

CAI, Congreso Argentino de AgrolInformática

## Identificação de Variáveis de Relevância no Índice de Contusões Associadas ao Transporte de Gado de Corte

Marcel da Silva Camargo<sup>1</sup>, Ana Paula Lüdtke Ferreira<sup>1</sup>, Naylor Bastiani Perez<sup>1,2</sup><sup>1</sup> Programa de Pós-Graduação em Computação Aplicada, Universidade Federal do Pampa, Bagé/RS, Brasil

marcel@sigasistemas.com.br, anaferreira@unipampa.edu.br

<sup>2</sup> EMBRAPA Pecuária Sul, Bagé/RS, Brasil

naylor.perez@embrapa.br

**Resumo.** Este trabalho objetiva contribuir para a melhoria das condições de bem-estar animal durante o processo de transporte da propriedade para o frigorífico, bem como diminuir as perdas financeiras relacionadas às contusões desenvolvidas e ao estresse animal durante este processo. A metodologia usada busca identificar as variáveis relacionadas ao transporte de gado de corte que tenham maior impacto no índice de contusões associadas ao processo. Os dados foram coletados durante um período de quatro meses em um frigorífico da região sul do Brasil. As conclusões do trabalho apontam que existem questões sobre organização das cargas que precisam ser levadas em consideração pelas empresas de transporte.

### 1 Introdução

A evolução das exigências dos consumidores e as políticas de garantia de tratamentos adequados dos animais têm motivado mudanças nos processos de produção do mercado da proteína animal e, especificamente, da carne bovina. A cadeia de abastecimento de carne tem sido bastante dinâmica em termos de melhorias na segurança alimentar, qualidade e ajustes em sistemas de produção. Os consumidores estão interessados em origem animal, sistemas ecológicos e sustentáveis, tipo de alimentação e manuseio correto durante o transporte e o abate [1]. A motivação para essa questão engloba, mas vai além da questão ética, à qual se soma uma perspectiva econômica: parte das perdas na cadeia produtiva de carne bovina está relacionada com a falta de cuidado no transporte dos animais para abate.

O processo de transporte tem início no embarque dos animais nas propriedades, passa pelo transporte em si e finaliza-se no frigorífico, com o abate dos animais. A falta de condições adequadas de embarque na propriedade, quantidade, peso e densidade de carga inadequados ao caminhão, temperatura, tempo de viagem, condições das estradas, tempos de descanso, etc. são variáveis que determinam as condições nas quais os animais chegam aos frigoríficos. Segundo [2], o manejo e transporte são estressantes para os animais e têm implicações para o seu bem-estar. Juntamente com os aspectos

éticos, o tratamento humano dos animais em toda a cadeia de produção é um componente importante na qualidade e segurança da carne. Manejo e transporte inadequados são responsáveis por problemas de qualidade de carne induzidos pelo estresse, como o encolhimento da carcaça, pH elevado e danos à carcaça causados por contusões. A carne de animais contundidos não é apropriada para consumo humano e precisa ser descartada ou destinada a outros fins que geram menor retorno financeiro. De acordo com [3], em grande parte das fazendas e matadouros atualmente, observa-se o abuso de animais onde a falta de supervisão nos empregados ou métodos de manejo abusivos se tornaram uma prática normal da indústria. Problemas relacionados ao bem-estar ocorrem durante o transporte, tais como perdas por morte, lesões e contusões. Fatores esses que poderiam ser reduzidos por meio da supervisão dos transportadores, adequando a densidade de estocagem no veículo e treinando os funcionários para um manejo consciente e ético dos animais.

O trabalho realizado em [4] utilizou modelos de análise estatística para buscar a identificação de fatores relacionados com o número de contusões em processos de pré-abate. A partir dos resultados foi possível reduzir o número de contusões a partir do melhoramento nas instalações de embarque e de uma conscientização dos envolvidos com o manejo de bovinos, especialmente nas fêmeas, que apresentaram uma maior vulnerabilidade ao desenvolvimento de hematomas na carcaça. Em [5], foi investigado fatores de risco como possíveis causas de hematomas e pH muscular elevado sob condições comerciais de operação na Colômbia. As carcaças foram analisadas em termos de pH muscular e hematomas (local, tamanho, severidade e forma). Resultados indicaram que a densidade de carga de caminhões, número e condições das paradas durante o transporte de gado e o tempo de espera na planta aumentaram o risco de hematomas. O tempo de descanso na planta de abate superior a 18 horas e interrupções intermitentes durante o trânsito são fatores de risco para o aumento da incidência de contusões. Outro aspecto importante foi que novilhos apresentaram menor risco de apresentar um alto pH muscular.

