anteriormente. Isso geralmente acontece em situações em que o conjunto de treinamento é pequeno demais, fazendo com que não se tenha conhecimento de toda a estrutura dos dados, ou quando for grande demais, situação em que o modelo fica muito especializado e não consegue classificar novos registros (TAN et.al, 2009).

Considerando a análise pelas distâncias acima calculadas, o paciente do sexo feminino com 50 anos de idade (paciente B) é mais similar ao paciente com 50 anos do sexo masculino (paciente A) do que o paciente do sexo masculino com 20 anos de idade (paciente C). Consideramos que a variação é 50, o valor mínimo é 10 e a média é 45. Devido à característica do atributo idade ser contínuo é necessário o cálculo ponderado deste atributo para confirmar a classificação acima.

A tabela abaixo apresenta o resultado das medidas de ponderação calculadas para o exemplo citado.

Tabela 2 - Cálculo das Medidas de Ponderação do Atributo Idade

Paciente	Idade	Idade <sub>MNN</sub>	Idade <sub>Zcore</sub>	Sexo
A	50	$\frac{50-10}{50} = 0,8$	$\frac{50-45}{15} = 0,33$	Masculino
B	20	$\frac{20-10}{50} = 0,2$	$\frac{20-45}{15} = -1,67$	Masculino
C	50	$\frac{50-10}{50} = 0,8$	$\frac{50-45}{15} = 0,33$	Feminino

Fonte: Larose (2004) (adaptado pelo autor)

Com o resultado ponderado do atributo idade e calculando as distâncias novamente, temos:

Utilizando a medida Min-Max, o paciente A é mais similar ao paciente B:

$$d_{MNN}(A, B) = \sqrt{(0,8 - 0,2)^2 + 0^2} = 0,6$$

$$d_{MNN}(A, C) = \sqrt{(0,8 - 0,8)^2 + 1^2} = 1,0$$

Utilizando a medida Z-score, o paciente A é mais similar ao paciente C:

$$d_{Zcore}(A, B) = \sqrt{(0,33 - (-1,67))^2 + 0^2} = 2,0$$

$$d_{Zcore}(A, C) = \sqrt{(0,33 - 0,33)^2 + 1^2} = 1,0$$

Considerando as medidas e a característica dos atributos (contínuos e discretos), o resultado escolhido para classificação é a medida *MinMax*, concluindo que o paciente A é mais similar ao paciente B.

## 2.2 Agentes e Sistemas Multiagente

A complexidade crescente das aplicações computacionais, tem levado ao surgimento de novos modelos de projeto e desenvolvimento de sistemas de computação, dentre eles o modelo baseado em agentes. A tecnologia de agentes tem adquirido nos últimos anos uma importância cada vez maior em muitos aspectos da computação, principalmente na área de

Inteligência Artificial Distribuída (IAD) que é formada pela intersecção entre as áreas de Inteligência Artificial e Computação Distribuída.

A área de Inteligência Artificial tem alcançado grandes progressos desde a década de 1970. Vários métodos de resolução de problemas, planejamento de ações e mecanismos de aprendizagem foram propostos, testados, analisados e utilizados (RUSSEL e NORVIG, 2003). Uma característica que se tornou comum nestes desenvolvimentos é a utilização de uma metáfora de computação baseada no comportamento de um ser humano como um agente inserido em um ambiente: um agente que percebe fatos e eventos, raciocina, planeja e toma decisões, age sobre o ambiente e interage com outros agentes, além de buscar aprender novos conhecimentos e habilidades (RUSSEL e NORVIG, 2003).

O modelo de agentes surge como um paradigma integrador dessas tecnologias. Além disso, o modelo de arquitetura de uma aplicação orientada a agentes é tipicamente distribuído em um conjunto de agente, formando um *Sistema Multiagente*. Os Sistemas Multiagente são caracterizados pela comunicação e, ao menos em termos da IAD, cooperação entre diversos agentes que interagem em busca de um objetivo, que pode ser comum a todos.

### 2.2.1 Agentes de Software

Os agentes de *software* são originados das pesquisas em IAD, sendo projetados para atender domínios complexos de problemas (WEISS, 1999). Desde seu início como área de pesquisa independente, mas relacionada a IAD (BOND A.H.; GASSER, L., 1988), a área de agentes e Sistemas Multiagente tem utilizado várias definições para o termo agente de software ou agente artificial.

Apesar disso, uma definição em particular tem se firmado em trabalhos mais recentes. Essa definição, dada originalmente por Wooldridge e Jennings (1995), tem sido adotada de maneira consistente em diversos textos mais recentes (BORDINI et al. 2007; RUSSEL S.J; NORVIG, P, 2003; WOOLDRIDGE, 2002; WEISS, 1999).

Seguindo a definição dada em Wooldridge (2002) um agente de *software* (ou um agente artificial) é um sistema computacional, situado em um determinado ambiente, que é capaz de executar ações de forma autônoma, a fim de satisfazer seus objetivos. Na figura abaixo, é ilustrado um agente e sua interação com o ambiente em que está inserido.

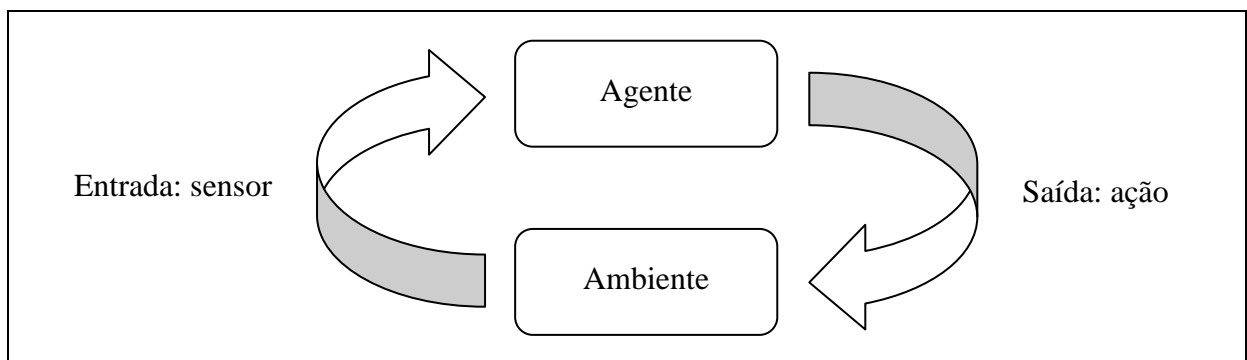


Figura 7 - Conceito de Agente (WOOLDRIDGE, 2002) (adaptado pelo autor)

### Características dos Agentes

Nesse contexto, as características básicas que distinguem a noção de agente de software das noções mais usuais de processos ou objetos de software são dadas pelos requisitos de autonomia, reatividade, pró-atividade e sociabilidade (WOOLDRIDGE e JENNINGS, 1995; BORDINI et al., 2007):

- *Autonomia*: capacidade de tomar iniciativa e ter um controle sobre suas próprias ações. Na sua forma mais elementar, o conceito de autonomia implica na capacidade de que o agente tem de se comportar de maneira independente, enquanto busca atingir os objetivos que lhe foram delegados. Portanto, quando um agente toma uma decisão específica a respeito de como atingir um dado objetivo, esta decisão e as ações implicadas estão sob o controle do agente, não sendo dirigidas por outros agentes artificiais ou humanos.
- *Reatividade*: propriedade que permite aos agentes perceberem seus ambientes e responderem adequadamente e em tempo hábil as mudanças nele ocorridas.
- *Pro-atividade*: o comportamento do agente deve ser dirigido a objetivos, com capacidade de tomar a iniciativa na execução das ações necessárias para alcançar seus objetivos.
- *Sociabilidade*: habilidade que os agentes possuem de se comunicar (interagir) com os outros agentes ou pessoas, no momento adequado, para concluir suas tarefas ou auxiliar outros agentes.

Um agente inteligente é um refinamento de um agente de software. Um agente inteligente é basicamente um agente de software com capacidade de inteligência, podendo ser definido como um agente com percepção dinâmica do ambiente e capacidade de *representação de conhecimento e raciocínio* (inferência) para interpretação das percepções, planejamento e decisão das ações autônomas que deverão ser feitas para cumprir o seu objetivo, além da avaliação dos resultados dessas ações (WOOLDRIDGE, 2002).

Um agente inteligente também pode requerer, eventualmente, capacidades adicionais, como aprendizagem, necessárias para suportar requisitos como flexibilidade e adaptabilidade de comportamentos. A seguir é apresentada uma lista de características secundárias de agentes de software, que apesar de não serem definidoras da noção de agência, são relevantes em vários domínios de aplicação (WOOLDRIDGE, 2002).

- *Continuidade temporal*: capacidade de continuarem ativos após uma mudança de ambiente;
- *Coerência*: propriedade que o agente possui de resolver conflitos entre objetivos conflitantes ou concorrentes;
- *Capacidade de Adaptação*: capacidade do agente de modificar seu comportamento em função de experiências anteriores;
- *Mobilidade*: capacidade do agente de transportar-se de uma máquina para outra;
- *Capacidade de Aprendizagem*: capacidade que um agente deve possuir para executar uma tarefa com maior eficiência do que em execuções anteriores;
- *Flexibilidade*: habilidade dos agentes de escolher dinamicamente as ações e a seqüência de execução das mesmas, em resposta a um estado do ambiente;

- Persistência: necessidade de manter um estado interno conciso através do tempo;
- Personalização: capacidade do agente de personalizar tarefas e ajudar as pessoas a desenvolvê-las da melhor maneira possível.

### Arquiteturas para Agentes Individuais

Existem diversas arquiteturas para agentes individuais. As mais conhecidas na literatura são: Reativa, *Subsumption*, Cognitiva e em Camadas (COPPIN, 2004):

- Arquitetura Reativa: baseia-se no modo estímulo-resposta, onde suas ações são provocadas pelo resultado de outra ação já realizada.
- Arquitetura Cognitiva: baseada em atitudes e estados mentais. O principal exemplo é a arquitetura BDI (do Inglês, *Belief, Desire and Intentions*), que utilizam crenças (conhecimento provisório), desejos (objetivos) e intenções (planos em execução) como principais atitudes e estados mentais.
- Arquitetura *Subsumption*: um agente nessa arquitetura possui conjuntos de entradas, saídas e módulos organizados em camadas, capazes de controlar aspectos do comportamento do agente. Os módulos são projetados na forma de máquinas aumentadas de estados finitos.
- Arquitetura em Camadas: vários tipos de arquitetura em camadas foram propostos na literatura. Elas podem ser divididas em arquiteturas de camadas horizontais, onde cada nível pode receber as percepções e contribuir para as ações, ou camadas verticais, onde as percepções são recebidas no nível inferior, sendo posteriormente repassadas aos níveis superiores. As ações são planejadas nos níveis superiores, podem ser executadas diretamente nesses níveis ou repassadas aos níveis inferiores.

### Arquitetura BDI

A arquitetura de interesse para este trabalho é a BDI, pois permite a representação de modelos cognitivos, podendo eventualmente ser combinada com mecanismos de inferência bayesiana de forma a criar uma arquitetura híbrida (VICARI, R.M.; GLUZ, J.C., 2007; GLUZ et al., 2006; VICARI et al., 2008).

O modelo BDI representa uma arquitetura cognitiva para agentes inteligentes, baseada em estados mentais, e tem sua origem no modelo de raciocínio prático humano (BRATMAN et al, 1988). O nome atribuído ao modelo é justificado pelos seus estados mentais: crença, desejo e intenção (*belief, desire and intention*). Uma arquitetura baseada no modelo BDI representa seus processos internos através dos estados mentais acima citados, e define um mecanismo de controle que seleciona de maneira racional o curso das ações (RAO, A.S.; GEORGEFF, M., 1991).

A idéia principal dessa abordagem se concentra no fato de que o agente cognitivo possui estados internos que se relacionam com estado do ambiente com o qual interage. Estes estados, usualmente denominados de estados mentais, apresentam um vínculo com mundo em termos de sua existência e significância. Os estados mentais podem ser classificados em duas categorias: estados mentais de informação e estados mentais pró-ativos. A primeira categoria



está relacionada à informação que o agente possui acerca do mundo que ele ocupa como, por exemplo, crença e conhecimento. Os estados mentais pró-ativos são aqueles que de alguma maneira guiam as ações do agente como, por exemplo, os desejos e as intenções. A seguir, detalhamos os três estados mentais adotados pelo modelo BDI.

As crenças representam o conhecimento sobre o mundo no qual o agente está inserido. Do ponto de vista computacional, crenças são apenas uma maneira de representar o estado do mundo, seja através de variáveis, uma base de dados relacional, ou expressões simbólicas em um cálculo de predicados. As crenças são essenciais porque o mundo é dinâmico (os eventos passados precisam ser lembrados), e os sistemas têm apenas uma visão local do mundo (eventos fora da sua esfera de percepção devem ser lembrados) (GEORGEFF et al., 1999). (Wooldridge, 2000) define as crenças como informações que um agente tem acerca do mundo no qual ele se encontra. Essas crenças podem ser incompletas ou incorretas, apresentando um conhecimento provisório a respeito do mundo, que pode ser modificado com o passar do tempo.

Os desejos são relacionados ao estado de mundos que o agente quer provocar. Os desejos não dirigem necessariamente o agente a agir, isto é, o fato de um agente possuir um desejo não significa agir para satisfazê-lo. Significa que antes de um determinado agente decidir o que fazer, ele passa por um processo de racionalização, confrontando os seus desejos com as suas crenças e escolhendo os desejos que são possíveis de acordo com um critério. Os desejos resultantes desse processo, passam a ser intenções do agente.

As intenções correspondem a estados de mundo que o agente quer atingir. Intenções podem ser consideradas um subconjunto dos desejos, mas ao contrário destes, devem ser consistentes. As intenções são formadas a partir de um processo de deliberação e a partir do refinamento de outras intenções. No entanto, um agente pode conter intenções iniciais inseridas pelo usuário. Normalmente, o termo intenção é empregado tanto para caracterizar um estado mental quanto para caracterizar uma ação. O estado mental intenção está direcionado para o futuro e não desencadeará obrigatoriamente uma ação.

O processo deliberativo consiste em formar as novas intenções do agente com base nas crenças, desejos e intenções atuais do mesmo. Normalmente, o processo é formado por duas etapas:

- Geração de opções: esta etapa consiste na escolha de um conjunto opções (desejos) levando em conta as crenças e as intenções atuais do agente;
- Filtragem: esta etapa tem como objetivo escolher a melhor alternativa gerada pela etapa anterior, a qual formará a nova intenção.

O raciocínio meio-fim é um processo que está associado à idéia de planos, os quais têm como objetivo realizar as intenções do agente. A IA elaborou diversas soluções para o problema de planejamento, como métodos de busca em espaços de possíveis seqüências de ações (BRATMAN et al, 1988). Porém, uma das abordagens mais comuns para o problema do planejamento consiste em uma biblioteca de planos prontos, que contém um conjunto de ações aplicável a uma determinada situação.

Quando uma intenção é selecionada através de um processo deliberativo, o agente deve se comprometer com a realização da mesma. A estratégia de comprometimento implica em persistência temporal e específica quando um agente deve abandonar uma intenção. Podemos classificar as estratégias de comprometimento em (WOOLDRIDGE, 2002):

- Cega (*blind*): um agente manterá uma intenção até que ela seja realizada;
- Obcecada (*single-minded*): um agente manterá uma intenção até que ela seja realizada ou que ele chegue a conclusão que não é mais possível realizá-la;

- Liberal (*open-minded*): uma intenção é mantida enquanto o agente crer que é possível realizá-la.

Normalmente, um agente reconsidera suas intenções após as seguintes condições serem satisfeitas: conclusão de um plano, realização de suas intenções atuais ou impossibilidade de realização das mesmas. Este agente é chamado de ousado. Caso um agente reconsidere suas intenções durante a execução de um plano, possivelmente após a execução de cada ação, ele é chamado de cauteloso (WOOLDRIDGE, 2000).

### **2.2.2 Sistemas Multiagente**

Os Sistemas Multiagente são basicamente sistemas distribuídos formados por vários agentes que interagem uns com os outros, tipicamente por meio de troca de mensagens (WOOLDRIDGE, 2002). No caso mais geral, cada agente possui objetivos distintos, que podem estar em conflito uns com os outros. Isso ocorre, por exemplo, quando agentes representam diferentes usuários, que possuem objetivos, motivações e interesses distintos a serem satisfeitos pelo sistema (um Sistema Multiagente que implementa um mercado eletrônico é um exemplo concreto dessa situação). Por outro lado, no caso da IAD o Sistema Multiagente busca atingir um objetivo comum, que deve ser distribuído entre os diversos agentes. Porém, independente dessas condições, os agentes de um Sistema Multiagente devem ser capazes de cooperar, coordenar e negociar uns com os outros para poder atingir seus objetivos (WOOLDRIDGE, 2002).

Em resumo, os Sistemas Multiagente são sistemas constituídos de múltiplos agentes que interagem ou trabalham em conjunto de forma a realizar um determinado conjunto de tarefas ou objetivos. Esses objetivos podem ser comuns a todos os agentes ou não. Dependendo da finalidade da aplicação, os agentes dentro de um sistema multiagente podem ser heterogêneos ou homogêneos, colaborativos ou competitivos, etc.

De acordo com (BELLIFEMINE et al, 2007) existem três principais desafios para o projeto e desenvolvimento de um Sistema Multiagente:

- O primeiro desafio diz respeito à forma de comunicação a ser empregada entre os agentes: como ela será efetivamente realizada entre os agentes e que tipo de protocolos de comunicação serão empregados.
- O segundo desafio é referente à interação entre os agentes: como essa interação ocorrerá, quais são seus efeitos e que linguagem os agentes devem usar para interagirem entre si e organizar seus esforços.
- Por último, o terceiro desafio diz respeito a coordenação do trabalho entre os agentes: como garantir a coordenação entre os agentes para que haja uma coerência na solução do problema que estão tentando resolver ou do objetivo que estão tentando alcançar.

#### Comunicação entre Agentes

A comunicação é fundamental para permitir que haja colaboração, negociação, cooperação entre entidades independentes. Nos Sistemas Multiagentes a comunicação deve ser até certo ponto padronizada para seja possível, necessitando assim uma linguagem que possa ser compreendida pelos agentes presentes no ambiente. Assim, a principal forma de

comunicação entre agentes ocorre por meio de uma Linguagem de Comunicação entre Agentes (*Agent Communication Language - ACL*) (LABROU et al., 1999; CHAIB-DRAA e DIGNUM, 2002).

A principal linguagem de comunicação utilizada atualmente é a linguagem FIPA-ACL (FIPA, 2002), disponível em diversas implementações, denominadas *Plataformas de Agente*. Uma plataforma amplamente disseminada atualmente é a plataforma JADE (BELLIFEMINE et al, 2007) que foi originalmente desenvolvida nos laboratórios da TILAB (Telecom Italia Lab, originalmente CSELT). A plataforma JADE foi implementada inteiramente em JAVA, suportando agentes desenvolvidos em tal linguagem. Para a implementação dos agentes, a plataforma JADE disponibiliza um conjunto de classes que devem ser especializadas para criar os agentes. Por exemplo, o comportamento de um agente em particular é criado através de uma especialização da classe pré-definida *Behaviour*. A plataforma JADE é distribuída sob a licença pública GNU.

### Interações

A forma de interação que ocorre entre os agentes é um fator muito importante na integração destes. As linguagens de comunicação e sua expressividade definem a capacidade de comunicação de cada agente. Ela deve ser universal e partilhada por todos os agentes, ser concisa e ter um número limitado de primitivas de comunicação (WEISS, 1999).

Além disso, protocolos de interação (FIPA, 2002) devem ser definidos de forma que as principais formas de interação que possam ocorrer em um Sistema Multiagente estejam devidamente documentadas. Isto deve incluir, dentre outras possibilidades, a forma como os agentes podem obter informações uns dos outros (consultas ou *queries*), a forma como distribuem tarefas, solicitam ações ou recurso, ou ainda como negociam objetivos ou atividades em comum.

