**TECH LEAD 2025.1**

**Projeto Final**

**Disciplina**: Data Science e IA

**Aluno1:** Sttiwe Washington F Sousa

**E-mail1:** [swfs@cesar.school](mailto:swfs@cesar.school)

**Aluno2:** Pedro William Bernardino

**E-mail2:** [pwbcf@cesar.school](mailto:pwbcf@cesar.school)

Student Performance: <https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression>

Inicialmente após ler o arquivo .csv com o CSV Reader podemos obter algumas informações sobre os dados. O arquivo contém um conjunto de dados de com 10 mil entradas, com as seguintes variáveis:

* ***Hours Studied***: O número total de horas dedicadas aos estudos por cada aluno.
* ***Previous Score***: As notas obtidas pelos alunos em provas anteriores.
* ***Extracurricular Activities***: Se o aluno participa de atividades extracurriculares (Sim ou Não).
* ***Sleep Hours***: O número médio de horas de sono que o aluno teve por dia.
* **Sample Question Papers Praticed**: O número de provas de exemplo que o aluno praticou.
* **Performance Index**: Média de desempenho de cada aluno.

Essa última variável será o nosso valor de referência para análise.

Iniciei com algumas visualizações. Incluí uma View Bar Chat para analisar e compara e visualizar graficamente a relação do desempenho com as outras variáveis.

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

Tabela

O conteúdo gerado por IA pode estar incorreto.

Tela de computador com fundo branco

O conteúdo gerado por IA pode estar incorreto.

Posteriormente incluí uma Linear Correlation para visualizar a correlação das variáveis com o valor de performance:

Interface gráfica do usuário, Texto, Aplicativo

O conteúdo gerado por IA pode estar incorreto.

Com a primeira análise podemos observar que os alunos com bom desempenho domem uma quantidade razoável de horas, o histórico de notas do aluno possui uma correlação muito alta no desempenho do aluno apresentando uma regularidade.

Tentei realizar a análise dos dados utilizando os exemplos visto em aula utilizando o a Regressão Linear Multivariável. Normalizei os dados e incluí o particionamento da tabela configurada em 75%, assim, utilizei o Linear Regression Learner e o Regression Predictor, denomalizei os dados e incluí o Numeric Scorer. Não compreendi bem por que o resultado ficou negativo em R^2 = -8.023.

Assim, segue imagem da utilização da regressão linear multivariável.

Interface gráfica do usuário

O conteúdo gerado por IA pode estar incorreto.

Realizei algumas pesquisas e umas das soluções recomenda a utilização da Simple Regression Tree Learner, a solução nos trouxe melhores resultados:

Diagrama

O conteúdo gerado por IA pode estar incorreto.

R^2 = 0.975

Mean ansolute error = 2.37

Tabela

O conteúdo gerado por IA pode estar incorreto.

Gráfico

O conteúdo gerado por IA pode estar incorreto.

Assim ficou toda solução:

Mapa de cidade

O conteúdo gerado por IA pode estar incorreto.