

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO E
SISTEMAS
NÍVEL MESTRADO

ROGÉRIO PEREIRA DE SOUSA

CLASSIFICAÇÃO LINEAR DE BOVINOS:

**Criação de um modelo de decisão baseado na conformação de tipo “*true type*”
como auxiliar a tomada de decisão na seleção de bovinos leiteiros**

São Leopoldo

2016

Rogério Pereira de Sousa

CLASSIFICAÇÃO LINEAR DE BOVINOS:

Criação de um modelo de decisão baseado na conformação de tipo "*true type*" como auxiliar a tomada de decisão na seleção de bovinos leiteiros

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Engenharia de produção e sistemas, pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Universidade do Vale do Rio dos Sinos - UNISINOS

Orientador: Dr. Guilherme Luís Roehe Vaccaro

São Leopoldo

2016

Ficha catalográfica

S725c Sousa, Rogério Pereira de
Classificação linear de bovinos: criação de um modelo de decisão baseado na conformação de tipo “true type” como auxiliar a tomada de decisão na seleção de bovinos leiteiros/ Rogério Pereira de Sousa.—São Leopoldo-RS, 2016.

93p. Il.

Dissertação (Mestrado em Engenharia de Produção e Sistemas). Universidade do Vale do Rio dos Sinos-UNISINOS. Programa de Pós-Graduação em Engenharia de Produção e sistemas.

Orientador: Prof.^o Dr. Guilherme Luís Roehe Vaccaro

1.Classificação Linear. 2. Data mining. 3. Lógica Fuzzi. I. Vaccaro, Guilherme Luís Roehe (Orient.). Título.

CDD 22.ed. 006.3

BIBLIOTECARIA: Elma Vital da Silva CRB2-1022

Rogério Pereira de Sousa

CLASSIFICAÇÃO LINEAR DE BOVINOS:

Criação de um modelo de decisão baseado na conformação de tipo “*true type*” como auxiliar a tomada de decisão na seleção de bovinos leiteiros:

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre, pelo Programa de Pós-Graduação em ENGENHARIA DE PRODUÇÃO E SISTEMAS da Universidade do Vale do Rio dos Sinos - UNISINOS

Aprovado em 29/08/ 2016

BANCA EXAMINADORA

Dr. Leonardo Dagnino Chiwiakowsky– UCS

Dr^a. Miriam Borchardt – Unisinos

Dr^a. Claudia Viviane Viegas– Unisinos

Dedico este trabalho a minha companheira
Elma, minhas filhas e a meus pais...

AGRADECIMENTOS

Agradeço a Deus pela oportunidade.

Um agradecimento especial ao meu amigo e orientador Prof. Dr. Guilherme Luís Roehe Vaccaro, por ter guiado no desenvolvimento da pesquisa e pelo incentivo e dedicação. Estendo o mesmo agradecimento ao professor Dr. Leonardo Dagnino Chiwiakowsky, a professora Dr^a. Miriam Borchardt e professora Dr^a. Claudia Viviane Viegas pela ajuda e contribuição.

Agradeço a todos os professores do Programa, pela dedicação, empenho e pela oportunidade de ter aprendido com cada um deles.

Agradeço a Lilian Amorim pela presteza, auxílio e pelas palavras de acalento.

Por fim, aos servidores do IFTO, alunos deste programa e meus companheiros, agradeço vossas presenças nos mais diversos momentos dessa jornada.

RESUMO

A seleção de bovinos leiteiros, através da utilização do sistema de classificação com características lineares de tipo, reflete no ganho de produção, na vida produtiva do animal, na padronização do rebanho, entre outros. Esta pesquisa operacional obteve suas informações através de pesquisas bibliográficas e análise de base de dados de classificações reais. O presente estudo, objetivou a geração de um modelo de classificação de bovinos leiteiros baseado em “*true type*”, para auxiliar os avaliadores no processamento e análise dos dados, ajudando na tomada de decisão quanto a seleção da vaca para aptidão leiteira, tornando os dados seguros para futuras consultas. Nesta pesquisa, aplica-se métodos computacionais à classificação de vacas leiteiras mediante a utilização mineração de dados e lógica fuzzy. Para tanto, realizou-se a análise em uma base de dado com 144 registros de animais classificados entre as categorias boa e excelente. A análise ocorreu com a utilização da ferramenta WEKA para extração de regras de associação com o algoritmo apriori, utilizando como métricas objetivas, suporte / confiança, e *lift* para determinar o grau de dependência da regra. Para criação do modelo de decisão com lógica fuzzy, fez-se uso da ferramenta R utilizando o pacote sets. Por meio dos resultados obtidos na mineração de regras, foi possível identificar regras relevantes ao modelo de classificação com confiança acima de 90%, indicando que as características avaliadas (antecedente) implicam em outras características (consequente), com uma confiança alta. Quanto aos resultados obtidos pelo modelo de decisão fuzzy, observa-se que, o modelo de classificação baseado em avaliações subjetivas fica suscetível a erros de classificação, sugerindo então o uso de resultados obtidos por regras de associação como forma de auxílio objetivo na classificação final da vaca para aptidão leiteira.

Palavras-chave: classificação linear, data mining, regras de associação, lógica fuzzy.

ABSTRACT

The selection of dairy cattle through the use of the rating system with linear type traits, reflected in increased production, the productive life of the animal, the standardization of the flock, among others. This operational research obtained their information through library research and basic analysis of actual ratings data. This study aimed to generate a dairy cattle classification model based on "true type" to assist the evaluators in the processing and analysis of data, helping in decision making and the selection of the cow to milk fitness, making the data safe for future reference. In this research, applies computational methods to the classification of dairy cows by using data mining and fuzzy logic. Therefore, we conducted the analysis on a data base with 144 animals records classified between good and excellent categories. Analysis is made with the use of WEKA tool for extraction of association rules with Apriori algorithm, using as objective metrics, support / confidence and lift to determine the degree of dependency rule. To create the decision model with fuzzy logic, it was made use of R using the tool sets package. Through the results obtained in the mining rules, it was possible to identify the relevant rules with confidence classification model above 90%, indicating that the characteristics assessed (antecedent) imply other characteristics (consequent), with a high confidence. As for the results obtained by the fuzzy decision model, it is observed that the classification model based on subjective assessments is susceptible to misclassification, suggesting then the use of results obtained by association rules as a way to aid goal in the final classification cow for dairy fitness

Key-words: linear classification, data mining, association rules , fuzzy logic

LISTA DE QUADROS

Quadro 1 - Sistema de Classificação até 2010	26
Quadro 2 - Sistema de Classificação Atual	27
Quadro 3 - Áreas de aplicação da Mineração de Dados.....	29
Quadro 4 - Quadro resumo do processo CRISP-DM	31
Quadro 5 – Classificações possíveis.....	42
Quadro 6 - Síntese teórica	48
Quadro 7 - Valores ideais de classificação	84

LISTA DE TABELAS

Tabela 1-Trabalhos relevantes.....	20
------------------------------------	----

LISTA DE FIGURA

Figura 1 -Modelo ideal de bovino leiteiro.....	25
Figura 2 - Exemplo de arquivo no formato .arff	39
Figura 3 - Representação de Pertinência.....	41
Figura 4 - Definição de meia-idade em conjuntos convencionais.....	42
Figura 5 - Definição de meia-idade com conjunto fuzzy	43
Figura 6 – Estrutura do sistema fuzzy	45
Figura 7 - Variáveis avaliadas em força leiteira.....	49
Figura 8 - Variáveis avaliadas em Garupa	50
Figura 9 - Variáveis avaliadas em Pernas e Pés.....	50
Figura 10 - Variáveis avaliadas em Sistema mamário.	51
Figura 11 - Processo de Modelagem	52
Figura 12 - Modelo de decisão	55
Figura 13 - Animais constantes na base de dados.....	57
Figura 14 - Categorização dos dados	57
Figura 15 - Definição de suporte e confiança no Weka.....	60
Figura 16 - Regras obtidas para todas categorias.....	62
Figura 17 - Regras obtidas para categoria força lombar	64
Figura 18 -Regras geradas da categoria garupa.....	65
Figura 19 - Regras geradas para categoria Perna e Pés	66
Figura 20 - Regras geradas para categoria Sistema Mamário.....	68
Figura 21 - Gráfico da variável linguística Estatura (est).....	69
Figura 22 - Gráfico da variável linguística nivelamento da linha superior (nsup)	70
Figura 23 - Gráfico da variável linguística largura do peito (lpeito).....	71
Figura 24- Gráfico da variável linguística Profundidade Corporal (pcorp)	71
Figura 25 - Gráfico da variável linguística Angulosidade (angulosi)	72
Figura 26 - Gráfico da variável linguística Condição Corporal (ccorp)	73
Figura 27 - Gráfico da variável linguística Ângulo da Garupa (angarupa).....	73
Figura 28 -Gráfico da variável linguística Largura da Garupa (lgarupa).....	74
Figura 29 - Gráfico da variável linguística Força Lombar (flomb).....	74
Figura 30 - Gráfico da variável linguística Condição Corporal (ccorp)	75
Figura 31 - Gráfico da variável linguística Profundidade do Talão (ptalao)	75
Figura 32 - Gráfico da variável linguística Qualidade óssea (qossea).....	76

Figura 33 - Gráfico da variável linguística Perna e Pés Vista Lateral (ppvlat).....	77
Figura 34 - Gráfico da variável Pernas e Pés Vista Posterior (ppvpost).....	77
Figura 35 - Gráfico da variável linguística Inserção do Úbere Anterior (insbant).....	78
Figura 36 - Gráfico da variável linguística Colocação dos tetos anteriores (ctetant) .	79
Figura 37 - Gráfico da variável linguística Comprimento do Teto (compteto).....	79
Figura 38 - Gráfico da variável linguística Profundidade do Úbere (profubb).....	80
Figura 39 - Gráfico da variável linguística Textura do Úbere (textub)	80
Figura 40 - Gráfico da variável linguística Ligamento médio (lmedio).....	81
Figura 41 - Gráfico da variável linguística Altura do Úbere Anterior (altubereant).....	81
Figura 42 - Gráfico da variável linguística Largura do Úbere Anterior (larubaant).....	82
Figura 43 - Gráfico da variável linguística Colocação dos Tetos Posteriores (coltepost).....	82
Figura 44 – Número de regras geradas e variáveis fuzzy	83
Figura 45 - Regras geradas.....	85

LISTA DE SIGLAS

ABCBRH	Associação Brasileira de Criadores de Bovinos da Raça Holandesa
ABEPRO	Associação Brasileira de Engenharia de Produção
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
EHFF	Federação Europeia de Gado Holandês
IBGE	Instituto Brasileiro de Geografia e Estatística
INTERBULL	Serviço de Avaliação Internacional de Touro
WHFF	Federação Mundial de Gado Holandês
ABNT	Associação Brasileira de Normas Técnicas
NBR	Normas Brasileiras de Regulação

SUMÁRIO

1 INTRODUÇÃO	13
1.1 DEFINIÇÃO DO TEMA.....	14
1.2 DELIMITAÇÕES DO TRABALHO	18
1.3 OBJETIVOS	19
1.3.1 Objetivo Geral	19
1.3.2 Objetivos Específicos	19
1.4 JUSTIFICATIVA	19
2 REVISÃO BIBLIOGRÁFICA	25
2.1 CARACTERÍSTICAS LINEARES DE TIPO EM REBANHOS DE BOVINOS LEITEIROS.....	25
2.2 MINERAÇÃO DE DADOS	28
2.3 TÉCNICAS SELECIONADAS DE MINERAÇÃO DE DADOS PARA ESTA PESQUISA.....	32
2.3.1 Regras de Associação	32
2.4 WEKA – WAIKATO ENVIRONMENT FOR KNOWEDGE ANALYSIS.....	37
2.5 LÓGICA FUZZY	39
2.7 SÍNTESE TEÓRICA	45
3 METODOLOGIA	45
3.1 MÉTODO DE PESQUISA	46
3.2 MÉTODO DE TRABALHO	46
3.3 LEVANTAMENTO BIBLIOGRÁFICO	52
3.4 CONSTRUÇÃO DO MODELO	53
3.4.1 Resultados preliminares	56
4 EXPERIMENTOS E RESULTADOS	59
4.1 APLICAÇÃO DE REGRAS DE ASSOCIAÇÃO	59
4.2 APLICAÇÃO DA TEORIA FUZZY	69
5 CONSIDERAÇÕES FINAIS	87
REFERÊNCIAS	88

1 INTRODUÇÃO

A evolução tecnológica e o surgimento contínuo de ferramentas e recursos computacionais fazem com que a aplicabilidade da computação alcance as mais diversas áreas de conhecimento para geração de valor, além de automatizar tarefas complexas do cotidiano. Os avanços no campo da tecnologia da informação e comunicação (TIC) possibilitaram às empresas armazenamento cada vez maior de dados oriundos de clientes, produtos e serviços, porém, com carências na extração de informação, resultando em pouco ou nenhum conhecimento atribuído a eles, deixando-os apenas como forma de base de dados para que sejam recuperados e/ou gerenciados. (DANTAS; DE SOUZA, 2008).

O setor pecuário, mesmo com a necessidade de manejar grandes quantidades de informações, possui um grau de resistência elevado às mudanças, não acompanhando a evolução dos demais setores da economia (COLBEK, 2011). Responsável pela cadeia produtiva animal, desenvolve recursos para uma maior produtividade e rentabilidade dos produtos derivados deste, partindo do princípio que a sobrevivência de um empreendimento rural, dentre outros fatores, dar-se-á pela sua capacidade de gerar lucro. No entanto, observa-se que analisar dados individualizados dos animais no tocante às características produtivas é uma tendência internacional, com vistas a gerar informações que garantam tomadas de decisões rápidas e precisas. (POLYCARPO, 2011). Endereçando parte dessas tendências, de incorporação de conceitos de gestão de informação e TICs para suporte à decisão, encontra-se a Zootecnia de Precisão.

A Zootecnia de Precisão dentro de suas especificidades, envolve a decisão sobre os fatores limitantes à produção, como custo e renda, visando cada vez mais à acurácia nas decisões a fim de evitar perdas de produção e conseqüentemente altos custos. Neste contexto, associada aos esforços da engenharia de produção buscam envolver um conjunto de técnicas aplicáveis a cada nível de planejamento para balizar o escopo desta cadeia produtiva por meio de Pesquisa Operacional e outras áreas do conhecimento. (MIGUEL et al., 2012).

Chaves (2011, p.13) conceitua a pesquisa operacional como sendo a “busca da melhor utilização de recursos e pessoas, através da aplicação de métodos científicos, visando uma melhor satisfação do cliente”. Segundo Miguel et al. (2012), a partir dos anos 90 com a expansão dos sistemas ERP (*Enterprise Resources*

Planning) nas empresas e a necessidade de TICs mais analíticas, teve início a incorporação de sistemas de apoio a decisão (DSS – *Decision Support Systems*), sistemas estes baseados em modelos de Pesquisa Operacional.

A inserção de princípios computacionais e de engenharia de produção aplicados à seleção de bovinos leiteiros abre caminho para o desenvolvimento de técnicas e tecnologias que auxiliem o pecuarista no manejo de seus animais, e preservação de dados de produtividade individualizados dos mesmos, com o intuito de aumentar a longevidade dos animais juntamente com a produtividade leiteira.

1.1 DEFINIÇÃO DO TEMA

No cenário Tocantinense, o Instituto Federal de Educação Ciência e Tecnologia do Tocantins (IFTO) é um difusor de tecnologia e mão de obra para as mais variadas áreas de conhecimento, inclusive para o desenvolvimento da pecuária leiteira nesta região. Segundo dados do IFTO Campus Araguatins, o mesmo disponibiliza para o mercado de trabalho, anualmente, 240 Técnicos em Agropecuária. (OLIVEIRA, 2014). Apesar do número de profissionais disponibilizados para o mercado de trabalho, no cenário nacional, de acordo com Venturini (2014), o Tocantins não se estabelece entre os dez maiores produtores de leite, ficando com menos de 1% da produção total de do país.

Considerado o contexto apresentado, esta pesquisa embasou-se na análise de estudos publicados que envolvem resultados de classificação linear de bovinos, mais especificamente nos modelos de tomada de decisão baseados em técnicas de pesquisa operacional, subárea da engenharia de produção. A pesquisa está enquadrada conforme documento elaborado pela comissão de graduação da ABEPRO (2008), no item 1.3.24, na área de concentração pesquisa operacional, subárea 1.3.24.01.01, Modelagem. Este estudo aborda o desenvolvimento de um modelo de decisão baseado na metodologia de seleção dos melhores animais para o rebanho leiteiro por meio do uso de um sistema baseado em regras de associação e lógica *fuzzy*, considerando os padrões estabelecidos pela associação brasileira de criadores de bovinos da raça holandesa

A seleção de bovinos leiteiros, através da utilização do sistema de características lineares de tipo, segundo Vallotto e Neto (2010), reflete não somente no ganho de produção, mas também em quesitos como a vida produtiva,

padronização do rebanho, sanidade, entre outros. Esteves et al. (2004a) conceituam a classificação linear como a metodologia aplicada para a avaliação das medidas de conformação dos animais utilizada pelas associações de criadores de gado leiteiro. Campos (2012) complementa afirmando que o termo “característica linear de tipo” é utilizado para referenciar aquelas características morfológicas que melhoram a eficiência da vaca não pelo aumento da produtividade, mas pela redução dos seus custos de produção. Assim sendo, na prática, a classificação é a atividade de mensurar as características externas, por meio de observações feitas por técnicos especializados. A utilização dessa técnica, conforme Mcmanus e Saueressig (1998), torna-se importante para que os produtores possam fazer uma gestão adequada da seleção e do melhoramento do rebanho, passando a entender detalhes dos pontos fortes e fracos, ajudando-os a identificar qual afinidade o animal possui, e onde devem focar para tê-lo com mais saúde, alta produção vitalícia e adequação a seu sistema de produção.

Os Estados Unidos, em 1929, foi o primeiro país a utilizar a técnica de classificação em bovinos leiteiros da raça holandesa. Para essa, foram consideradas quatro categorias sendo: (i) aparência geral; (ii) caracterização leiteira; (iii) condição de escore corporal; e (iv) sistema mamário. No ano de 1983, criou-se uma outra escala de pontuação, com o objetivo de auxiliar os técnicos classificadores a identificarem as vacas com maior longevidade por meio das melhores pontuações para as características com base em um modelo ideal. Denominado *True Type*, a pontuação passou a ser adotada pela Associação Norte Americana de Criadores da Raça Holandesa. (SHORT; LAWLOR, 1992).

Em 1988, a WHFF (Federação Mundial de Gado Holandês) aprovou a padronização das características lineares de tipo e o sistema de classificação proposto por grupo de trabalho da *European Holstein-Friesian Federation* – EHFF (Federação Europeia de Gado Holandês). A partir do ano 2000, a cada quatro anos, os representantes de diversos países integrantes da INTERBULL (Serviço de Avaliação Internacional de Touro) se reúnem para discutir e aprovar recomendações a nível internacional relacionadas a características lineares de tipo. A padronização com relação às características mais importantes a serem registradas pelos países é relevante pelo fato de facilitar a avaliação genética dos animais através do INTERBULL. (CAMPOS, 2012).

A mensuração das características lineares de tipo vem ganhando importância por estarem geneticamente relacionadas com a longevidade das vacas e também por características produtivas serem independentemente herdadas e ainda por auxiliarem produtores na tomada de decisão quanto à seleção de animais para produção. De modo geral, estudos (RENNÓ et al., 2003; VALLOTO; NETO, 2010; CAMPOS, 2012) apontam que a seleção natural para produção de leite pode, em longo prazo, deteriorar características funcionais (fertilidade, longevidade) e características de conformação do rebanho, diminuindo a vida produtiva das vacas.

No Brasil, a primeira classificação realizada ocorreu em 1983, pelos classificadores oficiais do Colégio Brasileiro de Classificadores (VALLOTO, 2014). De acordo com Campos (2012), o Brasil ainda não participa do INTERBULL, e, desde meados de 2010, a Associação Brasileira de Criadores de Bovinos da Raça Holandesa (ABCBRH) considera 23 características para seu sistema de classificação.

A ABCBRH usa o sistema linear de classificação para avaliar o tipo para o gado da raça holandesa no Brasil, sendo considerado por esta um modelo de classificação que possui escores que variam de 1(um) a 9 (nove) pontos. Neste modelo, as características estão divididas em quatro seções: (i) força leiteira; (ii) garupa; (iii) pernas e pés; e (iv) sistema mamário. Para esses, são atribuídos pesos percentuais de 22%, 10%, 26% e 42%, respectivamente para cada seção. (VALLOTO; NETO, 2010). Para a maioria das características, a pontuação máxima é o almejado, porém, pontuações intermediárias também são desejadas para outras características como por exemplo, 5 (cinco) pontos para comprimento dos tetos e nivelamento de garupa e 7 (sete) pontos para estatura, ângulo do casco e profundidade corporal. A classificação morfológica, forma que pode ser chamada de classificação linear, é realizada por técnicos classificadores formados pelo Colégio Brasileiro de Classificadores, e deve ser preferencialmente realizada no início da vida produtiva da vaca, ou seja, na primeira lactação. Quando não for possível, deverá a classificação realizar-se nas lactações seguintes. (MANUAL..., 2009).

A pontuação final deste modelo de seleção é a expressão do equilíbrio entre as 23 características lineares de tipo, cada uma delas ponderada dentro de cada seção. A sua importância está em refletir as características de um animal o mais próximo do tipo ideal, sendo que vacas classificadas entre 90 e 97 pontos são denominadas excelentes; entre 85 e 89 pontos, Muito Boa; entre 80 e 84, Boa para mais; entre 75 e 79, Boa; 65 a 74, Regular; e abaixo de 65 pontos, Fraca. (VALLOTO; NETO, 2010).

Segundo Valloto (2014), a classificação se torna importante por fornecer vantagens no manejo do rebanho uma vez que contribui com uma avaliação objetiva de cada animal, orienta e torna mais eficientes as realizações dos acasalamentos dirigidos, auxiliam na seleção das novilhas para reposição, facilita o descarte de vacas com defeitos importantes, avalia a eficiência do programa genético pela comparação da pontuação das mães e filhas, possibilita a evolução de animais e valoriza comercialmente famílias e rebanhos que utilizam este sistema.

Além da seleção de bovinos, utilizando a classificação para tipo com objetivos ao aumento da produção leiteira, torna-se necessário identificar e quantificar os fatores de meio ambiente em que o animal está sendo explorado. Estudos apontam que vários fatores influenciam na classificação do animal tais como ano e estação de classificação, estágio de lactação, idade, ordem do parto, classificador dentre outros. (WILCOX, 1959 apud ESTEVES, 2004a, p. 2).

Partindo da premissa de que o desenvolvimento produtivo de um rebanho leiteiro parte da seleção dos melhores animais para tal aptidão, entende-se que o melhoramento genético através de um modelo de classificação seria o passo inicial para o desenvolvimento da pecuária leiteira regional. Assim, e devido ao recente destaque obtido por esta instituição, através de projetos de desenvolvimento regional de produção leiteira em pequenas propriedades (SILVA et. al 2012), o presente estudo visa gerar um modelo de classificação baseado regras de associação e lógica *fuzzy* para atender demandas da região, com o intuito de melhorar a forma de análise de dados tornando o processo de tomada de decisão mais viável, ágil e seguro quanto às informações geradas. Com o intuito realizar um adequado armazenamento dos dados e análises para consultas posteriores, relatórios organizados e elaborados, com dados estatísticos também contribuem para a melhora das condições de suporte ao desenvolvimento da pecuária neste estado.

Considerados o contexto e os conceitos previamente apresentados, nesta pesquisa orienta-se a seguinte questão: Como obter maior eficiência na classificação de bovinos leiteiros por intermédio de técnicas de tomada de decisão?

1.2 DELIMITAÇÕES DO TRABALHO

A presente pesquisa gera modelo auxiliar de tomada de decisão, baseado em análises do sistema de seleção para classificação linear de tipo para bovinos leiteiros através da base de dados obtida junto a Associação Paranaense de Criadores de Bovinos da Raça Holandesa (APCBRH). A implementação do referido modelo também se utilizará da análise de dados de campo, junto à Unidade de Produção de Zootecnia III do campus Araguatins, reduzindo seu foco de pesquisa à aplicação do modelo a animais criados nesta unidade, na fase de testes.

O estudo está delimitado a pesquisar aspectos de seleção animal para tipo leiteiro baseado na classificação linear com avaliação de 23 (vinte e três) características, subdivididas em quatro seções conforme atualização do sistema de seleção para vacas da raça holandesa ocorrido em meados de 2010, como relatam Valloto e Neto (2010) em sua publicação. Outro aspecto da pesquisa está voltado à geração de conhecimento no que tange à verificação das variáveis que possuem maior peso na decisão final da seleção e o que leva o modelo a decidir por determinado resultado mediante os dados obtidos por coleta de campo.

Não serão explorados modelos de classificação para outras raças de bovinos leiteiros, senão a raça holandesa. Nas fronteiras desta pesquisa detêm-se apenas aspectos de seleção, não sendo verificados com o modelo vertentes de ganhos ou perdas de produção, não serão associados fatores de ambiente que possam influenciar na seleção animal como relatam em seu estudo Esteves et al. (2004a), além disso, fatores como herdabilidade genética e correlações de produção comparadas ao modelo estão fora do escopo desta pesquisa.

Para o modelo, serão usadas técnicas de mineração de dados baseadas em extração de regras de associação relacionadas ao resultado final da classificação, uma vez que as mesmas assumem que não se aplica a dados numéricos, ou seja, a tarefa de associação encontrará elementos que implicam na presença de outros elementos de uma mesma transação de forma a encontrar associações ou padrões frequentes no conjunto de dados. (BARANAUKAS, 2013).

O estudo em questão está direcionado a atender as demandas de classificação, podendo ser um instrumento auxiliar a classificação de bovinos leiteiros da raça holandesa com intuito de otimizar a padronização do rebanho e trabalhar em cima das deficiências e qualidades individuais de seus animais.