A disponibilidade de dados de embarque, transporte, desembarque e espera introduz o desafio de analisar, entender e usar os dados para melhorias no processo. Diferentemente de técnicas que apontam relações lineares entre variáveis aleatórias, técnicas de inteligência artificial, incluindo mineração de dados, permitem que os dados possam ser classificados de forma a apresentar padrões interessantes ou servir como preditores de tendências. Diversas técnicas algorítmicas foram desenvolvidas para extrair informações e discernir padrões que podem ser úteis para o suporte a decisões [6]. Nessa linha, este trabalho pretende aplicar técnicas clássicas de Inteligência Artificial, nomeadamente o algoritmo de classificação *K-means*, Redes Bayesianas e Árvores de Decisão, para identificar as variáveis críticas de produção de hematomas no processo de transporte de gado bovino. O trabalho foi feito sobre a mesma base de dados usados em [4] e comparações com os resultados atingidos nos dois trabalhos são efetuados. O texto a seguir encontra-se organizado da seguinte forma: na Seção 2 são apresentados os procedimentos metodológicos usados, as características da base de dados trabalhada e os algoritmos usados; na Seção 3 são discutidos os resultados obtidos; a Seção 4 apresenta as considerações finais e os potenciais trabalhos futuros.

## 2 Materiais e Métodos

### 2.1 Coleta de Dados

O processo de coleta dos dados usados neste trabalho, oriundo de [4], foi organizado com a quantificação de contusões e hematomas, feita por um ser humano e registrada por avaliação da carcaça na linha de abate, após esfolamento animal e identificação da carcaça por meio de sua sequência de abate e numeração individual de lotes. Cada carcaça foi classificada como machucada ou normal (não machucada). Em uma segunda etapa, o número total de contusões foi contado para cada carga, de modo que o gado com várias contusões recebeu uma pontuação correspondente ao número de contusões presentes em cada animal ou carcaça. Não houve diferenciação sobre a localização dos hematomas no corpo do animal. O conjunto de dados foi coletado entre os meses de agosto a novembro de 2013, no Frigorífico Silva Indústria e Comércio LTDA, Serviço de Inspeção Federal (SIF) nº 1733, localizado no município de Santa Maria, no estado do Rio Grande do Sul. Nas unidades experimentais foram utilizados dados de 142 cargas de gado de abate, com uma média de 31,25 animais por carga, variando de 18 a 50 por carregamento e até 215 animais de um mesmo produtor, totalizando 4438 animais e 49,9% de fêmeas e 50,1% de carcaças de novilho, que foram transportadas separadamente.

Os dados utilizados na avaliação *ante mortem* foram relacionados às instalações de embarque e manuseio, transporte de animais, desembarque e manejo no matadouro. Esses dados foram avaliados utilizando formulários de registro gerados pelo sistema de compra da indústria, que foram preenchidos e assinados pelo pessoal de transporte. A planta tinha uma capacidade de abate de 600 cabeças de gado/dia, processando aproximadamente 100 animais por hora.

### 2.2 Seleção de Algoritmos

Neste trabalho foram escolhidos um algoritmo de agrupamento (*K-means*) e dois algoritmos de inferência probabilística, um baseado em probabilidades condicionais (redes Bayesianas) e outro baseado em Teoria da Informação (árvore de decisão). Estes algoritmos foram escolhidos por sua robustez na análise de diversos tipos de dados. Sendo este um trabalho aplicado e exploratório, entendeu-se pela busca de resultados que pudessem ser comparados com outros já estabelecidos.

