

Patrick Ferreira dos Santos

Uso de técnicas de machine learning para análise de risco de crédito

Brasil

2022, v-1.9.7

Patrick Ferreira dos Santos

Uso de técnicas de machine learning para análise de risco de crédito

Dissertação apresentada ao Curso de Mestrado Profissional em Economia, Universidade de Brasília, como requisito parcial para a obtenção do título de Mestre em Economia

Universidade de Brasília - UnB

Faculdade de Administração Contabilidade e Economia - FACE

Departamento de Economia - ECO

Programa de Pós-Graduação

Orientador: Herbert Kimura

Brasil

2022, v-1.9.7

Patrick Ferreira dos Santos

Uso de técnicas de machine learning para análise de risco de crédito/ Patrick Ferreira dos Santos. – Brasil, 2022, v-1.9.7-55p. : il. (algumas color.) ; 30 cm.

Orientador: Herbert Kimura

Dissertação (Mestrado) – Universidade de Brasília - UnB
Faculdade de Administração Contabilidade e Economia - FACE
Departamento de Economia - ECO
Programa de Pós-Graduação, 2022, v-1.9.7.

1. Aprendizado de Máquina. 2. Risco de Crédito. 3. Cheque Especial. II. Universidade de Brasília. III. Faculdade de Administração, Contabilidade e Economia - FACE. IV. Departamento de Economia IV. Uso de técnicas de machine learning para análise de risco de crédito

Patrick Ferreira dos Santos

Uso de técnicas de machine learning para análise de risco de crédito

Dissertação apresentada ao Curso de Mestrado Profissional em Economia, Universidade de Brasília, como requisito parcial para a obtenção do título de Mestre em Economia

Trabalho aprovado. Brasil, 15 de junho de 2022:

Herbert Kimura
Orientador

Leonardo Fernando Cruz Basso
Convidado 1

João Gabriel de Moraes Souza
Convidado 2

Brasil
2022, v-1.9.7

*Dedico este trabalho aos meus filhos Pedro e Helena
e à minha esposa Sandra, por serem sempre o melhor do meu dia
e à minha falecida mãe Maria de Lourdes por ensinar o valor da educação.*

Agradecimentos

Agradeço em primeiro lugar à Deus e ao seu Amor infinito por mim, à minha família que soube entender e suportar minha necessária ausência nos estudos, e ao professor e orientador Herbert Kimura pela disponibilidade e atenção em me guiar.

Também agradeço aos colegas de curso ao vencermos juntos esse desafio, em especial ao nosso grupo, Francisco Almeida, Juelline Schelci e Lemonier Barbosa, pelo apoio em todos os momentos.

*“Comece fazendo o que é necessário,
depois o que é possível,
e de repente você estará fazendo o impossível.
(São Francisco de Assis)*

Resumo

Este trabalho objetivou identificar um modelo preditivo de classificação de risco de crédito em operações comerciais de cheque especial para pessoa física que possibilite o devido gerenciamento de riscos e mensuração mais precisa das situações de *default* proporcionando a evolução financeira do credor. A variável *target* considera o cliente “mau” caso ultrapasse 90 dias de atraso em algum mês, durante os 12 primeiros meses após a contratação, caso contrário, o cliente será considerado “bom”. Foram utilizados os algoritmos de *Machine Learning* Regressão Logística, *k-Nearest Neighbors* (k-NN), *Support Vector Machines* (SVM), Árvore de Decisão, Floresta Aleatória e *Gradient Boosting*. Para validação dos modelos fizemos uso das métricas Precisão, *Recall*, Especificidade, *F1-Score*, Média Geométrica, IBA (Index Balanced Accuracy) e a curva ROC (AUC). Os resultados apontaram melhor desempenho para o modelo Floresta Aleatória podendo ser incremento de resultado para instituições financeiras. O que aponta melhor performance de modelos de computação inteligente sobre modelos estatísticos tradicionais como Regressão Logística. Além disso, não foram encontrados outros trabalhos sobre a temática de cheque especial.

Palavras-chave: Aprendizagem de Máquinas, Risco de Crédito, *Machine Learning*, Cheque Especial, Floresta Aleatória, k-NN, *Support Vector Machines*, Árvore de Decisão, *Gradient Boosting*

Abstract

This study aimed to identify a predictive model of credit risk classification in commercial operations of overdraft for individuals that allows proper risk management and more accurate measurement of *default* situations, providing the financial evolution of the creditor. The *target* variable considers the customer “bad” if he exceeds 90 days of delay in any month, during the first 12 months after hiring, otherwise, the customer will be considered “good”. The algorithms of *Machine Learning* as Logistic Regression, *k-Nearest Neighbors* (k-NN), *Support Vector Machines* (SVM), Decision Tree, Random Forest and *Gradient Boosting* were used. To validate the models, we used the metrics Precision, Recall, Specificity, *F1-Score*, Geometric Mean and IBA (Index Balanced Accuracy) and the ROC curve (AUC). The results showed better performance for the Random Forest model, which could be an increase in results for financial institutions. This points to better performance of intelligent computing models over traditional statistical models such as Logistic Regression. In addition, no other works on the subject of overdraft were found.

Keywords: *Machine Learning*, Credit Risk, Overdraft, Random Forest, k-NN, *Support Vector Machines*, Decision Tree, *Gradient Boosting* .

Lista de ilustrações

| | |
|---|----|
| Figura 1 – Percentual do crédito em relação ao PIB para setor privado por bancos | 26 |
| Figura 2 – Percentual do crédito em relação ao PIB para setor privado | 26 |

Lista de tabelas

| | |
|---|----|
| Tabela 1 – Números da amostra | 34 |
| Tabela 2 – Variáveis | 36 |
| Tabela 3 – Matriz de Confusão | 41 |
| Tabela 4 – Resultados dos modelos | 45 |
| Tabela 5 – Resultados da classe <i>target</i> | 46 |

Lista de abreviaturas e siglas

| | |
|-------|--|
| ML | Machine Learnig |
| kNN | k-Nearest Neighbors |
| SVM | Support Vector Machines |
| IBA | Index Balanced Accuracy |
| AUC | Area under the ROC Curve |
| DRL | Deep Reinforcement Learning |
| BACEN | Banco Central do Brasil |
| BCB | Banco Central do Brasil |
| IPEA | Instituto de Pesquisa Econômica Aplicada |
| CMN | Conselho Monetário Nacional |
| SCR | Sistema de Informações de Crédito |
| SMOTE | Synthetic Minority Over-sampling Technique |

Lista de símbolos

α Letra grega Alfa

β Letra grega Beta

Sumário

| | | |
|------------|--|-----------|
| 1 | INTRODUÇÃO | 25 |
| 2 | REFERENCIAL TEÓRICO | 29 |
| 3 | DADOS E MÉTODOS | 33 |
| 3.1 | Base de Dados | 33 |
| 3.1.1 | Seleção de Amostras | 34 |
| 3.1.2 | Tratamento das Variáveis | 35 |
| 3.1.3 | Balanceamento da Base | 35 |
| 3.1.4 | Validação do Modelo | 36 |
| 3.2 | Metodologias de <i>Machine Learning</i> | 37 |
| 3.2.1 | Regressão Logística | 38 |
| 3.2.2 | Support Vector Machines | 38 |
| 3.2.3 | k-Nearest Neighbors | 38 |
| 3.2.4 | Árvore de Decisão | 39 |
| 3.2.5 | Gradient Boosting | 39 |
| 3.2.6 | Floresta Aleatória | 40 |
| 3.2.7 | Medidas de Avaliação de <i>ML</i> | 40 |
| 3.2.7.1 | Matriz de Confusão | 40 |
| 3.2.7.2 | Precisão | 41 |
| 3.2.7.3 | Recall | 41 |
| 3.2.7.4 | Especificidade | 42 |
| 3.2.7.5 | F1-Score | 42 |
| 3.2.7.6 | Média Geométrica | 42 |
| 3.2.7.7 | Index Balanced Accuracy | 42 |
| 3.2.7.8 | Curva AUC | 43 |
| 4 | RESULTADOS | 45 |
| 5 | CONCLUSÕES | 49 |
| | REFERÊNCIAS | 53 |

1 Introdução

A atividade bancária é de fundamental importância no desenvolvimento da economia mundial. A intermediação financeira, função típica dos bancos, permite que o fluxo dos recursos financeiros seja mais fluido entre agentes superavitários e deficitários no mercado, melhorando a alocação de capital. E dentro dessa perspectiva o risco de crédito é seguramente o risco de maior materialidade e tradicionalmente o mais explorado e discutido.

