 <https://doi.org/10.56238/tecavanaborda-019>

Mário Márcio Rodrigues

Ciência da Computação, UNEMAT

E-mail: mariomarciorodrigues2@gmail.com

Daniela Cabral de Oliveira

Ciência da Computação, UNEMAT

E-mail: daniela.cabral@unemat.br

Max Robert Marinho

Ciência da Computação, UNEMAT

E-mail: max.marinho@unemat.br

Weder Nunes Ferreira Junior

Ciências Agrárias, IF GOIANO

E-mail: wedernunesiftm@gmail.com

Miélle Silva Pestana

Ciência da Computação, UNEMAT

E-mail: miellepestana@gmail.com

Sérgio Santos Silva Filho

Ciência da Computação, UNEMAT

E-mail: santos@unemat.br

RESUMO**1 INTRODUÇÃO**

A COVID-19, doença causada pelo coronavírus SARS-CoV-2, teve origem na Província de Wuhan, na China, com a identificação do primeiro caso no Brasil no final de fevereiro de 2020 (ZHANG *et al.*, 2020). Logo após, no mês de março de 2020, a Organização Mundial da Saúde (OMS) declarou a doença como pandemia, devido a sua presença na maior parte dos países no mundo.

Muito embora a doença tenha se espalhado rapidamente pelo mundo e, não existissem informações suficientes para conter a sua disseminação, os países, em todo o mundo, tiveram que desenvolver de saúde pública de forma rápida, na busca por evitar um cenário caótico com relação ao contágio pelo coronavírus, causador da COVID-19. Assim, procedimentos médicos não considerados urgentes, foram cancelados e o atendimento hospitalar passou a ser de urgência e emergência no cuidado dos pacientes acometidos pelo novo vírus (SILVA *et al.*, 2022).

Este trabalho tem o objetivo de utilizar o aprendizado de máquina supervisionado e não supervisionado para classificar os resultados dos exames de pacientes do hospital Albert Einstein durante um período no contexto da pandemia. Para a classificação dos dados foram utilizados os algoritmos de aprendizado de máquina supervisionado de uma rede neural perceptron de múltiplas camadas, Support Vector Machine (SVM) para diferentes kernels e árvore de decisão. Assim, verificou-se que os dados foram classificados com taxa de acurácia em 100%, 99.187%, 86.1789% e 84.9583% para todos os classificadores do aprendizado de máquina supervisionado. E com a taxa de acurácia em 87.05% para clusterização quanto aos resultados dos exames dos pacientes do hospital Albert Einstein durante um período no contexto da pandemia utilizando o aprendizado de máquina não supervisionado.

Palavras-chave: Pacientes, Hospital Albert Einstein, Aprendizado de Máquina Supervisionado, Aprendizado de Máquina não Supervisionado.

Neste sentido, faz-se necessário analisar o banco de dados do Hospital Albert Einstein com finalidade de diagnosticar, por meio dos parâmetros dos exames realizados no contexto da pandemia, se os pacientes estavam saudáveis ou não saudáveis utilizando o Aprendizado de Máquina (AM).

O AM é uma subárea da Inteligência Artificial, que tem como foco de desenvolvimento um modelo capaz de aprender por meio da experiência. Nesse sentido, o aprendizado ocorre a partir de algoritmos dedutivos, os quais com base em cálculos estatístico, permitem a extração de regras e padrões em locais, em que se encontrem um grande volume de dados (HAYKIN, 1999).

Na literatura existem alguns tipos de aprendizado de máquina, sendo eles: aprendizado de máquina supervisionado, não supervisionado e por reforço. Desse modo, tem-se que no presente estudo, foram testados algoritmos de aprendizado de máquina supervisionado, tais como: árvore de decisão, *Support Vector Machine* (SVM) para diferentes *kernels*, rede neural *perceptron* de múltiplas camadas e o algoritmo *K-means* para o aprendizado de máquina não supervisionado.

Atenta-se para o fato de que os resultados com base em técnicas inteligentes foram eficientes, confiáveis e robustos para o diagnóstico de pacientes durante o contexto da pandemia, o que justifica o uso do AM para o desenvolvimento desse estudo e a apresentação de informações coletadas junto ao banco de dados do Hospital Albert Einstein.

2 MATERIAL E MÉTODOS

O banco de dados utilizado para os testes numéricos foi obtido pelo Hospital Albert Einstein divulgados na internet. Esse foi composto por três banco de dados primários, sendo eles: desfechos, pacientes e exames. O banco desfechos era composto por oito variáveis e 89.938 linhas totalizando em 719.504 dados; o banco de pacientes era composto por sete variáveis e 14.674 linhas totalizando em 102.718 dados, enquanto o banco de exames era composto por nove variáveis e 1.048.576 linhas totalizando em 9.437.184 dados.

Diante desse contexto, foi realizado uma mineração dos dados eliminando àqueles redundantes e os *outliers*. Nesse sentido, o banco de dados secundário foi composto por seis variáveis, sendo cinco oriundas do banco de dados primário e uma tendo sido determinada das variáveis resultado e valor de referência, para assim, ter sido possível a determinação da variável diagnóstico.

Em seguida, os dados foram ajustados para o formato arff e então processados pela ferramenta Weka na versão 3.8.6. O banco de dados foi composto por 1.470 dados, tendo como variáveis: identificações dos pacientes, exames, analitos, resultados, valores de referência e diagnósticos.

Para a classificação do aprendizado de máquina supervisionado foram utilizados os seguintes algoritmos: rede neural *perceptron* de múltiplas camadas, SVM e árvore de decisão. Quanto ao algoritmo SVM foram alterados os kernels para *poly kernel*, *normalize poly kernel* e RBF com intuito

de verificar os resultados da classificação em diferentes situações. Também foi realizado o teste para o aprendizado de máquina não supervisionado utilizando o algoritmo k-means.

