aper:180194_1

Aplicações de Mineração de Dados na Pecuária de Corte: Previsão de Indicadores de Qualidade de Carcaças

Rodrigo R. da Silva[1], Thales V. Maciel[1], Vinícius do N. Lampert[2], Denizar S. de Souza[3]

[1]Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense (IFSUL)
Campus Bagé – Av.Leonel de Moura Brizola, 2501 – 96.418-400 – Bagé – RS – Brasil

[2]Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)
Unidade Pecuária Sul - CPPSUL – BR 153, Km 603 – Bagé – RS – Brasil

[3]Centro de Ciências Exatas e Aplicadas (CCEA)
Universidade da Região da Campanha (URCAMP) – Bagé – RS – Brasil

orki2008@gmail.com, thalesmaciel@ifsul.edu.br

vinicius.lampert@embrapa.br, denizarsouza@urcamp.edu.br

Abstract. *Considering that cattle breeders are able to acquire influence variables along the breeding process, this paper aims to provide a method for predicting carcasse bonus, daily average weight gain, age at slaughter and weight at slaughter based on influce variables to be collected until the bovine ablactation. For such, data mining applications were performed through linear regression applied to a 167 bovine instances dataset. Obtained results showed that carcasse bonus and daily average weight gain may be predicted with zero or insignificant error, maywhile age and weight at slaughter produced higher error rates upon prediction.*

Resumo. *Considerando que o produtor rural pode obter algumas variáveis de influência ao longo do processo produtivo do gado de corte, objetiva-se prever se as variáveis de influência obtidas até o desmame dos bovinos podem explicar a bonificação, ganho médio diário, idade de abate e peso de fazenda. Para tanto procede-se a mineração de dados através da regressão linear, em um conjunto de dados de 167 bovinos. Deste modo, observa-se que para a bonificação e ganho médio diário os modelos gerados apresentaram erros baixos, enquanto que para idade de abate e peso de fazenda os erros foram maiores, o que permite concluir que os atributos não foram o suficientes para predizer a idade de abate e peso de fazenda, mas bons para a bonificação e ganho médio diário.*

Introdução

No contexto da pecuária de corte, o sistema produtivo pode ser conceituado como um conjunto de tecnologias e práticas de manejo. Bem como o perfil do animal, a intenção da criação, a raça ou grupamento genético e a região onde a atividade é desenvolvida [Euclides Filho 2000].

Segundo [Gomes et al. 1997], os negócios agropecuários revestem-se da mesma complexidade e dinâmica dos demais setores da economia, requerendo do produtor uma

nova visão diferenciada dos seus negócios, principalmente pela necessidade de se distanciar da posição de fazendeiro tradicional para assumir o papel de empresário rural.

Para analisar um sistema produtivo da pecuária de corte, é indispensável mensurar seus indicadores de qualidade, pois somente assim o produtor rural terá embasamento para tomada de decisão. Para [Oiagen 2010] a mensuração e análise de indicadores que retratam o funcionamento rural são fundamentais para a tomada de decisão. Estes indicadores, de acordo com [Oiagen and Barcellos 2008], são conhecidos como variáveis de influência, ou seja, informações gerenciais de ordem técnica ou econômica que contribuem com avaliações precisas dos processos internos da propriedade rural.

Ainda segundo [Oiagen 2010], deve ficar claro que para a empresa rural, interessa, sobretudo, a rentabilidade, que é o elemento mais importante na avaliação da atividade econômica praticada em moldes capitalistas. Este indicador de desempenho deve situar-se em nível adequado para que o investimento se justifique. No âmbito do criador e das informações que estão acessíveis a ele, os indicadores devem possuir relevância para serem aplicados nos casos de estudos de caso.

O problema de pesquisa abordado neste trabalho é "existem variáveis de cria, ou seja, dados coletados sobre indivíduos de rebanhos bovinos entre nascimentos e desmames, que explicam bons indicadores de qualidade zootécnicas?". A hipótese é que os dados mês de nascimento, mês de desmame, peso de desmame e idade de desmame têm correlação suficiente com o peso de fazenda e idade de abate para explicar bons valores em tais indicadores. Adicionalmente, ganho médio diário de peso e bonificação também são investigados.