O risco de crédito é a perda potencial de um banco sofrer se uma contraparte não cumprir suas obrigações de acordo com os termos acordados. É o maior risco que a maioria dos bancos enfrenta e surge da possibilidade de que os empréstimos ou títulos detidos por um banco não sejam reembolsados parcial ou totalmente (APOSTOLIK; DONOHUE; WENT, 2009).

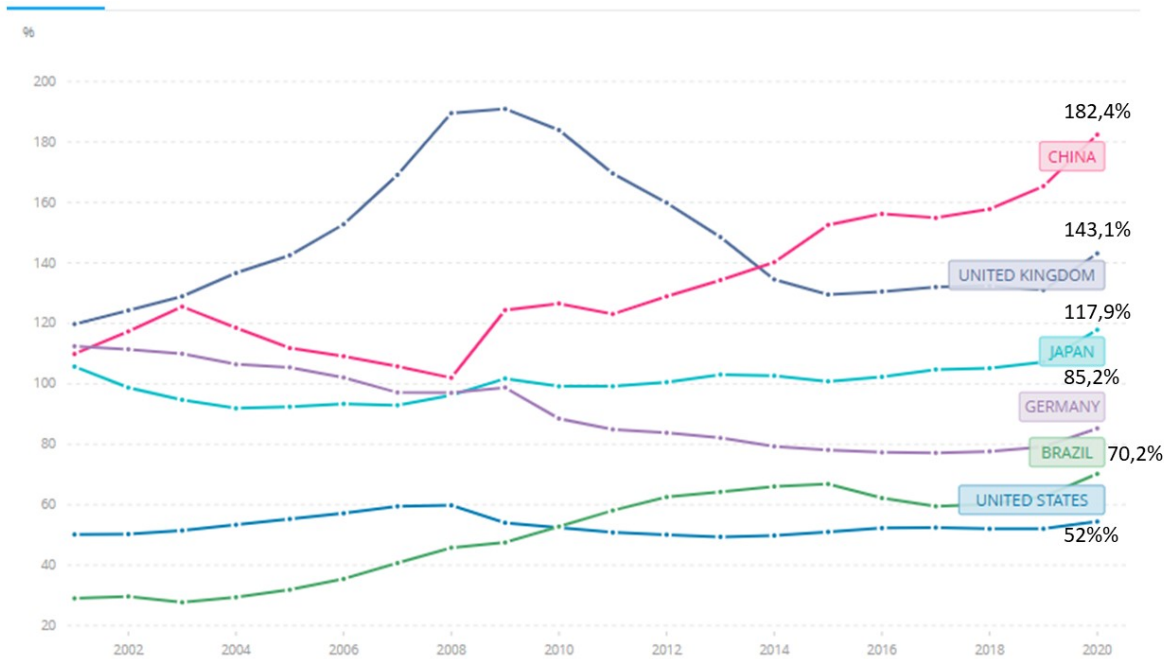
Calcular o crédito como proporção do PIB indica a importância do crédito relativamente ao tamanho da economia, permitindo comparações entre os países. O Brasil ainda não atingiu os patamares dos países desenvolvidos na relação entre o crédito e PIB. Na Figura 1 comparamos a relação crédito originado de bancos e PIB entre as cinco maiores economias do mundo e o Brasil, conforme dados do Banco Mundial em 2020 (BANK et al., 2021).

Na análise da Figura 1 o Brasil, com percentual de 70,2% na relação entre crédito e PIB, situa-se apenas acima dos Estados Unidos que possuem 52% quando o crédito é originado dos bancos. Porém, a situação é alterada quando outros fornecedores de crédito são incluídos, como por exemplo, empresas de arrendamento mercantil, sociedades de seguros, fundos de pensão e empresas de câmbio, conforme Figura 2.

Nesse caso são os americanos que assumem o topo com 215,9% e permanece o Brasil no mesmo patamar, porém agora em último, demonstrando que não foi alcançado o nível dos países desenvolvidos na relação crédito e PIB.

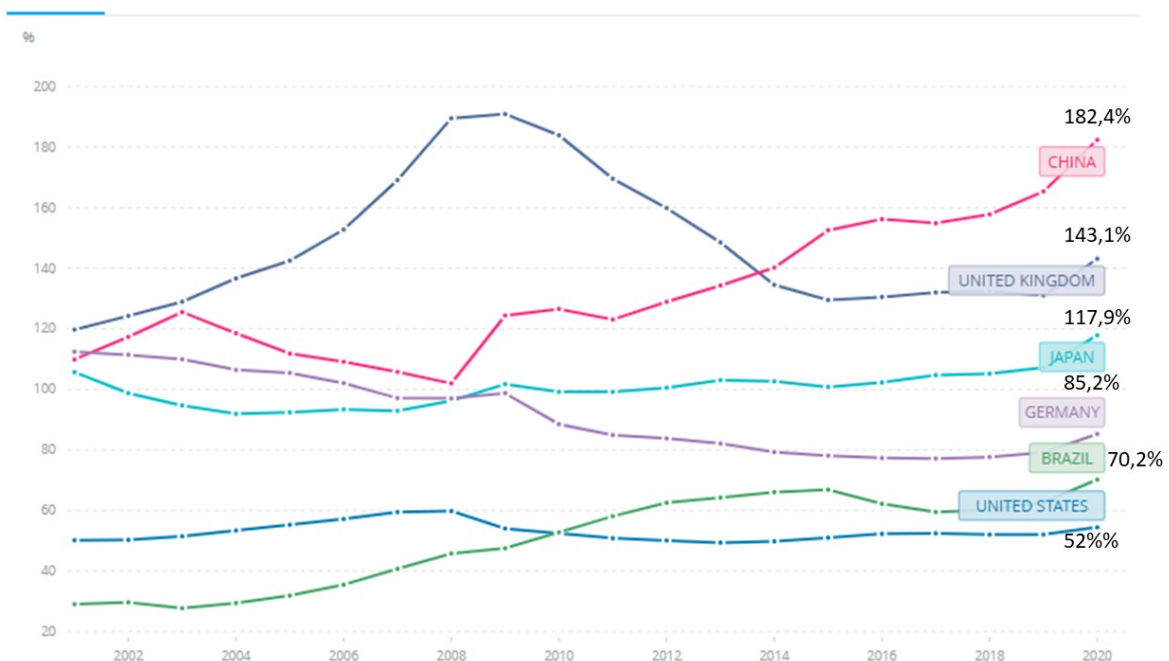
O cheque especial é umas das mais conhecidas e habituais operações de crédito no Brasil, caracterizado como um híbrido de produto de crédito com serviço financeiro, estando associado a uma conta bancária. Pode ser definido como um contrato de concessão de limite de crédito rotativo para cobertura de saldo devedor em conta de depósitos à vista, em geral, sem garantia, e que se destina a uso em caráter emergencial (BCB - Banco Central do Brasil, 2019).

Este produto bancário sofreu recente intervenção regulatória no Brasil com a Resolução CMN 4.765 de 2019 (BRASIL, 2019), pois, em razão de falhas de mercado que

Figura 1 – Percentual do crédito em relação ao PIB para setor privado por **bancos**

Fonte: Banco Mundial, página eletrônica (2022).

Figura 2 – Percentual do crédito em relação ao PIB para setor privado



Fonte: Banco Mundial, página eletrônica (2022).

impedem que a concorrência atinja todo o seu potencial benéfico para os consumidores, as taxas de juros remuneratórios cobradas sobre o valor utilizado do cheque especial

foram limitadas a, no máximo, 8% (oito por cento) ao mês para pessoas naturais e microempreendedores individuais. Em dezembro de 2019, antes dos efeitos da Resolução terem efeito, a taxa média de juros do cheque especial alcançou 272,7% a.a., conforme Relatório de Economia Brasileira de 2020 (BCB - Banco Central do Brasil, 2020).

Observando o Sistema Gerenciador de Séries Temporais do Banco Central do Brasil (BCB - Banco Central do Brasil, 2022b) o saldo da carteira de crédito de cheque especial para pessoas físicas alcançou o valor de R\$ 22,811 bilhões em dezembro de 2021, o que representou 5,96% do saldo da carteira de crédito rotativo com recursos livres para pessoas físicas e ainda 1,51% do saldo da carteira de crédito total com recursos livres para pessoas físicas.

Em vista da materialidade dos valores envolvidos e do impacto da inadimplência do crédito no balanço dos bancos e até na economia, diversos pesquisadores procuraram novas aplicações, cálculos ou ainda abordagens para o risco de crédito. Porém, não foi identificado trabalho similar com uso de dados de cheque especial.

Segundo [Moscatto, Picariello e Sperli \(2021\)](#) a ampla utilização de serviços financeiros tem focado a atenção de pesquisadores na gestão de risco de crédito para o desenvolvimento de modelos visando, por um lado, reduzir os riscos financeiros e, por outro, aumentar os lucros relacionados.

