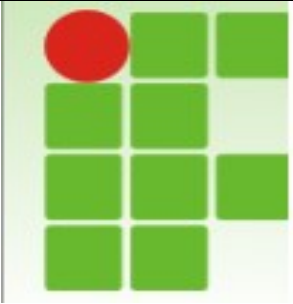


| | |
|--|---|
|  | <p>Instituto Federal de Educação Ciência e Tecnologia – Sudeste de Minas Gerais</p> <p>Disciplina: Machine Learning Data: 11/04/2023</p> <p>Aluno: _____ Valor: 3,0 pontos</p> <p>Data da Entrega: 28/04/2023 Grupo: 02 alunos Entrega: SIGAA</p> |
|--|---|

Considere o seguinte dataset:

https://github.com/gustavowillam/ML/blob/main/trabalhos/Trab1/StudentsPerformance_train.csv

Implemente um algoritmo de Machine Learning para realizar a predição do valor da Pontuação média do aluno. Considere como Pontuação média: (math score + reading score + writing score) / 3.

Execute as seguintes etapas da Análise Exploratória dos Dados (EDA):

Tratar dados ausentes (NaN)

Tratar dados categóricos.

Normalize/Padronize os dados de for necessário.

Implemente técnicas de Eliminação de Outliers

Seleção de Variáveis (*features*)

Implemente otimização de Parâmetros

Objetivos: Encontrar o melhor modelo que minimiza o RMSE e maximiza o R^2

Considere como variáveis independentes (X):

- a) gender (sexo)
- b) race/ethnicity (raça/etnia)
- c) parental level of education (nível de escolaridade dos pais)
- d) lunch (almoço)
- e) test preparation course (curso preparatório para testes)
- f) math score (pontuação de matemática)
- g) reading score (pontuação de leitura)
- h) writing score (pontuação de escrita)

Considere como variável dependente (y):

Mean score (Pontuação média) = (math score + reading score + writing score) / 3

1) Implemente uma solução utilizando as técnicas de Aprendizado de Máquina desenvolvidas em sala de aula.

2) Para medir a performance do seu modelo utilize o seguinte dataset de teste:

https://github.com/gustavowillam/ML/blob/main/trabalhos/Trab1/StudentsPerformance_test.csv

3) Implemente uma interface gráfica utilizando o framework python tkinter, para realizar o deploy do modelo de Machine Learning implementado. Seu sistema deverá aceitar como entrada dados categóricos (string). Realize o tratamento destes dados categóricos antes de enviá-los para o modelo de Machine Learning desenvolvido.

4) Apresentação do trabalho:

O trabalho deverá ser apresentado em sala de aula. Deverá ser explicado como foi implementado cada técnica: EDA, Eliminação de Outliers, Escalonamento dos dados, técnicas utilizadas para tratar dados categóricos, otimização de parâmetros, seleção de features, tratamento de valores nulos, etc...

5) Ponto Extra: (Com base na métrica: RMSE)

- a) 1,0 pontos (1º colocado do Ranking interno da sala de aula)
- b) 0,6 pontos (2º colocado do Ranking interno da sala de aula)
- c) 0,3 pontos (3º colocado do Ranking interno da sala de aula)