### Coordenação entre Agentes

A coordenação é uma das mais fundamentais formas de interação. Segundo (WEISS, 1999) a coordenação é uma característica fundamental para os agentes de um Sistema Multiagente que executam alguma atividade em um ambiente compartilhado. A coordenação está relacionada ao compartilhamento de conhecimento entre os agentes, tendo como principal objetivo tornar as ações de cada agente individual coordenadas para se atingir o objetivo final do sistema.

## **2.2.3 Engenharia de Software Orientada a Agentes**

O projeto e desenvolvimento de agentes individuais e de Sistemas Multiagente pode se tornar uma tarefa bastante complexa, necessitando, então, do apoio de ferramentas e métodos de Engenharia de *Software* (ES) (WOOLDRIDGE e JENNINGS, 1999; JENNINGS, 2001). Para tanto, nos últimos anos foram criadas várias metodologias de ES específicas para o projeto e desenvolvimento de sistemas orientados (ou baseados) em agentes (DE WOLF e HOLVOET, 2005; DELOACH, 2005; VICARI e GLUZ, 2007; PADGHAN e WINIKOFF,

2002, 2004). Em conjunto, estas metodologias formam a área de *Engenharia de Software Orientada a Agentes* que pesquisa como métodos e processos de ES poderiam ser aplicados no projeto e desenvolvimento de agentes e Sistemas Multiagente (HENDERSON-SELLERS e GIORGINI, 2005). A ES Orientada a Agentes se preocupa tanto com a adaptação de metodologias consagradas de ES para trabalhar com agentes, quanto com a criação de novas metodologias específicas para *softwares* e sistemas orientados a agentes.

Dentre as diversas metodologias de ES orientadas a agente atualmente existentes, neste trabalho se optou pelo uso da metodologia Prometheus porque esta metodologia fornece um suporte completo e detalhado para todas as fases do processo de projeto e desenvolvimento de software. A Prometheus fornece apoio para as fases de especificação, projeto, implementação e teste/depuração do sistema. Esta metodologia também fornece mecanismos de estruturação hierárquica que permitem que a modelagem de um dado sistema possa ser feita em vários níveis de abstração e, como ferramenta de desenvolvimento, disponibiliza a ferramenta denominada PDT (Prometheus Design Tool), fácil de utilizar e que permite a criação e edição dos diversos diagramas empregados na Prometheus (PADGHAM; WINIKOFF, 2004).

A metodologia Prometheus possui as seguintes fases: especificação do sistema, projeto da arquitetura e projeto detalhado dos agentes, que são apresentadas a seguir.

### Fase de Especificação do Sistema

É nesta fase que é feita a identificação dos objetivos e funcionalidades básicas do sistema, sendo definidos os dados de entrada (percepções) e de saída (ações), além de qualquer outra informação importante na troca de dados. Para definição dos objetivos, a metodologia prevê um diagrama para ilustrar os objetivos identificados, conforme exibido na Figura 8.

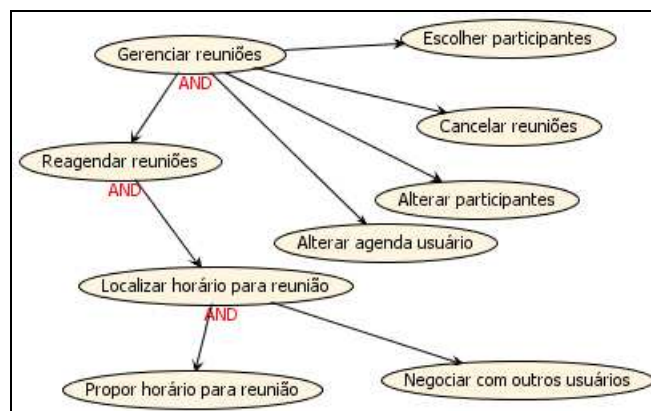


Figura 8 - Diagrama de Objetivos (HENDERSON-SELLERS e GIORGINI, 2005) (adaptado pelo autor)

A partir de um conjunto de objetivos, são definidas as funcionalidades e os cenários de caso de uso. Para cada um desses itens, a metodologia prevê um padrão de informações (descriptor) que deve ser preenchido. No caso das funcionalidades, no descriptor deve constar: nome, descrição, lista de ações e percepções relevantes, dados usados e produzidos e uma breve descrição das interações com outras funcionalidades. Já, na definição dos cenários, é importante descrever internamente os passos de cada funcionalidade do sistema.

## Fase de Projeto da Arquitetura do Sistema

Nesta fase que são identificados os agentes e como eles interagem. É nesta fase em que a metodologia prevê o maior número de diagramas para representação do projeto. A primeira etapa a ser desenvolvida nesta fase é identificar os agentes e isto é feito através do agrupamento de funcionalidades. O diagrama apresentado na Figura 9 ilustra esta etapa.

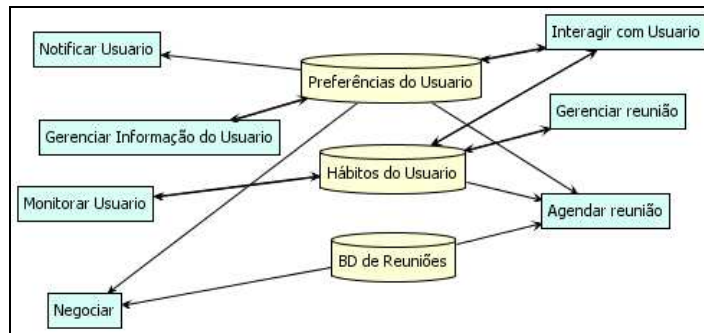


Figura 9 - Diagrama de Agrupamento de Dados (HENDERSON-SELLERS e GIORGINI, 2005) (adaptado pelo autor)

Uma vez identificado os agentes, o próximo passo consiste em definir a interação entre eles e isto é feito através da criação de protocolos. O diagrama de interação de agentes disponível pela metodologia é apresentado na Figura 10.

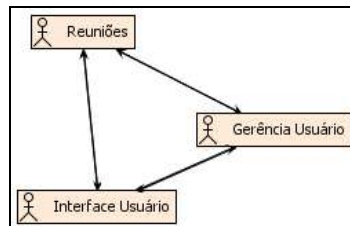


Figura 10 - Diagrama de Interação de Agentes (HENDERSON-SELLERS e GIORGINI, 2005) (adaptado pelo autor)

É nesta fase que um dos principais diagramas é desenvolvido: o diagrama geral do sistema. Neste diagrama são apresentados todos os agentes do sistema, as percepções, ações, mensagens, protocolos de interação, e dados compartilhados. Na Figura 11 é ilustrado um exemplo deste diagrama.

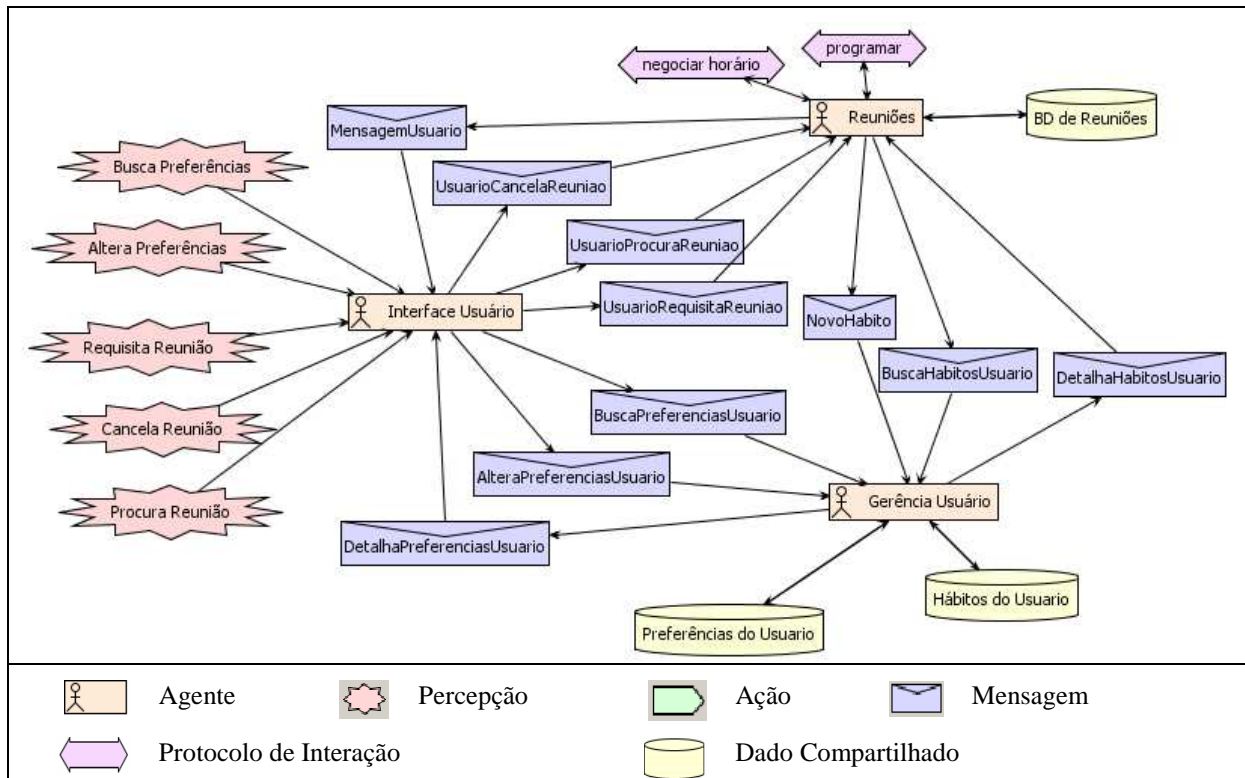


Figura 11 - Diagrama Geral do Sistema (HENDERSON-SELLERS e GIORGINI, 2005) (adaptado pelo autor)

Fase de Projeto Detalhado dos Agentes

Nesta fase são detalhadas as características de cada agente identificado na fase anterior, devendo ser especificado como cada agente irá cumprir suas tarefas. Assim são definidas as capacidades, eventos internos, e planos dos agentes, além dos detalhes das estruturas de dados usadas pelos agentes. Nesta fase leva-se em conta a plataforma de implementação. O primeiro passo nesta fase é definir as capacidades. Elas devem ser descritas com as informações: nome, descrição, informação sobre a interface externa (eventos de entrada e saída), interação com outras capacidades, informação sobre dados lidos e escritos pela capacidade. O Diagrama Geral do Agente apresentado na Figura 12, modela as capacidades de um agente e o fluxo de eventos entre as capacidades.

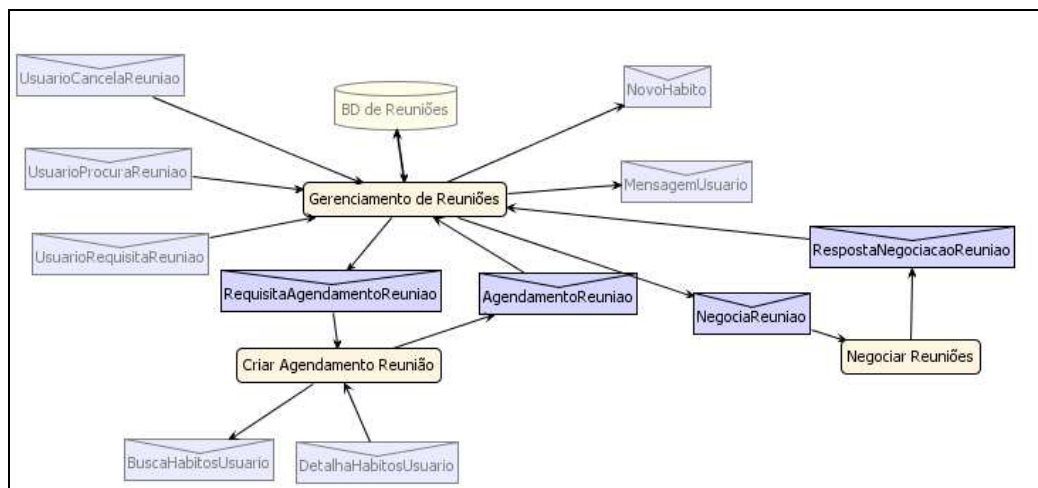


Figura 12 - Diagrama Geral do Agente *Reuniões* (HENDERSON-SELLERS e GIORGINI, 2005) (adaptado pelo autor)

A próxima etapa de modelagem da fase detalhada é a descrição dos planos de ação, eventos e dados, onde o descritor de planos contém o identificador do plano, descrição, evento que dispara o plano, passos do plano, contexto de uso e dados lidos e escritos. O descritor de eventos descreve o propósito do evento e os dados que este carrega e o descritor de dados, descreve os campos e os métodos utilizados para guardar os dados (HENDERSON-SELLERS e GIORGINI, 2005).

### **2.3 Doenças Crônicas Cardíacas**

Doenças cardíacas ou cardiovasculares são doenças que afetam o sistema circulatório. São as doenças constantes no capítulo IX da Classificação Internacional de Doenças (CID): doença reumática do coração, doença hipertensiva, doença isquêmica do coração, doenças cerebrovasculares, entre outras (LAURENTI, 1982).

No início do século XX percebeu-se um crescimento elevado da incidência de doenças cardiovasculares na população. Isso deve-se principalmente ao aumento da expectativa de vida que hoje supera os 70 anos de idade.

No entanto, nem todas as doenças cardíacas são doenças crônicas. Conforme estipulado pelo Ministério da Saúde (SAUDE, 2010), as doenças cardíacas que são consideradas crônicas são:

- Síndrome clínica de insuficiência cardíaca:

Insuficiência cardíaca é a incapacidade do coração de oferecer oxigênio suficiente aos tecidos. A insuficiência cardíaca crônica usualmente manifesta-se por falta de ar, edema e fadiga. Essa síndrome é ilustrada pela imagem de um coração dilatado com reduzida contração do ventrículo esquerdo (MESQUITA et al. 2004).

Segundo Mesquita (2004), a insuficiência cardíaca é uma das prioridades entre as doenças crônicas da Organização Mundial de Saúde, estimando que quinze milhões de pessoas no mundo inteiro sejam portadores desta doença, onde dois milhões e meio são brasileiros. Além disso, a insuficiência cardíaca é um das maiores causas de mortalidade, com um aumento de 127% nos Estados Unidos entre os anos de 1979 e 1997, além de aumentar em 40% o número de internações hospitalares no mesmo período.

- Hipertensão arterial:

A pressão arterial é a força que o sangue exerce para bombear o sangue. A hipertensão (ou pressão alta) ocorre quando esta tensão é maior, danificando os vasos sanguíneos, tornando-se um fator de risco para acidentes cardiovasculares e outras doenças (SBC, 2006).

A tabela 3 exhibe a classificação da pressão arterial segundo a Sociedade Brasileira de Cardiologia.

Tabela 3 - Classificação da Pressão Arterial

Classificação	Pressão sistólica (mmHg)	Pressão diastólica (mmHg)
Ótima	< 120	< 80
Normal	< 130	< 85
Limítrofe	130-139	85-89
Hipertensão estágio 1	140-159	90-99
Hipertensão estágio 2	160-179	100-109
Hipertensão estágio 3	≥ 180	≥ 110
Hipertensão sistólica isolada	≥ 140	< 90

*Quando as pressões sistólica e diastólica de um paciente situam-se em categorias diferentes, a maior deve ser utilizada para classificação da pressão arterial.*

Fonte: Sociedade Brasileira de Cardiologia (2006)

A hipertensão arterial é a principal causa da mortalidade por doenças cardíacas, respondendo a 40% das mortes por acidente vascular cerebral e 25% por doenças coronarianas (SBC, 2006).

- Hipertensão arterial pulmonar:

A hipertensão arterial pulmonar (HAP) é definida como uma anormalidade circulatória caracterizada por aumento da resistência vascular na pequena circulação, envolvendo vasoconstrição, remodelamento da parede arterial, trombose, causando insuficiência do ventrículo direito e morte precoce (CALLOU e RAMOS, 2009).

- Valvulopatias:

As válvulas ajudam o sangue a circular dentro do coração, abrindo e fechando para permitir o fluxo do sangue em apenas um sentido. As valvulopatias são problemas das válvulas cardíacas que impedem de se abrir ou fechar totalmente.<sup>3</sup>

- Anginas:

A angina é uma doença caracterizada pela dor ou desconforto no tórax, mandíbula, ombro, dorso ou membros superiores, geralmente desencadeada através de exercício físico ou estresse emocional. Nesta doença, existe o comprometimento de uma ou várias artérias coronárias (SBC, 2004).

A evolução do quadro clínico pode caracterizar Angina Instável (AI) e, posteriormente, Infarto Agudo do Miocárdio (NICOLAU et al. 2007).

- Cardiopatia congênita cianótica:

São doenças congênitas do coração, provavelmente resultado de malformações cardíacas, que apresentam cianose, ou seja, colocação de pele azulada, onde há mistura de sangue venoso e arterial.

Se realizada cirurgia precoce para correção, diminui o risco de mulheres na gestação, uma vez que existe risco fetal e neonatal, onde 50% causa aborto espontâneo, 30% a 50%, parto prematuro e baixo peso do bebê (BORN, 2009).

- Cardiopatia congênita acianótica, não corrigida cirurgicamente ou por intervenção percutânea:

<sup>3</sup> Clínica SAADI de Cirurgia Cardiovascular, Porto Alegre, RS. Disponível em <<http://www.clinicasaadi.com.br/sistema-cardiovascular/doencas/valvulopatias/>> acessado em 03 de junho de 2011.



São doenças congênitas do coração caracterizadas por problemas na comunicação interatrial (CIA) e interventricular (CIV).

No primeiro caso, a sobrecarga de uma gestação pode aumentar a hipertensão pulmonar, agravando outros problemas. Isto pode acontecer em função de embolia pulmonar ou hipotensão grave (síndromes hemorrágicas e complicações anestésicas).

Enquanto, a CIV faz com que a evolução da gestação venha acompanhada de insuficiência cardíaca, arritmia, entre outros (BORN, 2009).

- Miocardiopatias<sup>4</sup>:

A miocardiopatia é uma doença que altera e compromete a estrutura da parede muscular dos ventrículos do coração (miocárdio).

Existem três tipos de miocardiopatias: dilatada, hipertrófica e restritiva.

A miocardiopatia dilatada refere-se à dilatação dos ventrículos, impedindo-os de bombear o volume de sangue adequado acarretando em insuficiência cardíaca.

A miocardiopatia hipertrófica é caracterizada pelo espessamento das paredes ventriculares, acarretando aumento na carga de trabalho do músculo cardíaco.

E, a miocardiopatia restritiva é semelhante à hipertrófica no tangente ao efeito (aumento do trabalho do coração), mas é caracterizada pelo enrijecimento das paredes ventriculares, não espessamento.

---

<sup>4</sup> MANUAL MERK, Distúrbios do Coração e dos Vasos Sanguíneos, Miocardiopatia, Cap. 18. Disponível em <[http://mmspf.msdonline.com.br/pacientes/manual\\_merk/secao\\_03/cap\\_018.html](http://mmspf.msdonline.com.br/pacientes/manual_merk/secao_03/cap_018.html)>. Acessado em 04 de junho de 2011.

### 3 TRABALHOS RELACIONADOS

Esse capítulo apresenta uma revisão da literatura na área de sistemas de apoio ao acompanhamento de pacientes crônicos, mostrando os trabalhos mais importantes que estão relacionados com o objetivo de pesquisa da presente proposta.

Além de critérios gerais de importância e atualidade, os trabalhos analisados a seguir foram selecionados tendo em vista dois critérios mais específicos: a aplicabilidade das técnicas de mineração de dados no contexto brasileiro e a utilização da tecnologia de agentes e Sistemas Multiagente em sistemas médicos. Trabalhos que atendam ambos critérios têm um claro impacto nas pesquisas vinculadas a presente dissertação.

#### ***3.1 Utilizando Técnicas de Data Mining para Descoberta de Padrões de Comportamento das Hepatites Virais***

O controle epidemiológico de doenças é um dos principais objetivos dos órgãos públicos de saúde. A pesquisa realizada por Trindade (TRINDADE et al, 2007) foca a busca pelo perfil epidemiológico das hepatites virais.