Por sua vez, [Vieira et al. \(2019\)](#) referencia que o crédito tem uma importância fundamental para o desenvolvimento da economia, uma vez que provê os recursos financeiros para que consumidores possam realizar seus projetos e adquirir bens.

Em complemento [Leo, Sharma e Madduletya \(2019\)](#) sentencia que o risco de crédito tem sido tradicionalmente o maior risco enfrentado pelos bancos e, geralmente, aquele que exige mais capital.

Diante desse cenário, este estudo tem por objetivo identificar um modelo de classificação de risco de crédito que permita predição em operações comerciais rotativas de cheque especial para pessoa física possibilitando o devido gerenciamento de riscos e mensuração mais precisa das situações de *default* proporcionando a evolução financeira do credor.

Nossa pesquisa está construída em 4 capítulos, sendo o segundo formado pelo arcabouço teórico das pesquisas recentes sobre o tema *Machine Learning* para risco de crédito com o detalhamento das observações encontradas. No capítulo 3, falamos da base de dados disponível e das metodologias empregadas, assim são tratados os meios de seleção de amostra, tratamento das variáveis, balanceamento de base, validação dos modelos, metodologias de *ML* e medidas de acurácia. No capítulo 4 são apresentados os resultados encontrados após treino e teste da base e qual métrica destacou-se no desempenho, e finalmente no capítulo 5 concluímos o trabalho com sugestão de trabalhos futuros.

2 Referencial Teórico

Neste capítulo nós revisamos a literatura disponível sobre modelos de previsão de default em risco de crédito. O uso de inteligência computacional serviu de impulso aos modelos preditivos mais acurados alinhado ao crescente número de dados disponíveis de clientes e de mercado.

Pesquisas elaboradas nos últimos anos, em especial após a crise financeira de 2008-2009 ([MOSCATO; PICARIELLO; SPERLI, 2021](#)), concentraram-se na pontuação do risco de crédito, que tem crescido cada vez mais em importância, tornando-se um meio crítico na gestão do risco de crédito.

Modelos de pontuação de crédito, numa visão geral, estão relacionados a três métodos: (i) julgamento da experiência do especialista; (ii) análises estatísticas; e (iii) uso de algoritmos de computação inteligente ([WANG; JIA; TIAN, 2022](#)).

Enquanto no primeiro método destacavam-se os 5 C's do crédito (caráter, capacidade, capital, condições e colateral); no segundo, a regressão logística e a análise discriminante são mais utilizadas.

Os problemas encontrados nesses métodos referem-se a fadiga, erro de julgamento e resposta lenta sob forte pressão de trabalho, para o caso de especialistas.

A metodologia das análises estatísticas geralmente assume que as variáveis previstas devem ser lineares e normais e é difícil para os modelos atingirem um bom desempenho se essas suposições forem violadas ([WANG; JIA; TIAN, 2022](#)).

A fim de superar as deficiências do método de análise estatística, os estudos começaram a utilizar o método de computação inteligente para construir modelos de pontuação de crédito de clientes.

O objetivo de qualquer modelo de aprendizado de máquina, segundo [Khandani, Kim e Lo \(2010\)](#), é a identificação de relacionamentos estatisticamente confiáveis entre certos recursos dos dados de entrada e a variável ou resultado de destino.

Particularmente para a indústria bancária, machine learning pode permitir a construção de modelos de risco mais precisos, identificando padrões complexos e não lineares em grandes conjuntos de dados ([LEO; SHARMA; MADDULETYA, 2019](#)).

O Acordo de Capital de Basileia II exige que as instituições financeiras estimem, respectivamente, a probabilidade de inadimplência (PD), a exposição à inadimplência (EAD) e a perda por inadimplência (LGD). Modelagens de PD são especialmente bem pesquisadas e continuam a atrair muita atenção ([LESSMANN et al., 2015](#)).

Os métodos mais procurados para desenvolver modelos de PD são a classificação e a análise de sobrevivência. Sendo o segundo utilizado para estimar não apenas se, mas também quando um cliente torna-se inadimplente.

No primeiro caso, do tipo análise de classificação, representa a abordagem clássica e os benefícios de uma variedade incomparável de métodos de modelagem.

Diversos algoritmos podem ser utilizados em *machine learning* para risco de crédito, a exemplo de *Support Vector Machines* (SVM), árvore de decisão, floresta aleatória e redes neurais artificiais para os casos de modelos supervisionados.

Quanto aos problemas que não possuem um determinado rótulo (modelo não-supervisionado) são explorados algoritmos de agrupamento. O método mais utilizado no setor financeiro é o k-Means (MACHADO; KARRAY, 2022).

Recentemente, pesquisadores como Bao, Lianju e Yue (2019) concentraram-se na estratégia de conjunto (*ensemble*) para integrar diferentes modelos de ML para pontuação de crédito, e uma das principais estratégias de ensemble é tomar decisões de classificação de consenso com base em resultados preditivos de modelos de ML individuais.

Houve uma explosão de artigos que adotam a estratégia de ensemble que tem mostrado resultados promissores que podem ajudar a melhorar o desempenho de classificação de modelos de pontuação de crédito (BAO; LIANJU; YUE, 2019).

Uma literatura crescente que usa uma estratégia integrada de ensemble mostrou maior poder de previsão em modelos de pontuação de crédito (MACHADO; KARRAY, 2022).

Outra técnica recém utilizada é o sistema híbrido para modelo de pontuação de crédito, em que uma técnica é empregada para a previsão final após o uso de várias técnicas heterogêneas na análise (LEO; SHARMA; MADDULETYA, 2019).

Usam-se algoritmos de ML híbridos que combinam métodos de ML não supervisionados e supervisionados. Machado e Karray (2022), descobriram que os modelos híbridos superam suas contrapartes individuais na previsão de pontuações de crédito de clientes comerciais.

Além disso, embora a literatura existente ignore as pontuações de crédito anteriores, observaram que o desempenho preditivo dos modelos híbridos é maior quando esses recursos são incluídos.

Alguns estudos sumarizaram outras pesquisas tabelando os classificadores recomendados com as medidas de avaliação dos modelos. Wang et al. (2020) levantou 12 estudos entre 2003 a 2021, em que o algoritmo mais recomendado, em cinco casos, são as redes neurais artificiais.

Já Lessmann et al. (2015), em quarenta e oito estudos também observou que a

utilização de redes neurais foi mais frequente, seguido pelo *SVM*.

O julgamento de especialistas também pode ser combinado com métodos de computação, [Lappas e Yannacopoulos \(2021\)](#), observaram resultados que indicam que ao medir a consistência dos especialistas, suas opiniões podem contribuir para chegar a soluções quase ótimas em um tempo razoável.

O aprendizado por reforço profundo (DRL) desenvolvido nos últimos anos, é uma subclasse do aprendizado de máquina, que estuda como o agente seleciona a estratégia de ação ideal maximizando a recompensa cumulativa esperada ([WANG; JIA; TIAN, 2022](#)).

A ideia básica do DRL é aprender a estratégia de ação ótima que pode completar o objetivo da tarefa maximizando a recompensa cumulativa obtida pelo agente do ambiente de acordo com a função de recompensa.

Importante ressaltar que na literatura recente não foi encontrado estudo que foque especificamente na linha de crédito do cheque especial. Sendo encontrado na maioria das vezes estudos sobre risco de crédito para cartão de crédito ([FONSECA; WANKE; CORREA, 2021](#)), ([LI et al., 2014](#)), ([OH; JOHNSTON, 2014](#)) e ([CAMPBELL; GRANT; THORP, 2022](#))

Por fim, ainda no campo de machine learning mas desvirtuando da pontuação de crédito, outro interessante campo de estudo é o desenvolvimento de modelo de aprendizado de máquina para inadimplência de crédito ao consumidor que pode ser surpreendentemente preciso na previsão de eventos de crédito com 3 a 12 meses de antecedência ([KHANDANI; KIM; LO, 2010](#)).

3 Dados e Métodos

3.1 Base de Dados

A base de dados desta pesquisa utiliza amostra anonimizada de operações de crédito de uma carteira comercial de um banco brasileiro do Segmento 1 (S1). Trata-se de concessão de limite de crédito rotativo concedido por instituições financeiras em conta de depósitos à vista titulada por pessoas naturais, conhecido como cheque especial (BRASIL, 2019).

Foram consideradas as contratações a partir de 01/05/2020 apenas para as safras maduras (com 12 meses de desempenho), critério necessário para a marcação do cliente (bom e mau). Desta forma, foram analisadas seis safras maduras disponíveis para a construção da base de dados, período entre Mai/2020 e Out/2020.

