



PROJETO RISCO RELATIVO

Automatização da Análise de Crédito: Reduzindo riscos e agilizando decisões no banco Super Caja

▼ Ficha Técnica

▼ Objetivo

Desenvolver modelo preditivo de análise de crédito automatizado para o Banco Super Caja, utilizando machine learning e inteligência artificial, com foco em:

- Reduzir tempo de processamento de solicitações de crédito.
- Aumentar precisão e eficiência na tomada de decisões de crédito.
- Mitigar risco de inadimplência e proteger saúde financeira do banco.
- Integrar métricas de pagamentos em atraso para fortalecer modelo.
- Implementar modelo de análise de crédito responsável e transparente.

Benefícios:

- Serviço de crédito mais rápido, preciso e personalizado.
- Redução de tempo de espera para clientes.
- Menor risco de inadimplência e perdas por crédito mal parado.
- Maior solidez financeira e retorno sobre investimentos.
- Relação de confiança e transparência com clientes e stakeholders.

▼ Contexto

No cenário financeiro atual, a redução das taxas de juros tem impulsionado um aumento notável na demanda por crédito no banco Super Caja. Os clientes estão aproveitando as condições favoráveis resultando em um fluxo elevado de solicitações de empréstimo.

Entretanto, essa crescente demanda está sobrecarregando a equipe de análise de crédito, que se encontra presa a um processo manual ineficiente e demorado para avaliar cada solicitação. Este método manual não apenas retarda a velocidade de processamento das solicitações, mas também afeta negativamente a eficácia da equipe. Além disso, a taxa de inadimplência crescente é uma preocupação significativa, pressionando ainda mais os bancos a identificar e mitigar os riscos associados ao crédito.

- **Proposta de Solução: Automação da Análise de Crédito**

Para enfrentar esses desafios, propõe-se a automação do processo de análise de crédito utilizando técnicas avançadas de análise de dados. O principal objetivo desta inovação é melhorar a eficiência, precisão e rapidez na avaliação das solicitações de crédito, permitindo ao banco tomar decisões informadas e reduzir o risco de empréstimos não reembolsáveis.

- **Desenvolvimento de um Score de Crédito**

A análise de dados visa desenvolver um score de crédito que possa classificar os solicitantes em diferentes categorias de risco, com base na probabilidade de inadimplência. Essa classificação permitirá ao banco tomar decisões mais acertadas sobre a concessão de crédito. Além disso, a integração da métrica já existente de pagamentos em atraso no novo sistema automatizado fortalecerá a capacidade do modelo de identificar riscos, contribuindo para a solidez financeira e eficiência operacional do banco Super Caja.

Em resumo, a automação do processo de análise de crédito representa uma solução estratégica para enfrentar a sobrecarga da equipe de análise, melhorar a qualidade das decisões de crédito e mitigar os riscos de inadimplência, fortalecendo a posição do banco no mercado financeiro atual.

▼ Equipe

Ana Guimarães

▼ Ferramentas e Tecnologias

- **Google BigQuery:** Data warehouse onde foi realizado o processamento dos dados em SQL.

- **Google Colab:** Plataforma de desenvolvimento em linguagem de programação Python em Notebooks.
- **Apresentações Google:** Ferramenta para criação e edição da apresentação.
- **Google Looker Studio:** Ferramenta para criação e edição do dashboard.

▼ Processamento e Análises

▼ Preparação da base de dados

▼ Identificar e tratar valores nulo

- tabela `default` : 0 nulos - total de 36.000 registros
- tabela `loans_detail` : 0 nulos - total de 36.000 registros
- tabela `loans_outstanding` : 0 nulos - total de 305.335 registros
- tabela `user_info` : total de registros 36.000 - `null_last_month_salary` : 7199 | `null_last_number_dependents` : 943

Inserido a mediana de `last_month_salary` '5400' nos 7.199 registros nulos e inserido a mediana de `number_dependents` '0' nos 943 registros nulos.

