

GERENCIAMENTO DE RISCO DE CRÉDITO POR MEIO DA UTILIZAÇÃO DE APRENDIZADO DE MÁQUINA: O CASO DO BANCO BS2

CREDIT RISK MANAGEMENT THROUGH THE USE OF MACHINE LEARNING: THE CASE OF BANCO BS2

ALEX GIOVANI DE ASSIS¹

Faculdade Fipecafi

● <https://orcid.org/0009-0005-7070-362X>

alexgiovandeassis@gmail.com

SONIA ROSA ARBUES DECOSTER

Faculdade Fipecafi

● <https://orcid.org/0000-0002-0081-347X>

sonia.decoster@fipecafi.org

RESUMO

O risco de crédito tem desempenhado um papel central em várias crises financeiras globais nas últimas três décadas. O cenário financeiro, cada vez mais complexo e interconectado, faz com que o gerenciamento de risco se torne fundamental para a estabilidade e o crescimento das instituições financeiras. Este estudo de caso tem como objetivo analisar a utilização de aprendizado de máquina, especificamente o algoritmo *Gradient Boosting Decision Tree* (GBDT), em um modelo preditivo, que combina variáveis financeiras e não financeiras significativas e utiliza as consultas aos *bureaus* de crédito na gestão de risco de crédito pelo Banco BS2, com o intuito de adquirir maior acurácia na tomada de decisões e melhorias na mitigação de riscos. A métrica F1, utilizada como parâmetro para demonstrar a precisão do modelo, comparada com o modelo da Serasa, apresenta um índice superior, de 0,77. A capacidade de monitoramento contínuo oferecida por esse modelo preditivo tem proporcionado ao BS2, desde 2022, uma visão em tempo real da saúde financeira de sua base de clientes, ajudando na implementação de políticas mais assertivas. A taxa de inadimplência da Pessoa Jurídica do Banco BS2, registrada pelo BCB-CADOC (2024), tem se mostrado decrescente após a implementação do novo modelo, baseado no algoritmo GBDT. Este estudo contribui para a promoção da inovação e competitividade nas instituições financeiras, incentivando a transparência e fortalecendo a confiança de investidores, stakeholders e reguladores, como o Banco Central, ao adotar ferramentas de Inteligência Artificial (IA) que detectam precocemente riscos de crédito e previnem crises sistêmicas.

Palavras-chave: Gestão de Risco de Crédito. Inadimplência. Modelo Preditivo. Aprendizado de Máquina. IA.

Editado em português e inglês. Versão original em português.

¹ Endereço para correspondência: R. Maestro Cardim, 1170 | Bela Vista | 01323-001 | São Paulo/SP | Brasil.

Recebido em 19/08/2024. Revisado em 20/12/2024. Aceito em 21/01/2025 pelo Prof. Dr. Rogério João Lunkes (Editor-Chefe). Publicado em 18/02/2025.

Copyright © 2025 RCCC. Todos os direitos reservados. É permitida a citação de parte de artigos sem autorização prévia, desde que identificada a fonte.

ABSTRACT

Credit risk has played a central role in several global financial crises over the past three decades. An increasingly complex and interconnected financial landscape makes risk management essential for the stability and growth of financial institutions. This case study aims to analyze the use of machine learning specifically, the Gradient Boosting Decision Tree (GBDT) algorithm in a predictive model that combines significant financial and non-financial variables and incorporates credit bureau inquiries into Banco BS2's credit risk management process. The goal is to achieve greater accuracy in decision-making and improvements in risk mitigation. The F1 metric, employed as a measure of the model's precision, shows a superior value of 0.77 when compared with the model used by Serasa. Since 2022, the continuous monitoring capability offered by this predictive model has provided BS2 Bank with a real-time view of the financial health of its customer base, thereby facilitating the implementation of more assertive policies. Furthermore, the default rate among Banco BS2's corporate clients, as recorded by BCB-CADOC (2024), has been on a decline following the implementation of the new GBDT-based model. This study contributes to promoting innovation and competitiveness within financial institutions by encouraging transparency and strengthening the confidence of investors, stakeholders, and regulators such as the Central Bank through the adoption of Artificial Intelligence (AI) tools that detect credit risks early and help prevent systemic crises.

Keywords: *Credit Risk Management. Default. Predictive Model. Machine Learning. AI.*

1 INTRODUÇÃO

O risco de crédito, conforme destacado por Lassance e Ternoski (2021), refere-se à probabilidade da ocorrência de inadimplência em uma obrigação financeira, o que pode acarretar em perdas substanciais para as instituições financeiras. Nas últimas décadas, estudos têm demonstrado a relevância do gerenciamento de risco de crédito para a estabilidade financeira global, considerando os impactos dos riscos sistêmicos que a inadimplência pode trazer para o mercado financeiro como um todo (Reinhart et al., 2020).

De acordo com Dewasari et al. (2024), a crescente complexidade dos sistemas financeiros e alta volatilidade econômica geram a necessidade de metodologias robustas de análise de crédito para a mitigação de riscos. Os modelos de avaliação de risco tradicionais não conseguem acompanhar o ritmo acelerado advindo dos crescentes volume, velocidade e complexidade das transações e dados financeiros (Rahmani et al., 2023). Diante disso, o gerenciamento de crédito exige abordagens preditivas e tecnológicas que permitam às instituições financeiras tomar decisões informadas e eficazes (Khemakhem & Boujelbene, 2018).

No contexto brasileiro, os *bureaus* de crédito para a análise de crédito apresentam relevância por fornecerem informações detalhadas sobre o histórico financeiro e o comportamento de pagamento dos consumidores, especialmente em um mercado onde o acesso a dados financeiros confiáveis pode fazer a diferença entre uma concessão de crédito segura e uma exposição arriscada (Serasa Experian, 2023). Os *bureaus* atuam como fontes confiáveis de informações para auxiliar na avaliação da capacidade de pagamento dos clientes e na tomada de decisões informadas (Oliveira & Santos, 2022). A Resolução CMN n.º 5.037 do Banco Central, de 29/09/2022, apresenta “o embasamento dos acordos para compartilhamento de dados com cinco gestores de bancos de dados (*bureaus* de crédito), com vistas a contribuir para a ampliação do acesso ao crédito aos brasileiros e a um custo mais acessível” (Banco Central do Brasil, 2022).

Segundo Fosu et al. (2023), os *bureaus* oferecem um substancial apoio na tomada de decisão para aprovar ou reprová-la uma concessão de crédito através de um score de crédito, pontuando o cliente e, desta forma, as instituições podem aprovar, calibrar o volume concedido ou reprová-la o crédito. Em países em desenvolvimento, a gestão de *bureaus* pelo setor privado

demonstra-se frequentemente mais eficiente em comparação com a administração pública (Oliveira & Santos, 2022).