O problema de *clustering* (ou de agrupamento) é definido como o problema de encontrar grupos similares em um determinado conjunto de dados. Cada um desses grupos é chamado de *cluster* e pode ser definido como uma região na qual a densidade de objetos é localmente mais alta do que em outras regiões [7]. Os algoritmos de *clustering* são usualmente não supervisionados: eles são apresentados a um conjunto de instâncias de dados que devem ser agrupados de acordo com alguma métrica de similaridade. O algoritmo tem acesso apenas ao conjunto de recursos que descrevem cada objeto, não é dada qualquer informação (por exemplo, rótulos) sobre onde cada uma das instâncias deve ser colocada dentro da partição. *K-means* é um algoritmo de *clustering* popular

que tem sido usado em uma variedade de domínios de aplicativos, como segmentação de imagem e recuperação de informações [8].

Uma rede bayesiana é um modelo gráfico que representa uma distribuição de probabilidade conjunta e que captura propriedades de independência entre as variáveis de um modelo. Esses modelos são atraentes por sua capacidade de descrever processos estocásticos complexos e porque fornecem uma metodologia clara para aprender com observações [9]. Os algoritmos de inferência da rede bayesiana mostraram-se particularmente promissores, pois, diferentemente da maioria dos *frameworks* de modelagem, podem capturar muitos tipos de relacionamentos entre variáveis. Devido à sua natureza probabilística, os algoritmos também são capazes de lidar com dados ruidosos, como encontrado em experimentos biológicos, podendo lidar com centenas de variáveis [10],[12]. O algoritmo *Naïve Bayes* é um classificador probabilístico que calcula um conjunto de probabilidades contando a frequência e as combinações de valores em um determinado conjunto de dados. O algoritmo assume que todos os atributos são independentes entre si, dado o valor da variável de classe. Essa suposição de independência condicional raramente é verdadeira em aplicações no mundo real, daí a caracterização como *naïve* (ingênuo), embora o algoritmo tenha um bom desempenho e aprenda rapidamente em vários problemas de classificação supervisionada [11]. Essa "ingenuidade" permite que o algoritmo construa facilmente classificações a partir de grandes conjuntos de dados sem recorrer a esquemas complicados de estimativa de parâmetros iterativos [6]. Para os fins desse trabalho, contudo, considerou-se inicialmente essa independência. Investigações sobre esse aspecto das variáveis estão, presentemente, sendo conduzidas.

Em um problema de classificação, todos os eventos possíveis caem exatamente em uma das classes pertencentes a um conjunto finito  $C$ , denominadas  $c_1, \dots, c_k$  [13].  $C$  é uma variável aleatória que toma os valores correspondentes a essas classes, enquanto  $X$  é uma variável aleatória de vetor cujos valores são relacionados ao domínio de observações do problema.  $P(C = c_k | X)$  é a probabilidade condicional de um valor pertencer à classe  $c_k$ , dado que sabemos que possui características do vetor  $X$ . O teorema de Bayes especifica como essa probabilidade condicional pode ser calculada a partir das probabilidades condicionais de observação de determinadas características dos valores do vetor para documentos de cada classe e a probabilidade marginal de observação de um elemento de cada classe.

A abordagem da árvore de decisão é mais poderosa para problemas de classificação. Há duas etapas nas técnicas: a construção da árvore a partir de um conjunto de dados de treinamento e aplicação da árvore ao conjunto de dados. Existem diversos algoritmos de árvore de decisão, tais como CART, ID3, C4.5, CHAID e J48 [14]. O classificador J48 é uma árvore de decisão baseada no algoritmo C4.5. Com essa técnica, uma árvore é construída para modelar o processo de classificação. Depois que a árvore é construída, ela é aplicada a cada tupla no banco de dados e resulta em uma classificação para essa tupla [11]. O algoritmo J48 é usado neste trabalho. De acordo com [14], o algoritmo J48 usa o método de remoção para construir uma árvore. A poda é uma técnica que reduz o tamanho da árvore, removendo dados com excesso de ajuste. O algoritmo J48 recursivamente classifica os dados até que eles tenham sido categorizados da maneira mais perfeita possível. Essa técnica fornece máxima precisão nos dados.

Para a análise dos dados foi utilizada a linguagem R, desenvolvida de forma colaborativa para o uso em computação estatística, análise de dados e representações gráficas. A linguagem permite a utilização de extensões, as quais possuem um conjunto de funções para áreas específicas. Foi utilizada a ferramenta de interface gráfica RStudio, para criar e alterar o script de manipulação do conjunto de dados que executa as funções da linguagem R.