O objetivo deste trabalho é descobrir a relação estatística entre as variáveis de cria e os indicadores zootécnicos de qualidade de carcaças após abate. Quantificar o peso dos atributos e hipóteses dos respectivos domínios de valores nos indicadores de qualidade inferidos. Para tal, foram realizadas tarefas de mineração de dados no âmbito de descoberta de conhecimento em banco de dados. O foco da atividade ocorreu com experimentos de regressão, conforme descrito na metodologia.

Referencial Teórico

Nesta seção é apresentado um referencial teórico sobre descoberta de conhecimento com mineração de dados, seguido de um levantamento de suas aplicações na pecuária de corte.

Descoberta de Conhecimento em Banco de Dados

Observa-se que uma grande quantidade de dados cresce de forma acelerada em diversos campos de conhecimento, fato que dificulta a sua interpretação, pois o volume destes dados é maior que o poder de interpretá-los [Vieira and Oliveira 2014]. Desta forma, surgiu a necessidade do desenvolvimento de ferramentas e técnicas automatizadas para minimizar esta situação, as quais pudessem auxiliar o analista a transformar os dados em conhecimento [Han et al. 2011].

Grande parte dessas técnicas e ferramentas podem ser encontradas no processo de descoberta de conhecimento em bases de dados (DCBD). Segundo [Fayaad et al. 1996], DCBD é definida como um processo não trivial que busca identificar padrões novos, potencialmente úteis, válidos e compreensíveis, com o objetivo de melhorar o entendimento de um problema ou um procedimento de tomada de decisão.

O processo de DCBD compreende três principais etapas: pré-processamento, mineração de dados e pós-processamento [Tan et al. 2005]. No pré-processamento os dados são coletados e tratados para serem utilizados nas próximas etapas. A limpeza e a remoção de dados ruidosos também ocorre no pré-processamento, visando assegurar a qualidade dos dados selecionados. Subsequentemente, ocorre a mineração de dados, que são processos aplicados para explorar e analisar os dados em busca de padrões, previsões, erros, associações entre outros [Amaral 2016]. A etapa final consiste no pós-processamento, que engloba a interpretação dos padrões descobertos e a possibilidade de retorno a qualquer um dos passos anteriores. Assim, a informação extraída é analisada (ou interpretada) em relação ao objetivo proposto, sendo identificadas e apresentadas as melhores informações [Corrêa and Sferra 2003]. As tarefas de mineração de dados podem ser dividas em quatro grupos: classificação, regressão, agrupamentos e regras de associação.

A regressão é um tipo específico de classificação. Enquanto a classificação trata de previsão de valores nominais ou categóricos, chamados de classes, a regressão mantém o objetivo de realizar previsões, mas tem como alvo valores numéricos. No agrupamento não existe classe, o objetivo é criar grupos e atribuir instâncias a estes grupos a partir de características, ou atributos destas instâncias. Regras de associação buscam relações entre os ítens, gerando regras que determinam a associação entre esses ítens [Amaral 2016]. Este estudo tem foco em tarefas de regressão.

Revisão dos Trabalhos Correlatos

No âmbito da pecuária de corte, foram identificados trabalhos relacionados ao problema investigado nesta pesquisa.

No trabalho de [Mota et al. 2017] foi proposta uma abordagem de análise de dados com *data warehouse*, consultas analíticas online e mineração de dados, auxiliando o produtor na tomada de decisão do melhor momento para o abate. A abordagem se divide em 4 etapas: 1) responsável pela extração, transformação e carga dos dados; 2) etapa de criação do modelo multidimensional para armazenagem dos dados; 3) etapa de visualização e exploração dos dados armazenados no *data warehouse*; e 4) a aplicação de algoritmos de *data mining* por meio da ferramenta Weka. Na quarta etapa, há indícios de que a adoção de algoritmos de *data mining* fornecem uma taxa média de acerto acima de 62% em relação à predição do grau de acabamento e do rendimento de carcaça.

Já no estudo de [Costa 2016] foram analisados um conjunto de características zootécnicas para gerar um modelo afim de prever o rendimento dos bovinos, através das variáveis peso de fazenda (PF) e bonificação (BN). Para tanto o autor utilizou a técnica de Redes Neurais Artificiais (RNA's). Segundo aponta o autor, o resultado para o modelo de previsão de bonificação apresentou erro bem elevado, baixa correlação e generalização insatisfatória devido a uma limitação da ferramenta e da escolha dos dados utilizados na matriz de entrada da rede. Cabe ressaltar o trabalho não proveu comparações de desempenho com outros métodos de inferência de dados, tampouco indicações de peso de cada variável de entrada no produto de saída.

