



ISSN: 2230-9926

Available online at <http://www.journalijdr.com>

IJDR

International Journal of Development Research

Vol. 10, Issue, 05, pp. 35732-35735, May, 2020

<https://doi.org/10.37118/ijdr.18743.05.2020>



RESEARCH ARTICLE

OPEN ACCESS

IDENTIFICAÇÃO DE ARRITMIAS CARDÍACA PELO ELETROCARDIOGRAMA UTILIZANDO JRIP- RIPPER

*¹Walnizia Kessia Batista Olegário, ²Adriana Nascimento Gomes, ³Bárbara Meira de Oliveira, ⁴Kleyber Dantas Torres de Araujo, ⁵Dionarte Dantas de Araujo and ⁶Ronei Marcos de Moraes

¹Enfermeira. Mestre em Modelos de Decisão e Saúde pela Universidade Federal da Paraíba - UFPB, João Pessoa PB, Brasil; ²Fisioterapeuta. Doutoranda em Modelos de Decisão e Saúde pela Universidade Federal da Paraíba - UFPB, João Pessoa PB, Brasil; ³Gestora Ambiental. Mestre em Modelos de Decisão e Saúde pela Universidade Federal da Paraíba - UFPB, João Pessoa PB, Brasil; ⁴Analista de Sistemas. Doutorando em Modelos de Decisão e Saúde pela Universidade Federal da Paraíba - UFPB, João Pessoa PB, Brasil; ⁵Analista de Sistema, Mestrando em Informática pela Universidade Federal da Paraíba - UFPB, João Pessoa PB, Brasil; ⁶Estatístico. Doutor em Computação Aplicada pelo Instituto Nacional de Pesquisas Espaciais - UFPB, João Pessoa PB, Brasil

ARTICLE INFO

Article History:

Received 28th February, 2020

Received in revised form

03rd March, 2020

Accepted 16th April, 2020

Published online 25th May, 2020

Key Words:

Eletrocardiografia; Mineração de Dados; Informática.

*Corresponding author: Sáva Denise Silva Carlotto Herrera

ABSTRACT

Objetivo: O coração é um órgão vital para os seres humanos que pode ser acometido por várias disfunções, sendo o eletrocardiograma um exame fácil para detecção de alterações cardíacas. Este estudo teve como objetivo determinar a eficiência do algoritmo JRip, na classificação dos tipos de arritmia através de exames eletrocardiogramas gravados (ECG). **Metodologia:** Foram realizados exames de ECG em 452 sujeitos que alimentou o banco com 279 atributos e 16 variáveis classificatórias dos tipos de arritmia. Para analisar o banco de dados, foi utilizado um sistema de mineração WEKA com um modelo baseado em regras JRip. **Discussão:** O algoritmo JRip gerou 12 regras que indicaram os tipos de arritmias em 8 das 16 classes de arritmias. **Conclusão:** Observamos que o modelo JRIP não foi adequado e eficiente na classificação dos resultados

Copyright © 2020, Walnizia Kessia Batista Olegário et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Walnizia Kessia Batista Olegário, Adriana Nascimento Gomes, Bárbara Meira de Oliveira, et al. "Identificação de arritmias cardíaca pelo eletrocardiograma utilizando jrrip- Ripper", International Journal of Development Research, 10, (05), 35732-35735.

INTRODUCTION

As arritmias ou ritmos anormais do coração são distúrbios cardíacos comuns e podem causar sérios riscos à vida das pessoas, sendo uma das principais causas de óbito. Alguns tipos de arritmias podem ser identificadas a partir do exame do eletrocardiograma (ECG) que avalia o ritmo cardíaco e registra os fenômenos elétricos que a contração do músculo cardíaco gera e são captados pelo eletrocardiógrafo que analisa o formato das ondas características P, Q, R, S e T, e o intervalo entre batimentos consecutivos (Petrikicz, 2017). Para realizar maiores avanços em pesquisas na área com a utilização de banco de dados, os quais eram entendidos como um desafio a ser enfrentado, pois são geradores de inúmeras informações que necessitam de coleta, armazenamento, gerenciamento e