### 3 Resultados e Discussões

No banco de dados utilizado, o número de contusões por carregamento é uma variável discreta. Porém, como o número de animais por carregamento varia a cada registro, o número de contusões nos carregamentos com maior número de animais geralmente resulta em um número absoluto maior de contusões. Para gerar um coeficiente de contusões por carregamento foi dividido o número de contusões pelo número de animais, gerando um coeficiente, o qual foi adicionado no banco de dados com o nome “Contusões”.

Para a avaliação dos dados usando redes bayesianas, foi utilizado o pacote para a linguagem R chamado *naivebayes*, que possui funções para a construção de classificadores probabilísticos baseados na aplicação do teorema de Bayes. As funções do pacote *naivebayes* obrigam que as classes tenham valores discretos. Com isso, a predição da variável “Contusões” não seria possível sem transformá-la para uma variável discreta/catórica. Para atingir este fim foi utilizada a técnica de *clustering* por meio de outra extensão da linguagem R chamada *Ckmeans.1d.dp* que implementa funções do algoritmo K-Means.

A imagem a seguir ilustra os três *clusters* criados a partir dos valores do coeficiente de contusões, agrupando os valores em três categorias as quais foram utilizadas posteriormente nas funções de predição e árvores de decisão como classes de contusões, sendo a classe 1 os coeficientes com menor número de contusões por animal, a classe 2 número médio de contusões e a classe 3 com o maior coeficiente.

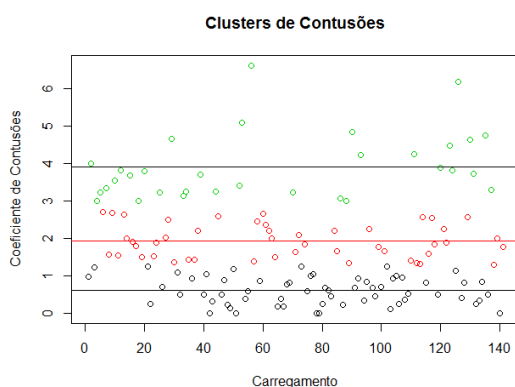
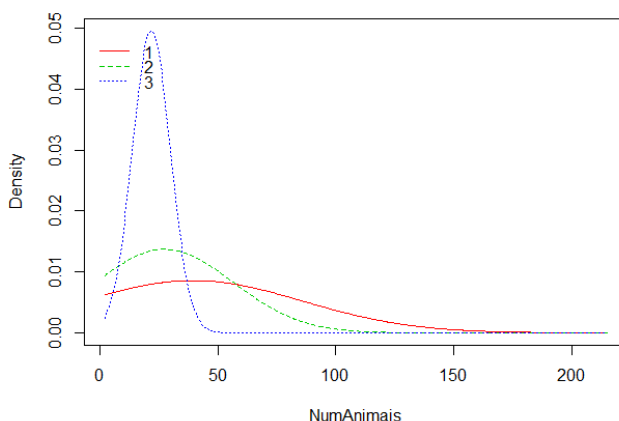


Fig. 1. Cluster K-Means.

O conjunto de dados após a classificação das contusões, foi submetido à função *naivebayes*, onde os preditores métricos são tratados assumindo a distribuição Normal, dado o rótulo de cada coluna do conjunto de dados. Na sequência é criada uma tabela que armazena o retorno da função *predict*, a qual calcula as probabilidades condicionais posteriores para cada classe usando o teorema de Bayes. Ao final da execução do script é executada a função *plot*, responsável pela impressão dos objetos da classe *naivebayes*, projetado para análise das probabilidades condicionais de variáveis predictoras, dada a classe.

A Fig. 2 apresenta a relação encontrada entre o número de animais do carregamento e as classes de contusões. É possível observar que a classe com maior número de contusões está relacionada com carregamentos de até 46 animais e as classes com número de contusões intermediária e baixa com uma menor concentração para carregamentos de até 100 e 150 animais de mesmo produtor, respectivamente. Esta informação é um indicativo que carregamentos de produtores rurais que possuem um menor número de animais tendem a apresentar um maior número de contusões, de acordo com os dados coletados. Note-se que, nesse caso, temos duas situações possíveis: a densidade de carga é menor no caminhão, o que indica que muito espaço entre os animais no transporte não é algo benéfico e existem carregamentos de produtores distintos, ocasionando que animais desconhecidos entre si viagem juntos.



