

# Detecção de Fraudes em Cartões de Crédito com Machine Learning





# Objetivo

Construir um modelo de machine learning para identificar transações fraudulentas em cartões de crédito, maximizando o desempenho em um cenário de dados desbalanceados. Sendo o objetivo aplicar uma combinação de abordagens para otimizar o modelo e garantir a capacidade de identificar fraudes com precisão.





# Descrição do Dataset

O dataset contém transações de cartão de crédito, sendo a classe "fraude" uma pequena porcentagem do total, o que cria um cenário de dados desbalanceados. A classe minoritária representa transações fraudulentas, enquanto a classe majoritária representa transações legítimas.





## **Você pode resolver este problema utilizando algumas técnicas como:**

1. Tuning de Hiperparâmetros
2. Técnicas de Balanceamento (Sampling)
3. Seleção de Atributos (Feature Selection)
4. Ensemble de Modelos





# Seleção de Atributos (Feature Selection)

- Selecionar as features mais relevantes para o problema, utilizando métodos como SelectKBest, Recursive Feature Elimination (RFE), ou a análise de Feature Importance com base em modelos de árvore (ex.: Random Forest).
- Remover features menos informativas para reduzir a complexidade e aumentar a interpretabilidade do modelo.





# Técnicas de Balanceamento (Sampling)

- Para lidar com o desbalanceamento, aplique técnicas de oversampling (como SMOTE) para aumentar a quantidade de instâncias da classe minoritária, ou undersampling para reduzir a classe majoritária.
- Explore diferentes abordagens e avalie o impacto de cada uma no desempenho do modelo.





# Tuning de Hiperparâmetros

- Realize ajustes finos nos hiperparâmetros dos modelos usando GridSearchCV ou RandomizedSearchCV com validação cruzada.
- Otimize modelos como Random Forest, SVM e XGBoost ajustando parâmetros importantes para maximizar as métricas de interesse.





# Ensemble de Modelos

- Experimente combinar diferentes modelos (ex.: Random Forest, Gradient Boosting, Voting Classifier) para formar um ensemble que aumente a robustez e melhore o desempenho.
- Crie um ensemble com Voting Classifier, combinando modelos como Decision Tree, SVM e KNN para comparar o impacto de cada combinação.







# Avaliação / Análise

Avalie o desempenho do seu modelo final usando métricas adequadas para dados desbalanceados:

- Acurácia
- Macro Avg F1-Score: Para equilibrar precisão e recall entre as classes.
- Matriz de Confusão: Avaliar acertos e erros por classe para identificar padrões de erro.
- Mostre quais técnicas foram aplicadas e quais abordagens mais contribuíram para o aumento do desempenho. Apresentando o impacto do desbalanceamento das classes e como as técnicas escolhidas afetaram o modelo.