▼ Identificar e tratar valores duplicados

- Não há dados para exibir na tabela `default`
- Não há dados para exibir na tabela `loans_detail`
- Não há dados para exibir na tabela `loans_outstanding`
- Não há dados para exibir na tabela `user_info`

▼ Identificar e gerenciar dados fora do escopo da análise

Afim de aumentar a inclusão financeira e evitar atitudes discriminatórias, a ponderação de diversos fatores éticos, legais e práticos, o uso da variável `sex` foi removida para não tomar decisões discriminatórias em relação a crédito ou outros serviços financeiros. Esta variável somente foi utilizada para visualização gráfica das relações entre gênero e inadimplência.

▼ Identificar e tratar dados inconsistentes em variáveis categóricas

Na tabela `loans_outstanding` foi utilizado o comando LOWER para converter para maiúsculo os valores de `loan_type`, pois existia uma não conformidade, como valores representados por 'real estate', 'REAL ESTATE', 'Real Estate', 'other', 'others', 'Other' e 'OTHER'. Após a modificação esses valores passaram para 'Real Estate' e 'Other'.

▼ Identificar e tratar dados discrepantes em variáveis numéricas

Na variável `last_month_salary` 2 outliers com valor em notação científica de '1.00E+05' foram substituídos pela mediana do salário, nos registros 4931 e 9466, afim de evitar impacto desses dados no manuseio dos dados gerais.

▼ Verificar e alterar o tipo de dados

- O tipo da variável `last_month_salary` foi alterado de FLOAT para INTEGER utilizando o comando CAST.

▼ Criar novas variáveis

- Usando o comando DISTINCT foram criadas 3 novas variáveis `total_loan`, `num_real_estate` e `num_other` para agrupar o total e o tipo de empréstimos.

▼ Unir tabelas

- Através da função LEFT JOIN as tabelas `user_info` e `default` foram unidas pelo `user_id` como correspondência.
- Com o mesmo comando, uma nova tabela foi criada e nomeada como `users_loans_full` para unir as tabelas `user_info_default`, `total_loan` e `loans_detail`.

Ao final da união foram identificados 425 registros nulos.

▼ Construir tabelas auxiliares

- Usando o comando CREATE VIEW foi criada a exibição de dados em `view_fulltable_nulls` para a análise posterior dos 425 registros nulls isolados após o join.

▼ Análise Exploratória

▼ Novas variáveis

- **total_loan**: somatório de todos os empréstimos realizados por `user_id`.
- **num_real_estate**: total de empréstimos de imóveis.
- **num_other**: total de empréstimos de diferentes tipos.

▼ Agrupar e visualizar dados de acordo com variáveis categóricas

- **Idade (age)**: Agrupamento dos clientes por faixa etária.
- **Sexo (sex)**: Distribuição dos clientes por gênero.
- **Salário do último mês (last_month_salary)**: Segmentação dos clientes por intervalos de salário.
- **Número de dependentes (number_dependents)**: Classificação dos clientes conforme o número de dependentes.
- **Indicador de inadimplência (default_flag)**: Identificação dos clientes que estão em inadimplência.
- **Número de vezes que atrasou o pagamento do empréstimo de 30 a 59 dias (number_times_delayed_payment_loan_30_59_days)**: Agrupamento dos clientes pelo número de atrasos de 30 a 59 dias.

- **Número de vezes que atrasou o pagamento do empréstimo de 60 a 89 dias (number_times_delayed_payment_loan_60_89_days):** Agrupamento dos clientes pelo número de atrasos de 60 a 89 dias.
- **Mais de 90 dias em atraso (more_90_days_overdue):** Contagem dos clientes com mais de 90 dias de atraso no pagamento.
- **Uso de linhas de crédito não garantidas por ativos pessoais (using_lines_not_secured_personal_assets):** Análise do uso de linhas de crédito sem garantia de ativos pessoais.
- **Relação dívida/renda (debt_ratio):** Classificação dos clientes por faixas de relação dívida/renda.
- **Total do empréstimo (total_loan):** Segmentação dos clientes por intervalos de quantidade de empréstimos.

Gráficos de barras, linhas e histogramas foram criadas no Looker Studio para visualizar a distribuição das variáveis categóricas descritas acima.

