View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Directory of Open Access Journals

Revista Brasileira de Computação Aplicada, Julho, 2018

DOI: 10.5335/rbca.v10i2.8078
Vol. 10, Nº 2, pp. 80–90
Homepage: seer.upf.br/index.php/rbca/index

ISSN 2176-6649

ARTIGO ORIGINAL

Avaliação da performance do algoritmo J48 para construção de modelos baseados em árvores de decisão

Elamara Marama de Araujo Vieira, Nívea Trindade de A. T. Neves, Ana Carolina C. de Oliveira, Ronei Marcos de Moraes, João Agnaldo do Nascimento[1]

[1]Universidade Federal da Paraíba

*elamaravieira@gmail.com; niveamaria2@hotmail.com; carolyneoliveira@gmail.com; ronei@de.ufpb.br; joaoagh@gmail.com

Submetido: 28/04/2018. **Revisado:** 14/06/2018. **Aceito:** 04/07/2018.

Resumo

As árvores de decisão são modelos hierárquicos utilizados em várias áreas do conhecimento por sua capacidade preditiva e de resolução de problemas de maneira simples e objetiva. Entretanto, apresentam algumas limitações relacionadas à sua adequação à base de dados e ao se atentar quanto aos procedimentos para seleção dos parâmetros de crescimento e poda a serem adotados. Desta forma, têm-se como objetivo avaliar e discutir a performance do algoritmo J48 para construção de modelos de tomada de decisão em árvore em base de dados com atributos de diferentes tipos. Para tanto, realizaram-se experimentos em 10 bases de dados disponíveis em repositório internacional, considerando como variantes os métodos de treinamento, teste e poda, aplicados em toda base de dados e com o uso dos métodos Wrapper e CFS (*Correlation-based Feature Selection*) para seleção de atributos. Identificou-se que na presença de dados contínuos, os únicos modelos que apresentaram boa capacidade preditiva estiveram presentes em situações em que a grande quantidade de exemplos puderam compensar tal deficiência. Os modos de treinamento "validação cruzada" e "divisão por porcentagem" mostraram-se similares em suas predições quando ajustados a 10 *folds* e 75%, respectivamente. Ademais, a seleção de atributos não foi capaz de gerar melhores predições denotando que tal método de forma isolada não compensa possíveis inadequações nas bases de dados. Pode-se constatar que os resultados referentes à capacidade preditiva dos modelos são fortemente direcionados pelo quantitativo de exemplos pertencentes à base, presença de dados contínuos e de dados com ruído.

Palavras-Chave: Árvore de decisão. J48. Modelo de decisão.

Abstract

Decision trees are hierarchical models used in several areas of knowledge due to their predictive capacity and problem solving in a simple and objective way. However, they present some limitations related to their adequacy to the database and in regard to paying attention to the procedures for selection of growth and pruning parameters to be adopted. In this way, the objective is to evaluate and discuss the performance of the J48 algorithm for the construction of tree decision-making models in databases with attributes of different types. Therefore, experiments in 10 databases available in international repository were carried out, considering as variants the training, testing and pruning methods, applied throughout the database and using the Wrapper and Correlation-based Feature Selection (CFS) methods for attribute selection. It was identified that in the presence of continuous data, the only models that presented good predictive capacity were present in situations in which the large number of examples could compensate for such deficiency. The cross-validation and percentage split training modes were similar in their predictions when adjusted to 10 folds and 75%, respectively. Furthermore, the selection of attributes was unable to generate better predictions denoting that such a method, in an isolated way, does not compensate for possible inadequacies in the database. It can be realized that the results regarding the predictive capacity of the models are strongly guided by the number of examples belonging to the base, presence of continuous data and noisy data.

keywords: Decision tree. J48. Decision model.

1 Introdução

A tomada de decisão é um processo de escolha entre duas ou mais alternativas para alcançar um ou mais objetivos de forma eficiente, podendo utilizar-se para tanto de modelos que subsidiem a fase de escolha e implementação da decisão mais apropriada (Turban et al.; 2011).As árvores de decisão têm sido usadas como técnicas confiáveis para desenvolvimento de modelos preditivos de apoio à tomada de decisão por se tratarem de estruturas gráficas hierárquicas de fácil entendimento e aplicação, caracterizadas por segmentar dados heterogêneos de acordo com suas similaridades de maneira que se tornem mais homogêneos em relação à variável alvo (Cervantes et al.; 2015); (Ramya et al.; 2015). A relevância de tais árvores para a tomada de decisão se dá por sua capacidade preditiva, ou seja, a capacidade do modelo em predizer no presente as interações que ocorrerão no futuro com um nível de certeza, auxiliando na resolução de problemas em diferentes áreas (Evangeline and Sudhasini (2016); Sousa et al. (2011)),incluindo a área da saúde como apresentado nos trabalhos de (Funchal et al. (2015), Maciel et al. (2015) e Kamadi et al. (2016)).

Um dos mais conhecidos algoritmos de indução de árvore de decisão, o C4.5 criado inicialmente por Ross Quinlan na década de 90 e implementado na linguagem JAVA sob o nome "J48" vem sendo popularmente utilizado por apresentar um determinado padrão de comportamento em conjuntos de dados de diferentes formas de representação, ou seja, apresenta poucas restrições quanto às características dos atributos adotados mostrando-se adequado para os procedimentos envolvendo atributos qualitativos, contínuos e discretos, além disso, não exige uma distribuição de probabilidade específica (Chauhan and Chauhan (2013); Lin and Chen (2012)).

As principais vantagens de tal técnica estão relacionadas ao fato de que têm a capacidade de processar valores em falta ou dados com ruídos (incertos), e ainda, assim, terem resultados de alto desempenho com baixo custo computacional (Cervantes et al. (2015);Bhargava et al. (2013)). Entretanto, este método possui algumas limitações relacionadas, principalmente, a sua avaliação focal do potencial preditor de uma covariável em relação à variável alvo, ou seja, o quanto um nó é informativo para o outro sem, no entanto, atentar para a sequência que produz a melhor predição, resultando em soluções simplistas para problemas complexos, e que associado ao fato de cada tipo de variável apresentar um tratamento distinto em relação à construção do modelo (Nong; 2014) pode resultar em dificuldades para adequação à determinadas bases de dados.