Considerando a grande quantidade de informações disponíveis nos bancos de dados de saúde, técnicas padrões para consulta não são aplicáveis e, por isso, nesta pesquisa foram utilizadas técnicas de KDD, nas suas respectivas etapas:

- **Estudo do Problema:** Analisando junto à Secretaria de Estado da Saúde do Paraná, um dos objetivos deste estudo foi identificar a classificação etiológica das hepatites virais com idade, imunizações, forma clínica, evolução e diagnóstico da doença.
- **Seleção dos Dados:** Foram extraídos 442 registros com 6 atributos da base de dados da SINAN (Sistema de Informações de Agravos de Notificação), dos cerca de 20000 registros com 126 atributos que constituem a base completa.
- **Pré-Processamento:** A retirada dos 120 atributos da análise foi realizada mediante validação de um profissional devidamente qualificado no estudo epidemiológico das hepatites virais.
- **Transformação:** Códigos foram substituídos pelas suas descrições e campos nulos preenchidos com a informação “não informados”.
- **Mineração de Dados:** Nesta etapa, a técnica proposta por (TRINDADE et al, 2007) foi o uso do algoritmo C4.5, resultando na árvore de decisão e nas regras de classificação, conforme exibido na figura e tabela abaixo.



Conforme exibido na Figura 13, o algoritmo identificou o atributo “Forma Clínica” como mais relevante, de forma que compôs o topo da árvore. Das formas clínicas existentes, a identificação do vírus deu-se diretamente na maioria. Na forma “Hepatite Aguda”, houve a divisão pela idade dos pacientes, onde maiores de 23 anos são prováveis do vírus B e menores de 23 anos, tiveram sua classificação relacionada ao atributo evolução. Portadores assintomáticos, o atributo Imunização foi mais relevante, identificando o paciente imunizado portador do vírus B e do vírus C o paciente sem imunização.

As regras de classificação geradas pelo algoritmo e exibidas na tabela 4, ilustram, de forma textual, a representação da árvore de decisão gerada.

Segundo os autores do artigo, a aplicação do algoritmo C4.5 no banco de dados do SINAN identificou padrões muito informativos (TRINDADE et al, 2007). Entre eles:

- Pode-se constatar que a identificação de algumas regras assemelha-se com a literatura. Um exemplo é a tendência do vírus B em pacientes com hepatite aguda maiores de 23 anos, enquanto que nos menores, quando a evolução é a cura, predomina a hepatite A.
- Na infecção assintomática, foram encontrados alguns casos com etiologia ignorada. Isto reflete provavelmente um erro de preenchimento, uma vez que este diagnóstico só pode ser feito através da imunização para Hepatite A.
- A regra 1 confirma a evidência do vírus A nas formas de hepatite aguda que evoluem para a cura.
- A regra 10 elucida a importância da vacinação e as consequências da não realização.
- As regras 6,7,8,9 e 11 que auxiliam na identificação das hepatites crônicas em maiores de 28 anos, que geralmente estão associadas à Hepatite C.

Os autores concluem que a pesquisa realizada e a aplicação das técnicas de KDD tornaram possível visualizar as variáveis significativas à identificação da classificação etiológica, permitindo comparar o resultado desta técnica com informações disponíveis na literatura.

### **3.2 Identificação dos Fatores de Risco em Doenças Cardíacas utilizando Técnicas de Mineração de Dados**

Outro estudo na área de saúde envolvendo aplicação de técnicas de mineração de dados foi desenvolvido por Cavalcante (CAVALCANTE, 2009), aplicando algoritmos de associação para identificar a prevalência dos fatores de risco cardiovasculares em pacientes com obstrução das artérias coronárias (DAC – Doença da Artéria Coronariana). A pesquisa foi realizada no Hospital Monte Sinai, em Goiânia, considerando entrevistas realizadas com pacientes no período de agosto a novembro de 2008. Foram entrevistados 85 pacientes portadores de DAC com idade superior a 20 anos.

A entrevista aplicada foi composta de questões derivadas do *Framingham Heart Study*. Este estudo, coordenado pelo *National Heart Institute*, iniciou-se em 1948 e tinha como objetivo identificar os principais fatores de risco para o desenvolvimento de doenças cardiovasculares. Para esta pesquisa, 5209 pessoas entre 30 e 60 anos de idade foram

acompanhadas por vários anos, de forma a identificar um grupo de pessoas cujo diagnóstico evoluiu para uma doença cardíaca e outro que não foi identificada a doença. O estudo de *Framingham* tornou-se, então, um padrão mundial para doenças cardiovasculares.

Para desenvolvimento da sua pesquisa, a autora utilizou as seguintes variáveis para análise dos fatores de risco (CAVALCANTE, 2009):

- Sexo; Idade (adulto jovem, adulto médio, velhice); Raça/cor; Estado Civil; Escolaridade.
- Trabalho (sim ou não); Ocupação (braçal, misto, intelectual e nenhuma; Renda Familiar Mensal (baixa, média baixa, média e alta).
- Número de Moradores por Residência; Saneamento Básico; Número de Cômodos; Área onde reside.
- Dieta quanto à gordura (rica, equilibrada e pobre); Atividade física; Tabagismo; Etilismo; Estresse;
- Hipercolesterolemia (colesterol alto ou não); Diabetes; Hipertensão; Cardiopatia anterior; Hipercolesterolemia familiar; Diabetes familiar;
- DAC familiar; Dor no peito; Número de infartos por paciente; Doenças afins (paciente teve outra doença cardiovascular); Obstrução das Coronárias;
- Peso; Altura; Índice de Massa Corpórea (IMC); Obesidade (normal, sobrepeso e obesidade); Circunferência abdominal.

As informações coletadas foram submetidas às técnicas de *data mining* através da ferramenta Weka<sup>5</sup>. Os resultados obtidos foram agrupados conforme a categoria das variáveis avaliadas. Através da análise sociodemográfica a autora identificou estimativas que convergem com os percentuais de estudos da área. Como por exemplo (CAVALCANTE, 2009):

- Probabilidade maior de presença da DAC em pacientes que residem na área urbana (inclusive isso se deve não somente à localização, mas principalmente aos hábitos e poluição característicos de quem reside nesta área);
- Probabilidade maior de presença da DAC em Adulto Médio e Velhice. Estudos comprovam que a incidência de DAC é maior em pacientes com idade superior a 50 anos;
- Probabilidade maior de presença da DAC em pessoas que trabalham. Estudos comprovam que a incidência de DAC é maior em quem trabalha de 40 a 48 horas/semana em média.

A análise do estilo de vida da população demonstrou melhor convergência aos percentuais já homologados. As tabelas a seguir exibem os resultados das análises aplicadas neste contexto.

---

<sup>5</sup> Ferramenta desenvolvida pela Universidade de Waikato para aplicação de técnicas de mineração de dados e algoritmos de aprendizagem de máquina, disponível em <<http://www.cs.waikato.ac.nz/ml/weka/>>

Tabela 5 - Análise do estilo de vida da população da incidência de DAC no Hospital Monte Sinai

<b>Dieta Quanto à Gordura</b>	<b>Ocorrência</b>	<b>Frequência%</b>
Rica	71	83,5
Equilibrada	11	13,0
Pobre	3	3,5
<b>Atividade Física Regular</b>		
Sim	19	22,3
Não	66	77,7
<b>Tabagismo</b>		
Sim	63	74,1
Não	22	25,9
<b>Etilismo</b>		
Sim	37	43,5
Não	48	56,5
<b>Estresse</b>		
Sim	66	77,7
Não	19	22,3

Fonte: Cavalcante (2009)

Concluindo as análises, as tabelas abaixo ilustram os resultados aplicados na caracterização da história clínica dos pacientes e das medidas antropométricas realizadas.

Tabela 6 - Análise do estilo de vida da população da incidência de DAC no Hospital Monte Sinai

<b>Hipercolesterolemia</b>	<b>Ocorrência</b>	<b>Frequência%</b>
Sim	54	63,6
Não	31	36,4
<b>Diabetes</b>		
Sim	22	25,9
Não	63	74,1
<b>Hipertensão</b>		
Sim	73	85,9
Não	12	14,1
<b>Cardiopatía ant. Coronariopatía</b>		
Sim	13	15,2
Não	72	84,8
<b>Hipercolesterolemia Familiar</b>		
Sim	60	70,6
Não	25	29,4
<b>Diabetes Familiar</b>		
Sim	39	45,9
Não	46	54,1
<b>Hipertensão Familiar</b>		
Sim	82	96,5
Não	3	3,5
<b>DAC Familiar</b>		
Sim	59	69,4
Não	26	30,6
<b>Dor no Peito</b>		
Sim	66	77,7
Não	19	22,3
<b>Número de Infartos\Pacientes</b>		
Pouco	76	89,4
Médio	9	10,6
Muito	0	0,0
<b>Doenças Afins</b>		
Sim	5	5,9
Não	80	94,1

Fonte: Cavalcante (2009)

Tabela 7 - Análise antropométrica da população da incidência de DAC no Hospital Monte Sinai

<b>Acúmulo de Gordura Abdominal</b>	<b>Ocorrência</b>	<b>Frequência%</b>
Sim	50	58,8
Não	35	41,2
<b>Obesidade – I.M.C.</b>		
Normal	33	38,9
Sobrepeso	44	51,7
Obesidade	8	9,4

Fonte: Cavalcante (2009)

A partir destas informações, utilizando a ferramenta Weka, foram associados 15 atributos, cujo seleção foi baseada nos fatores utilizados no Estudo de Framingham. Do resultado da aplicação da associação, foram selecionadas as que tiveram fator de confiança menor ou igual a 1. O algoritmo de associação utilizado não foi citado na pesquisa. O resultado é apresentado na figura abaixo.

1. Atividade física=Nao      Hipercolesterol=Sim      Hipertensao=Sim      Dac familia=Sim Gordura abdominal=Sim 28 ==> Dietagordura=Rica 28      Conf:(1)
2. Atividade física=Nao      Hipercolesterol=Sim      Hipertensao=Sim      Imc=Sobrepeso 26 ==> Gordura abdominal=Sim 26      Conf:(1)
3. Idade=Adulto medio      Dietagordura=Rica      Hipercolesterol=Sim      Gordura abdominal=Sim 22 ==> Hipertensao=Sim 22      Conf:(1)
4. Hipercolesterol=Sim      Diabetes=Sim 21 ==>      Hipertensao=Sim 21      Conf:(1)
5. Dietagordura=Rica      Atividade física=Nao      Hipercolesterol=Sim      Imc=Sobrepeso Gordura abdominal=Sim 26 ==>      Hipertensao=Sim 25      Conf:(0.96)
6. Sexo=M      Dietagordura=Rica      Hipercolesterol=Sim 23 ==>      Hipertensao=Sim 22      Conf:(0.96)
7. Sexo=F      Dietagordura=Rica      Alcool=Nao      Hipercolesterol=Sim      Gordura abdominal=Sim 24 ==> Hipertensao=Sim 23      Conf:(0.96)
8. Raca=Parda      Dietagordura=Rica 23 ==>      Hipertensao=Sim 22      Conf:(0.96)
9. Diabetes=Sim 22 ==>      Hipercolesterol=Sim      Hipertensao=Sim 21      Conf:(0.95)
10. Diabetes=Sim 22 ==>      Gordura abdominal=Sim 21      Conf:(0.95)
11. Alcool=Sim 37 ==>      Tabagismo=Sim 35      Conf:(0.95)
12. Hipercolesterol=Sim 54 ==>      Hipertensao=Sim 51      Conf:(0.94)
13. Hipercolesterol=Sim 54 ==>      Dietagordura=Rica 52      Conf:(0.94)
14. Diabetes=Nao      Imc=Normal 31 ==>      Gordura abdominal=Nao 29      Conf:(0.94)
15. Dietagordura=Rica      Tabagismo=Sim      Hipercolesterol=Sim 30 ==>      Hipertensao=Sim 28 Conf:(0.93)

Figura 14 – Melhores associações entre fatores de risco cardiovasculares dos pacientes portadores de DAC (CAVALCANTE, 2009)

A autora conclui que os dados resultantes do estudo convergem aos resultados que a literatura oferece sobre frequências de fatores de risco cardiovasculares, validando o modelo aplicado.

### 3.3 Modelo de Gestão do Relacionamento do Paciente Crônico (GRPC)

Considerando o foco de pesquisa no monitoramento de pacientes crônicos, existem estudos no desenvolvimento de sistemas que apóiam esse monitoramento. Entre as soluções, a aplicação de sistemas multiagentes (WERNECK et al, 2006) para acompanhamento de pacientes crônicos facilita a implementação do sistema de informação sem perder a capacidade de atuar em forma distribuída. A proposta de (WERNECK et al, 2006) é de um Modelo de Gestão do Relacionamento do Paciente Crônico (GRPC), no qual utiliza os conceitos de CRM (*Customer Relationship Management*) para acompanhamento e monitoramento dos pacientes.

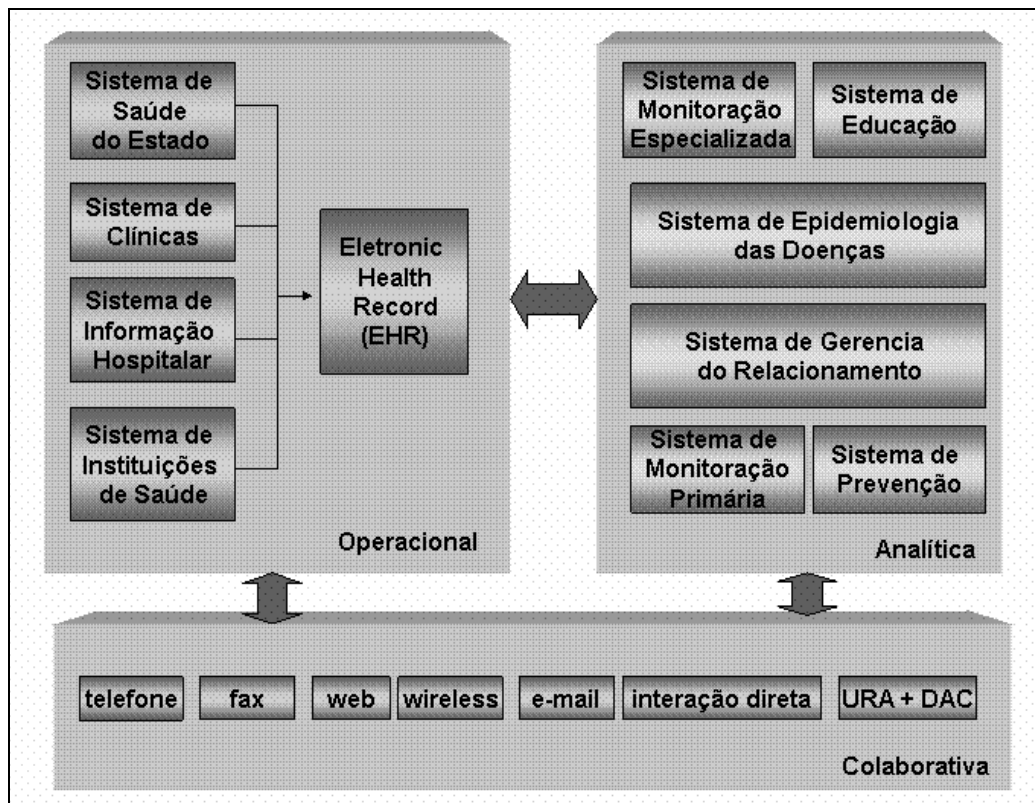


Figura 15 - Modelo GRPC utilizando os conceitos de CRM (WERNECK et al, 2006)

Os componentes apresentados na figura acima tem as seguintes características:

- Operacional: coleta os dados dos diversos níveis de interação com o paciente.
- Colaborativo: implementa os canais de comunicação direta;
- Analítico: efetua a análise dos dados nos níveis estratégicos, táticos e de inteligência do atendimento personalizado ao paciente.



Para operacionalizar o modelo GRPC, (WERNECK et al, 2006) propõe uma Central de Relacionamento com Pacientes Crônicos (CRC), que implementa o controle e monitoramento de diabéticos.

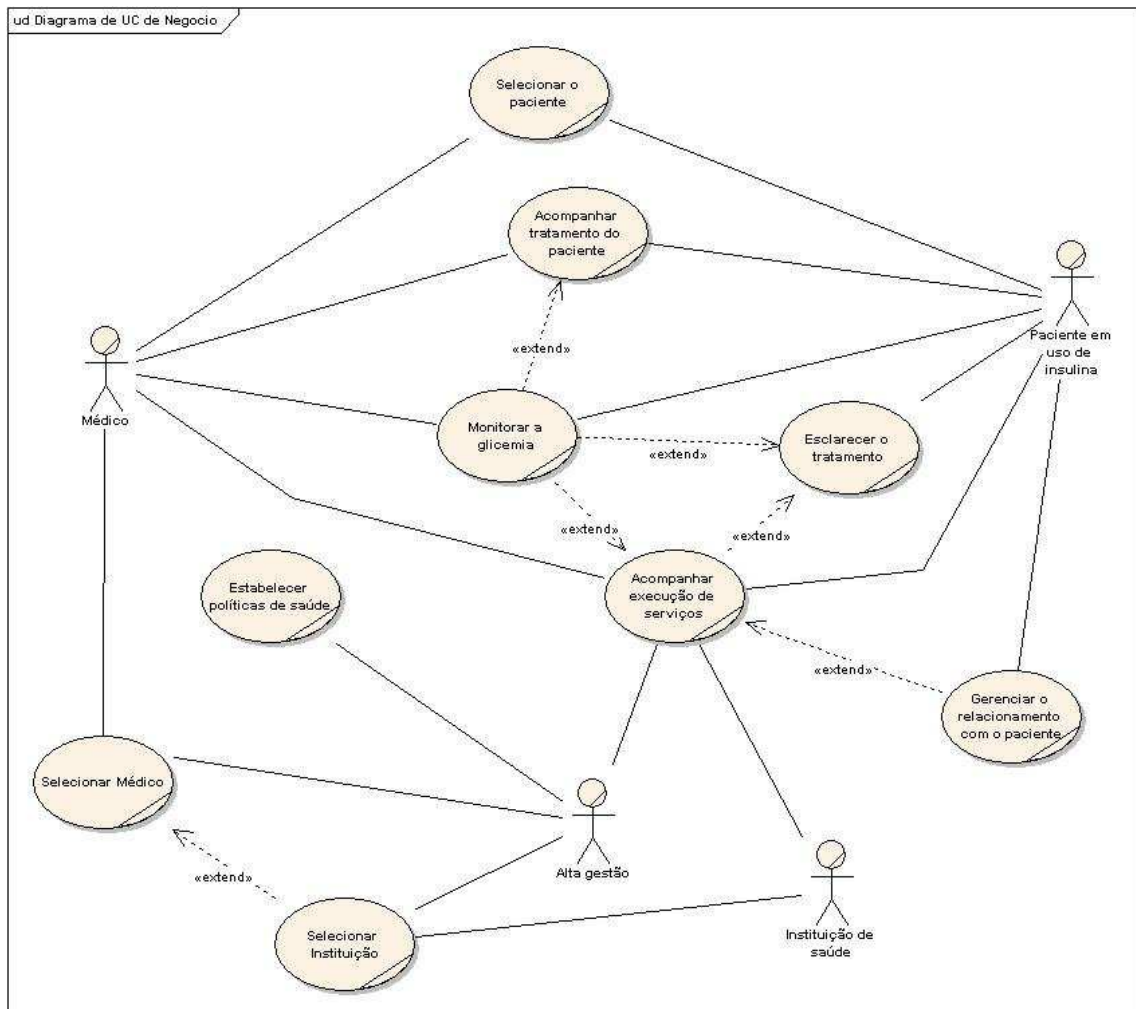


Figura 16 - Central de Monitoramento de Pacientes Crônicos (WERNECK et al, 2006)

O objetivo do projeto era viabilizar a interação informatizada da CRC com o modelo de gestão GRPC. Para isso, era necessário integrar a esse modelo uma arquitetura de agentes.

