

## Atividade Prática – Regressão Linear

**Nome(s):** \_\_\_\_\_ **Data:**    /    /

**Objetivo:** Explorar o conjunto de dados, visualizar os dados, aplicar um modelo de regressão linear e avaliar o desempenho do modelo.

### Instruções iniciais:

- O conjunto de dados contém informações sobre veículos fabricados no ano 2000, incluindo dados sobre consumo de combustível e emissões de CO2. Essas informações são importantes para entender o impacto ambiental dos veículos e discutir soluções mais sustentáveis. O foco desta aula será construir um modelo de Regressão Linear para prever as emissões de CO2 com base em variáveis como tamanho do motor e consumo de combustível.

### Configuração do Google Colab:

- **Passo 1:** Abra o Google Colab.
- **Passo 2:** Carregue o conjunto de dados disponível no GitHub em “Extra”
- **Passo 3:** Certifique-se de que as bibliotecas a seguir estão importadas:
  - Pandas (para manipulação de dados)
  - Matplotlib e Seaborn (para visualização)
  - Scikit-learn (para aplicar o modelo de Regressão Linear)

Dica: Se precisar importar as bibliotecas, use o seguinte código:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

### Exploração dos Dados:

- **Passo 1:** Carregue o conjunto de dados em um DataFrame usando o Pandas.
- **Passo 2:** Exiba as primeiras linhas para verificar o formato dos dados (`df.head()`).
- **Passo 3:** Verifique as informações do conjunto de dados, como tipos de variáveis e valores nulos (`df.info()`).
- **Passo 4:** Use `df.describe()` para visualizar estatísticas descritivas dos atributos numéricos.
- **Passo 5:** Verifique se há valores faltantes e, se houver, discuta possíveis maneiras de tratá-los (por exemplo, preenchimento ou remoção de linhas).

### Visualização de Dados

- **Passo 1:** Crie um gráfico de dispersão para explorar a relação entre:
  - Consumo de combustível e emissões de CO2.
  - Tamanho do motor e emissões de CO2.

### Preparação dos Dados para o Modelo de Regressão Linear

- **Passo 1:** Selecione as variáveis que você usará como independentes (por exemplo, tamanho do motor, consumo de combustível).
- **Passo 2:** Defina as emissões de CO2 como a variável dependente (y).
- **Passo 3:** Divida os dados em conjuntos de treino e teste. A divisão recomendada é 80% para treino e 20% para teste. Use o seguinte código:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

### Aplicação do Modelo de Regressão Linear

- **Passo 1:** Crie uma instância do modelo de regressão linear e ajuste-o aos dados de treino.

```
model = LinearRegression()  
model.fit(X_train, y_train)
```

- **Passo 2:** Use o modelo treinado para fazer previsões no conjunto de teste.

```
y_pred = model.predict(X_test)
```

### Avaliação do Desempenho do Modelo

- **Passo 1:** Calcule as métricas de desempenho do modelo:

- **R<sup>2</sup> (coeficiente de determinação):**

```
r2 = r2_score(y_test, y_pred)
```

- **Passo 2:** Interprete essas métricas para entender como o modelo está performando.

**Enviar o notebook do Colab até as 20h30 do dia 24/09/2024.**