3 RESULTADOS E DISCUSSÃO

Os dados foram diagnosticados para os pacientes do Hospital Albert Einstein quanto aos resultados de exames durante o contexto da pandemia para diferentes algoritmos de aprendizado de máquina supervisionado. A ferramenta Weka classificou os dados com taxa de acurácia em 100% para os algoritmos de rede neural perceptron e SVM para o kernel poly kernel, 99.187% para o algoritmo de árvore de decisão, 86.1789% para o algoritmo SVM para o kernel normalize poly kernel e 84.9583% para o algoritmo SVM para o kernel RBF. Os testes foram realizados para os seguintes algoritmos: rede neural *perceptron* de múltiplas camadas, SVM em diferentes kernels e árvore de decisão (Tabela 1).

Tabela 1 – Algoritmos de aprendizado de máquina supervisionado

Dados	Rede Neural Perceptron	Árvore de Decisão	SVM Kernel Poly	SVM Kernel Normalize Poly	SVM Kernel RBF
Hospital Albert Einstein	100	99.187	100	86.1789	84.9583

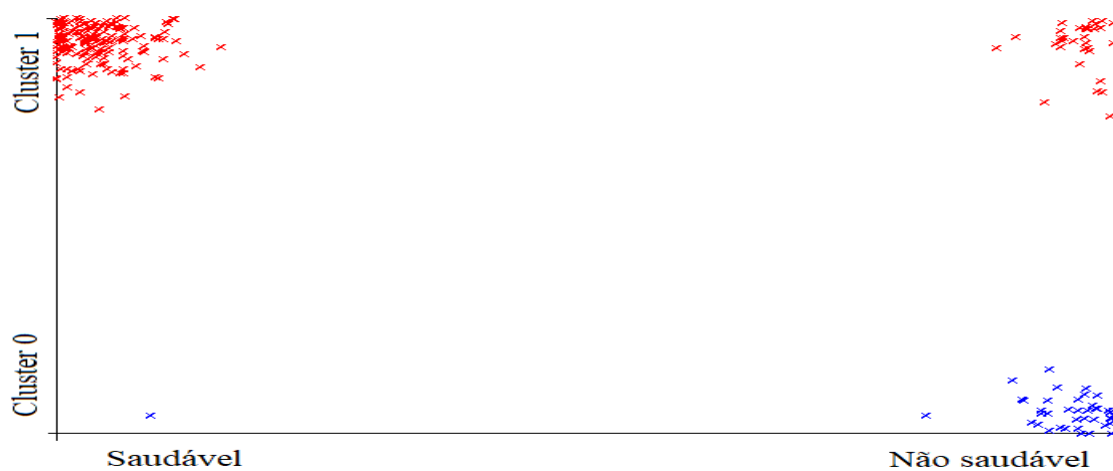
Nesse teste foi possível constatar que para os algoritmos de aprendizado de máquina supervisionado rede neural perceptron e SVM para o kernel poly tiveram excelentes taxas de acurácia em 100%, ou seja, os algoritmos classificaram corretamente todos os resultados dos exames dos pacientes do Hospital Albert Einstein durante o período da pandemia, tornando a tomada de decisão assertiva.

Em seguida foi realizado o mesmo teste para o algoritmo de aprendizado de máquina não supervisionado k-means com taxa de acurácia em 87,05%. Nesse teste foi possível verificar que para os resultados dos exames dos pacientes que estavam não saudáveis houve somente 1 dado clusterizado de forma errada. Já para os resultados dos exames que estavam saudáveis houve 32 dados clusterizados de forma errada, conforme ilustra a figura 1.

Observando os resultados obtidos neste teste (Figura 1), verificou-se que o aprendizado de máquina não supervisionado k-means clusterizou no cluster 0 os resultados dos exames dos pacientes que estavam não saudáveis e consequentemente no cluster 1 os pacientes saudáveis do Hospital Albert Einstein. Também foi possível observar que o algoritmo k-means teve uma taxa de acurácia melhor que os algoritmos SVM para o kernel normalize poly e kernel RBF. Vale ressaltar, que nesse caso não existe a parte de treinamento, ou seja, a taxa de acurácia se dá pela similaridade entre os dados. Ou

seja, o algoritmo k-means em alguns casos não conseguiu identificar a similaridade entre os dados dos pacientes saudáveis e não saudáveis.

Figura 1. Clusterização dos pacientes no contexto da pandemia



4 CONSIDERAÇÕES FINAIS

A partir da metodologia adotada para investigar a classificação dos resultados foi possível verificar uma classificação em 100% dos dados para os algoritmos rede neural perceptron e SVM para o kernel poly e clusterizou em 87,05% os dados para o algoritmo k-means. Portanto, os resultados encontrados neste estudo contribuem para o diagnóstico dos pacientes do Hospital Albert Einstein proporcionando um resultado robusto e com tomada de decisão assertiva

REFERÊNCIAS

Haykin, s. Neural networks: a comprehensive foundation. Ed. 2. Nova jersey: prentice-hall, 1999.

Silva, t. C.; fortes, r. C.; ferrão, p. A. Percepção de pacientes oncológicos quanto ao impacto da pandemia de covid-19 frente ao diagnóstico e tratamento de câncer. Brazilian journal od development, curitiba, v. 8, n. 1, p. 6508-6532, 2022.

Zhang, j.; dong, x.; cao, y.; yuan, y.; yang, y. Clinical characteristics of 140 patients infected with sars-co v-2 in wuhan, china. Allergy, v. 75, n. 7, p. 1730-1741, 2020. Doi: 10.1111/all.14238.