O trabalho difere-se dos demais por usar a tarefa de regressão no como técnica de processamento na descoberta de conhecimento, além disto, os atributos utilizados são diferentes, pois neste trabalho optou-se por analisar a influência das variáveis de cria em relação as variáveis de qualidade zootécnicas, contribuindo desta maneira para novas

abordagens, relatos e discussões sobre a temática da pecuária de corte.

Metodologia

O conjunto de dados analisado foi constituído por 167 instâncias de animais bovinos da raça Hereford. As nomenclaturas e respectivas descrições dos atributos do conjunto analisado são apresentadas na Tabela 1.

Tabela 1. Atributos utilizados

Nomenclatura	Tipo de Dado	Descrição
abate_peso	Numerico	Peso de abate na fazenda
nascimento_mes	Nominal	Mês de nascimento (1,8,9,10,11,12)
abate_idade	Numerico	Idade de abate
desmame_idade	Numerico	Idade de desmame
desmame_mes	Nominal	Mês de desmame (1,4,5)
desmame_peso	Numerico	Peso de desmame
gmd	Numerico	Ganho médio diário de peso
diff_abate_desmane	Numerico	Diferença entre abate/desmame
bonificação	Numerico	Bonificação

Como ferramenta para a realização das tarefas de pré-processamento e aplicações das tarefas de mineração de dados, foi utilizados o *software Waikato Environment for Knowledge Analysis* (WEKA), um ambiente para análise de conhecimento desenvolvido pela Universidade de Waikato, Nova Zelândia [Hall et al. 2009]. O WEKA tem como objetivo agregar algoritmos provenientes de diferentes abordagens dedicando-se ao estudo de aprendizagem de máquina. O grande número de algoritmos de aprendizado de máquina implementados pela WEKA é um dos maiores benefícios de usar a plataforma.

O experimento realizado dividiu-se em três etapas. Na primeira etapa os dados foram recuperados em formato .CSV afim de serem utilizados no *software* WEKA, o conjunto de dados original constava com 53 atributos e 1015 instâncias de bovinos de diversas raças. A etapa de pré-processamento deste conjunto de dados contou com tarefas de transformação, remoção de atributos irrelevantes, remoção atributos com dados faltantes e dos que não faziam parte do escopo dos experimentos, resultando no conjunto de dados descritos pela Tabela 1, realizou-se o calculo da diferença entre a data de abate e a data de desmame. Após o WEKA ser alimentado com os dados, foi aplicado o filtro *weka.filters.unsupervised.attribute.NumericToNominal* sobre os atribuitos desmame_mes e nascimento_mes de modo que os dados foram convertidos no formato numérico para nominal, afim de evitar que na forma numérica os meses constituíssem pesos quem afetassem os modelos descobertos.

A segunda etapa consistiu no processamento do conjunto de dados, que ocorreu com a tarefa de regressão linear, através do algoritmo *weka.classifiers.functions.LinearRegression* [Witten et al. 2016]. A regressão linear é utilizada basicamente com duas finalidades, prever o valor de y a partir do valor de x e estimar quanto x influencia ou modifica y. Adotou-se este algoritmo pois ele gera um modelo de comportamento, também produz o valor da correlação entre os atributos utilizados nos experimentos e o atributo alvo. Além disso, só usa as colunas que contribuem estatisticamente para a precisão, descartando e ignorando as colunas que não

ajudam a criar um bom modelo. Foram executados testes para cada uma das 4 variáveis alvo. Tabela 2 apresenta os atributos selecionados para cada um dos experimentos.

