

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

Rodrigo Rosa da Silva¹
Arthur de Oliveira Silveira²
Thales Vaz Maciel³

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 a partir dos bovinos podem explicar a bonificação, ganho médio diário, idade de abate e peso de abate 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 peso de abate na fazenda os modelos descobertos apresentaram baixa correlação e coeficiente de determinação, enquanto que para idade de abate e ganho médio diário, os resultados foram bons, o que permite concluir que os atributos não foram o suficiente para prever a idade de abate e peso de fazenda, mas bons para a idade de abate e ganho médio diário.

Palavras-chave: *Indicadores, mineração de dados, modelos*

1 INTRODUÇÃO

A produção sustentável de alimentos precisa ser ambientalmente correta, socialmente responsável e economicamente viável, priorizando cinco princípios: recursos naturais, pessoas e a comunidade, saúde e bem-estar animal, produção de alimento e eficiência e inovação, GRSB[1]. O setor agropecuário envolve essa multifuncionalidade da produção, pois além de desempenhar sua função básica de produção de alimentos, possui essas outras funções ambientais, econômicas, territoriais e sociais que determinam a renda para o setor. Desta forma, a produção de carcaças de qualidade pode impactar em toda a cadeia de valor, em que a função social de sistemas agrícolas está relacionada com a viabilidade socioeconômica das áreas rurais, êxodo e sucessão rural.

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[2].

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 Oaigen[3] a mensuração e análise de indicadores que retratam o funcionamento rural são fundamentais para a tomada de decisão. Estes indicadores, de acordo com Oaigen & Barcellos[4], 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.

¹ Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense - Campus Bagé - Autor

² Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense - Campus Bagé - Autor

³ Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense - Campus Bagé - Orientador

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

Ainda segundo Oaigen[3], 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 em situações 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 nascimento e desmame, que explicam bons indicadores de qualidade zootécnicas?". A hipótese é que os dados ano de nascimento, ano de abate, ano de desmame, mês de nascimento, mês de desmame, peso de desmame e idade de desmame têm correlação suficiente com o peso de abate na fazenda, idade de abate, ganho médio diário e bonificação explicar bons valores em tais indicadores.

O objetivo deste trabalho é descobrir a relação estatística entre as variáveis de cria e esses indicadores zootécnicos de qualidade de carcaças após abate e 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.

2 REFERENCIAL TEÓRICO

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

A. 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 & Oliveira[5]. 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.[6].

Grande parte dessas técnicas e ferramentas podem ser encontradas no processo de descoberta de conhecimento em bases de dados (DCBD). Segundo Fayaad et. al.[7], 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.[8]. 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

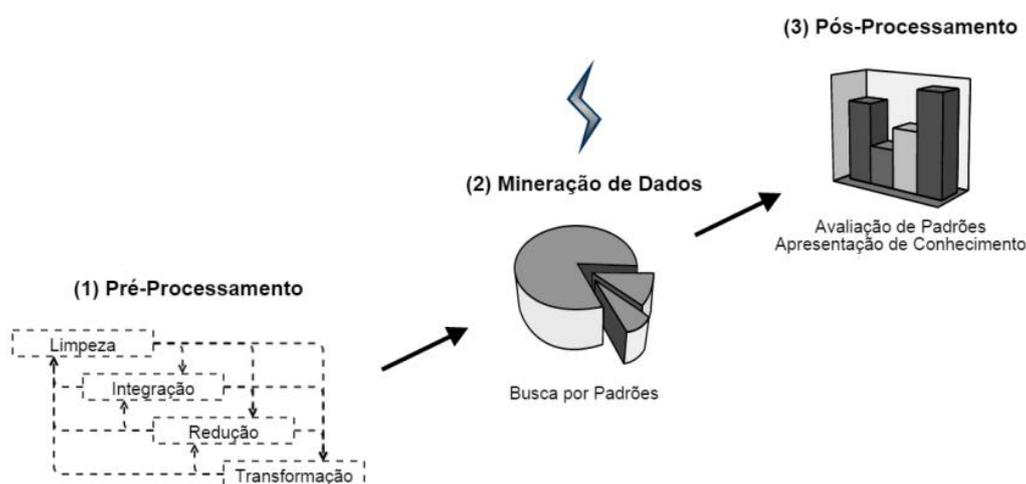
> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

mineração de dados, que são processos aplicados para explorar e analisar os dados em busca de padrões, previsões, erros, associações entre outros, Amaral[9]. 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 & Sferra [10]. A representação das etapas de mineração de dados dentro do contexto da DCBD poder ser observadas na Figura 1.

