

NOTABREVE

## MINERAÇÃO DE DADOS PARA ESTIMATIVAS DE MORTALIDADE PRÉ-ABATE DE FRANGOS DE CORTE

DATA MINING FOR PRESLAUGHTER MORTALITY PREVIEW OF BROILERS

Rodrigues, V.C.<sup>1\*</sup>; Vieira, F.M.C.<sup>1</sup> e Silva, I.J.O.<sup>1A</sup>

<sup>1</sup>Departamento de Engenharia de Biossistemas. Piracicaba-SP. Brasil. \*vcrodrig@esalq.usp.br; ^ijosilva@esalq.usp.br

### PALAVRAS CHAVE ADICIONAIS

Árvores de decisão. Avicultura. Suporte à decisão.

### ADDITIONAL KEYWORDS

Decision making. Decisions trees. Poultry.

### RESUMO

Objetivou-se analisar a eficiência de um sistema de informação simples para diagnosticar cenários de mortalidade em condições de espera pré-abate, pelo uso de árvores de regressão e classificação, estudando dados de um abatedouro comercial de frangos de corte no Estado de São Paulo, utilizando-se um sistema de mineração de dados (CART, classification and regression trees). Esta árvore de decisão, por meio de um conjunto de dados históricos, é capaz de prever variáveis categóricas e classificar efeitos, e apresentou 77 % de acerto e mostrou ser um meio prático e rápido para predições, além de ser de fácil interpretação para uso direto de técnicos.

### SUMMARY

The performance of a simple information system to identify possible mortality conditions in pre-slaughter operation through classification and regression trees was studied. Data referring a commercial broiler slaughterhouse, located in State of São Paulo, were studied using a data mining system (CART). This decision tree is able to predict categorical variables through a historical dataset and then classify effects. The classification tree had 77 % of success and showed to be a practical and faster way to predictions for this dataset and it is able to show an easy interpretation to direct use by the sector area experts.

### INTRODUÇÃO

As operações pré-abate são um dos

maiores problemas na produção de frango de corte, devido às perdas por mortalidade antes da chegada. O sistema de criação industrial apresenta elevada densidade de animais, condições bioclimáticas inadequadas como altas temperaturas e umidade relativa do ar, na maior parte do tempo nos países de clima tropical. O tempo de viagem, a distância entre as granjas e o abatedouro e aspectos relacionados à espera no abatedouro afetam as perdas ao final do processo (Warriss *et al.*, 2005). Com base nos estudos de Vieira *et al.* (2011), foi utilizado o modelo linear generalizado duplo que considera como preditoras: temperatura externa, umidade relativa externa, tempo de espera, turno, número de aves por caixa e estação do ano. No entanto, os modelos utilizados apresentaram problemas quanto à predição da mortalidade, abrindo espaço para o uso de sistemas especialistas. As árvores de decisão, chamadas de árvores de regressão e classificação, também são ótimas escolhas para obter resultados com precisão conhecida de forma rápida e compreensível (Sutton, 2005). Objetivou-se analisar a eficiência de um sistema de informação simples para diagnosticar possíveis condições de mortalidade em condições de espera pré-abate, através do uso de árvores de regressão e classificação.

Recibido: 7-6-10. Aceptado: 22-9-11.

Arch. Zootec. 62 (239): 469-472. 2013.

## MATERIALE MÉTODOS

CART é um método estatístico não-paramétrico, que usa árvores de decisão para solucionar problemas de regressão e classificação. Quando a variável dependente é categórica (não contínua) CART produz uma árvore de classificação. São três os passos a serem seguidos para o crescimento da árvore de decisão (Breiman *et al.*, 1984). A árvore máxima é construída a partir do procedimento binário de regras de divisão (*split*) que ocorre na raiz da árvore onde estão todas as informações sem qualquer tipo de organização, é o chamado primeiro grupo *mãe*. Inicialmente, este grupo é dividido em dois outros grupos *filhas* segundo alguns critérios de similaridades entre as mesmas variáveis de um grupo e cada *filha* torna-se *mãe* para que o processo ocorra até que cada uma das divisões (regras *if-then*) descreva respostas homogêneas quanto a variável categórica. Para que isto aconteça o crescimento da árvore é feito por meio de um descritor apropriado. O sistema CART usa um algoritmo cujo critério de divisão é apresentar a máxima redução de impureza dos dados em cada grupo, ou seja, grupos homogêneos de informações em cada ramo da árvore. Para tanto é utilizado o índice Gini (Breiman *et al.*, 1984) que é a medida de dispersão dos dados presentes dos nós da árvore e é comumente usado para variáveis categóricas.

O produto das probabilidades mede justamente a probabilidade de erro entre as classes. Obtidos os atributos que constituirão a construção da árvore, a árvore máxima é obtida, e o processo de *poda* (*pruning*) é necessário devido à ocorrência de ambigüidades (*overfitting*). Durante este processo, árvores menores e menos complexas são obtidas por cortes sucessivos dos ramos da árvore (regras) máxima. As diferentes subárvores são comparadas entre si para que a melhor seja encontrada. O critério para seleção é o custo que cada subárvore representa. A medida é feita com relação ao

parâmetro complexidade do custo ( $R_d(T)$ ), para a árvore T.