Apesar da pesquisa usar dados vinculados à seleção de animais no Estado do Tocantins, o modelo desenvolvido, com os devidos ajustes, possivelmente poderá atender outras propriedades com fins à produção leiteira no Brasil, pois se buscará baseá-lo em um sistema de classificação utilizado em nível nacional.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

O objetivo geral desta pesquisa é gerar um modelo de decisão baseado em um sistema de classificação linear de tipo ideal “*true type*”, para auxiliar a seleção de vacas para aptidão leiteira.

1.3.2 Objetivos Específicos

A fim de sustentar o objetivo geral, são apresentados os seguintes objetivos específicos:

- a) Encontrar padrões frequentes de características morfológicas em uma base de dados de classificação real em relação a vacas com classificação entre boa e excelente;
- b) Analisar os resultados de classificação posicionando a vaca por meio da lógica *fuzzy*, de acordo com sua pontuação final de classificação;
- c) Avaliar resultados obtidos pelo modelo.

1.4 JUSTIFICATIVA

Com o passar dos anos, as TICs têm evoluído a tal ponto que cada vez mais se tornam presentes em várias áreas produtivas, já que a capacidade computacional vem aumentando e as Instituições de Ensino (IEs) colaboram com pesquisas e com profissionais mais capacitados para lidar com esta realidade. Na pecuária, mais especificamente na pecuária leiteira, o sucesso dos programas de melhoramentos genéticos, que envolvem características de tipo e de produção dependem diretamente da qualidade da coleta de dados obtida em campo, sendo esta aliada a softwares específicos pode se tornar uma importante ferramenta no apoio à tomada de decisão.

A utilização de softwares aplicados à pecuária leiteira no tocante à seleção de animais com as melhores características de conformação pode aumentar substancialmente a rapidez na coleta, análise e confiabilidade nos resultados obtidos. A informatização e a automatização de rotinas de trabalho podem garantir maior produtividade e eficiência no manejo dos animais com custo de investimento relativamente mais baixo, visto que equipamentos com alto poder de processamento computacional estão cada dia mais acessíveis e com interfaces mais intuitivas.

Do ponto de vista acadêmico, observa-se que a construção de um modelo de decisão diminuiria as lacunas entre a avaliação e a produção animal, se observada a capacidade de processamento e eficiência na estruturação de dados, visando à possibilidade de estimativas de resposta animal ao modelo desenvolvido. Para sustentar essa afirmação, esta pesquisa concentrou-se em realizar buscas por artigos e dissertações nas bases de periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Google Scholar, com vistas a fomentar o embasamento teórico e objetivando a verificação da importância da presente pesquisa frente ao produtor e aos órgãos oficiais de classificação e produção animal. A busca compreendeu publicações feitas no período 2000 - 2015, utilizando como filtro as palavras chave "classificação linear de bovinos" leiteiros, associado aos termos (i) regressão linear múltipla, (ii) regressão linear simples e (iii) lógica *fuzzy*, e obtendo os resultados conforme Tabela 1.

Tabela 1-Trabalhos relevantes

Palavra chave	Palavras Associadas	Relevantes
	-	3
Classificação linear de bovinos leiteiros	Regressão linear múltipla	3
	Lógica <i>fuzzy</i>	6
	Regras de associação	1
TOTAL		13

Fonte: Elaborado pelo autor (2015).

Em meio aos trabalhos pesquisados, destacaram-se treze considerados relevantes a pesquisa, sendo que:

- a) Teodoro et al. (2000), em seu artigo com o título Estudo de características do Sistema mamário e suas relações com a produção de leite em vacas da raça Gir, utilizaram o método de Máxima Verossimilhança restrita para verificar a produção em relação a estimativas de herdabilidade e seleção pela característica do sistema mamário;
- b) Esteves et al. (2004a) estudaram os efeitos ambientais sobre as características lineares de tipo em rebanhos bovinos da raça holandesa. Neste estudo os autores utilizaram o método dos mínimos quadrados nas análises de variância das características;
- c) Bressan (2004) fez utilização de dados semanais correspondentes às cotações, para geração de um modelo de previsão para compra e venda de contratos futuros de boi gordo em datas próximas ao vencimento, empregando modelos ARIMA, redes neurais artificiais e modelos lineares;
- d) Esteves et al. (2004b), em sua pesquisa intitulada Correlações genéticas e fenotípicas entre características de tipo e produção de leite em bovinos da raça holandesa, fizeram correlações genéticas e fenotípicas entre as 21 características lineares para verificar se as características possuíam alta correlação com a produção;
- e) Berry et al. (2005), em publicação intitulada Relações entre características de tipo e a longevidade em vacas leiteiras na Nova Zelândia, analisaram a produção leiteira em relação às características de tipo em uma região onde as vacas permanecem geralmente ao ar livre de forma contínua, alimentando-se de grama e principalmente de pastagem;
- f) Faria et al. (2008) compararam as estimativas de parâmetros genéticos, obtidas por meio de análises bayesianas uni-característica e bi-característica, através dos modelos animal linear e de limiar, considerando-se características categóricas morfológicas de bovinos da raça Nelore. De modo geral, os modelos de limiar e linear foram eficientes na estimação dos parâmetros para escores visuais. Deste modo, observa-se que o uso de ambos os modelos não influenciou a classificação dos animais, o que indica que os mesmos podem ser utilizados em programas de melhoramento genético;
- g) Falcão et al. (2008) analisaram 108.702 registros de produção de leite de vacas da raça holandesa, utilizando um modelo multicaracter para estimar

os componentes de variância utilizando amostragem de Gibbs. Os dados pertencentes aos estados de Minas Gerais, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul sugeriram a existência de interação genótipo-ambiente na produção leiteira;

- h) Perissinotto (2009), em publicação de abrangência internacional, analisou o conforto térmico de bovinos leiteiros em clima subtropical e mediterrâneo, fazendo uso da teoria dos conjuntos de *fuzzy*, com dados de estação meteorológica ao longo de 24 horas, por meio de mineração de dados, com objetivo de formar uma árvore de decisão para indução de regras, com intuito de estabelecer alguns parâmetros ideais de conforto aos bovinos da raça holandesa;
- i) Silva (2010), em sua tese, realizou uma análise de agrupamento para classificar e selecionar modelos não lineares de crescimento de bovinos, tendo como base o resultado de diferentes avaliadores de qualidade e ajuste, por meio de coeficiente de determinação (R^2), quadrado médio do erro, critério de informação de Akaike, critério de informação Bayesiano, erro quadrático médio de predição e coeficiente de determinação;
- j) Valloto e Neto (2010) relataram alterações quanto ao modelo de classificação para avaliação de bovinos da raça holandesa no Brasil, resultante de 21 características individuais, divididas em 5 seções para 23 características divididas em 4 seções, com o objetivo a atender as demandas mundiais de seleção linear para tipo leiteiro;
- k) Felipe (2012), em sua monografia, tratou da revisão dos principais pontos na teoria de desenvolvimento e implementação das RNA, para mostrar resultados da mesma na aplicação de melhoramento genético animal. Em sua conclusão, ressaltou que as RNAs têm grande potencial para o melhoramento genético com ressalvas à existência apenas de publicações recentes sobre o tema, contendo alguns pontos a serem explorados e otimizados;
- l) Ventura et al. (2012), dados de 19.240 animais, por meio de redes neurais artificiais, para predição de valores genéticos para peso a 205 dias, indicando que existe riscos ao uso de RNAs para avaliação genética desta característica; e

m) Venturini (2014) traçou o mapa da produção leiteira no país, por meio de dados do IBGE e ranqueia os estados maiores produtores de leite, onde o estado do Tocantins encontra-se como terceiro maior produtor da região norte.

A partir dos resultados obtidos por meio do embasamento teórico, averiguando publicações nos últimos quinze anos em relação à classificação linear de bovinos, observa-se que existe um leque amplo de oportunidades de pesquisas, até por se verificar a existência de um volume maior de estudos publicados há sete anos se comparado a publicações recentes. No Brasil, ainda ficam evidenciadas lacunas a serem exploradas em relação ao melhoramento de bovinos leiteiros por meio de classificação linear, em observância à grande concentração de altas produções em poucas regiões do país. (VENTURINI, 2014).

Outro ponto justificador para realização do estudo, pela análise dos textos previamente citados, é a evidência de mudanças no modelo de classificação, relatado por Valloto e Neto (2010), não sendo identificada outra publicação recente com dados de seleção posterior ao ano de 2010.

Os trabalhos analisados acima estão diretamente relacionados à análise do modelo de classificação em relação a questões de herdabilidade de características genéticas com objetivos ao ganho na produção ou perda da mesma (BRESSAN, 2004; ESTEVES et al., 2004; FARIA et al. 2008; VENTURA et al., 2012). Partindo deste princípio, torna-se justificável a presente proposta de estudo por se tratar da geração de uma proposta de modelo auxiliar a tomada de decisão de classificação de bovinos leiteiros da raça holandesa, observando-se proposta similar apenas no trabalho de Silva (2010), porém com objetivo de criação de modelos de crescimento em bovinos da raça Gir.

Partindo da relevância dos trabalhos analisados nos parágrafos anteriores, a presente proposta está embasada no desenvolvimento de um modelo de classificação e a implementação de um protótipo para que seja possível a coleta e processamento de dados de forma mais ágil e confiável garantindo assim uma ferramenta poderosa na ajuda para tomada de decisão. Como relata Pollak (1984 apud CAMPOS, 2012, p. 63), “já se pensava na utilização de aplicações e equipamentos com maiores capacidades de processamento”, uma vez que análises de grandes volumes de dados

esbarravam na baixa capacidade de processamento dos computadores da época, fato que pode ser amenizado nos dias atuais.

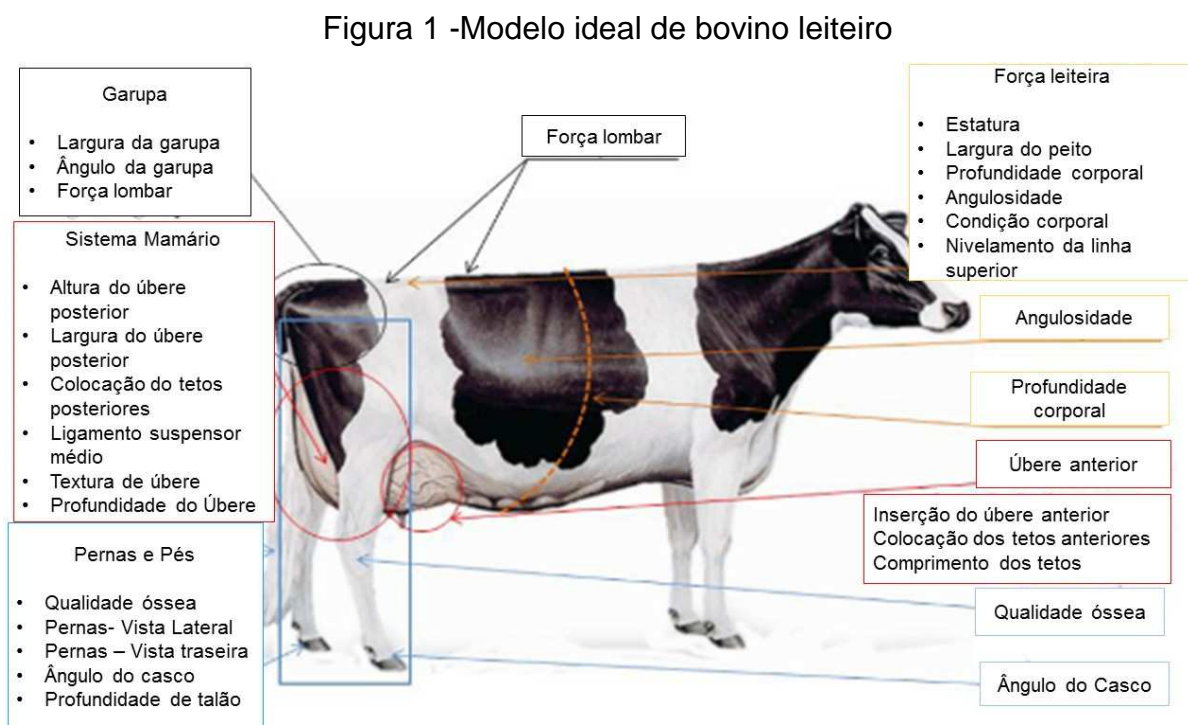
Com base no tema proposto e os objetivos a serem alcançados, esta pesquisa também visa a auxiliar a tomada de decisão no contexto da pecuária leiteira, através da tecnologia, e englobar áreas de conhecimentos distintos como Ciências Exatas e da Terra, Ciência da Computação, Sistema de Informação e Engenharia.

2 REVISÃO BIBLIOGRÁFICA

2.1 CARACTERÍSTICAS LINEARES DE TIPO EM REBANHOS DE BOVINOS LEITEIROS

A classificação linear de bovinos leiteiros é uma ferramenta de manejo e seleção que se baseia em um modelo de classificação criado originalmente, nos Estados Unidos, onde é utilizado desde 1929. No Brasil, a seleção por este sistema teve início na década de 60 e desde 1993 o modelo utilizado para classificação de tipo é o canadense. (VALLOTO, 2014). Deste período até o atual, este método de seleção veio sofrendo atualizações na quantidade de características a serem avaliadas e o peso de cada grupo destas, conforme figura 1.

No início da década de 80, criou-se uma escala de pontuação biológica linear que descreve os extremos biológicos para uma variedade de características morfológicas dos animais. Essa técnica teve por objetivo auxiliar na identificação de vacas com maior longevidade, através da expressão das melhores características com base em um modelo ideal “*True Type*”. (CAMPOS, 2012). A Figura 1, ilustra o modelo bovino leiteiro ideal e as características avaliadas.



Fonte: Valloto e Neto (2010).

A classificação linear para tipo ou registro seletivo, como também é chamado é o termo utilizado para definir a metodologia técnica de avaliação morfológica dos animais da raça holandesa. (VALLOTO; NETO, 2010). No Brasil, este sistema é representado pela Associação Brasileira de Criadores de Bovinos da Raça Holandesa (ABCBRH), que oferece este serviço de classificação dentre outros a seus associados.

Como mencionado anteriormente, o sistema de classificação linear adotado para a raça holandesa no Brasil é uma adaptação do modelo Canadense. Originalmente, neste sistema, as características lineares eram divididas em cinco seções que englobavam 21 (vinte e uma) características lineares (ESTEVES et al. , 2004a), conforme o Quadro 1.

Quadro 1 - Sistema de Classificação até 2010

	Categoria	Pontuação		
		1 Ponto	5 Pontos	9 Pontos
	Conformação/ Capacidade			
1	Estatura	≤ 1,27m	= 1,37m	≥1,47m
2	Nivelamento da Linha Superior	Baixa	Nivelada	Alta
3	Tamanho	≤430kg	=520kg	≥610kg
4	Largura Torácica	Estreita	Intermediária	Larga
5	Profundidade Corporal	Rasa	Intermediária	Profunda
6	Força Lombar	Fraca	Intermediária	Forte
	Garupa			
7	Nivelamento da Garupa	Alta	Nivelada (=5cm)	Baixa
8	Largura da Garupa	Estreita (≤8cm)	Média (=15cm)	Larga (≥20cm)
	Pés e Pernas			
9	Ângulo do Casco	Achinelado	Intermediária	Escarpado
10	Qualidade Óssea	Grosseiro	Intermediária	Plano e Limpo
11	Vista Lateral das Pernas e Pés	Retas	Intermediária	Curvas
	Sistema Mamário			
12	Inserção do úbere anterior	Fraca	Intermediária	Escarpado
13	Colocação das tetas anteriores	Afastadas	Centralizadas	Juntas
14	Cumprimento das tetas	Curtas	Intermediárias	Longas
15	Altura do úbere posterior	Baixo cm	Médio (=27cm)	Alto (36cm)
16	Largura do úbere posterior	Estreito (≤8cm)	Média (=15cm)	Largo (≥20cm)
17	Colocação das tetas posteriores	Afastadas	Centralizadas	Juntas
18	Profundidade do úbere	Profundo	Intermediário	Raso
19	Textura do úbere	Carnudo	Intermediária	Macio
20	Ligamento suspensório mediando	Pouco definida	Intermediária	Bem definida
	Caracterização Leiteira			
21	Angulosidade	Grosseira	Intermediária	Angulosa

Fonte : Adaptdo de Esteves (2004a).

O modelo, adotado pela ABCRH em meados de 2010 trouxe atualizações para atender a padrões internacionais de classificação. As principais modificações que ocorreram foram na diminuição das seções e adequações para cada seção, juntamente com a inserção de outras características. Nos moldes da classificação canadense, que englobava 21 características divididas em cinco seções, os animais poderiam receber pontuações que atingiriam no máximo 90 pontos, sendo considerados excelentes com este escore.

Atualmente, o sistema de classificação possui 23 (vinte e três) características a serem avaliadas (Quadro 2), mantendo os extremos biológicos para características individuais que continuam sendo os mesmos do quadro anterior, ou seja, no mínimo 1(um) e no máximo 9 (nove) pontos para características descritivas, sendo que, para algumas o escore de 5 (cinco) pontos é o ideal, em outras 7, porém, para grande maioria 9 (nove) pontos é o ideal.

Quadro 2 - Sistema de Classificação Atual

	Conformação/ Capacidade	1 Ponto	5 Pontos	9 Pontos
1	Estatuta	≤ 1,27m	= 1,37m	≥1,47 m
2	Nivelamento da linha superior	Baixa	Nivelada	Alta
		1 Ponto	3 Pontos	5 Pontos
3	Condição corporal	≤430 kg	=520 kg	≥610kg
		1 Ponto	5 Pontos	9 Pontos
4	Largura torácica	Estreita	Intermediária	Larga
5	Profundidade corporal	Rasa	Intermediária	Profunda
6	Angulosidade	Grosseira	Intermediária	Angulosa
	Garupa			
7	Ângulo da Garupa	Alta	Nivelada (=5 cm)	Baixa
8	Largura da Garupa	Estreita (≤8 cm)	Média (=15 cm)	Larga (≥20 cm)
9	Força de Lombo	Fraca	Intermediária	Forte
	Pés e Pernas			
10	Ângulo do Casco	Achinelado	Intermediária	Escarpado
11	Profundidade do Talão	Pequeno	Intermediário	Grande
12	Qualidade Óssea	Grosseiro	Intermediária	Plano e Limpo
13	Pernas Posteriores –vista lateral	Retas	Intermediária	Curvas
14	Pernas Posteriores – vista Posterior	Retas	Intermediária	Curvas
	Sistema Mamário			
15	Inserção do úbere anterior	Fraca	Intermediária	Escarpado
16	Colocação das tetas anteriores	Afastadas	Centralizadas	Juntas
17	Cumprimento das tetas	Curtas	Intermediárias	Longas
18	Profundidade do úbere	Profundo	Intermediário	Raso
19	Textura do úbere	Carnudo	Intermediária	Macio
20	Ligamento Médio	Pouco definida	Intermediária	Bem definida
21	Altura do úbere posterior	Baixo	Médio (=27)	Alto (36 cm)
22	Largura do úbere posterior	Estreito (≤8cm)	Média (=15)	Largo (≥20cm)
23	Colocação das tetas posteriores	Afastadas	Centralizadas	Juntas

Fonte: Adaptado de Valloto e Neto (2010).

De acordo com Valloto e Neto (2010), as alterações no sistema de avaliação das vacas, realizadas em 2010, tiveram por objetivo principal acompanhar as tendências mundiais na análise da conformação de vacas leiteiras, buscando as principais características para que a vaca produza muito leite, com maior vida produtiva (longevidade) e saúde. Os benefícios citados por Valloto e Neto (2010), em referência à seleção de bovinos por tipo estão relacionados à identificação dos melhores animais para o tipo da propriedade e conseqüentemente, os de pior conformação também.

Em contrapartida, Castro (1993) e Campos (2012) indicam que as razões para classificação do rebanho por tipo incluem melhorar o preço do animal em uma eventual venda e aumentar a qualidade do rebanho. Ao encontro desta afirmação, Valloto e Neto (2010) afirmam que esta classificação: identifica os pontos positivos e negativos de cada animal, direcionando o acasalamento dos mesmos, contribuindo diretamente com a valorização dos pedigrees e transferindo características genealógicas a seus descendentes e auxilia no descarte e evolução de animais pura por cruzada (PC) para puro por origem (PO).

Segundo Valloto (2014), nos Estados Unidos e Canadá é comum os criadores comprarem os animais somente avaliando a classificação para tipo. Se a avaliação for recente, muitos compram sem olhar o animal. No entanto, infelizmente, no Brasil, o número de animais classificados anualmente é muito pequeno se comparado a países como Estados Unidos, Holanda, França entre outros.

O embasamento teórico levantado sobre características lineares tem o intuito situar a presente pesquisa no atual sistema de classificação em uso para seleção de bovinos de tipo leiteiro no Brasil. Nesse sentido, através deste tenta-se obter uma visualização das variáveis para posteriormente decidir quais técnicas mais são indicadas para analisá-las. Técnicas que serão descritas nas próximas seções.

2.2 MINERAÇÃO DE DADOS

A mineração de dados (*Data Mining*) é uma linha de pesquisa pertencente à Ciência da Computação, que tem por objetivo analisar grandes bases de dados, procurando extrair informações gerando conhecimento através de análise de registros. Surgiu nos anos 90, derivada de conceitos pertencentes a diferentes áreas

como inteligência artificial, banco de dados, estatística e visualização de dados. (GONÇALVES, 2014a). A *data mining*, por assim dizer, veio com o intuito de resolver um problema gerado pelo surgimento de sistemas computacionais: com a evolução das tecnologias de hardware e software, a quantidade de dados aumentou significativamente; contudo, estes dados têm sido mantidos em fontes variadas, tais como banco de dados, páginas Web, arquivos convencionais e planilhas eletrônicas; e não necessariamente transformados em informação (útil para a tomada de decisão).

Com uma geração diária cada vez maior de dados, surgiu, segundo Camilo e Silva (2009), o questionamento do que se fazer com os dados, uma vez que as técnicas tradicionais de exploração não eram mais adequadas. Todavia, Macedo et al. (1996), apontam que a extração do conhecimento por meio da utilização de mineração de dados propõe novas técnicas e possibilita descobrir novas correlações contidas em bancos de dados, objetivando novos padrões e tendências entre as informações de uma empresa, podendo gerar conhecimento que estão ocultos em conjunto de dados. A mineração de dados é amplamente utilizada em vários segmentos sociais e produtivos, como exemplifica o Quadro 3.

Quadro 3 - Áreas de aplicação da Mineração de Dados

Áreas	Descrição
Retenção de Clientes	Identificação de perfis para determinados produtos, venda cruzada
Bancos	Identificação de padrões para gerenciamento de relacionamento com o cliente
Cartão de Crédito	Identificação de segmentos de mercado, identificação de padrões de rotatividade
Cobrança	Detecção de fraudes
Telemarketing	Acesso facilitado aos dados dos clientes
Eleitoral	Identificação de um perfil para possíveis votantes
Medicina	Indicação de diagnósticos mais precisos
Segurança	Detecção de atividades territoriais criminais
Biometria	Auxílio em pesquisas biométricas
RH	Identificação de competências em currículos
Tomada de Decisão	Filtro de informações relevantes, fornecendo indicadores de probabilidade

Fonte: Adaptado de Camilo e Silva, (2009).

Em relação a análise em bases de dados, torna-se relevante ressaltar que os mesmos são categorizados em quantitativos e qualitativos. Os quantitativos são representados por valores numéricos podendo assumir-se como discretos ou contínuos. Já os dados qualitativos contêm os valores nominais e ordinais, ou seja, as variáveis podem ser classificadas em quantitativa ou categórica, quantitativas assumem valores numérico e representam algum tipo de medição, os categóricos por outro lado, tomam uma categoria ou valores rotulados e colocam a observação do indivíduo em diversos grupos.

A mineração de dados visa coletar dados e fazer relações entre estes para, assim, gerar informação. A informação gerada pode ou não ser útil para o interessado, porém, pode chamar a atenção para outras áreas de interesse, às vezes esquecidas. Segundo Macedo e Matos (2010), o modelo de transformação dos dados em conhecimento consiste em um processamento manual de todas as informações por pessoas especializadas, produzindo relatórios que, então, deverão ser analisados. A mineração de dados trabalha com dois paradigmas de aprendizagem: supervisionada; e não-supervisionada. (SANCHES; MONARD, 2004). Na mineração de dados são aplicados algoritmos que objetivam identificar padrões nos dados originais, tais algoritmos se baseiam em técnicas estatísticas e inteligência artificial. Geralmente, os dados são volumosos e possuem grande variação entre os elementos, o que torna o processo de descoberta de informação trabalhoso, porém gratificante, pois resulta na descoberta de algo raro e preciso, daí o nome mineração de dados (SANTINI; TELLES, 2008).

Do ponto de vista processual, as fases que operacionalizam a mineração de dados são consideradas interativas. Rezende et al. (2011) relatam que atualmente diversos processos padronizam as fases de atividades da *data mining*. Para este referencial, é mostrado o Processo Padrão Interindústrias de Mineração de Dados ou CRISP-DM (*Cross-Industry Standard of Data Mining*), modelo que consiste em seis fases para o processo de conhecimento (Quadro 4).

Quadro 4 - Quadro resumo do processo CRISP-DM

<i>Fases</i>	<i>Descrição</i>
<i>Entendimento dos Negócios</i>	O foco desta etapa é entender qual o objetivo que se deve atingir com a mineração de dados
<i>Entendimento de dados</i>	Descrever de forma clara o problema; Identificar os dados relevantes para o problema em questão; Certifica-se de que variáveis relevantes para o projeto não são interdependentes.
<i>Preparação dos dados</i>	Devido origem dos dados, é comum os dados não estarem preparados para os métodos de mineração, envolve filtrar, combinar e preencher valores vazios.
<i>Modelagem</i>	Nesta fase são aplicadas as técnicas (algoritmos) de mineração. A escolha da técnica depende do objetivo a ser alcançado.
<i>Avaliação</i>	Fase crítica do processo de mineração, nesta etapa é necessária a presença de especialistas em dados, visando confiabilidade do modelo por meios de teste de validação.
<i>Distribuição</i>	Fase de apresentação dos resultados do modelo aos envolvidos.

