

TURMA	Inteligencia Artificial (Escolha 3) -N-ANALISE E DESENV. DE SISTEMAS			DISCIPLINA	IA
PERÍODO	NOTURNO	SEMESTRE	6º	PROFESSOR	VINICIUS GODOY
REF	EXERCICIO 3				
ALUNO	Elis Christina Gomes Mattosinho				
RA	0210482213005				

Exercício de Regressão Linear com Dados de Emissões de Veículos

Objetivo:

- Criar um modelo de regressão linear para prever a emissão de CO2 (CO2EMISSIONS) com base no tamanho do motor (ENGINE SIZE).
- Analisar os resultados da regressão linear e interpretar os coeficientes do modelo.
- Avaliar a qualidade do modelo e identificar possíveis outliers ou valores inconsistentes nos dados.

Recursos:

- Dataset “DATASET_CO2”
- Ambiente de desenvolvimento em Python (Google Colab).

Etapas:

1. **Importar os dados para o Google Colab.**
 - Certifique-se de que os dados estão formatados corretamente (colunas, tipos de dados, etc.).
2. **Explorar os dados.**
 - Visualizar a distribuição dos dados.
 - Identificar outliers ou valores inconsistentes.
3. **Dividir os dados em conjuntos de treinamento e teste.**
 - Utilize cerca de 75% dos dados para treinamento e 25% para teste.
4. **Treinar o modelo de regressão linear.**
 - Selecionar a variável `ENGINE SIZE` como preditora (X).
 - Ajustar o modelo e verificar os resultados.
5. **Avaliar o modelo de regressão linear.**
 - Calcular o SSE, MSE, MAE, RMSE e R2-score e interprete os resultados.
 - Verificar se os resultados são satisfatórios (R² entre 85% e 90%).
6. **Apresentar os resultados.**
 - Criar um relatório com os resultados das métricas e caso não forem satisfatórias apresente possíveis ações para melhorar o modelo.
 - Interpretar os resultados e tirar conclusões.
7. **Predições**
 - Qual o intercepto (b) e a inclinação (m)?
 - Apresente as emissões de CO2 para os veículos com os seguintes tamanhos de motor:

- | | |
|---------|---------|
| ▪ 2.0 = | ▪ 1.8 = |
| ▪ 1.5 = | ▪ 0.9 = |
| ▪ 3.0 = | ▪ 2.5 = |
| ▪ 4.5 = | |
-

Sem tratamento

5. Avaliar o modelo de regressão linear.

- Calcular o SSE, MSE, MAE, RMSE e R²-score e interprete os resultados.

Soma dos Erros ao Quadrado (SSE): 38896.15

Erro Quadrático Médio (MSE): 733.89

Erro Médio Absoluto (MAE): 21.78

Raiz do Erro Quadrático Médio (RMSE): 27.09

R²-score: 0.81

- Verificar se os resultados são satisfatórios (R² entre 85% e 90%).

O R² atingiu um resultado de 81%, indicando insatisfação com base na meta definida.

6. Apresentar os resultados.

- Criar um relatório com os resultados das métricas e caso não forem satisfatórias apresente possíveis ações para melhorar o modelo.
- Interpretar os resultados e tirar conclusões.

Relatório de Avaliação do Modelo de Regressão Linear

Introdução:

Neste relatório, avaliamos o desempenho de um modelo de regressão linear utilizando as métricas Soma dos Erros ao Quadrado (SSE), Erro Quadrático Médio (MSE), Erro Médio Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE) e R²-score. A análise foi realizada com o objetivo de determinar a eficácia do modelo e propor ações de melhoria caso os resultados não atendam às metas estabelecidas.

Métricas Calculadas:

- Soma dos Erros ao Quadrado (SSE): 38.896,15

- Erro Quadrático Médio (MSE): 733,89
- Erro Médio Absoluto (MAE): 21,78
- Raiz do Erro Quadrático Médio (RMSE): 27,09
- R^2 -score: 0,81

Interpretação dos Resultados:

- SSE (Soma dos Erros ao Quadrado): Esse valor reflete a soma total dos erros ao quadrado, com valores mais baixos indicando um modelo mais preciso. No entanto, por ser uma métrica dependente da escala dos dados, ela não permite comparações diretas com outros modelos ou datasets.
- MSE (Erro Quadrático Médio): Representa a média dos erros ao quadrado. Um valor de 733,89 sugere que, em média, as previsões do modelo estão erradas por 733,89 unidades quadráticas.
- MAE (Erro Médio Absoluto): Indica o erro médio absoluto nas previsões. Com um valor de 21,78, isso significa que o modelo, em média, se desvia das observações reais por 21,78 unidades.
- RMSE (Raiz do Erro Quadrático Médio): A RMSE, que é a raiz quadrada do MSE, fornece uma interpretação mais direta dos erros em termos da unidade de medida dos dados. Um RMSE de 27,09 implica que as previsões do modelo geralmente têm um erro de 27,09 unidades.
- R^2 -score: O valor de R^2 de 0,81 indica que 81% da variação nos dados é explicada pelo modelo. Embora um R^2 de 81% demonstre que o modelo captura a maior parte da variação nos dados, ele não atinge a meta estabelecida entre 85% e 90%.

Conclusões e Ações de Melhoria:

O R^2 de 0,81 sugere que o modelo de regressão linear sem tratamento possui uma capacidade considerável de explicação, mas não atende às expectativas estabelecidas. As outras métricas, como MSE, MAE e RMSE, indicam a presença de erros que poderiam ser otimizados.

Ações de Melhoria:

- Regularização: Implementar técnicas de regularização (Lasso, Ridge, Elastic Net) para penalizar coeficientes excessivamente grandes e reduzir o overfitting.
- Verificação de Outliers: Identificar e tratar outliers, que podem estar distorcendo as previsões do modelo.
- Divisão do Dataset: Revisar a divisão do dataset em treinamento, validação e teste, garantindo que seja representativa e que não haja vazamento de dados.

Essas ações podem ajudar a melhorar a precisão do modelo e aumentar o valor de R^2 , aproximando-o das metas desejadas.

7. Predições

- Qual o intercepto (b) e a inclinação (m)?

O intercepto (b) é igual a 140.74518994 e a inclinação (m) é igual a 35.75317017.

- Apresente as emissões de CO₂ para os veículos com os seguintes tamanhos de motor:
 - 2.0 = 212.25 g/km
 - 1.5 = 194.37 g/km
 - 3.0 = 248.00 g/km
 - 4.5 = 248.00 g/km
 - 1.8 = 205.10 g/km
 - 0.9 = 172.92 g/km
 - 2.5 = 230.13 g/km
-

Com tratamento

8. Avaliar o modelo de regressão linear.

- Calcular o SSE, MSE, MAE, RMSE e R²-score e interprete os resultados.
- Verificar se os resultados são satisfatórios (R^2 entre 85% e 90%).

9. Apresentar os resultados.

- Criar um relatório com os resultados das métricas e caso não forem satisfatórias apresente possíveis ações para melhorar o modelo.
- Interpretar os resultados e tirar conclusões.

10. Predições

- Qual o intercepto (b) e a inclinação (m)?
- Apresente as emissões de CO₂ para os veículos com os seguintes tamanhos de motor:
 - 2.0 =
 - 1.5 =
 - 3.0 =
 - 4.5 =
 - 1.8 =
 - 0.9 =
 - 2.5 =