análise desses dados, para obter-se maiores progressos no ramo da medicina, a informática surge com os sistemas artificiais que facilitam a tarefa que seriam impossíveis para os seres humanos. Descoberta de Conhecimento em Bases de Dados (Knowledge Discovery in Databases - KDD é sinônimo de "Mineração de Dados" no contexto amplo. A mineração de dados é a aplicação de algoritmos específicos para extrair padrões de dados. Segue abaixo o esquema da processo básico da mineralização dos dados:

Os mineradores de dados tentam adotar métodos inovadores para enfrentar o problema causado pelo tamanho elevado dos banco de dados, aplicando métodos de seleção de recursos. Mineração de dados é o processo de descoberta de conhecimento em bases de dados que utiliza ferramentas

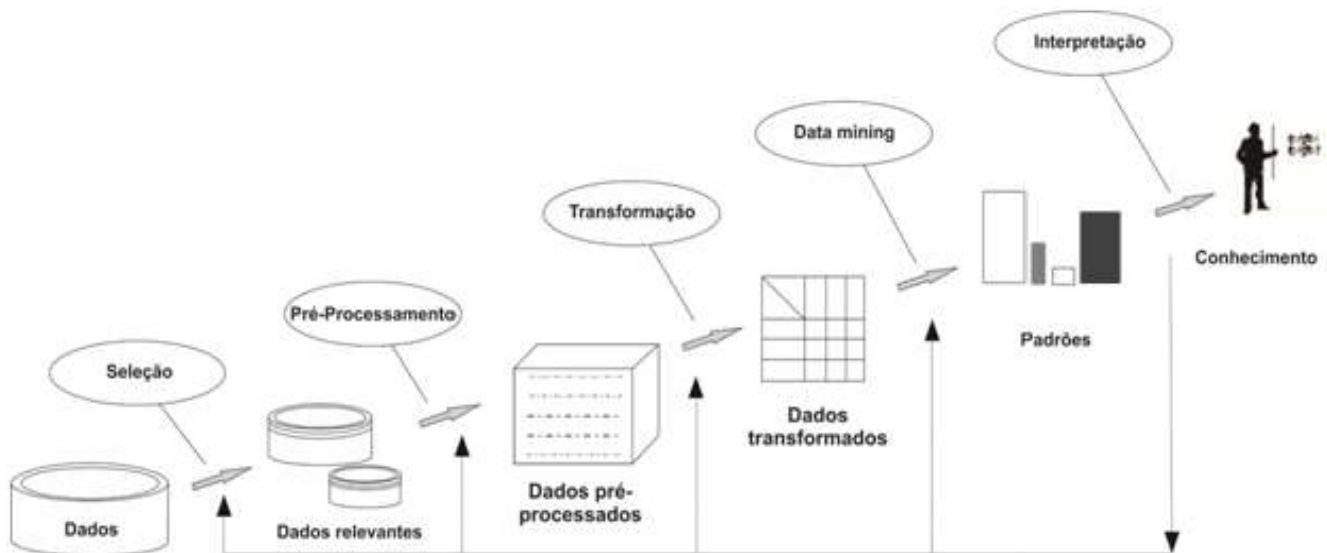


Figura 1. Etapas do processo de KDD (Fayyad et al., 1996) (adaptado)

