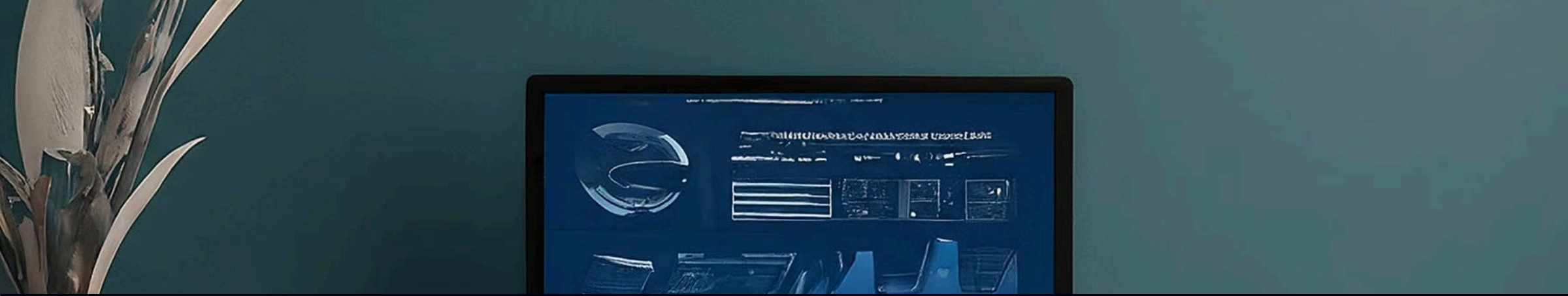


Análise de Risco de Crédito

Esta análise explora um conjunto de dados sobre risco de crédito, fornecendo insights valiosos sobre fatores que influenciam a inadimplência de empréstimos. Através de visualizações e técnicas de processamento de dados, podemos obter uma compreensão mais profunda desse tema crucial para instituições financeiras e tomadores de empréstimos.



by Tállyto Rodrigues



Visão Geral dos Dados

1

Tamanho do Conjunto de Dados

O conjunto de dados contém 32.576 registros, cada um representando um empréstimo individual com informações sobre a idade do tomador, renda, valor do empréstimo e status do empréstimo (inadimplente ou não).

2

Variáveis Principais

As variáveis-chave incluem a idade da pessoa (`person_age`), renda da pessoa (`person_income`), valor do empréstimo (`loan_amnt`) e status do empréstimo (`loan_status`).

3

Distribuição de Status do Empréstimo

Dos 32.576 empréstimos, 25.468 (78,2%) foram pagos (`loan_status` = 0) e 7.108 (21,8%) foram inadimplentes (`loan_status` = 1).



Distribuição de Idade

1

Faixa Etária

O histograma da distribuição de idades revela que a maioria dos tomadores de empréstimo está na faixa etária de 20 a 60 anos, com um pico por volta dos 30 anos.

2

Valores Extremos

Existem alguns valores extremos, com a pessoa mais nova tendo 20 anos e a mais velha 94 anos. No entanto, esses casos são raros e podem representar erros de dados.

3

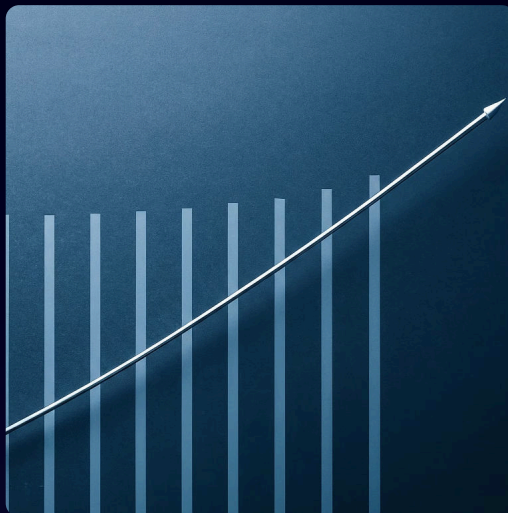
Impacto na Inadimplência

A idade pode ser um fator relevante na determinação do risco de crédito, pois tomadores mais jovens podem ter menor estabilidade financeira, enquanto tomadores mais velhos podem enfrentar desafios de renda na aposentadoria.

Renda dos Tomadores

Distribuição de Renda

O histograma da distribuição de renda mostra uma grande variação, com a maioria dos tomadores tendo renda anual entre R\$ 20.000 e R\$ 100.000. No entanto, existem alguns outliers com rendas extremamente altas, chegando a mais de R\$ 2 milhões por ano.



Renda Mínima e Máxima

A renda mínima no conjunto de dados é de R\$ 4.000 por ano, enquanto a renda máxima atinge impressionantes R\$ 2.039.784 por ano. Essa grande variação pode indicar diferentes perfis de risco de crédito entre os tomadores.

Impacto no Risco de Crédito

Em geral, espera-se que tomadores com rendas mais altas tenham menor risco de inadimplência, pois possuem maior capacidade de pagamento. No entanto, outros fatores como gerenciamento financeiro e obrigações também devem ser considerados.