▼ Aplicar medidas de tendência central, medidas de dispersão e ver as distribuições

Focando nas variáveis numéricas contínuas, foi aplicado medidas de tendência central (média, mediana) nas variáveis `age`, `last_month_salary` e `number_dependents` para entender a tendência central, dispersão e distribuição dos dados e identificar tendências temporais, como o comportamento de pagamento em diferentes grupos de renda e de idade.

Gráficos de barras, linhas e histogramas foram criadas no Looker Studio para visualizar a distribuição das variáveis numéricas descritas acima.

▼ Calcular quintis para variáveis de risco relativo

Com o objetivo de identificar os grupos com maior risco de inadimplência e com as variáveis numéricas definidas para avaliar o risco relativo associado a cada cliente (`age`, `last_month_salary`, `debt_ratio`, `using_lines_not_secured_personal_assets`, `total_loan` e `more_90_days_overdue`) foi possível classifica-los em cinco grupos usando com base em idade, salário do último mês, uso de linhas de crédito não garantidas por bens pessoais e total de empréstimo, em relação a `default_flag`, como também foi calculada a correlação entre as mesmas através do comando CORR usando SQL.

A função `NTILE(5) OVER (ORDER BY age)` foi usada para dividir os dados em 5 grupos de 7115 clientes. Os registros foram ordenados antes de serem divididos, e neste caso, os dados estão sendo ordenados pela variável `age` e depois divididos em grupos iguais para realizar as análises em diferentes segmentos de idade. O procedimento também foi aplicado para as variáveis `last_month_salary`, `debt_ratio`, `using_lines_not_secured_personal_assets` e `total_loan`. Na variável `more_90_days_overdue` o processo de agrupamento foi realizado de modo distinto das demais, e os registros foram divididos em apenas 2

grupos. Esta variável foi transformada para valores binários, onde 1 representa se o cliente atrasou o pagamento por mais de 90 dias e 0 quando não houve atraso.

▼ Técnicas de Análise

▼ Calcular Risco Relativo

O cálculo do risco relativo é uma ferramenta importante na análise de riscos, especialmente no contexto financeiro, que auxilia na criação da pontuação de crédito.

Os dados dos cálculos dos quintis foram reunidos em uma tabela, e os riscos relativos foram calculados para cada quintil de cada variável em relação à variável `default_flag`. Os resultados foram unidos em uma única consulta final, onde cada variável foi identificada com seu quintil e o risco relativo correspondente.

- Etapas
 - Calculados os quintis para cada uma das variáveis financeiras.
 - Combinados os resultados dos quintis em uma tabela unificada: `combined_quintiles`.
 - Calculados os riscos relativos de inadimplência para cada quintil de cada variável.

O risco relativo de inadimplência para cada variável financeira foi calculado envolvendo uma contagem do número total de clientes e inadimplentes em cada quintil, o cálculo da taxa média de inadimplência para cada quintil e a comparação da taxa de inadimplência de cada quintil com a taxa média de inadimplência para todos os clientes.

- Resultados combinados em uma consulta final usando `UNION ALL`, onde cada variável passou a ser identificada junto com seu quintil e risco relativo correspondente.

▼ Resultados

- Total de 7115 clientes em cada quintil

Quintil idade	Inadimplentes	Taxa de inadimplência	Risco relativo
1	219	3,08%	1,76
2	169	2,38%	1,36
3	135	1,90%	1,09
4	62	0,87%	0,50
5	37	0,52%	0,30

Quartil salário	Inadimplentes	Taxa de inadimplência	Risco relativo
1	200	2,81%	1,61
2	155	2,18%	1,25
3	102	1,43%	0,82
4	117	1,64%	0,94
5	48	0,67%	0,39

Quartil índice de endividamento	Inadimplentes	Taxa de inadimplência	Risco relativo
1	83	1,17%	0,67
2	123	1,73%	0,99
3	113	1,59%	0,91
4	199	2,80%	1,60
5	104	1,46%	0,84

Quartil limite usado	Inadimplentes	Taxa de inadimplência	Risco relativo
1	8	0,11%	0,06
2	0	0,00%	0,00
3	4	0,06%	0,03
4	52	0,73%	0,42
5	558	7,84%	4,49