computacionais para que possa extrair informações úteis, em um conjunto de dados biológicos, descritas na forma de padrões, a partir dos bancos de dados coletados e armazenados por instituições, serviços ou organizações (Edeki, 2012 and Correia, 2011). O sistema de mineração utilizado neste estudo para aplicação das técnicas de mineração de dados foi o WEKA (Waikato Environment for Knowledge Analysis). Esse sistema possui vários algoritmos de classificação, onde cada um deles possui padrões pré definidos. Alguns autores testaram métodos utilizando os algoritmos de classificação mais conhecidos (Naïve Bayes, Logística, Multicamada Perceptron, Melhor Primeira Árvore de Decisão e JRIP) para avaliar o banco de dados referente ao câncer de pulmão (Nasiripasa, 2013) com o objetivo de melhorar o desempenho do grupo de algoritmos de classificação no banco de dados que possuía muitas informações (Nasiripasa, 2013). Dentro dos vários algoritmos que são utilizados para tomadas de decisão utilizando banco de dados extenso, utilizaremos o modelo baseado em regras proposicionais ou condicionais, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), ou seja este algoritmo realiza a poda incremental repetida para produzir redução de erro, também conhecido como JRip (indução baseada em regras), implementação em JAVA proposto por William W. Cohen (Witten, 2005; Ferreira, 2015 and Vera, 2015).

É uma estrutura para a quantificação relativa livre de etiqueta e combina uma série de algoritmos para pré-processamento, quantificação de analito, alinhamento de tempo de retenção e agrupamento de analitos em corridas. É também a primeira estrutura de *software* para implementar a normalização de intensidade baseada em proximidade. Produz listas de sinais de analito (ou regras) com suas intensidades não normalizadas e normalizadas que podem servir como entrada para métodos de espectrometria de massa estatística, que pode detectar diferenças quantitativas entre amostras biológicas. Sendo assim, o RIPPER tornou-se uma ferramenta interna valiosa para um novo desenvolvimento de métodos quantitativos baseados em massa estatística muito utilizado na bioinformática (Susan, 2016 and University of Waikato, 2004). O método Ripper pode ser utilizado por várias áreas, ciência, economia, enologia, biologia, saúde, e outros. Alguns dos estudos interessantes foram: a detecção da quantidade de dióxido de enxofre na produção de vinhos, definição de

concessão de empréstimo bancário para pessoas jurídicas (Steiner, 2007), diagnóstico de pessoas com a doença de Alzheimer usando ao empregar o teste neuropsicológico, distinguir casos de não-casos de asma em crianças (Afzal, 2013) e análise educacional (Vera, 2015; Susan, 2016; University of Waikato, 2004; Steiner, 2007; Shree, 2016; Afzal, 2013 and Oliveira Junior, 2015). O modelo é avaliado usando o classificador baseado em regras JRip, que implementa uma aprendizagem de regra proposicional, usa as regras representadas por IF - THEN (SE - ENTÃO). Uma regra IF - THEN é uma expressão do formulário, condição IF, então, conclusão. As regras são uma boa maneira de representar informação ou pedaços de conhecimento⁽¹⁴⁾. O algoritmo basicamente divide-se em duas fases: a primeira gera um conjunto de regras para a comparação e a segunda otimiza o conjunto de regras iniciais para diminuir erros e tornar o processo mais seletivo, sendo esses passos repetidos inúmeras vezes no sistema WEKA (Oliveira Junior, 2015). Este estudo teve como objetivo determinar o tipo de arritmia de acordo com as gravações de ECG. Para isso utilizou-se o método da Lógica Clássica (JRip) condicional baseada em regras para ajudar na tomada de decisão quanto ao tipo de arritmia associada aos parâmetros do ECG.

MÉTODOS

Trata-se de um estudo do tipo experimental exploratório, transversal de abordagem quantitativa, realizado a partir da realização de experimentos e simulações de parametrização do algoritmo Jrip. Este estudo utilizou um banco de dados disponível em repositórios internacionais (<http://repository.seasr.org/Datasets/UCI/arff/>), sendo utilizado o banco arrhythmia em formato arff. O banco de dados foi composto por 452 pacientes, sendo 203 do sexo masculino e 249 do sexo feminino, formado por 279 atributos, mais 1 atributo de classificação. Do total de 279 atributos, 206 são do tipo lineares e 73 atributos nominais. O software utilizado para analisar este banco de dados foi o WEKA versão 3.8, desenvolvido pela Universidade de Waikato da Nova Zelândia, Java, distribuído sob os termos *General Public License* (GNU)⁽⁹⁾, sendo utilizado o JRip como modelo para tomar decisão sobre o tipo de arritmia. A seguir descreve-se a aplicação das etapas do processo de descoberta do conhecimento.