Com um total de 168.800 (cento e sessenta e oito mil e oitocentos) observações, os critérios para marcação de bom ou mau estão relacionados às variáveis de *default* chamadas M12EVER90 e M12OVER90.

O cliente é considerado “mau” caso ultrapasse 90 dias de atraso em algum mês, durante os 12 primeiros meses após a contratação. Caso contrário, o cliente será considerado “bom”. Trata-se da variável M12EVER90.

Já para a M12OVER90, o cliente é considerado “mau” caso ultrapasse 90 dias de atraso no 12º mês após a contratação. Caso contrário, o cliente será considerado “bom”.

Conforme Bao, Lianju e Yue (2019) um desafio para os modelos de pontuação de crédito é a indisponibilidade de dados de crédito do mundo real, uma vez que os dados de crédito dos clientes são confidenciais nas instituições financeiras e os pesquisadores não conseguem acesso a esses dados.

Como alternativa, pesquisas tem amplamente utilizado *datasets* de crédito mundiais publicados pela Universidade da Califórnia Irvine (UCI) (*dataset* australiano, 1987; *dataset* alemão, 1994) para lançar pesquisas de pontuação de crédito ao longo dos anos.

No entanto, com a atual dinâmica observada de clientes e do mercado de crédito, fatores relacionados ao crédito ao cliente mudaram significativamente ao longo do tempo, uma vez que o comportamento do cliente mudou muito e rapidamente. Desta forma, pesquisas baseadas em conjuntos de dados atualizados do mundo real são importantes e urgentes (BAO; LIANJU; YUE, 2019).

No presente estudo os dados coletados são reais advindos de um banco brasileiro de varejo de grande porte do Segmento 1 (S1).

Os dados foram trabalhados na linguagem computacional Python com o uso das bibliotecas Pandas para manipulação, Scikit-learn para modelos e Imblearn para *oversampling*.

3.1.1 Seleção de Amostras

Do total de observações disponíveis, 168.800, foram realizados filtros para retirada de números discrepantes que possam induzir a resultados enviesados. Das variáveis disponíveis algumas apresentaram baixa completude com poucos resultados disponíveis e pouca variabilidade.

Com quantil de 99%, mantendo quantidade significativa de observações, foram utilizados como filtro o tempo de renda do cliente, menor que 32 anos de renda; e a renda bruta do cliente, renda mensal menor que R\$ 31.593,00 (trinta e um mil, quinhentos e noventa e três reais). Além disso, outro filtro utilizado foram os analfabetos, que foram retirados. Esses três filtros excluíram observações das amostras para melhor aproveitamentos dos resultados.

Foi demonstrado que o uso de filtros para *outliers* pode ajudar a melhorar o desempenho do modelo. Como a presença de *outliers* afeta negativamente o desempenho de qualquer modelo de classificação ou previsão, é necessário filtrar os dados para eliminar as instâncias que atrapalham o processo de generalização (GARCÍA; MARQUÉS; SÁNCHEZ, 2012).

Assim, da base original restaram 165.041 (cento e sessenta e cinco mil e quarenta e uma) observações. E dessas, em razão do alto esforço computacional requerido, a amostra coletou 25% dos casos, representando 41.250 (quarenta e um mil, duzentos e cinquenta) observações para geração do modelo.

Na tabela 1 são demonstrados os números envolvidos nas observações da amostra utilizada.

Tabela 1 – Números da amostra

| Base | Quantidade |
|--|------------|
| Amostra original | 168.800 |
| Amostra após filtros | 165.041 |
| Observações utilizadas - 25% da amostra após filtro | 41.250 |
| Base de treino - 75% das observações | 30.945 |
| Base de teste - 25% das observações | 10.315 |
| Reamostragem - <i>default</i> 6,1% do treino | 29.050 |
| Base de treino com reamostragem | 58.100 |

Fonte: Autor.

3.1.2 Tratamento das Variáveis

A variável *target* do modelo foi nomeada como M12EVER90, em que é considerado cliente "mau" quando se ultrapassa 90 dias de atraso em algum mês, durante os 12 primeiros meses após a contratação, caso contrário, é identificado como "bom".

A escolha dessa variável está em consonância com a Resolução CMN 4.557/2017 (BRASIL, 2017) no qual se afirma que a exposição (carteira de crédito) deve ser caracterizada como ativo problemático quando verificado que uma obrigação está em atraso há mais de noventa dias.

Ressalta-se que a variável alvo também encontra similaridade com a definição de *Non-Performing Loan* (NPL), empréstimo bancário que passou mais de 90 dias sem que o mutuário tenha procedido ao pagamento das prestações acordadas ou dos juros, métrica internacionalmente conhecida e utilizada.

Segundo a metodologia do SCR.data do Painel de Operações de Crédito do Banco Central do Brasil (BCB - Banco Central do Brasil, 2022a), são consideradas como ativos problemáticos as operações de crédito em atraso há mais de noventa dias e as operações nas quais existem indícios de que respectiva obrigação não será integralmente honrada. Entende-se que há indicativos de que a obrigação poderá não ser integralmente honrada, entre outros eventos, quando a operação for objeto de reestruturação e a instituição financeira reconhecer contabilmente deterioração significativa da qualidade do crédito do tomador, classificando-o entre os níveis de risco E e H.

Essa definição de ativo problemático encontra forte correlação com situação de *default*, objeto desse estudo.

As variáveis independentes para o estudo foram gênero (masculino ou feminino), instrução (ensino fundamental, ensino médio, ensino superior e pós-graduação), tempo de renda (quantidade de dias da renda mais antiga do cliente, podendo ser formal ou informal) e renda bruta (valor total da renda bruta apurada, ou seja, a soma de todas as rendas brutas formais e informais).

A tabela 2 relaciona as variáveis utilizadas nesse estudo, tanto a variável dependente, quanto às independentes.

Lembramos que eventuais variações no cenário econômico do país e estratégias políticas de incentivo ao crédito, não foram objeto desse estudo e, portanto, não foram avaliados.

3.1.3 Balanceamento da Base

Um das dificuldades encontradas no mundo real em modelos de *machine learning* é utilizar bases de dados desbalanceados. Em grande parte dos casos que se pretende prever

Tabela 2 – Variáveis

| Variáveis | Descrição | Tipo |
|-----------------------|---|-----------|
| M12EVER90 | Cliente "mau"atraso de 90 dias em 12 meses contrato | Informado |
| Gênero | Identifica o gênero humano: masculino ou feminino | Informado |
| Instrução | Ensino fundamental, médio, superior e pós-graduação | Informado |
| Tempo de renda | Quantidade de dias da renda mais antiga | Informado |
| Renda bruta | Valor total da renda bruta apurada | Informado |

Fonte: Autor.

um modelo de classificação são as classes minoritárias o *target* do estudo.

Nesta pesquisa no universo amostral empregado, a classe objeto de estudo, ou seja, os clientes inadimplentes representavam apenas 6,1% das observações, o que totalizavam 1.895 de inadimplentes frente a um universo de 30.945 observações.

A técnica lidada para contornar essa situação chama-se *oversampling*, em que são adicionadas novas amostras (sintéticas) na classe minoritária, até que se iguale ao número de amostras da classe majoritária.

O método SMOTE (*Synthetic Minority Over-sampling Technique*) foi aplicado, assim, novas amostras sintéticas são geradas a partir de amostras vizinhas. Essa implementação de SMOTE não muda o número de casos majoritários.

A técnica SMOTE vem sendo usada com bastante frequência em estudo recentes sendo tanto no risco de crédito quanto na detecção de fraude (PANDEY; MITTAL; SUBBIAH, 2021), (RAWAT et al., 2021) e (ZHANG et al., 2022).

Os novos dados não são apenas cópias de casos minoritários existentes. Em vez disso, o algoritmo usa exemplos do espaço de recurso para cada classe de destino e seus vizinhos mais próximos. O algoritmo, em seguida, gera novos exemplos que combinam recursos do caso de destino com recursos de seus vizinhos. Essa abordagem aumenta os recursos disponíveis para cada classe e torna os exemplos mais gerais.

Utilizamos peso 1 para 1 no SMOTE para o número de vizinhos mais próximo, com $k = 5$ vizinhos e alcançamos uma base de treino com reamostragem de 58.100 observações, sendo 29.050 de cada classe.

3.1.4 Validação do Modelo

A base de dados foi dividida em dados para treino e para teste na razão 75 / 25, ou seja, 75% das observações empregados no treino, totalizando 30.945 observações e 25% ficou para teste, na quantidade de 10.315 observações.

Utilizamos *cross-validation* e hiperparâmetros na validação do modelo para avaliar sua capacidade de generalização.