Quartil total de empréstimos	Inadimplentes	Taxa de inadimplência	Risco relativo
1	201	2,83%	1,62
2	161	2,26%	1,29
3	81	1,14%	0,65
4	88	1,24%	0,71
5	91	1,28%	0,73

- Para calcular o risco relativo da variável `more_90_days_overdue`, não foram utilizados quintis, em vez disso, a divisão foi feita de forma binária, criando dois grupos distintos:

Grupo atraso superior a 90 dias	Inadimplentes	Taxa de inadimplência	Risco relativo
1	56	0,17%	0,09
2	566	31,87%	18,23

- 1776 clientes com atrasos superior a 90 dias
- 33799 clientes com atrasos inferior a 90 dias

▼ Validar Hipóteses

As validações das hipóteses foram realizadas através dos resultados do cálculo do risco relativo.

- **Hipótese 1: Clientes mais jovens correm um risco maior de não pagamento.**

Hipótese Confirmada: A análise demonstra que as faixas etárias mais jovens apresentam maior risco de inadimplência em relação à média geral.

Faixas etárias com maior risco:

- 1ª faixa etária: 1.76 vezes mais chances de inadimplência
- 2ª faixa etária: 1.36 vezes mais chances de inadimplência

Recomendações:

Explorar as possíveis causas da maior inadimplência entre os jovens (por exemplo, falta de experiência com crédito, instabilidade financeira).

- **Hipótese 2: Pessoas com mais empréstimos ativos correm maior risco de serem maus pagadores.**

Hipótese Refutada: Pessoas com mais empréstimos ativos correm maior risco de serem maus pagadores.

- **Q1 (1-4 empréstimos) e Q2 (4-7 empréstimos):**

Apresentam os maiores riscos relativos de inadimplência, com 1.62 e 1.29 vezes mais chances, respectivamente.

- **Grupos com mais de 7 empréstimos:**

Possuem menor risco relativo de inadimplência, contrariando a hipótese inicial.

- **Hipótese 3: Pessoas que atrasaram seus pagamentos por mais de 90 dias correm maior risco de serem maus pagadores.**

Hipótese Confirmada: Pessoas que atrasaram seus pagamentos por mais de 90 dias correm maior risco de serem maus pagadores.

Risco relativo da variável `more_90_days_overdue` :

- Indica que pessoas com atrasos acima de 90 dias têm 18 vezes mais chances de se tornarem inadimplentes.
- Esse valor elevado comprova a relação entre o atraso superior a 90 dias e o maior risco de inadimplência.

Recomendações:

Implementar medidas para reduzir o índice de atrasos superiores a 90 dias, como:

- Cobrança mais eficaz;
- Negociação de acordos de pagamento;
- Oferta de opções de refinanciamento.

▼ Aplicar Segmentação - Construção do modelo de score de crédito

Com os quintis de risco estabelecidos, foram identificados os grupos com maior probabilidade de inadimplência. Adicionalmente, foram definidos os limites da pontuação de crédito para assegurar a classificação adequada dos clientes com maior risco.

- Etapas
 - Cálculo do risco relativo.
 - Segmentação por risco de cliente.
 - Segmentação em quintis por perfil de risco.
 - Definição de pontos de corte para pontuação.
 - Cálculo da pontuação com base no risco e quintis.

A consulta calculou os quintis para diferentes variáveis financeiras e combinou esses resultados para cada cliente. Com base nesses quintis, foram criadas variáveis de risco binárias que são somadas para fornecer uma pontuação de risco total (**score**). Esta pontuação é utilizada para classificar os clientes como bons ou maus pagadores, facilitando a análise de risco e a tomada de decisão.

- Novas variáveis
 - **age_risk (idade)**: Clientes nos dois quintis mais baixos de idade (abaixo do 3º) são considerados de maior risco.
 - **salary_risk (salário)**: Clientes nos dois quintis mais baixos de salário (abaixo do 3º) são considerados de maior risco.
 - **debt_ratio_risk (índice de endividamento)**: Somente clientes no 4º quintil (mais alto) de endividamento são considerados de maior risco.
 - **using_lines_risk (linhas de crédito)**: Clientes no quintil mais alto (5º) de linhas de crédito não garantidas são considerados de maior risco.
 - **total_loan_risk (total de empréstimos)**: Usuários nos dois quintis mais baixos de total de empréstimos (abaixo do 3º) são considerados de maior risco.