Fonte : Adaptado de Rezende et al. (2011).

Vários recursos são necessários para promover a descoberta de conhecimento em bases de dados e a realização das atividades previstas no Quadro 4. Além disso, a exploração de dados por meio de técnicas de mineração para análise de respostas em um conjunto de dados, deve passar por etapas de preparação, trazendo os dados para a visão de negócios onde se pretende obter algum tipo de conhecimento. Segundo Silva et al. (2014), dependendo da técnica usada, as fases de preparação de dados para mineração são assim divididas:

1º Fase: Modelagem:

1. Questões para as quais se buscam respostas;
2. Informações de significância de utilização de cada uma das propriedades do problema;

2º Fase: Pré-processamento:

1. Seleção de dados a serem tratados (integrando, reduzindo e transformando-os com o objetivo, adicionar qualidade, tornando-os exequíveis);

3º Fase: Experimentos:

1. Experimentos da mineração de dados
2. Aplicação do algoritmo sobre os dados pré-processados com tratamento dos parâmetros e graus de confiança.

4º Fase: Pós- processamento

1. Considerada pós-processamento onde é compreendida a tarefa de refinamento e interpretação das regras extraídas;
2. Avaliação de especialistas do domínio para análise dos resultados.

Por fim, segundo Camilo e Silva (2009), o conhecimento dos dados é fundamental para a escolha do método adequado para extração de conhecimento de um conjunto de registros.

2.3 TÉCNICAS SELECIONADAS DE MINERAÇÃO DE DADOS PARA ESTA PESQUISA

Nesta seção são apresentadas as técnicas de mineração de dados selecionadas para a realização desta pesquisa. Inicialmente, é apresentado o conceito chamado regras de associação, com ênfase ao algoritmo apriori, introduzido por Agrawal et al. (1993). Para a técnica associada a esse conceito são descritas as principais medidas de interesse e definições para entendimento do problema.

2.3.1 Regras de Associação

A princípio, a mineração de dados surgiu como um método que se utilizava de estratégias automatizadas para extração de informações ocultas, tendo seus primeiros softwares para mineração de dados desenvolvidos em meados dos anos 90 ainda em ambiente acadêmico (GONÇALVES, 2014a). Uma dessas estratégias foi apresentada publicamente em 1993, sob a forma de extração de regras de associação, introduzido por Agrawal et al. (1993). Nesta época as regras eram

mineradas a partir de bases de dados transacionais (ou bases transacionais) que são formalizadas a seguir:

Seja $I = \{i_1, i_2, \dots, i_n\}$ um conjunto de literais chamados itens. Seja D um conjunto de transações, onde cada transação T é um conjunto de itens tal que $T \subseteq I$. Associado a cada transação existe um único identificador chamado de TID . Dizemos que a uma transação T contém X , um conjunto de alguns itens de I , se $X \subseteq T$. Uma Regra de Associação é uma implicação na forma $X \rightarrow Y$ onde $X \subset I$, $Y \subset I$ e $X \cap Y = \emptyset$. A regra $X \rightarrow Y$ mantida no conjunto de transações D com confiança c , se $c\%$ das transações em D que contém X também contém Y . A regra $X \rightarrow Y$ tem suporte s no conjunto de transações D , se $s\%$ das transações em D contém $X \cup Y$ [...].(AGRAWAL ; SRIKANT, 1994, p.2).

As regras de associação representam combinações de itens que ocorrem com determinada frequência em uma base de dados. A mineração de regras de associação torna-se atraente e popular por possuir uma forma fácil de ser compreendida, sendo descritas na expressão $A \rightarrow B$ podendo ser lidas como: o atributo A frequentemente implica no atributo B (AGRAWAL et al., 1993), onde, A é denominado antecedente e B consequente de uma regra de associação.

Dentre os algoritmos de mineração de regras de associação destaca-se o *apriori* (AGRAWAL et al., 1993) apontado pela IEEE na *International Conference on Data Mining (ICDM)* (WU, 2007) como o mais promissor algoritmo de geração de regras de associação e umas das abordagens mais populares em mineração de dados. A referência acima mencionada trata-se de uma versão seminal do algoritmo, desenvolvida para encontrar conjuntos de itens frequentes em uma base de dados. Geralmente utilizada para análise de cesta de compra (*market basket analysis*) por meio de uma base de dados que armazenam produtos comprados por clientes, as regras geradas são utilizadas para indicar que clientes que compram determinado produto também compram outros produtos no conjunto (GONÇALVES, 2005b). As regras, nesse caso, representam combinações de itens que ocorrem com determinada frequência em uma base de dados que poderiam gerar o seguinte exemplo: $\{\text{feijão}\} \{\text{couve}\} \rightarrow \{\text{linguiça}\}$, esta regra é usada para indicar que clientes que compram feijão e couve, tendem a também comprar linguiça. (GONÇALVES, 2005b).

Do ponto de vista da qualidade da informação gerada, o objetivo da mineração das regras de associação é pesquisar relacionamentos relevantes entre os itens de uma base de dados, permitindo que as regras geradas possam ser indicadoras

valiosos na tomada de decisão (XARVIER, 2010). Mais amplamente, para verificar quão forte é uma regra, introduziu-se maneiras de se avaliar a qualidade da mesma por meio da utilização de medidas de interesse, medidas estas, que podem ser divididas em dois tipos (GONÇALVES, 2005b, p. 1):

Medidas de interesse objetivas: são medidas que identificam estatisticamente a força de uma regra. Medidas de interesse subjetivas: são medidas que identificam a força da regra por meio de opinião de uma analista [...].

Para avaliar a relevância das regras, Agrawal et al. (1993) propuseram um modelo em que as regras devem atender a um suporte e confiança mínimo especificadas pelo decisor. Gonçalves (2005b, p.27) define modelo suporte/confiança da seguinte forma:

O suporte de um conjunto de itens Z , $Sup(Z)$, representa a porcentagem de transações da base de dados que contêm os itens de Z . O suporte de uma regra de associação $A \Rightarrow B$, $Sup(A \Rightarrow B)$, é dado por $Sup(A \cup B)$. Já a confiança desta regra, $Conf(A \Rightarrow B)$, representa, dentre as transações que contêm A , a porcentagem de transações que também contêm B , ou seja, $Conf(A \Rightarrow B) = Sup(A \cup B) \div Sup(A)$ [...].

Nesta equação, $Sup(A \cup B)$ representa o suporte real do conjunto de itens $A \cup B$, enquanto $Sup(A) \times Sup(B)$ é o suporte esperado do conjunto $A \cup B$. Conforme Gonçalves (2005b, p. 5)

Suporte Esperado para um Conjunto de itens. Seja D uma base de dados de transações definida sobre um conjunto de itens I . Sejam $A \subset I$ e $B \subset I$ dois conjuntos não vazios de itens, onde $A \cap B = \emptyset$. O suporte esperado ($SupEsp$) do conjunto $A \cup B$ é obtido por: $SupEsp(A \cup B) = Sup(A) \times Sup(B)$

Mais especificamente, encontrar todas as regras que possuam suporte e confiança maiores ou iguais, respectivamente a suporte mínimo (chamado de *minsup*) e confiança mínima (chamado de *minconf*), especificada pelo usuário, chamado de modelo suporte /confiança.

Apesar de ser um modelo comumente utilizado em extração de regra de associação, o modelo suporte/confiança tem recebido muitas críticas ao longo dos anos motivadas pelo grande número de regras geradas dificultando o processo de análise por parte do usuário e, em alguns casos, as regras não parecem expressar

relacionamentos válidos mesmo quando os valores de confiança são muito altos. Este fato, segundo Gonçalves (2005b), ocorre uma vez que a medida de confiança não considera a dependência entre os itens de dados, podendo gerar um grande número de relacionamentos falsos ou ilusórios.

A aferição do grau de importância de uma regra se dá por meio de indicadores e da análise de um especialista. Uma vez feita a mineração de regras, é realizada a análise da saída dos resultados com o intuito de verificar se a mesma é válida (forte) para a resolução do problema. Esta, consiste em identificar as regras que são demasiadamente óbvias, regras independentes e regras com dependência negativa deixando apenas as que possuem dependência positivas entre os itens de dados. Conforme descreve Gonçalves (2005b, p.30) as definições para os tipos de regras encontradas na mineração de associações podem ser descritas como:

Independência entre dados:

Seja D uma base de dados de transações definida sobre um conjunto de itens I . Sejam $A \subset I$ e $B \subset I$ dois conjuntos não vazios de itens, onde $A \cap B = \emptyset$. Os conjuntos de itens A e B são independentes se:
 $Sup(A \cup B) = SupEsp(A \cup B)$

Dependência Negativa entre dados:

Seja D uma base de dados de transações definida sobre um conjunto de itens I . Sejam $A \subset I$ e $B \subset I$ dois conjuntos não vazios de itens, onde $A \cap B = \emptyset$. Os conjuntos de itens A e B possuem dependência negativa se:
 $Sup(A \cup B) < SupEsp(A \cup B)$.

Dependência positiva entre itens:

Seja D uma base de dados de transações definida sobre um conjunto de itens I . Sejam $A \subset I$ e $B \subset I$ dois conjuntos não vazios de itens, onde $A \cap B = \emptyset$. Os conjuntos de itens A e B possuem dependência positiva se:
 $Sup(A \cup B) > SupEsp(A \cup B)$.
 [...].

As medidas citadas anteriormente, assim como o suporte e confiança, são medidas de interesse objetivas que podem ser definidas como índices estatísticos para selecionar regras, dentre as muitas que podem ser geradas por um algoritmo de mineração.

Para contornar as fragilidades dos indicadores anteriormente mencionados, outras medidas de interesse objetivas foram propostas. Podem ser destacadas o *lift* (BRIN; MOTOWANI; ULLMAN, 1997) e o *Rule Interest* (RI), introduzido em Piatetsky-Shapiro (1991), que são medidas comumente usadas para medir o grau de interesse de uma regra de associação.

O interesse *lift* é uma medida de avaliação, também conhecida como interest. Esta permite eliminar regras com confiança elevada, mas com pouco interesse. É utilizada para encontrar dependências, pois indica quanto mais frequente torna-se B quando A ocorre. Varia entre 0 e ∞ . (SILVA et al. 2014). Conforme (BRIN; MOTOWANI; ULLMAN, 1997), o valor do *lift* para $A \rightarrow B$ é calculado pela expressão (2):

$$Lift(A \rightarrow B) = \frac{Conf(A \rightarrow B)}{Sup(B)} \quad (2)$$

Se $Lift(A \rightarrow B) = 1$, então A e B são independentes, portanto a ocorrência de A não aumenta e nem diminui a probabilidade de B ocorrer. Se $Lift(A \rightarrow B) > 1$, então A e B são positivamente dependentes, assim, a ocorrência de A aumenta a probabilidade de B ocorrer. Por fim, se $Lift(A \rightarrow B) < 1$, então A e B são negativamente dependentes, ou seja, a ocorrência de A diminui a probabilidade de B ocorrer.

O índice *Rule Interest* (RI), também conhecido na literatura como PS (Piatetsky-Shapiro's), é outra medida de interesse que pode ser utilizada para avaliar as dependências. Esta medida calcula a diferença entre os valores de suporte real e suporte esperado de uma regra de associação (GOLÇALVES, 2005b). O valor do RI para $A \rightarrow B$ é calculado pela expressão (3):

$$RI(A \Rightarrow B) = Sup(A \Rightarrow B) - SupEsp(A \Rightarrow B) \quad (3)$$

Se $RI(A \rightarrow B) = 0$ significa dizer que A e B são independentes, em outras palavras diz-se que a ocorrência de A não influencia na probabilidade de B ocorrer. Se $RI(A \rightarrow B) > 0$, então A e B são positivamente dependentes, assim, a ocorrência de A influencia aumentando a probabilidade de B ocorrer. Senão, A e B são negativamente dependentes, ou seja, a ocorrência de A diminui a probabilidade de B ocorrer (GONÇALVES; PLASTINO, 2004).

Gonçalves (2005b, p.33) destaca que:

É importante observar que o *lift* consegue destacar com maior facilidade a dependência positiva entre conjuntos de itens que possuem suporte baixo. Já a medida de RI é especialmente útil para destacar a dependência positiva entre conjuntos de itens que possuam suporte médio ou alto.[...].

As medidas de interesse objetivas suporte e confiança, conforme relata Gonçalves (2005b), demonstram deficiência de não considerar questões de dependência entre os itens. Em sua pesquisa, o autor sugere a utilização conjunta das medidas de interesse suporte/confiança, *lift* e *RI* como forma de diminuir a chance de mineração de regras óbvias e irrelevantes, enriquecendo o poder de entendimento a respeito das associações geradas. O *lift* e *RI* são as duas das medidas de interesse mais comumente usadas para mensurar as dependências entre itens de uma regra de associação, mas vale destacar que, além deste listado anteriormente, diversas outras medidas de interesse objetivas podem ser identificadas, como: convicção, coeficiente de correlação, resultado do teste *chi-quadrado* para independência, Medida-J, entre outras. (GONÇALVES,2005b).

Devido a sua grande aplicabilidade, as regras de associação encontram-se entre um dos mais importantes tipos de conhecimento que pode ser minerado de uma base de dados. Isso de fato se deve pela automatização de rotinas de mineração de dados por meio de ferramentas especializados para tal atividade.

2.4 WEKA – WAIKATO ENVIRONMENT FOR KNOWLEDGE ANALYSIS

O Weka é uma ferramenta desenvolvida na linguagem de programação Java e distribuído sob os termos da licença GNU (*General Public License*). É um sistema multiplataforma podendo rodar na plataforma Windows, Macintosh e distribuições Linux. Foi projetado para que o usuário possa rapidamente experimentar métodos existentes em novos conjuntos de dados de maneira flexível fornecendo suporte extensivo para todo o processo de mineração de dados experimental.

O sistema possui um vasto conjunto de algoritmos de aprendizagem que são disponibilizados por meio de uma plataforma amigável e intuitiva. A sua interface possibilita que o usuário possa comparar vários métodos diferentes facilitando a escolha do que é mais adequado para resolução do seu problema, incluído uma ampla gama de ferramentas de pré e pós-processamento. Possui métodos para os principais problemas de mineração de dados como: regressão, classificação, clusterização, mineração de regras de associação e seleção de atributos.

Embora a ferramenta possibilite uma quantidade considerável de métodos para resolução de problemas em mineração, a preparação dos dados consome normalmente a maior parte do esforço. A experiência mostra que naturalmente as bases de dados de onde se pretende extrair alguma forma de conhecimento não estão condizentes com o problema que se pretende resolver, tornando praticamente impossível obter o sucesso na primeira vez em que se utiliza uma base de dados “crua” tornando relevante a utilização de alguns passos para a preparação dos mesmos para extração de conhecimento.

A preparação dos dados envolve ainda a transformação de um arquivo de entrada para que os mesmos possam ser reconhecidos pelo aplicativo. A ferramenta trabalha a entrada por meio de um arquivo no formato ARFF, que corresponde a um arquivo de texto contendo um conjunto de informações precedido por uma sessão *Header* (cabeçalho) que contém os comandos *@relation* e *@attribute*. Geralmente as linhas iniciais do arquivo são antecidas de % (símbolo de porcentagem) que representa os comentários do mesmo, ou seja, onde existem estes símbolos significa dizer que esta linha não será computada pela ferramenta como um dado, em seguida é nomeada a relação com a sintaxe *@relation<nome_da_relação>*. Os atributos são descritos em forma de uma tabela relacional, onde cada atributo é identificado pela palavra reservada *@attribute* seguido do nome e do tipo de dado a ser investigado, para declaração de mais de um atributo são declarados com uma sequência deste comando, um por linha com a sintaxe *@attribute < nome_do_atributo > < tipo de dado >* onde o nome do atributo não deve ser iniciado por caracteres especiais e caso haja espaços dentro do nome estes devem ficar entre aspas.

Quanto aos tipos de dados, Xavier (2010) relata a existência atualmente de 4(quatro) tipos suportados pelo Weka, sendo, para valores numéricos os tipos *integer* e *real* e para nominais *string* e *date*. Após a definição dos atributos, são relacionadas as entidades representadas a partir de então por um bloco definido por *@data*. As instancias são relacionadas uma por linha com dados para cada atributo separados por vírgula e na ordem que foram declaradas na seção *Header* caso o dado seja desconhecido ou esteja ausente deve ser substituído por um “?” ponto de interrogação para evitar erros por parte do algoritmo aplicado. A Figura 2 a seguir mostra um exemplo de um arquivo arff.

Figura 2 - Exemplo de arquivo no formato .arff

```
@relation "Transacoes"

@attribute I1 {y, n}
@attribute I2 {y, n}
@attribute I3 {y, n}
@attribute I4 {y, n}
@attribute I5 {y, n}
@attribute I6 {y, n}

@data
y,y,?,?,y,?
?,y,?,y,?,?
```

Fonte: Gonçalves (2014a).

Uma característica importante da ferramenta é a sua simplicidade, sendo possível fazer a utilização de seus algoritmos por meio de uma interface intuitiva com alto nível de usabilidade tornando seu uso agradável. Para Gonçalves (2014a), encontra-se como característica negativa o fato dos dados a serem minerados precisarem ser estruturados em um formato não muito prático, uma vez que é preciso especificar todos os itens do domínio no *Header* e indicar os itens ausentes.

2.5 LÓGICA FUZZY

Com os avanços tecnológicos e, conseqüentemente, com o aumento da complexidade de controle de operações atreladas a estes avanços, ações que antes eram realizadas de forma manual puderam ser automatizadas, porém, a complexidade trazida pela automatização de processos propiciou a busca de novas técnicas de controlar estas ações que já não eram mais possíveis de serem feitas apenas com a lógica clássica.

A lógica clássica não possibilita que sistemas possam tomar decisões de maneira mais aproximada com o raciocínio humano ou automatizar processos de maneira semelhante. Houve então a necessidade de buscar novos sistemas de controle, como a utilização de inteligência Artificial (IA), e o desenvolvimento de novas lógicas, como a lógica fuzzy. Fazendo um paralelo, Gomes e Rodrigues (2016 p.5)

relatam que muitas vezes o ser humano se depara com informações e conceitos aos quais não podem ser representados por valores numéricos e precisos, tais como:

O que fazer quando um processo de decisão inerentemente humano deve ser automatizado? Como modelar e implementar uma aplicação de forma que ela responda a estímulos de natureza imprecisa? Que ferramentas e técnicas podemos utilizar para simular computacionalmente o comportamento de um motorista de carro, ou mesmo de um operador de bolsa de valores?[...].

Partindo deste princípio, a lógica fuzzy, também conhecida como lógica nebulosa ou difusa, tenta responder questões em que apenas parâmetros como verdadeiro e falso não conseguem responder. É uma técnica da área da inteligência computacional que simula o comportamento do raciocínio humano e permite representar modelos que tenham certo grau de incerteza ou imprecisão.

Gomes e Rodrigues (2016, p.1) destaca o paradoxo Sorites como o primeiro raciocínio a questionar a imprecisão do pensamento humano:

[...]Este raciocínio foi empregado no século IV a.c. pelo filósofo Ebulides de Mileto, na proposição que ficou conhecida como Paradoxo Sorites. Ele questiona: “quando um monte de areia deixa de ser um monte de areia, caso retiremos um grão de cada vez? ”. Em sua proposição, ele define que se n grãos de areia são um monte, também o seriam $(n - 1)$ grãos. Ou seja, um único grão de areia não deveria fazer diferença nesta classificação. O mesmo se diz do cenário oposto: um grão de areia não faz um monte, e adicionando mais um único grão, também não faria diferença. Mas isso significa, por indução, que nunca teremos um monte de areia, por mais que juntemos os grãos um a um[...].

Para Rignel, Chenci e Lucas (2011, p. 3), os princípios de lógica fuzzy foram desenvolvidos primeiramente por Jan Lukasiewicz (1878-1956), que em 1920 desenvolveu e introduziu conjuntos com grau de pertinência que combinados aos conceitos da lógica clássica, desenvolvida por Aristóteles, deu embasamento suficiente para que, na década de 60, Lofti Asker Zadeh, chegasse a ser o primeiro autor de uma publicação sobre lógica fuzzy. Zadeh foi professor de engenharia elétrica e ciência da computação da Universidade de Berkeley na Califórnia. Em 1965, lançou o artigo “Fuzzy sets”, introduzindo o assunto no meio acadêmico, sendo muito

questionada e criticada inicialmente, porém, muito utilizada com o passar do tempo. (POSSELT et al. 2015).

A lógica *fuzzy* diferentemente da lógica clássica booleana que se detém apenas a valores binários ou bivalentes, verdadeiro e falso ou 0 e 1, baseia-se em conjuntos difusos e operações difusas que possibilita a modelagem realista e flexível de sistema. Posselt et al. (2015) explicam que esta lógica foi desenvolvida para determinar quanto um elemento pertence ou não a um determinado conjunto, com uso de graus de pertinência que são valores reais no intervalo de [0;1] suportando o raciocínio não exato e as quantidades imprecisas.

Quanto função de pertinência, na teoria dos conjuntos clássica a pertinência de um elemento x a um conjunto A é indicado pela relação $x \in A$ e é expressa por função de pertinência bivalente $\mu_A(x)$, ou seja, verdadeiro ou falso. Na lógica *fuzzy*, emprega-se o grau de pertinência ou pertinência parcial de elementos a um determinado conjunto, ou seja, uma pertinência de 0,5 pode representar meia verdade onde 0,9 e 0,1, representam quase verdade e quase falsidade. (RIGNEL; CHENCI; LUCAS, 2011; FELICIANO, 2012).

Cada elemento fuzzy possui um grau de pertinência também chamado de grau de inclusão, podendo haver uma transição gradual do elemento entre um conjunto e outro, podendo assumir diversos valores, não apenas, verdadeiro ou falso, conforme ilustra a Figura 3.

Figura 3 - Representação de Pertinência



Fonte: Google Imagens (2016).

A necessidade de lidar com as incertezas e com a complexidade da natureza dos problemas em diversas situações do cotidiano é uma atividade comum do ser humano. Por exemplo, quando é necessário classificar determinado indivíduo por sua faixa etária, produção ou pontuação física. Para melhor ilustrar tomar-se-á

como base o quadro final de classificação de bovinos leiteiros, conforme apresentado no Quadro 5:

Quadro 5 – Classificações possíveis

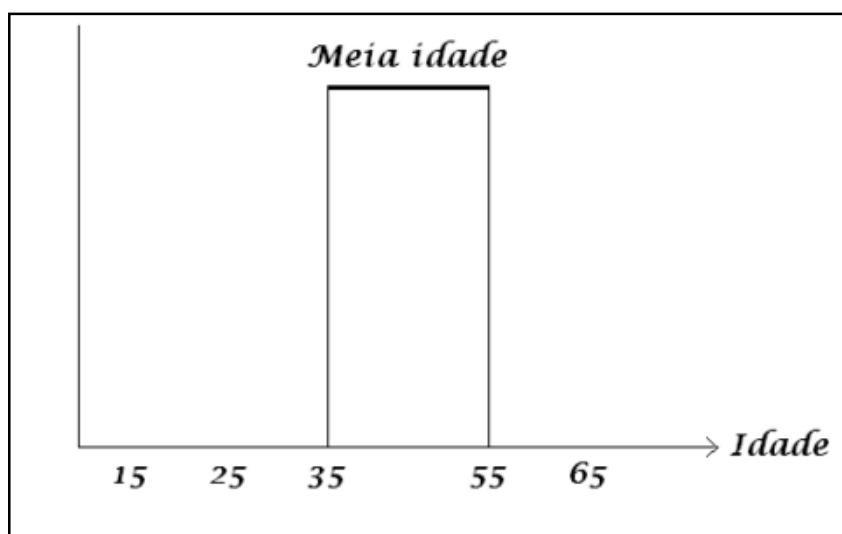
Classes	Pontuação Final
FRACA	50 a 64 pontos
REGULAR	65 a 74 pontos
BOA	75 a 79 pontos
BOA PARA MAIS	80 a 84 pontos
MUITO BOA	85 a 89 pontos
EXCELENTE	90 a 97 pontos

Fonte: Valloto e Neto (2010).

Utilizando a lógica tradicional uma vaca só mudaria da classe Fraca para a classe Regular quando sua pontuação fosse igual ou maior a 65 (sessenta e cinco) pontos. Por ser uma classificação subjetiva baseada além das características biológicas do animal, também, na experiência do avaliador por 1(um) ponto o animal poderia mudar de categoria sem que houvesse qualquer tipo de mudança física no mesmo.

Situação semelhante ocorre quando classificamos indivíduos por idade, Rignel, Chenci e Lucas (2011) levam em consideração um exemplo de quando se considera um período de meia idade, que começa em 35 anos e termina em 55 anos, conforme ilustra Figura 4.

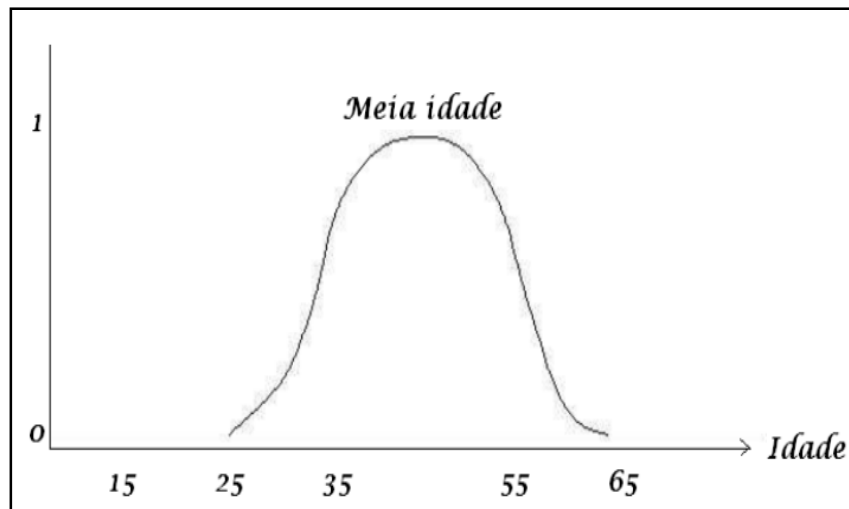
Figura 4 - Definição de meia-idade em conjuntos convencionais



Fonte: Rignel, Chenci e Lucas (2011).