Durante o processo de *poda* este parâmetro varia de 0 a 1 o que minimiza o valor de  $R_d(T)$  reduzindo o custo de treinamento do modelo. A seleção da melhor árvore é feita com base em informações do menor custo de complexidade da validação cruzada que testa o desempenho do modelo, analisando uma parcela retirada dos dados em relação aos dados de referência. É calculado, para tanto, a estimativa de erro de classificação (*resubstitution error*), ou seja, a proporção de observações originais que foram erroneamente classificadas em vários subgrupos da árvore original. A validação cruzada é utilizada para estimar o verdadeiro erro, para as árvores de vários tamanhos. A árvore que apresenta menor erro de validação cruzada é escolhida. Após a escolha, é possível calcular o custo estimado a partir da análise discriminante e verificar a matriz de classificação dos dados originais. O conjunto de dados foi obtido por meio de um experimento realizado em um abatedouro no Estado de São Paulo, Brasil, para 218 carregamentos de frangos de corte (Vieira *et al.*, 2010). Foram analisados os caminhos de transporte sob a condição de espera, em cada período estudado (manhã, tarde e noite). Dessa forma, considerou-se para a análise dos dados 4 repetições para o turno da manhã, 5 para o turno da tarde e 5 para o turno da noite.

O modelo de árvore de regressão foi construído a partir do *Matlab Statistics toolbox*® R2006a. Foram utilizadas 4 classes para representar a taxa de mortalidade, de forma categórica, conforme Silva e Vieira (2010).

## RESULTADOS E DISCUSSÃO

A melhor árvore, que apresenta o menor valor de variância residual (0,57), foi escolhida pelo sistema e apresenta 13 folhas (nós terminais), sendo que a árvore máxima apresentou 35 folhas devido sua tamanha

## MINERAÇÃO DE DADOS E MORTALIDADE DE FRANGOS DE CORTE

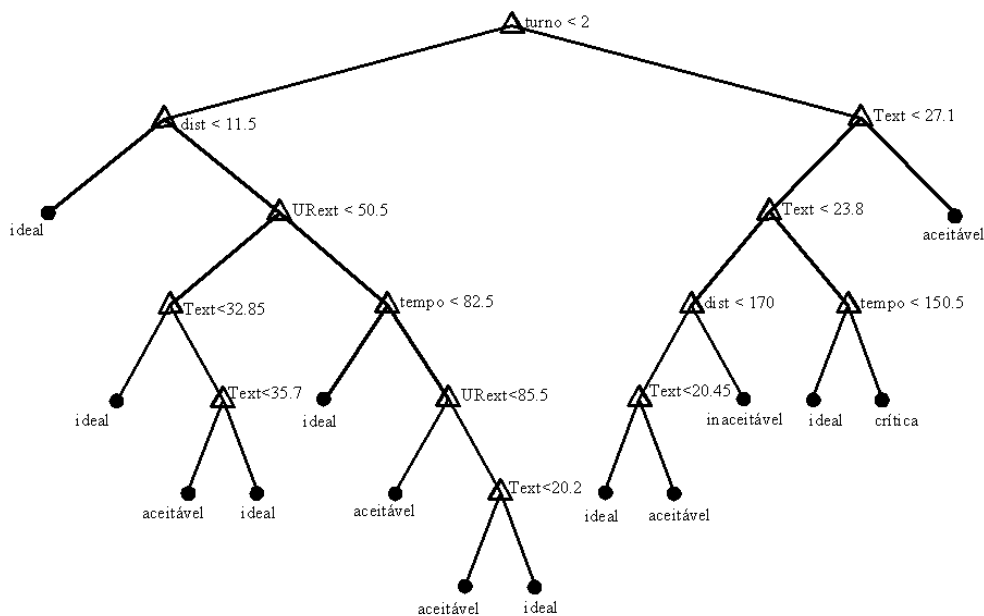
complexidade (**figura 1**).

Pela formação da árvore, é possível afirmar que o sistema reconheceu como variável de melhor retratação dos grupos o turno de viagem, pois o algoritmo conseguiu separar dois grupos bem delimitados. Partindo do nó do topo da árvore, também chamado de raiz, foi possível observar que para o turno 1 (manhã) o suporte de decisão é dado pelo ramo esquerdo da árvore. Dessa forma, a interpretação do sistema de suporte a decisão mostra que a variável turno foi selecionada como de maior ganho (índice Gini) para dar suporte às demais informações, dessa forma, para os turnos 1 e 2, as informações estão contidas na ramificação esquerda da árvore, onde maiores detalhes são mostrados para a necessária separação das classes. Para o turno 3, ou seja, noturno, a ordem das variáveis preditoras mostrou-se diferentes quando comparadas aos turnos 1 e 2.

Dessa forma, o desempenho preditivo da árvore de classificação foi de 77 %. Foram

167 conjuntos de dados classificados adequadamente. A classe que apresentou problemas de perdas de classificação (*misclassification*) foi a *inaceitável* por ter apresentado poucos conjuntos de treino (11) em comparação aos demais.