- `more_90_days_risk` (**atraso superior a 90 dias**): Nessa variável a divisão de grupos baseou-se em uma classificação binária: alto risco (atrasos > 90 dias) e baixo risco (atrasos <= 90 dias).
- `risk_score` (**pontuação de risco total**): A pontuação total de risco é a soma das variáveis de risco binárias

A criação das novas variáveis e a análise dos quintis de risco representam um avanço significativo na avaliação do risco de inadimplência. Esses métodos servem de base para a construção de modelos preditivos mais precisos, contribuindo para tomadas de decisões mais assertivas e a otimização dos resultados financeiros do banco.

▼ Matriz de Confusão - Validação da classificação

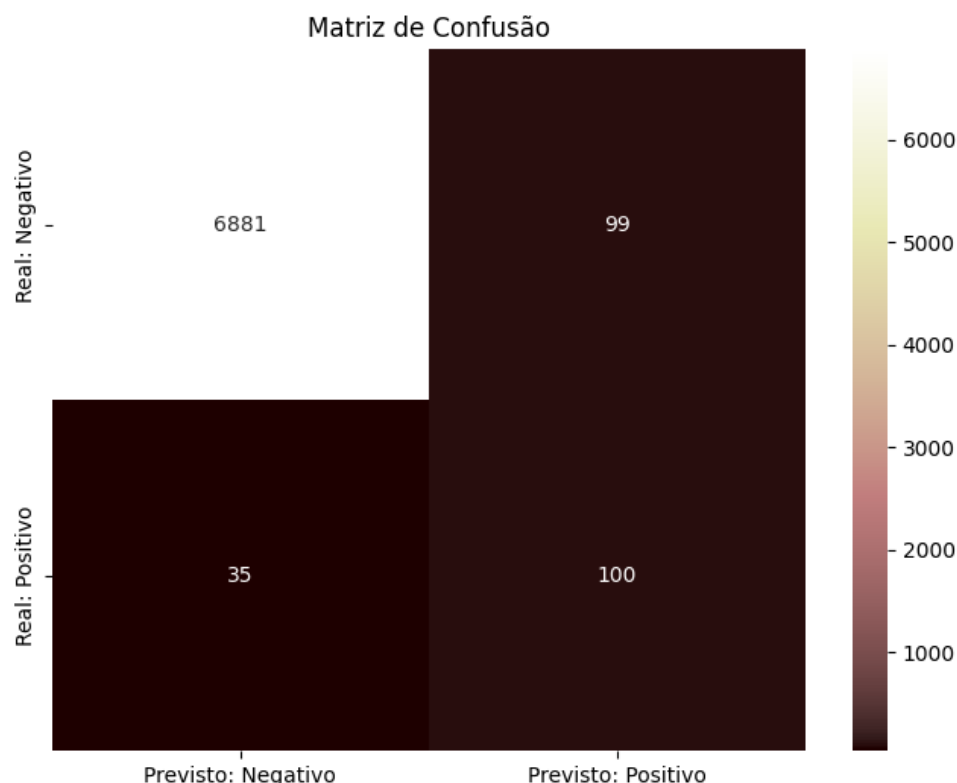
Para validar a classificação de risco de crédito dos clientes, foi utilizado uma matriz de confusão para calcular as métricas de avaliação da eficácia do modelo preditivo. A matriz de confusão é uma ferramenta essencial na avaliação do desempenho de modelos de classificação em aprendizado de máquina, permitindo visualizar a performance do modelo na previsão das classes de um conjunto de dados.

Foi gerada uma variável binária (`risk_score_binary`) para classificar os clientes em dois grupos: baixo risco (0) e alto risco (1). Um ponto de corte foi definido para essa classificação e os dados foram divididos em conjuntos de treino e teste, com 80% dos dados destinados ao treino e 20% ao teste. O cálculo do risco é realizado com base nas variáveis fornecidas, e o resultado é armazenado na coluna `risk_score`.