Desta forma, considerando tais limitações e a gama de possibilidades que o método oferece em termos de modo de crescimento e de poda da árvore, questiona-se sobre qual comportamento de tais modelos ao se atentar para os procedimentos de seleção dos atributos relevantes, tendo em vista que problemas que envolvem uma grande quantidade de atributos podem se tornar demasiadamente complexos. Ademais, sabendo-se que critérios de partição dos dados e a escolha dos parâmetros a serem adotados para cada tipo de base refletem diretamente no sucesso e fracasso dos modelos questiona-se sobre qual a melhor configuração dos modos de treinamento, teste e poda da árvore de decisão, se a caracterização dos dados afeta a acurácia da classificação/predição e, finalmente, se o número de exemplos da base de dados afeta a acurácia da classificação. Logo, têm-se como objetivo avaliar e discutir a performance do algoritmo J48 para construção de modelos de tomada de decisão em árvore em bases de dados com atributos de diferentes tipos.

Nas seções subsequentes são apresentados um breve referencial teórico da literatura pertinente e atualizada em relação ao tema em questão, em seguida são expostos os procedimentos metodológicos que contemplam aspectos de treinamento e teste da árvore de decisão, métodos de seleção de atributos, processo de poda e medidas de acurácia. A seção dos resultados e discussão é descrita em dois momentos: Fase I, que compõe experimentos sem seleção de atributos e a fase II, que compõe experimentos com seleção de atributos. Por fim, uma breve conclusão dos achados relevantes.

2 Referencial teórico

Uma árvore de decisão é constituída de uma cadeia de nós de decisão, conectados por ramificações, estendendo-se desde o nó raiz até os nós folhas, tendo como requisitos básicos a existência de um atributo alvo (Last et al. (2016); Larose (2014)). Sua construção é viabilizada por algoritmos, dentre os quais o J48, um algoritmo de código aberto, implementado pelo *software* WEKA (Witten et al.; 2016), em que a estruturação do modelo adota a estratégia "dividir para conquistar", baseando-se no conceito de razão de ganho de informação que identifica por meio da redução de entropia o quanto informativo um atributo é, para então selecionar a separação ótima, ou seja, o quanto espera-se que a entropia se reduza caso um determinado nó seja escolhido para fazer a partição dos dados (Larose; 2014); (Bhargava et al. (2013); Lin and Chen (2012)). Para cada atributo no conjunto de dados a razão de ganho de informação é calculada como pode ser vista na equação 4 e utilizada como critério de partição do atributo, e com maior razão de ganho de informação o nó será utilizado como raiz (Lavanya and Rani (2011); Venkatadri and Lokanatha (2010)).

O J48 é um algoritmo que pode lidar tanto com atributos contínuos e discretos, quanto com valores categóricos e valores ausentes. O tratamento de atributos contínuos envolve a consideração de todos os valores presentes no conjunto de treinamento, fazendo com que sejam ordenados de forma crescente considerando todos os valores presente nos dados de treinamento e, após esta ordenação, seja selecionado o valor que favorecerá a redução da entropia (Camargo et al. (2016); Ramya et al. (2015)).)). Os cálculos para escolha do atributo referente ao nó raiz são realizados da seguinte maneira:

Equação 1 – Redução de Entropia: Cálculo Info(S) para identificar a classe no conjunto de treinamento

S:

$$Info(S) = -\sum_{i=1}^{k} \left\{ \left[\frac{freq(Ci,S)}{|S|} \right] log2 \left[\frac{freq(Cj,S)}{|S|} \right] \right\} \quad (1)$$

|S| é o número de casos no conjunto de treinamento; Ci é a classe: i = 1, 2, 3, ..., k, k = número de classes; freq (Ci,S) = número de casos em Ci.

Equação 2 – Redução de Entropia: Cálculo do valor da informação esperada, Infox(S), para o atributo X da partição S. Onde *n* é o número de valores possíveis que o atributo pode assumir, ou seja, o número de nós-filhos, *N* é o número total de objetos do nó-pai e *N* (*Si*) é o número de exemplos associados ao nó filho *Si*.

$$Info_X(S) = -\sum_{i=1}^{n} \left[\left(\frac{|Si|}{|S|} Info(Si) \right) \right] \quad (2)$$

Assim, na equação 3, o ganho de informação será dado por:

$$Ganho(X) = Info(S) - Info_X(S) \quad (3)$$

O uso do critério de ganho de informação para escolha do nó raiz da árvore favorece dados com grandes variações nos valores, podendo representar um viés. Assim, a razão de ganho de informação, representada pela equação 4, em que o denominador normaliza o conjunto de amostra de atributos que apresentam grandes variações, pode superar tal limitação ao suavizar favorecimentos que por ventura venham a acontecer e certificar que a melhor escolha tenha sido feita (Quinlan; 1993).

Equação 4 – Razão do ganho de informação (RG): ganho de informação de um atributo X normalizado pela medida de informação dividida.

$$RG = \frac{Ganho(X)}{\sum_{i=1}^{k} \left[\frac{|Si|}{|S|} log_2 \frac{|Si|}{|S|} \right]} \quad (4)$$

A partir deste processo, duas fases de construção da árvore podem ser identificadas: o crescimento, sub composto pelas fases de treinamento e teste; e a poda. Na fase de crescimento os dados são divididos em grupos, podendo ser direcionados para treinamento e aprendizado da estrutura, e ainda para o teste que identifica a capacidade preditiva da árvore (Last et al. (2016); Larose (2014)). A sequência de passos para construção e poda da árvore está abordada no pseudocódigo do algoritmo descrito por Camilo and Silva (2009), como exposto na Fig. 1.

aborda três possibilidades: 1) teste padrão em um atributo discreto, em que há um resultado e um ramo da árvore para cada possível valor desse atributo; 2) teste mais complexo baseado em um atributo discreto, em que os valores possíveis são atribuídos a um número variável de grupos; e 3) se um atributo "A" têm valores numéricos contínuos, um teste binário com resultados A<=Z e A>Z. Todos estes testes são avaliados da mesma forma e além disso, revelaram-se úteis para introduzir uma restrição adicional, ou seja, para qualquer divisão, pelo menos dois dos subconjuntos de um conjunto maior deve conter um número razoável de casos. Esta restrição, só acontece quando um conjunto for pequeno.

```
Input: um conjunto de dados D
begin
    Árvore={};
    if D é "puro" OU existe outro critério de parada then
        encerrar;
    foreach atributo a ∈ D do
        Calcular ganho de informação;
    end
    a_melhor = Melhor atributo de acordo com o calculo do ganho de informação;
    Árvore = Cria um nó baseado no a_melhor;
    D_v = Divide o subconjunto de D baseado no a_melhor;
    foreach D_v do
        Árvore_v = J48(D_v);
        Fixe a Árvore_v no galho obtido no passo anterior;
    end
end
return [Árvore]
```

Figura 1: Pseudocódigo do algoritmo J48

Quando as árvores de decisão são construídas, muitas das sub-árvores podem possuir ruídos (erros), isso acarreta em um problema conhecido como sobreajuste, que significa um aprendizado muito específico do conjunto de treinamento, ou ainda quando há muitos exemplos de uma classe e pouco de outras, não permitindo ao modelo generalizar e nem expressar o verdadeiro potencial de predição, o que é bastante comum em atributos com grande quantidade de valores possíveis. Para detectar e excluir essas sub-árvores, são utilizados métodos de poda da árvore, que fornecem uma maior precisão nas estimações na presença de conjunto de dados ruidosos, e que pode ser realizada durante o treinamento dos dados, chamada de poda em tempo real ou poda por redução de erros que avalia a taxa de erros da árvore em um conjunto de casos separados (*folds*) ou ainda após a total indução da árvore, chamada de pós-poda. Tais técnicas têm o objetivo de melhorar a taxa de acerto do modelo para novos exemplos e tornar a árvore mais simples e facilmente interpretável por parte do usuário (Ahlemeyer-Stubbe and Coleman (2014); Witten et al. (2016); Last et al. (2016)).