Utilizando-se da notação dos conjuntos fuzzy, este mesmo cenário poderia ser representado conforme ilustrado na Figura 5 destacando -se que as fronteiras entre os elementos não estão bem definidas, por isso, ocorre uma transição gradativa dos graus de pertinência, à medida que o indivíduo de meia-idade avança. (GOMES ; RODRIGUES, 2016).

Figura 5 - Definição de meia-idade com conjunto fuzzy



Fonte: Rignel, Chenci e Lucas (2011).

O modelo de representação apresentado anteriormente se aproxima da forma com o entendimento humano compreende as decisões presentes no dia-a-dia, imagina-se que não há forma definida nos conjuntos, podendo assim capturar informações vagas, descritas em linguagem natural e converte-las para o formato numérico sem precisar ater-se a conhecimentos complexos de matemática e suas implicações lógicas adotando a estratégia do tipo SE <condição> ENTÃO <ação> para modelagem e controle de ações. (FELICIANO, 2012; GOMES & RODRIGUES, 2016).

Segundo Rignel, Chenci e Lucas (2011), operações com conjuntos nebulosos podem ser assim apresentadas:

- O conjunto *fuzzy* A é subconjunto de um conjunto *fuzzy* B se o grau de pertinência do conjunto universo U no conjunto A é menor ou igual que seu grau de pertinência no conjunto B; ou seja, para todo $x \in U$, $\mu_A(x) \leq \mu_B(x)$ e indica-se $A \subseteq B$.

- Os conjuntos *fuzzy* A e B **são iguais** se $\mu_A(x) = \mu_B(x)$ para todo elemento $x \in U$ e indica-se $A=B$.
- Os conjuntos *fuzzy* A e B **são diferentes** se $\mu_A(x) \neq \mu_B(x)$ para no mínimo um $x \in U$ e indica-se $A \neq B$.
- O conjunto *fuzzy* A é um **subconjunto próprio** do conjunto *fuzzy* B quando A é subconjunto de B e $A \neq B$, isto é, $\mu_A(x) \leq \mu_B(x)$ para todo $x \in U$ e $\mu_A(x) < \mu_B(x)$ para no mínimo um $x \in U$ e indica-se $A \subset B$ se somente se $A \subseteq B$ e $A \neq B$.
- O **complemento** de um conjunto *fuzzy* A em relação ao conjunto universo U é indicado por A^c e a função de pertinência é definida como: $\mu_{A^c}(x) = 1 - \mu_A(x)$ para todo $x \in U$.
- A **união** de dois conjuntos *fuzzy* A e B é um conjunto *fuzzy* $A \cup B$ tal que para todo $x \in U$ $\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)]$.
- A **intersecção** de dois conjuntos *fuzzy* A e B é um conjunto *fuzzy* $A \cap B$ tal que para todo $x \in U$ $\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)]$.

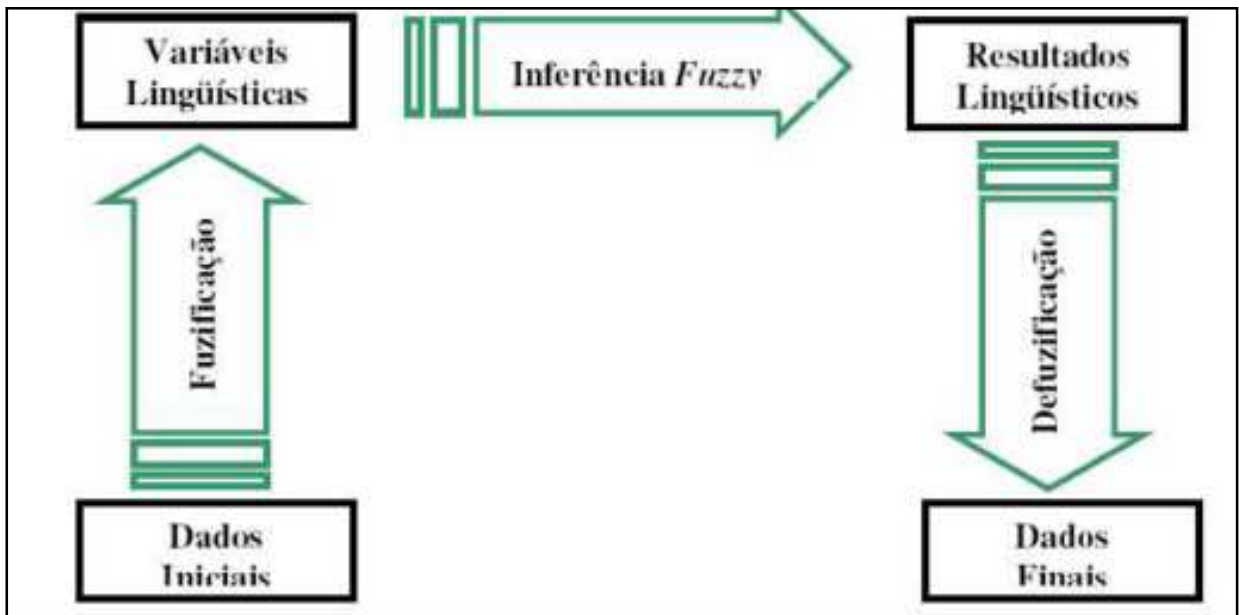
A lógica *fuzzy* possui características especiais na forma de representar o manuseio de informações imprecisas, provendo um método de traduzir expressões verbais, vagas, imprecisas e qualitativas, comuns na comunicação humana. Por meio de variáveis linguísticas que são em outras palavras nomes de conjuntos *fuzzy*, tem por função fornecer uma maneira sistemática de aproximação de fenômenos complexos ou mal definidos. Usando como exemplo o peso de uma pessoa que pode ser uma variável linguística assumindo valores baixo, médio e alto. (RIGNEL; CHENCI; LUCAS 2011)

De acordo com Gonçalves (2005b), em 1970, Mamdani propôs um método de inferência *fuzzy* que foi por muitos anos um padrão para utilização dos conceitos da lógica *fuzzy* em sistemas de controle, uma regra típica do modelo é mostrada a seguir:

IF Temperatura **IS** AumentandoMuito **AND** Pressão **IS** MuitoAlta , **THEN**
PerigoDeExplosão **IS** Grande

Nessa mesma linha, o raciocínio fuzzy é composto por três etapas, ou processos, que são: a fuzificação, a inferência e a defuzificação (VIDAL; LANDIN; COSTA, 2012), como ilustrado na Figura 6.

Figura 6 – Estrutura do sistema fuzzy



Fonte: Vidal; Landin e Costa (2012).

A primeira etapa chamada fuzificação consiste em transformar os dados iniciais de entrada em suas respectivas variáveis linguísticas. A inferência é a fase em que é tomada a decisão conforme regras similares a anteriormente apresentadas e onde são geradas as regras que irão reger o processo de inferência. A terceira e última fase, onde, a variável linguística de saída inferida pelas regras fuzzy traduzida num número, em uma única saída.

2.7 SÍNTESE TEÓRICA

3 METODOLOGIA

Este capítulo trata do desenvolvimento do trabalho de pesquisa e as formas que serão utilizadas para que se alcancem os objetivos esperados em consonância

com que se exige em um trabalho de pesquisa, o mesmo está dividido em Método de pesquisa, Método de Trabalho, e subdivisões inerentes etapas que o estudo exige.

3.1 MÉTODO DE PESQUISA

Esta pesquisa quanto ao método é classificada como pesquisa operacional. Quanto aos objetivos a serem alcançados é aplicada e de caráter exploratório. Possui natureza qualiquantitativa, uma vez que pesquisa operacional está voltada principalmente para manipulação de variáveis quantitativas, mas julgamentos qualitativos (subjetivos) serão necessários para sua realização. Os procedimentos de pesquisa envolvem entendimento do contexto de classificação de bovinos por intermédio de levantamento bibliográfico e análise de dados de classificação real no ano de 2015, fornecida pela Associação Paranaense de Criadores de Bovinos da Raça Holandesa. O estudo também será balizado por entrevistas que serão aplicadas a professores e técnicos especializados do setor de zootecnia III (bovinocultura) do IFTO - Campus Araguatins, setor este que servirá como laboratório para as fases de teste e implementação do modelo.

O método da pesquisa deve apresentar os procedimentos empregados para realizar cientificamente o estudo, ou seja, o propósito do trabalho, o método de delineamento, as técnicas de coleta de dados e a técnica de análise.

3.2 MÉTODO DE TRABALHO

Esta pesquisa busca um modelo de classificação que tenta avaliar a capacidade de uma ou mais variáveis usadas como sinalizadoras de um determinado padrão de animal, classificando-o como bom ou ruim para atividade leiteira. Para cumprir os objetivos, utilizam-se critérios estatísticos, matemáticos e computacionais na construção deste modelo sobre as variáveis escolhidas. Por meio de técnicas de mineração de dados e inferência de regras fuzzy, pretende identificar as variáveis que melhor explicam o resultado final de classificação e selecioná-las, posicionando, o bovino entre as categorias apresentadas no Quadro 5, por meio da lógica nebulosa.

O Quadro 6 tem por objetivo descrever de forma sucinta as técnicas e conceitos utilizados nesta pesquisa bem como citar alguns autores importantes para

desenvolvimento deste estudo. Este traz, em síntese, os passos a serem desenvolvidos baseados nos conteúdos apresentados no capítulo do referencial teórico.

A primeira macro-etapa diz respeito ao entendimento do modelo de classificação vigente, visando compreensão da metodologia e técnicas de classificação para melhor atacar o problema. Este conhecimento viabilizou a identificação dos resultados finais de classificação como sendo baseado na combinação de 22 (vinte e duas) variáveis que representam características produtivas do animal e 1 (uma) que indica a sua condição corporal. Outra possibilidade dentro desta macro-etapa, foi em relação ao resultado final de classificação pois, observa-se que a classificação do animal se baseia em lógica booleana [0,1] categorizando o animal em uma das seis classes possíveis conforme quadro 5. Partindo deste princípio, a macro-etapa dois almeja a possibilidade de aplicação de mineração de dados, mais especificamente a mineração de regras de associação para descoberta das características que frequentemente existem em função de outras. Para a mineração de regras de associação, segundo Rezende et al. (2011), é necessário que haja uma preparação dos dados. A preparação ou pré-processamento dos dados consiste na filtragem dos registros considerados interessantes para o modelo, transformando as informações ou registros em um formato possível de ser lido como uma regra, no caso deste estudo em um arquivo “. arff ” compatível com Weka, para que seja feita a indução das regras de associação por meio do algoritmo apriori, ambos descritos neste referencial.

A macro-etapa três, define os procedimentos adotados para organização dos dados para geração de classificação dos animais baseada em lógica *fuzzy*. Semelhante ao pré-processamento da mineração de regras de associação, tratando os dados, criando variáveis linguísticas, definindo universo de cada uma e gerando as regras.

Quadro 6 - Síntese teórica

Macro-etapa	Etapa	Conceitos
<i>Entendimento do modelo atualmente em uso</i>	1. Realizar pesquisa em bases de periódicos em busca de referências sobre melhoramento de bovinos leiteiros e classificação linear.	Atualmente o sistema de classificação possui 23 características a serem avaliadas mantendo os extremos biológicos para características individuais que continuam sendo os mesmos do quadro anterior, ou seja, no mínimo 1(um) e no máximo 9 (nove) pontos para características descritivas, sendo que, para algumas o escore de 5 (cinco) pontos é o ideal, em outras 7, porém para grande maioria 9 (nove) pontos é o ideal. (VALLOTO E NETO,2010)
	2. Estudo de técnicas e ferramentas para que seja possível gerar um modelo de classificação.	
	3. Obtenção de dados de classificação de bovinos leiteiros com modelo atual de classificação.	
<i>Organização dos dados para mineração de regras de associação</i>	1. Preparação dos dados fazendo filtragem dos dados eliminando os registros incompletos e preencher os valores vazios. 2. Modelagem escolha e aplicação da técnica e algoritmo apriori para gerar regras de associação.	Do ponto de vista processual, as fases que operacionalizam a mineração de dados são consideradas interativas. Rezende et al. (2011) relatam que atualmente diversos processos padronizam as fases de atividades da <i>data mining</i> .
<i>Organização dos dados para utilização de lógica fuzzy</i>	1. Transformar os dados de classificação em variáveis linguísticas. 2. Criação das regras de classificação linear. 3. Obtenção de resultados linguísticos.	A lógica <i>fuzzy</i> possui características especiais na forma de representar o manuseio de informações imprecisas, provendo um método de traduzir expressões verbais, vagas, imprecisas e qualitativas, comuns na comunicação humana. Por meio de variáveis linguísticas que são em outras palavras nomes de conjuntos <i>fuzzy</i> , tem por função fornecer uma maneira sistemática de aproximação de fenômenos complexos ou mal definidos. Usando como exemplo o peso de uma pessoa que pode ser uma variável linguística assumindo valores baixo, médio e alto. (RIGNEL; CHENCI; LUCAS 2011)
		[...] o raciocínio fuzzy é composto por três etapas, ou processos, que são: a fuzificação, a inferência e a defuzzificação (VIDAL; LANDIN; COSTA, 2012)
<i>Análise de resultados</i>	1. Identificação de regras de associação que indiquem quais características corporais do bovino ocorrem com mais frequência em detrimento de outras utilizando a medida de interesse objetiva <i>lift</i> da regra.	O <i>lift</i> e <i>RI</i> são as duas das medidas de interesses mais comumente usadas para mensurar as dependências entre itens de uma regra de associação, mas vale destacar que, além destes listados anteriormente, diversas outras medidas de interesse objetivas podem ser identificadas, como: convicção, coeficiente de correlação, resultado do teste <i>chi-quadrado</i> para independência, Medida-J, entre outras. (GONÇALVES,2005b).
	2. Identificar a quantas classes o mesmo animal pertence por meio de aplicação da lógica fuzzy.	Posselt et al. (2015) explicam que esta lógica foi desenvolvida para determinar quanto um elemento pertence ou não a um determinado conjunto, com uso de graus de pertinência que são valores reais no intervalo de [0;1] suportando o raciocínio não exato e as quantidades imprecisas.

Fonte: Elaborada pelo autor (2016).

Na macro-etapa quatro, são almejados os resultados por meio da relação dos resultados obtidos nas macro-etapas 1, 2 e 3. Busca-se uma relação frequente entre as variáveis, que serão obtidas por mineração de regras de associação indicando por exemplo {estatura = alta} => {ângulo_garupa = intermediário}, ou seja, significaria dizer que um animal que possua a característica estatura = alta possivelmente terá a característica ângulo da garupa intermediária. Ainda nesta macro-etapa, pretende-se classificar o animal usando regras de inferência fuzzy gerando um modelo onde seria possível visualizar além do quadro 5, o quanto um animal que possui uma classificação final excelente, possui da categoria regular, bom, bom para mais etc., conforme será descrito nos capítulos a seguir.

A análise dos dados se dará por meio da ferramenta Weka e R Project com a análise de uma base de dados de classificação real. Os dados serão avaliados em cinco etapas sendo:

1ª Etapa:

Serão analisados dados correspondentes a todas as variáveis avaliadas na Figura 7, Figura 8, Figura 9 e Figura 10.

2ª Etapa:

Serão analisados dados correspondentes a variáveis pertencentes as características de Força Leiteira que, somadas, equivalem a 22% da pontuação final de classificação conforme Figura 7:

Figura 7 - Variáveis avaliadas em força leiteira

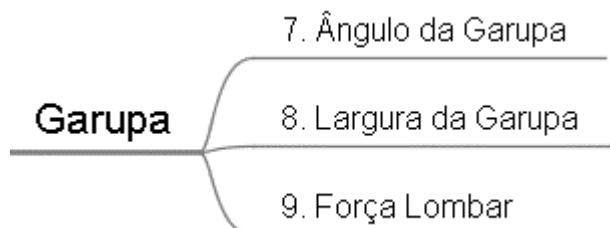


Fonte: Elaborada pelo autor (2016).

3ª Etapa:

Nesta etapa serão analisados dados correspondentes as variáveis pertencentes as características de Garupa que somadas equivalem a 10% da pontuação final de classificação conforme Figura 8:

Figura 8 - Variáveis avaliadas em Garupa

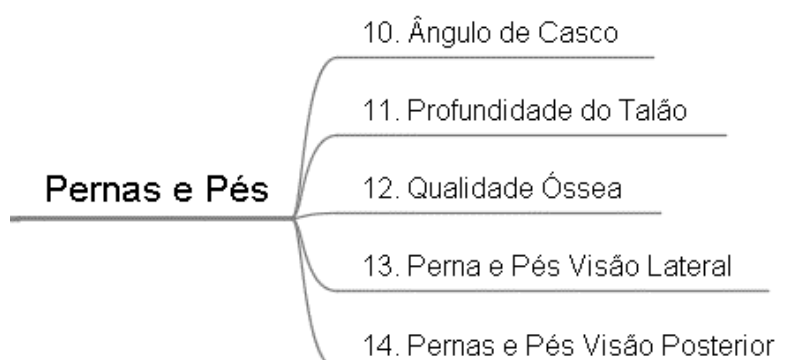


Fonte: Elaborada pelo autor (2016).

4ª Etapa:

Serão analisados os dados correspondentes as variáveis pertencentes as características de Pernas e Pés que, somadas, equivalem a 26% da pontuação final de classificação conforme Figura 9:

Figura 9 - Variáveis avaliadas em Pernas e Pés

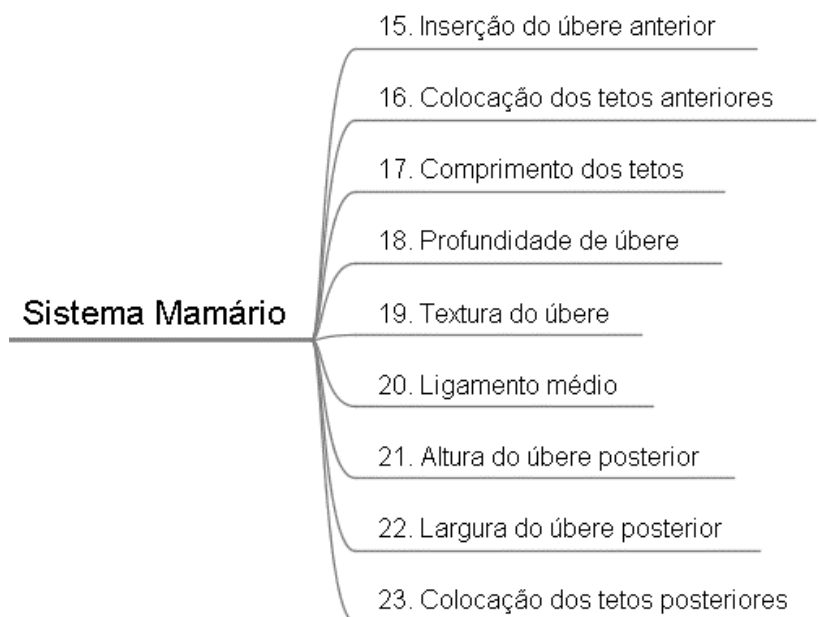


Fonte: Elaborada pelo autor (2016).

5ª Etapa:

Serão analisados dados correspondentes as variáveis pertencentes as características de Sistema Mamário que somam 42% da pontuação final de classificação conforme Figura 10:

Figura 10 - Variáveis avaliadas em Sistema mamário.

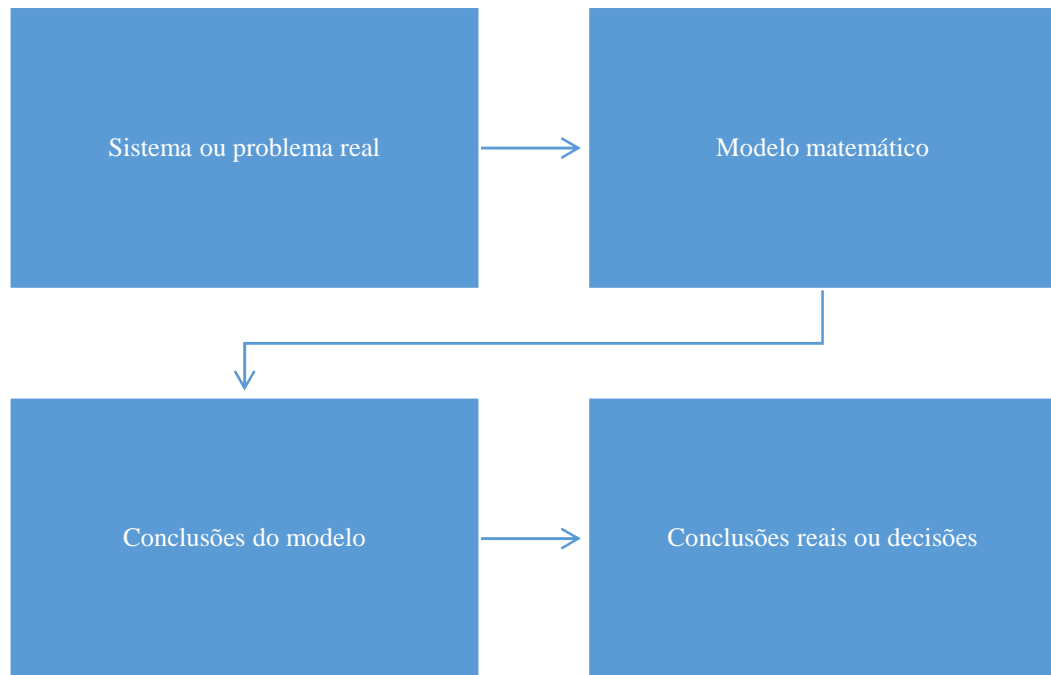


Fonte: Elaborada pelo autor (2016).

As cinco etapas citadas anteriormente são parte integrantes dos resultados preliminares, nas quais são apresentadas, resultados de mineração de regras de associação indicando quais características surgem em conjunto com outras regras com maior frequência.

Segundo Miguel et al. (2012), a solução de problemas utilizando modelagem matemática, para representar satisfatoriamente o sistema real, deve ser feita em diferentes níveis, tendo em vista, que a validação seja coerente com os objetivos e restrições do sistema real. O diagrama apresentado na Figura 11 ilustra o processo simplificado da solução de problemas por meio da utilização de modelagem matemática. Porém, este modelo pode também ser interpretado para modelo de simulação. Deste modo, a construção do modelo seguirá a metodologia da abordagem de pesquisa operacional envolvendo várias fases baseadas na Figura 11.

Figura 11 - Processo de Modelagem



Fonte: Adaptado de Miguel et al. (2012).

3.3 LEVANTAMENTO BIBLIOGRÁFICO

O levantamento bibliográfico visa à ciência de contexto e geração de conhecimento por meio do entendimento da seleção de bovinos com o sistema de classificação linear através de levantamento das principais fontes de referências.

Buscou-se com esta etapa verificar publicações recentes acerca da seleção de bovinos por classificação linear de tipo por meio de teses, dissertações, trabalho de conclusão de graduação e sites especializados. Buscou-se também referências sobre criação de modelos de apoio a decisão por meio de várias técnicas estatísticas e não-estatísticas, em periódicos relevantes para a pesquisa operacional, livros de referências básicas sobre estatística, mineração de dados em bases especializadas e de domínio público. Nesta, também foi definido o escopo do problema em estudo, as decisões de interesse e os objetivos geral e específicos além da justificção dos mesmos. Os detalhes resultantes foram apresentados no Capítulo 1.

3.4 CONSTRUÇÃO DO MODELO

A pesquisa em gestão de produção e operações quando baseada em modelos quantitativos pode ser classificada como uma abordagem de geração de conhecimento racional, partindo do princípio que é possível construir modelos que expliquem pelo menos parte de processos reais. (MIGUEL et al., 2012).

Para construção do modelo de decisão, foi considerado importante a utilização de mineração de dados para análise, e posterior geração de conhecimento sobre os bovinos classificados constantes em uma base de dados. A importância para análise dos dados por meio de *data mining* se fez por tratar-se de uma técnica amplamente aplicada em vários segmentos sociais e produtivos, conforme Quadro 3 do referencial teórico.

Para este modelo de decisão, a utilização da mineração torna-se interessante para viabilizar o conhecimento e descoberta de padrões em base de dado. Assim sendo, um dos padrões mais comuns que pode ser descoberto a partir do processo de mineração de dados é um conjunto de regras de associação que representa combinações de itens que ocorrem simultaneamente com determinada frequência. (GONÇALVES, 2005b).

Foi considerado para este modelo a hipótese de que os animais pertencentes a esta base de dados e classificados segundo o método de classificação linear, possuem características que são comuns ou frequentes para cada nível de classificação, conforme é ilustra o Quadro 5. Partindo desta hipótese, a mineração de dados com regras de associação tem por objetivo encontrar todas as regras relevantes e que serão utilizadas para indicar prováveis característica antecedentes que implicam na característica consequente.

Para análise e extração das regras de associação utilizou-se a ferramenta WEKA com algoritmo apriori introduzido por Agrawal et al. (1993). A escolha por esta ferramenta se deu por se tratar de uma ferramenta amplamente utilizada no meio acadêmico, por possuir um conjunto de algoritmos para extração de padrões de uma base de dados além de uma interface intuitiva e ampla documentação sobre a mesma.

Paralelamente com a mineração de dados optou-se por utilizar a lógica *fuzzy*, como uma outra forma de analisar os dados e ajudar na tomada de decisão para a classificação de bovinos. A lógica fuzzy, neste contexto, tem por função automatizar as ações do modelo por meio de geração de resultados, resultados estes, obtidos por

valores de entrada e as regras de interferência do modelo *fuzzy*, sendo a saída da classificação.