Conforme Louzada et al. (2016), a necessidade de um gerenciamento de risco efetivo tem significado que as instituições financeiras começam a buscar uma melhoria contínua nas técnicas usadas para análise de crédito, o que tem resultado em um desenvolvimento e na aplicação de numerosos modelos quantitativos. A utilização de métodos de pontuação de crédito (*credit scoring*) tem crescido substancialmente na última década devido ao acesso a dados, aumento do poder computacional, requisitos regulatórios e a demanda por crescimento econômico (Demirgüç-Kunt et al., 2017). Nesse contexto, as inovações em aprendizado de máquina e inteligência artificial são apresentadas como ferramentas na construção de modelos preditivos para aumentar a acurácia, maior assertividade e prevenção (Lessman et al., 2015).

Conforme descrito por Provost e Fawcett (2001), o impacto dos riscos sistêmicos sobre a economia mundial reflete a importância dos modelos de aprendizado supervisionado para a detecção de padrões de comportamento financeiro que podem prever a inadimplência. O Acordo da Basileia III, lançado em 2013, demonstrou um maior compromisso do *Basel Committee* em incentivar modelos mais sofisticados para o cálculo padrão de risco de crédito (*Bank for International Settlements*, 2024). À medida que o mercado se torna cada vez mais competitivo, dinâmico e interoperável, tecnologias como inteligência artificial e aprendizado de máquina são essenciais para a aplicação prática de tecnologias emergentes no monitoramento de riscos financeiros, abordando a importância da adaptação dos modelos de crédito em cenários de incerteza e mudanças rápidas (Mashrur et al., 2020).

A aplicação desses modelos avançados, como as árvores de decisão e técnicas de *boosting*, facilita a adaptação das instituições financeiras às mudanças econômicas globais e aumenta a resiliência frente às crises (Reinhart et al., 2020). E, adicionalmente, conforme artigo de revisão da literatura realizada por Montevechi et al. (2024), atualmente, os avanços nos modelos desenvolvidos por meio de aprendizado de máquina permitem lidar com grandes bases de dados com maior poder preditivo, de forma que algoritmos, tais como árvores de decisão, *Support Vector Machines* (SVM) e redes neurais, apresentam potencial de evidenciar o processo de modelagem de risco de crédito. Entretanto, ainda conforme salientado por Montevechi et al. (2024), quando se trata de modelos classificadores, há evidências suficientes para afirmar que não há uma escolha considerada a melhor.

O gerenciamento de risco orientado por IA estende o seu alcance por meio de avaliação de risco de crédito, predição de risco de mercado, detecção de fraudes e na gestão do compliance (Rahman et al., 2021). O uso de aprendizado de máquina, através de algoritmos, dentre eles, o GBDT (*Gradient Boosting Decision Tree*), salienta como a inovação tecnológica pode melhorar a resiliência das instituições financeiras e apoiar a estabilidade econômica global de longo prazo (Lessman et al., 2015). Segundo Zhou et al. (2019), algoritmos baseados em árvores de decisão GBDT, XGBoost and LightGBM são os mais avançados algoritmos de aprendizado de máquina desenvolvidos nos últimos anos, por terem alcançado resultados esperados de predição em tarefas de predição com dados desbalanceados.

Além do mais, a precisão de um modelo de crédito não depende apenas do algoritmo utilizado, mas também da seleção das variáveis certas. E, desta forma, o GBDT se destaca, permitindo identificar quais variáveis mais impactam a previsão de inadimplência, aprimorando a capacidade do modelo de classificar corretamente os perfis de risco (Chen & Guestrin, 2016). À luz dos elementos previamente contextualizados, o objetivo deste estudo é analisar a utilização de aprendizado de máquina, especificamente o algoritmo *Gradient Boosting Decision Tree* (GBDT), na gestão de risco de crédito pelo Banco BS2, o qual desde 2022 vem utilizando o uso da IA. A questão de pesquisa que norteará o estudo é: **Como a utilização de um modelo de predição**

baseado em um algoritmo de aprendizado de máquina, especificamente o algoritmo *Gradient Boosting Decision Tree (GBDT)*, pode auxiliar na prevenção do risco de crédito?

Este estudo pretende contribuir com as pesquisas relativas a gerenciamento do risco da inadimplência, promovendo uma concessão de crédito mais responsável e empregando um algoritmo de aprendizado de máquina com o intuito de revelar os preditores do risco, bem como buscar propiciar uma disseminação do conhecimento nas ferramentas de aprendizado de máquina no cenário do mercado financeiro brasileiro, e mais especificamente, o algoritmo *Gradient Boosting Decision Tree (GBDT)*, por conta também de um volume incipiente de estudos que utilizem este algoritmo em análise de risco de crédito. O restante deste artigo é composto de mais cinco seções. A seção 2 fornece uma detalhada revisão da literatura a respeito do risco de crédito no mundo, a concessão de crédito no Brasil por meio dos *bureaus* de crédito e a utilização do aprendizado de máquina na análise de risco de crédito. Os procedimentos metodológicos usados no estudo são descritos na seção 3. A seção 4 apresenta e interpreta os resultados. E, por fim, a seção 5 apresenta observações finais, limitações do estudo e recomendações para artigos futuros, seguida pelas referências bibliográficas.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Risco de Crédito no Mundo

O risco de crédito, segundo Reinhart et al. (2020), é um dos principais fatores de análise na prevenção de crises financeiras globais, exigindo monitoramento constante e adaptações pelos reguladores e instituições financeiras. Compreender os eventos passados auxilia no desenvolvimento de modelos preditivos que fortalecem a resiliência econômica e institucional (Reinhart et al., 2020). Neste sentido, pode-se relacionar desde a crise financeira asiática de 1997, que, conforme Agarwal e Vandana (2022), destacou a vulnerabilidade de economias emergentes, em face de empréstimos excessivos e fraquezas sistêmicas. Já, conforme Calvo (2008), a crise financeira russa em 1998 foi marcada por uma série de fatores adversos, de forma que a incapacidade do governo russo em estabilizar a economia e financiar sua dívida resultou em default e na desvalorização do rublo, e reforçou a importância de uma gestão eficaz do risco de crédito soberano (Dabrowski, 2023).

A crise *dot.com*, ocorrida no início dos anos 2000, teve seu foco no mercado de ações das empresas de tecnologia, e caracterizou-se pelo colapso de muitas *startups* de internet que haviam sido supervalorizadas (Rizvi et al., 2015) e pela necessidade de uma maior diligência na concessão de crédito a empresas emergentes (Rizvi et al., 2015). A crise financeira de 2008, desencadeada pela concessão de hipotecas *subprime* nos Estados Unidos, levou a uma das maiores recessões globais da história recente e conduziu ao colapso grandes instituições financeiras, demonstrando as falhas sistêmicas na gestão de risco de crédito (Mian & Sufi, 2009). Essa crise impulsionou os Acordos de Basileia com o intuito de fortalecer as práticas de concessão de crédito e gerenciamento de risco, e incentivou modelos mais sofisticados para a avaliação e mitigação do risco de crédito (*Bank for International Settlements*, 2024).