Valores dos Empréstimos



1

Distribuição dos Valores

O histograma dos valores dos empréstimos revela uma distribuição assimétrica, com a maioria dos empréstimos variando entre R\$ 5.000 e R\$ 20.000. No entanto, existem alguns empréstimos com valores muito altos, chegando a R\$ 35.000.

2

Valor Mínimo e Máximo

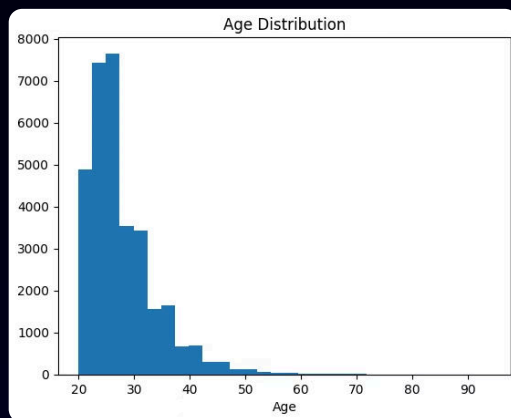
O menor valor de empréstimo no conjunto de dados é de R\$ 500, enquanto o maior valor é de R\$ 35.000. Essa ampla variação pode estar relacionada a diferentes propósitos dos empréstimos, como financiamento de veículos, reformas ou consolidação de dívidas.

3

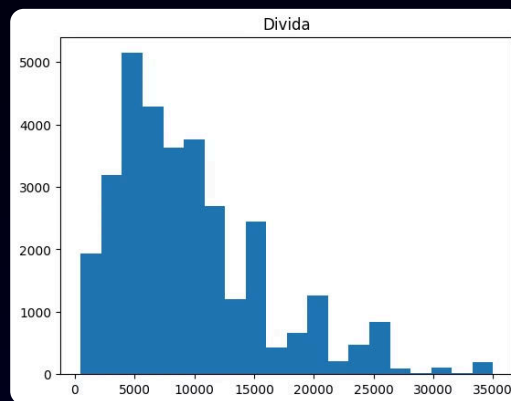
Impacto no Risco de Crédito

Empréstimos de valores mais altos geralmente representam um risco maior para os credores, pois os tomadores podem ter dificuldades em lidar com parcelas maiores. No entanto, outros fatores como renda e histórico de crédito também devem ser considerados na avaliação do risco.

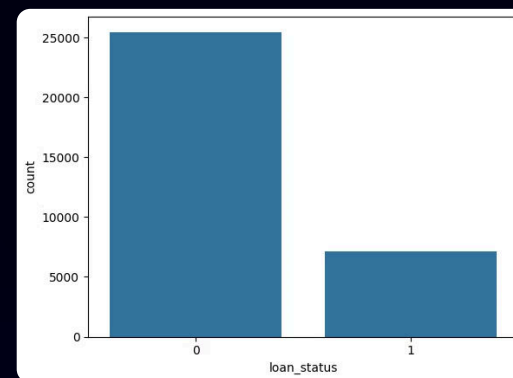
Graficos



Idade

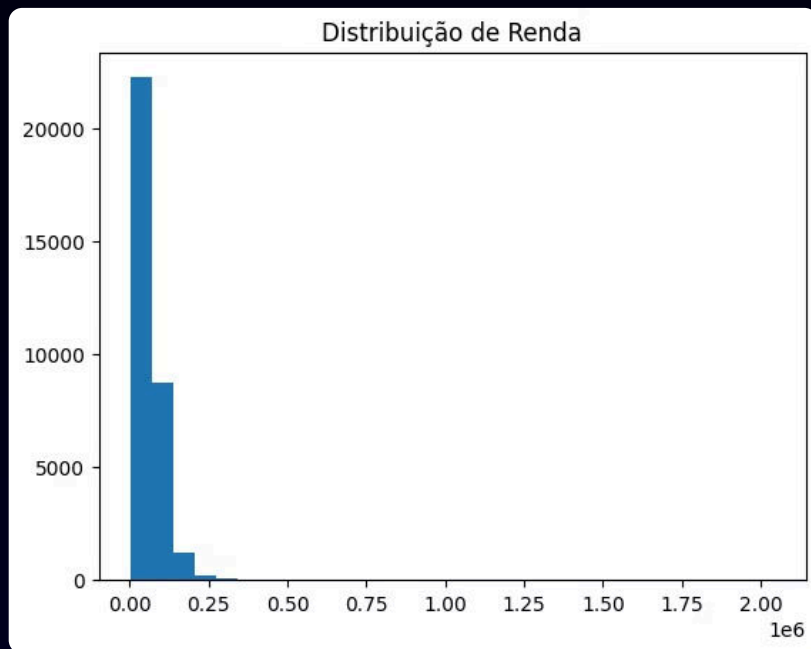


Dividas



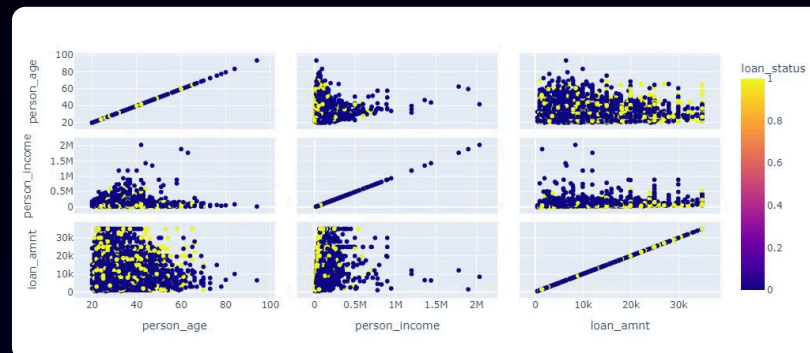
Inadimplência do empréstimo

E mais gráficos...