Na validação cruzada, repartimos aleatoriamente o conjunto de dados diversas vezes entre partes de treino e partes de teste para garantir que o modelo performa bem em todo o conjunto.

A técnica escolhida de validação cruzada chama-se *k-fold*, que consiste em dividir o conjunto total de dados em k subconjuntos mutuamente exclusivos do mesmo tamanho e, a partir daí, um subconjunto é utilizado para teste e os $k-1$ restantes são utilizados para estimação dos parâmetros, fazendo-se o cálculo da acurácia do modelo. Este processo é realizado k vezes alternando de forma circular o subconjunto de teste. No presente estudo $k = 3$ e a reamostragem foi realizada dentro de cada *fold*.

Em um algoritmo de *ML*, os parâmetros são ajustados diretamente pelo processo de aprendizado e influenciam diretamente na performance do algoritmo. Já os hiperparâmetros são variáveis do algoritmo definidas antes do treinamento que representam características mais construtivas do modelo.

Neste estudo a técnica selecionada de otimização de hiperparâmetros foi a *GridSearch*, ou pesquisa em grade, um algoritmo de busca que recebe um conjunto de valores de um ou mais hiperparâmetros e testa todas as combinações possíveis, exaustivamente. Fornece alguns valores de input e testa todas as combinações plotando em um plano cartesiano. Em seguida, selecionará os hiperparâmetros que obtiveram o menor erro.

3.2 Metodologias de *Machine Learning*

Diversos algoritmos de Machine Learning (*ML*) foram criados e estudados pelos pesquisadores ao longo dos últimos anos para novas aplicações em *credit scoring*.

Num amplo estudo que levantou quarenta e oito pesquisas e cobriu diversas abordagens (linear / não linear, paramétrica, não-paramétrica etc) em que objetivou explorar a eficácia relativa de algoritmos de classificação alternativos na pontuação de crédito de varejo, [Lessmann et al. \(2015\)](#) compararam 41 (quarenta e um) classificadores em termos de seis medidas de desempenho em oito conjuntos de dados de pontuação de crédito do mundo real. Concluíram que vários classificadores preveem o risco de crédito de forma significativamente mais precisa do que o padrão utilizado na indústria bancária, ou seja, a regressão logística. Afirmaram ainda que os classificadores de conjuntos (*ensemble*) especialmente heterogêneos têm bom desempenho.

Dentre os classificadores de *ML* os principais pesquisados em modelos supervisionados são Regressão Logística, Support Vector Machines, Redes Neurais Artificiais, Boosting, Bagging, Árvore de Decisão, Floresta Aleatória e K-NN. Já para os modelos

não-supervisionados destaca-se o K-Means e por fim temos os modelos híbridos / *ensemble*.

3.2.1 Regressão Logística

Modelo mais tradicional para pontuação de crédito, utilizado largamente na indústria bancária, a regressão logística pode ser observada como base para problemas de classificação binária. Além disso, descrever e estimar a relação entre uma variável binária dependente e variáveis independentes torna esse modelo instrutivo em aprendizado profundo (WANG et al., 2020).

Com base nos dados de treinamento, deve-se estimar os coeficientes da equação de regressão logística, o que em geral é feito pela estimativa de máxima verossimilhança. Os coeficientes otimizados levarão a um modelo que preveria o valor da classe *default* próximo a um, e o valor para as demais classes seria zero. Portanto, os pesquisadores perseguem o valor dos coeficientes que poderiam minimizar o erro nas probabilidades previstas. As previsões são feitas inserindo números na equação recém-criada e computando o resultado. Sua fórmula é dada por:

$$\log[p(1 - p)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Onde $\log[p(1 - p)]$ como variável dependente é o logaritmo da razão de dois resultados de probabilidade de interesse, sendo p a probabilidade de *default*. O β_i como termo de interceptação é o coeficiente relacionado às variáveis independentes X .

3.2.2 Support Vector Machines

A ideia principal do algoritmo *Support Vector Machines* (SVM) é projetar os dados de entrada em um espaço de características de alta dimensão e então encontrar um hiperplano suportado pelos vetores de suporte para separar as duas classes com uma margem máxima (BAO; LIANJU; YUE, 2019). Com base nas características dos vetores de suporte, o rótulo da nova amostra de entrada pode ser previsto.

Destaca-se em comparação com outros algoritmos por causa de suas soluções superiores para resolver o problema de esparsidade. Este modelo tem sido amplamente aplicado no campo da pontuação de crédito, devido às suas poderosas capacidades preditivas.

3.2.3 k-Nearest Neighbors

Conforme Wang et al. (2020) o algoritmo kNN possui o seguinte mecanismo de trabalho: calcula a distância, encontra os vizinhos mais próximos e vota nos rótulos. Neste método, "K" é o número de vizinhos mais próximos e é uma variável de controle para este método. Cada conjunto de dados tem seu próprio número ótimo de "K". Se "K" for pequeno, o ruído influenciaria muito o resultado, enquanto que se "K" for grande, seria caro

computar. Observaram que um "K"pequeno é mais flexível com alta variância e baixo viés, enquanto um grande "K"não é flexível com menor variância, mas maior viés.

A ideia principal por trás do modelo kNN é que ele prediz os rótulos das novas amostras de entrada de acordo com o conjunto mais próximo (ou k -vizinhos mais próximos) de amostras previamente rotuladas (BAO; LIANJU; YUE, 2019).

A vantagem do método é que a fase de treinamento do kNN é muito mais rápida que outros classificadores, e é bastante intuitiva e simples, pois não contém suposições. A fraqueza desse método é que a fase de teste da classificação K-vizinho mais próximo é mais cara e mais lenta, pois requer grande memória para armazenar todos os conjuntos de dados de treinamento. Além disso, o kNN é sensível a magnitudes e outliers porque escolhe diretamente os vizinhos com base na distância (WANG et al., 2020).

3.2.4 Árvore de Decisão

Os modelos de Árvore de Decisão (DT - *Decision Tree*) não requerem cálculos poderosos, são fáceis de entender e explicar, lidam bem com informações ausentes e trabalham com variáveis numéricas e nominais (MACHADO; KARRAY, 2022).

É um tipo de método de aprendizado supervisionado não paramétrico e ferramenta de apoio à decisão usado para regressão e classificação. Este algoritmo usa a representação em árvore para resolver o problema, nós internos na árvore para representar um atributo e nós na folha para representar rótulos de classe. Inicia-se da raiz da árvore, separando a amostra em grupos de conjuntos homogêneos de acordo com os divisores mais essenciais das variáveis de entrada. Repita esse processo até que os nós na folha em todos os ramos da árvore sejam encontrados (WANG et al., 2020).

Os resultados preditivos do DT podem ser representados como um gráfico acíclico direcionado em forma de árvore, assim os resultados do modelo podem ser facilmente compreendidos por humanos (BAO; LIANJU; YUE, 2019).

Comparado a outros métodos, esse método requer pouca limpeza de dados, pois não é afetado por valores discrepantes. No entanto, existem pontos fracos, como o problema de ajuste excessivo, que pode ser eliminado usando a floresta aleatória.

3.2.5 Gradient Boosting

Modelos de *Gradient Boosting* podem ser usados para classificação ou regressão. Esse algoritmo faz uso de um *learner* forte – um classificador arbitrariamente bem correlacionado com a classificação verdadeira – que é construído a partir da combinação de diferentes *leaners* fracos, ou seja, classificadores que são apenas ligeiramente correlacionados com a classificação verdadeira. Ao combinar *leaners* fracos e fortes, a saída do modelo de

gradient boosting é uma previsão (classificação ou regressão) (MACHADO; KARRAY, 2022). Utilizamos Árvore de Decisão como *learner* nesta pesquisa.

O termo “*gradient boosting*” vem da ideia de “boosting” (aumento) que é agrupar *leaners* de base fraca com alto viés e baixa variância, tipicamente árvores de decisão, a fim de gerar um modelo mais confiável e preciso. *Gradient boosting* é uma extensão do *boosting* onde a estratégia de minimização de erros baseada em *boosting* é usada para gerar modelos de forma aditiva. O processo de modelagem é adicionar árvores de decisão por vez, então a próxima árvore é adicionada e treinada para reduzir a perda movendo-se na direção certa. O modelo continua adicionando árvores até que o número de árvores atinja um número fixo ou a perda atinja um nível aceitável ou não melhora mais.

3.2.6 Floresta Aleatória

Os modelos de floresta aleatória são uma categoria de modelos aditivos que fazem previsões mesclando as saídas do modelo árvore de decisão, que vêm de uma sequência de modelos básicos (MACHADO; KARRAY, 2022).