Para criação do modelo *fuzzy*, utilizou-se como base a estrutura proposta por Vidal; Landin e Costa (2012), conforme ilustrado na Figura 6 do referencial teórico. A ferramenta utilizada para tal, foi o R GUI (32-bit), com a biblioteca *sets*, desenvolvida exclusivamente para lógica *fuzzy*. Esta ferramenta de interface amigável opera através de linhas de comandos com métodos e parâmetros que podem ser compilados, executados e salvos em vários formatos de texto, facilitando assim a manipulação do mesmo.

Inicialmente foi criado um modelo conceitual através da abstração das variáveis presentes na base de dados fornecida pela associação de criadores, extraíndo as variáveis que definem o comportamento do sistema de classificação. Neste sentido, a divisão dos processos em passos trouxe a possibilidade de segmentação das atividades e a criação de um artefato com documentação mais detalhada.

O modelo de decisão como expressado na Figura 12, envolve dois atores no processo de classificação, sendo, o classificador e a vaca a ser classificada. Neste modelo toda ação de classificação é iniciada pelo classificador onde, por meio da identificação das características e atribuição de pontuações o modelo pode auxiliar o classificador até a última etapa, que é o resultado final de classificação.

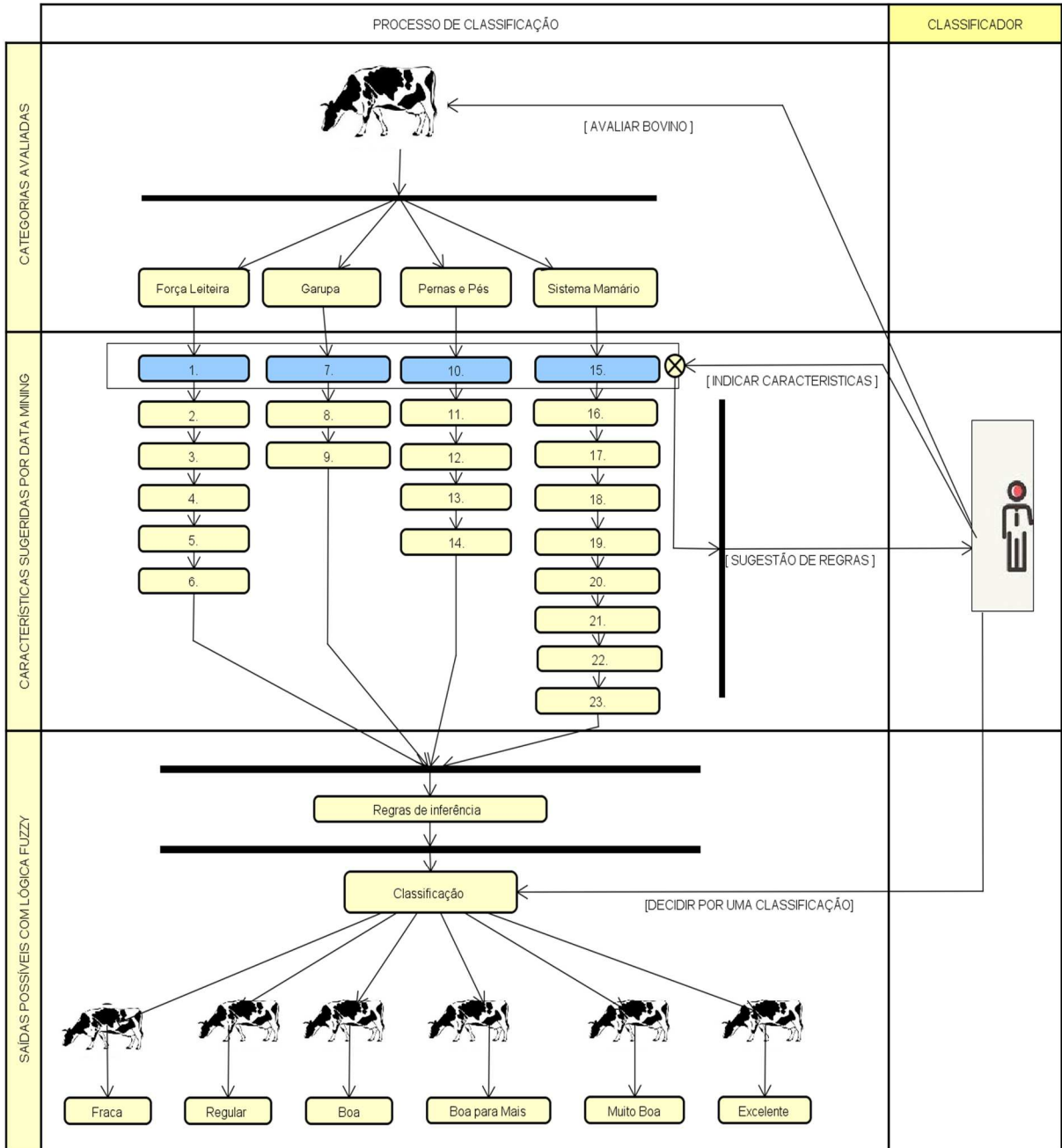
A primeira divisão do modelo tem por título, “Categorias Avaliadas”, e compreende a etapa do processo de classificação onde o classificador avalia as categorias, força leiteira, garupa, pernas e pés e sistema mamária.

Iniciando a classificação pela categoria força leiteira, garupa, pernas e pés ou sistema mamária, o classificador passará para a segunda divisão do modelo. Esta, tem por título “Características sugerida por *data mining*”, nesta, o classificador atribuirá uma pontuação para a primeira característica representada no diagrama por um retângulo com cantos arredondados na cor azul onde, a partir de então modelo passará a sugerir características obtidas por meio de regras de associação, que podem ser aceitas ou não pelo classificador.

Na terceira divisão do modelo, intitulada “Saídas possíveis com lógica *fuzzy*”, o modelo com base em regras de inferência *fuzzy* irá resultar em uma das categorias possíveis de classificação, de acordo com as pontuações obtidas por cada uma das características avaliadas na etapa anterior, cabendo ao classificador o poder de

decisão para aptidão da vaca. A Figura 12, representa o modelo conceitual desenvolvido por este estudo.

Figura 12 - Modelo de decisão



Fonte: Elaborado pelo autor (2016).

- 1.Estatura, 2.Largura do peito, 3. Profundidade corporal, 4. Angulosidade,5. Condição corporal,
- 6.Nivelamento da linha superior, 7. Largura da garupa,8.Angulo da garupa, 9.Força lombar,10.Qualidade óssea, 11.Pernas vista lateral,12. Pernas Vista traseira, 13. Ângulo do casco,
14. Profundidade de talão, 15.Inserção do úbere anterior, 16. Colocação dos tetos anteriores, 17. Comprimento dos tetos, 18.Profundidade do úbere,19. Textura do úbere, 20. Ligamento suspensor médio,21. Altura do úbere posterior, 22. Largura do úbere posterior, 23. Colocação dos tetos posteriores.

3.4.1 Resultados preliminares

Para utilização da base de dados tanto para análise de regras de associação quanto para geração de regras de inferência fuzzy, a princípio foi feita a escolha das variáveis que fazem parte do modelo atual de classificação de bovinos leiteiros de acordo com relatos de Valloto e Neto (2010). O segundo passo visou à identificação de dados considerados inconsistentes, realizando o descarte dos mesmos. Para a análise preliminar do modelo, fez-se necessário o uso do procedimento conhecido como pré-processamento que envolve basicamente a seleção, limpeza, codificação e enriquecimento dos dados, ou seja, quatro etapas antes de iniciar a análise propriamente dita. (GOLDSCHMIDT; PASSOS,2005).

Tendo os dados limpos, o próximo passo foi a seleção dos dados que foram necessários para o processo de mineração de dados. O processo de associação de características em transações leva em conta apenas o número da transação e as características lineares nelas compostas. Diante disso, dados como nome do animal, sangue, nome do proprietário e outras informações características com finalidade de informação sobre a transação foram excluídos.

Como resultados preliminares, para esta base de dados foram contabilizados 144 animais com as maiores produções vitalícias do mês de abril de 2015 do estado do Paraná. Fazendo uso da ferramenta WEKA com o uso da aba classificação em uma análise preliminar nos registros dos 144 animais classificados, observou-se que a maioria dos animais foram classificados como bom para mais (BM) seguido de animais classificados como muito bom (MB). Em uma análise mais detalhada obteve-se como resultado que 81,9% dos animais pertencentes a essa base de dados possuem pontuações elevadas não contendo nenhuma vaca com pontuação que lhe conferisse a condição de um animal regular (R) para aptidão leiteira, conforme na Figura 13.

Figura 13 - Animais constantes na base de dados

```

a  b  c  d  <-- classified as
0 15  0  0 | a = BOA
0 69  0  0 | b = BM
0 11  0  0 | c = EXCELENTE
0 49  0  0 | d = MB

```

Fonte: elaborado pelo autor (2016)

O raciocínio para isso é que se algumas características morfológicas presentes em animais classificados entre boa (B) e excelente (EX) para produção leiteira são muito presentes, a probabilidade de estarem associadas é maior que em classificações mais baixas. Diante dessa premissa e da constatação das características morfológicas mais presentes possuírem pontuações mais próximas de ideais e altas.

Para os dados da planilha de classificação foi preciso categorizar os dados mudando a estrutura da tabela, onde antes eram pontos de 1 a 9 passaram a receber o nome da característica mais uma das categorias baixa, intermediária, alta ou ideal ficando conforme na Figura 14.

Figura 14 - Categorização dos dados

	A	B	C	D	E	F
1	ESTATURA	NSUP	LARGPEITO	PROFCORP	ANGULOSIDADE	CONDICI
2	ESTIDEAL	NSUPIDEAL	LARGPINTER	PROFCINTER	ANGULOSIBAIXA	CONDICI
3	ESTALTA	NSUPINTER	LARGPINTER	PROFCINTER	ANGULOSIIDEAL	CONDICI
4	ESTINTER	NSUPIDEAL	LARGPIDEAL	PROFCIDEAL	ANGULOSIIDEAL	CONDICI
5	ESTALTA	NSUPINTER	LARGPINTER	PROFCINTER	ANGULOSIIDEAL	CONDICI
6	ESTALTA	NSUPINTER	LARGPINTER	PROFCINTER	ANGULOSIIDEAL	CONDICI
7	ESTIDEAL	NSUPALTA	LARGPINTER	PROFCIDEAL	ANGULOSIALTA	CONDICI
8	ESTINTER	NSUPIDEAL	LARGPIDEAL	PROFCIDEAL	ANGULOSIIDEAL	CONDICI
9	ESTIDEAL	NSUPALTA	LARGPINTER	PROFCIDEAL	ANGULOSIALTA	CONDICI
10	ESTINTER	NSUPINTER	LARGPIDEAL	PROFCIDEAL	ANGULOSIIDEAL	CONDICI
11	ESTIDEAL	NSUPIDEAL	LARGPINTER	PROFCINTER	ANGULOSIBAIXA	CONDICI

Fonte: Elaborado pelo autor (2016).

Para mineração destes dados, foram consideradas as 23 características de conformação mais a coluna linear que recebe a classificação, para as 144 vacas registradas nesta base. A etapa de mineração de dados é a etapa mais importante de todo o processo de geração de conhecimento. Para se chegar nela, é essencial dar uma atenção especial às outras fases, pois esta é uma fase crítica do processo. Se o sistema entrar com variáveis inconsistentes, falhas ou ausentes, provavelmente todo o conhecimento extraído não será verdadeiro. Como definido antes, o escopo inicial deveria fazer tarefas de mineração para:

- Identificar as características mais frequentes e pontuações em animais classificados entre BOA e EXCELENTE; e
- Tentar conhecer as características que estão associadas em transações, para cada categoria avaliada como por exemplo avaliar quais pontuações são mais comuns às características do sistema mamário.

4 EXPERIMENTOS E RESULTADOS

Este capítulo discorre sobre os experimentos e resultados obtidos pelo modelo de classificação. Os experimentos realizados iniciam com a análise de regras de associação com algoritmo apriori, por meio da ferramenta WEKA em uma base de dados com 144 registros de vacas classificadas no modelo de classificação linear do tipo “true type”. Esta etapa, objetiva extrair regras com suporte e confiança relevantes para utilização na segunda fase do modelo conforme está ilustrado na Figura 12.

A segunda parte deste capítulo compreende a etapa de modelagem da lógica fuzzy com o uso da ferramenta R Gui (32 bit), tomando por base a Figura 6. Objetiva-se com esta etapa a finalização do modelo de decisão, fazendo com que as regras de inferência fuzzy atuem como decisivas a classificação final da vaca avaliada.

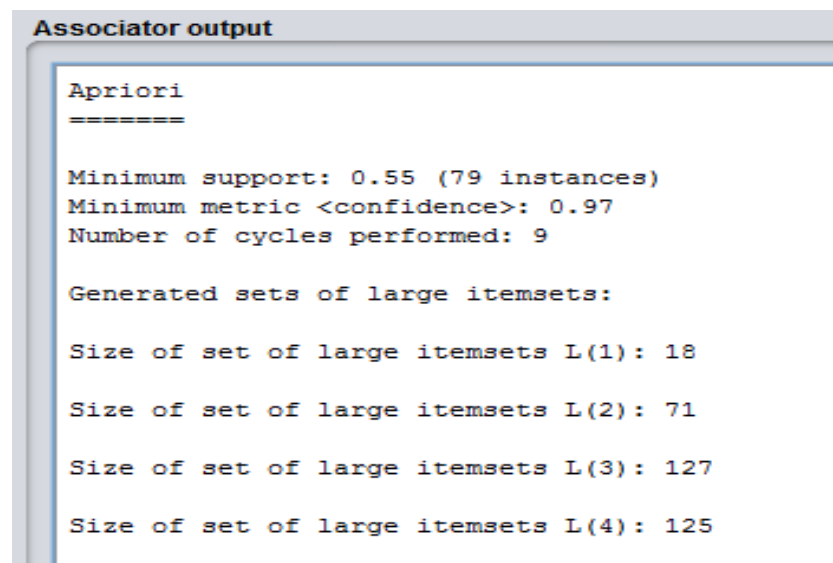
4.1 APLICAÇÃO DE REGRAS DE ASSOCIAÇÃO

Foram realizados cinco experimentos, cada um com três testes alterando as métricas: suporte mínimo para 30%, 45% e 55% sucessivamente e mantendo a confiança em 97%. Para registro, o suporte determina a porcentagem mínima exigida em que as características estejam contidos nas transações, já a confiança determina a porcentagem das transações que contendo uma característica ou mais de uma, contenha outra associada. Os experimentos foram assim encaminhados para que não houvessem uma geração demasiada de regras consideradas irrelevantes para o problema. Assim sendo, os experimentos foram os seguintes:

1. Aplicação de regras de associação em todas as categorias
2. Aplicação de regras de associação em todas as variáveis da categoria força leiteira;
3. Aplicação de regras de associação em todas as variáveis da categoria garupa
4. Aplicação de regras de associação em todas as variáveis da categoria pernas e pés; e
5. Aplicação de regras de associação em todas as variáveis da categoria sistema mamário.

Para os experimentos realizados com as características pertencentes a categorias listadas anteriormente foram mantidas as configurações da ferramenta Weka com suporte de 55% e a confiança de 97%, isso significa que a ferramenta procurou em todos os registros, aqueles que estava presentes em no mínimo 55% das transações (suporte) e que estivessem juntos em, pelo menos, 97% das transações (confiança). Na Figura 15, definições de suporte e confiança são apresentadas.

Figura 15 - Definição de suporte e confiança no Weka



```

Associator output

Apriori
=====

Minimum support: 0.55 (79 instances)
Minimum metric <confidence>: 0.97
Number of cycles performed: 9

Generated sets of large itemsets:

Size of set of large itemsets L(1): 18
Size of set of large itemsets L(2): 71
Size of set of large itemsets L(3): 127
Size of set of large itemsets L(4): 125
  
```

Fonte: Elaborada pelo autor (2016).

Diante dos resultados obtidos, foram observados alguns pontos: foram solicitadas ao programa mil regras de associação considerando apenas as regras que contenham *lift* superior a um ($lift > 1$), indicando que as ocorrências são dependentes. Algumas regras foram encontradas podem dizer alguma coisa sobre a implicação de pontuações para cada característica. Por exemplo, as regras

ALTUBANT = ALTUBAIDEAL

==>

LARGUBPOST = LARGUBPIDEAL

ou

ANGGARUPA = ANGGINTER,

FLOMBO = FLOMIDEAL,

ALTUBANT = ALTUBAIDEAL

==>

LARGUBPOST = LARGUBPIDEAL,

podem indicar no primeiro caso que os animais que possuem altura do úbere anterior ideal também possuem largura do úbere posterior ideal e na outra regra, que animais que possuem ângulo da garupa intermediário, força de lombo ideal, altura do úbere anterior ideal provavelmente terá largura do úbere posterior ideal. Sugere que algumas características implicam em outras com a confiança de 100% para estas transações. A Figura 16 apresenta as regras geradas neste experimento.

Figura 16 - Regras obtidas para todas categorias

-
1. FLOMBO=FLOMIDEAL 135 ==> ANGGARUPA=ANGGINTER 135 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 2. CONDCORP=CONDCIDEAL 130 ==> ANGGARUPA=ANGGINTER 130 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 3. LARGUBPOST=LARGUBPIDEAL 123 ==> ANGGARUPA=ANGGINTER 123 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 4. CONDCORP=CONDCIDEAL FLOMBO=FLOMIDEAL 123 ==> ANGGARUPA=ANGGINTER 123 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 5. ALTUBANT=ALTUBAIDEAL 119 ==> ANGGARUPA=ANGGINTER 119 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 6. ALTUBANT=ALTUBAIDEAL 119 ==> LARGUBPOST=LARGUBPIDEAL 119 <conf:(1)> lift:(1.17) lev:(0.12) [17] conv:(17.35)
 7. CONDCORP=CONDCIDEAL LARGUBPOST=LARGUBPIDEAL 119 ==> ANGGARUPA=ANGGINTER 119 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 8. ALTUBANT=ALTUBAIDEAL LARGUBPOST=LARGUBPIDEAL 119 ==> ANGGARUPA=ANGGINTER 119 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 9. ANGGARUPA=ANGGINTER ALTUBANT=ALTUBAIDEAL 119 ==> LARGUBPOST=LARGUBPIDEAL 119 <conf:(1)> lift:(1.17) lev:(0.12) [17] conv:(17.35)
 10. ALTUBANT=ALTUBAIDEAL 119 ==> ANGGARUPA=ANGGINTER LARGUBPOST=LARGUBPIDEAL 119 <conf:(1)> lift:(1.17) lev:(0.12) [17] conv:(17.35)
 11. CONDCORP=CONDCIDEAL ALTUBANT=ALTUBAIDEAL 115 ==> ANGGARUPA=ANGGINTER 115 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 12. CONDCORP=CONDCIDEAL ALTUBANT=ALTUBAIDEAL 115 ==> LARGUBPOST=LARGUBPIDEAL 115 <conf:(1)> lift:(1.17) lev:(0.12) [16] conv:(16.77)
 13. FLOMBO=FLOMIDEAL LARGUBPOST=LARGUBPIDEAL 115 ==> ANGGARUPA=ANGGINTER 115 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 14. CONDCORP=CONDCIDEAL ALTUBANT=ALTUBAIDEAL LARGUBPOST=LARGUBPIDEAL 115 ==> ANGGARUPA=ANGGINTER 115 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 15. CONDCORP=CONDCIDEAL ANGGARUPA=ANGGINTER ALTUBANT=ALTUBAIDEAL 115 ==> LARGUBPOST=LARGUBPIDEAL 115 <conf:(1)> lift:(1.17) lev:(0.12) [16] conv:(16.77)
 16. CONDCORP=CONDCIDEAL ALTUBANT=ALTUBAIDEAL 115 ==> ANGGARUPA=ANGGINTER LARGUBPOST=LARGUBPIDEAL 115 <conf:(1)> lift:(1.17) lev:(0.12) [16] conv:(16.77)
 17. CONDCORP=CONDCIDEAL FLOMBO=FLOMIDEAL LARGUBPOST=LARGUBPIDEAL 113 ==> ANGGARUPA=ANGGINTER 113 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 18. FLOMBO=FLOMIDEAL ALTUBANT=ALTUBAIDEAL 111 ==> ANGGARUPA=ANGGINTER 111 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 19. FLOMBO=FLOMIDEAL ALTUBANT=ALTUBAIDEAL 111 ==> LARGUBPOST=LARGUBPIDEAL 111 <conf:(1)> lift:(1.17) lev:(0.11) [16] conv:(16.19)
 20. FLOMBO=FLOMIDEAL ALTUBANT=ALTUBAIDEAL LARGUBPOST=LARGUBPIDEAL 111 ==> ANGGARUPA=ANGGINTER 111 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 21. ANGGARUPA=ANGGINTER FLOMBO=FLOMIDEAL ALTUBANT=ALTUBAIDEAL 111 ==> LARGUBPOST=LARGUBPIDEAL 111 <conf:(1)> lift:(1.17) lev:(0.11) [16] conv:(16.19)
 22. FLOMBO=FLOMIDEAL ALTUBANT=ALTUBAIDEAL 111 ==> ANGGARUPA=ANGGINTER LARGUBPOST=LARGUBPIDEAL 111 <conf:(1)> lift:(1.17) lev:(0.11) [16] conv:(16.19)
 23. CONDCORP=CONDCIDEAL FLOMBO=FLOMIDEAL ALTUBANT=ALTUBAIDEAL 109 ==> ANGGARUPA=ANGGINTER 109 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 24. CONDCORP=CONDCIDEAL FLOMBO=FLOMIDEAL ALTUBANT=ALTUBAIDEAL 109 ==> LARGUBPOST=LARGUBPIDEAL 109 <conf:(1)> lift:(1.17) lev:(0.11) [15] conv:(15.9)
 25. CONDCORP=CONDCIDEAL FLOMBO=FLOMIDEAL ALTUBANT=ALTUBAIDEAL LARGUBPOST=LARGUBPIDEAL 109 ==> ANGGARUPA=ANGGINTER 109 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 26. CONDCORP=CONDCIDEAL ANGGARUPA=ANGGINTER FLOMBO=FLOMIDEAL ALTUBANT=ALTUBAIDEAL 109 ==> LARGUBPOST=LARGUBPIDEAL 109 <conf:(1)> lift:(1.17) lev:(0.11) [15] conv:(15.9)
 27. CONDCORP=CONDCIDEAL FLOMBO=FLOMIDEAL ALTUBANT=ALTUBAIDEAL 109 ==> ANGGARUPA=ANGGINTER LARGUBPOST=LARGUBPIDEAL 109 <conf:(1)> lift:(1.17) lev:(0.11) [15] conv:(15.9)

Fonte: Elaborada pelo autor (2016).

No segundo experimento, onde foi realizado a mineração de regras de associação apenas com dados da categoria força lombar seguindo os mesmos parâmetros utilizados no experimento anterior, obteve-se algumas regras como por exemplo, a regra 15

NSUP = NSUPALTA,
CONDCORP = CONDCIDEAL

==>

LARGPEITO = LARGPINTER

que indica que o animal que possui nivelamento da linha superior (NSUP) alta e condição corporal (CONDCORP) ideal provavelmente terá largura do peito (LARGPEITO) intermediária. As regras obtidas são apresentadas na Figura 17.

Já no experimento utilizando a categoria garupa, utilizando suas três variáveis, ângulo da garupa, largura da garupa e força lombar seguindo os mesmos parâmetros para suporte e confiança, apenas três regras com *lift* superior a um (1), tendo como exemplo, a regra

LARGGARUPA = LARGGIDEAL

==>

FLOMBO = FLOMIDEAL,

indicando que a vaca que possui largura da garupa ideal também provavelmente terá força de lombo ideal. Na Figura 18 são mostradas as regras geradas para esta categoria.

No quarto experimento, utilizando a categoria pernas e pés e todas as suas variáveis, obteve-se um total de 29 regras acima de 97% de confiança, como de associação para esta categoria temos a regra gerada pelo Weka,

QOSSEA = QOSSEAINTER,

PPVPOST = PPVPBAIXA

==>

PROFTALAO = PROFTINTER,

onde um animal com qualidade óssea intermediária, pernas e pés visão posterior igual a baixa pode também possuir profundidade do talão intermediário. Na Figura 19 são ilustradas algumas regras desta categoria.