Figura 1. Fases de um processo DCBD, Maciel et al.[11]



As tarefas de mineração de dados podem ser divididas 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 itens, gerando regras que determinam a associação entre esses itens, Amaral[9]. Este estudo tem foco em tarefas de regressão.

B. Revisão dos Trabalhos Correlatos

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

No trabalho de Meirelles[12], foram utilizadas duas ferramentas computacionais para fins de auxiliar tomadas de decisões na produção de bovinos de corte, criados de maneira extensivas. A primeira parte do trabalho visou à construção de um *software* utilizando a técnica de simulação monte carlo para analisar características de produção (ganho de peso) e manejo (fertilidade, anestro pós-parto, taxa de

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

natalidade e puberdade). Na segunda parte do trabalho foi aplicada a técnica de redes neurais artificiais para classificar animais, segundo ganho de peso nas fases de crescimento (nascimento ao desmame, do desmame ao sobreano) relacionado com o valor genético do ganho de peso do desmame ao sobreano (GP345) obtidos pelo método de melhor preditor linear não viesado (BLUP - do inglês *Best Linear Unbiased Predictor*). Ambos modelos mostraram potencial para auxiliar a produção de gado de corte.

Na pesquisa de Ventura et al.[13] foram utilizados dados de 19240 animais Tabapuã, nascidos entre 1976 e 1995, foram utilizados para predição do valor genético do peso aos 205 dias de idade (VG_P205) por meio de redes neurais artificiais (RNA's) e usando o algoritmo LM - *Levenberg Marquardt* - para treinamento dos dados de entrada. Por se tratar de rede com aprendizado supervisionado, foram utilizados, como saída desejada, os valores genéticos preditos pelo melhor preditor linear não viesado (BLUP) para a característica P205. Os valores genéticos do P205 obtidos pela RNA e os preditos pelo BLUP foram altamente correlacionados. A ordenação dos valores genéticos do P205 oriundos das RNA's e os valores preditos pelo BLUP (VG_P205_RNA) sugeriram que houve variação na classificação dos animais, indicando riscos no uso de RNA's para avaliação genética dessa característica. Inserções de novos animais necessitam de novo treinamento dos dados, sempre dependentes do BLUP.

Já no estudo de Costa[14] foi analisado 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.

No trabalho de Mota et al.[15] 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 *Waikato Environment for Knowledge Analysis* (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.

Nota-se que trabalhos correlatos publicados recentemente, mesmo que parcialmente eficazes segundo os respectivos autores, não explicam as predições realizadas pelos experimentos que documentam, ou pela impossibilidade disto ser característica do algoritmo empregado (caixa-preta) ou por não apresentar a totalidade dos resultados da classificação nos resultados obtidos nos testes (matrizes de confusão, por exemplo). O trabalho difere-se dos demais por usar a tarefa de regressão no

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

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.

3 METODOLOGIA

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

Tabela 1 - Nomenclatura e Descrição dos Atributos Utilizados com Regressão Linear

Nomenclatura	Tipo de Dado	Descrição
abate_ano	Nominal	Ano de abate (2013,2014)
desmame_ano	Nominal	Ano de desmame (2011,2012,2013)
nascimento_ano	Nominal	Ano de nascimento (2010,2011,2012)
nascimento_mes	Nominal	Mês de nascimento (1,8,9,10,11,12)
desmame_peso	Numérico	Peso de desmame
desmame_idade	Numérico	Idade de desmame
desmame_mes	Nominal	Mês de desmame (1,4,5)
abate_idade	Numérico	Idade de abate
Gmd	Numérico	Ganho médio diário de peso
abate_peso	Numérico	Peso de abate na fazenda
bonificação	Numérico	Bonificação

Como ferramenta para a realização das tarefas de pré-processamento e aplicações das tarefas de mineração de dados, foi utilizado o software *Waikato Environment for Knowledge Analysis* (WEKA), um ambiente para análise de conhecimento desenvolvido pela Universidade de Waikato, Nova Zelândia, Hall et al.[16]. 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, Amaral[9].