A poda em tempo real tenta decidir quando parar o desenvolvimento de sub-árvores enquanto a árvore ainda está sendo induzida. Nesta, o crescimento da árvore para quando não existe uma associação estatisticamente significativa entre um atributo e a classe de um nó em particular, ou seja, apenas atributos estatisticamente significativos mensurados pelo *fator de confiança* são permitidos de serem selecionados para o cálculo de razão de ganho de informação, conduzindo-se a um limite superior de confiança para a verdadeira taxa de erro, que será usada como uma estimativa da taxa de erro no nó. Para tanto, é necessário estimar a taxa de erro esperada de um determinado nó retendo dados e usando-os como um conjunto independente de teste. O verdadeiro problema irá acontecer quando não existir uma grande oferta de dados disponíveis,

visto que a árvore aprende a estrutura dos dados e se há grande quantidade em falta, as estimativas podem ser prejudicadas por limitar a quantidade dos dados que podem ser usados para o treinamento e para os testes (Ahlemeyer-Stubbe and Coleman (2014) Witten et al. (2016)).

Entretanto, a pós-poda atua de forma que a remoção dos nós não significantes seja feita por "*bottom up*" de maneira que a taxa de erro de cada nó filho seja usada para derivar o total de erros do nó pai. Duas operações bastante distintas têm sido consideradas: *substituição de subárvores*, que é a operação de poda primária em que a ideia é selecionar subárvores e substituí-las por folhas individuais e, a *elevação de subárvores* que é mais complexa e operacionalizada de modo que toda a subárvore de um dado nó seja elevada se fundindo com a subárvore precedente (Witten et al.; 2016).

Desta forma, a capacidade preditiva de uma árvore de decisão é o resultado da maneira como ela é construída considerando os métodos de crescimento (treinamento e teste) e poda, e a escolha de tais métodos a depender do ajuste a determinadas características das bases de dados podem otimizar os resultados de desempenho da árvore, gerar estruturas inteligíveis e aplicáveis à complexidade do mundo real ou, caso contrário gerar resultados desfavoráveis.

3 Metodologia

Trata-se de um estudo experimental envolvendo a aplicação do algoritmo classificador J48 para construção de modelos preditivos em árvores de decisão realizados em dez (10) bases de dados públicas (Tab. 1) disponíveis em repositório internacional (http://repository.seasr.org/Datasets/UCI/arff/) através do *software* WEKA (*Waikato Environment for Knowledge Analysis*) versão 3.8. As etapas do método de indução de árvore de decisão e testes realizados são expostas sequencialmente na Fig. 2.

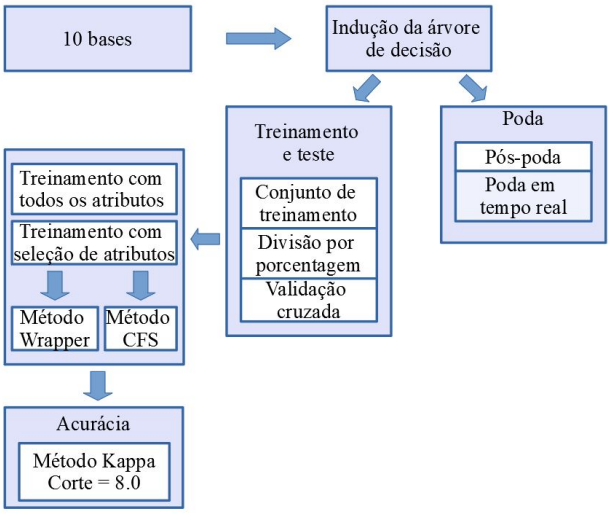

Figura 2: Fluxograma da pesquisa

3.1 Treinamento e teste da árvore

Os experimentos em cada base de dados ocorreram em duas fases: (1) com todos os atributos originais da base; e, (2) apenas com os atributos mais relevantes das bases identificadas através de métodos de seleção de atributos descritos na sequência.

Em cada fase supracitada a construção das árvores foi realizada nos modos "Conjunto de treinamento", que opera usando o conjunto de dados completo para treinamento e novamente para o teste, "Validação cruzada" que por sua vez, decidindo-se por um número fixo de partições de dados, mede-se a taxa de erro de um esquema de aprendizagem de uma partição de dados em particular em relação aos demais usados para o treinamento, para tanto usou-se 5, 10 e 15 *folds*, e por fim a "Divisão por porcentagem" que fornece um percentual dos dados para treinamento e outra para teste, adotando-se os valores de 25%, 50% e 75% para treinamento. Tais valores foram adotados considerando testes prévios em que se obtiveram as melhores estimativas.

3.2 Métodos de Seleção de atributos

No intuito de otimizar as estimativas das árvores e identificar os itens mais importantes para o tratamento dos dados, aplicou-se métodos de seleção de atributos, em que o grupo de atributos que melhor otimizasse a árvore seria selecionado para novas estimações. Para tanto, a seleção de atributos foi feita de acordo com os métodos Wrapper (5 *folds*) e CFS (*Correlation Feature Selection*) e a opção com o melhor valor do mérito do subconjunto de atributos identificado foi considerado para as novas predições.

é um método mais simples de avaliação em que se treina um classificador usando um conjunto de recursos desejado, e utiliza-se a precisão do classificador como uma medida da aptidão, dando um *feedback* direto sobre a capacidade de usar o conjunto de recursos, ou seja, executa o algoritmo de aprendizagem envolvido no processo de seleção, neste caso o J48, realizando uma avaliação independente e gerando um subconjunto de atributos mais relevante para a escolha da classe. O processo ocorre com as estimativas de precisão fornecidas pelo algoritmo de aprendizagem alvo em que o subconjunto com melhor pontuação (maior taxa de sucesso) é selecionado, esse processo é repetido para cada subconjunto de atributos até que o critério de parada determinado seja satisfeito (Li et al.; 2014).