O sistema multiagente proposto em (WERNECK et al, 2006) abstrai o mecanismo de comportamento de agentes, a racionalização, para fora dos agentes e concentra em papéis distribuídos para outros agentes, denominados *dummies*. O ambiente físico do sistema é representado por objetos, configurado em tempo de execução com a utilização do conceito de agente tipo *broker* aplicado a uma agência baseada em papéis e a proposição de um *framework* de racionalização, para resolução facilitada de problemas complexos. A arquitetura implementada prevê utilização em qualquer outro domínio dentro do modelo GRPC. Mais tarde, Silva (SILVA, L.R; ITO, M., 2007) propõe, através do modelo GRPC, a integração de dados de saúde considerando o padrão EHR ou TISS. A integração proposta dá-se através da ação de agentes para buscar informações nas diferentes bases de dados e compor o componente operacional do modelo GRPC.

### 3.4 Sistema HCDMP

Outra solução utilizando a tecnologia de agentes é o Sistema HCDMP (*Hospitals Chronic Disease Management and Purchasing System*), proposto por Musavi (2008), cujo foco é utilizar a tecnologia de agentes para auxiliar no gerenciamento de doenças crônicas, fornecendo, entre outros, alguns recursos para auxílio dos pacientes:

- *Guidebooks*: orientações ao paciente sobre a doença, medicamentos, entre outros cuidados;
- *Prompts and Reminders*: avisos e alertas ao paciente;
- Ambiente para interação entre pacientes portadores da mesma doença, mediante autorização prévia do paciente.

A arquitetura do sistema proposta é dividida nos cenários *Initial*, *Admission*, *Home*, *Dr. Room*, *Office*, *Treatment*, *Pharmacy*, *Purchase*, *Stock*, *Supplier* e *Final*. Cada cenário envolvido possui uma característica e papéis específicos. Considerando um paciente que deseja consultar um médico, o primeiro cenário a ser trabalhado é o *Admission*, na qual ele vai solicitar um atendimento médico. No estado inicial, ao solicitar a consulta com um profissional médico, o sistema solicita a identificação do paciente (ID) para verificar se trata-se de um paciente já existente na base de dados. Uma vez validada (aceita) a identificação, o paciente terá sua consulta agendada, caso contrário, será solicitado o preenchimento de um formulário cujas informações serão submetidas a aceitação. Se for aceito o registro, um novo paciente é armazenado e seu agendamento de consulta é registrado. Caso contrário, o agendamento de consulta é recusado. Esta descrição do protocolo do cenário *Admission* é ilustrada na Figura 17.

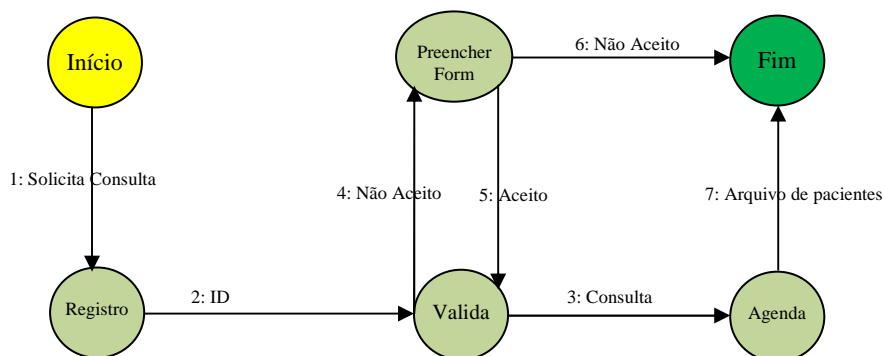


Figura 17 - Estrutura do Cenário *Admission*, onde os nodos representam o protocolo e as setas as transições (MUSAVI et al. 2008) (adaptado pelo autor)

Uma vez aceito o atendimento, o próximo cenário a ser executado é *DrRoom*, na qual o médico questiona o paciente e ele informa seu problema/doença. Se for um novo paciente, o protocolo sugere perguntas diferenciadas de uma consulta de um paciente com histórico médico. Dependendo do diagnóstico, outros cenários são requisitados: *Treatment* e/ou *Pharmacy*, cada um com protocolos distintos (MUSAVI et al. 2008).

Musavi (2008) propôs um modelo para gerenciamento de doenças crônicas genérico. A pesquisa e o trabalho desenvolvidos não tiveram foco em alguma doença específica, de modo que a proposta poderá ser estendida, se necessário, para atender um escopo definido.

### 3.5 Análise Comparativa dos Trabalhos

A seleção dos trabalhos relacionados a presente pesquisa deu-se devido ao uso de técnicas de mineração de dados e agentes. Nenhum dos trabalhos cita o uso destas técnicas concomitantemente. E, tratando-se de mineração de dados, não foram encontrados trabalhos com o uso da técnicas de Bayes ou KNN. Embora, tecnologicamente distintos, todos os trabalhos citados possuem como foco principal de pesquisa tecnologias para auxílio na gestão de pacientes portadores de doenças crônicas.

A tabela 8 apresenta um comparativo das principais características encontradas nas pesquisas citadas neste capítulo que utilizam tecnologia de mineração de dados:

Tabela 8 - Análise Comparativa dos Trabalhos Relacionados que utilizam Técnicas de Mineração de Dados

Trabalho	Objetivo de Pesquisa	Base de Dados	Técnica Utilizada
(A) Descoberta de Padrões de Comportamento das Hepatites Virais	Busca do Perfil Epidemiológico das Hepatites Virais	442 registros com 6 atributos extraídos da base de dados SINAN	Classificação por Árvores de Decisão e Regras de Classificação, utilizando algoritmo C4.5.
(B) Identificação dos Fatores de Risco em Doenças Cardíacas	Risco em Pacientes com DAC	Entrevistas com 85 pacientes portadores de DAC do Hospital Monte Sinai - Goiás	Algoritmos de associação (o artigo não cita o algoritmo utilizado).

Fonte: Desenvolvido pelo autor

A tabela 9 apresenta um resumo comparativo das principais características encontradas nas pesquisas que utilizam tecnologia de agentes:

Tabela 9 - Análise Comparativa dos Trabalhos Relacionados que utilizam Tecnologia de Agentes

Trabalho	Objetivo de Pesquisa	Tecnologia	Doença Crônica
(C) Modelo GRPC	Definição de um Modelo de Agentes baseado no conceito do CRM.	Sistema Multiagente Genérico, estendendo a padrões de saúde já estabelecidos como TISS e EHR	Diabetes
(D) Sistema HCDMP	Definição de um Modelo de Agentes Genérico para Monitoramento de Pacientes Crônicos (foco em avisos, alertas e orientações gerais ao paciente).	Arquitetura de sistema multiagente genérico	Nenhuma específica

Fonte: Desenvolvido pelo autor

O diferencial da presente pesquisa em relação aos trabalhos (A) e (B) se situa em dois aspectos principais: o foco no domínio das Doenças Cardíacas Crônicas e o uso de modelos preditivos probabilísticos sobre a passagem de uma situação crônica para uma situação de risco.

Em particular, além da diferença em termos de domínio de aplicação, o uso de modelos probabilísticos de classificação, aptos a lidar com situações de incerteza, é um diferencial importante em relação ao trabalho (A) que está focado em modelos mais tradicionais baseados em árvores de decisão ou regras de classificação.

O presente trabalho também estende e generaliza as técnicas empregadas em (B), propondo um modelo preditivo sobre risco em Doenças Cardíacas Crônicas capaz de integrar os critérios do Escore de Framingham (EF) e da Sociedade Brasileira de Cardiologia (SBC) a conhecimentos de especialistas médicos.

Quanto aos trabalhos (C) e (D) centrados na aplicação da tecnologia de agentes em sistemas da área médica, um dos principais diferenciais do presente trabalho está em sua aplicação concreta em um sistema integrado de apoio e acompanhamento de Doenças Cardíacas Crônicas.

O presente trabalho não propõe uma arquitetura genérica para este tipo de sistema, mas busca verificar se o uso da tecnologia de agentes pode efetivamente ser útil em sua concretização. Os resultados, apresentados nos capítulos a seguir, mostram que o uso de agentes permitiu construir um modelo preditivo híbrido que combinou diversos mecanismos de classificação (KNN e bayesiano) e com critérios distintos de classificação (especialista médico, EF e SBC) em um modelo preditivo híbrido. O uso de agentes também foi fundamental para implementar as características pró-ativas do sistema de acompanhamento, além de ajudar na integração deste sistema com os sistemas legados hospitalares existentes.

## 4 ARQUITETURA DO SISTEMA CARDIAC-M

Neste capítulo serão apresentados os casos de uso, o modelo de arquitetura definido para o sistema Cardiac-M, além das características do protótipo do sistema desenvolvido para apoio ao diagnóstico de riscos em pacientes crônicos cardíacos.

### 4.1 Casos de Uso

O objetivo mais geral do sistema Cardiac-M é identificar características de pacientes cardíacos de risco. Tal identificação pode ocorrer durante o registro de uma nova consulta do paciente ou após um processo de análise da base de dados atual que detecta indícios de uma situação. Essa segunda situação caracteriza um comportamento proativo do sistema, na detecção e aviso de ocorrência de situações de risco.

Dessa forma, foram considerados dois cenários distintos para utilização deste sistema: um primeiro cenário reativo, que é diretamente vinculado ao processo do atendimento de um paciente e um segundo cenário pró-ativo, relacionado a interação assíncrona do sistema com os demais sistemas e bancos de dados de prontuário médico legados.

No primeiro caso, considera-se a interação *on-line*, onde o médico, durante um atendimento, informa dados de saúde do paciente no seu prontuário. Uma vez registrada esta informação, um agente de *software* é acionado para verificar se o paciente em questão possui algum risco cardíaco. Para classificar o paciente como tal, verifica-se o resultado da aplicação de dois modelos de classificação probabilística: *Naive Bayes* e KNN. Se classificado como risco, em qualquer um dos modelos, o paciente é classificado como tal e o médico é imediatamente comunicado. O diagrama representado este caso de uso é apresentado na figura abaixo.

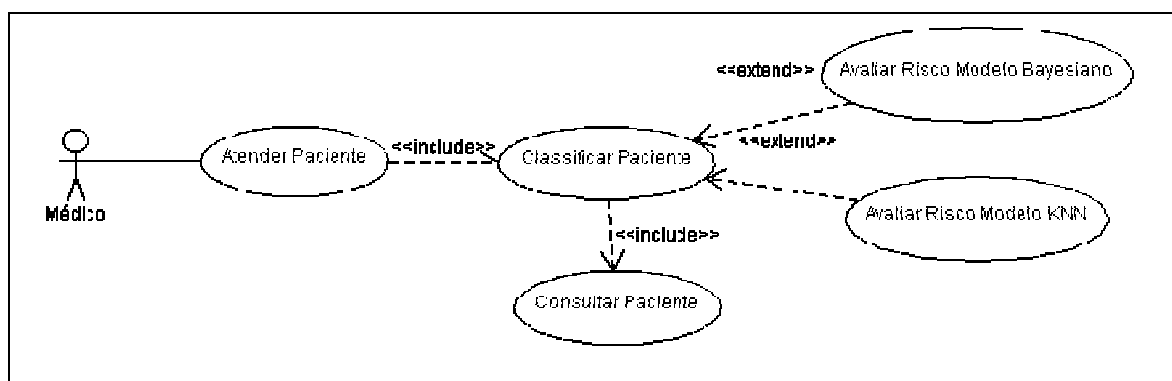


Figura 18 – Diagrama de Caso de Uso do Cardiac-M no Atendimento do Paciente

Com relação à interação assíncrona, a aplicação do sistema está relacionada à ação de um outro agente que, periodicamente, consulta na base de dados pacientes cujas informações foram alteradas. Esses pacientes são então submetidos à análise de risco e, equivalente ao

caso de uso anterior, se for classificado por qualquer um dos modelos, o médico responsável pelo último atendimento será notificado através de *e-mail*, conforme ilustrado na Figura 19.

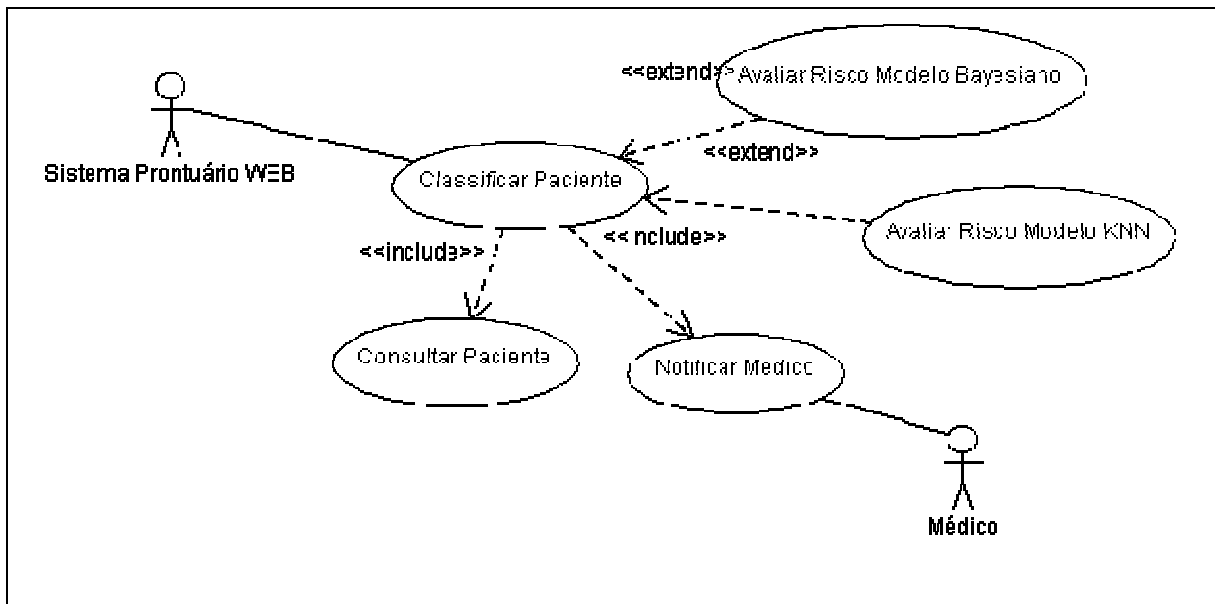


Figura 19 - Diagrama de Caso de Uso da Interação Pró-ativa com Sistemas Legados

## 4.2 Modelo de Arquitetura

Considerando a análise dos casos de uso, os principais requisitos do sistema Cardiac-M de apoio à identificação de riscos em pacientes crônicos cardíacos são os seguintes:

- Permitir interação on-line: durante a consulta médica, o médico registra informações no prontuário do paciente que está disponível na WEB. Uma vez se tratando de um paciente crônico, com os dados informados, é possível identificar um paciente de risco. Realizada esta análise, o médico deverá ser notificado on-line.
- Obter dados do prontuário médico do paciente: consultar, no sistema de prontuário médico do paciente, informações necessárias para classificação do paciente crônico.
- Identificar risco de paciente crônico: com o conhecimento dos dados de saúde de um determinado paciente crônico, o agente deverá tentar classificá-lo como paciente de risco ou não. Para isso, utilizará os modelos de classificação gerados a partir das técnicas de mineração de dados avaliadas anteriormente. Uma vez identificado o risco, em qualquer um dos modelos, o usuário que estiver efetuando registros no prontuário (neste caso, o médico), deverá ser notificado por e-mail.

Os requisitos acima são identificados como funcionalidades do sistema e, para atender, agentes de *software* são projetados para atender esses objetivos. Os agentes de software definidos no modelo estão apresentados na Figura 20, através de um diagrama de agentes da metodologia *Prometheus* (PADGHAN e WINIKOFF, 2004).

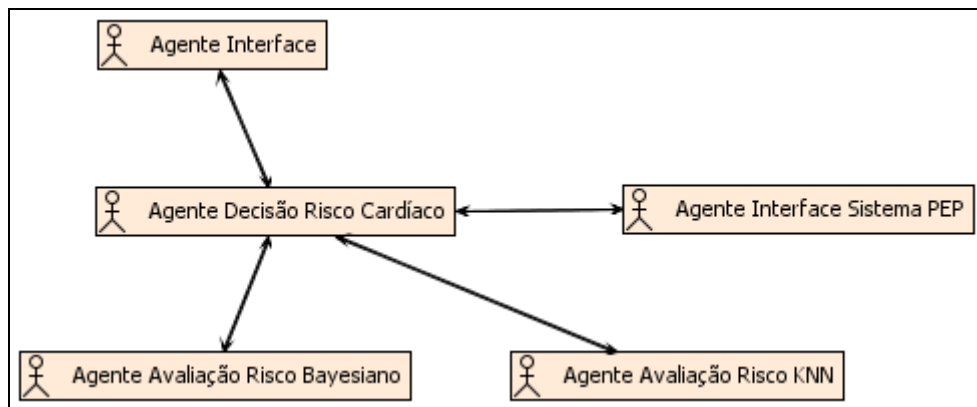


Figura 20 – Diagrama de Relacionamento entre Agentes

O Agente de Interface é responsável pela comunicação entre o médico e o sistema, recebendo informações do prontuário médico do paciente e notificando o médico caso o paciente se enquadre numa situação de risco cardíaco.

O Agente de Decisão de Risco Cardíaco é quem define se o paciente está em situação de risco cardíaco, baseado em interações deste agente com os agentes de estimativa de risco. Atualmente estão previstos dois tipos de agentes de estimativa de risco:

- Agente de Avaliação de Risco Bayesiano: verifica o risco de um paciente crônico cardíaco se tornar um paciente de risco aplicando o algoritmo de classificação *Naive Bayes*.
- Agente de Avaliação de Risco KNN: verifica o mesmo tipo de risco aplicando o algoritmo de classificação KNN.

Outros agentes de estimativa de risco podem ser eventualmente adicionados ao sistema, flexibilizando e generalizando a operação do mesmo.

O Agente de Interface com o Sistema PEP é responsável pela integração com o sistema de prontuário médico, consultando os dados de saúde necessários para a aplicação dos modelos de classificação.

A arquitetura geral deste sistema é apresentada na figura 21 (o diagrama segue a notação definida pela metodologia *Prometheus*).

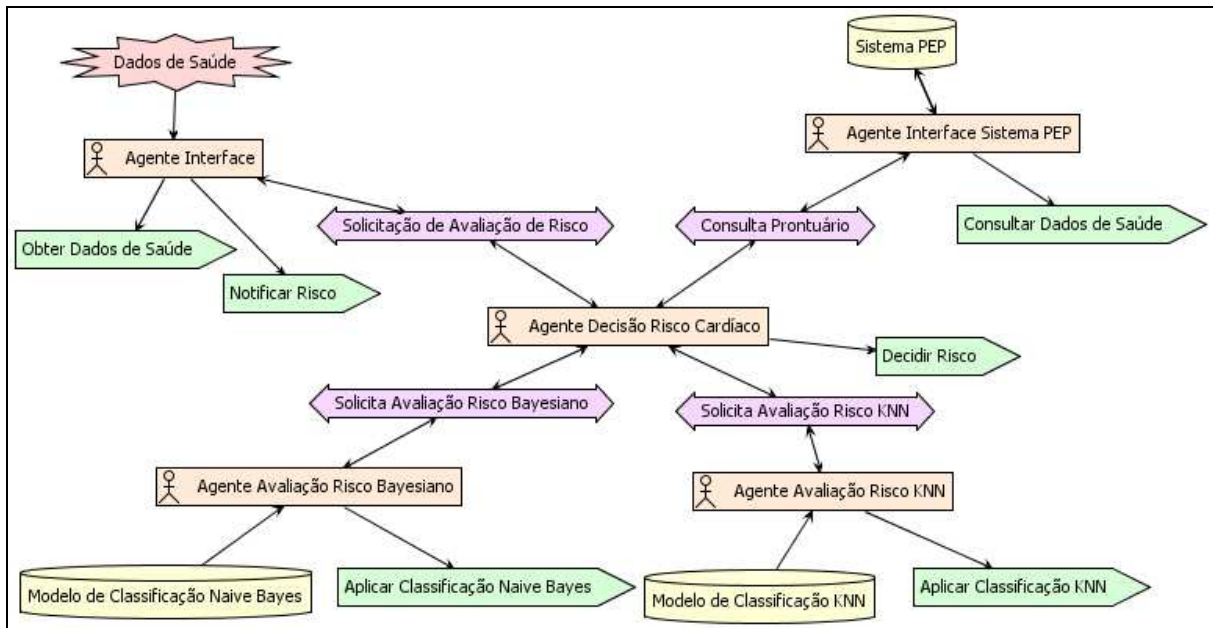


Figura 21 - Arquitetura Geral do Sistema CARDIAC-M

Na arquitetura apresentada, são representados os agentes de software, suas principais ações e percepções, informações compartilhadas e protocolos de interação.