Conforme Freeman et al. (2017), o gerenciamento de risco é crucial para a estabilidade financeira, destacando-se na indústria bancária, na qual o risco de inadimplência pode comprometer a sustentabilidade de uma instituição. E a análise histórica de eventos de inadimplência revela a importância de práticas regulatórias e sobretudo desenvolvimento e adoção de tecnologias de predição avançadas para a gestão de risco, conforme indicado por Rahman et al. (2021). De maneira similar, Dewasiri et al. (2024) avaliaram que as instituições financeiras que utilizam inteligência artificial e aprendizado de máquina têm demonstrado melhor capacidade de adaptação aos riscos e de tomada de decisões preventivas em contextos voláteis.

2.2 Concessão de crédito no Brasil por meio de consultas aos *bureaus* de crédito

No Brasil, a concessão de crédito é regulamentada pelo Banco Central do Brasil (BCB), que utiliza *bureaus* de crédito para consolidar dados de clientes (BCB, 2023), facilitando análises de risco mais precisas e preservando a liquidez do sistema financeiro (Mendonça & Deos, 2020). Além disso, a Resolução CMN n.º 5.037, de 2022, dispõe sobre o Sistema de Informações de Créditos (SCR), constituído por informações remetidas ao Banco Central do Brasil sobre operações de crédito e busca propiciar o intercâmbio de informações entre instituições financeiras e entre demais entidades (BCB, 2022).

Cada *bureau* utiliza uma metodologia de pontuação, conhecida como *score* de crédito, que considera o histórico financeiro dos clientes; pontuações mais baixas indicam maior risco de inadimplência e essa análise é fundamental para garantir a integridade das operações de crédito no mercado financeiro (Oliveira & Santos, 2022). A concessão de crédito é apoiada por consultas a *bureaus* de crédito, que fornecem dados essenciais para a avaliação do risco de crédito de clientes e empresas (Mendonça & Deos, 2020). Segundo Sfeir (2023), o mercado financeiro brasileiro frequentemente cruza análises internas com as informações disponibilizadas por *bureaus* de crédito, aproveitando dados como o Cadastro Positivo, e ressalta que a inovação no setor de *bureaus*, incluindo a integração com o *Open Banking*, tem permitido às instituições financeiras expandir suas capacidades analíticas, oferecendo produtos mais personalizados e melhorando a experiência do consumidor.

A utilização das pontuações, por parte dos *bureaus*, permite uma abordagem padronizada na tomada de decisão, além de facilitar o acesso a dados essenciais para a definição de taxas e limites de crédito (Oliveira & Santos, 2022). A adoção dessas práticas, regulamentadas pelo BCB (2023), visa promover a segurança e estabilidade do sistema financeiro brasileiro (BCB, 2023). Essa classificação permite que as instituições financeiras avaliem a probabilidade de inadimplência dos clientes, facilitando a definição de limites e taxas de crédito (Marini & Manfrin, 2020). O uso de dados alternativos, como comportamento de consumo e histórico de pagamentos, tem se mostrado uma ferramenta apropriada para complementar as informações financeiras tradicionais e melhorar a precisão das avaliações de crédito (Lassance & Ternoski, 2021).

Grunert et al. (2005) e Altman et al. (2010) relatam a importância da inclusão de variáveis qualitativas, além das quantitativas, nos modelos de análise de crédito, principalmente para pequenas e médias empresas. Essa integração proporciona um entendimento mais abrangente da solvência e reduz os riscos de inadimplência. De maneira similar, Khemakehem e Boujelbene (2017) destacam a utilização não somente de variáveis financeiras, mas também as não financeiras, o que permite uma análise mais holística do perfil de crédito dos consumidores, especialmente para aqueles sem histórico financeiro formal, conhecidos como “invisíveis ao crédito”.

As condições de comprometimento de crédito são traçadas mediante os riscos, ponderando os fatores favoráveis e os desfavoráveis. As condições estão intrinsecamente ligadas ao cenário micro e macroeconômico, isso permite às instituições ter maior segurança ao conceder crédito a prazo, parcelado, no cheque, com boleto, crediário ou cartão próprio, reduzindo significativamente os riscos de inadimplência, conforme apontado por Lassance & Ternoski, 2021.

2.3 Utilização do aprendizado de máquina na análise de risco de crédito

Segundo Wanzeller et al. (2023), Big Data representa um vasto conjunto de dados que inclui informações estruturadas e não estruturadas de fontes como transações financeiras, mídias sociais e sensores IoT. Na concessão de crédito, o uso de Big Data possibilita a análise de dados diversificados, incluindo históricos tradicionais e padrões de comportamento (Jordan & Mitchell, 2015). A implementação de Big Data, Inteligência Artificial, *Data Mining* e aprendizado de máquina traz uma transformação substancial nesse processo, proporcionando uma análise robusta

(Timotio et al., 2024), e um sistema de crédito eficiente e adaptado às necessidades dos clientes e do mercado financeiro (Wanzeller et al., 2023).

A pontuação de crédito é principalmente um problema de classificação, ou seja, os candidatos a crédito devem ser enquadrados em uma classe resultante da probabilidade de *default*, derivada dos parâmetros definidos do Acordo de Basel II (Lessmann, 2015). A maior parte dos métodos de classificação empregados são divididos em duas categorias: estatística e aprendizado de máquina (Dumitrescu et al., 2022). Enquanto o método estatístico visa inferir a relação entre atributos, a prioridade do aprendizado de máquina é o desempenho preditivo (Montevechi et al., 2024), produzindo um grande impacto nos métodos de classificação habituais, beneficiando-se do tratamento de grandes quantidades de dados (Lessmann, 2015).

Dumitrescu et al. (2022) enfatizam que, no contexto da pontuação de crédito, os métodos *ensemble* baseados em árvores de decisão, entre eles, *random forest*, apresentam um maior desempenho do que modelos baseados em regressão logística. Por outro lado, os autores salientam que este último permanece sendo o *benchmark* na indústria do gerenciamento do risco de crédito, principalmente, porque a falta de interpretabilidade dos métodos *ensemble* é incompatível com os requerimentos dos reguladores financeiros (Dumitrescu et al., 2022).

Segundo Leo et al. (2019), o aprendizado de máquina apresenta a capacidade de detectar padrões significativos em dados, tornando-se uma ferramenta para qualquer tarefa confrontada com o requisito de extrair informações significativas de conjuntos de dados. De acordo com os autores, a crescente adoção do aprendizado de máquina tem sido impulsionada pelas oportunidades potenciais na redução de custos, bem como no aperfeiçoamento da produtividade e do gerenciamento de riscos.

Jordan e Mitchell (2015) apontam que o aprendizado de máquina pode identificar padrões complexos, proporcionando previsões mais precisas e que o uso combinado de diferentes tipos de dados não apenas melhora a avaliação de crédito, mas também contribui para uma decisão mais informada, promovendo a inclusão financeira e a redução do risco de crédito. De maneira similar, Lassance e Ternoski (2021) destacam que o uso dessas tecnologias na avaliação de risco é um avanço essencial para as instituições financeiras, por aprimorar a precisão das avaliações, otimizar processos e personalizar o atendimento ao cliente. Com isso, o sistema de concessão de crédito se torna mais eficiente, adaptado às necessidades do mercado financeiro (Marini & Manfrin, 2020).