Enquanto que método CFS, apesar de ser chamada de seleção de características de correlação, usa informação mútua como medida de correlação, que contorna o problema da característica correlacionada com valores negativos, tornando a medida mais estável e menos propensa a selecionar características aleatórias, ou seja, é um método que tenta encontrar um conjunto de características tais que estas avaliam a capacidade preditiva atribuindo pontuações mais altas de cada atributo individual e o grau de consistência nos valores entre eles, preferindo conjuntos de atributos que são altamente correlacionados com a classe, e baixa intercorrelação com outros conjuntos de atributos. Quando os exemplos de treinamento são projetados sobre o conjunto, a consistência de qualquer subconjunto de

Tabela 1: Características das bases de dados

Base de dados	Atributo	Qto	Tipo de dados	% dados ausentes	Classes
Arrhythmia	280	452	Discretos, contínuos e categóricos	2 – 83%	16
Breast-cancer	10	286	Categóricos	3%	2
Dermatology	35	366	Discretos e categóricos	2%	6
Diabetes	9	768	Discretos, contínuos e categóricos	0%	2
Ecoli	8	336	Discretos, contínuos e categóricos	0%	8
Hypothyroid	30	3772	Discretos, contínuos e categóricos	0% - 20%	4
Nursery	9	12960	Categóricos	0%	4
Postoperativepatient-data	9	90	Categóricos	0% - 3%	3
Vote	17	435	Categóricos	0% - 24%	2
Zoo	18	101	Discretos e categóricos	0%	4

Legenda: Qto = quantitativo de exemplos.

atributos nunca pode ser melhor do que ele todo. É um método em que a correlação é medida usando a incerteza simétrica (medida de correlação bastante utilizada na avaliação de atributos, e que é calculada utilizando-se outras duas medidas: o ganho de informação e entropia) medida pela Equação 5, em que F é o número de variáveis ou "características"em cada amostra, X é uma matriz F x N, onde N é o número de amostras, Y é um vetor de dimensão N de classes e, finalmente "j̄sū" refere-se as incertezas simétricas entre os atributos (França et al. (2015); Freeman et al. (2015); Witten et al. (2016)).

$$J_{CFS}(S_m) = \frac{Fjsu(\bar{X_i}, Y)}{\sqrt{F + F(F-1)jsu(\bar{X_i}, X_j)}} \quad (5)$$

Tem-se como *output* de ambos os métodos de seleção o número de conjuntos avaliados e o de atributos a ser considerado com seu correspondente valor de "mérito" que se refere a uma medida de correlação entre o conjunto de atributos e as classes; quanto maior o valor de mérito, maior será a correlação entre as classes. Após a filtragem dos atributos o processo de geração da árvore de decisão foi reiniciado aplicando os mesmos padrões da fase anterior.

3.3 Processo de poda

A poda foi aplicada considerando as categorias pós-poda no modo *"elevação de subárvores"* e poda em tempo real. Na pós-poda estabeleceu-se um fator de confiança de 1.0 e número mínimo de casos por folha (MNO) com valores variantes de 2, 5 e 10, enquanto que na poda em tempo real dois modos de testes foram realizados: (1) estabelecendo-se o número de fixo 2 *folds* e testando-se para cada número mínimo de casos por folha (considerando os valores de 2, 5 e 10) os fatores de confiança de 0.1 a 0.9 com incrementos de 0.2; e (2) estabelecendo-se o número de fixo de fator de confiança de 0,1 e testando-se para cada número mínimo de casos por folha (considerando os valores de 2, 5 e 10) aplicados para número de *folds* de 3, 5, 7 e 9.

3.4 Medidas de acurácia

Para avaliação do potencial preditivo dos modelos foi usado como indicador a estatística Kappa, que compara o sistema de exatidão com o sistema aleatório, ou seja, é obtida a partir da matriz de classificação (ou de confusão), que é formada pelos erros e acertos das decisões do modelo, de tal modo que os acertos estão dispostos na diagonal principal da matriz e os erros fora dela. É uma medida de qualidade do modelo na qual se verifica o comportamento das decisões, e testa sua confiabilidade e precisão quando pondera sua concordância considerando os erros e acertos da decisão a partir de uma referência (Duda et al. (2012); Ramya et al. (2015)). Tal medida é obtida pela equação 6, a acurácia total e aleatória é dada pelas equações 7 e 8, respectivamente, em que o n é o número de colunas e linhas em uma matriz de classificação; m_{ij} é o elemento (i, j) da matriz de classificação; N é o número total de observações.

$$Kappa = \frac{Totalaccuracy - Randomaccuracy}{1 - Randomaccuracy} \quad (6)$$

$$Totalaccuracy = \frac{1}{N}\sum_{i=1}^{n} mii \quad (7)$$

$$Randomaccuracy = \frac{1}{N^2}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} m_{jk} * \sum_{j=1}^{n} m_{kj}\right) \quad (8)$$

O coeficiente Kappa varia entre 0 (discordância total) e 1 (total concordância), e seu valor é padronizado, sendo assim interpretado da mesma forma em vários estudos. Segundo Cohen (1960), o idealizador do coeficiente Kappa, seu valor sendo superior a 0,8 é tido como associação quase perfeita entre predição do modelo e o real. No presente estudo, considerou-se como ponto de corte de predições aceitáveis valores superiores à 0,8 (McHugh; 2012).

4 Resultados e discussão

Testou-se em cada fase de construção das árvores de decisão um total de 216 experimentos em cada base de dados considerando as várias formas de aplicação dos métodos de treinamento e poda propostos. Os métodos mais eficazes, que produziram experimentos com valor de estatística Kappa mais altos, serão