| Atributos | |
|-----------|---------------------------|
| 1 | QTinterval |
| 2 | Heartrate |
| 3 | ChDII_DD_RRwaveExists |
| 4 | ChV1_intrinsicReflections |
| 5 | ChV2_RPwaveExists |
| 6 | ChV3_Swave |
| 7 | ChDII_QwaveAmp |
| 8 | ChV1_RPwaveAmp |
| 9 | ChV3_RwaveAmp |
| 10 | ChV4_TwaveAmp |
| 11 | ChV5_TwaveAmp |
| 12 | chV6_RPwaveAmp |
| 13 | Class |

Figura 1. Tabela dos 13 atributos

| Classes | Número de indivíduos | |
|---------|---------------------------------------|-----|
| 1 | Normal | 245 |
| 2 | Isquemia de artéria coronária | 44 |
| 3 | Infarto crônico de parede anterior | 15 |
| 4 | Infarto crônico de parede inferior | 15 |
| 5 | Taquicardia sinusal | 13 |
| 6 | Bradicardia sinusal | 25 |
| 7 | Contração ventricular prematura | 3 |
| 8 | Contração supraventricular prematura | 2 |
| 9 | Bloqueio de ramo esquerdo | 9 |
| 10 | Bloqueio de ramo direito | 50 |
| 11 | Bloqueio atrioventricular de grau I | 0 |
| 12 | Bloqueio atrioventricular de grau II | 0 |
| 13 | Bloqueio atrioventricular de grau III | 0 |
| 14 | Hipertrofia de ventrículo esquerdo | 4 |
| 15 | Fibrilação atrial ou flutter | 5 |
| 16 | Outros | 22 |

Figura 2. Tipos de Arritmias

No WEKA estão incluídos quatro avaliadores de desempenho de aprendizagem, porém para este trabalho, somente foram utilizados o Conjunto de treinamento (*Use Training Set*), onde o classificador divide o conjunto de dados em dados de treinamento e teste. E a validação cruzada (*Cross Validation*), onde o WEKA desenvolve “n” modelos, encontra o desempenho médio desses modelos e exibe os resultados. Os modelos restantes são excluídos. A classificação é o segundo passo para saber como os dados estão sendo classificados. A avaliação é geralmente descrita pela precisão. As informações de execução também são exibidas, para uma rápida inspeção de quão bom funciona o classificador. O conjunto de dados é pré-processado e os resultados são comparados com diferentes avaliadores. A classificação é aplicada aos dados pré-processados antes e depois de selecionar os recursos e os resultados são comparados (Shree, 2016).

Preparação dos Dados: Após os procedimentos supracitados, foi realizado no WEKA uma filtragem (“limpeza”) dos dados, o que gerou uma redução do número dos atributos que passaram de 279 para 13 atributos (Figura 1). Porém, dentro do atributo de classes, onde foi identificado na tabela 1 que três (03) das 16 classificações de arritmia (Figura 2), não possuíam valores referenciados. E o programa não permitiu que fossem excluídos estes atributos para realizar a análise. Após seleção e limpeza os dados foram transformados para se enquadrar ao formato do modelo escolhido. As classes de arritmias foram previamente divididas em 16, sendo estas: normal, isquemia de artéria coronária, infarto crônico de parede anterior, infarto crônico de parede inferior, taquicardia sinusal, bradicardia sinusal, contração ventricular prematura, contração supraventricular prematura, bloqueio de ramo esquerdo, bloqueio de ramo direito, bloqueio atrioventricular de grau I, bloqueio atrioventricular de grau II, bloqueio