Conforme destacado por Ambavat (2021), a adoção de técnicas avançadas como aprendizado de máquina permite a incorporação de uma ampla gama de variáveis não financeiras, resultando em modelos de risco mais robustos e abrangentes. A literatura recente tem enfatizado a importância de integrar tanto dados financeiros quanto não financeiros para criar uma visão holística do perfil de crédito dos consumidores. Grunert et al. (2005) e Altman et al. (2010) salientam que a inclusão de variáveis não financeiras (qualitativas) nos modelos de análise de crédito de pequenas e médias empresas conjuntamente a variáveis financeiras (quantitativas) melhora a acurácia destes modelos.

No estudo de Khemakhem e Boujelbene (2018), no contexto da solvência de empresas, é realizada a comparação da acurácia das técnicas de redes neurais artificiais (ANN) e árvores de decisão em um grupo de empresas tunisianas, levando em consideração variáveis financeiras e não financeiras. Os resultados obtidos pelos autores mostram que as árvores de decisão são mais eficientes que ANN em termos de predição de risco de crédito usando dados balanceados. Adicionalmente, o estudo contribui no entendimento do fenômeno de previsão do risco de crédito, a partir do conhecimento das variáveis e seu relacionamento com a variável dependente, que pode auxiliar os analistas financeiros numa melhor previsão do risco de crédito (Khemakhem & Boujelbene, 2018).

Lee e Shin (2020) descrevem abordagens distintas para o aprendizado de máquina: aprendizado supervisionado, não supervisionado e por reforço. Segundo os autores, no primeiro caso, o modelo é treinado com dados rotulados, auxiliando na categorização e previsão. A

classificação, nesse contexto, permite a categorização das observações, enquanto a predição sustenta a tomada de decisão. Lassance e Ternoski (2021) complementam que, na análise de risco de crédito, a regressão logística é amplamente utilizada para prever inadimplência, sendo o aprendizado supervisionado empregado em algoritmos como K-vizinhos, *Naive Bayes*, árvores de decisão, florestas aleatórias (*Random Forest*) e redes neurais.

Conforme Zöllner e Huber (2021), entre os métodos supervisionados, destacam-se a árvore de decisão, que, apesar de simples de interpretar, pode sofrer ajustes e as redes neurais artificiais, por outro lado, capturam relações complexas entre variáveis, enquanto a floresta aleatória (*Random forest*) aumenta a precisão preditiva ao combinar várias árvores de decisão. Segundo Lee e Shin (2020), o *Gradient Boosting Decision Trees (GBDT)* surge como uma técnica forte para aumentar a precisão em dados complexos. Em outro estudo, de Zhang & Song (2022), que investiga a avaliação de crédito para PMEs, é proposto um modelo baseado no algoritmo GBDT combinado com o algoritmo Rede Neural Convolutacional (CNN) e com a regressão logística (LR). A simulação do experimento foi realizada em uma amostra de 14.366 PMEs, conduzindo a um resultado que demonstra que o modelo GBDT-CNN-LR tem a melhor performance quando se compara com os métodos estatísticos individuais, por conta da acurácia destes modelos tradicionais de serem afetados pelo trabalho inicial de engenharia não apresentarem a mesma eficiência que o GBDT.

Diante disso, a Tabela 1 apresenta uma síntese das técnicas de aprendizado de máquina classificadas por tipos de aprendizado.

Tabela 1
Técnicas de aprendizado de máquina

Aprendizado supervisionado
<p>A Árvore de decisão é uma técnica popular, onde os dados são divididos em segmentos baseados em variáveis de entrada, criando um modelo simples de entender e interpretar. No entanto, as árvores de decisão podem sofrer sobreajuste, o que limita sua capacidade de generalização (Zöllner, & Huber, 2021).</p> <p>Redes neurais artificiais imitam o funcionamento do cérebro humano e são usadas para capturar relações complexas entre variáveis de entrada e saída. Elas são compostas por camadas de neurônios conectados, capazes de aprender padrões complexos, sendo especialmente úteis quando os dados possuem uma estrutura não linear.</p> <p>Florestas aleatórias (<i>Random Forest</i>), por sua vez, são um conjunto de árvores de decisão treinadas com diferentes subconjuntos de dados, aumentando a precisão e reduzindo o sobreajuste. Elas combinam as previsões de várias árvores para produzir uma previsão mais robusta e precisa (Avelar et al., 2022).</p> <p>O Support Vector Machine (SVM) é um método de aprendizado supervisionado que busca encontrar o hiperplano ótimo que separa as classes de dados com a maior margem possível. Ele é eficaz em espaços de alta dimensão e útil para casos em que o número de dimensões é maior que o número de amostras (Avelar et al., 2022).</p> <p>O Gradient Boosting Decision Trees (GBDT) é um método de <i>Boosting</i>, dentro da categoria de <i>Ensemble Learning</i>, que combina múltiplos modelos fracos, geralmente árvores de decisão simples, de forma sequencial, corrigindo os erros dos modelos anteriores. Esse método é poderoso para melhorar a precisão preditiva, especialmente em conjuntos de dados complexos e grandes (Avelar et al., 2022). Trata-se de uma técnica avançada de aprendizado supervisionado, combina múltiplos modelos fracos, geralmente árvores de decisão, para criar um modelo forte.</p>
Aprendizado não supervisionado
<p>O algoritmo trabalha com dados não rotulados e tenta identificar padrões ou estruturas intrínsecas nos dados. Técnicas como <i>clustering</i> (agrupamento) e redução de dimensionalidade são comuns aqui, sendo os algoritmos de k-means e análise de componentes principais (PCA) exemplos típicos.</p>

Fonte: Zöllner e Huber (2021) e Avelar et al. (2022).

Conforme relatado por O’Neil (2016), a evolução do aprendizado de máquina possui potencial transformador, especialmente na análise de crédito, mas também traz desafios, como o sobreajuste (*overfitting*), a transparência e a ética. A pesquisa contínua nessas áreas é crucial para garantir modelos confiáveis e aplicáveis ao contexto financeiro e com a crescente complexidade do mercado de crédito, técnicas avançadas como aprendizado de máquina têm sido integradas aos

modelos tradicionais de risco de crédito para aprimorar as previsões e identificar padrões de risco com maior precisão (Ambavat, 2021).

3 PROCEDIMENTOS METODOLÓGICOS

A classificação quanto aos objetivos de pesquisa é descritiva, a qual é uma categoria poderosa de pesquisa para descrever características, comportamentos ou fenômenos, procurando estabelecer relações entre variáveis (Gil, 2009).

A classificação quanto à natureza da pesquisa é qualitativa, em que o método empregado é o estudo de caso, escolha justificada pelo fato de tal abordagem proporcionar maior compreensão do fenômeno em sua abrangência e complexidade (Creswell & Poth, 2016). O estudo de caso é um método altamente reconhecido quando se busca uma compreensão profunda e detalhada de um fenômeno específico, de suas particularidades e que incluem categorias com relações complexas entre si (Godoy, 2006). Conforme Yin (2009), o estudo de caso proporciona uma análise intensiva de uma situação específica, buscando focar o caso em suas particularidades, com o propósito predominantemente descritivo da realidade pesquisada, ampliando os caminhos para a sua compreensão.