▼ Resultados do melhor modelo alcançado

Os resultados mostram uma melhoria significativa na sensibilidade (recall) após a aplicação da técnica de sobreamostragem com SMOTE:

1. Verdadeiros Negativos: 6881
2. Falsos Positivos: 99
3. Falsos Negativos: 35
4. Verdadeiros Positivos: 100



A matriz de confusão e as métricas de avaliação indicam a performance do modelo de classificação.

Métricas de Avaliação:

- Exatidão (Accuracy): 0.981
- Precisão (Precision): 0.503
- Sensibilidade (Recall): 0.741
- F1-score: 0.599

A exatidão está alta, indicando que o modelo está acertando a maioria das previsões. A precisão diminuiu um pouco, comparado ao primeiro modelo, mas ainda é significativa, significando que, quando o modelo prevê um inadimplente, ele acerta 50.3% das vezes. A sensibilidade melhorou bastante, indicando que o modelo está capturando 74.1% dos inadimplentes reais. O F1-score também melhorou, indicando um equilíbrio entre precisão e sensibilidade.

Conclusão: O balanceamento das classes com a técnica de sobreamostragem usando SMOTE foi eficaz em melhorar a sensibilidade do modelo, permitindo que ele capture mais casos positivos. Isso é crucial para problemas onde a detecção de casos positivos é de alta importância, como no atual caso de detecção de inadimplência em análises de risco de crédito.

Os modelos gerados estão disponíveis no arquivo `analise_riscocredito_matrizconfusao_regressao.ipynb`

▼ Regressão Logística - Automação do processo de análise de crédito

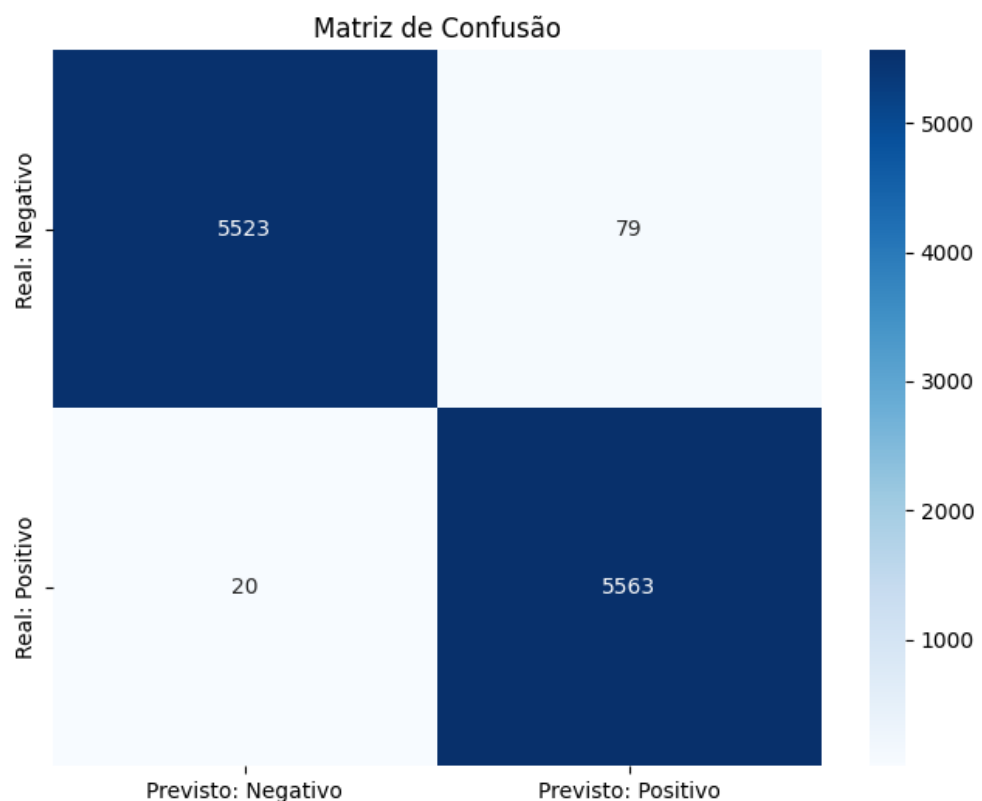
A regressão logística foi utilizada para refinar a matriz de confusão e aprimorar a precisão da classificação de risco.