Renda

Matriz de Dispersão



Preparação dos Dados



Remoção de Outliers

Foram removidos registros com idades negativas ou maiores que 100 anos, pois provavelmente representam erros de dados. Também foram removidos registros com valores extremos de renda e valor do empréstimo.



Padronização de Escalas

As variáveis numéricas foram padronizadas usando a técnica de StandardScaler do scikit-learn, para que todas tenham a mesma escala e evitar que variáveis com valores maiores dominem o modelo.



Separação de Variáveis

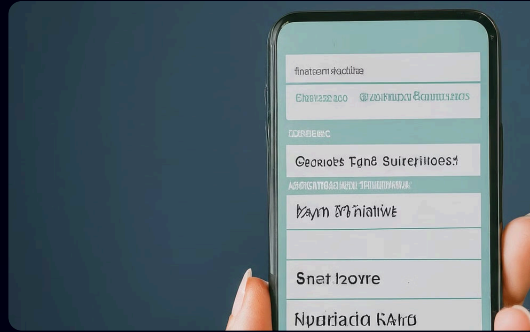
As variáveis preditoras (idade, renda e valor do empréstimo) foram separadas da variável-alvo (status do empréstimo) para preparar os dados para modelagem de aprendizado de máquina.

Modelagem de Risco de Crédito



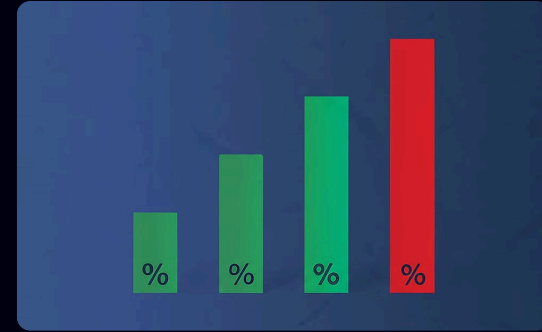
Treinamento de Modelos

Diferentes algoritmos de aprendizado de máquina, como redes neurais, árvores de decisão e regressão logística, podem ser treinados usando os dados preparados para prever o risco de inadimplência de novos empréstimos.



Scorecard de Risco

Os modelos treinados podem ser implementados em um scorecard de risco de crédito, que avalia as informações do solicitante e fornece um score de risco, auxiliando na tomada de decisão de aprovação ou negação do empréstimo.



Monitoramento de Desempenho

O desempenho dos modelos de risco de crédito deve ser monitorado continuamente, comparando as previsões com os resultados reais e ajustando os modelos conforme necessário para manter a precisão e a eficácia.

Impacto e Considerações Éticas

Acesso ao Crédito

Os modelos de risco de crédito desempenham um papel crucial no acesso ao crédito, determinando quem pode ou não obter um empréstimo. É essencial garantir que esses modelos sejam justos e não discriminem indevidamente grupos específicos.

Privacidade de Dados

Os dados utilizados para treinar os modelos de risco de crédito podem conter informações pessoais sensíveis dos solicitantes. É fundamental proteger a privacidade desses dados e garantir que sejam utilizados de forma ética e responsável.

Transparência e Explicabilidade

Os modelos de risco de crédito devem ser transparentes e explicáveis, permitindo que os solicitantes entendam os fatores que influenciam as decisões de aprovação ou negação de empréstimos.

Monitoramento Contínuo

É necessário monitorar continuamente os modelos de risco de crédito para garantir que permaneçam precisos e não desenvolvam vieses indesejados ao longo do tempo, devido a mudanças nas condições econômicas ou outros fatores.

Próximos Passos e Conclusão

Exploração Adicional de Dados

Realizar uma análise mais aprofundada dos dados, investigando outras variáveis potencialmente relevantes, como histórico de crédito, propósito do empréstimo e localização geográfica.

1

2

Treinamento e Validação de Modelos

Treinar e validar diferentes modelos de aprendizado de máquina usando técnicas como validação cruzada e conjuntos de teste separados, a fim de selecionar o modelo mais preciso para prever o risco de inadimplência.

Implementação e Monitoramento

Implementar o modelo selecionado em um sistema de scorecard de risco de crédito e monitorar continuamente seu desempenho, ajustando-o conforme necessário para manter a precisão e a relevância.

3

4

Considerações Éticas e Regulatórias

Garantir que o uso dos modelos de risco de crédito esteja em conformidade com as regulamentações e diretrizes éticas, evitando discriminação e protegendo a privacidade dos dados dos solicitantes.