O sistema de suporte à decisão mostra que se o turno avaliado for manhã, por exemplo, o ramo da árvore a ser utilizado pela análise de previsão será o da esquerda, pois o turno de número 1 (manhã) foi selecionado. Assim, prosseguindo a análise, tem-se que para distâncias percorridas inferiores a 11,5 km (*galho* à esquerda do ramo esquerdo) a taxa de mortalidade esperada é classificada como ideal, caso a distância seja igual ou superior a 11,5 km, outros critérios devem ser tomados como orientação para se prever as possíveis classificações das taxas de mortalidade. Nesse caso, se a umidade relativa do ar for menor que 50,5 % o *galho* a ser utilizado é o da esquerda e assim sucessivamente.



**Figura 1.** Representação gráfica da melhor árvore de classificação para os 218 conjuntos de dados analisados. (Graphic representation of best decision tree for 218 analyzed dataset).

Tais resultados acima foram confirmados anteriormente por diversos trabalhos que envolveram as operações pré-abate de frangos de corte, especialmente às de transporte e espera em abatedouro. Quanto ao turno em que as operações foram executadas, Barbosa Filho *et al.* (2009) encontraram que o melhor turno sob o ponto de vista bioclimático é o turno da manhã, por oferecer as melhores condições térmicas para o transporte de aves. Isto explica o resultado de melhor ganho do turno 1 (manhã) segundo o índice Gini. Juntamente com as demais informações, este turno ofereceu melhores opções para a tomada de decisão. Com relação à distância, também a árvore obteve resultados similares e concordantes com Vecerek *et al.* (2006), os quais recomendaram que as menores distâncias são melhores e com baixas mortalidades associadas às distâncias abaixo de 50 km. No entanto, considerando a realidade da maioria das empresas, cujas granjas integradas se encontram em um raio acima de 11,5 km, como foi evidenciada na CART, outros fatores passam a ser importantes para a tomada de decisão, como por exemplo, o tempo de espera.

O tempo de espera pré-abate só é eficaz se realizado dentro de galpões climatizados com nebulização e climatização (Silva e Vieira, 2010). Considerando a relação entre o tempo de espera e a distância, foi evidenciado por meio da árvore de decisão que para distâncias acima de 11,5 km e umidade relativa do ar acima de 50,5 %, o tempo de espera estimado pela CART visando o nível ideal de mortalidade foi abaixo de 82,5 minutos.

A árvore de decisão é uma plataforma de fácil compreensão a partir de um conhecimento introdutório, o que viabiliza o uso desse sistema para possíveis aferições de produtores que estejam interessados em simular algumas condições de campo.

## CONCLUSÕES

O sistema de suporte à decisão mostrou-se eficiente e de fácil análise. Para o aumento da performance do sistema outros algoritmos deverão ser empregados para que o acerto ultrapasse os 77 % atingido. Deve-se considerar que a ferramenta apresentada tem grande utilidade na produção animal como novo instrumento de avaliação de dados reais e históricos de produção.

## BIBLIOGRAFIA

- Barbosa Filho, J.A.D.; Vieira, F.M.C.; Silva, I.J.O.; Garcia, D.B.; Silva, M.A.N. e Fonseca, B.H.F. 2009. Transporte de frangos: caracterização do microclima na carga durante o inverno. *Rev Bras Zootecn*, 38: 2442-2446.
- Breiman, L.; Friedman, J.H.; Olshen, R.A. and Stone, C.J. 1984. Classification and Regression Trees. Wadsworth. Pacific Grove. 368 pp.
- Silva, I.J.O. e Vieira, F.M.C. 2010. Ambiência animal e as perdas produtivas no manejo pré-abate: o caso da avicultura de corte brasileira. *Arch Zootec*, 59: 113-131.
- Sutton, C.D. 2005. Handbook of Statistics. Penn State University. Pennsylvania. 24: 303-329.
- Vecerek, V.; Grbalova, S.; Voslarova, E.; Janackova, B. and Malena, M. 2006. Effects of travel distance and the season of the year on death rates of broilers transported to poultry processing plants. *Poultry Sci*, 85: 1881-1884.
- Vieira, F.M.C.; Silva, I.J.O.; Barbosa Filho, J.A.D. and Vieira, A.M.C. 2010. Productive losses on broiler preslaughter operations: effect of the distance from farms to abattoirs and of lairage time in a climatized holding area. *Rev Bras Zootecn*, 39: 2471-2476.
- Vieira, F.M.C.; Silva, I.J.O.; Barbosa Filho, J.A.D.; Vieira, A.M.C.; Rodrigues-Sarnighausen, V.C. and Garcia, D.B. 2011. Thermal stress related with mortality rates on broiler's preslaughter operations: a lairage time effect study. *Ciência Rural*, 41: 1639-1644.
- Warriss, P.D., Pagazaurtundua, A. and Brown, S.N. 2005. Relationship between maximum daily temperature and mortality of broiler chickens during transport and lairage. *Brit Poultry Sci*, 46: 6647-6651.