

**UNIVERSIDADE CATÓLICA DE BRASÍLIA**  
**CIÊNCIA DA COMPUTAÇÃO**

FELIPE PORTO DA SILVEIRA PRATES

**PREVIC**

Previsão de Inadimplência de Crédito

Brasília  
2025

## SUMÁRIO

INTRODUÇÃO.....	3
OBJETIVOS.....	4
ANÁLISE DE DADOS (EDA).....	5
MODELOS UTILIZADOS.....	6
CONCLUSÕES E RECOMENDAÇÕES.....	7
REFERÊNCIAS.....	8

# INTRODUÇÃO

Este relatório apresenta o desenvolvimento de um modelo preditivo para avaliar o risco de inadimplência de clientes de uma instituição financeira.

O objetivo é identificar clientes com maior risco de não pagar os empréstimos contratados, utilizando dados financeiros e comportamentais.

Para isso, foi realizado um trabalho de análise exploratória (EDA), seguido pela aplicação de técnicas de machine learning, com o intuito de melhorar a precisão da previsão de inadimplência.

O conjunto de dados utilizado contém variáveis como renda, idade, saldo bancário, histórico de crédito e a informação sobre inadimplência (default).

A ferramenta desenvolvida utiliza a interface interativa do Streamlit, permitindo ajustes de modelos e parâmetros em tempo real.

# OBJETIVOS

O objetivo deste projeto é desenvolver um modelo preditivo de inadimplência utilizando um conjunto de dados financeiros e comportamentais. As metas do projeto são:

- Exercitar técnicas de leitura e limpeza de dados utilizando a biblioteca Pandas.
- Desenvolver uma análise exploratória dos dados (EDA) e criar visualizações utilizando Matplotlib e Seaborn.
- Aplicar algoritmos de classificação para prever a inadimplência, com o uso do scikit-learn.
- Avaliar a performance dos modelos utilizando métricas como matriz de confusão, AUC, recall e precisão.
- Criar uma apresentação interativa, por meio de um notebook bem documentado ou um aplicativo simples utilizando o Streamlit.

# ANÁLISE DE DADOS (EDA)

Durante a análise exploratória dos dados, as seguintes conclusões foram extraídas:

- A maioria dos clientes apresenta limite de crédito entre R\$ 50.000 e R\$ 200.000.
- A inadimplência é mais comum entre clientes com menos de 30 anos.
- Solteiros(as) representam uma maior proporção de inadimplentes.
- Clientes com menor escolaridade tendem a ter maiores taxas de inadimplência.
- Existe uma correlação significativa entre os valores das faturas de meses consecutivos, o que sugere um comportamento de inadimplência repetitiva.

Essas descobertas foram fundamentais para orientar a escolha dos modelos preditivos e definir quais características seriam mais importantes para prever o risco de inadimplência

# MODELOS UTILIZADOS

Para a tarefa de previsão de inadimplência, foram aplicados os seguintes modelos de machine learning:

## **XGBoost:**

- Acurácia: 80.86%
- Recall (inadimplentes): 50.97%
- Precision (inadimplentes): 56.16%
- F1-score (inadimplentes): 53.44%

## **Random Forest (Padrão):**

- Acurácia: 81.98%
- Recall (inadimplentes): 25.46%
- Precision (inadimplentes): 73.73%
- F1-score (inadimplentes): 37.85%

Embora o modelo Random Forest tenha mostrado uma ligeira vantagem em termos de acurácia, o XGBoost se destacou no recall, o que é mais relevante para identificar inadimplentes.

## CONCLUSÕES E RECOMENDAÇÕES

Concluimos que o modelo XGBoost apresentou melhor desempenho para identificar inadimplentes, devido ao seu maior recall, que é crucial para identificar o risco de inadimplência de maneira mais eficiente.

O modelo Random Forest, embora tenha apresentado uma acurácia ligeiramente maior, demonstrou menor sensibilidade para a classe de inadimplentes, resultando em um recall mais baixo.

Recomendamos a implementação de um sistema de previsão de inadimplência utilizando o modelo XGBoost, com a possibilidade de ajustar o limiar de decisão conforme a tolerância da instituição a falsos negativos. Isso permitirá um ajuste dinâmico da política de crédito, considerando diferentes perfis de risco dos clientes.

Por fim, destacamos a importância de uma avaliação contínua do modelo para garantir sua eficácia ao longo do tempo, adaptando-se a mudanças nos padrões de comportamento de inadimplência.

# REFERÊNCIAS

[Matplotlib documentation — Matplotlib 3.10.1 documentation](#)

[Streamlit documentation](#)

[seaborn: statistical data visualization — seaborn 0.13.2 documentation](#)

[pandas documentation — pandas 2.2.3 documentation](#)

[scikit-learn: machine learning in Python — scikit-learn 1.6.1 documentation](#)

[XGBoost Documentation — xgboost 3.1.0-dev documentation](#)

[Plotly Python Graphing Library](#)

[Random Forest Algorithm in Machine Learning | GeeksforGeeks](#)

[Visualize Anything in Python: A Crash Course in Matplotlib & Seaborn | by Logupradeep | May, 2025 | Python in Plain English](#)