O Agente de Interface, responsável pela interação entre o sistema e o médico, identifica o registro de informações de saúde (percepção), registra as informações obtidas e notifica o médico quando necessário (ações). O Agente de Interface interage com o Agente de Decisão de Risco Cardíaco que, por sua vez, interage com o Agente de Interface Sistema PEP e os agentes de avaliação de risco, nesta ordem. Através do protocolo *Consulta Prontuário*, o Agente de Interface Sistema PEP consulta os dados de saúde e resulta as informações a serem submetidas para avaliação. Com estas informações o Agente de Decisão de Risco Cardíaco interage com os agentes de Avaliação de Risco Bayesiano e Avaliação de Risco KNN. Estes agentes, por sua vez, são responsáveis por aplicar os modelos de classificação *Naive Bayes* e KNN, respectivamente. O retorno esperado da interação através dos protocolos *Solicita Avaliação Risco Bayesiano* e *Solicita Avaliação Risco KNN* é a classificação dos modelos, neste caso, os valores *Sim* ou *Não*. A partir do retorno destes resultados, o Agente de Decisão de Risco Cardíaco decide o risco e envia esta informação ao Agente de Interface como retorno da interação através do protocolo *Solicitação de Avaliação de Risco*.

O diagrama de interações que representa a interação entre os agentes citados nos diagramas anteriores é apresentado na Figura 22.



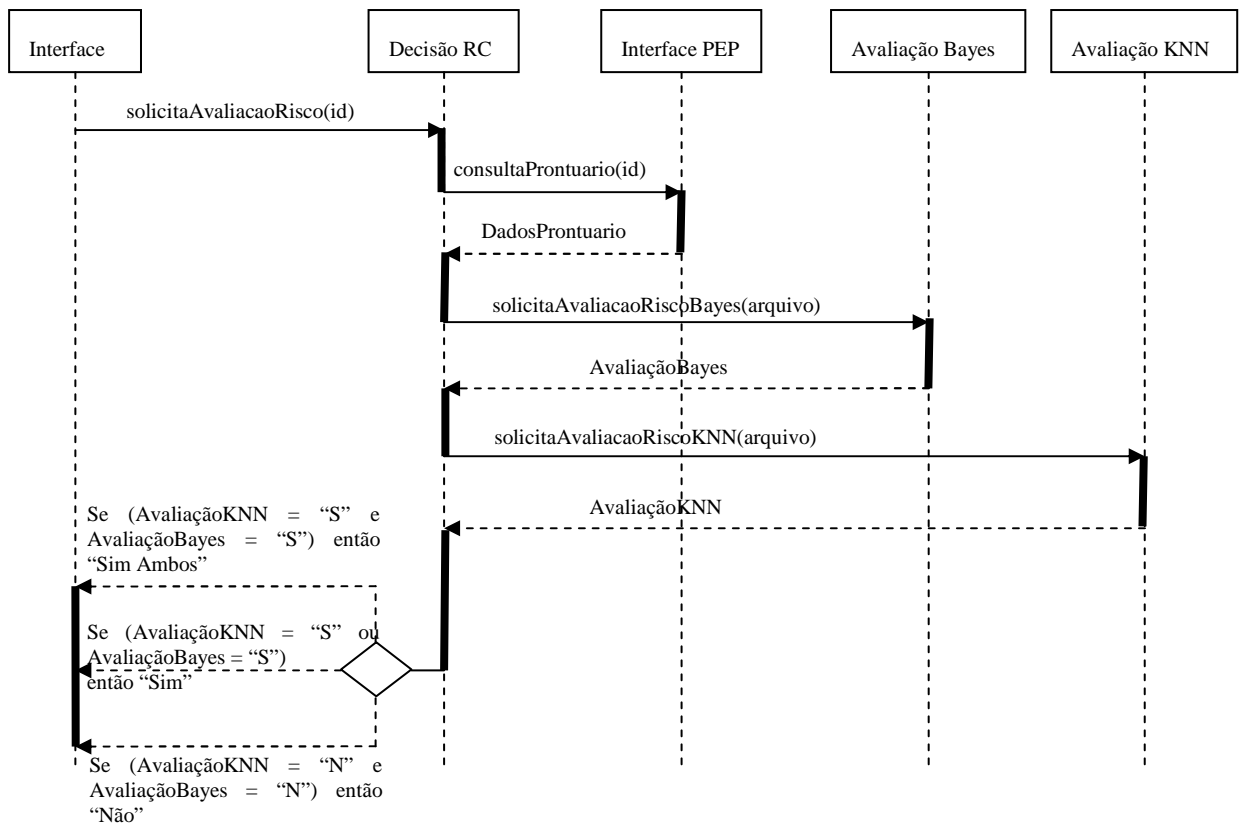
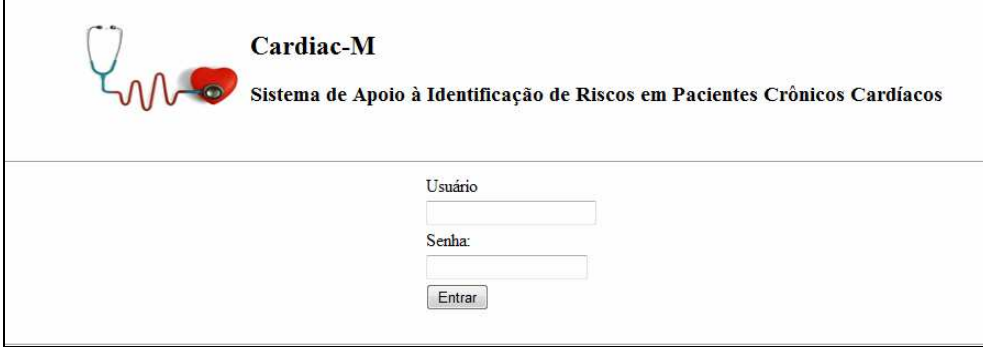


Figura 22 - Diagrama de Interação CARDIAC-M

### 4.3 Protótipo do Cardiac-M

Para avaliação do sistema proposto neste trabalho foi desenvolvido um protótipo em Java, cuja interface WEB simula o sistema de prontuário médico atualmente utilizado para teste do sistema. Por conta do período exíguo de tempo para desenvolvimento e teste dos modelos de estimativa de risco e do sistema de avaliação de risco, tanto a interface do sistema quanto as classes correspondentes aos agentes foram desenvolvidos em JSP (*Java Servlets Pages*), que é uma versão da linguagem JAVA projetada para desenvolvimento de aplicações WEB. As figuras 23 e 24 ilustram a interface desenvolvida para autenticação de usuário no sistema e registro de atendimento.

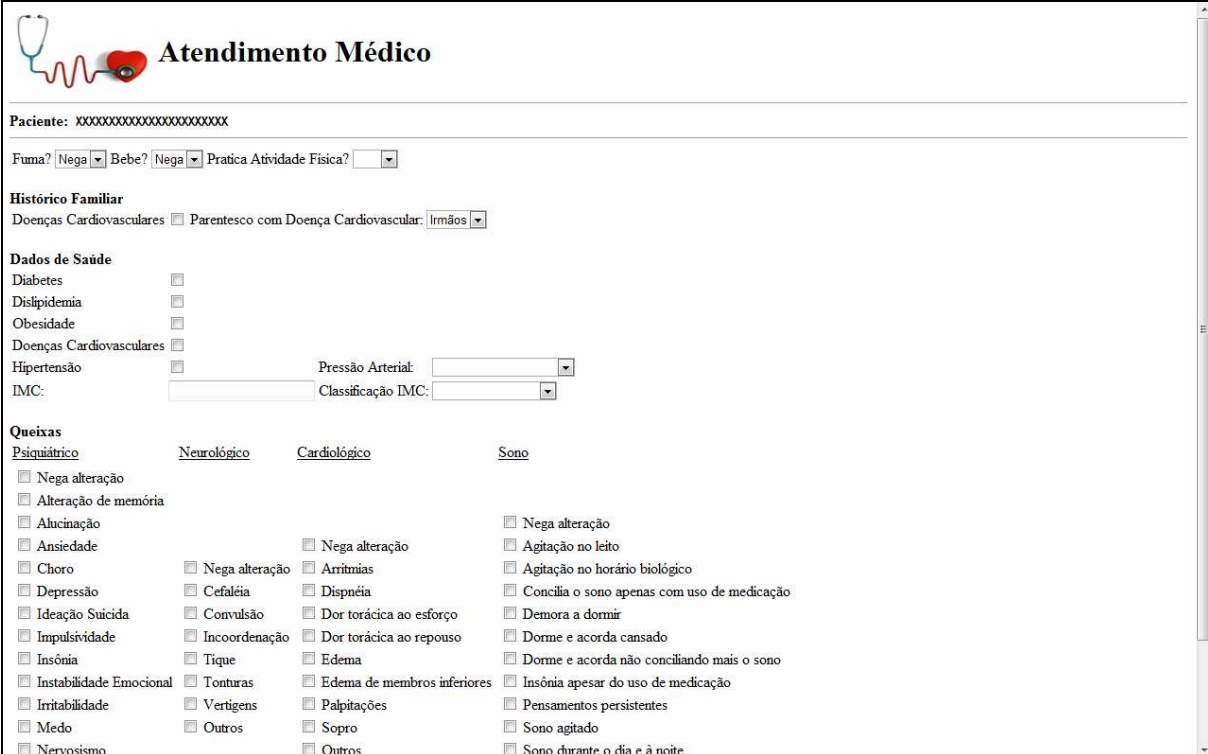


**Cardiac-M**  
Sistema de Apoio à Identificação de Riscos em Pacientes Crônicos Cardíacos

Usuário

Senha:

Figura 23 - Tela de Login da Aplicação Cardiac-M



**Atendimento Médico**

Paciente: XXXXXXXXXXXXXXXXXXXXXXX

Fuma?  | Bebe?  | Prática Atividade Física?

**Histórico Familiar**  
Doenças Cardiovasculares  Parentesco com Doença Cardiovascular:

**Dados de Saúde**  
 Diabetes   
 Dislipidemia   
 Obesidade   
 Doenças Cardiovasculares   
 Hipertensão  Pressão Arterial:   
 IMC:  Classificação IMC:

**Queixas**

Psiquiátrico	Neurológico	Cardiológico	Sono
<input type="checkbox"/> Nega alteração			<input type="checkbox"/> Nega alteração
<input type="checkbox"/> Alteração de memória			<input type="checkbox"/> Agitação no leito
<input type="checkbox"/> Alucinação			<input type="checkbox"/> Agitação no horário biológico
<input type="checkbox"/> Ansiedade		<input type="checkbox"/> Nega alteração	<input type="checkbox"/> Concilia o sono apenas com uso de medicação
<input type="checkbox"/> Choro	<input type="checkbox"/> Nega alteração	<input type="checkbox"/> Arritmias	<input type="checkbox"/> Demora a dormir
<input type="checkbox"/> Depressão	<input type="checkbox"/> Cefaléia	<input type="checkbox"/> Dispneia	<input type="checkbox"/> Dorme e acorda cansado
<input type="checkbox"/> Ideação Suicida	<input type="checkbox"/> Convulsão	<input type="checkbox"/> Dor torácica ao esforço	<input type="checkbox"/> Dorme e acorda não conciliando mais o sono
<input type="checkbox"/> Impulsividade	<input type="checkbox"/> Incoordenação	<input type="checkbox"/> Dor torácica ao repouso	<input type="checkbox"/> Insônia apesar do uso de medicação
<input type="checkbox"/> Insônia	<input type="checkbox"/> Tique	<input type="checkbox"/> Edema	<input type="checkbox"/> Pensamentos persistentes
<input type="checkbox"/> Instabilidade Emocional	<input type="checkbox"/> Tonturas	<input type="checkbox"/> Edema de membros inferiores	<input type="checkbox"/> Sono agitado
<input type="checkbox"/> Irritabilidade	<input type="checkbox"/> Vertigens	<input type="checkbox"/> Palpitações	<input type="checkbox"/> Sono durante o dia e à noite
<input type="checkbox"/> Medo	<input type="checkbox"/> Outros	<input type="checkbox"/> Sopros	
<input type="checkbox"/> Nervosismo		<input type="checkbox"/> Outros	

Figura 24 - Tela de Registro de Atendimento da Aplicação Cardiac-M

O protótipo desenvolvido permite que o médico, uma vez conectado, selecione o paciente a ser atendido e registre as informações do prontuário. As funcionalidades previstas nas ações de agentes de software definidos na arquitetura do sistema foram implementadas em classes JAVA. As classes definidas na aplicação são apresentadas a seguir.

#### Classe *conexaoBD*:

Esta classe contém os métodos responsáveis para conexão e manipulação da base de dados do sistema de prontuário médico de onde foram extraídas as informações para o treinamento dos modelos de classificação empregados neste trabalho. Entre os métodos criados, o principal método é *consultaProntuario*, que é apresentado de forma resumida na figura 25.

```

public static ResultSet consultaProntuario (String codPessoa) throws SQLException {
    conecta();
    Statement stm = conexao.createStatement();
    String sql =
        "select top 1 convert(char,e.dtInicio,103) as dtAtendimento "+
        "    ,datediff(YEAR,pp.dtNascimento,e.dtInicio) as idade "+
        "    ,pp.fgSexo as sexo "+
        "    ,isnull((select top 1 ree.nmAlternativa "+
        "        from    dsResposta rr "+
        "        inner join dsPergunta pgg on pgg.cdPergunta = rr.cdPergunta "+
        "        inner join dsRespostaEscolha ree on ree.cdResposta = rr.cdResposta "+
        "        where   pgg.nmPergunta = 'Tabagismo' "+
        "        and    rr.cdEntrevista = e.cdEntrevista),'Sem Resposta') as Tabagismo "+
        "        .
        .
        .
        ,isnull((select COUNT(distinct re.nmAlternativa) "+
        "        from    dsResposta rr "+
        "        inner join dsPergunta pgg on pgg.cdPergunta = rr.cdPergunta "+
        "        inner join dsRespostaEscolha re on re.cdResposta = rr.cdResposta "+
        "        where   re.nmAlternativa != 'Nega alteração' "+
        "        and    pgg.nmPergunta = 'Sono' "+
        "        and    rr.cdEntrevista = e.cdEntrevista),0) as Sono "+
    "from    dsEntrevista e "+
    "inner join adPessoa pp on pp.cdPessoa = e.cdPessoa "+
    "where e.fgFinalizada = 1 "+
    "and    convert(char,e.dtInicio,103) = convert(char,(select MAX(e1.dtInicio) "+
    "        from    dsEntrevista e1 "+
    "        inner join adPessoa ppl on ppl.cdPessoa = e1.cdPessoa "+
    "        where   ppl.nrDocumento = pp.nrDocumento),103) "+
    "and e.cdPessoa = " + codPessoa;

    ResultSet resultado = stm.executeQuery(sql);
    return resultado;
}

```

Figura 25 – Resumo do Método *consultaProntuario* da Aplicação Cardiac-M

O método *consultaProntuario* é responsável por consultar as informações do prontuário médico de um paciente necessárias para submissão à avaliação de risco cardíaco. O registro de saúde de um paciente (prontuário médico) possui um volume muito grande de informações e nem todas são utilizadas para análise de risco. A chamada do método *consultaProntuario* dá-se no momento em que foi realizada uma alteração no prontuário do paciente e é necessário avaliar se o paciente está em risco. A partir do código do paciente, o método consulta na base de dados o registro de saúde mais recente do paciente, executando um pré-processamento destes dados, ou seja, seleciona atributos relevantes para análise de risco, conforme exibido na tabela 10, e substitui valores nulos. O retorno do método são as informações de saúde do paciente que, posteriormente, serão submetidas à análise de risco.

#### Classe *classificador*:

Nesta classe são implementados os métodos de avaliação de risco. Existem dois métodos desenvolvidos para efetuar essa avaliação: *classificar* e *solicitaAvaliacaoRisco*. O primeiro método é responsável pela aplicação do modelo para avaliação de risco de um determinado paciente, estando integrado com as bibliotecas do *RapidMiner*. Quanto ao método *solicitaAvaliacaoRisco*, este é responsável pela decisão se o paciente possui risco cardíaco ou não, baseado nos retornos das chamadas do método *classificar* com o modelo bayesiano e com o modelo KNN. Este método também prepara as informações para serem classificadas, invocando o método de consulta aos dados e gerando o arquivo CSV com estas informações no formato de leitura dos dados requerido pelo método *classificar*. Os códigos

fonte dos métodos *classificar* e *solicitaAvaliacaoRisco* estão ilustrados nas figuras 26 e 28 respectivamente.

```

private static String classificar (String nomeArquivoNormalizado, String
arquivoModClassificador) throws
    OperatorCreationException, OperatorException {
    // Seta as propriedades para apontar para o diretório de plugins do rapidminer
    String pluginDirString = new File("/lib/plugins").getAbsolutePath();
    System.setProperty(RapidMiner.PROPERTY_RAPIDMINER_INIT_PLUGINS_LOCATION,
pluginDirString);
    // Inicializar o RapidMiner
    RapidMiner.init();
    // Criar o processo
    com.rapidminer.Process processo = new com.rapidminer.Process();
    // OPERADOR 1: Aplicador do modelo no classificador (Apply Model)
    Operator aplicadorModelo = OperatorService.createOperator(ModelApplier.class);
    // Seta as portas do operador
    InputPort portaEntradaAplicadorModelo =
aplicadorModelo.getInputPorts().getPortByName("model");
    InputPort portaEntrada2AplicadorModelo =
aplicadorModelo.getInputPorts().getPortByName("unlabelled data");
    OutputPort portaSaidaAplicadorModelo =
aplicadorModelo.getOutputPorts().getPortByName("labelled data");
    // OPERADOR 2: Carrega arquivo CSV para classificar (Read CSV)
    Operator leitorCSV = OperatorService.createOperator(CSVDataReader.class);
    leitorCSV.setParameter("file_name", nomeArquivoNormalizado);
    leitorCSV.setParameter("encoding", "SYSTEM");
    leitorCSV.setParameter("trim_lines", "false");
    leitorCSV.setParameter("skip_comments", "true");
    leitorCSV.setParameter("comment_characters", "#");
    leitorCSV.setParameter("use_first_row_as_attribute_names", "true");
    .
    .
    leitorCSV.setParameter("attribute_names_already_defined", "false");
    // A lista abaixo mantém os metadados do conteúdo que está contido no arquivo .CSV
    List<String[]> listaMetaDados = new LinkedList<String[]>();
    listaMetaDados.add(new String[] { "0", "idade.true.3.regular" });
    listaMetaDados.add(new String[] { "1", "sexo.true.4.regular" });
    .
    .
    leitorCSV.setListParameter("data_set_meta_data_information", listaMetaDados);
    // Setar a porta de saída do operador
    OutputPort portaSaidaLeitorCSV = leitorCSV.getOutputPorts().getPortByName("output");
    // OPERADOR 3: Carregar o modelo treinado no RapidMiner
    Operator carregadorModelo = OperatorService.createOperator(ModelLoader.class);
    carregadorModelo.setParameter("model_file", arquivoModClassificador);
    // Setar a porta de saída do operador
    OutputPort portaSaidaCarregadorModelo =
carregadorModelo.getOutputPorts().getPortByName("output");
    // Efetuar as conexões entre os operadores
    portaSaidaLeitorCSV.connectTo(portaEntrada2AplicadorModelo);
    portaSaidaCarregadorModelo.connectTo(portaEntradaAplicadorModelo);
    // Adicionar os operadores ao processo
    processo.getRootOperator().getSubprocess(0).addOperator(leitorCSV);
    processo.getRootOperator().getSubprocess(0).addOperator(carregadorModelo);
    processo.getRootOperator().getSubprocess(0).addOperator(aplicadorModelo);
    InputPort result1InputPort =
processo.getRootOperator().getSubprocess(0).getInnerSinks().getPortByName("result 1");
    portaSaidaAplicadorModelo.connectTo(result1InputPort);
    IOContainer container = processo.run();
    ExampleSet resultSetFinal = container.get(ExampleSet.class);
    // Buscar os resultados
    if (resultSetFinal.size()>0) {
        Example e = resultSetFinal.getExample(0);
        String r = e.getAttributes().getPredictedLabel().getMapping().mapIndex((int)
e.getPredictedLabel());
        return (r);
    } else return (null);
}

```

Figura 26 - Método *classificar* da Aplicação Cardiac-M

No método *classificar* é implementado o processo de classificação do *RapidMiner*. A figura 27 mostra um processo de classificação modelado no *RapidMiner*, onde são inseridos três operadores: *Read Model*, *Read CSV* e *Apply Model*.