Tabela 2: Experimentos mais eficazes sem seleção de atributos

Base	Treinamento e teste	Kappa	Fator de confiança	Folds	NMO
Arrhythmia	Conjunto de treinamento	0,894*	1.0	NA	2
	Validação cruzada - 15 *folds*	0,576	0,1	3	5
	Divisão por porcentagem – 75%	0,629	0,1	5	2
Breast-cancer	Conjunto de treinamento	0,692*	0,7 - 1.0	2	2
	Validação cruzada - 15 *folds*	0,289	0,1	2	10
	Divisão por porcentagem – 25%	0,88	0,1	3	5
Dermatology	Conjunto de treinamento	0,979*	0,7 - 1.0	2	2
	Validação cruzada - 15 *folds*	0.935	0.1	3/7	2
	Divisão por porcentagem – 50%	0,890*	0,1 - 1.0	3/9	2
Diabetes	Conjunto de treinamento	0,638*	0,3 - 1.0	2	2
	Validação cruzada - 15 *folds*	0.469	0.3	2	10
	Divisão por porcentagem – 75%	0,529	NA	3	10
Ecoli	Conjunto de treinamento	0,913*	0,7 - 1.0	2	2
	Validação cruzada - 10 *folds*	0.782	0,3	2	2
	Divisão por porcentagem – 75%	0,779	0,7 - 1.0	2	5
Hypothyroid	Conjunto de treinamento	0,993	0,5	2	2
	Validação cruzada - 10 *folds*	0.971*	0,3	2	2/5
	Divisão por porcentagem – 75%	0,984*	0,3	3	2/10
Nursery	Conjunto de treinamento	0,993*	NA	NA	NA
	Validação cruzada - 10 *folds*	0.983	0,1	2	2
	Divisão por porcentagem – 75%	0,971	0,1	3/9	2/5/10
Postoperativepatient-data	Conjunto de treinamento	0,528*	NA	NA	NA
	Validação cruzada - 10/15 *folds*	0*	0,1	3	2
	Divisão por porcentagem – 75/50%	0*	0,1-0,3	3	2
Vote	Conjunto de treinamento	0,956*	NA	NA	NA
	Validação cruzada - 10*folds*	0,937*	NA	NA	NA
	Divisão por porcentagem – 75%	0,943*	0,1	3	5
Zoo	Conjunto de treinamento	0,987	0,3	2	2
	Validação cruzada - 10*folds*	0,922	0,1	7	10
	Divisão por porcentagem – 75%	0,947*	0,3/0,7/0,9	2	2/5/10

*O mesmo valor esteve presente na ausência de poda.

dispostos e comentados nas seções seguintes.

4.1 Qual a melhor configuração dos modos de treinamento, teste e poda da árvore de decisão?

A Tab. 2 apresenta os experimentos que resultaram nos melhores índices Kappa para cada base de dados, cada modo de treinamento e teste. Nesta tabela pode-se identificar, além do valor Kappa mais alto para cada base, o modo de treinamento e teste relacionado a tal valor, e os métodos de poda administrados, ou seja, por meio da pós-poda (usando o fator de confiança de 1.0 e variações do valor de NMO), poda em tempo real com 2 *folds* (usando o fator de confiança entre 0,1 e 0,9 e variações do valor de NMO) e a poda em tempo real com número de *folds* variável (usando fator de confiança de 0.1 e variações do valor de NMO). Por vezes o mesmo valor Kappa pôde ser encontrado em mais de um teste e desta forma os modos identificados são considerados na tabela podendo estar no formato de intervalo de valores, como frequentemente visualizado na coluna "Fator de confiança" (como por exemplo o intervalo de fatores entre 0,7 - 1,0 na base *Arrhythmia*, modo de teste "conjunto de treinamento") ou em valores individuais (denotados pela "/"). O termo "NA" refere-se a valores que não se aplicam para determinada poda ou ainda que não foram podados.

A partir da observação de alguns detalhes referentes à Tab. 2 pode-se identificar que os experimentos com melhores índices Kappa estiveram presentes quando o modo de conjunto de treinamento foi utilizado. Entretanto, os valores Kappa deste modo de treinamento não são susceptíveis para serem bons indicadores de desempenho futuro, visto que classificam dados que foram treinados e testados com o mesmo conjunto de exemplos fazendo com que as estimativas de acurácia sejam demasiadamente otimistas. Desta forma, e como pode-se constatar nos dados expostos, o desempenho do modo de "conjunto de treinamento" tem um indicador de acurácia muito satisfatório se comparado a um conjunto que testa dados independentes, tal como a validação cruzada ou divisão por porcentagem, porém pouco confiáveis pois são necessárias formas de prever os limites de acurácia com base em experiências que não tenham sido testados pelos mesmos dados que os modelaram. Ademais, o modo "conjunto de treinamento" apresenta os melhores resultados associados à parâmetros de pós-poda ou ainda à poda em tempo real com *fold* fixo de 2, denotando que a poda em tempo real com *fold* variável, ou seja, com valores acima de 2, não seria a melhor escolha nestes casos.

4.2 A caracterização dos dados afeta a acurácia da classificação/predição?

Os tipos de dados de cada base foram fatores influentes tendo em vista que os conjuntos de dados em que há a presença de três tipos de

dados, ou seja, categóricos, discretos e contínuos, apenas conseguiram apresentar um bom valor Kappa no modo "conjunto de treinamento", enquanto que as bases que apresentaram unicamente variáveis categóricas e discretas, apresentaram bons indicadores em todos os três modos de treinamento. As bases que tiveram apenas dados contínuos (Tab. 1) conseguiram uma boa predição apenas quando continham um grande número de exemplos. Tal constatação é corroborada por pesquisa de Drakakis et al. (2016) que identifica a deficiência do J48 em lidar com dados contínuos, podendo apresentar uma redução em até 20 pontos na acurácia do modelo. Entretanto, para bases que contavam apenas com variáveis discretas Salami et al. (2016) identificou uma acurácia sempre superior a 90%.

4.3 O número de exemplos da base de dados afeta a acurácia da classificação?

As duas bases com maiores quantitativo de exemplos ("*Hypothyroid*" e "*Nursery*"), foram os que conseguiram ter as melhores predições nos três modos de teste. Este quantitativo de exemplos se tornou um diferencial em relação às demais bases. Segundo Ismail et al. (2012) o quantitativo de exemplos de dados de fato afeta a precisão, assim um conjunto com uma maior quantidade de exemplos factíveis, ou seja, exemplos sem dados perdidos (ausência de dados) e sem erros, produzem indicadores de desempenho mais elevados. Dessa forma, quanto menor o quantitativo de exemplos de dados disponíveis ou maior quantidade de valores perdidos, menor será a taxa de acurácia do modelo (Ahlemeyer-Stubbe and Coleman (2014) Witten et al. (2016)). Ademais, a porcentagem de dados perdidos não mostrou nestes experimentos ter forte influência no poder preditivo dos testes, tendo em vista que as melhores predições foram obtidas em bases com perdas significativas de dados, de até 24%. Assim, observa-se claramente que outros fatores puderam compensar tais perdas, como por exemplo, a grande quantidade de exemplos ou ainda a ausência de variáveis contínuas.