atrioventricular de grau III, hipertrofia de ventrículo esquerdo, fibrilação atrial ou flutter e outros. Foram observados atributos que atingiram valor 0 na modelagem, apresentando valores nulos na matriz de confusão, sendo assim retirados, tais atributos totalizaram um valor de 139 observações. O modelo foi treinado 5 vezes com os atributos que restaram, totalizando 142 valores. Observou-se que o modelo não apresentou diferença no valor do Kappa, antes e após a retirada dos atributos. A figura 2 apresenta os tipos de arritmia apresentados no estudo e a quantidade de sujeitos relacionados a cada um dos tipos de arritmia.

RESULTADOS E DISCUSSÃO

Foram utilizados os parâmetros padronizados do método JRip para o banco de dados, e recomendados para o WEKA (retirados do próprio manual do WEKA): Número de dobras (Folds= 3); Pesos mínimos das instâncias dentro de uma divisão (N= 2.0); Número de execuções otimizadas (O = 2); Semente de randomização (S=1), mas devido ao seu resultado não satisfatório para o método *crossvalidation*, foram feitos agrupamentos de 3 até 6 *folds*, tanto no modo *use training set* (1) como *nocross-validation*(2), com 66% da amostra, com intenção de obter-se os melhores valores na quantidade de regras geradas, no índice do Kappa e na matriz de confusão.

Modo Cross Validation

| Folds | Erro Mín. | Optimizations | Seed | Kappa |
|-------|-----------|---------------|------|--------|
| 3 | 2.0 | 2 | 1 | 0,5736 |
| 4 | 2.0 | 2 | 1 | 0,5966 |
| 5 | 2.0 | 2 | 1 | 0,5613 |
| 6 | 2.0 | 2 | 1 | 0,5939 |

Modo Use training set

| Folds | Erro Mín. | Optimizations | Seed | Kappa |
|-------|-----------|---------------|------|---------|
| 3 | 2.0 | 2 | 1 | 0,69999 |
| 4 | 2.0 | 2 | 1 | 0,6702 |
| 5 | 2.0 | 2 | 1 | 0,6647 |
| 6 | 2.0 | 2 | 1 | 0,6998 |

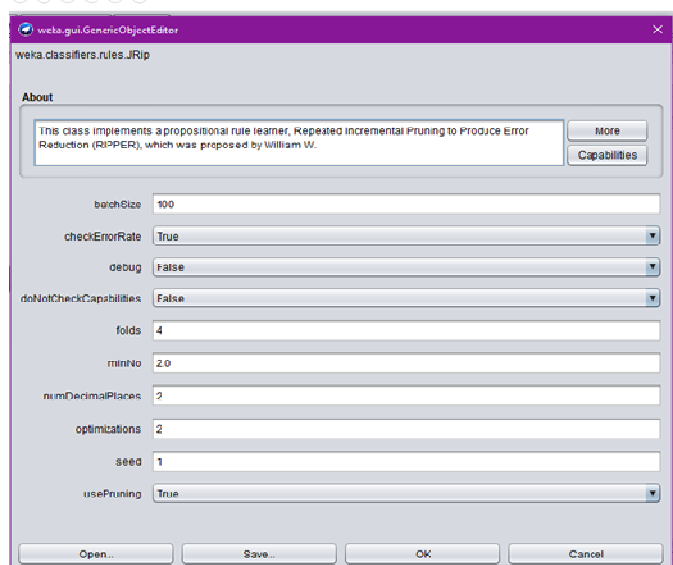


Figura 3. Parâmetros utilizados com melhor índice Kappa

Identificou-se que as modelagens que apresentaram melhor resultado, evidenciados pelo maior valor de Kappa foram, para o modo 1, os que foram rodados com o parâmetro padrão: Folds 3, Erro mínimo: 2.0, Optimizations 2, Seed 1. Já para o modo 2, o melhor resultado foi atingido com os seguintes parâmetros: Folds 4, Seed 1, Optimizer 2, Erro mínimo 2.0 (Figura 3).