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. O pré-processamento do 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

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

dados descritos pela Tabela 1. Após o WEKA ser alimentado com os dados, foi aplicado o filtro *weka.filters.unsupervised.attribute.NumericToNominal* sobre os atributos *desmame_mes* e *nascimento_mes* de modo que os dados foram convertidos do 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.[17]. 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 , Amaral[9]. 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 zootécnicas de qualidade, onde para cada teste, as outras variáveis zootécnicas de qualidade foram eliminadas. Tabela 2 apresenta os atributos selecionados para cada um dos experimentos.

Tabela 2 - Variáveis Utilizadas nos Experimentos com Regressão Linear

Variáveis dos experimentos	Bonificação	Peso de abate	Idade de abate	Ganho médio diário
abate_ano				
desmame_ano				
nascimento_ano				
Idade de desmame				
Mês de desmame				
Peso de desmame				
Mês de nascimento				
Bonificação	Alvo	Removido	Removido	Removido
Peso de abate	Removido	Alvo	Removido	Removido
Idade de abate	Removido	Removido	Alvo	Removido
Ganho médio diário	Removido	Removido	Removido	Alvo

Com as variáveis zootécnicas de qualidade selecionadas, foram realizados os experimentos com o uso da técnica de validação cruzada, mecanismo usado para avaliar a taxa de acerto dos algoritmos de mineração de dados. Na validação cruzada, os dados são divididos em um número n de conjuntos definidos pelo usuário, conhecidos com partições. Cada um destes conjuntos será utilizado uma vez como teste do modelo. Se o número de conjuntos definido foi de 10, significa que haverá 10 testes sobre

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

o modelo. A validação cruzada é a forma mais recomendada para avaliar modelos, Amaral[9]. Os experimentos foram realizados com partições iguais a 10, 5 e 3 para cada uma das variáveis zootécnicas de qualidade testadas.

Além disso, para cada partição testada, também foi modificado a semente aleatória, sendo utilizada uma variação de 1 até 5. Basicamente a semente aleatória é uma semente geradora, que determina quais parte do conjunto de dados serão utilizadas para treinamento e teste. A opção pelo uso de sementes aleatórias, variando entre 1 e 5, foi proposta com a finalidade de determinar que o algoritmo se utiliza partes diferentes do conjunto para realizar o treinamento e teste, juntamente com a variação das partições, e por fim, analisar o desempenho do algoritmo predictor.

Na Figura 2 observa-se os modelos descobertos para os quatro experimentos, bonificação, peso de abate fazenda, ganho médio diário e idade de abate. 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. A terceira etapa consistiu na análise e interpretação dos resultados obtidos na etapa anterior, onde foi utilizado o algoritmo de regressão linear no conjunto de dados.

Figura 2. Modelos Descobertos

```
bonificacao =  
  
0.0067 * nascimento_ano=2010,2011 +  
0.0117 * nascimento_mes=1,11,10 +  
0.0001 * desmame_peso +  
0.0001 * desmame_idade +  
0.0174 * desmame_mes=1 +  
0.0125  
  
abate_peso =  
  
30.5924 * abate_ano=2014 +  
26.0388 * desmame_ano=2012,2011 +  
57.5075 * nascimento_ano=2010 +  
31.8451 * nascimento_mes=8,9,11,10,1 +  
13.2494 * nascimento_mes=10,1 +  
0.4496 * desmame_peso +  
21.8136 * desmame_mes=1 +  
292.871
```

