

Uma solução computacional baseada em inteligência artificial como ferramenta de apoio para o processo de tomada de decisão em ambientes médicos**A computational solution based on artificial intelligence as a support tool for the decision-making process in medical environments**

DOI:10.34117/bjdv6n7-430

Recebimento dos originais: 10/06/2020

Aceitação para publicação: 17/07/2020

Icaro Rodrigues Costa Vinciaqui

Graduando em Ciências da Computação

Instituição: Universidade Federal de Jataí

Endereço: BR 364, KM 195, nº 3800, Jataí - GO, Brasil

E-mail: icarorc@hotmai.com

Joslaine Cristina Jeske de Fretias

Doutora em Ciências da Computação pela Universidade Federal de Uberlândia

Instituição: Universidade Federal de Jataí

Endereço: BR 364, KM 195, nº 3800, Jataí - GO, Brasil

E-mail: joslaine@ufg.br

Laurence Rodrigues do Amaral

Doutor em Ciências da Computação pela Universidade Federal de São Carlos

Instituição: Universidade Federal de Uberlândia

Endereço: Rua Getúlio Vargas, nº 230, Centro, Patos de Minas - MG, Brasil

E-mail: laurence@ufu.br

Franciny Medeiros Barreto

Mestre em Ciências da Computação pela Universidade Federal de Uberlândia

Instituição: Universidade Federal de Jataí

Endereço: BR 364, KM 195, nº 3800, Jataí - GO, Brasil

E-mail: franciny@ufg.br

RESUMO

A introdução da tecnologia nos processos da rotina humana gera diariamente uma grande quantidade de dados de diversas fontes (lojas, supermercados, exames médicos, etc.). Técnicas de Mineração de dados aplicadas a esses tipos de dados se tornam particularmente interessantes para extrair informações úteis. Neste contexto, o presente trabalho tem como objetivo aplicar técnicas de Mineração de Dados e Aprendizado de Máquina em uma base de dados clínicos para extrair informações relevantes para auxiliar na definição de diagnósticos e prognósticos de pacientes. Portanto, este trabalho apresenta uma solução computacional que servirá como ferramenta de apoio no processo de tomada de decisão em ambientes biológicos e médicos. Para ilustrar a abordagem dessa pesquisa foi escolhida uma base de dados que consiste de dados clínicos relativos a procedimentos de transplante renal.

Palavras-chave: Mineração de Dados, Aprendizado de Máquina, Medicina de Precisão

ABSTRACT

The introduction of technology in human routine processes generates a large amount of data from different sources (stores, supermarkets, medical exams, etc.) on a daily basis. Data mining techniques applied to these types of data become particularly interesting for extracting useful information. In this context, the present work aims to apply Data Mining and Machine Learning techniques in a clinical database to extract relevant information to assist in the definition of diagnoses and prognosis of patients. Therefore, this work presents a computational solution that will serve as a support tool in the decision-making process in biological and medical environments. To illustrate the approach of this research, a database was chosen that consists of clinical data related to kidney transplant procedures.

Keywords: Data Mining, Machine Learning, Precision Medicine

1 INTRODUÇÃO

O surgimento de processos metodológicos que incluíram a Mineração de Dados (MD) seguiram o intuito de não somente extrair informações escondidas nas Bases de Dados mas também de fazer a máquina adquirir conhecimento e se tornar capaz de inferir conclusões a respeito dos dados. Uma dessas áreas ganhou o nome de Descoberta de Conhecimento em Bases de Dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Diante do fato de que a tecnologia está cada vez mais inserida nos processos da rotina humana, grandes quantidades de dados são gerados todos os dias (SOWMYA; SUNEETHA, 2017). Esses dados formam Bases de Dados que aplicadas a um processo de descoberta de informações, podem trazer informações que processadas se tornam conhecimento para a máquina.

Com um futuro promissor, os processos de Aprendizagem de Máquina que utilizam a Mineração de Dados como subprocesso foram aplicados em diversas áreas (SOWMYA; SUNEETHA, 2017) tornando possível a predição do comportamento na bolsa de valores, por exemplo, até a descoberta das relações entre a associação de genes ligados ao vírus HIV (ESPÍNDOLA et al., 2018).

Trabalhos que utilizaram dados biológicos como objeto de estudo descobriram informações importantes que podem ser decisoras quando auxiliam o profissional da saúde (GRECO et al., 2010). Segundo (PORTO, 2001), o poder da Mineração de Dados e do Aprendizado de Máquina podem contribuir de forma positiva fornecendo mais informações para o especialista, visto que um profissional da saúde precisa de adquirir o máximo de informações possíveis, na menor quantidade de tempo, para a realização correta de seus prognósticos.

