

RELATÓRIO: MODELO PREDITIVO PARA DOADORES DE SANGUE

1 INTRODUÇÃO

As ferramentas de *machine learning* apresentam métodos de suma importância na análise, construção de informações e alternativas para meios de trabalho entre várias áreas da sociedade, tornando-se uma tendência global. Os métodos de aprendizado são bastante úteis na criação de modelos de respostas baseado em dados históricos, a partir disso este estudo se baseia no método preditivo da classificação como fonte de aprendizado e teste da ferramenta.

2 OBJETIVOS

Este estudo teve por objetivo aplicar o modelo preditivo de classificação para buscar respostas para determinar se doadores do Centro de Serviços de Transfusão de Sangue na cidade de Hsin-Chu, em Taiwan, doaram sangue em março de 2007 baseado em seu conjunto de dados históricos de doação.

3 MÉTODOS E RESULTADOS

Através da base de dados disponibilizadas na *UCI Machine Learning Repository* (ANEXOS) foi extraída a tabela no formato .data, os dados dessa tabela continha registros de 748 doadores de sangue aleatórios do Centro de Serviços de Transfusão de Sangue na cidade de Hsin-Chu, em Taiwan, cada doador apresenta dados numéricos separadas em colunas, R (Recência - meses desde a última doação), F (Frequência - número total de doações), M (Monetário - sangue total doado em cc), T (Tempo - meses desde a primeira doação) e uma variável binária que representa se doou sangue em março de 2007 (1 significa doou sangue; 0 significa não doou sangue).

O algoritmo se baseou na importação as bibliotecas ‘pandas’ e ‘numpy’ na *IDE Spyder* (Python versão 3.7), os dados dos doadores foram chamados para leitura criando um *DataFrame*, a partir disso se pode realizar o tratamento dos dados, utilizando ‘.loc’ para achar dados negativos, sendo inexistente. Assim, pelo ‘slice’ da tabela foi criada os dados dos previsores - utilizados no aprendizado e previsão das repostas - e a classe, as respostas binárias de 1: doou e 0: não doou.

Mesmo não apresentado dados ausentes utilizou-se um ‘SimpleImputer’ como cautela, preenchendo com a média do conjunto de valores referentes a cada coluna. Após esses passos, foi realizado o escalonamento dos previsores por um ‘StandardScaler’, para reduzir as distância entre os dados nas colunas auferindo valores mais precisos.

Para o aprendizado da máquina, os previsores e a classe foram ambos separados através do ‘train_test_split’ em _teste (25%) e _treinamento (75%). Com isso, o modelo escolhido dentro da classificação para a tomada de decisão foi o algoritmo *Naive Bayes*, que possibilita através da análise de dados históricos montar uma previsão probabilística dos dados. Nesse método de aprendizagem, o algoritmo ‘GaussianNB’ aplicou sobre o _treinamento entre a classe e previsores possibilitando o uso de um ‘.predict’ nos _teste de previsores, gerando a tabela de previsões de respostas do método naive_bayes, com isso pode-se comparar as respostas com as do _teste da classe, as respostas corretas.

Logo gerou-se a tabela de acertos e erros pela importação de ‘confusion_matrix, accuracy_score’, vista na Figura 1.

Figura 1. Matriz de erros e acertos de registros de doadores de sangue

	0	1
0	124	8
1	42	13

Fonte: Elaboração própria, 2019.

Os dados permitiram apurar uma precisão de 73,26% de acertos do algoritmo, lembrando que 0 representam não doaram e 1 representam doaram, o algoritmo assim errou com 8 marcados como doaram mas realmente não doaram, e com 42 erros marcados como não doaram, mas realmente doaram.

CONCLUSÕES

Conclui-se que a aplicação do algoritmo permitiu o aprendizado de máquina sobre a doação de sangue baseado nos dados históricos de 748 doadores aleatórios. Observou-se durante a implementação do estudo, que o uso de mais atributos correlacionados aos dados possibilita uma apuração mais fiel das previsões do método naive bayes, a retirada também dificulta o grau de precisão de acertos na aprendizagem, outras situações como melhor limpeza de dados, escalonamento, também ajudam na melhora das respostas e qualidade do algoritmo.