Também foi feita uma segmentação do risco usando a cláusula `CASE` para categorizar o risco com base no valor da pontuação em `risk_score`:

- Se a pontuação estiver entre 0 e 1, o cliente é classificado como **"Menor Risco de Inadimplência"**.
- Se a pontuação estiver entre 2 e 3, o cliente é classificado como **"Risco Moderado a Baixo de Inadimplência"**.
- Caso contrário, o cliente é classificado como **"Maior Risco de Inadimplência"**.

▼ Resultados do melhor modelo alcançado

Otimização do modelo RandomForestClassifier para classificação de Risco de Crédito com GridSearchCV, uma técnica que busca os melhores parâmetros do modelo para melhor desempenho.



Métricas de Avaliação:

- Exatidão (Accuracy): 0.991

- Precisão (Precision): 0.986
- Sensibilidade (Recall): 0.996
- F1-score: 0.991
- AUC-ROC: 1.000
- Log-Loss: 0.031

Os resultados da regressão logística com ajuste de pesos e validação cruzada foram bastante significativos.

- Aproximadamente 99,1% dos casos foram classificados corretamente pelo modelo. Isso indica uma excelente capacidade de previsão.
- A precisão é de 98,6%, o que significa que quando o modelo prevê que um cliente será inadimplente, ele está correto em 98,6% das vezes.
- A sensibilidade é de 99,6%, indicando que o modelo é capaz de detectar 99,6% dos casos reais de inadimplência.
- O F1-score, que é a média harmônica entre precisão e recall, é de 99,1%. Isso indica um bom equilíbrio entre precisão e recall.
- A área sob a curva ROC é 1, o que indica um modelo perfeito de classificação.
- A log-loss é 0,031, o que é um valor muito baixo e indica que o modelo está fazendo previsões com alta confiança.

Conclusão: Esses últimos resultados sugerem que o modelo é altamente eficaz na classificação de clientes em relação à inadimplência. Ele tem uma precisão muito alta, bem como uma alta sensibilidade, o que é crucial para um modelo de análise de risco de crédito.

Os modelos gerados estão disponíveis no arquivo `analise_riscocredito_matrizconfusao_regressao.ipynb`

▼ Resultados e Conclusões

O banco Super Caja vivenciou um cenário que sobrecarregou a equipe de análise de crédito, que ainda utilizava um processo manual e ineficiente para avaliar cada solicitação. Essa ineficiência não apenas retardava o tempo de resposta, como também comprometia a acurácia das análises e aumentava o risco de inadimplência. Diante desse desafio do crescimento impulsionado pela queda das taxas de juros, resultando em um aumento exponencial na demanda por crédito, o Super Caja buscou uma solução inovadora: a automação da análise de crédito.

Através de um estudo aprofundado em sua base de dados com 35.575 registros, foi possível identificar um perfil de cliente com alto risco de inadimplência:

- Faixa etária: Entre 21 e 48 anos
- Número de empréstimos: Entre 1 e 7
- Histórico de inadimplência: Atrasos superiores a 90 dias

Com base nesse perfil, foi desenvolvido um modelo de inteligência artificial capaz de analisar automaticamente cada solicitação de crédito, considerando diversos fatores como histórico financeiro e comportamento de pagamento. Essa solução proporcionará diversos benefícios ao banco:

- Maior agilidade: A automação propõe uma redução drástica do tempo de análise, permitindo que o banco responda às solicitações de crédito de forma muito mais rápida, proporcionando uma melhor experiência para os clientes.
- Maior precisão: O modelo de inteligência artificial será capaz de analisar um conjunto de dados muito maior e com mais precisão do que os analistas humanos, resultando em decisões de crédito mais assertivas e na redução da taxa de inadimplência.
- Maior eficiência: A automação pode liberar a equipe de análise de crédito para se concentrar em tarefas mais complexas que exigem julgamento humano, otimizando o uso dos recursos do banco.
- Redução de custos: A automação também deve levar a uma redução natural e significativa dos custos operacionais, pois eliminará a necessidade de mão de obra manual para a análise de crédito.