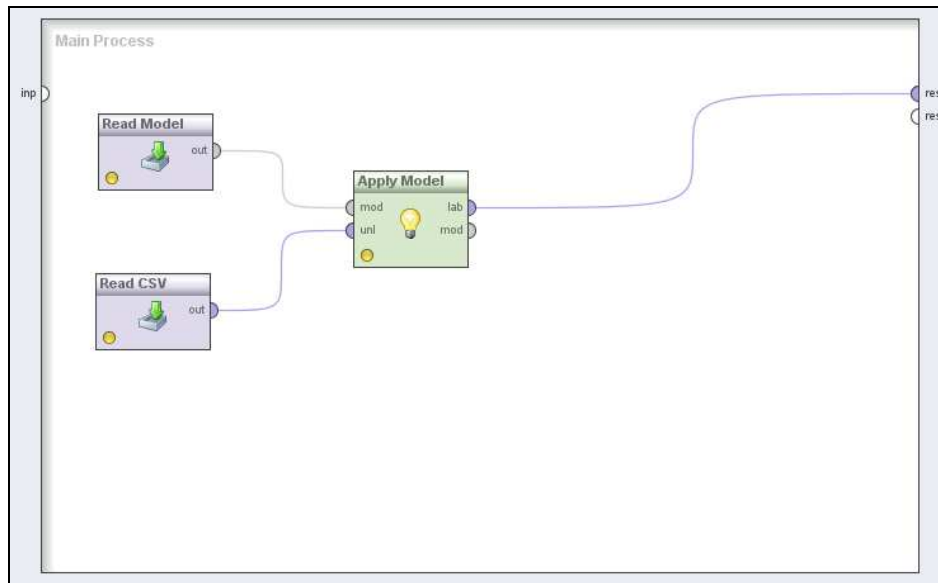


Figura 27 - Processo de Classificação Modelado no *RapidMiner*

O operador *Read Model* contém informação do modelo de classificação a ser utilizado. No código do método *classificar*, corresponde às linhas de código apresentadas abaixo:

```
// OPERADOR 3: Carregar o modelo treinado no RapidMiner
Operator carregadorModelo = OperatorService.createOperator(ModelLoader.class);
carregadorModelo.setParameter("model_file", arquivoModClassificador);
OutputPort portaSaidaCarregadorModelo =
carregadorModelo.getOutputPorts().getPortByName("output");
```

O operador *Read CSV* contém informação do arquivo CSV a ser classificado. No código do método *classificar*, este operador corresponde às linhas de código apresentadas abaixo:

```
// OPERADOR 2: Carrega arquivo CSV para classificar (Read CSV)
Operator leitorCSV = OperatorService.createOperator(CSVDataReader.class);
leitorCSV.setParameter("file_name", nomeArquivoNormalizado);
leitorCSV.setParameter("encoding", "SYSTEM");
leitorCSV.setParameter("trim_lines", "false");
leitorCSV.setParameter("skip_comments", "true");
leitorCSV.setParameter("comment_characters", "#");
leitorCSV.setParameter("use_first_row_as_attribute_names", "true");
.
.
leitorCSV.setParameter("attribute_names_already_defined", "false");
List<String[]> listaMetaDados = new LinkedList<String[]>();
listaMetaDados.add(new String[] { "0", "idade.true.3.regular" });
listaMetaDados.add(new String[] { "1", "sexo.true.4.regular" });
.
.
leitorCSV.setListParameter("data_set_meta_data_information", listaMetaDados);
OutputPort portaSaidaLeitorCSV = leitorCSV.getOutputPorts().getPortByName("output");
```

E, por fim, o operador *Apply Model* aplica o modelo de classificação ao arquivo, resultando na avaliação de risco que, neste caso, será “Sim” ou “Não”. O trecho do código fonte do método *classificar* que corresponde à definição deste operador é exibido abaixo:

```
// OPERADOR 1: Aplicador do modelo no classificador (Apply Model)
Operator aplicadorModelo = OperatorService.createOperator(ModelApplier.class);
InputPort portaEntrada1AplicadorModelo =
aplicadorModelo.getInputPorts().getPortByName("model");
InputPort portaEntrada2AplicadorModelo =
aplicadorModelo.getInputPorts().getPortByName("unlabelled data");
OutputPort portaSaida1AplicadorModelo =
aplicadorModelo.getOutputPorts().getPortByName("labelled data");
.
.
```

```
public static int solicitaAvaliacaoRisco(String codPessoa, String dir) throws SQLException,
IOException, OperatorCreationException, OperatorException {
    // Lê registro do BD com informações para classificação
    conexaoBD con = new conexaoBD();
    ResultSet dados = con.consultaProntuario(codPessoa);
    // Gera o arquivo CSV com os dados para
    // enviar ao classificador
    String arquivoCSV = dir + "/cronicos.csv";
    File file = new File(arquivoCSV);
    Writer writer = new BufferedWriter(new FileWriter(file));
    writer.write("idade;sexo;Tabagismo;" +
        "Etilismo;AtividadeFisica;" +
        "DoencasCardiovascularesHF;" +
        .
        "Cardiologico;Sono");
    while (dados.next()) {
        writer.write("\n");
        writer.write(dados.getString("idade")+";"+
            dados.getString("sexo")+";"+
            dados.getString("tabagismo")+";"+
            .
            dados.getString("cardiologico")+";"+
            dados.getString("sono"));
    }
    writer.flush();
    writer.close();
    // Chama o classificador
    String arquivoModBayes = dir + "/cronicos_naiveb.mod";
    String arquivoModKnn = dir + "/cronicos_knn.mod";
    String classBayes = classificar(arquivoCSV,arquivoModBayes);
    String classKNN = classificar(arquivoCSV, arquivoModKnn);
    if (classBayes==null) {
        classBayes = "";
    }
    if (classKNN==null) {
        classKNN = "";
    }
    if (classBayes.equals("Sim")) {
        if (classKNN.equals("Sim")) {
            return (2);
        } else return (1);
    } else if (classKNN.equals("Sim")) {
        return (1); }
    else return (0);
}
}
```

Figura 28 - Método *solicitaAvaliacaoRisco* da Aplicação Cardiac-M

O método *solicitaAvaliacaoRisco* simula o comportamento do agente *Decisão de Risco Cardíaco* definido no projeto. A principal função deste método é submeter o paciente à avaliação de risco utilizando os modelos *Naive Bayes* e *KNN*. A partir destes resultados, definir se o paciente está em risco, retornando zero (0) se o paciente não possui risco, um (1) se o paciente foi avaliado com risco por apenas um dos classificadores e dois (2) se o paciente foi avaliado com risco por ambos classificadores. Antes de solicitar a avaliação dos modelos, é necessário preparar os dados e o código fonte inicial do método prevê isso, invocando o

método *consultaProntuario* para retornar as informações de saúde do paciente e gerando um arquivo CSV com essas informações.

Finalmente, para utilizar as funcionalidades do protótipo, o primeiro passo é se conectar no sistema. A conexão só é permitida para usuários médicos que estejam cadastrados no sistema e possuam informação de código de CRM válido.

Uma vez conectado, o próximo passo é a seleção do paciente, cuja busca é efetuada pelo nome do paciente, podendo o médico informar todo ou parte dele. A figura 29 ilustra esta interface.



Figura 29 - Tela para Seleção de Paciente da Protótipo da Aplicação Cardiac-M

Após selecionado o paciente, o médico seleciona a opção *Atender* e é exibida a tela de registro de atendimento do paciente, ilustrada no início da seção na figura 24. Neste momento, também são exibidos os dados de saúde do paciente, caso já tenham sido informados no sistema. Após a conclusão do atendimento, o médico faz as alterações necessárias nos dados de saúde do paciente e solicita a gravação do registro. Neste momento, o sistema solicita avaliação de risco, conforme explicado anteriormente.

Uma vez classificado como paciente com risco cardíaco, a sistema alerta o médico conforme exibido na figura abaixo. A mensagem do alerta pode variar dependendo do resultado da análise de risco: identificado com risco por ambos modelos de análise, por apenas um ou por nenhum. O alerta previsto para envio de mail não foi implementado neste protótipo.

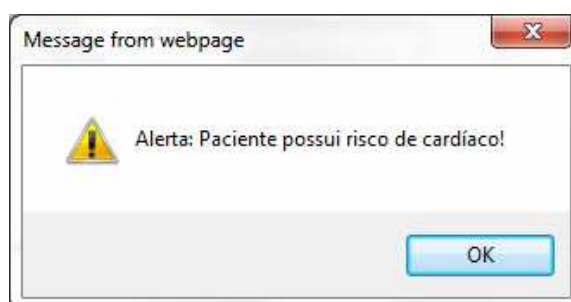


Figura 30 - Alerta exibido ao médico

## 5 MODELOS DE ESTIMATIVA DE RISCO

Este capítulo apresenta as propriedades dos modelos probabilísticos de classificação que foram construídos para estimar o risco de agravamento de doença cardíaca crônica de um paciente. Para tanto, através da utilização de técnicas e ferramentas de *Data Mining*, foram construídos dois modelos de classificação baseados, respectivamente, nos algoritmos *Naive Bayes* e *KNN*.

Neste capítulo são apresentadas as características das bases de dados, os critérios de análise destes dados e as ferramentas usadas para construir os modelos. Também são mostradas as principais medidas de desempenho dos modelos, bem como as características que justificam o uso de ambos modelos de classificação para estimar a possibilidade de risco.

### 5.1 Bases de Dados

As informações de saúde utilizadas para construção dos modelos de classificação são oriundas do sistema de prontuário eletrônico fornecido por uma operadora de saúde localizada na serra gaúcha. O prontuário médico do paciente contém as principais informações relacionadas à saúde do paciente, seus hábitos e estilo de vida, histórico de doenças do paciente e de seus familiares, exame físico, entre outros.

Inicialmente foram efetuadas análises das informações disponíveis de setembro de 2009 até 31 de agosto de 2010, de onde foram extraídos 24255 registros, onde 681 registros são dados de pacientes crônicos cardíacos. Considerando a característica da aplicação e dos dados disponíveis, para análise do modelo foram extraídos alguns atributos relevantes para o estudo a ser realizado. Abaixo é exibida a tabela com os atributos utilizados para treino do algoritmo de classificação:

Tabela 10 – Tabela com os atributos extraídos do Prontuário Médico para treino do algoritmo de classificação.

Nome do Atributo	Domínio	Descrição
Idade	0 a 110	Indica a idade do paciente no momento do atendimento.
Sexo	F ou M	Indica o sexo do paciente.
Tabagismo	Nega, Sim, Ex-fumante, Sem Resposta	Indica se o paciente possui hábito de tabagismo.
Etilismo	Nega, Sim, Etilismo no passado, Sem Resposta	Indica se o paciente possui hábito de etilismo.
Atividade Física	Não, Sim, Sem Resposta	Indica se o paciente pratica alguma atividade física.
Doenças Cardiovasculares (Histórico Familiar)	Não, Sim, Sem Resposta	Indica se algum membro da família (pais ou irmãos) possui alguma doença cardiovascular (IAM, HAS, etc.), exceto Hipotensão.



ParentescoDoençasCardiovasculares (Histórico Familiar)	Pai, Mãe, Irmãos, Sem Resposta	Indica qual membro da família (pai, mãe ou irmãos) tem alguma doença cardiovascular.
Hipertenso	Não, Sim	Indica se o paciente é hipertenso.
Classificação Pressão Arterial	Ótima, Normal, Normal-alta, Hipertensão Estágio I, Hipertensão Estágio II, Hipertensão Estágio III, Sem Resposta	Classificação da medição da pressão arterial do paciente.
Doenças Cardiovasculares	Não, Sim, Sem Resposta	Indica se o paciente possui alguma doença cardiovascular, exceto Hipotensão.
Diabetes	Não, Sim, Sem Resposta	Indica se o paciente possui Diabetes.
Dislipidemia	Não, Sim, Sem Resposta	Indica se o paciente possui alguma Dislipidemia (colesterol, triglicérido, etc.).
Obesidade	Não, Sim, Sem Resposta	Indica se o paciente possui obesidade.
IMC	Numérico	IMC do paciente no momento do atendimento.
Classificação do IMC	Magreza, Normal, Sobrepeso, Obesidade Grau I, Obesidade Grau II, Obesidade Grau III, Sem Resposta, Inválido	Classificação do IMC do paciente (Inválido corresponde à informação incorreta de peso ou altura fornecida pelo médico).
Psiquiátrico	Numérico	Quantidade de queixas de origem psiquiátrica apontadas pelo paciente (alteração de memória, alucinações, ansiedade, etc.)
Neurológico	Numérico	Quantidade de queixas de origem neurológica apontadas pelo paciente (dor de cabeça, tonturas, etc.)
Cardiológico	Numérico	Quantidade de queixas de origem cardiológica apontadas pelo paciente (arritmias, dor torácica, palpitações, etc.)
Sono	Numérico	Quantidade de queixas com relação ao sono (agitação, insônia, uso de medicação para dormir, etc.)

---

Fonte: Desenvolvido pelo autor

## 5.2 Critérios de Análise dos Dados

Esses registros foram classificados segundo três conjuntos distintos de critérios, para estabelecer quais pacientes crônicos cardíacos estavam em situação de risco:

- Fatores de Risco Cardiovascular estabelecidos pela Sociedade Brasileira de Cardiologia (SBC) (ver Figura 31).

- Critérios utilizados para calcular o Escore de Framingham<sup>6</sup> (ver Figura 32).
- Critérios declarados por um especialista médico, além de pacientes específicos selecionados por este especialista.

Fatores de risco maiores
Tabagismo
Dislipidemias
Diabetes melito
Nefropatia
Idade acima de 60 anos
História familiar de doença cardiovascular em: <ul style="list-style-type: none"> <li>• Mulheres com menos de 65 anos</li> <li>• Homens com menos de 55 anos</li> </ul>
Outros fatores
Relação cintura/quadril aumentada
Circunferência da cintura aumentada
Microalbuminúria
Tolerância à glicose diminuída/glicemia de jejum alterada
Hiperuricemia
PCR ultra-sensível aumentada <sup>62,63</sup>

Figura 31 - Fatores de Risco Cardiovascular (SBC, 2006)

	Homens	Mulheres
<b>Idade (anos)</b>		
<34	-1	-9
35-39	0	-4
40-44	1	0
45-49	2	3
50-54	3	6
55-59	4	7
60-64	5	8
65-69	6	8
70-74	7	8
<b>Colesterol total (mg/dL)</b>		
<160	-3	-2
160-199	0	0
200-239	1	1
240-279	2	2
>280	3	3
<b>HDL colesterol (mg/dL)</b>		
<35	2	5
35-44	1	2
45-49	0	1
50-59	0	0
>60	-2	-3
<b>Pressão arterial sistólica</b>		
< 120	0	0
120-129	1	1
130-139	2	2
140-159	3	3
>160		
<b>Diabetes</b>		
Não	0	0
Sim	2	4
<b>Tabagismo</b>		
Não	0	0
Sim	2	2

Figura 32 – Critérios para Cálculo do Escore de Framingham (LOTUFO, 2008)

<sup>6</sup> O Escore de Framingham é um estudo que iniciou na década de 60 para identificação dos principais fatores de risco de doenças coronarianas e, atualmente, é aplicado em programas do Ministério da Saúde para prevenção destas doenças.

O especialista médico que ajudou na seleção dos pacientes em risco, possui uma ampla experiência da área de gestão de pacientes crônicos, trabalhando como médico do trabalho há 14 anos no acompanhamento de funcionários com doenças crônicas, principalmente cardíacos.

A intervenção do especialista se deu através de duas formas distintas: inicialmente o especialista apenas selecionou, dentro da base de informações a ser utilizada para treino, um conjunto de pacientes considerados em risco. Posteriormente, após análises e discussões com o especialista a respeito dos pacientes selecionados pelo mesmo, mas que não seriam selecionados pelos critérios da SBC ou do Escore de Framingham, foi estabelecida a importância que atributos indicativos de situações de estresse, (atributos Psiquiátrico, Neurológico e Sono, nos dados de treinamento) teriam na detecção do risco.

Assim, utilizando os critérios estabelecidos pelo especialista médico, pela SBC e pelo Escore de Framingham, 492 pacientes crônicos cardíacos da base de informações usada para treino do modelos foram considerados em situação de risco.

A base de informações foi modificada, então, para incorporar esse atributo de classificação. Após esse processo, foi utilizada a ferramenta RapidMiner, versão 5.0, para executar os processos de treinamento e avaliação dos modelos.

### **5.3 Ferramenta de Mineração de Dados**

A ferramenta de mineração de dados utilizada neste trabalho foi o *RapidMiner 5.0*. O *RapidMiner*<sup>7</sup> é uma ferramenta *open source* para mineração de dados. Desenvolvida em JAVA, pode ser executada a partir de qualquer plataforma e sistema operacional. Possui uma interface gráfica, conforme apresentado na figura 33, uma interface por linha de comando ou suas bibliotecas podem ser acessadas através de API JAVA.

---

<sup>7</sup> Ferramenta de mineração de dados desenvolvida pela empresa Rapid-i, disponível em <[http://rapid-i.com/component/option,com\\_frontpage/Itemid,1/lang,en/](http://rapid-i.com/component/option,com_frontpage/Itemid,1/lang,en/)>

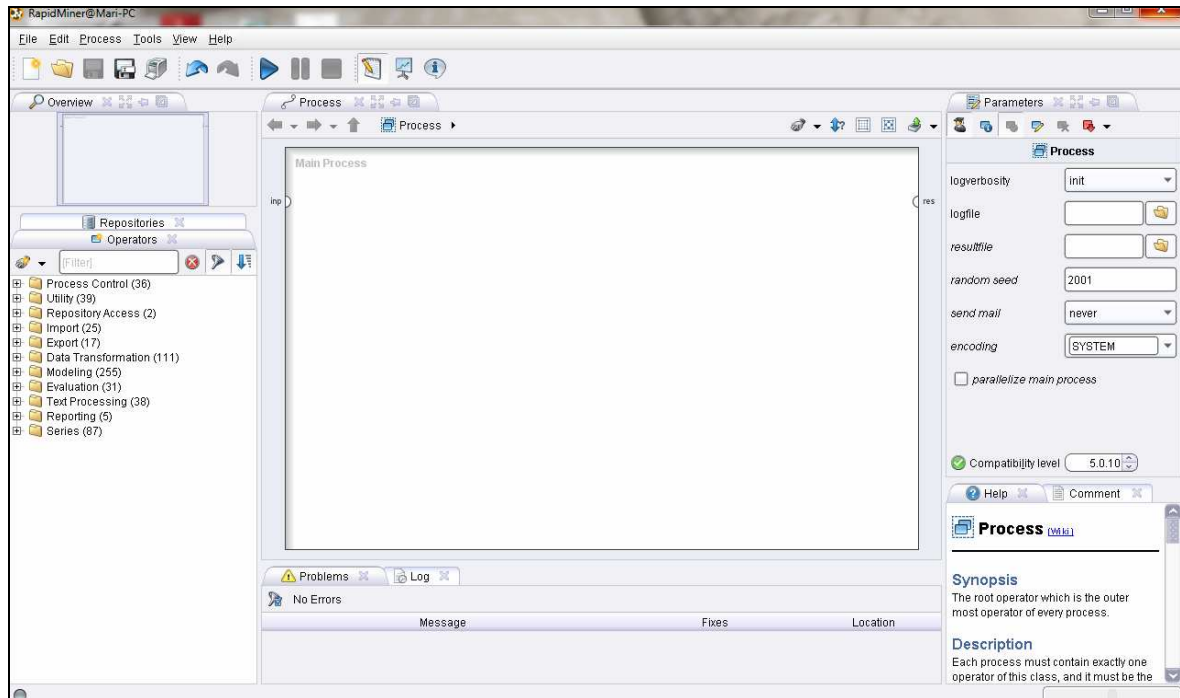


Figura 33 - Interface Gráfica da Ferramenta RapidMiner

A interface gráfica do *RapidMiner*, oferecida a partir da versão 5.0, permite que as principais funções, operações e subprocessos de um processo de mineração sejam especificados através de diagramas similares aos Diagramas de Fluxo de Dados (DFD), onde cada elemento (nó) do diagrama é um operador de mineração de dados, ou um subprocesso que ainda pode ser refinado.