A floresta aleatória monta uma grande quantidade de árvores de decisão a partir do conjunto de dados de treinamento, e também usa uma ferramenta chamada *bagging* para realizar tarefas de classificação e regressão. Cada árvore de decisão representa uma previsão de classe, esse método coleta os votos dessas árvores de decisão e a classe com mais votos é considerada a classe final (WANG et al., 2020).

O modelo de floresta aleatória usa aleatoriedade em dois estágios: primeiro, seleciona aleatoriamente subconjuntos do conjunto de dados original; em segundo lugar, seleciona aleatoriamente subconjuntos de recursos derivados das dimensões originais de recursos completos. Desta forma, a correlação entre as árvores de decisão na floresta aleatória é reduzida (BAO; LIANJU; YUE, 2019).

3.2.7 Medidas de Avaliação de ML

Para a avaliação dos modelos foram utilizadas medidas-padrão para classificação de crédito (WANG et al., 2020). Estas medidas são: precisão, *recall*, especificidade, F1-Score, média geométrica e index balanced accuracy.

3.2.7.1 Matriz de Confusão

Antes de entrar no conceito de cada métrica demonstramos a Matriz de Confusão, uma tabela que mostra as frequências de classificação para cada classe do modelo, ou seja, permite a visualização do desempenho de um algoritmo de classificação.

Na Tabela 3 a seguir, cada linha da matriz representa instâncias de uma classe prevista enquanto cada coluna representa instâncias da classe atual.

Tabela 3 – Matriz de Confusão

| | Condição Positiva Atual | Condição Negativa Atual |
|----------------------------|--------------------------|--------------------------|
| Condição Positiva Prevista | Verdadeiro Positivo (VP) | Falso Positivo (FP) |
| Condição Negativa Prevista | Falso Negativo (FN) | Verdadeiro Negativo (VN) |

Fonte: Autor.

O resultado **Verdadeiro Positivo** (VP) ocorre quando no conjunto real, a classe que estamos buscando foi prevista corretamente. No caso presente, se o cliente previsto como inadimplente, de fato, verificou-se inadimplente.

Já o resultado **Falso Positivo** (FP) ocorre quando no conjunto real, a classe que estamos buscando prever foi prevista incorretamente. Desta forma, nesse estudo verificamos se o cliente previsto como inadimplente foi classificado incorretamente, ou seja, o cliente não se encontra em atraso. Também é conhecido como Erro Tipo 1.

O resultado **Falso Verdadeiro** (FV) ocorre quando no conjunto real, a classe que não estamos buscando prever foi prevista corretamente. Ou seja, se o cliente previsto como adimplente foi verificado como de fato sem atraso.

Por fim, o resultado **Falso Negativo** (FN) ocorre quando no conjunto real, a classe que não estamos buscando prever foi prevista incorretamente. Nesse estudo, verificamos se os clientes previstos como inadimplentes apresentaram atraso na linha de crédito. Também é conhecido como Erro Tipo 2.

3.2.7.2 Precisão

A precisão é uma das métricas mais comuns para avaliar modelos de classificação de *machine learning*. Esta métrica é calculada pela razão entre a quantidade de exemplos classificados corretamente como positivos e o total de exemplos classificados como positivos. Dá maior ênfase para os erros por falso positivo.

Responde a pergunta: Qual a proporção de identificações positivas foi realmente correta? Em outras palavras, o qual bem meu modelo trabalhou. O que em nosso estudo significa, de todos os clientes classificados como inadimplentes quantos foram corretamente classificados.

$$Precisão = \frac{VP}{VP+FP}$$

3.2.7.3 Recall

Por outro lado, a revocação, ou *recall* em inglês e também conhecida como sensibilidade, dá maior ênfase para os erros por falso negativo. Esta métrica é calculada pela razão

entre a quantidade de exemplos classificados corretamente como positivos e a quantidade de exemplos que são realmente positivos.

Busca responder a seguinte pergunta: de todos os exemplos que são positivos, quantos foram classificados corretamente como positivos? No caso desse estudo, de todos os clientes que foram verificados serem inadimplentes, quantos tiveram a previsão acertada.

$$Recall = \frac{VP}{VP+FN}$$

3.2.7.4 Especificidade

A especificidade avalia a capacidade do método de detectar resultados negativos. Também é conhecida como *True Negative Rate*. É a razão de classes negativas corretamente previstas para o total de classes negativas reais.

$$Especificidade = \frac{VN}{VN+FP}$$

3.2.7.5 F1-Score

A métrica F1, ou *F1-Score* em inglês e também conhecida como *F-measure*, leva em consideração tanto a precisão quanto a revocação. Ela é calculada pela média harmônica entre as duas.

Traz bons resultados quando utilizar em *dataset* com classes desproporcionais, e o seu modelo não emite probabilidades. Em geral, quanto maior o F1-Score, melhor.

Um modelo que apresenta um bom *F1-Score* é um modelo capaz tanto de acertar suas predições (precisão alta) quanto de recuperar os exemplos da classe de interesse (revocação alta). Portanto, esta métrica tende a ser um resumo melhor da qualidade do modelo.

$$F1 = 2 * \frac{Precisão * Sensibilidade}{Precisão + Sensibilidade}$$

3.2.7.6 Média Geométrica

A média geométrica é a raiz quadrada do produto de Revocação e Especificidade. Ou seja, raiz quadrada do produto entre True Positive Rate e True Negative Rate.

$$MédiaGeométrica = \sqrt{Sensibilidade * Especificidade}$$

3.2.7.7 Index Balanced Accuracy

Trata-se de métrica para medir o desempenho e para avaliar processos de aprendizagem em domínios desequilibrados de duas classes. O método combina um índice imparcial de sua precisão geral e uma medida sobre quão dominante é a classe com a maior taxa de precisão individual.

Primeiramente, é calculada a Dominância, que é a diferença da Taxa Verdadeiramente Positiva e da Taxa Verdadeira Negativa. A fórmula é $Dominance = Recall - Specificity$. Ao valor de dominância é atribuído um peso α .

$$IBA = (1 + \alpha * (Sensibilidade - Especificidade)) * (Sensibilidade * Especificidade)$$

3.2.7.8 Curva AUC

AUC é um parâmetro amplamente utilizado derivado da curva ROC (Receiver Operating Characteristic) (XIAO et al., 2017) e (LOYOLA-GONZÁLEZ et al., 2016). AUC representa a probabilidade de um modelo classificar aleatoriamente um bom candidato em vez de classificar um candidato ruim aleatoriamente. Varia de 0 a 1 e quanto maior a AUC, melhor é o desempenho do modelo (BAO; LIANJU; YUE, 2019).

Pode ser usado como métrica de qualidade de um modelo, pois quanto mais próxima a curva estiver do canto superior esquerdo, maior será a área sob a curva e melhor será o modelo. Tem como vantagem não ser sensível ao desbalanço de classes.

Nos problemas de classificação binária, AUC pode ser aproximadamente denotada como a seguir (WANG; JIA; TIAN, 2022):

$$AUC = \frac{Sensibilidade + Especificidade}{2}$$

4 Resultados

Das 10.315 observações separadas para teste, observamos as medidas de avaliação do modelo para cada uma das técnicas de *ML* empregadas, conforme Tabela 4. A referência 1 é a classe dos inadimplentes, *target* dessa pesquisa.

Tabela 4 – Resultados dos modelos

| pre | rec | spe | f1 | geo | iba | sup | referencia | modelo |
|------|------|------|------|------|------|----------|------------|-----------|
| 0.96 | 0.43 | 0.75 | 0.59 | 0.57 | 0.31 | 9696.00 | 0 | logistica |
| 0.08 | 0.75 | 0.43 | 0.14 | 0.57 | 0.33 | 619.00 | 1 | logistica |
| 0.91 | 0.45 | 0.73 | 0.57 | 0.57 | 0.31 | 10315.00 | medias | logistica |
| 0.94 | 0.79 | 0.27 | 0.86 | 0.46 | 0.22 | 9696.00 | 0 | knn |
| 0.07 | 0.27 | 0.79 | 0.12 | 0.46 | 0.20 | 619.00 | 1 | knn |
| 0.89 | 0.76 | 0.30 | 0.82 | 0.46 | 0.22 | 10315.00 | medias | knn |
| 0.97 | 0.35 | 0.83 | 0.52 | 0.54 | 0.28 | 9696.00 | 0 | svc |
| 0.08 | 0.83 | 0.35 | 0.14 | 0.54 | 0.30 | 619.00 | 1 | svc |
| 0.92 | 0.38 | 0.80 | 0.49 | 0.54 | 0.28 | 10315.00 | medias | svc |
| 0.96 | 0.43 | 0.73 | 0.60 | 0.56 | 0.31 | 9696.00 | 0 | tree |
| 0.08 | 0.73 | 0.43 | 0.14 | 0.56 | 0.32 | 619.00 | 1 | tree |
| 0.91 | 0.45 | 0.71 | 0.57 | 0.56 | 0.31 | 10315.00 | medias | tree |
| 0.96 | 0.44 | 0.74 | 0.61 | 0.57 | 0.32 | 9696.00 | 0 | rsf |
| 0.08 | 0.74 | 0.44 | 0.14 | 0.57 | 0.34 | 619.00 | 1 | rsf |
| 0.91 | 0.46 | 0.72 | 0.58 | 0.57 | 0.32 | 10315.00 | medias | rsf |
| 0.96 | 0.48 | 0.67 | 0.64 | 0.57 | 0.32 | 9696.00 | 0 | boost |
| 0.08 | 0.67 | 0.48 | 0.14 | 0.57 | 0.33 | 619.00 | 1 | boost |
| 0.91 | 0.49 | 0.66 | 0.61 | 0.57 | 0.32 | 10315.00 | medias | boost |