| Regras | | | Resultados | Arritmias | Acertos | Erros |
|--------|---|-------|------------|------------------------------------|---------|-------|
| 1 | SE Onda RP existente na derivação V2 =1E FC >= 82 | ENTÃO | Classe 14 | Hipertrofia de ventrículo esquerdo | 0 | 4 |
| 2 | SE onda S na derivação V3 >= 84 | ENTÃO | Classe 9 | Bloqueio de ramo esquerdo | 9 | 0 |
| 3 | SE FC >=98 E intervalo QT <= 287 | ENTÃO | Classe 5 | Taquicardia Sinusal | 5 | 8 |
| 4 | SE FC >= 96 E amplitude da onda T na derivação V5 >= 1.6 | ENTÃO | Classe 5 | Taquicardia Sinusal | 5 | 8 |
| 5 | SE FC >= 103 E intervalo QT >=328 | ENTÃO | Classe 5 | Taquicardia Sinusal | 5 | 8 |
| 6 | SE amplitude da onda R na derivação V3 <=0 E reflexos intrínsecos na derivação V1 <=8 | ENTÃO | Classe 3 | Infarto crônico de parede anterior | 11 | 4 |
| 7 | SE amplitude da onda Q na derivação DII <= -1.4 | ENTÃO | Classe 4 | Infarto crônico de parede inferior | 9 | 6 |
| 8 | SE FC <= 58 | ENTÃO | Classe 6 | Bradycardia Sinusal | 22 | 3 |
| 9 | SE amplitude da onda T na derivação V5 <= 0.3 | ENTÃO | Classe 2 | Isquemia de artéria coronária | 27 | 17 |
| 10 | SE amplitude da onda RP na derivação na derivação V1 >= 0.7 | ENTÃO | Classe10 | Bloqueio de ramo direito | 30 | 20 |
| 11 | Reflexos Intrínsecos na derivação V1 >=40 | ENTÃO | Classe 10 | Bloqueio de ramo direito | 30 | 20 |
| 12 | | ENTÃO | Classe 1 | Normal | 223 | 23 |

Figura 4. Conjunto de regras geradas com o classificador JRip no WEKA e acertos e erros para as classes supracitadas

O classificador utilizado como parâmetro para atingir o objetivo deste estudo, e que melhor apresentou resultado, foi o *cross-validation k-fold* (validação cruzada com k=4 número de grupos), por apresentar resultado mais fidedigno, no qual o programa fez com que os dados fossem divididos em um conjunto de 4 subclasses (grupos), que foram combinados entre si sendo um deles utilizado para validação e o restante usado para treinamento, isso continuou até que todos os 4 grupos tivessem sido comparados entre si, tanto para treinamento como para validação, resultando em 336 instâncias classificadas como corretas (74.3363%), 116 instâncias classificadas como incorretas (25.6637%), Kappa 0.5966 e média do erro absoluto 0.0493, considerados não aceitáveis.

O modelo apresentou também as regras que compuseram a lógica clássica, permitindo dessa forma a divisão da amostra nas 16 classificações de arritmias presentes no estudo (Figura 2). Para a modelagem foi realizado uma filtragem dentro do *software* WEKA, que forneceu 13 atributos para realizar a processamento. Com relação aos 16 atributos de classe, foi observado que 3 classes apresentaram valores nulos, inclusive na matriz de confusão, classes 11, 12 e 13 (bloqueio atrioventricular de grau I, bloqueio atrioventricular de grau II e bloqueio atrioventricular de grau III), respectivamente. Nesse experimento o algoritmo JRipper gerou como resultado um conjunto de doze regras, que são visualizadas abaixo (Figura 5). O modelo treinado apresentou uma matriz de confusão, onde pudemos observar que três Classes (11,12,13), previamente descritos no banco de dados analisados, possuem informação nula. Após os dados serem processados observou-se que outras 5 classes com acertos zero, evidenciando dessa maneira que o modelo não se adequou ao banco de dados, evitando que o valor Kappa fosse maior que o encontrado no estudo.