Cada um dos nós do diagrama de processo do *RapidMiner* tem indicado em seu ícone os pontos (ou portas) de entrada e saída de dados necessários para o funcionamento do operador. Estas portas são conectadas através de enlaces que definem o fluxo de dados entre os operadores e subprocessos de um processo.

Além disso, a ferramenta é totalmente compatível com padrão XML, facilitando a troca de informações entre os sistemas, além de acessar diversos tipos de bases de dados, tais como: Excel, Access, Oracle, IBM DB2, Microsoft SQL, Sybase, Ingres, MySQL, Postgres, SPSS, dBase, arquivos texto, entre outros.

*RapidMiner* implementa vários algoritmos de mineração de dados de classificação, regressão, associação, clusterização, entre outros. Entre os algoritmos de classificação e regressão, método utilizado no presente trabalho, constam os modelos probabilísticos *Naive Bayes*, KNN e Redes Neurais, e os modelos baseados em regras e árvores de decisão.

O *RapidMiner* também integra bibliotecas *Weka*, software de mineração de dados *open source* com características semelhantes ao RapidMiner. *Weka* é uma ferramenta também desenvolvida em JAVA, com recursos para pré-processamento das informações, algoritmos de classificação, regressão, clusterização e regras de associação.

## 5.4 Construção dos Modelos

A estrutura de alto nível dos processos de construção dos modelos de classificação NaiveBayes e K-NN, utilizados no presente trabalho, é idêntica, sendo apresentada na figura 34.

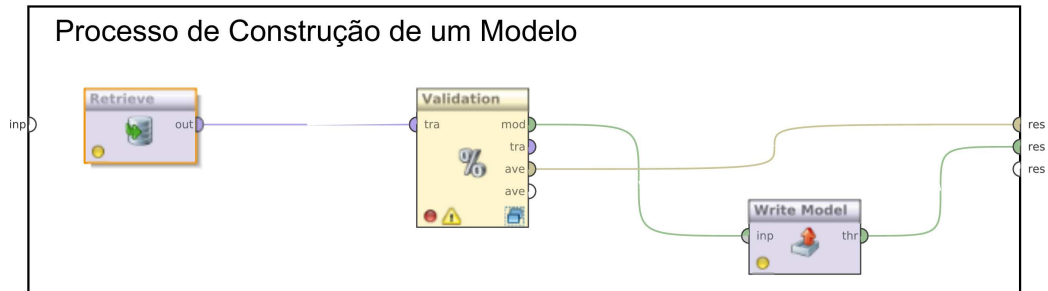


Figura 34 – Diagrama do Processo de Construção do Modelos de Classificação

O operador *Retrieve* é responsável por ler os registros do banco de dados de treinamento e o operador *WriteModel*, salva o modelo obtido do sub-processo de treinamento e avaliação do modelo (denominado de *Validation* no diagrama). Este sub-processo de treinamento e avaliação é diferente para cada um dos modelos de classificação. A Figura 35 apresenta a estrutura do processo empregado para treinar e testar o modelo NaiveBayes.

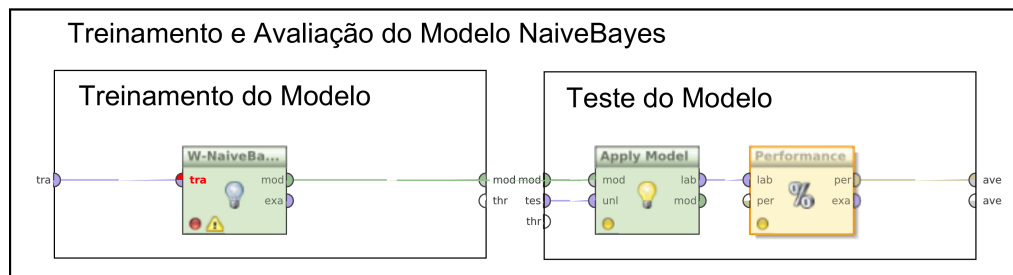


Figura 35 – Estrutura do Processo de Treinamento e Avaliação do Modelo NaiveBayes

O operador *W-NaiveBayesSimple* do RapidMiner que implementa o algoritmo de aprendizagem supervisionada de modelos *NaiveBayes* foi empregado utilizado para criar o modelo de classificação a partir dos dados de treinamento. Este operador é, na verdade, importado do software *Weka*, uma vez que o algoritmo de aprendizagem nativo do *RapidMiner* apresenta um comportamento e desempenho inferior a versão *Weka* deste algoritmo.

O operador *ApplyModel* efetua a aplicação do modelo obtido pelo algoritmo de treinamento, sobre dados de teste gerados pelo operador *Performance*. O tipo de validação selecionada no modelo é *Cross Validation*, configurada para utilizar dez por cento dos dados de treinamento para testes, escolhidos de forma aleatória. Ou seja, para treinamento do algoritmo o processo seleciona 90% dos dados aleatoriamente. Os demais são utilizados para testar o modelo, apresentando os resultados em uma matriz, cujo formato será explicado posteriormente.

O processo de treinamento do modelo KNN é similar ao empregado no modelo *Naive Bayes*. A Figura 36 mostra a estrutura interna do processo de treinamento e teste do modelo KNN.

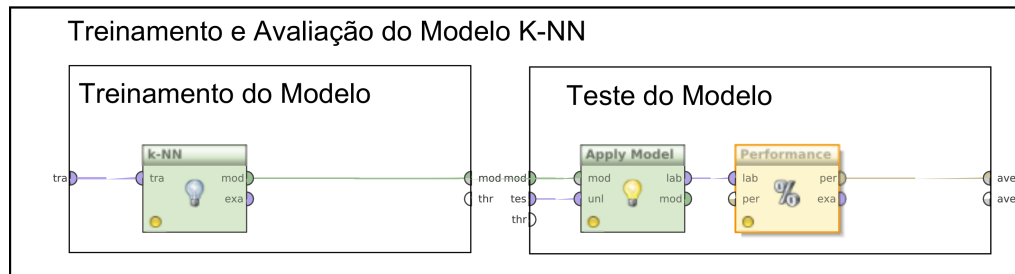


Figura 36 – Estrutura do Processo de Treinamento e Avaliação do Modelo KNN

A maior diferença é a utilização do operador KNN do *RapidMiner* para executar o algoritmo de aprendizagem supervisionada KNN. Nesse caso foi utilizado o operador nativo do *RapidMiner*, que é o único disponibilizado pela ferramenta. O processo de avaliação (teste) do modelo foi o mesmo empregado para o modelo *Naive Bayes*.

A saída básica do processo de treinamento e avaliação de um modelo de classificação é composta da *Matriz de Confusão* (ver Tabela 11) que indica os acertos e erros do classificador, considerando uma classificação elementar onde os exemplos pertencentes a classe são usualmente rotulados como “+” (positivos), enquanto os exemplos não pertencentes a classe são rotulados como “-” (negativos). Os acertos incluem VP, o número de exemplos positivos da classe que foram previstos corretamente pelo classificador e VN, o número de exemplos negativos da classe que também foram previstos corretamente. Os erros incluem os falsos positivos (FP) que são realmente exemplos negativos da classe, mas que foram previstos incorretamente pelo classificador como sendo positivos e os falsos negativos (FN) que são exemplos positivos da classe previstos incorretamente pelo classificador como exemplos negativos.

Tabela 11 – Matriz de Confusão

Classe	Verdadeira +	Verdadeira -
Prevista +	Verdadeiros Positivos VP	Falsos Positivos FP
Prevista -	Falsos Negativos FN	Verdadeiros Negativos VN

Fonte: Desenvolvido pelo Autor

Com base na Matriz de Confusão podem ser geradas várias de medidas de desempenho do modelo de classificação:

1. *Precisão* – a porcentagem dos documentos que foram corretamente rotulados como pertencentes à classe (6).
2. *Recall* - a porcentagem de todos os documentos pertencentes à classe em questão que conseguiram ser recuperados (7).
3. *Acurácia* - a porcentagem dos documentos que foram corretamente classificados (8).

A Tabela 12 mostra os resultados das medidas de desempenho dos classificadores desenvolvidos neste trabalho. Com base nestes dados é possível obter alguns resultados iniciais que indicam a viabilidade da metodologia de pesquisa sendo empregada neste trabalho.

A taxa de erro de ambos classificadores é bastante baixa (inferior a 5%). O valor alto de acurácia indica que a grande maioria dos registros foi corretamente classificado, indicando que é viável a utilização de ambos classificadores para identificar casos de risco.

A taxa de precisão de classificações corretas do paciente de risco (Sim) é alta, sendo superior a 70% (quase 80% no caso do *Naïve Bayes*). Esse é um ponto importante do presente trabalho, porque um índice de precisão alto é muito desejável para evitar a perda de pacientes que estejam realmente em caso de risco.

O modelo bayesiano possui um índice de *recall* maior que o classificador KNN. O classificador *Naïve Bayes* classificou corretamente um número maior de pacientes dentro do conjunto de pacientes que realmente possui risco.

Por outro lado, o índice de 99% de acurácia do classificador *Naïve Bayes*, pode indicar uma situação de *overfitting* deste classificador. Idealmente um classificador construído pelo algoritmo de aprendizagem deveria ter a capacidade de generalizar o conhecimento adquirido para novas situações que não foram apresentadas durante o treinamento. O *overfitting* ocorre quando o classificador fica demasiado especializado nos casos recebidos no treinamento, mas não consegue generalizar para outros casos. A solução encontrada para esse problema, justamente foi a utilização de ambos classificadores, uma vez que o K-NN, dado seus índices de acurácia, precisão e *recall*, parece menos sujeito ao problema.

Tabela 12 – Métricas dos Modelos de Classificação

Classificador	Medidas		
	Acurácia	Precisão (Sim)	Recall (Sim)
Naive Bayes	99,08%	78,33%	94,33%
K-NN	97,94%	72,85%	50,23%

Fonte: Desenvolvido pelo autor

Assim o sistema desenvolvido para atender o problema de pesquisa utilizará ambos modelos de classificação para prever se o paciente que o médico está atendendo possui algum fator de risco de agravamento de sua doença cardíaca. Por isso, uma vez classificado como paciente com risco cardiológico por qualquer um dos modelos, o médico será comunicado.

## 6 EXPERIMENTOS E AVALIAÇÃO

Neste capítulo serão apresentados os experimentos conduzidos para avaliar o modelo de identificação de pacientes cardíacos de risco, além dos testes funcionais efetuados com o intuito de validar o protótipo do sistema Cardiac-M. Ao final da descrição dos experimentos e testes, são apresentados e analisados os resultados obtidos durante esse processo de avaliação. O processo de avaliação do modelo e do protótipo do sistema Cardiac-M contou com o suporte de um médico especialista na área da medicina do trabalho e cardiologia.

### 6.1 Experimentos de Avaliação do Modelo de Risco

A avaliação do modelo de identificação de risco, que é composto dos classificadores *Naive Bayes* e KNN, foi executada em uma fase posterior ao processo de treinamento e criação destes classificadores (descrito no Capítulo 4), levando em conta apenas a comparação da previsão destes classificadores com o diagnóstico fornecido por um médico especialista. Todas as informações utilizadas no teste pertencem a casos reais que foram atendidos nos meses de setembro e outubro de 2010.

Para avaliação do modelo de identificação de risco foram extraídas informações de novos pacientes que foram atendidos nos meses de setembro e outubro de 2010. Pacientes que foram classificados como crônicos cardíacos de risco em um determinado mês, não foram avaliados nos demais meses. Este grupo de registros foi submetido inicialmente à avaliação do especialista médico que identificou 73 pacientes com risco no mês de setembro e 68 pacientes no mês de outubro de 2010. A seguir, este mesmo conjunto de registros foi submetido à classificação dos modelos *Naive Bayes* e KNN. A tabela abaixo mostra o resultado da comparação da previsão dos modelos com o diagnóstico do médico especialista.

Tabela 13 – Avaliação da Previsão dos Modelos de Classificação

Mês	Total de Registros	Previsão Médico	Previsão Classificador		Acertos Classificador		
			Naive Bayes	KNN	Naive Bayes	KNN	Naive Bayes + KNN
09/2010	2853	73	55	63	38 (52,05%)	42 (57,53%)	56 (76,71%)
10/2010	2530	68	45	51	35 (51,47%)	34 (50,00%)	45 (66,17%)

Fonte: Desenvolvido pelo autor

Analisando os resultados obtidos, é visível que ambos classificadores são equivalentes quanto ao percentual de acertos. No entanto, confirma-se que o uso de ambos classificadores é mais eficaz, uma vez que o número de pacientes identificados como risco é muito maior (76,71% dos pacientes identificados como risco em setembro de 2010 e 66,17% dos pacientes identificados como risco no mês de outubro de 2010).



Considerando a avaliação médica e os resultados obtidos, foi definido então o uso de um modelo de classificação probabilística híbrido, que utiliza vários critérios de classificação e apresenta o resultado como junção do retorno da aplicação dos métodos de classificação *Naive Bayes* e KNN.

## 6.2 Avaliação do Protótipo do Sistema Cardiac-M

Considerando o tempo disponível para avaliação do protótipo desenvolvido, foram selecionados cinco pacientes aleatoriamente para efetuar os testes funcionais. O médico especialista participou destes testes, principalmente com relação à avaliação da análise de risco informada pelo sistema. Como as informações de saúde são de sigilo médico, os pacientes foram anonimizados.

Dos cinco pacientes avaliados, quatro eram potenciais pacientes com risco cardíaco. A seguir serão exibidos os testes efetuados e os resultados obtidos.

Depois de conectado no sistema, o médico precisa selecionar o paciente a ser atendido. A figura 37 apresenta a interface onde o médico realiza esta pesquisa.



Seleção de Paciente

Paciente:

LOURDES ABC  
 LOURDES DEF  
 LOURDES GHI  
 LOURDES JLM

Figura 37 - Seleção do Paciente para Atendimento no Protótipo do Sistema Cardiac-M

Uma vez selecionado o paciente, o sistema apresenta a tela de atendimento, ilustrada na figura 38, onde o médico informa dados do atendimento do paciente e grava a informação no prontuário médico.

**Atendimento Médico**

Paciente: LOURDES DEF

Fuma?  Sim  Não  Bebe?  Não  Sim  Pratica Atividade Física?  Não  Sim

**Histórico Familiar**  
Doenças Cardiovasculares  Parentesco com Doença Cardiovascular: Mãe

**Dados de Saúde**  
Diabetes   
Dislipidemia   
Obesidade   
Doenças Cardiovasculares   
Hipertensão  Pressão Arterial:   
IMC: 24,00 Classificação IMC: Sobrepeso

**Queixas**

Psiquiátrico	Neurológico	Cardiológico	Sono
<input type="checkbox"/> Nega alteração			<input type="checkbox"/> Nega alteração
<input checked="" type="checkbox"/> Alteração de memória			<input type="checkbox"/> Agitação no leito
<input type="checkbox"/> Alucinação			<input type="checkbox"/> Agitação no horário biológico
<input type="checkbox"/> Ansiedade	<input type="checkbox"/> Nega alteração	<input type="checkbox"/> Nega alteração	<input type="checkbox"/> Concilia o sono apenas com uso de medicação
<input type="checkbox"/> Choro	<input type="checkbox"/> Arritmias	<input type="checkbox"/> Arritmias	<input type="checkbox"/> Demora a dormir
<input type="checkbox"/> Depressão	<input type="checkbox"/> Cefaléia	<input type="checkbox"/> Dispneia	<input type="checkbox"/> Dorme e acorda cansado
<input type="checkbox"/> Ideação Suicida	<input type="checkbox"/> Convulsão	<input type="checkbox"/> Dor torácica ao esforço	<input type="checkbox"/> Dorme e acorda não conciliando mais o sono
<input type="checkbox"/> Impulsividade	<input type="checkbox"/> Incoordenação	<input type="checkbox"/> Dor torácica ao repouso	<input type="checkbox"/> Insônia apesar do uso de medicação
<input type="checkbox"/> Insônia	<input type="checkbox"/> Tique	<input type="checkbox"/> Edema	<input type="checkbox"/> Pensamentos persistentes
<input type="checkbox"/> Instabilidade Emocional	<input type="checkbox"/> Tonturas	<input type="checkbox"/> Edema de membros inferiores	<input type="checkbox"/> Sono agitado
<input type="checkbox"/> Irritabilidade	<input type="checkbox"/> Vertigens	<input type="checkbox"/> Palpações	<input type="checkbox"/> Sono durante o dia e à noite
<input type="checkbox"/> Medo	<input type="checkbox"/> Outros	<input type="checkbox"/> Sopros	
<input type="checkbox"/> Nervosismo		<input type="checkbox"/> Outros	

Figura 38 - Interface para Registro de Atendimento Médico do Protótipo do Sistema Cardiac-M

Uma vez que esta informação é gravada no prontuário médico do paciente, o sistema chama o método para avaliar a situação de risco do paciente. O retorno obtido para o paciente *Lourdes DEF* (A) é de que identificou uma situação de risco e este diagnóstico foi validado pelo médico especialista. Ambos classificadores identificaram o risco. A mensagem exibida neste caso é apresentada na figura 39.

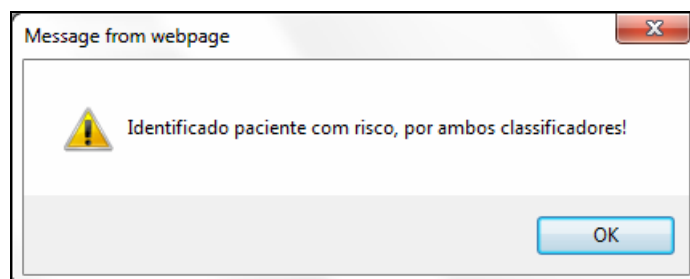


Figura 39 - Alerta emitido ao médico quando o paciente é avaliado com risco por ambos classificadores

Outro exemplo utilizado para validação foi o paciente *H123* (B), também diagnosticado como paciente de risco pelo médico especialista e cujo registro de atendimento é exibido na figura 40.



Figura 40 - Interface do Registro de Atendimento do Paciente *H123*

O paciente *H123* foi avaliado com risco cardíaco por apenas um dos classificadores. A figura 41 exibe o alerta exibido neste caso.

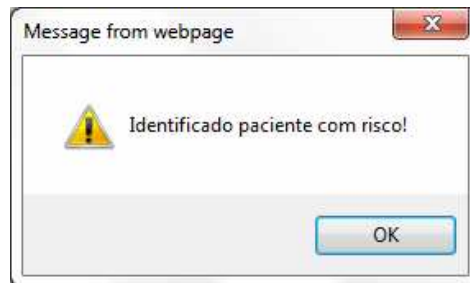


Figura 41 - Alerta emitido ao médico quando o paciente é avaliado com risco por apenas um dos classificadores

Além destes, outros três pacientes foram utilizados nos testes. Os pacientes (C) e (D), segundo o especialista médico são considerados pacientes de risco, enquanto o paciente (E) não é considerado de risco. A tabela abaixo exibe um resumo das características dos pacientes submetidos aos testes, a avaliação realizada pelo protótipo e a avaliação do médico especialista.