Tabela 2. Variáveis utilizadas nos experimentos

Variáveis	Bonificação	Peso de fazenda	Idade de abate	Ganho médio diário
Idade de desmame				
Mês de desmame				
Peso de desmame				
Mês de nascimento				
Diferença abate/desmane			Removido	
Bonificação	————	Removido	Removido	Removido
Peso de Fazenda	Removido	————	Removido	Removido
Idade de abate	Removido	Removido	————	Removido
Ganho médio diário	Removido	Removido	Removido	————

Na Figura 1 observa-se os modelos descobertos para os quatro experimentos realizados. O modelo é o resultado gerado pela tarefa de regressão linear. Nele, os atributos relevantes têm pesos atribuídos, de forma a comporem uma fórmula matemática para o cálculo do atributo alvo.

```
abate_peso =

    22.6907 * nascimento_mes=11,10,1 +
    51.6212 * nascimento_mes=1 +
     0.5664 * desmame_peso +
     0.3494 * desmame_idade +
    34.22   * desmame_mes=1 +
     0.1935 * diff_abate_desmame +
   153.0857
```

(a) PF

```
gmd =

     0.0335 * nascimento_mes=11,10,9,1 +
    -0.0163 * nascimento_mes=9,1 +
     0.0675 * nascimento_mes=1 +
     0.0007 * desmame_peso +
    -0.0002 * desmame_idade +
     0.0243 * desmame_mes=1,5 +
    -0.025  * desmame_mes=5 +
    -0.0003 * diff_abate_desmame +
     0.6084
```

(b) GMD

```
bonificacao =

     0.0112 * nascimento_mes=1,11,10 +
     0.0002 * desmame_peso +
     0.0001 * desmame_idade +
     0.0155 * desmame_mes=1 +
     0      * diff_abate_desmame +
    -0.0139
```

(c) BN

```
abate_idade =

    -1.7433 * desmame_peso +
  1138.5243
```

(d) IA

Figura 1. Modelos descobertos

Para o peso de fazenda o experimento gerou um modelo onde os atributos utilizados foram mês de nascimento, peso de desmame, idade de desmane, mês de desmane e diferença abate/demame, sendo o atributo mês de nascimentos = 1 o mais relevante pois apresenta dois coeficientes no modelo, dando um maior peso a este atributo, outro fato a se ressaltar está na circunstância de os mês de nascimento = 8, 9 e 12 não serem utilizados no modelo, o mesmo ocorre com o atributo mês de desmame = 4 e 5, sendo utilizado apenas o mês de desmame = 1.

Para o ganho médio diário o modelo gerado também apresenta o mês de nascimento = 1 como atributo de maior relevância, porém neste modelo, é gerado três coeficientes para este atributo, sendo que o mês de nascimento = 8 ou 12 não foram utilizados no modelo. Nota-se também que o mês de desmame = 5 apresenta dois coeficientes enquanto mês de desmame = 4 não foi utilizado no modelo.

O modelo gerado para a bonificação assemelha-se ao do peso de abate, salvo pelos valores dos coeficientes e de que o atributo mês de nascimento = 1 aparecer apenas uma vez no modelo, não sendo utilizado o mês de nascimento = 8, 9 ou 12, comportamento semelhante ocorre com mês de desmame. O modelo de idade de abate foi o mais simples, levando em conta apenas o atributo peso de desmame desconsiderando os outros atributos.

Análise dos Resultados

Além dos modelos, cada experimento apresentou o relatório de valores reais para cada valor previsto, o valor previsto e a diferença entre eles(erro na previsão). Analisando os erros de cada instância com o valor real, observa-se que os erros para o ganho médio diário e bonificação foram baixos, as maiores diferenças entre o valor real e o previsto ocorreram na idade de abate. O peso de fazenda apresentou um desenho razoável por não apresentar uma variação muito elevada do erro.

A Tabela 3 apresenta os valores para comparação dos coeficientes de correlação e erros médios absolutos, calculados pelo algoritmo de regressão linear.

Tabela 3. Correlação e erro médio absoluto

Classes	Correlation coefficient	Mean absolute error
Bonificação	0.4067	0.0136
GMD	0.7736	0.0316
Idade de Abate	0.4092	83.9369
Peso de Fazenda	0.5882	27.2316

A correlação é uma medida estatística que indica a força e a direção da relação entre variáveis numéricas [Amaral 2016]. Ou seja, a correlação é um índice que indica o quanto duas variáveis estão relacionadas, sendo os valores retornados sempre dentro do intervalo de -1 e 1. Quanto mais próximas de -1 e 1, maior será a correlação entre as variáveis, e da mesma forma, quanto mais próxima de 0, mais fraca ela é.

O indicador de direção é dado pelo sinal da correlação, uma correlação positiva indica que enquanto uma variável cresce, a outra, correlacionada, também cresce, já na correlação negativa, enquanto uma variável cresce a outra diminui [Amaral 2016].