Conclusão

Diante do exposto observou-se que com o modelo JRIP ao ser utilizado para analisar os arritmias cardíacas, através do ECG não foi muito eficaz. Porém, é importante considerar a realização de uma melhor limpeza dos atributos com um melhor processamento prévio dos dados, eliminando os valores que não possuíam observações. Como poderia ter sido feito nos dados referentes a variável classificatória dos tipos de arritmias. Foi observado ainda que em diversas literaturas encontradas, foi possível identificar a utilização do JRIP com folds divergente do que se tem disponível como padronizado na documentação do WEKA. isto permite uma melhor adequação do algoritmo ao banco de dados.

REFERÊNCIAS

- Afzal Z, et al. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. 2013; 22(8): 826–833.
- Correia C F, Pimentel E P. Mineração de Dados na Formação de Turmas para a Recuperação Paralela na Educação Básica. In: Anais do XXII SBIE - XVII WIE;2011. Nov. 21 – 25; Aracaju, Sergipe. [internet] [citado 2017 ago 24]. Disponível em: <http://www.br-ie.org/pub/index.php/sbie/article/view/1824/1586>.
- Edeki C, Pandya S. Comparison of Data Mining Techniques used to Predict Cancer Survivability. International Journal of Computer Science and Information Security. 2012;10(6): 1-6.
- Ferreira E A L. Mineração de Dados Aplicada à dados Médicos. [dissertação] Belo Horizonte: Universidade Federal de Minas Gerais. 2015.
- Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. 3.ed. San Francisco: Morgan Kaufmann Publishers Inc. CA, USA, 2011.
- Nasiripasa M, Bidgoli A, Varae T. Improving Performance of a Group of Classification Algorithms Using Resampling and Feature Selection. World of Computer Science and Information Technology Journal. 2013;(4): 70-76.
- Oliveira Junior J G. Pattern Identification for Dropout Analysis in Undergraduate Courses using Educational Data Mining. [dissertação] Curitiba: Universidade Tecnológica Federal do Paraná. 2015.
- Petrikicz D H, Oliveira B R, Abreu C C. Aplicação do algoritmo de seleção negativa na detecção de arritmias cardíacas. In: Anais do XXXIV Simpósio Brasileiro de Telecomunicações; 2016. ago. 30 – set. 02; Santarém, Pará. [internet] [citado 2017 ago 24]. Disponível em: <http://sbirt.org.br/sbirt2016/anais/IC02/1570274252.pdf>.
- Shree S R B, Sheshadri H S, Muralikrishna. Diagnosis of Alzheimer's Disease using Rule based Approach. Indian Journal of Science and Technology. 2016;9(13): 1-6.
- Steiner M T A, et al. Extração de Regras de Classificação a partir de Redes Neurais para o auxílio na Concessão de Crédito Bancário. Pesquisa Operacional. 2007;27(3): 407-426.
- Susan K R V, Higgins L A, Carlis J V, Griffin T J. RIPPER: a framework for MS1 only metabolomics and proteomics label-free relative quantification. Bioinformatics. 2016;32(13): 2035–2037.
- University of Waikato. Weka 3 – Machine Learning Software in Java. 2004 [citado 2017 ago 19]. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka>
- Vera C M. Predicción del Fracaso y el Abandono Escolar Mediante Técnicas de Minería de Datos. [tesis doctoral] Universidad de Córdoba. 2015.
- Witten H, Frank E. Data Mining: Practical Machine Learning Tools and Techniques. 2.ed. San Francisco: Morgan Kaufmann; 2005.