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

```
gmd =  
-0.0741 * abate_ano=2014 +  
0.11 * desmame_ano=2012,2013 +  
0.1922 * desmame_ano=2013 +  
0.0322 * nascimento_mes=11,10,8,9,1 +  
-0.0315 * nascimento_mes=8,9,1 +  
0.0731 * nascimento_mes=1 +  
0.0007 * desmame_peso +  
-0.0002 * desmame_idade +  
0.3563  
  
abate_idade =  
-179.3713 * abate_ano=2013 +  
284.7182 * desmame_ano=2012,2011 +  
339.5202 * desmame_ano=2011 +  
-26.8786 * nascimento_mes=11,10,12 +  
16.2893 * nascimento_mes=10,12 +  
-29.1108 * nascimento_mes=12 +  
-0.3178 * desmame_peso +  
0.6733 * desmame_idade +  
-10.9038 * desmame_mes=4,1 +  
78.6673 * desmame_mes=1 +  
516.0642
```

4 ANÁLISE DOS RESULTADOS

O modelo descoberto para a bonificação utilizou o ano de nascimento = 2010 e 2011, mês de nascimento = 1, 11 ou 10, peso de desmame, idade de desmame e mês de desmame = 1, desconsiderando as outras informações. Observa-se que o ano e desmame e ano de bate sequer foram utilizados no modelo descoberto.

Para o peso de abate na fazenda o experimento gerou um modelo onde os atributos utilizados foram ano de abate, ano de desmame, ano de nascimento, mês de nascimento, peso de desmame, idade de desmame e mês de desmame, sendo o atributo mês de nascimentos = 10 e 1, o mais relevante, pois apresenta dois coeficientes no modelo para estes dois meses, dando um maior peso a este atributo, outro fato a se ressaltar está na circunstância de os meses de nascimento = 12 não ser utilizado 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 apresenta o mês de nascimento = 1 como atributo de maior relevância, pois neste modelo são gerados três coeficientes para este atributo, sendo que o mês de nascimento = 12 não foi utilizado no modelo. Nota-se também que o ano de desmame = 2013

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

apresenta dois coeficientes enquanto mês de desmame e ano de nascimento não foram utilizados no modelo.

O modelo descoberto para idade de abate foi utilizado o maior número de atributos. O mês de nascimento = 12 apresenta-se como o de mais relevância, sendo utilizado três vezes. Ano de desmame = 2011 e mês de desmame = 1 foram utilizados duas vezes no modelo.

Com referência aos modelos descobertos, os atributos não utilizados podem ter influenciado o resultado dos experimentos. Observa-se ainda que, mês de nascimento, peso de desmame foram utilizados nos quatro modelos gerados. No contraponto, mês de nascimento = 12 e mês de desmame = 5 não foram utilizados e nenhum dos modelos descobertos.

Além dos modelos, cada experimento apresentou o relatório de valores reais, os valores previstos e a diferença entre eles (erro na previsão). A Tabela 3 apresenta os melhores resultados dos experimentos para os coeficientes de correlação e erros médios absolutos, calculados pelo algoritmo de regressão linear.

Tabela 3 - Melhores Resultados com Regressão Linear

Resultados	Coefficiente de correlação	Erro médio absoluto
Bonificação	0,2491	0,0158
GMD	0,8514	0,0342
Idade de abate	0,9715	25,0985
Peso de abate	0,402	31,4487

Para a bonificação, os melhores resultados foram obtidos através da validação cruzada com 5 partições e $seed = 1$. Já para o ganho médio diário e idade de abate, os dados preferíveis foram observados com a validação cruzada com 10 partições e $seed = 5$. No peso de abate os dados supremos foram obtidos com a validação cruzada com 3 partições e $seed = 1$.

A correlação é uma medida estatística que indica a força e a direção da relação entre variáveis numéricas, Amaral[9]. 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[9].

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

Analisando a Tabela 3 nota-se que a idade de abate foi o que apresentou a maior correlação entre as variáveis preditoras, indicando que o modelo gerado teve uma ótima métrica de qualidade, pois todas as variáveis utilizadas possuem uma boa correlação, além disso o erro médio ficou relativamente baixo, em torno de 25 dias de diferença. Para o ganho médio diário, o coeficiente de correlação também pode ser considerado bom, indicando que as variáveis preditoras possuem uma boa correlação. Os piores resultados foram para a bonificação e peso de abate, onde se observa uma baixa correlação entre as variáveis.