Algumas outras observações pertinentes a partir dos resultados podem ser elencadas: (1) No modo de teste de validação cruzada, os resultados que obtiveram os melhores índices Kappa foram realizados em bases de dados com grande quantidade de exemplos e especialmente em experimentos com 10 *folds*, podendo estar associados ao método de poda em tempo real com *folds* fixo com fator de confiança de 0,1 a 0,3, poda em tempo real com *folds* variável de 3 a 7 *folds* ou ainda sem utilização da poda. Segundo Larose and Larose (2015), alguns testes com vários conjuntos de dados e técnicas de aprendizagem diferentes, mostraram que o uso de 10 *folds* na validação cruzada tem a melhor estimativa de erro. Embora não seja uma forma conclusiva e ainda estejam sendo feitos estudos para adequar o melhor esquema para avaliação, a validação cruzada com 10 *folds* torna-se o método mais usado, mas não se pode também descartar que 5 e 20 *folds* de validação cruzada pode ser tão bom quanto o valor de 10. Independentemente do número de *folds* utilizado na validação cruzada, na pesquisa de Sharma and Sahay (2016) o J48 apresentou acurácia semelhante, sempre acima 0,97, estando entre os cinco melhores classificadores dos treze investigados, isto quando a bases de dados apresentou variáveis apenas categóricas e discretas, entretanto quando a base de dados apresentou os três tipos de variáveis e, especialmente com número reduzido de exemplos distribuídos para cada classe a acurácia do modelo foi insuficiente Maciel et al. (2015), que, novamente, se considerando os nossos resultados pode denotar uma deficiência deste algoritmo para lidar com variáveis contínuas; (2) A validação por porcentagem trouxe melhores resultados quando adotado uma divisão à 75% dos dados podendo estar associada aos três tipos de poda e favorecida se utilizada com um fator de confiança de 0,3, número mínimo de objetos 2 e, se a escolha for na poda em tempo real utilizando-se 3 *folds* ; (3) Ao adotar-se o método de poda em tempo real com número fixo de 2 *folds*, os valores que otimizaram a preditividade do modelo estiveram associados à índices de confiança menores ou iguais a 0,5; (4) A pós-poda quando usada em associação ao modo de treinamento e teste de validação cruzada não trouxe bons resultados; e, finalmente (5) a poda em tempo real com 3 e 5 *folds* e, 2 a 10 números mínimos de objetos estiveram presente sempre associada aos mesmo valor kappa da poda em tempo real com número fixo de 2 *folds*, denotando a presença de resultados semelhantes em condições com dados de boa qualidade.

Vale ressaltar que os valores de Kappa resultantes do uso da validação cruzada com 10 *folds* e da divisão por porcentagem à 75% são bastante similares entre si ao usar-se a mesma medida de poda, ou seja, 0,1 a 0,3 com 2 números mínimos de objetos. Um outro ponto a ser considerado é que as bases com grandes quantidades de atributos e dados ruidosos, a exemplo da base *Arrhythmia* e *Diabetes*, apresentaram desempenho insatisfatório, o que segundo Kandhasamy and Balamurali (2015) são causas determinantes para a acurácia do modelo e capacidade preditiva.

4.4 Os métodos de seleção de atributos melhoram os resultados do algoritmo J48?

Dois métodos para seleção de atributos foram testados para otimizar os resultados de treinamento das árvores e, para efetiva seleção considerou-se o método com melhores previsões de acurácia e potenciais resultados da árvore de cada base de dados. A Tab. 3 expõe o valor de mérito, o número de conjuntos de atributos avaliados por cada método e o número final de atributos selecionados.

O método Wrapper apresentou melhores indicadores em 9 das 10 bases avaliados e desta forma foi predominantemente o método de escolha para a seleção de atributos com melhor aderência em relação ao atributo-alvo. Faz-se exceção à base "*Zoo*" em que o método CFS mostrou melhor valor de mérito, e sabendo que tal método opera identificando a capacidade de previsão individual baseando-se em característica de correlação, entende-se que tal base de dados é composta por atributos altamente correlacionados com a variável alvo. As bases

Tabela 3: Métodos de seleção de atributos

Base	Wrapper		CFS	
Arrhythmia	Mérito	0.742	Mérito	0.472
	N° conjuntos avaliados	3556	N° conjuntos avaliados	7941
	N° atributos selecionados	8	N° atributos selecionados	26
Breast-cancer	Mérito	0.759	Mérito	0.097
	N° conjuntos avaliados	70	N° conjuntos avaliados	47
	N° atributos selecionados	3	N° atributos selecionados	5
Dermatology	Mérito	0.962	Mérito	0.769
	N° conjuntos avaliados	368	N° conjuntos avaliados	506
	N° atributos selecionados	8	N° atributos selecionados	19
Diabetes	Mérito	0.746	Mérito	0.164
	N° conjuntos avaliados	55	N° conjuntos avaliados	37
	N° atributos selecionados	4	N° atributos selecionados	4
Ecoli	Mérito	0.827	Mérito	0.649
	N° conjuntos avaliados	28	N° conjuntos avaliados	30
	N° atributos selecionados	5	N° atributos selecionados	6
Hypothyroid	Mérito	0.997	Mérito	0.604
	N° conjuntos avaliados	303	N° conjuntos avaliados	165
	N° atributos selecionados	8	N° atributos selecionados	5
Nursery	Mérito	0.966	Mérito	0.581
	N° conjuntos avaliados	37	N° conjuntos avaliados	37
	N° atributos selecionados	8	N° atributos selecionados	1
Postoperativepatient-data	Mérito	0.711	Mérito	0.053
	N° conjuntos avaliados	36	N° conjuntos avaliados	39
	N° atributos selecionados	0	N° atributos selecionados	6
Vote	Mérito	0.968	Mérito	0.729
	N° conjuntos avaliados	147	N° conjuntos avaliados	85
	N° atributos selecionados	5	N° atributos selecionados	4
Zoo	Mérito	0.406	Mérito	0.864
	N° conjuntos avaliados	81	N° conjuntos avaliados	149
	N° atributos selecionados	0	N° atributos selecionados	10

Tabela 4: Experimentos mais eficazes sem seleção de atributos

Base	Treinamento e teste	Kappa	Fator de confiança	Folds	NMO
Arrhythmia	Conjunto de treinamento	0,822*	0,7-1,0	2	2
	Validação cruzada - 15 *folds*	0,646	0,1	2	2
	Divisão por porcentagem – 75%	0,622	0,1	2	2
Breast-cancer	Conjunto de treinamento	0,289*	NA	NA	NA
	Validação cruzada - 10/15 *folds*	0,289*	NA	NA	NA
	Divisão por porcentagem – 75%	0,331*	NA	NA	NA
Dermatology	Conjunto de treinamento	0,969*	0,1 - 1.0	3/7	2/5
	Validação cruzada - 15 *folds*	0.955	0,1	5	2
	Divisão por porcentagem – 75%	0,959*	0,1 - 1.0	2	2
Diabetes	Conjunto de treinamento	0,545*	0,1 - 1.0	2	2
	Validação cruzada - 5 *folds*	0.468	0.3	2	2/5
	Divisão por porcentagem – 75%	0,539	0,1	3	2/5
Ecoli	Conjunto de treinamento	0,897	0,9	5/7/9	2/5/10
	Validação cruzada - 10 *folds*	0.774	0,1	9	10
	Divisão por porcentagem – 25%	0,779	0,7 - 1.0	7/9	5/10
Hypothyroid	Conjunto de treinamento	0,987*	NA	NA	NA
	Validação cruzada - 10/15 *folds*	0.971	0,1	3/7	2
	Divisão por porcentagem – 75%	0,964	0,1-0,5	5	2/5
Vote	Conjunto de treinamento	0,942	0,1-0,9	5/7/9	2/5/10
	Validação cruzada - 10*folds*	0,942	0,1-0,3	3/9	2/10
	Divisão por porcentagem – 25%	0,961	0,1/0,7-1.0	7/9	2/5/10
Zoo	Conjunto de treinamento	0,974	0,1/0,9	5/7/9	2/5/10
	Validação cruzada - 10/15*folds*	0,909	0,1-1.0	2	2/5/10
	Divisão por porcentagem – 25%	0,947	0,7-1.0	7/9	2/5/10