Conforme Yin (2009), os estudos de caso oferecem a melhor abordagem para investigar questões de pesquisa que buscam explicar "como" e "por que" um fenômeno ocorre, permitindo uma compreensão mais profunda e abrangente do tema em questão. O "como" tem natureza descritiva, quando os processos analisados originam relatos sobre os fatos observados e, como sugere Godoy (2006, p. 128), escolhe uma "unidade de análise para estabelecer as fronteiras de interesse do pesquisador". Com a delimitação do foco da pesquisa se define se o estudo ocorrerá sobre um caso único ou sobre múltiplos casos, estes últimos permitindo comparações e resultados mais robustos (Yin, 2009). Flick (2004) indica que "sempre devemos partir de um caso único, estudado em profundidade antes de realizarmos análises comparativas".

Conforme Stake (1995), um caso é estudado quando demonstra ser de interesse especial e busca-se detalhes da sua interação com o contexto. O estudo de caso "é o estudo da particularidade e complexidade de um único caso, chegando à sua compreensão dentro de circunstâncias importantes" (Stake, 1995, p. 11).

No estudo de caso, a investigação parte de uma unidade de análise específica, com a adoção de critérios predeterminados e múltiplas fontes de dados, como documentos e registros em arquivo normalmente quantitativos da empresa, e de órgãos relacionados, entrevistas com foco direcionado aos tópicos do estudo de caso, observação direta ou quando o observador participa por fazer parte do contexto sob estudo ou outros artefatos físicos (Yin, 2009, p. 102). Em vez de delimitar um fenômeno, o estudo de caso se revela mais eficaz para ampliar a compreensão sobre ele, permitindo uma análise aprofundada de suas nuances e complexidades. Segundo Stake (2000), a capacidade de integrar dados quantitativos provenientes de fontes diversas é um diferencial dos estudos de caso em relação a outras metodologias qualitativas, permitindo uma compreensão mais abrangente do fenômeno investigado. Segundo Yazan (2016), Robert Yin é uma referência no método de estudo de caso, com uma perspectiva objetiva, vista como positivista, procurando abordar no escopo do método tanto estudos qualitativos como também os quantitativos.

A diversidade de fontes de evidência busca mitigar as potenciais fragilidades apontadas por aqueles que questionam a validade do método. Stake (1995) defende a importância da triangulação para a validação dos dados. Denzin (1978) identificou 4 tipos básicos de triangulação:

- 1) triangulação de dados – o uso de uma variedade de fontes de dados em um estudo;
 - 2) triangulação para investigação – o uso de vários e diferentes pesquisadores;
 - 3) triangulação de teoria – o uso de múltiplas perspectivas para interpretar um único grupo de dados; e
 - 4) triangulação metodológica – o uso de métodos múltiplos para estudar um mesmo problema.
- Segundo Denzin (1978), a triangulação pode combinar métodos e fontes de coleta de dados

qualitativos e quantitativos (entrevistas, questionários, observação e notas de campo, documentos, além de outras).

Segundo Yin (2009, p. 49), a escolha por um único caso em estudos de caso pode ser indicada em determinadas circunstâncias, “justificadas sob cinco aspectos: caso crítico, peculiar, comum, revelador ou longitudinal”. Sendo assim, face a este embasamento metodológico, a empresa investigada nesta pesquisa representa um caso de interesse especial (Stake, 1995), exigindo um estudo de caso único em profundidade, com o objetivo de viabilizar uma análise rica do trabalho. Ainda, conforme Yin (2009, p. 49), o caso se enquadra no aspecto revelador, “por permitir o acesso a informações não facilmente disponíveis e a possibilidade de divulgação”, não comum no segmento financeiro. Yin (2009) pondera também que, embora estudos de caso único possam ser valiosos, é crucial ter cautela ao extrapolar suas conclusões para outros contextos, sob o risco de generalizações infundadas.

Corroborando Stake (1995), a unidade de análise desta pesquisa, o Banco BS2, é de interesse especial, também, em virtude da possibilidade de aprofundamento na compreensão de um sistema de gestão de riscos de crédito baseado em IA, especificamente, aprendizado de máquina, que por sua natureza técnica, é complexa. O intuito é o de se analisar os impactos com relação ao efeito da inadimplência, e que, normalmente, seria de grande dificuldade devido ao acesso aos dados.

3.1 O Banco BS2

O Banco BS2 (2024) foi fundado pela família Pentagna Guimarães como Banco Bonsucesso S.A. nos anos 90. De origem mineira, que em 2015 formou uma *joint venture* com o banco Santander constituindo o Banco Olé Consignado. O Banco Bonsucesso, em 2017, reposicionou-se no mercado focando em produtos digitais e mudando o nome para Banco BS2 S.A., com sua sede em Belo Horizonte (MG). Na transformação digital do banco, a plataforma tecnológica desenvolvida tem foco no atendimento ao público PMEs e *corporate* (Banco BS2, 2024).

A oferta de produtos do Banco BS2 está fundamentada em quatro pilares: (1) produtos de crédito, (2) soluções de câmbio, (3) cash management e (4) seguros, com uma parcela importante do negócio representada por serviços (Banco BS2, 2024). Em abril de 2024, o banco obteve um duplo *upgrade* pela Moodys Local (2024) passando de rating BBB+ para A, em função dos resultados e níveis de capitalização. Em dezembro de 2023, o BS2 reportou um total de ativos consolidado de R\$ 12,7 bilhões e um patrimônio líquido de R\$ 741 milhões. Em comparação com o ano de 2022, o banco registrou um crescimento de 61% no lucro líquido em 2023, alcançando R\$ 85 milhões. Além disso, houve um aumento significativo na carteira de crédito. O volume de câmbio cresceu 35%, enquanto as transações de *cash management* aumentaram 26%.

3.2 Coleta de Dados

Stake (1995) defende a importância da triangulação para a validação dos dados. As evidências empíricas foram coletadas por meio de uma triangulação de dados (Denzin, 1978) provenientes das técnicas de observação, entrevistas e análise documental dentro de um determinado método, no caso o estudo de caso, para coletar e interpretar os dados. A utilização de várias fontes de evidência busca evitar erros apontados por aqueles que são críticos ao método, a partir de análise de documentos internos da empresa (manuais técnicos, relatórios relativos a critérios adotados, variáveis financeiras selecionadas etc.) referentes ao sistema de gestão de riscos de crédito, observação direta e participativa (o pesquisador faz parte da realidade sob estudo) e por meio de entrevistas realizadas durante o mês de junho de 2024.

