

Análise e Predição de Ataque do coração

Aluno: José de Ribamar Aguiar dos Santos Júnior

Profº: Selmo Eduardo

Disciplina: Fundamentos de Machine Learning

-Resumo:

As doenças cardiovasculares continuam sendo a primeira causa de morte no Brasil, responsáveis por quase 32% de todos os óbitos. Este artigo propõe-se ao estudo e análise de um dataset que se refere a dados de pacientes e sua propensão a serem acometidos por um ataque cardíaco seguido por aplicação de modelos de Machine Learning que prevê a chance de uma pessoa ser acometido por essa disfunção.

Palavras chaves: Machine Learning, ataque cardíaco, Predição

1-Introdução

O infarto do miocárdio, ou ataque cardíaco, é a morte das células de uma região do músculo do coração por conta da formação de um coágulo que interrompe o fluxo sanguíneo de forma súbita e intensa. A principal causa do infarto é a aterosclerose, doença em que placas de gordura se acumulam no interior das artérias coronárias, chegando a obstruí-las. Na maioria dos casos o infarto ocorre quando há o rompimento de uma dessas placas, levando à formação do coágulo e interrupção do fluxo sanguíneo. (Serrano, 2004)

O processo de prevenção de doenças é bem importante em vários cenários como em diminuição de mortes, menos pessoas com sintomas permanentes, menos gasto com saúde pública, etc.

Os métodos de aprendizado de máquina utilizados no presente trabalho serão: Support Vector Machine (SVM), k-nearest neighbors (KNN), Regressão Logística, Árvore de Decisão e Floresta Aleatória.

O SVM é um modelo de aprendizado de máquina supervisionado onde, no caso da classificação, busca se construir um classificador de acordo com um conjunto de padrões por ele identificado nos exemplos de treinamento, onde a classificação é conhecida. Busca se obter o hiperplano que maximiza a margem de separação entre os conjuntos. (Géron, 2019)

O kNN é um método de aprendizado supervisionado não paramétrico desenvolvido em 1951 e pode ser utilizado tanto para regressão como para classificação. Este algoritmo utiliza a proximidade para fazer classificações e previsões sobre os pontos de um ponto individual.

A Regressão Logística é comumente utilizada para estimar a probabilidade de uma instância pertencer a uma determinada classe. Se a probabilidade estimada for maior que 50%, então o modelo prevê que a instância

pertence a esta classe, ou então, ela prevê que não. Isso o transforma em um classificador binário. (Géron, 2019)

A Árvore de Decisão é um algoritmo bastante versátil que pode executar tarefas de classificação, regressão e, até mesmo, tarefas multioutput. São algoritmos muito poderosos capazes de moldar conjuntos complexos de dados.

Sabendo da importância do tema para nossa sociedade este artigo tem como perspectiva a aplicação de modelos de machine learning para predição desses possíveis ataques observando que métodos de aprendizagem terão os melhores resultados.

2- Metodologia

A metodologia utilizada no presente trabalho parte do estudo e análise do dataset com informações de pacientes através de uma Análise exploratória dos dados e a partir daí, com os resultados obtidos, aplica-se modelos de aprendizado de máquina para verificação dos resultados obtidos quanto à eficiência do modelo.

O dataset utilizado provém do site kaggle, uma comunidade on line de cientistas de dados e praticantes de aprendizado de máquina que permite colaborar com outros cientistas, encontrar e publicar dataset's e participar de competições de resolução de desafios sobre ciência de dados.

O Google Colab foi a plataforma escolhida no decorrer da pesquisa devido a facilidade que nos propõe a programar de forma gratuita sem precisar baixar softwares, ele permite que qualquer pessoa escreva e execute código Python arbitrário pelo navegador e é especialmente adequado para aprendizado de máquina, análise de dados e educação.

Para uma melhor aplicação dos dados ele foram padronizados para sua aplicação em todos os modelos de aprendizagem. A seguir calculou se através de métricas o resultado dos modelos para a posterior comparação entre os modelos.

A métrica utilizada para verificação da eficiência dos modelos é a acurácia, que é calculada através da seguinte fórmula:

$$A = \frac{(VN + VP)}{(VP + FN + VN + FP)}$$

Em que:

A- Acurácia

VN- Verdadeiros negativos

VP- Verdadeiros Positivos

FN- Falsos Negativos

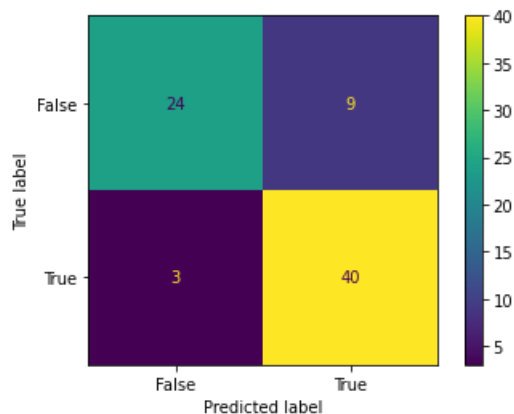
FP- Falsos Positivos

Para um efeito mais visual e melhor entendimento do leitor é utilizada a matriz de confusão para vermos esse resultado de forma mais explícita.