Legenda: * O mesmo valor esteve presente na ausência de poda; % todos os experimentos realizados neste modo resultaram no mesmo valor Kappa; NA = não se aplica

"*Nursery*" e "*Postoperativepatient-data*" apresentaram resultados peculiares em relação aos demais pois seu melhor valor de mérito esteve presente quando todos os atributos existentes na base original foram utilizados indicando a não necessidade de uma seleção. Desta forma em ambos as bases apenas a primeira fase foi completada.

Diversos trabalhos têm divulgado a superioridade do método Wrapper em relação a outros métodos para seleção de atributos. A exemplo, França et al. (2015) mostra em seu trabalho um incremento no poder preditivo do modelo em árvore de decisão com a seleção de atributos por meio do método Wrapper reduzindo em cerca de 9% o erro em relação às predições sem seleção de atributos e, em cerca de 2,5% em relação à seleção realizada pelo método CFS. A Tab. 4 expõe os novos testes em que os melhores valores Kappa foram encontrados considerado a etapa de seleção de atributos.

A partir destes valores o que se pode observar é que nas bases de dados em que os índices Kappa foram inferiores a 0,8 a seleção de atributos não foi capaz de gerar novos experimentos que superassem os experimentos da primeira fase, ou seja, não foi eficaz para otimizar as predições realizadas pela rede, o que é verdadeiro para quase todas as estruturas de base de dados considerada. Uma exceção encontra-se no melhor teste da base "*Arrhythmia*" em que usando o modo de aprendizagem e teste por meio da validação cruzada com número de 15 *folds* o índice Kappa obteve um leve incremento de 0,576 para 0,646, e também para a base "*Dermatology*" no modo de aprendizagem e teste por meio da "divisão por porcentagem" que aumentou o índice Kappa de 0,890 para 0,959.

Contraditoriamente, a base "*Breast-cancer*", apresentou uma redução do valor Kappa em relação a fase prévia. Isto pode ter ocorrido devido a uma substancial redução do número de atributos, passando de 10 na fase anterior para 3 nesta nova fase, tornando-se um número muito baixo para construção de um modelo que tenha boa capacidade preditiva. É importante ressaltar que o conhecimento do problema a ser investigado tem especial importância na fase de seleção de atributos. Apesar da existência e eficácia de meios automáticos como os apresentados nesta seção o conhecimento dos pormenores torna o modelo mais adequado à complexidade do mundo real.

Desta forma, não se observa na Tab. 4 grandes variações nos valores das melhores predições, entretanto, entende-se que são valores individuais que, em alguns casos, já mostram uma capacidade preditiva muito alta e, desta forma, podem não representar o real efeito dos métodos de seleção sobre a gama de possibilidades metodológicas que o modelo pode oferecer. Por este motivo, investigou-se os possíveis incrementos preditos dos métodos de seleção de atributos considerando todos os experimentos realizados. Isso foi feito por meio de uma plotagem em gráfico de dispersão pré e pós seleção de atributos em todos as bases de dados considerando o valor Kappa com ponto de corte de 0,8, como mostrado na Fig. 3. Observa-se em tal gráfico que existe uma diferença substancial no que diz respeito à quantidade de possibilidades com boa capacidade preditiva quando o método de seleção de atributos foi realizado, ou seja, a seleção de atributos não pôde gerar incrementos nos melhores teste das bases de dados, mas possibilitou que uma ampla gama de possibilidades, incluindo o método de treinamento e teste sem poda fosse usado, garantindo uma boa capacidade preditiva por aumentar a quantidade

de testes que gerem valores Kappa acima de 0,8. Tais resultados foram verdadeiros apenas para as bases *Dermatology*, *Vote* e *Zoo* que, vale ressaltar, não apresentam variáveis contínuas ou dados com ruído.

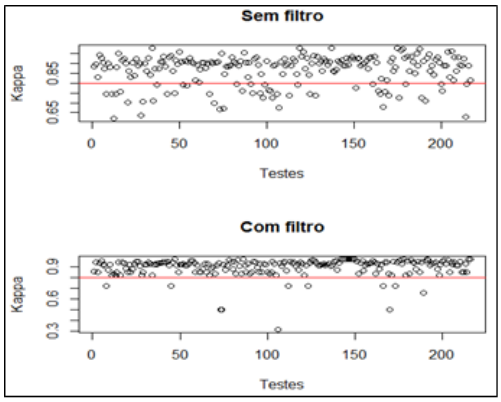

Figura 3: Poder preditivo dos testes com e sem seleção de atributos

Desta forma, pode-se constatar que o método de seleção de atributo não é uma solução única que poderá resultar em grandes incrementos na capacidade preditiva dos modelos em árvores de decisão independentemente de como os dados estão arranjados. Tais medidas requerem uma análise detalhada, além da monitorização da maneira como é realizada no intuito de que atenda as expectativas e se adeque à realidade e complexidade de um problema.

5 Conclusão

O objetivo deste trabalho foi avaliar e discutir a performance do algoritmo J48 para construção de modelos de tomada de decisão em árvore em bases de dados com atributos de diferentes tipos. Para tanto, realizou-se um total de 216 experimentos em diversas bases de dados, em duas fases, ou seja, considerando todos os atributos das bases e também os selecionando através de métodos automáticos. Pode-se constatar que os resultados referentes a capacidade preditiva dos modelos são fortemente direcionados pelo quantitativo de exemplos pertencentes à base, presença de dados contínuos e de dados com ruído. Em conjuntos de dados em que a presença de dados contínuos foi constatada (Bases "*Arrhythmia*, *Diabetes*", "*Ecoli*" e "*Hypothyroid*") o modelo apenas apresentou boa capacidade preditiva em situações de que a grande quantidade de exemplo compensou tal deficiência. Os modelos de treinamento "validação cruzada" e "divisão por porcentagem" mostraram-se bastante similares em suas predições quando ajustados à 10 *folds* e 75%, respectivamente, sendo no primeiro caso o método de poda mais indicado a modalidade poda em tempo real com 2 *folds* e fator de confiança inferior a 0,5, ou ainda com 3 e 5 *folds* utilizando-se 2 a 10 números mínimos de objetos. Ademais, a seleção de atributos não pôde compensar as imperfeições de bases de dados inadequados, entretanto, tal medida pode ofertar uma maior gama de resultados consistentes considerando o leque de possibilidades metodológicas disponíveis.