Analisando a Tabela 3 nota-se que GMD foi o que apresentou a maior correlação entre as variáveis preditoras, indicando que o modelo gerado teve uma boa métrica de qualidade, pois todas as variáveis utilizadas possuem uma boa correlação, além disso o erro médio ficou baixo. Mesmo caso do erro ocorreu com a bonificação,ou seja, o algoritmo quando não acertou o valor, errou por pouca diferença,para mais ou para menos. Para a idade de abate a média de erro ficou em 83.9369 dias e o peso de abate na fazenda em 27.2316 quilos.

As figuras 2, 3, 4 e 5 apresentam as distribuições de frequências para os erros

ocorridos nos experimentos referentes a bonificação, ganho médio diário, idade de abate e peso de fazenda.

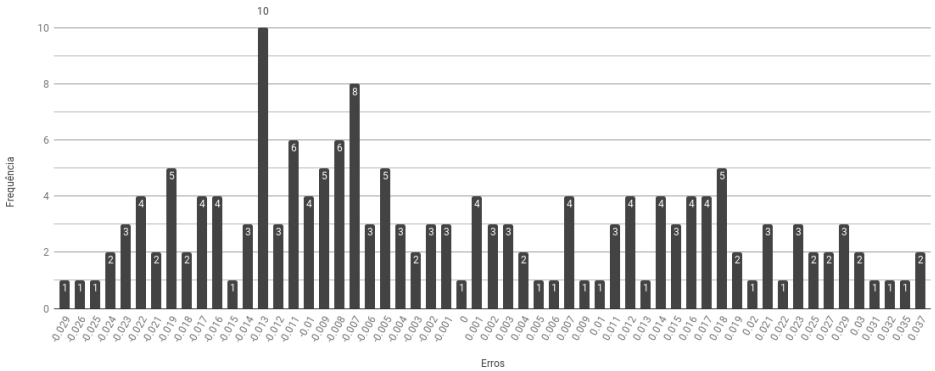

Figura 2. Frequência de Erros para Bonificação

Observando a Figura 2, nota-se que para a bonificação as maiores frequências então entre o intervalo $-0.013 \vdash -0.005$, indicando que o modelo calculou os valores e os erros concentraram-se neste intervalo, além disso a tendência linear indica que para os maiores valores positivos dos erros, a tendência da frequência é o valor 2(dois), caso contrário, para os menores valores negativos dos erros a tendência é 4(quatro), para o erro 0(zero), ou seja, que o modelo acerto o valore real, a tendência é 3(três), nota-se que a tendência corrobora com o intervalo das maiores frequências.

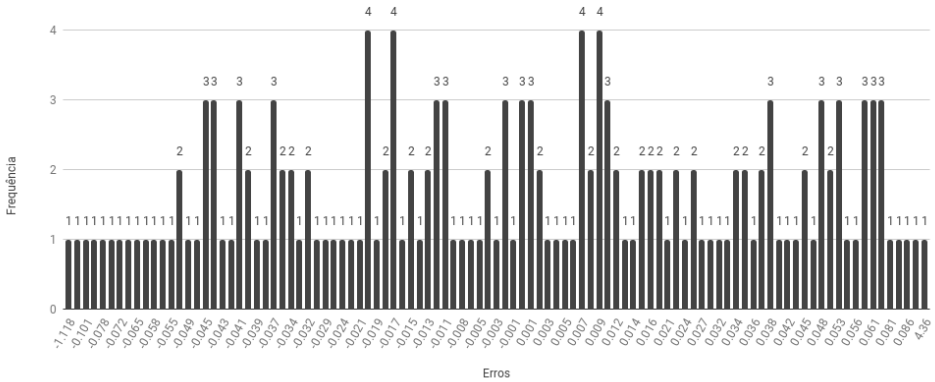

Figura 3. Frequência de Erros para GMD

Conforme a Figura 3, para o ganho médio diário, as maiores frequências então entre o intervalo $-0.02 \vdash 0.01$, demonstrando uma distribuição homogênea em torno do valor 0(zero), diferentemente da bonificação. No caso do GMD, a tendência linear para os menores valores negativos dos erros a tendência é 1.5(um ponto cinco) e para os maiores valores positivos dos erros, a tendência da frequência é o valor 2(dois), para o erro 0(zero) a tendência ficou aproximada a 1.75(um ponto setenta e cinco).