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	0.1380
GMD	0.7675
Idade de abate	0.9497
Peso de abate	0.2536

Analisando os valores de R^2 encontrados, observa-se que idade de abate e ganho médio diário apresentaram os coeficientes com bons índices. Os outros indicam que os demais modelos descobertos não são adequados para descrever as variáveis zootécnicas de qualidade.

Analisando o coeficiente de correlação e o de determinação dos experimentos realizados, fica evidente que para idade de abate e ganho médio diário de peso, os modelos descobertos são válidos para descrever os fenômenos testados e que as variáveis preditoras são suficientes para prever estas duas variáveis zootécnicas de qualidade. O mesmo não se pode dizer a respeito da bonificação e peso de abate na fazenda, pois os coeficientes demonstraram-se insatisfatórios.

5 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 mesmos. Todos os resultados foram discutidos

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

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 para a idade de abate e ganho médio diário de peso, além de um bom coeficiente de determinação, desta forma conclui-se que os modelos são bons para prever os fenômenos. Os modelos descobertos para a bonificação e peso de abate na fazenda apresentaram diferenças maiores entre o valor real e o valor previsto, baixa correlação e coeficiente de determinação, significando que apenas as variáveis de cria usadas não são suficientes para explicar o peso de abate na fazenda e bonificação.

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. Apresentar aos produtores os resultados obtidos e demonstrar que é possível aumentar seu rendimento com técnicas adequadas.

REFERÊNCIAS

- [1] Global Roundtable For Sustainable Beef, "Principles and criteria for defining global sustainable beef," 2015.
- [2] K. Euclides Filho, "Produção de bovinos de corte e o trinômio genótipoambiente-mercado." Embrapa Gado de Corte-Documents (INFOTECAE), 2000.
- [3] R. P. Oaigen, "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, 2010.
- [4] R. Oaigen and J. Barcellos, "Gerenciamento e custo de produção," MOURA, JA et al. Programa de atualização em medicina veterinária. Porto Alegre: ARTMED, pp. 51–88, 2008.
- [5] F. Vieira and S. Oliveira, "Minerac, ~ao de dados: conceitos e um estudo de caso sobre certificação racial de ovinos." *Embrapa Informática Agropecuária-Capítulo em livro científico (ALICE)*, 2014.
- [6] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
- [7] U. Fayaad, G. Shapiro, and P. Smyth, "From data mining to Knowledge discovery: An overview," 1996.

> Aplicação de Mineração de Dados na Predição de Indicadores de Qualidade de Carcaças Bovinas



2018 | Volume 1 | Nº 1

- [8] P. Tan, M. Steinbach, and V. Kumar, "Association analysis: basic concepts and algorithms," *Introduction to Data mining*, pp. 327–414, 2005.
- [9] F. Amaral, *Aprenda Mineração de Dados - Teoria e Prática*, 1st ed. Rio de Janeiro: Alta Books, 2016.
- [10] Â. Correa and H. Sferra, "Conceitos e aplicações de data mining," *Revista de ciência & tecnologia*, vol. 11, pp. 19–34, 2003.
- [11] T. V. Maciel, V. Da Rosa Seus, K. Dos Santos Machado, and E. N. Borges, "Mineração de dados em triagem de risco de saúde," *Revista Brasileira de Computação Aplicada*, vol. 7, no. 2, pp. 26–40, 2015.
- [12] F. Meirelles, "Modelo computacional de um rebanho bovino de corte virtual utilizando simulação monte carlo e redes neurais artificiais," Ph.D. dissertation, Universidade de São Paulo, 2005.
- [13] R. Ventura, M. Silva, T. Medeiros, N. Dionello, F. Madalena, A. Fridrich, B. Valente, G. Santos, L. Freitas, R. Wenceslau 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*, vol. 64, no. 2, pp. 411–418, 2012.
- [14] C. L. Costa, "Utilização de características zootécnicas e de manejo na pecuária para previsão do peso final e bonificação de bovinos empregando redes neurais artificiais," Trabalho de Conclusão de Curso, Universidade Federal do Pampa, 2016.
- [15] F. D. Mota, K. Souza, R. Ishii, and R. D. C. Gomes, "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., 2017.
- [16] M. Hall, E. Frank, G. Holmes, B. Pfhringer, P. Reutemann, and I. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [17] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.