**Fig. 2.** Contusões por número de animais carregados de cada produtor.

A Fig. 4 apresenta a relação encontrada entre o sexo e as classes de contusões. A classe 3 (contusões mais graves) possui uma maior concentração nas fêmeas, enquanto que as contusões dos machos concentram-se na classe 1 (contusões leves). A classe intermediária de contusões apresenta o mesmo tamanho em ambos os sexos. É possível identificar que os carregamentos compostos por fêmeas possuem um maior número de contusões se comparado com os machos. Essa questão está sendo investigada correntemente, para verificação das causas, junto a especialistas da área de Agrárias.

O mesmo conjunto de dados utilizado na abordagem *naivebayes* foi submetido à função J48 da extensão RWeka, que utiliza aprendizado em árvores de regressão e

classificação. Uma árvore de classificação é um modelo que prevê o rótulo de classe dos atributos. A árvore é construída dividindo repetidamente os dados em grupos com base em valores de atributos. O atributo dividido é selecionado pelo ganho de informações, uma técnica de Teoria da Informação, com objetivo de determinar qual divisão de atributos equilibrará melhor os dados.

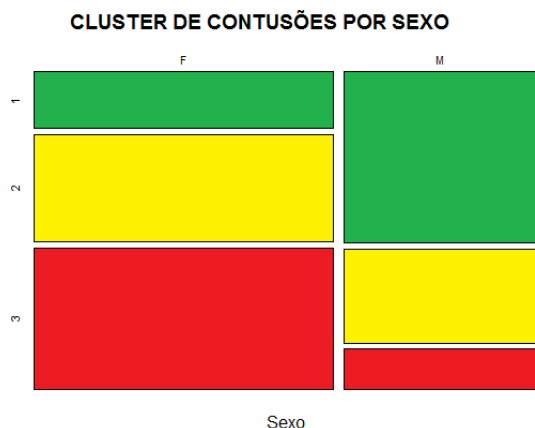


Fig. 3. Contusões por sexo.

Após a execução da função J48, os dados de saída são passados para a função *rpart plot*, responsável pela impressão dos nós e das folhas da árvore de decisão. O resultado obtido é apresentado na Fig. 4.

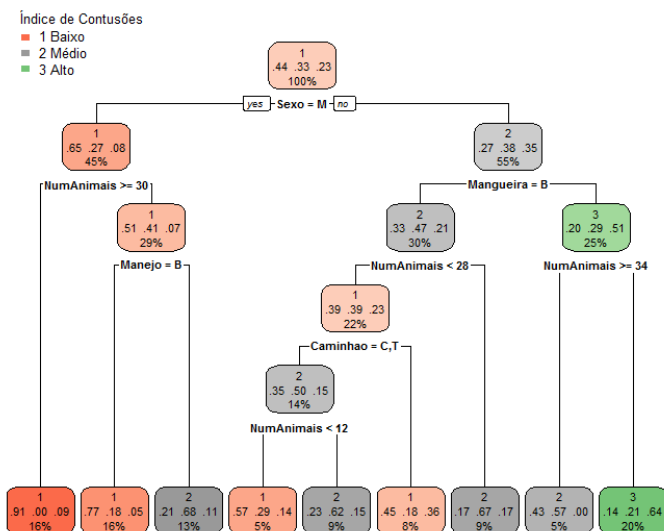


Fig. 4. Árvore de decisão

## 4 Conclusão

A análise efetuada das informações existentes no conjunto de dados coletado aponta a existência de variáveis com maior influência no número de contusões decorrentes do processo de transporte (que inclui o embarque na propriedade, o transporte propriamente dito, o desembarque e o período de espera no frigorífico). A partir dos resultados obtidos com as técnicas Naïve Bayes e Árvore de Decisão, observou-se que a densidade da carga e o sexo dos animais são os elementos mais determinantes. Não foram analisados, neste trabalho, as condições de embarque e o período posterior à chegada ao frigorífico. Também não foram levadas em consideração a localização das lesões no corpo do animal. As fêmeas tiveram um número de contusões superior aos novilhos machos, e as razões para esse achado encontram-se sob investigação, sendo considerado trabalho futuro. As baixas quantidades de animais embarcados resultam em um grande número de contusões, o que pode ser explicado tanto pela baixa densidade da carga (ocasionando quedas decorrentes da movimentação do caminhão) como pela mistura de diferentes lotes/produtores no mesmo caminhão, fazendo com que haja disputas entre animais que não se conhecem, fazendo com que os animais dominantes acabem contundindo os outros.