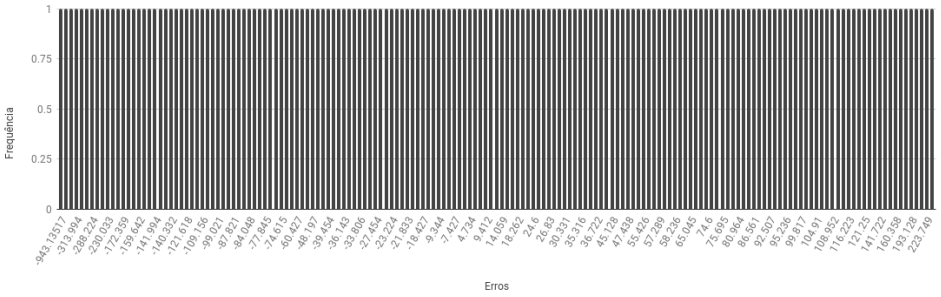

Figura 4. Frequência de Erros para Idade de Abate

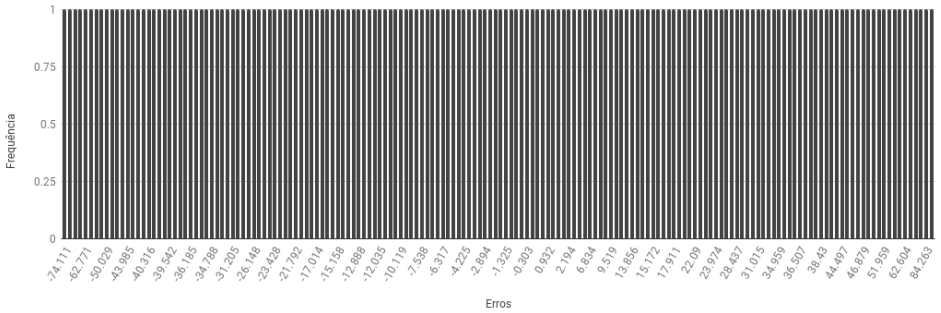

Figura 5. Frequência de Erros para Peso de Fazenda

Analisando as Figuras 4 e 5, a idade de abate e peso de fazenda, não houve repetição de erros, o que gerou uma frequência de 1(um) para todas as instâncias.

Com referência aos modelos gerados, os atributos não utilizados podem ter influenciado o resultado dos experimentos. Conforme Figura 1(d), para idade de abate o único atributo considerado relevante no modelo foi peso de desmame, desconsiderando as informações constantes nos outros atributos, podendo ter sido gerado um modelo pobre de previsão. Os outros modelos, Figuras 1(a), 1(b) e 1(c), também desconsideram algumas informações como mês de nascimento = 8 e 12 e mês de desmane = 4, podendo estes terem seus desempenhos afetados por essas exclusões de informações.

O algoritmo de regressão linear ainda fornece o valor de R^2, que é o coeficiente de determinação. Ele fornece uma informação auxiliar ao resultado da análise de variância da regressão, como maneira de se verificar se o modelo proposto é adequado ou não para descrever o fenômeno estudado. O valor de R^2 varia no intervalo de 0 a 1. Valores próximos de 1 indicam que o modelo proposto é adequado para descrever o fenômeno. Tabela 4 apresenta os valores do R^2 para os experimentos realizados.

Tabela 4. Valores dos coeficientes de determinação

Classes	Coeficiente de determinação - R^2
Bonificação	0.1654
GMD	0.5985
Idade de Abate	0.1675
Peso de Fazenda	0.3464

Analisando os valores de R^2 encontrados, observa-se que apenas o ganho médio diário apresentou um coeficiente relativamente elevado. Os outros indicam que os modelos descobertos não são adequados para descrever as variáveis zootécnicas de qualidade.

Conclusão

O presente trabalho teve como objetivo descobrir a influência e a relação das variáveis de produção e manejo na bonificação, peso de fazenda, ganho médio diário e idade de abate, em relação as variáveis de cria com o uso de aplicações de mineração de dados, almejando que os modelos descobertos possam empregados com a finalidade de auxiliar os produtores na gestão eficiente do negócio.