Figura 17 - Regras obtidas para categoria força lombar

Best rules found:

1. ANGULOSIDADE=ANGULOSIALTA 76 ==> CONDCORP=CONDCIDEAL 76 <conf:(1)> lift:(1.11) lev:(0.05) [7] conv:(7.39)
2. PROFCORP=PROFCINTER CONDCORP=CONDCIDEAL 63 ==> LARGPEITO=LARGPINTER 63 <conf:(1)> lift:(1.36) lev:(0.12) [16] conv:(16.63)
3. LARGPEITO=LARGPINTER PROFCORP=PROFCINTER 63 ==> CONDCORP=CONDCIDEAL 63 <conf:(1)> lift:(1.11) lev:(0.04) [6] conv:(6.13)
4. LARGPEITO=LARGPINTER LINEAR=BM 63 ==> CONDCORP=CONDCIDEAL 63 <conf:(1)> lift:(1.11) lev:(0.04) [6] conv:(6.13)
5. LARGPEITO=LARGPINTER ANGULOSIDADE=ANGULOSIALTA 58 ==> CONDCORP=CONDCIDEAL 58 <conf:(1)> lift:(1.11) lev:(0.04) [5] conv:(5.64)
6. PROFCORP=PROFCIDEAL ANGULOSIDADE=ANGULOSIALTA 57 ==> CONDCORP=CONDCIDEAL 57 <conf:(1)> lift:(1.11) lev:(0.04) [5] conv:(5.54)
7. PROFCORP=PROFCINTER CONDCORP=CONDCIDEAL LINEAR=BM 53 ==> LARGPEITO=LARGPINTER 53 <conf:(1)> lift:(1.36) lev:(0.1) [13] conv:(13.99)
8. LARGPEITO=LARGPINTER PROFCORP=PROFCINTER LINEAR=BM 53 ==> CONDCORP=CONDCIDEAL 53 <conf:(1)> lift:(1.11) lev:(0.04) [5] conv:(5.15)
9. NSUP=NSUPALTA 50 ==> LARGPEITO=LARGPINTER 50 <conf:(1)> lift:(1.36) lev:(0.09) [13] conv:(13.19)
10. NSUP=NSUPALTA 50 ==> ANGULOSIDADE=ANGULOSIALTA 50 <conf:(1)> lift:(1.89) lev:(0.16) [23] conv:(23.61)
11. NSUP=NSUPALTA 50 ==> CONDCORP=CONDCIDEAL 50 <conf:(1)> lift:(1.11) lev:(0.03) [4] conv:(4.86)
12. NSUP=NSUPALTA ANGULOSIDADE=ANGULOSIALTA 50 ==> LARGPEITO=LARGPINTER 50 <conf:(1)> lift:(1.36) lev:(0.09) [13] conv:(13.19)
13. NSUP=NSUPALTA LARGPEITO=LARGPINTER 50 ==> ANGULOSIDADE=ANGULOSIALTA 50 <conf:(1)> lift:(1.89) lev:(0.16) [23] conv:(23.61)
14. NSUP=NSUPALTA 50 ==> LARGPEITO=LARGPINTER ANGULOSIDADE=ANGULOSIALTA 50 <conf:(1)> lift:(2.48) lev:(0.21) [29] conv:(29.86)
15. NSUP=NSUPALTA CONDCORP=CONDCIDEAL 50 ==> LARGPEITO=LARGPINTER 50 <conf:(1)> lift:(1.36) lev:(0.09) [13] conv:(13.19)
16. NSUP=NSUPALTA LARGPEITO=LARGPINTER 50 ==> CONDCORP=CONDCIDEAL 50 <conf:(1)> lift:(1.11) lev:(0.04) [6] conv:(6.13)

Fonte: Elaborada pelo autor (2016).

Figura 18 -Regras geradas da categoria garupa

Best rules found:

1. FLOMBO=FLOMIDEAL 135 ==> ANGGARUPA=ANGGINTER 135 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. LARGGARUPA=LARGGINTER 77 ==> ANGGARUPA=ANGGINTER 77 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. LARGGARUPA=LARGGINTER FLOMBO=FLOMIDEAL 68 ==> ANGGARUPA=ANGGINTER 68 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. LARGGARUPA=LARGGIDEAL 67 ==> ANGGARUPA=ANGGINTER 67 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. LARGGARUPA=LARGGIDEAL 67 ==> FLOMBO=FLOMIDEAL 67 <conf:(1)> lift:(1.07) lev:(0.03) [4] conv:(4.19)
6. LARGGARUPA=LARGGIDEAL FLOMBO=FLOMIDEAL 67 ==> ANGGARUPA=ANGGINTER 67 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. ANGGARUPA=ANGGINTER LARGGARUPA=LARGGIDEAL 67 ==> FLOMBO=FLOMIDEAL 67 <conf:(1)> lift:(1.07) lev:(0.03) [4] conv:(4.19)
8. LARGGARUPA=LARGGIDEAL 67 ==> ANGGARUPA=ANGGINTER FLOMBO=FLOMIDEAL 67 <conf:(1)> lift:(1.07) lev:(0.03) [4] conv:(4.19)

Fonte: Elaborada pelo autor (2016).

Figura 19 - Regras geradas para categoria Perna e Pés

Best rules found:

1. QOSSEA=QOSSEAINTER PFVPOST=PFVPBAIXA 76 ==> PROFTALAO=PROFTINTER 76 <conf:(1)> lift:(1.44) lev:(0.16) [23] conv:(23.22)
2. PROFTALAO=PROFTINTER PFVPOST=PFVPBAIXA 76 ==> QOSSEA=QOSSEAINTER 76 <conf:(1)> lift:(1.57) lev:(0.19) [27] conv:(27.44)
3. QOSSEA=QOSSEAINTER PVLATERAL=PPVLIDEAL 67 ==> PROFTALAO=PROFTINTER 67 <conf:(1)> lift:(1.44) lev:(0.14) [20] conv:(20.47)
4. PVLATERAL=PPVLIDEAL PFVPOST=PFVPBAIXA 65 ==> PROFTALAO=PROFTINTER 65 <conf:(1)> lift:(1.44) lev:(0.14) [19] conv:(19.86)
5. PVLATERAL=PPVLIDEAL PFVPOST=PFVPBAIXA 65 ==> QOSSEA=QOSSEAINTER 65 <conf:(1)> lift:(1.57) lev:(0.16) [23] conv:(23.47)
6. QOSSEA=QOSSEAINTER PVLATERAL=PPVLIDEAL PFVPOST=PFVPBAIXA 65 ==> PROFTALAO=PROFTINTER 65 <conf:(1)> lift:(1.44) lev:(0.14) [19] †
7. PROFTALAO=PROFTINTER PVLATERAL=PPVLIDEAL PFVPOST=PFVPBAIXA 65 ==> QOSSEA=QOSSEAINTER 65 <conf:(1)> lift:(1.57) lev:(0.16) [23] †
8. PVLATERAL=PPVLIDEAL PFVPOST=PFVPBAIXA 65 ==> PROFTALAO=PROFTINTER QOSSEA=QOSSEAINTER 65 <conf:(1)> lift:(1.85) lev:(0.21) [29] †
9. ANGCASCO=ANGCINTER QOSSEA=QOSSEAINTER 57 ==> PROFTALAO=PROFTINTER 57 <conf:(1)> lift:(1.44) lev:(0.12) [17] conv:(17.42)
10. ANGCASCO=ANGCINTER QOSSEA=QOSSEAINTER 57 ==> PVLATERAL=PPVLIDEAL 57 <conf:(1)> lift:(1.4) lev:(0.11) [16] conv:(16.23)

Fonte: Elaborada pelo autor (2016).

Por fim, no quinto experimento, utilizando a categoria sistema mamário e todas as suas variáveis, obteve-se um total de 1000 regras acima de 97% de confiança, como de associação para esta categoria foi gerada, por exemplo, a regra gerada pelo Weka,

INSUBANT = INSUBALINTER,

COMPTETOS = COMPTBAIXA,

LIGMEDIO = LIGMIDEAL

==>

COLTANT = COLTANTINTER,

TEXUBERE = TEXUBINTER,

onde um animal com inserção do úbere anterior intermediária, comprimento dos tetos baixa, ligamento médio ideal, e também possuir colocação dos tetos anteriores intermediaria, textura do úbere intermediaria, com 100% de confiança e 2.4 de *lift*. Na Figura 20 são ilustradas algumas regras da categoria sistema mamário.

Figura 20 - Regras geradas para categoria Sistema Mamário

Best rules found:

1. ALTUBANT=ALTUBAIDEAL 119 ==> LARGUBPOST=LARGUBPIDEAL 119 <conf:(1)> lift:(1.17) lev:(0.12) [17] conv:(17.35)
2. COLTANT=COLTANTINTER ALTUBANT=ALTUBAIDEAL 89 ==> LARGUBPOST=LARGUBPIDEAL 89 <conf:(1)> lift:(1.17) lev:(0.09) [12] conv:(12.98)
3. LIGMEDIO=LIGMIDEAL LARGUBPOST=LARGUBPIDEAL 83 ==> ALTUBANT=ALTUBAIDEAL 83 <conf:(1)> lift:(1.21) lev:(0.1) [14] conv:(14.41)
4. LIGMEDIO=LIGMIDEAL ALTUBANT=ALTUBAIDEAL 83 ==> LARGUBPOST=LARGUBPIDEAL 83 <conf:(1)> lift:(1.17) lev:(0.08) [12] conv:(12.1)
5. INSUBANT=INSUBALINTER ALTUBANT=ALTUBAIDEAL 78 ==> LARGUBPOST=LARGUBPIDEAL 78 <conf:(1)> lift:(1.17) lev:(0.08) [11] conv:(11.38)
6. TEXUBERE=TEXUBINTER LARGUBPOST=LARGUBPIDEAL 77 ==> ALTUBANT=ALTUBAIDEAL 77 <conf:(1)> lift:(1.21) lev:(0.09) [13] conv:(13.37)
7. TEXUBERE=TEXUBINTER ALTUBANT=ALTUBAIDEAL 77 ==> LARGUBPOST=LARGUBPIDEAL 77 <conf:(1)> lift:(1.17) lev:(0.08) [11] conv:(11.23)
8. INSUBANT=INSUBALINTER COLTANT=COLTANTINTER ALTUBANT=ALTUBAIDEAL 76 ==> LARGUBPOST=LARGUBPIDEAL 76 <conf:(1)> lift:(1.17) lev:(0.08)
9. COMPTETOS=COMPTBAIXA 75 ==> TEXUBERE=TEXUBINTER 75 <conf:(1)> lift:(1.67) lev:(0.21) [30] conv:(30.21)
10. COMPTETOS=COMPTBAIXA 75 ==> ALTUBANT=ALTUBAIDEAL 75 <conf:(1)> lift:(1.21) lev:(0.09) [13] conv:(13.02)
11. COMPTETOS=COMPTBAIXA 75 ==> LARGUBPOST=LARGUBPIDEAL 75 <conf:(1)> lift:(1.17) lev:(0.08) [10] conv:(10.94)
12. COLTETPOST=COLTETPIDEAL 75 ==> ALTUBANT=ALTUBAIDEAL 75 <conf:(1)> lift:(1.21) lev:(0.09) [13] conv:(13.02)
13. COLTETPOST=COLTETPIDEAL 75 ==> LARGUBPOST=LARGUBPIDEAL 75 <conf:(1)> lift:(1.17) lev:(0.08) [10] conv:(10.94)
14. COMPTETOS=COMPTBAIXA ALTUBANT=ALTUBAIDEAL 75 ==> TEXUBERE=TEXUBINTER 75 <conf:(1)> lift:(1.67) lev:(0.21) [30] conv:(30.21)
15. COMPTETOS=COMPTBAIXA TEXUBERE=TEXUBINTER 75 ==> ALTUBANT=ALTUBAIDEAL 75 <conf:(1)> lift:(1.21) lev:(0.09) [13] conv:(13.02)
16. COMPTETOS=COMPTBAIXA 75 ==> TEXUBERE=TEXUBINTER ALTUBANT=ALTUBAIDEAL 75 <conf:(1)> lift:(1.87) lev:(0.24) [34] conv:(34.9)
17. COMPTETOS=COMPTBAIXA LARGUBPOST=LARGUBPIDEAL 75 ==> TEXUBERE=TEXUBINTER 75 <conf:(1)> lift:(1.67) lev:(0.21) [30] conv:(30.21)
18. COMPTETOS=COMPTBAIXA TEXUBERE=TEXUBINTER 75 ==> LARGUBPOST=LARGUBPIDEAL 75 <conf:(1)> lift:(1.17) lev:(0.08) [10] conv:(10.94)
19. COMPTETOS=COMPTBAIXA 75 ==> TEXUBERE=TEXUBINTER LARGUBPOST=LARGUBPIDEAL 75 <conf:(1)> lift:(1.87) lev:(0.24) [34] conv:(34.9)
20. COMPTETOS=COMPTBAIXA LARGUBPOST=LARGUBPIDEAL 75 ==> ALTUBANT=ALTUBAIDEAL 75 <conf:(1)> lift:(1.21) lev:(0.09) [13] conv:(13.02)

Fonte : Elaborada pelo autor (2016).

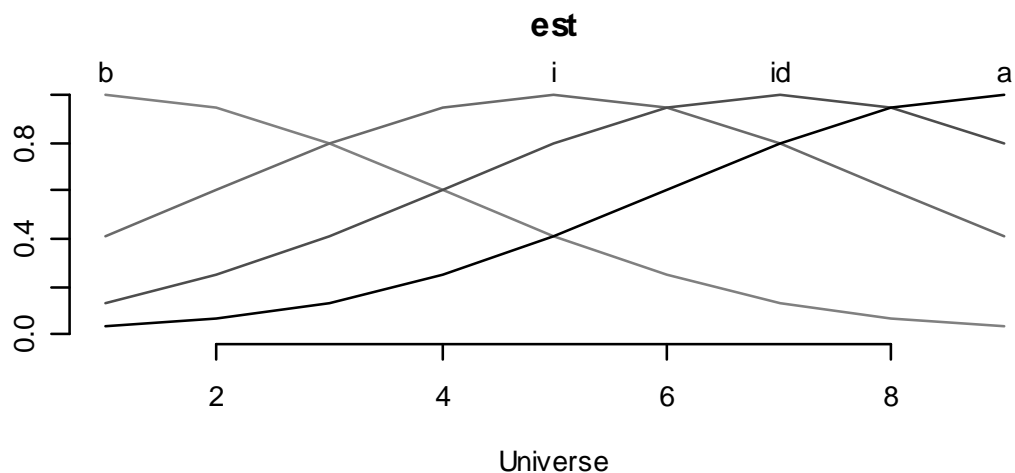
4.2 APLICAÇÃO DA TEORIA FUZZY

Na análise por meio da lógica nebulosa em registros em conjuntos de dados, para esta pesquisa foi necessário categorizar os dados de acordo com referencial teórico, buscando indicar por meio desta técnica a classificação em que determinado animal é pertencente, visualizando além da classificação final a porcentagem de quanto a vaca pertence as demais classes.

Para esse procedimento primeiramente foram transformadas as variáveis da base de dados em variáveis linguísticas nomeando-as de forma abreviada para facilitar a codificação no R. Posteriormente definiu-se para as variáveis pertencentes ao sistema fuzzy a função de pertinência do tipo gaussiana, universos de 1 até 5 para a variável linguística ccorp (condição corporal) e de 1 até 9 para as demais variáveis, conforme descrito abaixo.

À variável “**est**” (acrônimo para estatura) foram associadas quatro funções pertinência, assumindo que a estatura pode obter as funções, baixa (b), intermediária (i), alta (a) e ideal (id), baseada em Valloto e Neto (2010) e manual de classificação morfológica (MANUAL...,2009). A baixa (b) recebe o valor 1, a intermediária (i) recebe 5, alta (a) o valor 9 e ideal (id) o valor 7, em um universo de 1 a 9. As Na Figura 21 apresentam-se as funções de pertinência da variável linguística **est** (estatura).

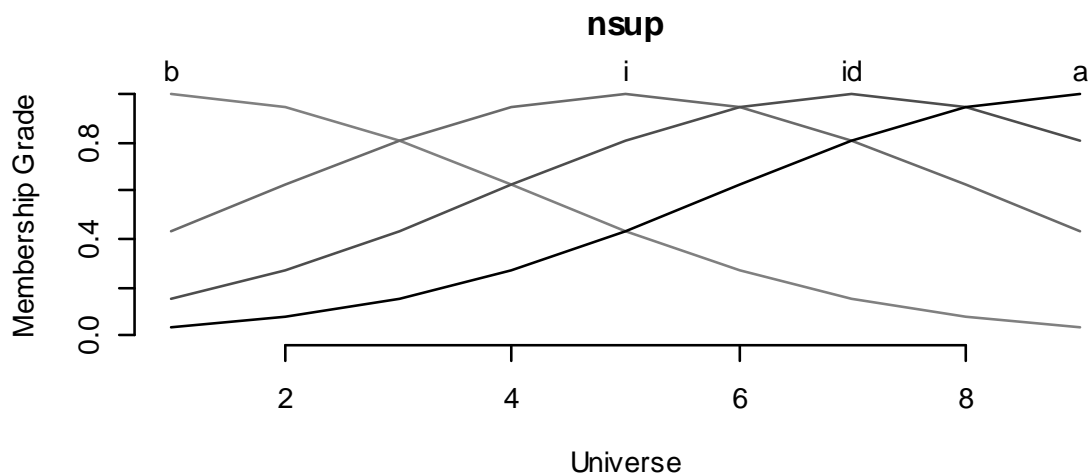
Figura 21 - Gráfico da variável linguística Estatura (est)



Fonte: Elaborada pelo autor (2016).

À variável “**nsup**” (acrônimo para nivelamento da linha superior) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alta (a) e ideal (id). A baixa (b) recebe o valor 1, a intermediária (i) o valor 5, alta (a) o valor 9 a e ideal (id) o valor 7, em um universo de 1 a 9. Na Figura 22 apresenta-se as funções de pertinência da variável linguística **nsup** (nivelamento da linha superior).

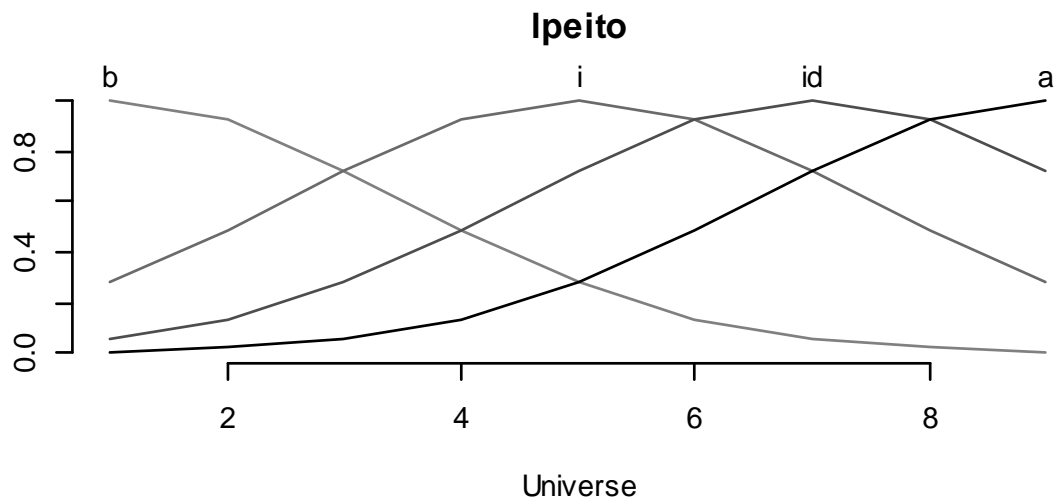
Figura 22 - Gráfico da variável linguística nivelamento da linha superior (nsup)



Fonte: Elaborada pelo autor (2016).

À variável “**lpeito**” (acrônimo para largura do peito) foram associadas também quatro funções pertinência, as funções, baixo (b), intermediária (i), alta (a) e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 7, em um universo de 1 a 9. Na Figura 23 apresenta-se as funções de pertinência da variável linguística **lpeito** (largura do peito).

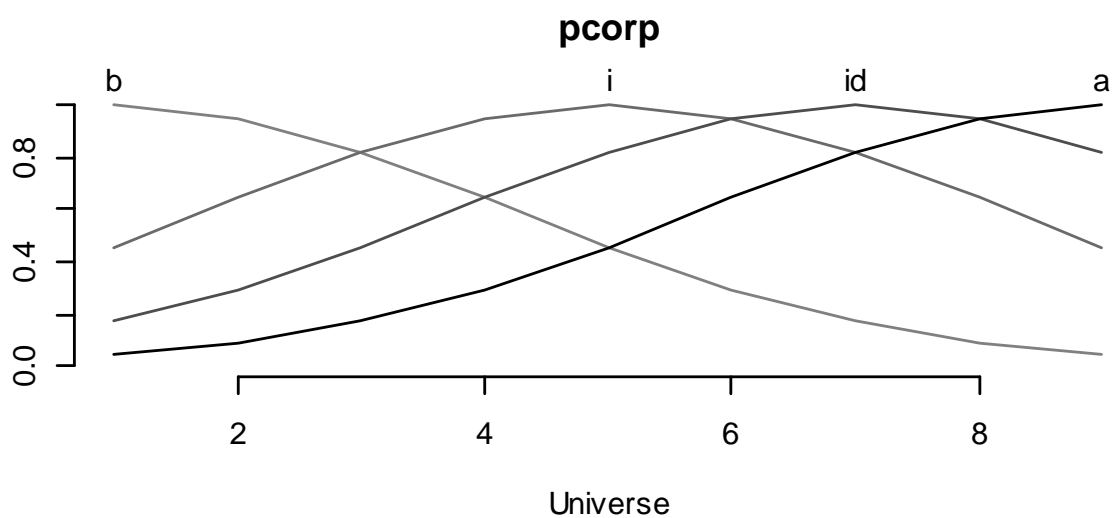
Figura 23 - Gráfico da variável linguística largura do peito (lpeito)



Fonte: Elaborada pelo autor (2016).

À variável “**pcorp**” (acrônimo para profundidade corporal) foram associadas também quatro funções pertinência, as funções, baixo (b), intermediário (i), alto (a) e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 7, em um universo de 1 a 9. Na Figura 24 apresenta-se as funções de pertinência da variável linguística **pcorp** (profundidade corporal).

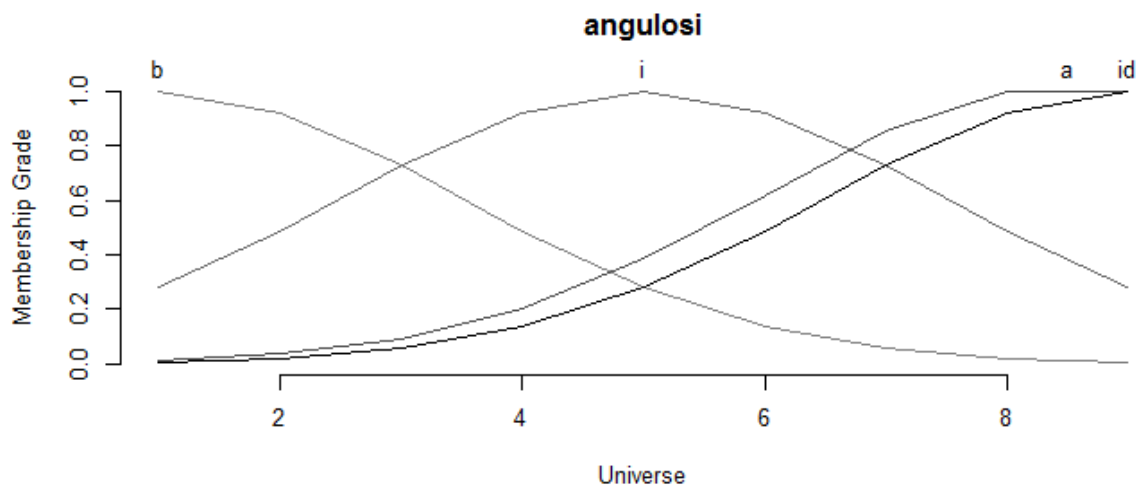
Figura 24- Gráfico da variável linguística Profundidade Corporal (pcorp)



Fonte: Elaborada pelo autor (2016).

À variável “**angulosi**” (acrônimo para angulosidade) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alta (a) e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em um universo de 1 a 9. Na Figura 25 apresenta-se as funções de pertinência da variável linguística **angulosi** (angulosidade).

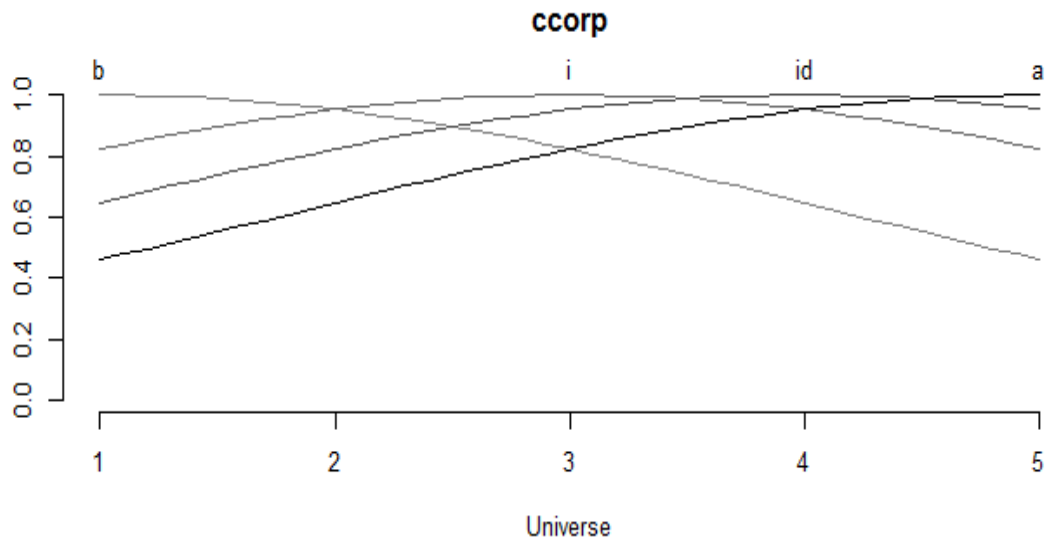
Figura 25 - Gráfico da variável linguística Angulosidade (angulosi)



Fonte: Elaborada pelo autor (2016).

À variável “**ccorp**” (acrônimo para condição corporal), foram associadas também quatro funções pertinência, as funções baixo (b), intermediário (i), alto(a) e ideal (id). A baixa recebe o valor 1, a intermediária o valor 3, alta o valor 5 a e ideal o valor 4, em um universo de 1 a 5. Na Figura 26 apresenta-se as funções de pertinência da variável linguística **ccorp** (condição corporal).

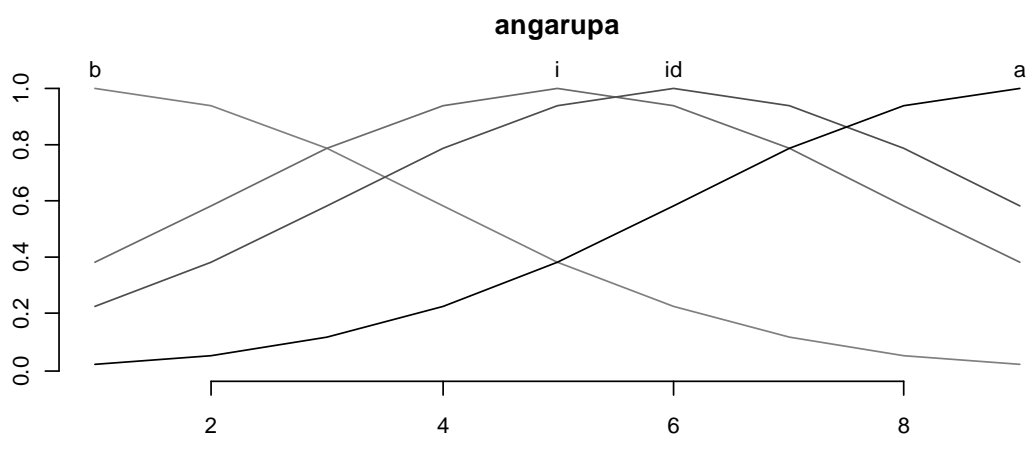
Figura 26 - Gráfico da variável linguística Condição Corporal (ccorp)



Fonte: Elaborada pelo autor (2016).