Fonte: Autor.

pre = Precisão; rec = Recall; spe = Especificidade; f1 = F1-Score; geo = Média Geométrica; iba = Index Balanced Accuracy

Percebe-se que os valores de Precisão podem ser considerados altos para a classe dos inadimplentes e também na média, porém para os dados em *default* os valores estão baixos, com valor máximo de 8% para 5 dos 6 modelos.

Para *Recall* percebe-se grande variedade nos valores, com sinal contraditório entre *SVM* e *kNN*, enquanto o primeiro obteve o maior resultado (83%) para a classe dos inadimplentes e o pior na média (38%), o *kNN* ficou em último (27%) quando dos inadimplentes e em primeiro (76%) na média.

A medida de avaliação Especificidade apresenta comportamento similar com grande variedade nos dados, porém, aqui quem teve melhor desempenho entre os dados em *default* foi o *kNN* (79%) e sendo o pior na média (30%), já o o *SVM* teve desempenho de 35% na

classe 1 e de 80% na média, com melhor resultado.

A métrica F1-Score é uma média harmônica entre Precisão e Revocação e no caso em questão obteve resultados bastante divergentes entre os inadimplentes, que tiveram valor máximo de 14% para Regressão Logística, SVM, Árvore de Decisão, Floresta Aleatória e *Boosting*, enquanto *kNN* alcançou 82%, bem acima dos demais, quando calculado pela média.

Média Geométrica não teve diferença entre os valores para inadimplentes ou médias, registrando exatamente os mesmo valores, sendo Regressão Logística, Floresta Aleatória e *Gradient Boosting* com melhor desempenho de 57%.

Outra medida de avaliação com baixa nota, o IBA alcançou seus maiores patamares com a Floresta Aleatória, sendo 34% para os em *default* e 32% para a média dos dados.

Em resumo quando se concentra nos dados do *target* e é incluído o indicador AUC, derivado da curva ROC, é perceptível observar por pequena diferença o melhor resultado do indicador Floresta Aleatória, que converge 5 dos 7 melhores resultados, conforme Tabela 5.

Tabela 5 – Resultados da classe *target*

| pre | rec | spe | f1 | geo | iba | AUC | referencia | modelo |
|------|-------------|-------------|------|------|-------------|------|------------|------------|
| 0.08 | 0.75 | 0.43 | 0.14 | 0.57 | 0.33 | 0.59 | 1 | logistica |
| 0.07 | 0.27 | 0.79 | 0.12 | 0.46 | 0.20 | 0.53 | 1 | knn |
| 0.08 | 0.83 | 0.35 | 0.14 | 0.54 | 0.30 | 0.59 | 1 | svc |
| 0.08 | 0.73 | 0.43 | 0.14 | 0.56 | 0.32 | 0.58 | 1 | tree |
| 0.08 | 0.74 | 0.44 | 0.14 | 0.57 | 0.34 | 0.59 | 1 | rsf |
| 0.08 | 0.67 | 0.48 | 0.14 | 0.57 | 0.33 | 0.58 | 1 | boost |

Fonte: Autor.

pre = Precisão; rec = Recall; spe = Especificidade; f1 = F1-Score; geo = Média Geométrica; AUC = Area under the Curve ROC; iba = Index Balanced Accuracy

A métrica AUC foi calculada com valor máximo de 59% para Regressão Logística, SVM e Floresta Aleatória e pouca variabilidade pois o valor mínimo foi 53%.

Comparando o o resultado de Floresta Aleatória (FA) contra Regressão Logística (RL), percebe-se que os dois obtiveram a mesma Precisão (8%), RL superou FA em 1% na Sensibilidade (75% - 74%), já FA superou por 1% RL em Especificidade (44% - 43%), empate em *F1-Score* (14%), empate em Média Geométrica (57%), FA supera por 1% RL em Index Balanced Accuracy e para AUC empatam em 59%.

Na avaliação das métricas importante ressaltar [Khandani, Kim e Lo \(2010\)](#) ao se tentar referenciar o indicador mais adequando ao seu problema: se um credor deseja identificar agressivamente contas ruins usando um limite mais baixo, essa agressividade

resultará em classificações mais incorretas de contas boas como ruins, levando ao custo de oportunidade da perda de receita de juros de cortar as linhas de crédito de clientes de baixo risco.

Destaca-se que para um modelo de risco de crédito a taxa de erro tipo 2, ou seja, aqueles clientes previstos como adimplentes mas posteriormente identificados como em atraso são bastante preocupantes. Pois, numa análise conservadora, cada empréstimo concedido para um inadimplente pode eliminar os ganhos da operação de crédito de vários clientes adimplentes, visto a margem de ganho, juros do empréstimo, ser um valor percentual do capital emprestado.

Outro ponto a considerar, observando a Tabela 5, são os valores extremamente baixos do indicador Precisão, não passando de 8%, o que indica que a técnica de SMOTE de geração de dados sintéticos pode ter melhorado o resultado da Sensibilidade, que possui o menor valor em 67% chegando até 83% para *SVM*, porém deixando a Precisão seriamente restringida.

Pelos resultados da Tabela 5, afetado pelos baixos valores de Precisão, o *F1-Score*, também registrou baixos valores, com pouca variabilidade entre 12% a 14%.

5 Conclusões

No mundo contemporâneo em que a atuação de instituições financeiras serve para impulsionador a economia mundial, o desempenho preditivo dos modelos de pontuação de crédito tem um grande efeito na rentabilidade das instituições financeiras (BAO; LIANJU; YUE, 2019). Desta forma, o objetivo deste estudo de identificar um modelo de classificação de risco de crédito que permita predição em operações comerciais rotativas de cheque especial para pessoa física possibilita o devido gerenciamento de riscos nos bancos e mensuração mais precisa das situações de *default* proporcionando a evolução financeira do credor.

Das 6 técnicas utilizadas na avaliação do modelo, todas bastante difundidas, conhecidas e estudadas, a que mais se destacou foi Floresta Aleatória com desempenho melhor em 5 das 7 métricas de avaliação. A Flores Aleatória teve melhor desempenho isolado nos indicadores IBA e Especificação e empatou em primeiro em Precisão, *F1-Score* e Média Geométrica.

Importante destacar a situação anômala do indicador *kNN* que possui o melhor resultado em Especificação, porém, teve o pior desempenho em todos os outros.

A situação atípica deste indicador pode ter relação com o uso da técnica SMOTE para *oversampling* em que foram gerados dados sintéticos para equilibrar a quantidade de dados *target* dos normalizados.

Há de se destacar os baixos valores encontrados no modelo para Precisão, máxima de 8%; *F1-Score*, com até 14% e IBA de até 34%.

Em média a métrica com melhores valores (67%) foi *Recall* que responde a pergunta: de todos os exemplos que são positivos, quantos foram classificados corretamente como positivos.

Já a Precisão obteve os piores valores em média (8%), que responde a pergunta: dos exemplos classificados como positivos, quantos realmente são positivos.

Destaca-se a pequena superação da Floresta Aleatória contra a Regressão Logística, o que sugere que bancos possam ter melhores resultados com a aplicação de modelos de risco de crédito por algoritmos de computação inteligente versus modelos estatísticos tradicionalmente usados como a Regressão Logística.

Este estudo contribui no meio acadêmico ao se aprofundar numa linha de crédito específica, caso do cheque especial, que é bastante utilizada no Brasil com valores robustos, saldo na casa de R\$ 22 bilhões (BCB - Banco Central do Brasil, 2022b) e para o qual não foram encontrados outras pesquisas correlatas, sendo um contraponto ao tradicional

modelo de risco para cartão de crédito.