Foi realizada a aquisição e seleção de indicadores de qualidade zootécnicos e de cria, o tratamento dos dados e uso da tarefa de mineração de dados sobre os mesmo. Todos os resultados foram discutidos na análise dos resultados, evidenciando a razão pela qual os mesmos foram obtidos.

Pode-se concluir que os resultados foram parcialmente alcançados, pois com as tarefas de regressão configuradas conforme descritas na metodologia, mostraram que as variáveis de cria usadas possuem boa correlação apenas para o ganho médio diário, e os modelos gerados para a bonificação e ganho médio diário apresentaram erros baixos, indicando que as variáveis de cria usadas podem explicar um ganho médio diário alto ou baixo, assim como a bonificação. Para a idade de abate e peso de fazenda, os modelos apresentaram diferenças maiores entre o valor real e o valor previsto, e baixa correlação, podendo significar que as variáveis de cria usadas nos experimentos não sejam suficientes para explicar o peso de fazenda e a idade de abate, além disso a média dos erros ficou elevada e fora dos pradões esperados. A frequência dos erros ficou heterogênea, com a tendência linear igual a 1(um) para as instâncias destes dois atributos. O coeficiente de determinação indica que no contexto da pecuária de corte, o modelo descoberto para o ganho médio diário poderá ser utilizado como ferramenta de consulta para os produtores, o outros modelos necessitam que um melhor tratamento.

Trabalhos futuros envolvem à adoção de novos indicadores, como por exemplo o peso de nascimento, tipo de alimentação da mãe do bovino enquanto este ainda mama, entre outros, para testar e observar como os modelos se comportam, pode-se também empregar outros tipos de técnicas de mineração de dados como o algoritmo M5P e redes neurais, com treinamento e configuração das camadas ocultas. Também se sugere a expansão do banco de dados, através de parcerias com outros produtores rurais, e do estudo para consideração de outras raças bovinas. Apresentando ao produtores os resultados obtidos e demonstrando que é possível aumentar seu rendimento com técnicas adequadas.

Referências

Amaral, F. (2016). *Aprenda Mineração de Dados - Teoria e Prática*. Rio de Janeiro: Alta Books, 1th edition.

Corrêa, Â. M. J. and Sferra, H. (2003). Conceitos e aplicações de data mining. *Revista de ciência & tecnologia*, 11:19–34.

Costa, C. L. (2016). *Utilização de caracteríticas zootécnicas e de manejo na pecuária para previsão do peso final e bonificação de bovinos empregando redes neurais artificiais*. Tabalho de conclusão de curso, Universidade Federal do Pampa.

Euclides Filho, K. (2000). Produção de bovinos de corte e o trinômio genótipo-ambiente-mercado. *Embrapa Gado de Corte-Documentos (INFOTECA-E)*.

Fayaad, U. M., Shapiro, G. P., and Smyth, P. (1996). From data mining to knowledge discovery: An overview.

Gomes, A., Carneiro, A., Yamaguchi, L., Passos, L., Carvalho, M., and Campos, O. d. (1997). Acompanhamento de fazendas produtoras de leite.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Mota, F. d., Souza, K., Ishii, R., and Gomes, R. d. C. (2017). Bovreveals: uma plataforma olap e data mining para tomada de decisão na pecuária de corte. In *Embrapa Gado de Corte-Artigo em anais de congresso (ALICE)*. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 11., 2017, Campinas. Anais... Campinas: Embrapa Informática Agropecuária; Unicamp, 2017.

Oiagen, R. and Barcellos, J. (2008). Gerenciamento e custo de produção. *MOURA, JA et al. Programa de atualização em medicina veterinária. Porto Alegre: ARTMED*, pages 51–88.

Oiagen, R. P. (2010). *Avaliação da competitividade em sistemas de produção de bovinocultura de corte nas regiões sul e norte do Brasil*. Tese de doutorado em zootecnia, Universidade Federal do Rio Grande do Sul.

Tan, P.-N., Steinbach, M., and Kumar, V. (2005). Association analysis: basic concepts and algorithms. *Introduction to Data mining*, pages 327–414.

Vieira, F. and Oliveira, S. d. M. (2014). Mineração de dados: conceitos e um estudo de caso sobre certificação racial de ovinos. *Embrapa Informática Agropecuária-Capítulo em livro científico (ALICE)*.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.