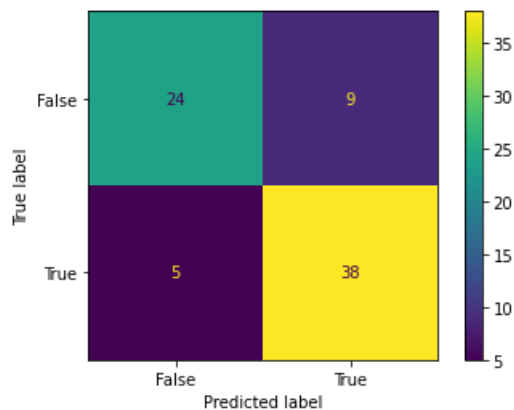
3- Resultados

Como resultados são mostrados a seguir as matrizes de confusão e a respectiva acurácia para cada modelo.

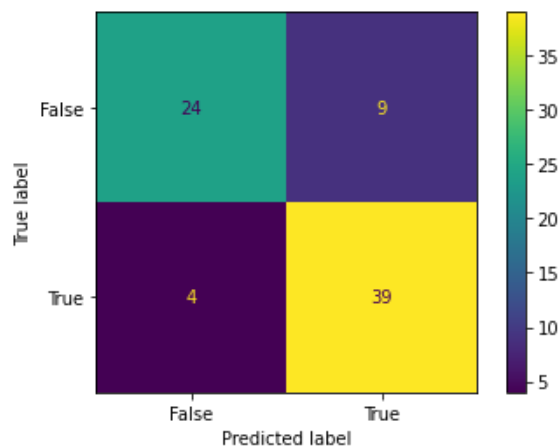
SVM:



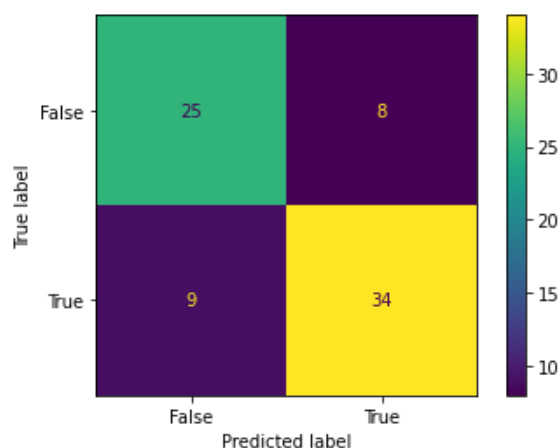
KNN:



Regressão Logística:



Árvore de Decisão:



Floresta Aleatória:

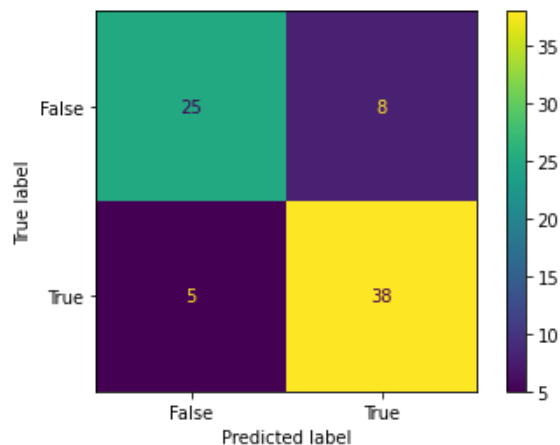


Tabela com Resultados de Acurácia

Modelo	Acurácia
SVM	0.8421
KNN	0.8158
Regressão Logística	0.8289
Árvore de Decisão	0.7763
Floresta Aleatória	0.8289

4- Conclusão

O presente artigo teve a proposta de utilizar vários métodos de aprendizado de máquina para classificar o o nível de chance de um paciente ter ou não um ataque cardíaco analisando as métricas para se observar os melhores resultados.

A média obtida com os modelos utilizados é uma acurácia de 0,8184 e o melhor resultado para os modelos foi utilizando o algoritmo SVM onde se obteve uma acurácia de 0,8421.

Para a área da saúde, do qual trata os dados utilizados este resultado para acurácia é considerado baixo, o ideal seria 98% pois essa previsão pode significar decisões importantes na vida de determinada pessoa.

Para trabalhos futuros uma sugestão seria a otimização dos parâmetros dos modelos utilizados e a aplicação de outros modelos para um melhor resultado de acurácia.

Referências

Seleção de recursos usando SelectKBest.

Disponível em:

“<https://www.kaggle.com/code/jepsds/feature-selection-using-selectkbest/notebook>”.

Acesso em: 13/12/2022.

MATOS, David. **Conceitos Fundamentais de Machine Learning**. [S. l.], 22 set. 2015.

Géron, Aurelio. **Mãos á obra: Aprendizado de Máquina com Scikit- Learn & TensorFlow**