À variável “**angarupa**” (acrônimo para ângulo da garupa) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 6, em um universo de 1 a 9. Na Figura 27 apresenta-se as funções de pertinência da variável linguística **angarupa** (ângulo da garupa).

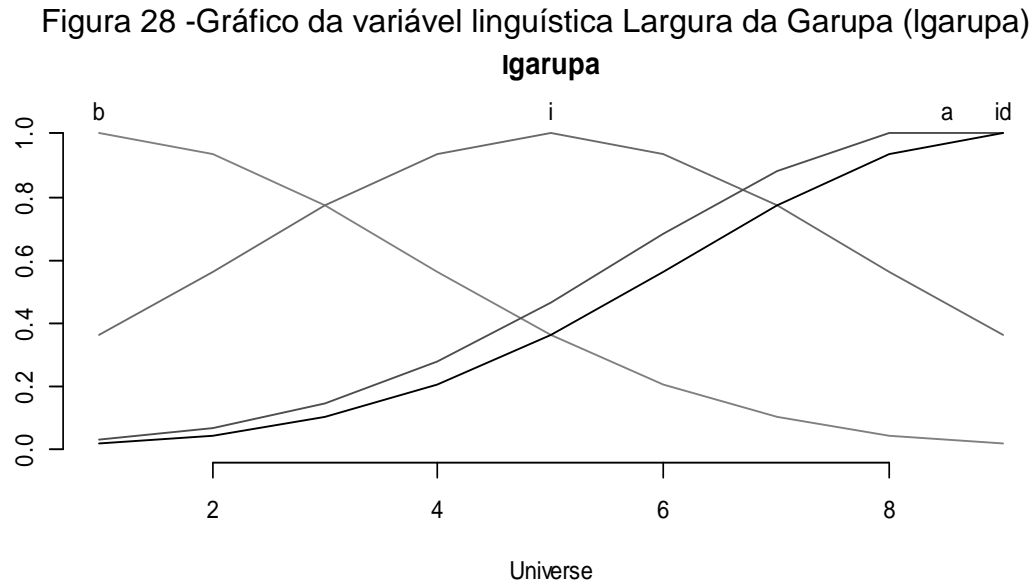
Figura 27 - Gráfico da variável linguística Ângulo da Garupa (angarupa)



Fonte: Elaborada pelo autor (2016).

À variável “**lgarupa**” (largura da garupa) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa

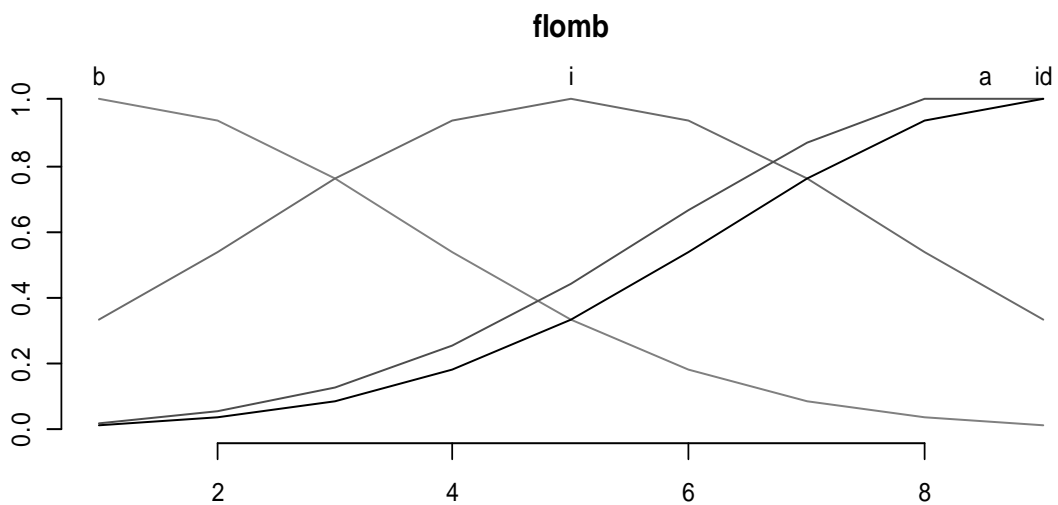
recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em um universo de 1 a 9. Na Figura 28 apresenta-se as funções de pertinência da variável linguística **Igarupa** (largura da garupa).



Fonte: Elaborada pelo autor (2016).

À variável “**flombar**” (força lombar) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em um universo de 1 a 9. Na Figura 29 apresenta-se as funções de pertinência da variável linguística **flombar** (força lombar).

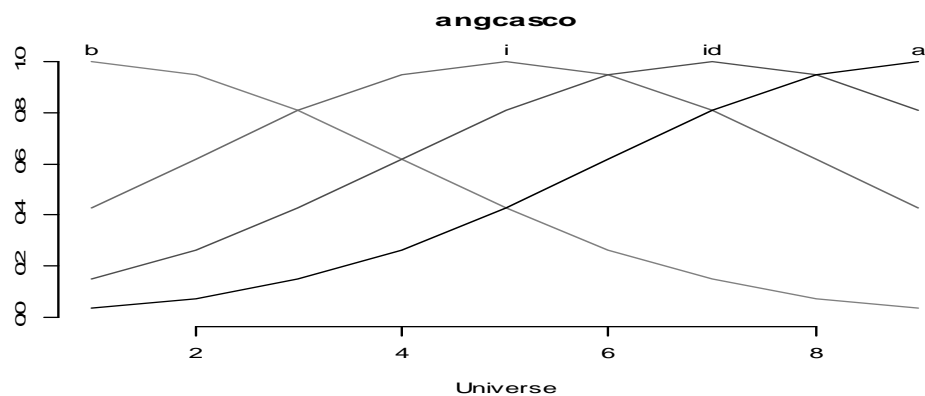
Figura 29 - Gráfico da variável linguística Força Lombar (flomb)



Fonte: Elaborada pelo autor (2016).

À variável “**angcasco**” (ângulo do casco) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 7, em um universo de 1 a 9. Na Figura 30 apresenta-se as funções de pertinência da variável linguística “**angcasco**” (ângulo do caco).

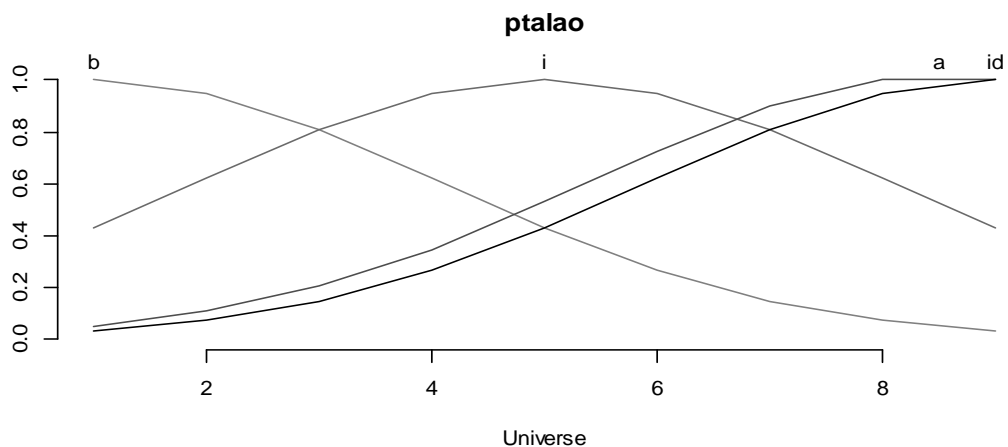
Figura 30 - Gráfico da variável linguística Condição Corporal (ccorp)



Fonte: Elaborada pelo autor (2016).

À variável “**ptalao**” (profundidade do talão) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em um universo de 1 a 9. Na Figura 31 apresenta-se as funções de pertinência da variável linguística **ptalao** (profundidade do talão).

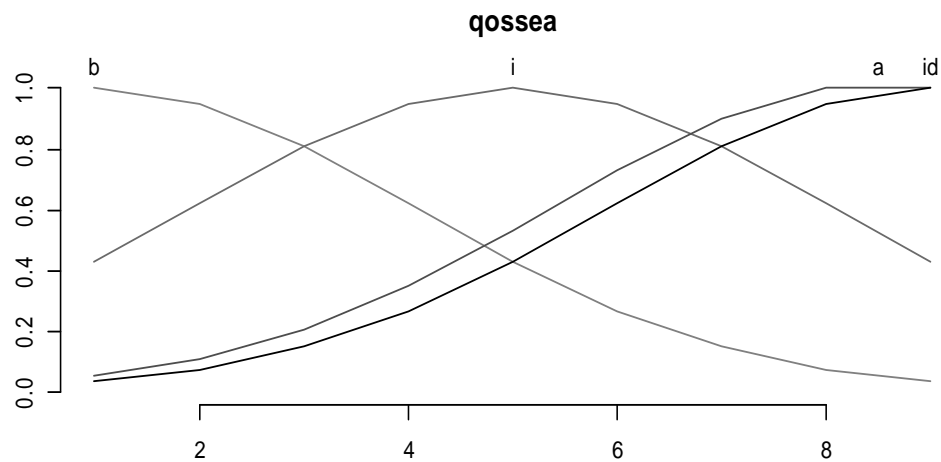
Figura 31 - Gráfico da variável linguística Profundidade do Talão (ptalao)



Fonte: Elaborada pelo autor (2016).

À variável “**qossea**”(qualidade óssea) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em um universo de 1 a 9. Na Figura 32 apresenta-se as funções de pertinência da variável linguística **qossea** (qualidade óssea).

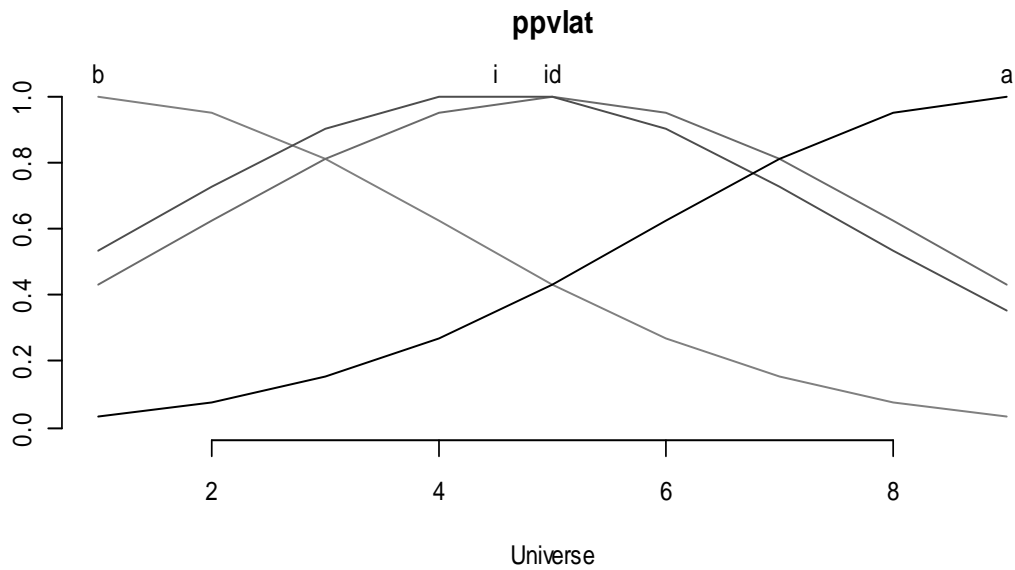
Figura 32 - Gráfico da variável linguística Qualidade óssea (qossea)



Fonte: Elaborada pelo autor (2016).

À variável “**ppvlat**” (pernas e pés vista lateral) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 5, em um universo de 1 a 9. Na Figura 33 apresenta-se as funções de pertinência da variável linguística **ppvlat** (pernas e pés vista lateral).

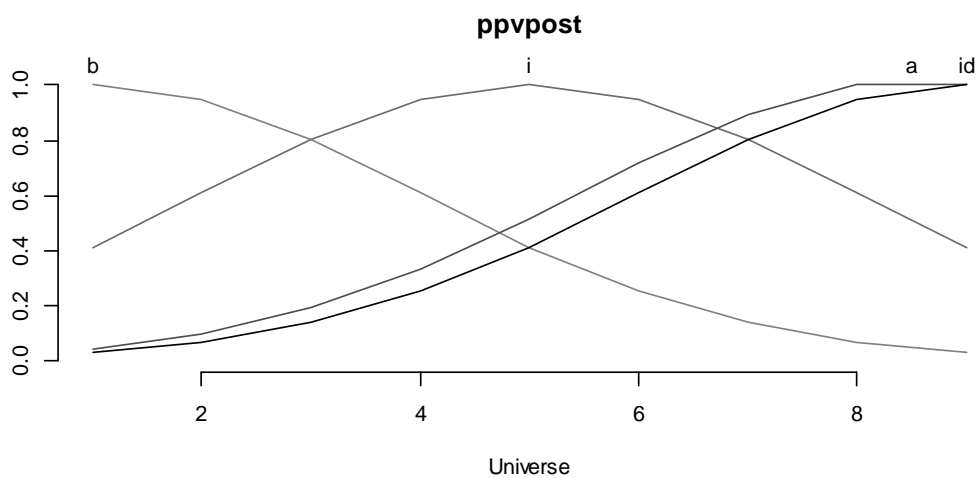
Figura 33 - Gráfico da variável linguística Perna e Pés Vista Lateral (ppvlat)



Fonte: Elaborada pelo autor (2016).

À variável “**ppvpost**” (pernas e pés vista posterior) foram associadas também quatro funções pertinências funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em um universo de 1 a 9. Na Figura 34 apresenta-se as funções de pertinência da variável linguística **ppvpost** (pernas e pés vista posterior).

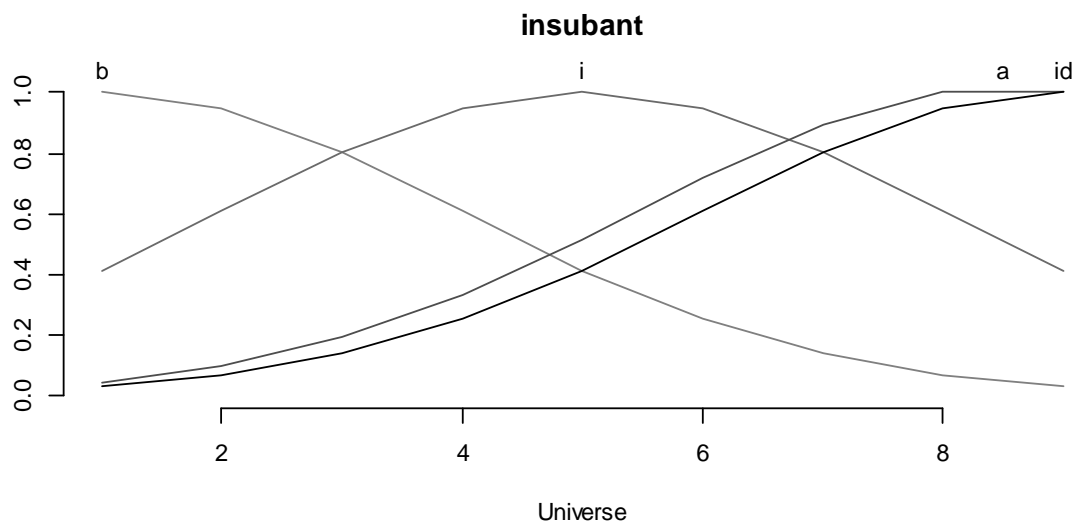
Figura 34 - Gráfico da variável Pernas e Pés Vista Posterior (ppvpost)



Fonte: Elaborada pelo autor (2016).

À variável “**insubant**” (inserção do úbere anterior) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alta (a) e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em um universo de 1 a 9. Na Figura 35 apresenta-se as funções de pertinência da variável linguística **insubant** (inserção do úbere anterior).

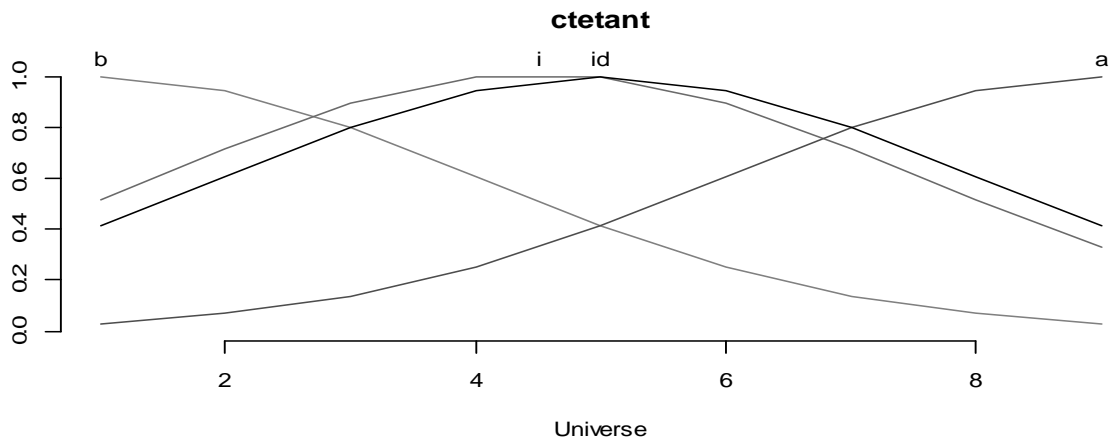
Figura 35 - Gráfico da variável linguística Inserção do Úbere Anterior (insbant)



Fonte: Elaborada pelo autor (2016).

À variável “**ctetant**” (colocação dos tetos anteriores) foram associadas também quatro funções pertinência, as funções baixo, intermediário, alto e ideal. A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 5, em um universo de 1 a 9. Na Figura 36 apresenta-se as funções de pertinência da variável linguística **cteant** (colocação dos tetos anteriores).

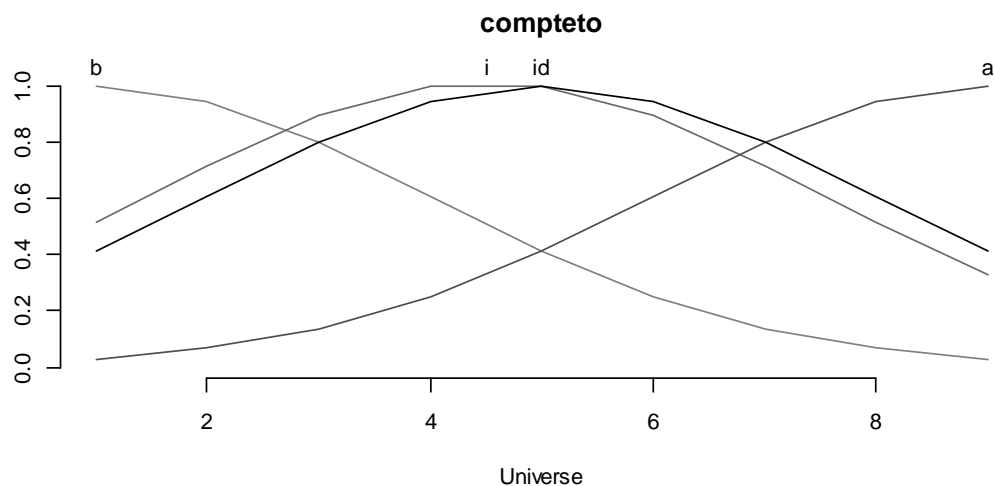
Figura 36 - Gráfico da variável linguística Colocação dos tetos anteriores (ctetant)



Fonte: Elaborada pelo autor (2016).

À variável “**compteto**” (comprimento dos tetos) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 5, em um universo de 1 a 9. Na Figura 37 apresenta-se as funções de pertinência da variável linguística **compteto** (comprimento dos tetos).

Figura 37 - Gráfico da variável linguística Comprimento do Teto (compteto)

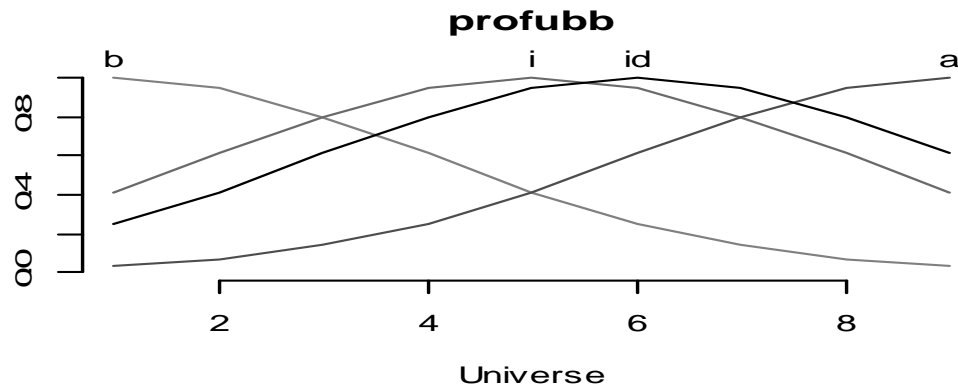


Fonte: Elaborada pelo autor (2016).

À variável “**profubb**” (profundidade do úbere) foram associadas também quatro funções pertinência, as funções baixo, intermediário, alto e ideal. A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 6, em um universo de

1 a 9. Na Figura 38 apresenta-se as funções de pertinência da variável linguística **profubb** (profundidade do úbere).

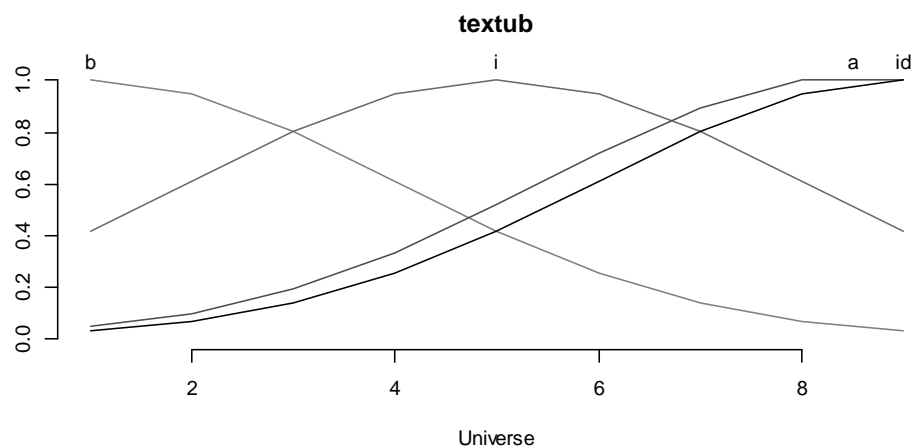
Figura 38 - Gráfico da variável linguística Profundidade do Úbere (profubb)



Fonte: Elaborada pelo autor (2016).

À variável “**textub**” (textura do úbere) foram associadas também quatro funções de pertinência, as funções, baixa (b), intermediária (i), alta e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em um universo de 1 a 9. Na Figura 39 apresenta-se as funções de pertinência da variável linguística **textub** (textura do úbere).

Figura 39 - Gráfico da variável linguística Textura do Úbere (textub)

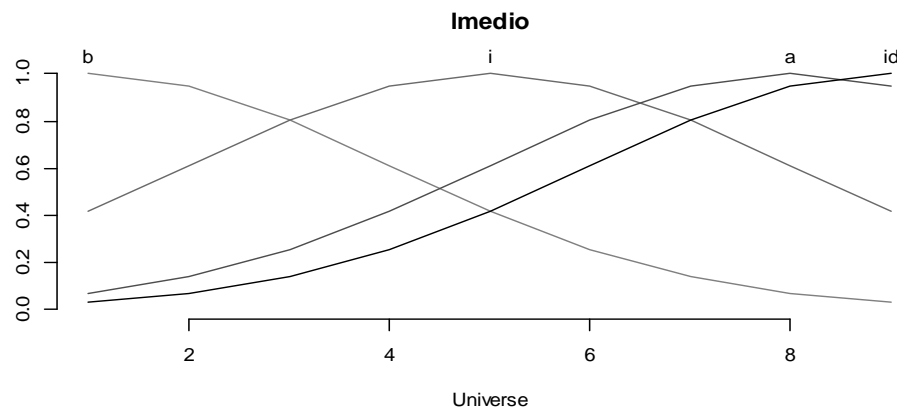


Fonte: Elaborada pelo autor (2016).

À variável “**Imedio**” (ligamento médio) foram associadas também quatro funções de pertinência, as funções, baixa (b), intermediária (i), alta e ideal (id). A baixa recebe o

valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em um universo de 1 a 9. Na Figura 40 apresenta-se as funções de pertinência da variável linguística **Imedio** (ligamento médio).

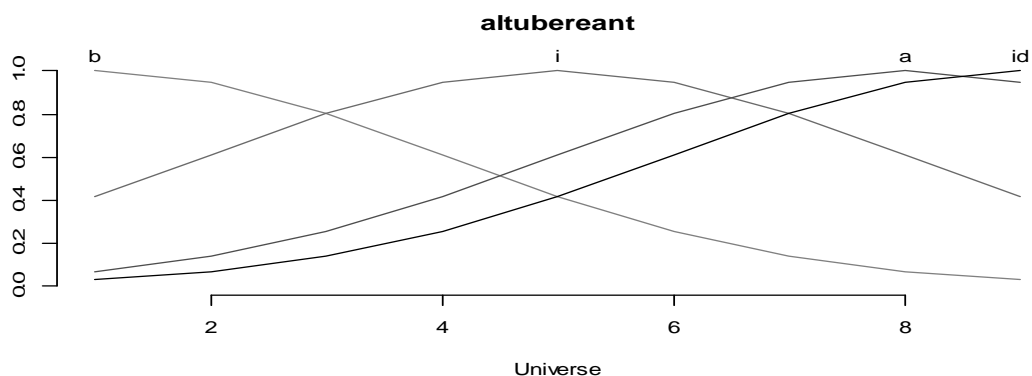
Figura 40 - Gráfico da variável linguística Ligamento médio (Imedio)



Fonte: Elaborada pelo autor (2016).

