

Relatório da Base de dados

1. Descrição do Banco de Dados

Este conjunto de dados é uma versão sintética baseada no dataset original de Risco de Crédito disponível no Kaggle, enriquecida com variáveis adicionais relacionadas ao Risco Financeiro para Aprovação de Empréstimos. O dataset foi ampliado usando a técnica SMOTENC para gerar novos pontos de dados, aumentando o número de instâncias e garantindo uma diversidade de amostras.

A estrutura da base de dados inclui tanto variáveis categóricas quanto contínuas, totalizando 45.000 registros com 14 variáveis.

2. Descrição das Variáveis

1. **person_age**: Idade da pessoa (Tipo: Float)
2. **person_gender**: Gênero da pessoa (Tipo: Categórico)
3. **person_education**: Nível mais alto de educação alcançado (Tipo: Categórico)
4. **person_income**: Renda anual (Tipo: Float)
5. **person_emp_exp**: Anos de experiência de trabalho (Tipo: Inteiro)
6. **person_home_ownership**: Status de posse da casa (ex.: alugada, própria, hipotecada) (Tipo: Categórico)
7. **loan_amnt**: Valor do empréstimo solicitado (Tipo: Float)
8. **loan_intent**: Finalidade do empréstimo (Tipo: Categórico)
9. **loan_int_rate**: Taxa de juros do empréstimo (Tipo: Float)
10. **loan_percent_income**: Percentual do valor do empréstimo em relação à renda anual (Tipo: Float)
11. **cb_person_cred_hist_length**: Duração do histórico de crédito em anos (Tipo: Float)
12. **credit_score**: Pontuação de crédito da pessoa (Tipo: Inteiro)
13. **previous_loan_defaults_on_file**: Indicador de inadimplência em empréstimos anteriores (Tipo: Categórico)
14. **loan_status**: Status de aprovação do empréstimo: 1 = aprovado; 0 = rejeitado (Tipo: Inteiro, variável alvo)

3. Objetivo da Análise

O objetivo principal desta análise é construir modelos de classificação para prever o status de aprovação do empréstimo (**loan_status**) com base nos atributos pessoais e relacionados ao empréstimo.

Grupo:

- Carlos Moura
- Clara Mesquita
- Davi Castro
- Marcelo Gonçalves