Dentro deste contexto, a pesquisa descrita neste trabalho tem como objetivo desenvolver uma solução computacional que servirá como ferramenta de apoio à tomada de decisão em ambientes biológicos e médicos. Especificamente para esse trabalho a Base de Dados escolhida para testes

consiste de dados clínicos relativos a procedimentos de transplante renal. Assim, a solução computacional desenvolvida servirá como ferramenta de apoio à tomada de decisão classificando as amostras em relação ao risco imunológico, mais precisamente a respeito de sobrevida do enxerto e incidência de rejeição em um transplante renal. Auxiliando então no prognóstico de transplantes renais, fornecendo informações sobre o risco de rejeição.

2 BASE TEÓRICA

2.1 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

A Descoberta de Conhecimento em Bases de Dados, mais conhecido pelo termo em inglês Knowledge Discovery in Data Bases ou KDD, é descrito como todo o processo de extração de informação útil de uma base de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). Tal processo envolve a Mineração de dados e o Aprendizado de Máquina como passos para extração do conhecimento.

Para se conseguir extrair informação de uma base de dados, é necessário executar passos que vão desde o pré-processamento como a remoção de ruídos por exemplo, até o pós-processamento que é onde acontece a visualização e geração de informações.

2.2 MINERAÇÃO DE DADOS

A Mineração de Dados é o processo de descobrimento automático de informações, no qual se faz o uso de regras de inferências e descobrimento de padrões (TAN, 2018). Existem várias técnicas que a Mineração de Dados utiliza, como exemplo a clusterização, que tem como base o agrupamento de padrões com características em comum. Uma outra técnica é o descobrimento de padrões, no qual a ideia é dizer qual o melhor padrão que infere na conclusão, árvores de decisão, entre outras.

Visto que a quantidade de Bases de Dados geradas no dia-a-dia é vasta existem processos que antecedem a aplicação da Mineração de Dados. O primeiro processo é chamado de pré-processamento dos dados, onde o objetivo é transformar os dados brutos em dados apropriados para análises subsequentes (TAN, 2018). Após a passagem da Base de Dados pelo pré-processamento, o mesmo é submetido a ferramenta mineradora. A ferramenta mineradora consiste em agrupar diferentes técnicas da Mineração de Dados para prover um melhor ambiente de trabalho para o especialista que está utilizando (HOLMES; DONKIN; WITTEN, 1994).

2.3 APRENDIZADO DE MÁQUINA

O Aprendizado de Máquina é uma tecnologia voltada para o desenvolvimento de técnicas de aprendizado automático, que adquire conhecimento através de treinamento (MONARD;

BARANAUSKAS, 2003). As técnicas de aprendizado, são técnicas que inferem, dependendo das experiências bem sucedidas anteriormente. Tais experiências são baseadas em inserir Base de Dados, onde a máquina consegue aprender e fazer suposições a respeito dos dados.

2.4 MEDICINA DE PRECISÃO

A Medicina de Precisão utiliza técnicas da Mineração de Dados e Aprendizado de Máquina aplicadas aos dados clínicos, que podem ser desde dados demográficos, por exemplo, até dados genômicos, com o objetivo de traçar prognósticos e diagnósticos precisos para cada paciente (MIRNEZAMI; NICHOLSON; DARZI, 2012) auxiliando em um tratamento mais eficaz.

Tal área começou a ganhar espaço pelo fato de que a tarefa de inferir um prognóstico correto demanda tempo e "bagagem clínica", pois o prognóstico médico, de acordo com (PORTO, 2001), consiste em saber previamente a provável evolução ou o atual estado de um quadro clínico, como também seu tratamento e os cuidados a serem tomados. Para isso, os médicos precisam ter uma vasta bagagem para possibilitar uma maior assertividade no prognóstico, acarretando também a demanda de tempo e raciocínio diferente para cada caso clínico.

3 OBJETIVOS

A pesquisa descrita neste trabalho tem como objetivo desenvolver uma solução computacional que seja uma ferramenta de apoio à tomada de decisão em ambientes biológicos e médicos. Especificamente objetiva-se:

- selecionar uma Base de Dados;
- executar a fase de pré-processamento com o objetivo de uma gerar uma Base de Dados mais consistente e precisa;
- implementar um algoritmo de cruzamento dos dados baseado nos princípios de um para um, dois para dois e três para três visando reduzir variáveis/atributos;
- aplicar os resultados gerados no item anterior na ferramenta mineradora WEKA;
- realizar uma análise quantitativa a respeito das classificações realizadas.