Referências

Ahlemeyer-Stubbe, A. and Coleman, S. (2014). *A practical guide to data mining for business and industry*, John Wiley & Sons.

Bhargava, N., Girja, S., Ritu, D. B. and Manisha, M. (2013). Decision tree analysis on j48 algorithm for data mining, *International Journal of Advanced Research in Computer Science and Software Engineering (JARCSSE)* **3**(6).

Camargo, A., Silva, R., Amaral, É., Heinen, M. and Pereira, F. (2016). Mineração de dados eleitorais: descoberta de padrões de candidatos a vereador na região da campanha do rio grande do sul, *Revista Brasileira de Computação Aplicada* **8**(1): 64–73.

Camilo, C. O. and Silva, J. C. d. (2009). Mineração de dados: Conceitos, tarefas, métodos e ferramentas, *Universidade Federal de Goiás (UFC)* pp. 1–29.

Cervantes, J., Lamont, F. G., López-Chau, A., Mazahua, L. R. and Ruíz, J. S. (2015). Data selection based on decision tree for svm classification on large data sets, *Applied Soft Computing* **37**: 787–798.

Chauhan, H. and Chauhan, A. (2013). Implementation of decision tree algorithm c4. 5, *International Journal of Scientific and Research Publications* **3**(10).

Drakakis, G., Moledina, S., Chomenidis, C., Doganis, P. and Sarimveis, H. (2016). Decision trees for continuous data and conditional mutual information as a criterion for splitting instances, *Combinatorial chemistry & high throughput screening* **19**(5): 423–428.

Duda, R. O., Hart, P. E. and Stork, D. G. (2012). *Pattern classification*, John Wiley & Sons.

Evangeline, S. B. and Sudhasini, P. (2016). An introduction to decision tree algorithm on various field of applications, *International Journal of Digital Communication and Networks (IJDCN)* **13**(3).

França, D. G., Lotte, R. G., de Almeida, C. M., Siani, S. M., Körting, T. S., Fonseca, L. G. and da Silva, L. T. (2015). Object-based image analysis for urban land cover classification in the city of campinas-sp, brazil, *Urban Remote Sensing Event (JURSE), 2015 Joint*, IEEE, pp. 1–4.

Freeman, C., Kulić, D. and Basir, O. (2015). An evaluation of classifier-specific filter measure performance for feature selection, *Pattern Recognition* **48**(5): 1812–1826.

Funchal, J. P., Madsen, C. A. C. and Adamatti, D. F. (2015). Classificação automática de dados para descoberta de conhecimento: um estudo de caso para classificação de risco na área da saúde, *Revista Brasileira de Computação Aplicada* **7**(2): 41–51.

Ismail, S. A., Matin, A. F. A., Mantoro, T. et al. (2012). A comparison study of classifier algorithms for mobile-phone's accelerometer based activity recognition, *Procedia Engineering* **41**: 224–229.

Kamadi, V. V., Allam, A. R., Thummala, S. M. et al. (2016). A computational intelligence technique for the effective diagnosis of diabetic patients using principal component analysis (pca) and modified fuzzy sliq decision tree approach, *Applied Soft Computing* **49**: 137–145.

Kandhasamy, J. P. and Balamurali, S. (2015). Performance analysis of classifier models to predict diabetes mellitus, *Procedia Computer Science* **47**: 45–51.

Larose, D. T. (2014). *Discovering knowledge in data: an introduction to data mining*, John Wiley & Sons.

Larose, D. T. and Larose, C. D. (2015). *Data mining and predictive analytics*, John Wiley & Sons.

Last, M., Tosas, O., Cassarino, T. G., Kozlakidis, Z. and Edgeworth, J. (2016). Evolving classification of intensive care patients from event data, *Artificial intelligence in medicine* **69**: 22–32.

Lavanya, D. and Rani, K. U. (2011). Performance evaluation of decision tree classifiers on medical datasets, *International Journal of Computer Applications* **26**(4).

Li, H., Li, C.-J., Wu, X.-J. and Sun, J. (2014). Statistics-based wrapper for feature selection: An implementation on financial distress identification with support vector machine, *Applied Soft Computing* **19**: 57–67.

Lin, S.-W. and Chen, S.-C. (2012). Parameter determination and feature selection for c4. 5 algorithm using scatter search approach, *Soft Computing* **16**(1): 63–75.

Maciel, T. V., da Rosa Seus, V., dos Santos Machado, K. and Borges, E. N. (2015). Mineração de dados em triagem de risco de saúde, *Revista Brasileira de Computação Aplicada* **7**(2): 26–40.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic, *Biochemia medica: Biochemia medica* **22**(3): 276–282.

Nong, Y. (2014). *Data mining: theories, algorithms, and examples*, CRC press.

Quinlan, J. R. (1993). *C4.5:Programing for machine learning*, Morgan Kauffmann.

Ramya, M., Lokesh, V., Manjunath, T. and Hegadi, R. S. (2015). A predictive model construction for mulberry crop productivity, *Procedia Computer Science* **45**: 156–165.

Salami, H. O., Ibrahim, R. S. and Yahaya, M. O. (2016). Detecting anomalies in students' results using decision trees, *International Journal of Modern Education and Computer Science* **8**(7): 31.

Sharma, A. and Sahay, S. K. (2016). An effective approach for classification of advanced malware with high accuracy, *arXiv preprint arXiv:1606.06897* .

Sousa, A. L., Leal, A. B., Martins, R. F. and de Sá, C. C. (2011). Proposta de um sistema de apoio à tomada de decisão para o monitoramento remoto de centrais de alarme patrimoniais, *Revista Brasileira de Computação Aplicada* **3**(2): 17–29.

Turban, E., Sharda, R. and Delen, D. (2011). *Decision support and business intelligence systems*, Pearson Education India.

Venkatadri, M. and Lokanatha, C. (2010). A comparative study on decision tree classification algorithms in data mining, *International Journal of Computer Applications Engineering, Technology and Sciences (Ij-Ca-Ets) Issn: 0974-3596* **2**(2): 24.

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