À variável “**altubereant**” (altura do úbere anterior) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em um universo de 1 a 9. Na Figura 41 apresenta-se as funções de pertinência da variável linguística **altubereant** (altura do úbere anterior).

Figura 41 - Gráfico da variável linguística Altura do Úbere Anterior (altubereant)

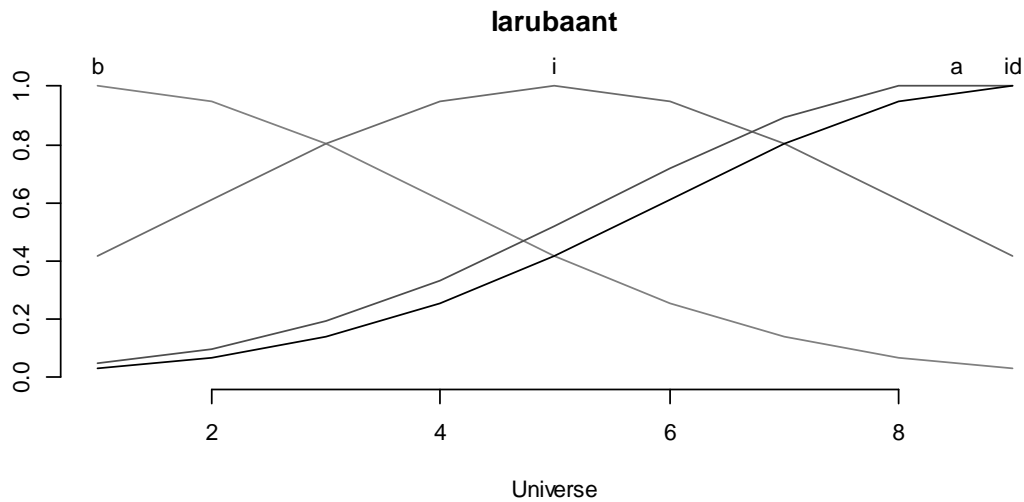


Fonte: Elaborada pelo autor (2016).

À variável “**larubaant**” (largura do úbere anterior) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 9, em

um universo de 1 a 9. Na Figura 42 apresenta-se as funções de pertinência da variável linguística **larubaant** (largura do úbere anterior).

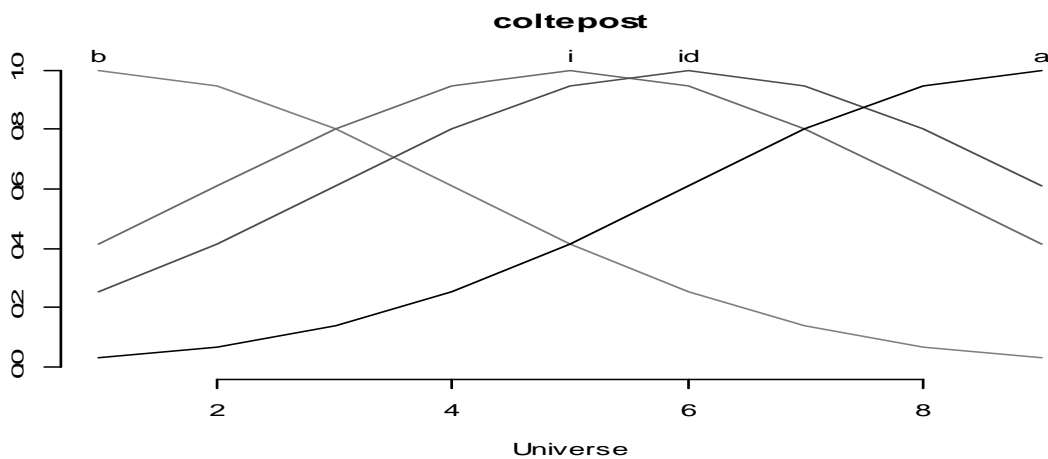
Figura 42 - Gráfico da variável linguística Largura do Úbere Anterior (larubaant)



Fonte: Elaborada pelo autor (2016).

À variável “**coltepost**” (colocação dos tetos posteriores) foram associadas também quatro funções pertinência, as funções, baixa (b), intermediária (i), alto e ideal (id). A baixa recebe o valor 1, a intermediária o valor 5, alta o valor 9 a e ideal o valor 6, em um universo de 1 a 9. Na Figura 43 apresenta-se as funções de pertinência da variável linguística **coltepost** (colocação dos tetos posteriores).


Figura 43 - Gráfico da variável linguística Colocação dos Tetos Posteriores (coltepost)



Fonte: Elaborada pelo autor (2016).

Os gráficos das variáveis linguísticas foram criados na ferramenta R, utilizando a biblioteca **sets**. Para esse conjunto de variáveis com suas funções de pertinência, foram geradas 267 regras, conforme indicado na Figura 44.

Figura 44 – Número de regras geradas e variáveis fuzzy



```

RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help

> print(sistema)
A fuzzy system consisting of 24 variables and 267 rules.

Variables:

angulosi(b, i, a, id)
lpeito(b, i, id, a)
flomb(b, i, a, id)
lgarupa(b, i, a, id)
angarupa(b, i, id, a)
est(b, i, id, a)
ppvpost(b, i, id, a)
insubant(b, i, a, id)
ctetant(b, i, a, id)
compteto(b, i, a, id)
profubb(b, i, a, id)
textub(b, i, a, id)
lmedio(b, i, a, id)
altubereant(b, i, a, id)
larubaant(b, i, a, id)
coltepost(b, i, id, a)
angcasco(b, i, id, a)
ptalao(b, i, a, id)
qossea(b, i, a, id)
ppvlat(b, i, id, a)
nsup(b, i, id, a)
pcorp(b, i, id, a)
ccorp(b, i, id, a)
linear(F, R, BOA, BPM, MBOA, EX)

```

Fonte: Elaborada pelo autor (2016).

Após a geração das variáveis regras do sistema fuzzy, é preciso fazer a inferência do sistema, que consiste na tomada de decisão conforme as regras que irão reger este processo. Para a inferência deste modelo baseado em regras foram utilizados, dados de um animal com suas características com notas consideradas

ideais, assim sendo o sistema deveria tomar a decisão para um animal classificado como excelente, como ilustrado no Quadro 7.

Quadro 7 - Valores ideais de classificação

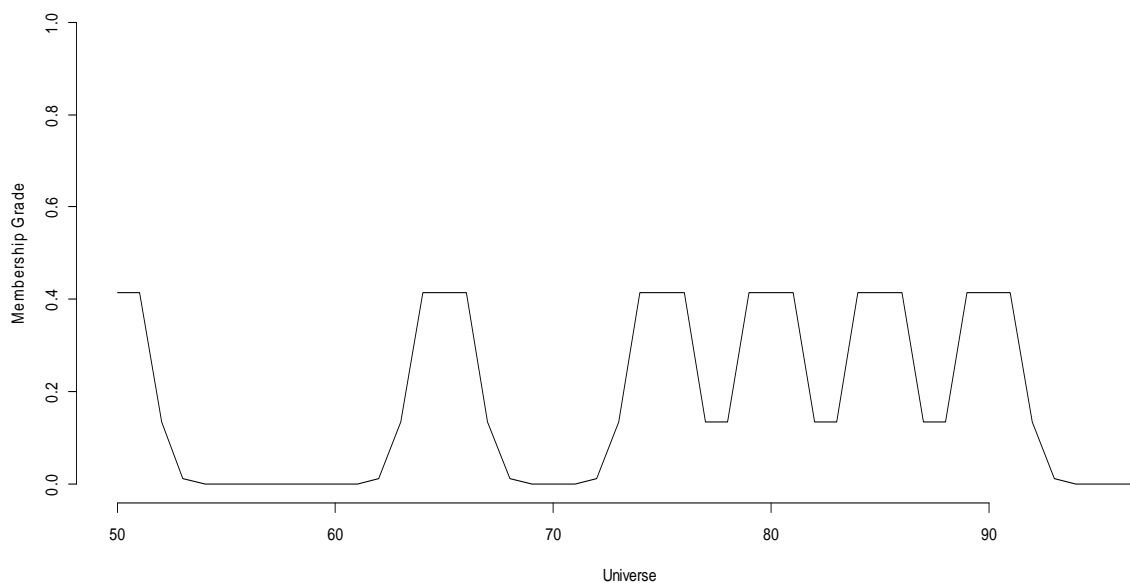
Secção	Área analisada	Valores ideais
Foça leiteira	Estatuta	7
	Nivelamento superior	7
	Largura do peito	7
	Prof. Corporal	7
	Angulosidade	9
	Condição Corporal	4
Garupa	Ângulo da garupa	6
	Largura da Garupa	9
	Força lombar	9
Pernas e Pés	Ângulo do casco	7
	Profundidade do talão	9
	Qualidade óssea	9
	Pernas posteriores vista lateral	5
	Pernas posteriores vista posterior	9
Sistema mamário	Inserção do úbere anterior	9
	Colocação do teto anterior	5
	Comprimento dos Tetos	5
	Profundidade	6
	Textura do úbere	9
	Ligamento médio	9
	Altura úbere posterior	9
	Largura úbere	9
	Colocação dos tetos	6

Fonte: adaptado de Valloto e Neto (2010).

A inferência executada na ferramenta R, feita pelo método `fuzzy_inference()`, recebe uma lista de valores de acordo com as variáveis linguísticas, seguido do método `plot()` para gerar o gráfico de inferência, que é exibido na Figura 45. Para ilustração, o comando usado no R foi

```
inferência <- fuzzy_inference(system, list(est=7, nsup=7, lpeito=7, pcorp=7,
angulosi=9, ccorp=4, angarupa=6, lgarupa=9, flomb=9, angcasco=7, ptalao=9,
qossea=9, ppvlat=5, ppvpost=9, insubant=9, ctetant=5, compteto=5, profubb=6,
textub=9, lmedio=9, altubereant=9, larubaant=9, coltepost=6)).
```

Figura 45 - Regras geradas



Fonte: Elaborada pelo autor (2016).

O gráfico da Figura 45 indica como varia o grau de pertinência do animal com as características consideradas ideais pelas associações de criadores em relação a seu escore total, considerando uma avaliação difusa sobre as 23 características lineares para o tipo leiteiro. O eixo horizontal deste gráfico mostra a pontuação que a vaca pode receber, variando de 50 até 97 pontos, em contrapartida o eixo vertical indica o grau de pertinência da inferência fuzzy, variando de 0 até 1. Os picos indicam as pertinências máximas alcançadas para cada categoria de classificação, iniciando da esquerda para a direita, temos o primeiro pico com a categoria fraca (F), o segundo com regular (R) o terceiro com a boa (B), o quarto com boa para mais (BM), o quinto com muito boa (MB) e o sexto com a categoria excelente (EX). Neste sentido, os vales das curvas indicam o grau de pertinência mínima alcançada para cada uma das categorias de classificação. Observa-se que esse escore fuzzy pode alcançar mais de noventa pontos conforme especificação do modelo de classificação. Segundo o modelo, um animal considerado ideal pode ser visto como um animal que tem o equilíbrio de suas características. No entanto, se aceitas as hipóteses usadas na

fuzzificação sobre a forma de classificação (subjetiva) para cada critério, observa-se que o animal 'ideal' pode obter escores baixos, sendo classificado como regular, boa, boa para mais, muito boa ou excelente, conforme apresentado anteriormente no Quadro 5. Essa percepção faz inferir que o sistema de classificação pode ser falho, se depender apenas de classificações subjetivas, reforçando o uso de abordagens como a de regras de associação apresentada anteriormente.

5 CONSIDERAÇÕES FINAIS

Em geral, a classificação de bovinos leiteiros no estado do Tocantins é uma prática não muito comum, sendo que o foco da produção agropecuária na região é voltado para a criação de bovinos de corte, tendo a produção leiteira como meio secundário as demais a serem exploradas na região. As práticas de classificação são bem mais aplicadas em instituições de ensino na formação de técnicos em agropecuária como é o caso do Instituto Federal do Tocantins que funciona como um difusor de tecnologias e técnicas para desenvolvimento regional nesta área e de poucas propriedades rurais que também utilizam essa técnica por intermédio de profissionais formados nesta instituição.

Desta forma o modelo vem ao encontro da necessidade da instituição e da região no que concerne à avaliação individualizada dos animais do plantel. Um modelo de classificação que possua uma interface amigável e intuitiva para com o usuário e uma parte lógica que utiliza métodos estatísticos e de mineração de dados para garantir um grau de maior confiabilidade aos resultados gerados.

A utilização da lógica fuzzy como saída do resultado de classificação possibilitou perceber o impacto que a avaliação subjetiva pode ter sobre os resultados da classificação, reforçando a importância de usar outras formas de classificação, como a baseada em regras de associação. Porém, o modelo proposto visa auxiliar a tomada de decisão quanto à escolha da melhor vaca para característica leiteira, sendo importante ressaltar que passa a ser uma proposta que auxilia e não substitui a função do avaliador, pois o mesmo tem parte importante neste processo.

O acompanhamento individual do animal leiteiro possibilita ao produtor verificar para os próximos cruzamentos quais características devem ser mantidas para o rebanho e quais devem ser atacadas para melhoramento do mesmo. O modelo também traz a alternativa de um acompanhamento geral do rebanho, podendo retornar uma média de classificação, e da mesma forma identificar por meio de gráficos e valores, quais animais estão fora do padrão aceitável ou os que estão acima deste.

As regras de associação encontradas por meio da análise da base de dados podem, portanto, servir para indicar outras formas de classificar animais, bem como servir como base para estudos futuros.

REFERÊNCIAS

- ABEPRO. Disponível em: <<http://www.abepro.org.br/interna.asp?c=362>>. Acesso em: 12. maio. 2016.
- AGRAWAL, Rakesh; SRIKANT, Ramakhishnan. **Fast algorithms for mining association rules**. Disponível em: <https://www.it.uu.se/edu/course/homepage/infoutv/ht08/vldb94_rj.pdf>. Acesso em: 05. jul. 2016.
- BARANAUKAS, J. A. Regras de Associação Introdução. ,São Paulo, 2013.
- BERRY, D. P. et al. Phenotypic associations between traits other than production and longevity in New Zealand dairy cattle. **Journal of dairy science**, v. 88, n. 8, p. 2962-2974, 2005.
- BERRY, D.P. **Phenotypic relationships between type traits and longevity in New Zealand**. Annual Meeting of the EAAP, 56th., June 5-8, 2005, Uppsala, Sweden. Session CG2.23.
- BRESSAN, A. A. Tomada de decisão em futuros agropecuários com modelos de previsão de séries temporais. **RAE eletrônica**, v. 3, n. 1, p. 1–20, 2004.
- BRIN, Sergey et. al. **Dynamic Itemset Counting and Implication Rules for Market Basket Data**. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.6476&rep=rep1&type=pdf>> . Acesso em: 16. jul. 2016.
- CAMILO, C.; SILVA, J. **Mineração de Dados: Conceitos, tarefas, métodos e ferramentas**. Universidade Federal de Goiás (UFG), p. 29, 2009. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf>. Acesso em: 3.jun. 2016.
- CAMPOS, Rafael Viégas. **Parametros genéticos para características lineares de tipo e produtivas em vacas da raça holandesa no Brasil**. Porto Alegre: UFRGS, 2012. (Tese de Doutorado)
- CASTRO, R.P. de. Sistema Linear de Classificação In: Peixoto, A.M., Moura, J.C. de, Faria, V.P. de. **Bovinocultura leiteira: fundamentos da exploração racional**. Piracicaba: FEALQ, 1993. p. 431-453.
- CHAVES, Viviane Hengler Corrêa. **Perspectivas históricas da Pesquisa Operacional**. Rio Claro/SP: UESP, 2011. (Dissertação de Mestrado)
- COLBEK, Adelmo Paulo. **Uso de uma ferramenta de gestão rural na agricultura familiar: um estudo de caso no município de Sede Nova/RS**. 2011.
- DANTAS, R. F.; SOUZA, S. A. de. Modelo de risco e decisão de crédito baseado em estrutura de capital com informação assimétrica. **Pesquisa Operacional**, v. 28, n. 2, p. 263–284, 2008.

ESTEVEES, A. M. C.; BERGMANN, J. A. G.; DURÃES, M. C.; COSTA, C. N.; SILVA, H. M. Estudo dos efeitos de ambiente sobre as características lineares de tipo em rebanhos bovinos da raça Holandesa. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 56, n. 4, p. 522–528, 2004

ESTEVEES, A. M. C.; BERGMANN, J. A. G.; DURÃES, M. C.; COSTA, C. N.; SILVA, H. M. Correlações genéticas e fenotípicas entre características de tipo e produção de leite em bovinos da raça Holandesa. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 56, n. 4, p. 529–535, 2004.

FALCÃO, Alencariano José da Silva et. al. Interação genótipo-ambiente na produção de leite de vacas da raça Holandesa. **Acta. Sci. Anim. Sci. Maringá**, v. 30, n.2, p. 225-231, 2008.

FARIA, Carina Ubirajara de Faria et. al. Análise genética de escores de avaliação visual de bovinos com modelos bayesianos de limiar e linear. **Pesquisa Agropecuária Brasileira**, Brasília, v.43, n.7, p.835-841, jul. 2008.

FELICIANO, Rafaelle Aguiar Correia. **Controle fuzzy espacialmente diferenciado para um sistema de irrigação**. Natal/RN: Universidade Federal do Rio Grande do Norte, 2012. (Dissertação de Mestrado)

FELIPE, V. P. S. **Redes Neurais Artificiais** : Revisão e Aplicação para predição de características complexas, 2012.

GOLDSCHIMIDT, R; PASSOS, E. **Data mining**: Um guia prático. Rio de Janeiro: Campus, 2005.

GOMES, Diogo da Silva Magalhães; RODRIGUES, Max de Castro O. **Introdução a lógica fuzzy com java**. 2016. Disponível em: <http://www.devmedia.com.br/articles/viewcomp_forprint.asp?comp=32444>. Acesso em: 22 abr. 2016.

GONÇALVES, E. C. ; PLASTINO, A. Mining Strong Associations and Exceptions in the STULONG Data Set. In: ECML/PKDD 2004 - 15th European Conf. on Machine Learning. **8th European Conf. on Principles and Practice of Knowledge Discovery in Databases**, 2004, Pisa. Proceedings of the Discovery Challenge 2004, 2004. p. 44-55.

GONÇALVES, E. C. **Data Mining de Regras de Associação**. , 2014. Disponível em: <<http://www.devmedia.com.br/data-mining-de-regras-de-associação-parte-1/6533>>. Acesso em: 6. Maio. 2016.

GONÇALVES, EDUARDO CORRÊA. **Regras de Associação e suas Medidas de Interesse Objetivas e Subjetivas**. Niteroi/RJ: Universidade Federal Fluminense, 2005.

MACEDO, Dayana Carla de; MATOS, Simone Nasser. Extração de conhecimento através da mineração de dados. **Revista de Engenharia e Tecnologia**, v. 2, n. 2, p.22-30, ago. 2010.

MANUAL de classificação morfológica. Disponível em: <http://www.apcrf.pt/fotos/editor2/manual_i.pdf>. Acesso em: 8. jun. 2016.

MCMANUS, C.; SAUERESSIG, M. G. Study of Linear Type Traits in Holstein-Friesian Cattle in Total Confinement in the Distrito Federal, **Revista brasileira de zootecnia**. V.25, n.5, p. 906–915, 1998.

MIGUEL, Paulo Augusto Cauchick et al. **Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações**. 2. ed. Rio de Janeiro: Elsevier, 2012. 277 p.

OLIVEIRA, Kelinne. Campus Araguatins realiza formatura de estudantes dos cursos técnicos. Disponível em: <<http://araguatins.ifto.edu.br/portal/index.php/ultimas-noticias/410-campus-araguatins-realiza-formatura-de-estudantes-dos-cursos-tecnicos>>. Acesso em: 10. jun. 2016.

PERISSINOTTO, Mauricio; DE MOURA, Daniella Jorge; DA CRUZ, Vasco Fitas. Applied Intelligent System for Environmental Control in Dairy Housing. In: **Livestock Environment VIII, 31 August–4 September 2008, Iguassu Falls, Brazil**. American Society of Agricultural and Biological Engineers, 2009. p. 128.

PIATETSKY-SHAPIRO, G. Discovery, Analysis and Presentation of Strong Rules: knowledge Discovery in Databases. [s.l.]: AAAI/MIT Press, 1991.

POLYCARPO, Rafaela Carareto. **Zootecnia de Precisão na Pecuária Leiteira: Ficção científica ou necessidade para o futuro?** Disponível em: <<http://www.milkpoint.com.br/radar-tecnico/sistemas-de-producao/zootecnia-de-precisao-na-pecuaria-leiteira-ficcao-cientifica-ou-necessidade-para-o-futuro-75979n.aspx>>. Acesso em: 20. jul. 2016.

POSSELT, Ederson Luis; FROZZA, Rejane; MOLZ, Rolf Fredi. INFUZZY: ferramenta para desenvolvimento de aplicações de sistemas difusos. **Revista Brasileira de Computação Aplicada**, v. 7, n. 1, p. 42-52, abr. 2015.

RENNÓ, Francisco Palma et. al. Correlações Genéticas e Fenotípicas entre Características de Conformação e Produção de Leite em Bovinos da Raça Pardo-Suíça no Brasil. **Revista Brasileira de Zootecnia**. v.32, n.6, p.1419-1430, 2003.

REZENDE, Solange O.; MARCACINI, Ricardo M.; MOURA, Maria F. Moura. O uso da Mineração de Textos para Extração e Organização Não Supervisionada de Conhecimento. **Revista de Sistemas de Informação da FSMA**, n. 7, P.7-21. 2011.

RIGNEL, DG de S.; CHENCI, G. P.; LUCAS, C. A. Uma introdução a Lógica Fuzzy. **Revista Eletrônica de Sistemas de Informação e Gestão Tecnológica**, v. 1, n. 1, p. 17-28, 2011.

SANCHES, Marcelo Kaminski; MONARD, Maria Carolina. Proposta de um Algoritmo de Clustering Semisupervisionado para Rotular Exemplos a Partir de Poucos Exemplos Rotulados. In: Workshop in Artificial Intelligence, Arica-Chile. **Jornadas Chilenas de Computación**. Chile: Sociedad Chilena de Ciências de la Computación. 2004. p. 1-9.

SANTINI, Paulo Henrique; TELLES, Marcelo Josué. **Mineração de Dados com WEKA**. 2008. Disponível em: <<http://www.devmedia.com.br/mineracao-de-regras-de-associacao-com-a-ferramenta-de-data-mining-weka/20478>>. Acesso em: 28 mar. 2016.

SHORT, T. H.; LAWLOR, T. J. Genetics And Breeding: Genetic Parameters of Conformation Traits, Milk Yield, and Herd Life in Holsteins. **Journal of Dairy Science**, Vol. 75, No. 7, 1992.

SILVA, Gaspar Ferreira da et. al. O Projeto Balde Cheio: estudo de caso no Campus Araguatins-TO. In: Congresso Norte e Nordeste de Pesquisa e inovação, 7., 19 A 21 de outubro, 2012. Palmas. Ciencia, Tecnologia e inovação: ações sustentáveis para o desenvolvimento regional. **Anais...** Palmas, 2012.

SILVA, Michel de A.; TREVISAN, Daniela; PRATA, David N.; MARQUES, Elineide E. **Aplicação do algoritmo Apriori para uma base de dados de ictioplâncton em um reservatório de água doce da Amazônia Legal**. Palmas: UFT, 2014.

SILVA, Natascha Almeida Marques da. **Seleção de modelos de regressão não lineares e aplicação do algoritmo saem na avaliação genética do crescimento de bovinos nelore**. Belo Horizonte: UFMG, 2010. (Tese de Doutorado)

TEODORO, Roberto Luiz. et. al. Estudo de Características do Sistema Mamário e suas Relações com a Produção de Leite em Vacas da Raça Gir. **Revista brasileira de Zootecnia**, v. 29, pag. 131-135, 2000.

VALLOTO, Altair Antonio. Disponível em: <<http://www.gadolando.com.br/site/pagina.php?codPag=213>>. Acesso em: 12. maio. 2016.

VALLOTO, Altair Antonio; RIBAS NETO, Pedro Guimarães. **Raça Holandesa moderniza e atualiza sistema de avaliação da conformação das vacas (classificação para tipo)**. Associação Paranaense de criadores de bovinos da raça holandesa. Disponível em: <www.holandeparana.com.br>. Acesso em: 16.02.2015.

VENTURA, R. V. et al. Use of artificial neural networks in breeding values prediction for weight at 205 days in Tabapuã beef cattle. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 64, n. 2, p. 411-418, 2012.

VENTURINI, Carlos Eduardo Pullis. A geografia do leite brasileiro. Disponível em: <<http://www.milkpoint.com.br/cadeia-do-leite/artigos-especiais/a-geografia-do-leite-brasileiro-87327n.aspx>>. Acesso em: 05. jun. 2016.

VIDAL, Leonardo de Carvalho; LANDIM, Wellington Carlos de Almeida; COSTA, Alessandro Pires. Aplicação de Lógica Fuzzy no Controle de Plantas Industriais. In: **IX Simpósio de Excelência em Gestão e Tecnologia**, 15f., 2012.

XAVIER, Sandro Luiz da Paixão. **Samira: uma proposta de sistema de apoio à mineração de regras de associação**. Fortaleza: Universidade Estadual do Ceará, 2010. (Dissertação de Mestrado).