

# Mineração de dados para prever o comportamento futuro de doadores de sangue

Jairo Marciano Silva<sup>1</sup>, Ana Maria Monteiro<sup>1</sup>

<sup>1</sup>Centro Universitário Campo Limpo Paulista - UNIFACCAMP  
jmarciano1971@gmail.com, anammont.alu@gmail.com

**Resumo:** *Serviços de Hemoterapia, chamados comumente de Unidades Hemoterápicas, muitas vezes tem problemas para convocar doadores de sangue em épocas de crise ou em situações de emergência como a recente pandemia do COVID-19. Além disso, nesses serviços existe uma preocupação constante em manter os estoques de sangue em patamares seguros e aceitáveis. Este projeto de pesquisa aqui apresentado, busca verificar aplicabilidade de técnicas de mineração de dados para prever doadores que possuam um maior potencial para ser tornarem doadores regulares e assim poder guiar políticas e decisões a serem adotadas para convocar doadores em épocas de crise ou em situações de emergência.*

**Palavras-chave:** *Aprendizado de Máquina, Classificadores, Doação de Sangue.*

## 1. Introdução

A transfusão sanguínea é fundamental para o tratamento de pacientes em casos de emergências graves causadas, por exemplo, por acidentes. Também é um recurso terapêutico de extrema importância para pacientes com anemias falciformes, hemofilia, leucemia aguda, entre outras, que necessitam de transfusões regulares (DE MAGALHÃES *et al.*, 2016). Outro aspecto importante que vale a pena ressaltar, é a natureza perecível do sangue que possui um prazo de validade de aproximadamente de 42 dias a partir da data de coleta. O sangue coletado é frequentemente dividido em concentrado de hemácias, concentrado de plaquetas randômicas e plasma com prazo de validade de 42, 5 e 365 dias respectivamente.

Recentemente, o sangue e seus componentes foram adicionados à lista de medicamentos essenciais da Organização Mundial da Saúde (OMS). Por outro lado, sua disponibilidade é insatisfatória nos países de renda per capita baixa ou média, levando a mortalidade e morbidade que poderiam ser evitadas (ROBERTS *et al.*, 2019).

Segundo a OMS, em 2018, ocorreram aproximadamente 118,5 milhões de doações de sangue no mundo todo. Desse total, somente 60% foram efetuadas em países de baixa e média renda, sendo que estes concentram 84% da população mundial [OMS 2022]. No Brasil, considerando dados de 2017 do Ministério da Saúde, ocorreram aproximadamente 3.4 milhões de doações de sangue e foram realizadas 2.8 milhões de transfusões (MINISTÉRIO DA SAÚDE DO BRASIL, 2019).

A motivação principal para doação de sangue é devida, principalmente, a um comportamento pró-social, a uma ideologia sem vantagens óbvias para o doador além da

satisfação pessoal. Outros fatores que levam as pessoas a doar é já ter precisado de sangue para si ou para pessoas próximas, algumas pessoas só vão atrás de exames sorológicos e outras são motivadas devido a estarem envolvidas com grupos que participam de campanhas de doação (GILLESPIE; HILLYER, 2002).

## **2. Objetivos e etapas do projeto de pesquisa**

O presente projeto de pesquisa utiliza uma abordagem para extração, tratamento e análise de dados, para predizer doadores que possuem um maior potencial para se tornarem doadores regulares. Para alcançar esses objetivos, utilizaremos técnicas fundamentalmente da área de aprendizado de máquina, em geral, e mineração de dados, em particular. Entre outros benefícios, os resultados da pesquisa podem guiar políticas e decisões a serem adotadas por hemocentros para convocar doadores em épocas de crise ou em situações de emergência.

Após a aprovação do uso dos dados por parte do Comitê Científico da Fundação Pró-Sangue/Hemocentro de São Paulo, iniciou-se a Revisão Sistemática de Literatura (RSL) (KITCHENHAM et al., 2009; KITCHENHAM, 2004; MIAN et al., 2005) e em paralelo foi realizada a análise e limpeza dos dados.

Com a RSL finalizada, providenciamos a escrita de um artigo que foi submetido em março de 2022 para o XXII Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS), realizado no XV Encontro Unificado de Computação do Piauí. O artigo foi aceito e apresentado em junho de 2022.

Após a apresentação do artigo, o esforço concentrou-se no processo de escrita final do projeto de pesquisa e a realização de uma análise estatística dos dados de doações de sangue total efetuadas nos 6 postos de coleta da Fundação Pró-Sangue – Hemocentro de São Paulo (FPS) entre os anos de 2016 e 2020. Isto nos remete a um total de 323.276 doadores de sangue.

A etapa de preparação ou pré-processamento dos dados está sendo concluída. A remoção de ruídos/dados inconsistentes (limpeza) e a escolha dos dados referentes à análise (seleção ou redução) está em andamento. Como próximo passo, iniciaremos a etapa de transformação dos dados para adequação do formato de alguns atributos, segundo necessidades específicas dos algoritmos de mineração de dados que serão utilizados no decorrer da pesquisa.

Em uma análise preliminar, podemos considerar o problema de predição de doadores com maior propensão a se transformarem em doadores regulares, como um problema de classificação binário. Com essa premissa, definimos como próxima atividade revisar os métodos de aprendizado de máquina tais como os algoritmos de árvores de decisão ID3, C4.5, Naive Bayes (NB), Support Vector Machine (SVM) e k-Nearest Neighbor (kNN), para sua posterior aplicação na base de dados da FPS pré-processada. Após a aplicação dos algoritmos, os resultados devem ser analisados para embasar a escrita final da dissertação de mestrado bem com de um artigo para divulgação dos resultados obtidos.

### 3. Referências

BRASIL. MINISTÉRIO DA SAÚDE. **Dezesseis a cada mil brasileiros doam sangue**. Disponível em: <<https://www.saude.gov.br/noticias/agencia-saude/45520-dezesseis-a-cada-mil-brasileiros-fazem-doacao-de-sangue>>.

DE MAGALHÃES, T. A. et al. Prevalência de inaptidão sorológica dos doadores de sangue no hemocentro regional de Montes Claros, Minas Gerais. **Revista Online de Pesquisa**, v. 8, n. 3, p. 4864–4871, 2016.

GILLESPIE, T.; HILLYER, C. Blood donors and factors impacting the blood donation process. **Transfusion medicine reviews**, v. 16, p. 115–130, 2002.

KITCHENHAM, B. Procedures for Performing Systematic Reviews. **Keele, UK, Keele Univ.**, v. 33, 2004.

KITCHENHAM, B. et al. Systematic literature reviews in software engineering-A systematic literature review. **Information and Software Technology**, v. 51, p. 7–15, 2009.

MIAN, P. et al. A systematic review process to software engineering. v. 32, 2005. .

ROBERTS, N. et al. The global need and availability of blood products: a modelling s