

RELATÓRIO TÉCNICO

Título: Sistema Preditivo de Inapimplênciam Infantil

Aluno:	Matheus da Silva Fontes
Disciplina:	Inteligência Artificial
Professor:	Professor Ronierison Maciel
Data:	16/11/2025

RESUMO EXECUTIVO

O projeto estabeleceu um pipeline completo de Machine Learning (ML) focado na previsão de inadimplência estudantil em financiamentos acadêmicos.

A estratégia central do sistema reside na escalabilidade e na utilização de um conjunto de variáveis financeiras e demográficas capazes de estimar o risco de um aluno se tornar inadimplente.

Para garantir robustez, o dataset foi padronizado, balanceado e enriquecido por meio de técnicas de engenharia de atributos, incluindo métricas financeiras derivadas da relação entre renda, valor financiado e parcelas estimadas.

A entrega final é um Dashboard interativo, totalmente funcional, que utiliza um Modelo de Regressão para classificar estudantes entre adimplentes (0) e inadimplentes (1).

METODOLOGIA E ARQUITETURA DE DADOS

Feature Engineering e Coerência Estratégica

O dataset foi construído a partir de atributos típicos da avaliação de crédito estudantil.

Além das variáveis principais, o pipeline aplicou engenharia de atributos para criar métricas financeiras essenciais ao entendimento do risco de inadimplência:

loan_to_income (razão financiamento / renda)

estimated_monthly_payment (estimativa de parcela mensal)

pct_income_commitment (comprometimento percentual da renda)

age_bucket (faixas etárias)

score_bucket (faixas de score)

A base limpa resultante (loan_features.csv) serviu de input para todo o pipeline: análise exploratória (EDA), treinamento do modelo e execução do Dashboard.

Isso garante coerência técnica de ponta a ponta, sem divergência entre desenvolvimento e ambiente final.

Modelo de Machine Learning

Algoritmo: Regressão Logística (com `class_weight="balanced"`).

O modelo foi escolhido por ser adequado para classificação binária e por atender às exigências do professor.

Features de Entrada (X):

- renda
- idade
- score
- valor
- meses_atraso
- loan_to_income
- estimated_monthly_payment
- pct_income_commitment
- age_bucket (OneHot)
- score_bucket (OneHot)

Target (Y):

- **0 = Adimplente**
- **1 = Inadimplente**

A escolha da Regressão Logística também se justifica pela boa interpretabilidade dos coeficientes e forte aderência a cenários de risco financeiro.

ANÁLISE E RESULTADOS DE PERFORMANCE

A avaliação foi realizada utilizando o conjunto de teste.

O modelo demonstrou desempenho consistente, medindo discriminabilidade entre bons e maus pagadores.

As principais métricas calculadas foram:

Métrica	Descrição
AUC	Medida global da separação entre classes
Gini	Derivado do AUC, muito usado em crédito
KS (Kolmogoro v-Smirnov)	Distância entre distribuições previstas
Lift@10%	Efetividade no top 10% mais arriscado
Precisão, Recall e F1	Qualidade da classificação binária

A saída das métricas foi salva em:

models/metrics_summary_noleak.csv

Esses indicadores permitem validar se o modelo consegue distinguir adequadamente perfis de risco.

ENTREGA FINAL: DASHBOARD DE PRODUÇÃO

O sistema é entregue com um Dashboard Interativo (Streamlit) que facilita a utilização prática do modelo.

Escalabilidade e Interatividade

O Dashboard permite que o usuário forneça informações individuais de um estudante:

Renda mensal

Idade

Score de crédito

Valor do financiamento

Meses em atraso

Com isso, o sistema:

- Realiza a predição em tempo real
- Exibe um gráfico interativo (Plotly) com os valores inseridos
- Apresenta, de forma clara e intuitiva:
 - ADIMPLENTE (0)
 - INADIMPLENTE (1)

Diferencial: Explicabilidade e Valor de Negócio

O dashboard foi projetado para fornecer clareza operacional ao usuário, permitindo que equipes acadêmicas identifiquem rapidamente alunos de maior risco e adotem ações preventivas, como:

- renegociação antecipada
- alertas financeiros
- análise individual de crédito

Embora não possua PLN contextual, a estrutura do sistema permite fácil expansão.

CONCLUSÃO

O Sistema Preditivo de Inadimplência Estudantil é uma solução completa que integra dados, modelagem estatística e engenharia de software.

O pipeline segue boas práticas de ML e garante coerência de dados do início ao fim.

O modelo foi entregue funcional com:

- Pipeline automatizado
- Feature engineering completa
- Balanceamento das classes
- Métricas salvas
- Dashboard operacional
- Interface intuitiva

Apesar do bom desempenho, recomenda-se que novas iterações sejam incorporadas como modelos não-lineares (Random Forest, Gradient Boosting), seleção de variáveis e ampliação da base de dados.