O instrumento de coleta de dados das entrevistas foi um roteiro semiestruturado, elaborado a partir do protocolo de pesquisa, baseado em estudos da literatura, com o objetivo de “dar ao entrevistador flexibilidade para ordenar e formular as perguntas durante a entrevista” (Godoi, Bandeira-de-Melo & Silva, 2010, p. 304), sendo aplicado ao *Head of Decision Science & Analytics* da empresa. Buscou-se captar em profundidade a perspectiva do ator pesquisado, tendo como objetivo obter uma compreensão detalhada das práticas e desafios na implementação de modelos de aprendizado de máquina no contexto do Banco BS2. As entrevistas foram gravadas e transcritas em arquivos. A análise destes dados foi providenciada mediante a técnica de análise de conteúdo.

A análise dos dados quantitativos foi conduzida posteriormente às entrevistas, em que foram utilizados os dados históricos de crédito do Banco BS2 que aplica métodos estatísticos e técnicas de aprendizado de máquina. Foi autorizado, por parte do Banco, o acesso aos dados e documentos por meio de um termo de consentimento. Com o intuito de avaliar qual técnica seria utilizada, métricas de desempenho, tais como Acurácia, F1 Score, Área Sob a Curva ROC (AUC-ROC), Precisão e *Recall* foram obtidas. Os algoritmos testados foram: Redes Neurais, Árvores de Decisão, Florestas Aleatórias, *Gradient Boosting Decision Trees* e Regressão Logística. O objetivo foi identificar o modelo mais adequado para prever o risco de crédito. Sendo assim, para a escolha do algoritmo que apresente as melhores métricas de desempenho, possibilitando uma análise comparativa entre as abordagens adotadas, foi necessário rodar, para cada modelo, a amostra selecionada no software utilizado neste estudo, o *Scikit Learn* (<https://scikit-learn.org/stable/>), ferramenta específica para aprendizado de máquina em Python. O algoritmo *Gradient Boosting Decision Trees (GBDT)* foi o escolhido como pode ser constatado na seção 5, referente à análise e à apresentação dos resultados.

3.2.1 Amostra utilizada para aferição do algoritmo

A amostra refere-se a dados de 10.000 empresas de pequeno e médio portes com contas abertas no Banco BS2, consideradas adimplentes e 700 empresas com *default* de crédito. A avaliação de risco de crédito para as pequenas e médias empresas é classificada pelo faturamento anual, sendo que entre R\$ 1.000.000 até R\$ 10.000.000,00 o BS2 considera empresas pequenas e empresas entre R\$ 10.000.000,00 até R\$ 30.000.000,00 de faturamento anual são classificadas como empresas médias, e o público total é de aproximadamente 150.000 clientes. Não é feita análise de pessoa física (PF), porque o Banco é focado em empresas e, portanto, não há carteira de PF.

3.2.2 Roteiro de Entrevista

O protocolo de entrevista foi elaborado com base nos estudos referenciados na Tabela 2 e como pode ser constatado, cobriu sete tópicos. Concentrou-se, principalmente, em perguntas relacionadas à técnica de aprendizado de máquina, tais como a escolha e implementação do algoritmo, preparação dos dados, métricas, integração dos modelos no sistema de decisão de crédito do banco, segurança, desafios e perspectivas.

Tabela 2

Roteiro de entrevista

Perguntas por tópicos	Artigos de referência
1 Contexto e Implementação de Aprendizado de Máquina (<i>Gradient Boosting Decision Trees -GBDT</i>)	Friedman (2001)
1.1 Quais foram os principais motivos que levaram o Banco BS2 a adotar o algoritmo GBDT para análise de crédito?	
1.2 Como foi o processo de implementação do algoritmo GBDT no sistema de crédito do banco?	

1.3. Quais foram os principais desafios enfrentados durante a implementação do GBDT?	
2. Seleção e Importância das Variáveis	Chen e Guestrin (2016)
2.1. Quais técnicas foram utilizadas pelo algoritmo GBDT para selecionar as variáveis mais importantes?	
2.2. Como o algoritmo GBDT ajudou na identificação das variáveis mais relevantes para o modelo de crédito?	
2.3. Como a inclusão de variáveis não financeiras impactou a precisão e eficácia do modelo de crédito?	
3. Indicadores de Desempenho do GBDT	Berrar (2019)
3.1. Quais métricas de desempenho (por exemplo, taxa de precisão, recall, F1-score, AUC-ROC) foram usadas para avaliar a eficácia do modelo?	
3.2. Quais foram os resultados mais significativos obtidos através dessas métricas?	
3.3. Como o modelo lida com o desbalanceamento de dados nas amostras?	
4. Benefícios e Resultados Práticos	Lessmann et al. (2015)
4.1. Quais foram os principais benefícios observados após a implementação do GBDT na análise de crédito?	
4.2. Como o uso desse algoritmo melhorou a capacidade do banco de identificar e mitigar riscos de crédito?	
4.3. Pode fornecer exemplos específicos de como a análise em tempo real afetou as políticas de crédito do banco?	
5. Feedback e Perspectivas Futuras	Ribeiro, Singh e Guestrin (2016)
5.1. Qual foi o feedback dos analistas de crédito e outras partes interessadas sobre o uso do GBDT?	
5.2. Existem planos para incorporar outras técnicas de IA e ML na análise de crédito no futuro?	
5.3. Quais melhorias ou ajustes adicionais são planejados para otimizar ainda mais o modelo de crédito?	
6. Considerações Éticas e Regulatórias	Wachter, Mittelstadt e Floridi, 2017
6.1. Como o banco assegura a conformidade com as regulamentações financeiras ao usar modelos de IA e ML?	
6.2. Quais medidas foram adotadas para garantir a transparência e a ética na utilização de dados dos clientes?	
6.3. Houve alguma preocupação relacionada à privacidade dos dados durante a implementação do modelo? Se sim, como foram abordadas?	
7. Integração e Treinamento da Equipe	Davenport e Harris (2007)
7.1. Como foi o processo de integração da equipe de dados com as demais áreas do banco?	
7.2. Quais treinamentos e capacitações foram oferecidos aos funcionários para que pudessem utilizar e interpretar os resultados do GBDT de maneira eficaz?	
7.3. Quais foram as principais dificuldades enfrentadas pela equipe durante a transição para o uso desse novo modelo?	

3.3 Preparação dos Dados

3.3.1 Seleção de variáveis e as métricas de desempenho do algoritmo

A seleção das variáveis é uma etapa essencial para a realização de um modelo preditivo, pois garante a qualidade e a validade dos resultados obtidos, buscando trazer a significância de cada variável para o modelo desenvolvido (vide item 5.1.2).

A análise de desempenho de algoritmos de aprendizado de máquina, especialmente em contextos de risco de crédito, exige o uso de métricas específicas que avaliam a eficácia do modelo em prever eventos críticos, como inadimplência. Entre as métricas mais utilizadas estão a acurácia, precisão, *recall* e F1-score, que, segundo Provost e Fawcett (2001), são fundamentais para

identificar a robustez de classificadores em cenários de dados desbalanceados, comuns no setor financeiro. Para garantir uma análise detalhada, o presente estudo aplicou para os algoritmos selecionados pelo Banco BS2, o software Scikit-learn, uma biblioteca amplamente reconhecida em Python por sua capacidade de facilitar o desenvolvimento e a avaliação de modelos de aprendizado de máquina (Pedregosa et al., 2011). Cada um dos modelos foi ajustado com base em variáveis financeiras e não financeiras previamente identificadas como relevantes no artigo, ponderadas conforme seu impacto na previsão de risco, proporcionando uma visão abrangente sobre a performance dos algoritmos testados.