Conclusão:

A automação da análise de crédito será um passo crucial na transformação digital do Super Caja. Através dessa solução inovadora, o banco poderá ser capaz de aumentar a agilidade, precisão e eficiência do processo de concessão de crédito, além de reduzir custos e riscos. Essa iniciativa demonstra o compromisso do Super Caja em oferecer serviços de alta qualidade e atender às necessidades de seus clientes de forma cada vez mais eficiente e segura.

Recomendações:

É importante que o Super Caja continue monitorando o desempenho do modelo de inteligência artificial e faça ajustes conforme necessário para garantir sua precisão e eficácia.

O banco também pode explorar outras aplicações da inteligência artificial para aprimorar ainda mais seus serviços, como na prevenção de fraudes ou na oferta de produtos personalizados aos clientes.

É fundamental que o Super Caja comunique os benefícios da automação da análise de crédito aos seus clientes para que eles possam entender como essa iniciativa contribui para um atendimento mais rápido e eficiente.

▼ Limitações/Próximos Passos

O banco de dados possui certas limitações que podem ser melhoradas para uma análise robusta e direcionada a concessão de crédito.

A ausência de métricas temporais na análise automatizada de crédito compromete a qualidade da avaliação do risco e impede a tomada de decisões estratégicas para otimizar a gestão de crédito e reduzir perdas. Para uma análise completa e eficaz, a inclusão de métricas como o período do empréstimo e a data de inadimplência são fundamentais.

A falta da variável data impediu a análise da inadimplência ao longo do tempo. Sem essa informação, é impossível observar como a inadimplência muda ao longo de meses ou anos, limitando a compreensão da dinâmica do problema e dificultando a tomada de decisões precisas. A inclusão da data de empréstimo permitiria analisar tendências, como sazonalidade ou o impacto de mudanças nas políticas de crédito. Com essa informação, seria possível identificar padrões e tomar medidas para reduzir a inadimplência de forma mais eficaz.

A análise da inadimplência enfrentou mais outro desafio crucial: a ausência de métricas de recuperação. Sem dados sobre a taxa de recuperação (percentual de inadimplências recuperadas em relação ao total), a compreensão completa do cenário fica comprometida, limitando a capacidade de tomar decisões assertivas para reduzir a inadimplência e otimizar a avaliação de risco e a concessão de crédito.

Impacto da Falta de Métricas Temporais:

- Visão limitada da inadimplência
- Dificuldade na segmentação de clientes
- Ineficiência na definição de prazos
- Estratégias de cobrança ineficazes

Recomendações:

- Implementar mecanismos para capturar e integrar dados de período do empréstimo e data de inadimplência nos sistemas de análise de crédito.
- Desenvolver modelos de análise que considerem a relação entre o período do empréstimo e a taxa de inadimplência.
- Promover a cultura data-driven na tomada de decisões de crédito.

Mais dados sobre o comportamento de consumo dos clientes facilitará uma atuação mais efetiva, direcionada e voltada para a experiência do cliente.

Ao superar essas limitações, o banco pode aprimorar significativamente a efetividade da análise automatizada de crédito, otimizando a gestão de risco e

impulsionando sua saúde financeira.

▼ Links de interesse

Github

- https://github.com/anacjuriti/projeto_risco_relativo

Dashboard

- <https://lookerstudio.google.com/reporting/d3fcf3ed-a5b1-4424-b321-e5336f8d04db>

Google Colab

- https://colab.research.google.com/drive/1dBtsGAzf2z-DdiTdlqVSGwECj2nD_rhw?usp=sharing

Apresentação de Slides

- https://docs.google.com/presentation/d/1svbVHFDL9oGB28rXs-vUz3vUwR8rvF8CJHmt_nP8loA/edit?usp=sharing

Vídeo de Apresentação da Análise

- <https://www.loom.com/share/c3110791bca941c9a4e3808369b005db?sid=8de23c85-2fca-4167-8ac6-3bd4add12382>
-