

## Projeto de Machine Learning – AP1

### Objetivo:

Aplicar técnicas estatísticas e modelos básicos de Machine Learning a um conjunto de dados real, explorando desde a análise inicial até a disponibilização dos modelos via API REST.

### Formação de Grupos:

Duplas.

### Descrição da Atividade:

Cada dupla deverá escolher **um dataset disponível publicamente no Kaggle** (ou similar), contendo ao menos uma variável numérica contínua e uma variável categórica binária, e realizar as seguintes etapas:

### Etapas do Trabalho

1. **Escolha e apresentação do dataset**
  - Escolha um dataset do Kaggle que possibilite a aplicação de regressão linear simples (predição) e regressão logística (classificação);
  - Apresente brevemente o conjunto de dados escolhido, incluindo: link para acesso, número de registros e variáveis, e a justificativa da escolha.
2. **Pré-processamento e análise exploratória**
  - Trate valores ausentes e/ou inconsistentes, se houver.
  - Gere estatísticas descritivas básicas e gráficos de dispersão e distribuição.
3. **Testes de Normalidade**
  - Aplique **ao menos um teste de normalidade** (por exemplo: Shapiro-Wilk, Kolmogorov-Smirnov) à variável numérica escolhida.
  - Interprete os resultados e discuta a adequação à regressão linear.
4. **Coeficiente de Correlação**
  - Calcule e interprete o **coeficiente de correlação (Pearson ou Spearman)** entre duas variáveis numéricas do dataset.
  - Explique a direção e a força da correlação encontrada.
5. **Regressão Linear Simples (predição)**
  - Selecione uma variável como preditora (X) e outra como resposta (Y contínua).
  - Crie e treine um modelo de regressão linear simples.
  - Avalie o modelo com métricas apropriadas ( $R^2$ , MAE, RMSE) e gere um gráfico de dispersão com a linha de regressão.
6. **Regressão Logística (classificação)**
  - Selecione uma variável binária como variável-alvo (Y binária) e uma variável preditora (X).
  - Treine um modelo de regressão logística.
  - Avalie a performance com métricas como **acurácia e matriz de confusão**.

## UNIDADES

### MINAS GERAIS

Rua Rio Grande do Norte, 300  
Funcionários  
Belo Horizonte - Minas Gerais  
Telefone: 0800 771 8020

### DISTRITO FEDERAL

SIG Quadra 4 Bl. A  
Edifício Capital Financial Center  
Brasília - Distrito Federal  
Telefone: 0800 771 8020

### RIO DE JANEIRO | BARRA

Av. Armando Lombardi, 940  
Barra da Tijuca  
Rio de Janeiro - Rio de Janeiro  
Telefone: 0800 771 8020

### RIO DE JANEIRO | CENTRO

Av. Presidente Wilson, 118  
Centro  
Rio de Janeiro - Rio de Janeiro  
Telefone: 0800 771 8020

### SÃO PAULO

AL. Santos, 2356  
Jd. Paulista  
São Paulo - São Paulo  
Telefone: 0800 771 8020

## 7. Disponibilização via API REST

- Desenvolva uma API REST em R com dois endpoints:
  - `/predicao`: recebe um valor de entrada X e retorna a predição Y do modelo de regressão linear.
  - `/classificacao`: recebe um valor de entrada X e retorna a classe prevista pelo modelo de regressão logística.
- Documente a API usando Swagger ou similar (opcional, mas recomendado).

---

## Entrega

**Prazo final:** 09/04/2025 (23h59min)

**Forma de entrega:** Cada membro da dupla deve enviar link para repositório no GitHub contendo:

- Código-fonte do projeto;
- Notebooks ou scripts com os testes e modelos;
- Documentação da API;
- Relatório final (em PDF ou Markdown) com explicações, gráficos e interpretações dos resultados.

---

## UNIDADES

### MINAS GERAIS

Rua Rio Grande do Norte, 300  
Funcionários  
Belo Horizonte - Minas Gerais  
Telefone: 0800 771 8020

### DISTRITO FEDERAL

SIG Quadra 4 Bl. A  
Edifício Capital Financial Center  
Brasília - Distrito Federal  
Telefone: 0800 771 8020

### RIO DE JANEIRO | BARRA

Av. Armando Lombardi, 940  
Barra da Tijuca  
Rio de Janeiro - Rio de Janeiro  
Telefone: 0800 771 8020

### RIO DE JANEIRO | CENTRO

Av. Presidente Wilson, 118  
Centro  
Rio de Janeiro - Rio de Janeiro  
Telefone: 0800 771 8020

### SÃO PAULO

Al. Santos, 2356  
Jd. Paulista  
São Paulo - São Paulo  
Telefone: 0800 771 8020