O trabalho mostra que é possível utilizar técnicas de Inteligência Artificial para determinar variáveis relevantes referentes aos índices de contusões no transporte de gado de corte. Os resultados obtidos, contudo, devem passar pelo crivo de especialistas na área, para verificação das explicações possíveis para os eventos descobertos. A partir dos resultados pode-se também detalhar outras variáveis e/ou qualificar o processo de coleta de dados. Este é um trabalho em andamento e um novo conjunto de dados coletados passará a ser analisado, de forma a permitir que o processo possa ser melhorado, em atenção às descobertas trazidas pelas técnicas utilizadas, atendendo ao objetivo traçado de auxiliar os produtores, empresas de transporte e matadouros a aprimorarem suas técnicas buscando um menor número de contusões e, conseqüentemente, uma maior qualidade e um maior rendimento de sua produção.

### Agradecimentos

Agradecemos ao pesquisador Fábio Mendonça pela cessão dos dados que possibilitaram a execução deste trabalho.

### Referências

1. G. C. Miranda-de la Lama, W. S. Sepúlveda, M. Villarroel, and G. A. María, Attitudes of meat retailers to animal welfare in Spain, *Meat Sci.*, vol. 95, no. 3, pp. 569–575, 2013.
2. T. Grandin, Animal welfare and society concerns finding the missing link, *Meat Sci.*, vol. 98, no. 3, pp. 461–469, 2014.
3. F. Adzitey, Effect of pre-slaughter animal handling on carcass and meat quality, *Int.*



- Food Res. J., vol. 18, no. 2, pp. 485–491, 2011.
4. F. S. Mendonça *et al.*, Pre-slaughtering factors related to bruises on cattle carcasses, *Anim. Prod. Sci.*, vol. 58, no. 2, pp. 385–392, 2016.
  5. M. H. Romero, L. F. Uribe-Velásquez, J. A. Sánchez, and G. C. Miranda-de la Lama, Risk factors influencing bruising and high muscle pH in Colombian cattle carcasses due to transport and pre-slaughter operations, *Meat Sci.*, vol. 95, no. 2, pp. 256–263, 2013.
  6. G. Dimitoglou, J. a Adams, and C. M. Jim, Comparison of the C4.5 and a Naive Bayes Classifier for the Prediction of Lung Cancer Survivability, *J. Neural Comput.*, vol. 4, no. 8, pp. 1–9, 2012.
  7. A. Likas, N. Vlassis, J. Verbeek, A. Likas, N. Vlassis, and J. Verbeek, The global k-means clustering algorithm, *Pattern Recognit*, vol. 36, no. 2, pp. 451–461, 2003.
  8. K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl, Constrained K-means Clustering with Background Knowledge, *Int. Conf. Mach. Learn.*, pp. 577–584, 2001.
  9. N. Friedman, M. Linial, I. Nachman, and D. Pe’er, Using Bayesian Networks to Analyze Expression Data, *J. Comput. Biol.*, vol. 7, pp. 601–620, 2000.
  10. J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink, and E. D. Jarvis, Advances to Bayesian network inference for generating causal networks from observational biological data, *Bioinformatics*, vol. 20, no. 18, pp. 3594–3603, 2004.
  11. T. R. Patil, Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification, *Int. J. Comput. Sci. Appl. ISSN 0974-1011*, vol. 6, no. 2, pp. 256–261, 2013.
  12. S. Dumais, J. Platt, D. Heckerman, and M. Sahami, Inductive learning algorithms and representations for text categorization, *7th International Conference on Information and Knowledge Management (CIKM'98)*. pp. 148–155, 1998.
  13. D. D. Lewis, Naive (Bayes) at forty: The independence assumption in information retrieval, *10th European Conference on Machine Learning*. pp. 4–15, 1998.
  14. C. S. Dangare and S. S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques, *Int. J. Comput. Appl.*, vol. 47, no. 10, pp. 44–48, 2012.