A eficácia de um modelo de classificação, tradicionalmente, é medida pela acurácia, a qual fornece a proporção de mutuários corretamente classificados (verdadeiros positivos e verdadeiros negativos) sobre o conjunto total de mutuários (verdadeiros positivos e verdadeiros negativos, falsos positivos e falsos negativos), apesar de que esta medida pode ser ineficaz quando é desbalanceada (Khandani et al., 2010). As métricas de desempenho avaliadas incluíram Acurácia, F1 Score, Área Sob a Curva ROC (AUC-ROC), Precisão e *Recall* ou Sensibilidade. O objetivo é identificar o modelo mais adequado para prever o risco de crédito.

Estudos com o uso da acurácia em situações de dados desbalanceados têm provocado discussões (Provost, F & Fawcett, T., 2001, Sun et al., 2007). A partir da matriz da confusão, o desempenho pode ser obtido (vide Tabela 3). Assim, os elementos ao longo da diagonal principal representam as decisões corretas: número de verdadeiros negativos (VN) e verdadeiros positivos (VP); e os elementos que se encontram fora dessa diagonal, representam os erros cometidos: número de falsos positivos (FP) e falsos negativos (FN) (Castro & Braga, 2011).

Tabela 3

Matriz de Confusão para um classificador binário.

	predição (y = 0)	predição (y = 1)
real (y = 0)	VN	FP
real (y = 1)	FN	VP

Luque et al. (2019) salientam que métricas essenciais podem ser extraídas da matriz de confusão, conforme descritas na Tabela 4.

Tabela 4

Medidas de desempenho do modelo

Acurácia ou Taxa de Casos Corretamente Classificados	Indica a performance geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente. $Acurácia = \frac{VP + VN}{VP + VN + FP + FN}$
Precisão ou Taxa de Positivos Previstos	Dentre todas as classificações de classe Positivo que o modelo fez, quantas estão corretas. $Precisão = \frac{VP}{VP + FP}$
Recall/Sensibilidade/Revocação ou Taxa de Verdadeiros Positivos	Dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas. $Recall = \frac{VP}{VP + FN}$
Especificidade	Mede a proporção de casos negativos corretamente classificados por negativos $Especificidade = \frac{VN}{VN + FP}$
F1-Score	Média harmônica entre precisão e <i>recall</i> . $F1\ Score = 2 * \frac{Precisão * Recall}{Precisão + Recall}$

Já, a Curva ROC (*Receiver Operating Characteristic*) é advinda também da Matriz de Confusão, gerada em um plano gráfico e representando o trade off da relação entre os indicadores de Sensibilidade e Especificidade. A Curva ROC demonstra uma estimativa da capacidade discriminativa do classificador em termos de probabilidade de erro (Castro & Braga, 2011). A AUC – *Area Under Curve* fornece uma métrica para avaliar, em média, qual o melhor algoritmo, de forma que quanto maior a área entre a Curva ROC e a diagonal principal, melhor performance do modelo.

Segundo Botelho e Tostes (2011), o Teste Kolmogorov-Smirnov (KS) consiste em uma estatística não paramétrica que tem como objetivo testar se as distribuições de dois grupos são iguais, sendo que em problemas de classificação é frequentemente usado para medir a capacidade de um modelo de classificação de distinguir entre duas classes. Além disso, o valor máximo do KS indica o ponto de corte onde o modelo faz a melhor distinção entre as duas classes (Botelho & Tostes, 2011).

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Nesta seção são apresentados os resultados advindos da triangulação dos dados coletados nas narrativas das entrevistas, na observação direta e participativa, com as respectivas notas de campo, e nos documentos levantados, proporcionando elementos suficientes para fomentar a análise conjunta.

4.1. Preparação dos Dados

4.1.1. Escolha do algoritmo Gradient Boosting Decision Trees (GBDT)

A Tabela 5 apresenta os resultados das métricas de desempenho para cada modelo, possibilitando uma análise comparativa entre as seguintes abordagens adotadas: redes neurais, árvore de decisão, floresta aleatória, GBDT e regressão logística. O GBDT apresentou os melhores resultados, contudo as diferenças observadas em relação às métricas aplicadas não geraram diferenças estatisticamente significativas quando comparadas com os outros algoritmos.

Tabela 5

Resultados estatísticos dos algoritmos

Algoritmos	Acurácia	Precisão	Recall	F1 - Score
Redes Neurais	0,95	0,85	0,79	0,78
Árvore de decisão	0,94	0,84	0,76	0,75
Floresta aleatória	0,96	0,90	0,80	0,81
GBDT	0,97	0,92	0,83	0,83
Regressão Logística	0,95	0,85	0,79	0,78

Conforme elucidado pelo entrevistado, com relação **ao contexto e seleção de Aprendizado de Máquina (*Gradient Boosting Decision Trees -GBDT*)**:

Após a realização dos testes com os algoritmos redes neurais, árvore de decisão, floresta aleatória, GBDT e regressão logística, o GBDT apresentou os melhores resultados no que tange às métricas. Adicionalmente, o GBDT é conhecido por sua alta precisão e robustez, especialmente em tarefas de classificação e regressão. Ele combina múltiplos modelos fracos, geralmente árvores de decisão, para criar um modelo forte, resultando em melhor desempenho preditivo. Além disso, a interpretabilidade das árvores de decisão facilita a compreensão dos fatores que influenciam as decisões de crédito, o que é crucial para conformidade regulatória e transparência.

Além do resultado obtido, o algoritmo GBDT, conforme Sun et al. (2007), também foi escolhido pela razão de ser conhecido ao testar os vários conjuntos de dados para determinar sua capacidade de prever inadimplências, mitigar riscos e fornecer a medição precisa de métricas como a taxa de precisão, *recall*, *F1-score* e a área sob a curva ROC (*Receiver Operating Characteristic*).

Esclarecendo, o método *boosting* apresenta n árvores de decisões com amostras aleatórias, porém elas não são independentes, pois a aprendizagem é sequencial. Cada árvore é treinada de forma a minimizar o erro cometido pelas árvores anteriores. Uma extensão dos algoritmos de *boosting* é o GBDT, porém o erro é minimizado utilizando os erros residuais das árvores anteriores. Ou seja, o GBDT baseia-se no treinamento de várias árvores de decisão de forma sequencial, onde cada nova árvore busca corrigir os erros das anteriores. Outro fator que ratifica a escolha do GBDT é que, segundo Lee e Shin (2020), o *Gradient Boosting Decision Trees (GBDT)* surge como uma técnica forte para aumentar a precisão em dados complexos.

