**Projeto A3**

A DataBase escolhida foi “Qualidade do Vinho Tinto”, que busca testar a qualidade de um vinho tinto português. Segue o link da DataBase:

<https://www.kaggle.com/datasets/uciml/redwine-quality-cortez-et-al-2009>

Ela possui as seguintes variáveis:

* Fixed Acidity (float64) - descreve a acidez;
* Volatile Acidity(float64) – descreve a quantidade de ácido acético;
* Citric Acid(float64) – descreve a quantidade de ácido cítrico;
* Residual Sugar(float64) – descreve a quantidade de açúcar restante após a fermentação;
* Chlorides(float64) – descreve a quantidade de sal no vinho;
* Free Sulfur Dioxide(float64) – descreve a quantidade de uma substância que previne micróbios e a oxidação do vinho;
* Total Sulfur Dioxide(float64) – descreve a quantidade da substância SO2, que acima de 50ppm é perceptível ao nariz e ao gosto;
* Density(float64) - descreve a densidade da água;
* pH(float64) – em uma escala de 0 a 14 descreve quão ácido o vinho é;
* Sulphates(float64) - descreve a quantidade deste aditivo que é antimicrobiano e antioxidante;
* Alcohol(float64) – descreve a quantidade de álcool contido no vinho;
* Quality(int64) – Escala de 0 a 10, descrevendo a qualidade.

A variável target escolhida foi a “Quality (Qualidade)”, porque busca classificar a qualidade do vinho que será testado, através das outras variáveis. Ela varia de 0 a 10, com valores mais altos indicando melhor qualidade.

**Aplicação do Primeiro Método:** No primeiro método, Regressão Linear, obtivemos um valor de **0.50** de variação pelo Mean Absolute Error;

**Aplicação do Segundo Método:** No segundo método, KNN, através do MAE, o resultado foi de **0.56**. Comparado com o primeiro método, uma leve alteração negativa de previsão;

**Aplicação de Terceiro Método:** No terceiro e último método, Redes Neurais, foi obtido o valor de **0.52**. No nosso caso melhor que o segundo método, mas abaixo do primeiro.

**Comparação:**

Todos os valores obtidos pelos métodos testados ficaram numa variação relativamente pequena, com o melhor sendo a Regressão Linear, pois temos variáveis que são independentes, mas que no conjunto fazem diferença para nossa variável target.