4 METODOLOGIA

Como passo inicial, é necessário obter a Base de Dados clínicos que servirá para a aplicação. Em seguida é preciso realizar o pré-processamento na Base de Dados visando a limpeza e remoção de ruídos. Após a limpeza da Base de dados, é preciso desenvolver um algoritmo que efetuará o cruzamento dos atributos entre si. O cruzamento entre os dados gerará como resultado várias bases

que, a partir desse ponto serão aplicados em uma ferramenta mineradora seguindo algum método de Aprendizado de Máquina.

O intuito de utilizar uma ferramenta mineradora é conseguir identificar na Base de Dados os pacientes com risco imunológico em quatro categorias: grupo 1 (baixo risco), grupo 2 (médio risco 1), grupo 3 (médio risco 2) e grupo 4 (alto risco). Após a classificação será realizada uma análise quantitativa das taxas de assertividades do algoritmo, bem como a sobrevida do enxerto, do paciente e da rejeição dos pacientes que foram submetidos a transplante renal cujo os dados estão contidos na Base de Dados.

5 RESULTADOS

A Base de Dados selecionada para esta pesquisa foi disponibilizada pela iniciativa privada para ser utilizada neste projeto. A base em questão é composta com informações de 404 pacientes que foram submetidos ao procedimento de transplante renal. O processo de pré-processamento consistiu em eliminar linhas e colunas faltosas na Base de Dados que poderiam afetar o desempenho da classificação quando aplicadas na ferramenta mineradora.

O algoritmo para cruzamentos dados foi implementado usando a linguagem de programação Perl. Sabendo que o algoritmo receberá como entrada arquivos de texto, a linguagem foi escolhida devido a sua facilidade de manipulação de arquivos, possuindo ferramentas úteis que facilita a construção da solução computacional.

O algoritmo se baseia na afirmação de Porto (2001), onde o autor diz que a assertividade no prognóstico está ligada a quantidade de sintomas. Portanto, o algoritmo utilizará a Base de Dados original e se baseará em pré-selecionar os atributos de um a um, dois a dois e três a três que serão cruzados entre si.

Após o cruzamento desses dados, como resultado obteve-se novos conjuntos de dados que serviram como entrada para a ferramenta mineradora. A ferramenta escolhida foi o WEKA, por ser uma ferramenta completa que implementa todas as principais técnicas de Aprendizado de Máquina. O método de Aprendizado de Máquina selecionado foi o C4.5, que é uma árvore de decisão já implementada na ferramenta. Para o trabalho, o modelo de classificação baseado em árvore de decisão se mostrou o mais correto, pois além da tarefa que o presente trabalho propõe ser descritiva, o modelo de árvores de decisão é transparente, tornando possível a observação da estrutura da árvore e como cada decisão foi tomada.

6 CONSIDERAÇÕES FINAIS

A base de dados escolhida para o desenvolvimento dessa pesquisa consiste de dados clínicos relativos a procedimentos de transplante renal. Apesar de usar uma Base de Dados específica de transplante renal, a solução computacional desenvolvida poderá ser aplicada em outras bases de dados clínicos, utilizando o mesmo princípio de cruzamento entre os atributos buscando um prognóstico com um menor número de variáveis/atributos, servindo como ferramenta auxiliadora para o especialista da saúde avaliando características específicas de pacientes em tratamento.

REFERÊNCIAS

- ESPÍNDOLA, Milena S. et al. Epigenetic alterations are associated with monocyte immune dysfunctions in HIV-1 infection. **Scientific reports**, v. 8, n. 1, p. 5505, 2018.
- FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**, v. 39, n. 11, p. 27-34, 1996.
- GRECO, R. et al. Decisional trees in renal transplant follow-up. In: **Transplantation proceedings**. Elsevier, 2010. p. 1134-1136.
- MIRNEZAMI, R.; NICHOLSON, J.; DARZI, A. Preparing for precision medicine. **New England Journal of Medicine**, Mass Medical Soc, v. 366, n. 6, p. 489–491, 2012.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.
- PORTO, C. C. Semiologia médica. Em: **Semiologia médica**. Rio de Janeiro ,RJ: Guanabara Koogan, 2001.
- SOWMYA, R.; SUNEETHA, K. R. Data mining with big data. In: **2017 11th International Conference on Intelligent Systems and Control (ISCO)**. IEEE, 2017. p. 246-250.
- TAN, P.-N. Introduction to data mining. Chennai, India: **Pearson Education India**, 2018.