Tabela 14 - Resultados dos Casos de Teste do Protótipo do Sistema Cardiac-M

Paciente	Informações de Saúde	Retorno Protótipo	Avaliação Médico
(A)	Idade = 52 anos Tabagismo = Sim Atividade Física = Não Parentesco = Mãe Hipertensão = Sim Diabetes = SR Obesidade = SR	Sexo = Feminino Etilismo = Nega HF DC = Sim DC = Sim PA = SR Dislipidemia = SR IMC = 24,60	Risco identificado pelos dois modelos de classificação  Paciente com risco

	Cl. IMC = Sobrepeso Q. Neurológicas = 0 Q. Sono = 0	Q. Psiquiátricas = 1 Q. Cardiológicas = 0		
(B)	Idade = 43 anos Tabagismo = Sim Atividade Física = Não Parentesco = Irmãos Hipertensão = Não Diabetes = SR Obesidade = SR Cl. IMC = Obes. Grau I Q. Neurológicas = 0 Q. Sono = 0	Sexo = Masculino Etilismo = Nega HF DC = SR DC = Não PA = Hip. estágio 1 Dislipidemia = SR IMC = 30,47 Q. Psiquiátricas = 0 Q. Cardiológicas = 0	Risco identificado por um dos modelos de classificação	Paciente com risco
(C)	Idade = 57 anos Tabagismo = Ex-fumante Atividade Física = Não Parentesco = Pai Hipertensão = Sim Diabetes = SR Obesidade = SR Cl. IMC = Sobrepeso Q. Neurológicas = 0 Q. Sono = 0	Sexo = Masculino Etilismo = Nega HF DC = Sim DC = Sim PA = Hip. estágio 2 Dislipidemia = SR IMC = 27,39 Q. Psiquiátricas = 0 Q. Cardiológicas = 0	Risco identificado pelos dois modelos de classificação	Paciente com risco
(D)	Idade = 19 anos Tabagismo = Sim Atividade Física = Não Parentesco = SR Hipertensão = Não Diabetes = SR Obesidade = SR Cl. IMC = Normal Q. Neurológicas = 0 Q. Sono = 1	Sexo = Masculino Etilismo = Sim HF DC = SR DC = SR PA = Normal Dislipidemia = SR IMC = 24,67 Q. Psiquiátricas = 0 Q. Cardiológicas = 0	Não identificou risco	Paciente com risco
(E)	Idade = 38 Tabagismo = Ex-fumante Atividade Física = Não Parentesco = Pai Hipertensão = Não Diabetes = SR Obesidade = SR Cl. IMC = Sobrepeso Q. Neurológicas = 0 Q. Sono = 0	Sexo = Feminino Etilismo = Nega HF DC = Sim DC = SR PA = SR Dislipidemia = SR IMC = 26,7 Q. Psiquiátricas = 0 Q. Cardiológicas = 0	Não identificado risco	Paciente sem risco

Fonte: Desenvolvido pelo autor

Dando continuidade aos testes funcionais do sistema, o médico especialista avaliou a usabilidade do mesmo considerando alguns requisitos: facilidade de uso, agilidade e resultado. Com relação ao resultado, a análise apresentada na tabela 14 demonstra a acurácia dos modelos aplicados. A interface desenvolvida é equivalente ao sistema que o médico avaliador utiliza habitualmente. Por isso, não houve dificuldades no uso do protótipo.

O requisito que, segundo a avaliação do médico, não foi atendido é a agilidade. O registro de atendimento médico precisa ser algo ágil e a aplicação dos modelos para avaliação de risco demonstrou-se lento na perspectiva do médico. Nesse caso é importante salientar que o protótipo do Sistema Cardiac-M ainda não está implementado como um Sistema

Multiagente Distribuído, tal como previsto na arquitetura definida no capítulo 4. Uma vez que a versão final do sistema esteja completamente distribuída em vários agentes o problema de desempenho deve ser bastante minorado ou eliminado, uma vez que cada agente pode ser implementado como um processo concorrente distinto, que pode executar em máquinas diferentes da do servidor WEB onde atualmente está implantado o protótipo.

Segundo opinião deste especialista, independente da agilidade no instante do atendimento, a análise posterior é mais indicada, pois, uma vez identificado um risco, é necessário fazer uma análise detalhada, inclusive de todo histórico de saúde do paciente em questão, que nem sempre é possível durante o período da consulta médica. Esse problema também deve ser resolvido com o processo assíncrono, com envio de alertas através de e-mail. O processo assíncrono está previsto na arquitetura do sistema, mas não foi desenvolvido no protótipo devido à falta de tempo, uma vez que a elaboração e avaliação dos modelos de classificação foi muito além do tempo previsto.

## 7 CONSIDERAÇÕES FINAIS

Neste capítulo são apresentadas as conclusões finais sobre o trabalho desenvolvido. Para tanto, na seção inicial, é feita uma recapitulação dos objetivos definidos para o trabalho, seguida de uma análise das principais contribuições alcançadas pelo trabalho e dos principais problemas encontrados em sua execução. Logo após, a seção de trabalhos futuros apresenta sugestões de tópicos de pesquisa que pode ser desenvolvidos com base nos resultados obtidos com o presente trabalho, bem como possibilidades de integração com outros temas de pesquisa.

### 7.1 Conclusões

Conforme citado anteriormente, as doenças crônicas, sobretudo as doenças crônicas cardíacas, aumentaram expressivamente nos últimos anos. Não somente a incidência destas doenças, mas principalmente o aumento no percentual de mortalidade envolvendo doentes cardíacos. O presente trabalho foi desenvolvido para propor uma solução com o objetivo de auxiliar o médico na identificação de riscos em pacientes crônicos cardíacos. A principal questão de pesquisa buscava verificar se, através do uso da tecnologia de agentes, mineração de dados e utilização de modelos de classificação híbridos, seria possível desenvolver um sistema capaz de atender essa necessidade.

Para tanto foi desenvolvido um modelo de estimativa de risco e implementado um protótipo do sistema Cardiac-M de identificação de risco que opera integrada a um sistema hospitalar previamente existente. A arquitetura e protótipo do sistema Cardiac-M são apresentados no Capítulo 4 e o modelo de identificação de risco é descrito no Capítulo 5. Os experimentos conduzidos com o modelo e com o protótipo do sistema Cardiac-M, apresentados no Capítulo 6, mostram a viabilidade do uso dessas tecnologias para resolver a questão de pesquisa proposta para a dissertação.

O processo de construção dos modelos de avaliação de risco tomou grande parte do tempo de desenvolvimento do trabalho por tratar-se de um item fundamental para avaliação da solução. Com a base de informações de saúde extraída de um banco de dados com informações reais, a avaliação de um médico especialista foi muito importante para o sucesso do trabalho. E, decorrente da análise deste profissional, ocorreram várias mudanças nos critérios de classificação dessas informações durante o desenvolvimento do trabalho. Em consequência, muitas análises e modelagens precisaram ser refeitas, bem como a escolha da técnica de classificação a ser aplicada. Isto resultou na construção de um modelo de classificação híbrido, cuja base de treino utilizou três critérios de classificação: diretrizes da Sociedade Brasileira de Cardiologia, Escore de Framingham e critérios acrescentados pelo médico especialista.

Não obstante essas dificuldades, o resultado da aplicação deste modelo foi preciso e aprovado pelo médico especialista. Essa é a maior contribuição da presente pesquisa, mostrando a viabilidade da utilização de modelos probabilísticos de classificação, construídos através de técnicas de mineração de dados, na previsão de estimativa de risco de pacientes cardíacos crônicos.

Outra contribuição importante é a verificação, ainda que em nível de projeto, de que a tecnologia de agentes é útil na concepção de sistemas de acompanhamento de pacientes crônicos. A solução para questões de desempenho e pro-atividade, efetivamente identificadas no teste funcional do protótipo, estão previstas na arquitetura multiagente definida no Capítulo 4, através da distribuição da carga de trabalho em agentes independentes e da criação de agentes específicos para acompanhamento do estado de saúde dos pacientes crônicos, com envio de alerta por *e-mail* em caso de ocorrência de risco.

## **7.2 Trabalhos Futuros**

Para dar continuidade a este trabalho, existem algumas sugestões de trabalhos futuros. Entre elas, a primeira é implementar a solução utilizando agentes, conforme proposto na arquitetura, uma vez que, devido ao longo intervalo de tempo necessário para a elaboração do modelo de estimativa de risco, não houve tempo suficiente para fazer essa implementação durante este trabalho.

O projeto de arquitetura multiagente desenvolvido prevê a integração de novos modelos de avaliação de risco, através da criação e incorporação de novos agentes de estimativa de risco. Esta característica cria outra possibilidade de evolução do presente trabalho, inserindo novos modelos de classificação, além dos probabilísticos KNN e *Naive Bayes* utilizados no presente.

Outra sugestão interessante, é a utilização de algoritmos de avaliação de risco baseados em modelos de classificação incrementais, de forma que o modelo seja re-treinado a cada inclusão de dados na base de dados de saúde. Esta é uma das vantagens de se empregar modelos probabilísticos, como o *Naive Bayes*, que suportam a utilização de mecanismos incrementais.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BELLIFEMINE, F. L.; CAIRE, G.; GREENWOOD, D. **Developing Multi-Agent Systems with JADE**. John Wiley & Sons, 2007.
- BOND, A. H.; GASSER, L. (eds) **Readings in Distributed Artificial Intelligence**. Morgan Kaufmann, San Mateo, CA, 1988.
- BORDINI, R.; HÜBNER, J. F.; WOOLDRIDGE, Michael. **Programming Multi-Agent Systems in AgentSpeak using Jason**. 2007. ISBN: 970-0-470-02900-8.
- BORN, D., Cardiopatia Congênita. **Arq. Bras. Cardiol.**, v.93, n.6, supl.1, São Paulo, Dez., 2009.
- BRASIL, Ministério da Saúde. Secretaria de Vigilância à Saúde. Secretaria de Atenção à Saúde. **Diretrizes e recomendações para o cuidado integral de doenças crônicas não-transmissíveis: promoção da saúde, vigilância, prevenção e assistência**. Brasília: Ministério da Saúde, 2008.
- BRASIL, Ministério da Saúde. Mortes por doenças crônicas caem 17% no Brasil. **Portal da Saúde**. Brasília: Ministério da Saúde: 2010. Disponível em <<http://www.saude.gov.br/>>. Acesso em julho de 2011.
- BRATMAN, M. E.; ISRAEL, D. J.; POLLACK. Plans and Resource-Bounded Practical Reasoning. **Computational Intelligence**, v.4, n.3, p.349-355. 1988.
- CALLOU, M.R.A, RAMOS, P.R.M, Hipertensão Arterial Pulmonar. **Arquivos Brasileiros de Cardiologia**, v.93, n.6, supl.1, São Paulo, Dez., 2009.
- CAVALCANTE, P., **A importância dos fatores de risco na obstrução das artérias coronárias utilizando técnicas de mineração de dados**. Goiás: Universidade Católica de Goiás, 2009.
- CHAIB-DRAA, B; DIGNUM, F. Trends in agent communication language. **Computational Intelligence**. v.2, n.5, p.1-14. 2002.
- COPPIN, J. **Artificial Intelligence Illuminated**. Jones and Bartlett Publ. 2004. ISBN 0-7637-3230-3.
- COSTA, E.A., KLEIN, C.H., Meio urbano e doenças cardiovasculares. **Cadernos de Saúde Pública**, v.1, n.3, p.305-312, jul./set., Rio de Janeiro, 1985.
- DE WOLF, T.; HOLVOET, T. Towards a Full Life-cycle Methodology for Engineering Decentralised Multi-Agent Systems. In: OOPSLA 2005, **Proceedings**, 2005.
- DELOACH, S. A. Multiagent Systems Engineering Organization-Based Multiagent Systems. In: 4th Workshop on Software Engineering for Large-Scale Multi-Agent Systems (SELMAS'05), **Proceedings**, 2005.



- DOMINGOS, P., PAZZANI, M. On the optimality of the simple Bayesian classifier under zero-one loss. **Machine Learning**, v.29, p.103-130, 1997.
- DUNBAR-JACOB, J.; MORTIMER-STEPHENS, MK. Treatment adherence in chronic disease. **J. Clin. Epidemiol.**, v. 54, 2001.
- DEWULF, N.L.S., MONTEIRO, R.A., PASSOS, A.D.C., VIEIRA, E.M., TRONCON, L.E.A. Adesão ao tratamento medicamentoso em pacientes com doenças gastrintestinais crônicas acompanhados no ambulatório de um hospital universitário. **Revista Brasileira de Ciências Farmacêuticas**, v.42, n.4, out./dez., 2006.
- FIPA [Foundation for Intelligent Physical Agents]. SC00037J: **FIPA Communicative Act Library Specification**. Dezembro de 2002. Disponível em: <<http://www.fipa.org/specs/fipa00037/>> Acesso em abril de 2010.
- GEORGEFF, M.; PELL, B. et al. The Belief-Desire-Intention Model of Agency. Proceedings of the 5th International Workshop on Intelligent Agents V, Agent Theories, Architectures, and Languages. **Lecture Notes in Computer Science**. London, UK: Springer-Verlag. v.1555: p.1-10, 1999.
- GLUZ, J. C. ; FLORES, C. D ; VICARI, R. M . Formal Aspects of Pedagogical Negotiation in AMPLIA System. In: NEDJA, N; MOURELLE, L.; DE ALMEIDA, N.; BORGES, M. (Org.). **Intelligent Educational Machines**. New York: Springer, 2006, v. 44, p. 117-146.
- HAN, J., KAMBER, M., **Data Mining: Concepts and Techniques**. Simon Fraser University, 2000.
- HENDERSON-SELLERS, B.; GIORGINI P. (Eds) **Agent Oriented Methodologies**. Idea Group, 2005.
- IBGE, Instituto Brasileiro de Geografia e Estatística. **Pesquisa Nacional por Amostra de Domicílios 2008 – Um Panorama da Saúde no Brasil**, 2008. Disponível em <[http://www.ibge.gov.br/home/presidencia/noticias/noticia\\_visualiza.php?id\\_noticia=1580&id\\_pagina=1](http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=1580&id_pagina=1)>. Acesso em 17 de maio de 2011.
- JENNINGS, N.R., An Agent-Based Approach for Building Complex Software Systems. **Comm. the ACM**, v. 44, n. 4, p. 35-41, Apr. 2001.
- LABROU, Y.; FININ T.; PENG Y. Agent communication languages: the current landscape. **IEEE Intelligent Systems**, p.45-52, Março-abril de 1999.
- LAROSE, D., **Discovering Knowledge in Data An Introduction to Data Mining**. Wiley, Nov, 2004.
- LAURENTI, R., Epidemiologia das Doenças Cardiovasculares no Brasil. **Arq. Bras. Cardiol.**, v.38, n.4, p.243-248, abril, 1982.
- LOTUFO, P. A., O escore do risco de Framingham para doenças cardiovasculares. **Rev. Med (São Paulo)**. v. 87, n. 4, p. 232-7, out.-dez., 2008.

- MESQUITA, E.T., SOCRATES, J., RASSI, S., VILLACORTA, H., MADY, C., Insuficiência Cardíaca com Função Sistólica Preservada. **Arquivos Brasileiros de Cardiologia**, v.82, n.4, abril, 2004.
- MITCHELL, T. **Machine Learning**. Nova Iorque: McGraw-Hill, 1997.
- MUSAVI, A.; SANCHEZ, M.L. Using Electronic Institutions for Hospitals Chronic Disease Management and Purchasing System. In: 11th International Conference of the Catalan Association for Artificial Intelligence, **Proceedings**. Amsterdam, p. 80-87, 2008. Disponível em <<http://portal.acm.org/citation.cfm?id=1566913>> Acesso em 20 de maio de 2011.
- NICOLAU, JC, TIMERMAN A, PIEGAS, L.S., MARIN-NETO K.A, RASSI, A.Jr., Guidelines for Unstable Angina and Non-ST-Segment Elevation Myocardial Infarction of the Brazilian Society of Cardiology, 2a. ed. **Arq. Bras. Cardiol**, v.89, n.4: p.89-131, 2007.
- PADGHAN, L.; WINIKOFF, M. Prometheus: A pragmatic methodology for engineering intelligent agents. In: OOPSLA 2002 Workshop on Agent-Oriented Methodologies, **Proceedings**, p. 97-108, Seattle, 2002.
- PADGHAN, L.; WINIKOFF, M. **Developing Intelligent Agent Systems : A practical guide**. Australia: Wiley, 2004.
- RAO, A. S.; GEORGEFF, M. Modeling rational agents within a BDI-architecture. International Conference on Principles of Knowledge Representation and Reasoning, **Proceedings**. Cambridge, Massachusetts: Morgan Kaufmann, p. 473-484, 1991.
- RUSSEL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. Second Edition. Prentice Hall. 2003.
- SAHAMI, M. A bayesian approach to filtering junk E-mail. In: Learning for Text Categorization Workshop, **Proceedings**. Madison, Wisconsin: AAAI Technical Report WS-98-05, 1998.
- SAUDE, Ministério da Saúde. **Estratégia Nacional de Vacinação Contra o Vírus Influenza Pandêmico (H1N1) 2009**, Brasília, 2010. Disponível em <[http://portal.saude.gov.br/portal/arquivos/pdf/informe\\_tecnico\\_vacina\\_18\\_03.pdf](http://portal.saude.gov.br/portal/arquivos/pdf/informe_tecnico_vacina_18_03.pdf)>. Acessado em 02 de junho de 2006.
- SBC, Sociedade Brasileira de Cardiologia, **Diretrizes de Doença Coronariana Crônica Angina Estável**, v.83, n.2, set., 2004.
- SBC, Sociedade Brasileira de Cardiologia, **V Diretrizes Brasileiras de Hipertensão Arterial**. São Paulo, SP, fevereiro de 2006.
- SILVA, L.; ITO, M. **Proposta de Integração das Unidades de Saúde Utilizando um Sistema Multiagente para Acompanhamento de Pacientes Crônicos**, Centro Estadual de Educação Tecnológica Paula Souza, CEETEPS, São Paulo, 2007.
- TAN, P., STEINBACH, M., KUMAR, V. **Introdução ao DataMining: Mineração de Dados**. Rio de Janeiro, Ed. Ciência Moderna, 2009.

- TRINDADE, C.M., Descoberta de padrões de comportamento das Hepatites Virais aplicando Datamining. **Boletim epidemiológico, Secretaria de Estado da Saúde do Paraná**, v.21, p.8-9, 2004.
- VICARI, R. M.; GLUZ, J. C. An Intelligent Tutoring System (ITS) View on AOSE. **International Journal of Agent-Oriented Software Engineering (IJAOSE)** (Print), v.1, p.295 - 333, 2007.
- VICARI, R. M.; FLORES, C. D.; SEIXAS, L. ; GLUZ, J. C. ; COELHO, H. AMPLIA: A Probabilistic Learning Environment. **International Journal of Artificial Intelligence in Education**, v. 18, p. 347-373, 2008.
- WEISS, G. Multiagent systems: A modern approach to distributed artificial intelligence. Londres: MIT Press, 1999.
- WERNECK, V.M.B. Monitor Glicêmico: Um Sistema Multi-Agentes para Controle de Diabetes. In: X CBIS - Congresso Nacional de Informática na Saúde, **Anais**. Florianópolis, 2006. CD do X Congresso Nacional de Informática na Saúde. v. 1. p. 1200-1210.
- WHO, World Health Organization. **Chronic Diseases**. Disponível em <[http://www.who.int/topics/chronic\\_diseases/en/](http://www.who.int/topics/chronic_diseases/en/)>. Acesso em 17 de maio de 2011.
- WHO, World Health Organization. **The Impact of Chronic Disease in The United States**. Disponível em <[http://www.who.int/chp/chronic\\_disease\\_report/usa.pdf](http://www.who.int/chp/chronic_disease_report/usa.pdf)>. Acesso em 17 de maio de 2011.
- WITTEN, I.; FRANK, E. **Data Mining: Practical Machine Learning Tools**. Elsevier, 2005.
- WOOLDRIDGE, M. **An Introduction to Multiagent Systems**. John Wiley & Sons, 2002.
- WOOLDRIDGE, M. **Reasoning about Rational Agents**. Cambridge, Massachusetts: MIT Press. 2000.
- WOOLDRIDGE, M.; JENNINGS, N. Intelligent Agents: Theory and Practice. **Knowledge Engineering Review**, v. 10, n. 2, 1995.
- WOOLDRIDGE, M.J.; JENNINGS, N.R. Software engineering with agents: pitfalls and pratfalls. **IEEE Internet Computing**, v. 3, n. 6, p.20-27, May/June 1999.