

PRÓ-REITORIA ACADÊMICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO
NOVAS TECNOLOGIAS

**Relatório Detalhado sobre Análise de Inadimplência de
Crédito**

Autoras: Ingrid de Oliveira Bonifácio
Camila Mendes Osterno
Cariny Saldanha Oliveira

Orientador: William Roberto Malvezzi

Sumário

| | |
|---|----------|
| Introdução | 3 |
| Principais Insights do EDA | 3 |
| Descrição dos Modelos e Métricas Obtidas | 4 |
| Conclusões e Recomendação de Negócio | 5 |
| Considerações Finais | 6 |

1. Introdução

O trabalho tem como objetivo analisar conjuntos de dados sintéticos para prever inadimplência de crédito usando técnicas de Análise Exploratória de Dados (EDA) e modelagem usando algoritmos de aprendizado de máquina, como regressão logística e florestas aleatórias.

A previsão de inadimplência de crédito é crucial para o gerenciamento de risco de instituições financeiras, podendo ajudá-las a prever a probabilidade de inadimplência do cliente e ajustar estratégias de crédito e cobrança.

2. Principais Insights do EDA (Análise Exploratória de Dados)

A análise exploratória foi realizada com a finalidade de entender a composição, a distribuição e as principais interações entre as variáveis do conjunto de dados. As etapas principais incluíram:

- **Verificação e tratamento de dados:**
Foram identificados dados ausentes, duplicados e inconsistentes, que foram devidamente corrigidos. A análise dos tipos de variáveis indicou a necessidade de converter algumas colunas categóricas.
- **Estatísticas descritivas:**
Foram analisadas medidas como média, mediana, desvio padrão e a distribuição das variáveis. Em algumas delas, foram encontrados valores extremos (outliers), apontando para a importância da normalização.
- **Visualizações gráficas:**
Através de gráficos de dispersão, histogramas e mapas de calor, foi possível identificar relações relevantes entre variáveis, como correlações positivas entre a renda e a aprovação de crédito, além de uma correlação negativa com a inadimplência.

3. Descrição dos Modelos e Métricas Obtidas

- **Regressão Logística:**

A **Regressão Logística** foi feita e aplicada para prever a probabilidade de inadimplência com base nas variáveis financeiras. Este modelo assume uma relação linear entre as variáveis preditoras e a probabilidade de ocorrência de inadimplência.

- **Random Forest:**

O modelo **Random Forest** foi utilizado como uma técnica mais robusta, construindo múltiplas árvores de decisão e combinando-as para melhorar a precisão e reduzir o overfitting.

- **Métricas Obtidas:**

AUC (Área sob a Curva ROC): 0.85. O modelo Random Forest superou a regressão logística, apresentando uma AUC significativamente maior, indicando que ele tem uma maior capacidade de discriminação entre as classes de inadimplência.

4. Conclusões e Recomendações de Negócio

- **Desempenho dos Modelos:**

O Random Forest se destacou como o modelo mais eficaz para a previsão de inadimplência, com uma AUC de 0.85, em comparação com a Regressão Logística, que obteve uma AUC de 0.73. com isso o Random Forest consegue capturar melhor as complexas relações não lineares entre as variáveis e a variável de inadimplência.

- **Importância das Variáveis:**

As variáveis mais influentes na previsão de inadimplência foram dívida, limite de crédito e pontuação de crédito. Essas variáveis têm uma relação direta com a capacidade de pagamento do cliente e o risco de inadimplência. A dívida e o limite de crédito são fatores essenciais que indicam a exposição do cliente ao risco de não pagamento.

- **Recomendações de Negócio:**

Segmentação de Clientes: As variáveis mais importantes para a previsão de inadimplência, as instituições financeiras podem desenvolver estratégias de segmentação de clientes. Por exemplo, clientes com alta dívida e baixa pontuação de crédito podem ser classificados como de alto risco e receber ofertas de produtos financeiros com condições mais restritivas.

Estratégias de Mitigação de Risco: Os bancos e instituições financeiras devem adotar estratégias diferenciadas para clientes com características de alto risco, como aumentar a cobrança antecipada ou oferecer planos de pagamento alternativos, a fim de minimizar as perdas financeiras.

Aprimoramento de Modelos: Embora florestas aleatórias tenham bom desempenho, a análise de outras técnicas, como aumento de gradiente ou redes neurais, pode ser explorada para melhorar ainda mais o poder preditivo do modelo. Além disso, coletar mais dados, como histórico de pagamentos, pode ajudar em uma melhor modelagem.

5. Considerações Finais

O uso de modelos de aprendizado de máquina, como florestas aleatórias, demonstrou ser eficaz na previsão de inadimplência de crédito, superando modelos mais simples, como regressão logística. A análise de variáveis financeiras e sua correlação com o comportamento de inadimplência fornece insights valiosos para a tomada de decisões de crédito e gestão de riscos. Recomendações de segmentação de clientes e estratégias de mitigação de riscos podem ser empregadas para reduzir o impacto da inadimplência nas instituições financeiras.

- Random Forest mostrou robustez, ainda que com menor sensibilidade na classe minoritária.
- Fatores como renda e histórico de crédito mostraram forte influência na inadimplência

Recomendações de Negócios

- Aprimorar a coleta de dados (ex: estabilidade de renda, histórico bancário)
- Aplicar segmentações preditivas para personalizar ofertas e mitigar riscos