O uso de dados reais reforça o entendimento de que fatores relacionados ao crédito ao cliente mudaram significativamente ao longo do tempo, uma vez que o comportamento do cliente mudou muito e rapidamente. Sendo portante, importante e urgente ter pesquisas baseadas em conjuntos de dados atualizados do mundo real (BAO; LIANJU; YUE, 2019).

Para a indústria bancária, torna-se interessante a descoberta de que a pequena variabilidade das variáveis, com baixa completude e não incluindo os tradicionais dados de patrimônio e situação cadastral sugerem uma baixa precisão no desempenho de modelos preditivos. Portanto, o cadastro do cliente torna-se fonte de informação essencial.

Uma limitação deste trabalho que pode ter influenciado o resultado do modelo é a baixa quantidade de variáveis. Utilizamos apenas uma variável independente, M12EVER90, que afirma que o cliente é considerado “mau” caso ultrapasse 90 dias de atraso em algum mês, durante os 12 primeiros meses após a contratação.

Já as variáveis dependentes, (i) gênero (masculino ou feminino), (ii) instrução (ensino fundamental, ensino médio, ensino superior e pós-graduação), (iii) tempo de renda (quantidade de dias da renda mais antiga do cliente, podendo ser formal ou informal) e (iv) renda bruta (valor total da renda bruta apurada, ou seja, a soma de todas as rendas brutas formais e informais) podem estar aquém do necessário para avaliação de crédito.

A base de dados real advinda de um grande banco brasileiro foi inicialmente anonimizada não permitindo a identificação ou individualização dos dados, o que não está no escopo deste trabalho.

Trabalhos futuros podem contribuir no refinamento de modelo preditivo desse tipo de operação de crédito, cheque especial, utilizando mais variáveis, como às relacionadas a patrimônio, situação cadastral e negativados.

Deverá ainda ser verificada a relação do SMOTE com o kNN em *Machine Learning* para fins de permanência no modelo de previsão.

Outro caminho a ser seguido é testar modelo sem a utilização de dados discriminantes ou que afetem a privacidade, mesmo sendo permitido pela Lei Geral de Proteção de Dados - LGPD, mas que possam conter eventual restrição e ainda acompanhando a lei americana Equal Credit Opportunity Act (ECOA) que torna ilegal qualquer credor discriminar qualquer requerente, com relação a qualquer aspecto de uma transação de crédito com base na raça, cor, religião, origem nacional, sexo, estado civil, idade, ou o fato de que a totalidade ou parte do rendimento provenha de um programa de assistência pública. Com isso fica a pergunta se esse trabalho pode gerar discriminação.

O uso de modelos *ensemble* ou híbridos também é recomendado para aprimoramento do valor de previsão.

Alternativamente poderão ser gerados novos resultados com a adição dos *outliers* que foram previamente retirados.

Por fim, uma possibilidade em trabalhos futuros seria explicar as variáveis mais importantes para o modelo utilizando Shapley Value para identificar o sinal e suas principais variáveis.

Referências

APOSTOLIK, R.; DONOHUE, C.; WENT, P. *Foundations of Banking Risk: An Overview of Banking, Banking Risks, and Risk-Based Banking Regulation*. [S.l.]: John Wiley & Sons, 2009. Citado na página 25.

BANK, W. et al. *Domestic credit to private sector by banks (% of GDP)*. 2021. Disponível em: <<https://data.worldbank.org/indicator/FD.AST.PRVT.GD.ZS?locations=US-BR-GB-CN-JP-DE>>. Citado na página 25.

BAO, W.; LIANJU, N.; YUE, K. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Systems with Applications*, p. 301–315, 2019. Citado 8 vezes nas páginas 30, 33, 38, 39, 40, 43, 49 e 50.

BCB - Banco Central do Brasil. *Exposição de Motivos da Resolução 4.765/2019*,. [S.l.]: Voto 246/2019 - BCB, de 27/11/2019, 2019. Disponível em: <https://normativos.bcb.gov.br/Votos/CMN/201981/Voto_0812019_CMN.pdf>. Citado na página 25.

BCB - Banco Central do Brasil. *Relatório de Economia Bancária de 2020*. 2020. Disponível em: <https://www.bcb.gov.br/content/publicacoes/relatorioeconomiabancaria/reb_2020.pdf>. Citado na página 27.

BCB - Banco Central do Brasil. *SCR.data - Painel de Operações de Crédito - Metodologia*. 2022. Disponível em: <<https://www.bcb.gov.br/content/estabilidadefinanceira/scr/scr.data/metodologia.pdf>>. Citado na página 35.

BCB - Banco Central do Brasil. *SGS - Sistema Gerenciador de Séries Temporais*. 2022. Disponível em: <<https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>>. Citado 2 vezes nas páginas 27 e 49.

BRASIL. *Resolução do Conselho Monetário Nacional nº 4.557*. [S.l.]: Diário Oficial da União, 01/03/2017, Seção 1, p. 41 - 46, 2017. Disponível em: <<https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o&numero=4557>>. Citado na página 35.

BRASIL. *Resolução do Conselho Monetário Nacional nº 4.765*. [S.l.]: Diário Oficial da União, 28/11/2019, Seção 1, p. 56, 2019. Disponível em: <<https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o&numero=4765>>. Citado 2 vezes nas páginas 25 e 33.

CAMPBELL, D.; GRANT, A.; THORP, S. Reducing credit card delinquency using repayment reminders. *Journal of Banking & Finance*, 2022. Citado na página 31.

FONSECA, D. P.; WANKE, P. F.; CORREA, H. L. A two-stage fuzzy neural approach for credit risk assessment in a brazilian credit card company. *Applied Soft Computing*, 2021. Citado na página 31.

GARCÍA, V.; MARQUÉS, A.; SÁNCHEZ, J. On the use of data filtering techniques for credit risk prediction with instance-based models. *Expert Systems with Applications*, 2012. Citado na página 34.

- KHANDANI, A. E.; KIM, A. J.; LO, A. W. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, p. 2767–2787, 2010. Citado 3 vezes nas páginas 29, 31 e 46.
- LAPPAS, P. Z.; YANNAKOPOULOS, A. N. A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment. *Applied Soft Computing*, 2021. Citado na página 31.
- LEO, M.; SHARMA, S.; MADDULETYA, K. Machine learning in banking risk management: A literature review. MDPI, 2019. Citado 3 vezes nas páginas 27, 29 e 30.
- LESSMANN, S. et al. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, p. 124–136, 2015. Citado 3 vezes nas páginas 29, 30 e 37.
- LI, S. et al. Research on simulation of credit risk of credit card based on multi-agent. *Procedia Computer Science*, 2014. Citado na página 31.
- LOYOLA-GONZÁLEZ, O. et al. Study of the impact of resampling methods for contrast pattern based classifiers in imbalanced databases. *Neurocomputing*, 2016. Citado na página 43.
- MACHADO, M. R.; KARRAY, S. Assessing credit risk of commercial customers using hybrid machine learning algorithms. *Expert Systems with Applications*, 2022. Citado 3 vezes nas páginas 30, 39 e 40.
- MOSCATO, V.; PICARIELLO, A.; SPERLI, G. A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications*, 2021. Citado 2 vezes nas páginas 27 e 29.
- OH, J.-H.; JOHNSTON, W. J. Credit lender–borrower relationship in the credit card market – implications for credit risk management strategy and relationship marketing. *International Business Review*, 2014. Citado na página 31.
- PANDEY, M. K.; MITTAL, M.; SUBBIAH, K. Optimal balancing and efficient feature ranking approach to minimize credit risk. *International Journal of Information Management Data Insights*, 2021. Citado na página 36.
- RAWAT, S. et al. Application of machine learning and data visualization techniques for decision support in the insurance sector. *International Journal of Information Management Data Insights*, 2021. Citado na página 36.
- VIEIRA, J. R. de C. et al. Machine learning models for credit analysis improvements: Predicting low-income families’ default. *Applied Soft Computing*, 2019. Citado na página 27.
- WANG, Y.; JIA, Y.; TIAN, J. X. Y. Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring. *Expert Systems with Applications*, 2022. Citado 3 vezes nas páginas 29, 31 e 43.
- WANG, Y. et al. A comparative assessment of credit risk model based on machine learning — a case study of bank loan data. *Procedia Computer Science*, p. 141–149, 2020. Citado 4 vezes nas páginas 30, 38, 39 e 40.

XIAO, J. et al. Gmdh-based semi-supervised feature selection for customer classification. Knowledge-Based Systems, 2017. Citado na página 43.

ZHANG, A. et al. Smote-rknn: A hybrid re-sampling method based on smote and reverse k-nearest neighbors. Information Sciences, 2022. Citado na página 36.