Para rodar a modelagem da análise de crédito com a utilização do algoritmo GBDT, no banco BS2 selecionou-se aproximadamente 10.000 amostras de clientes adimplentes e 700 clientes inadimplentes. O banco considera um cliente inadimplente um cliente que possui pelo menos um recebível com R\$ 100 de principal em aberto e os recebíveis que estiverem em atraso por 60 dias após a data de vencimento. **Com relação ao contexto e à implementação do GBDT no Banco BS2**, o entrevistado informou que envolveu 06 etapas:

1. Realização de uma análise de requisitos para identificar as necessidades específicas do modelo de crédito;
2. Limpeza e pré-processados para garantir a qualidade e a consistência dos dados;
3. Treinamento: utilização de um conjunto de dados históricos de crédito para treinar o algoritmo;
4. Aplicação de técnicas como validação cruzada para evitar *overfitting*;
5. Validação e teste do modelo para avaliar seu desempenho;
6. Integração da modelagem ao sistema de crédito do banco, com monitoramento contínuo para ajustes e melhorias.

Ainda conforme o entrevistado, alguns desafios foram enfrentados durante a implementação do GBDT, tais como, “a necessidade de lidar com grandes volumes de dados e as múltiplas de fontes de dados que o Banco BS2 acessa para análise de crédito”. Outro desafio significativo foi o desbalanceamento das classes no conjunto de dados, onde a maioria dos registros correspondia a clientes adimplentes. Além disso, houve “a necessidade de ajustar os hiperparâmetros do modelo para otimizar seu desempenho, o que exigiu tempo e recursos computacionais consideráveis”.

Com relação ao desbalanceamento dos dados, o entrevistado enfatizou:

O modelo utilizou técnicas de reamostragem, como *oversampling* das classes minoritárias ou *undersampling* das classes majoritárias, para lidar com o desbalanceamento de dados. Além disso, foram aplicadas técnicas de ajuste de pesos nas classes durante o treinamento do modelo para dar mais importância às classes minoritárias. Essas abordagens ajudaram a garantir que o modelo não fosse tendencioso em favor da classe majoritária, melhorando a capacidade de prever corretamente as instâncias da classe minoritária.

4.1.2 A seleção e aferição de importância das variáveis pelo GBDT

A seleção das variáveis para a análise de crédito no Banco BS2 foi conduzida com rigor acessando múltiplas fontes de dados: 04 *bureaus* de crédito, informações internas, informações públicas e pesquisa web reputacional, essencial para capturar um panorama completo para análise de crédito e para expressar de maneira técnica e científica, após a coleta de dados, um conjunto extenso de variáveis foi inicialmente considerado para a análise que após processo de limpeza e tratamento dos dados faltantes ou incompletos, seguiram para a seleção das amostras. O processo envolveu técnicas intrínsecas do algoritmo para determinar a importância das variáveis e garantir que as mais relevantes fossem incluídas no modelo final.

Com relação à seleção e a importância das variáveis, o entrevistado informou:

O algoritmo GBDT utiliza um processo iterativo para selecionar variáveis significativas, baseando-se em sua contribuição para a redução do erro no modelo. Durante o treinamento, o algoritmo calcula a importância das variáveis com base na melhoria que cada variável

proporciona às divisões em cada árvore de decisão. A métrica de importância das variáveis pode ser analisada através da soma das reduções dos critérios de divisão (como Gini ou entropia) para cada variável em todas as árvores da floresta. Essas importâncias são então normalizadas, proporcionando uma medida relativa da relevância de cada variável no modelo.

Ou seja, para um maior esclarecimento técnico, durante o treinamento, o algoritmo cria várias árvores de decisão, onde cada árvore corrige os erros das anteriores. Em cada nó de decisão, o algoritmo escolhe a variável que mais reduz o erro preditivo. A importância de cada variável é então medida pela soma das reduções nos critérios de divisão através da entropia ao longo de todas as árvores. A significância das variáveis é calculada com base na melhoria que cada variável proporciona às divisões em cada árvore de decisão. Isso é feito somando a redução de erro proporcionada por cada variável ao longo de todas as árvores do modelo. As variáveis com maiores somas são consideradas mais importantes. Este processo permite identificar rapidamente quais variáveis têm maior impacto na previsão do risco de crédito, proporcionando um modelo mais eficiente e preciso.

A seleção das variáveis é uma etapa essencial para a realização de uma pesquisa quantitativa robusta, pois garante a qualidade e a validade dos resultados obtidos. Para identificar as variáveis mais relevantes e significativas para o modelo o algoritmo GBDT, as variáveis financeiras e não financeiras foram selecionadas mediante a significância extraída, obtidas por regressão múltipla e relacionadas na Tabela 6.

Com relação ao impacto da inclusão de variáveis não financeiras na precisão e eficácia do modelo de crédito, o entrevistado salientou:

A inclusão de variáveis não financeiras aumentou a precisão e a eficácia do modelo ao fornecer uma visão mais holística do perfil de crédito dos clientes. Essas variáveis adicionaram uma camada de informações que complementa os dados financeiros tradicionais, permitindo ao modelo capturar aspectos comportamentais e socioeconômicos dos clientes. Como resultado, o modelo tornou-se mais robusto e capaz de prever com maior precisão o risco de inadimplência, especialmente em casos onde as variáveis financeiras sozinhas não eram suficientes para uma avaliação completa.

Esta constatação (do entrevistado) vem de encontro com estudos tais como Ambavat (2021), Grunert et al. (2005) Altman et al. (2010), que ratificam a adoção de técnicas de aprendizado de máquina permitindo a incorporação de uma ampla gama de variáveis não financeiras, resultando em modelos de risco mais robustos e abrangentes, melhorando a acurácia destes modelos.

Tabela 6

Variáveis financeiras e não financeiras utilizadas no modelo com as respectivas significâncias Extraídas pelo Scikit Learn

Variável	Signif. (%)	Descrição
general_lost_pct_risco	32,27	Representa % de risco em inadimplência (vencido + prejuízo) geral. É uma métrica direta de risco que reflete a probabilidade de inadimplência, sendo crucial para prever o comportamento de crédito.
years_since_first_relationship (não financeira)	8,84	Representa o tempo desde o início do relacionamento bancário , clientes com relacionamentos mais longos tendem a ter um histórico mais confiável, mais sólido e previsível, reduzindo a incerteza na análise de crédito.
chk_esp_cred_rot_a_vencer_pct_risco_max_prev_10m	8,73	Indica o máximo de cheque especial e crédito rotativo a vencer nos próximos 10 meses . Altos valores podem

		indicar problemas de liquidez e maior risco de inadimplência.
cnt_declined_reasons (não financeira)	7,21	Indica a quantidade de recusa de crédito , sendo que um alto número de recusas pode sinalizar problemas de crédito recorrentes, pois múltiplas recusas de crédito estão associadas a um maior risco de default.
general_limite_do_cred_360m_sum	5,47	Refere-se à soma dos limites de crédito até 365 dias , tomados em outras instituições. Este índice reflete a capacidade de acesso a crédito, influenciando a liquidez e o risco financeiro. Limites de crédito elevados podem sinalizar comprometimento de capacidade financeira quanto exposição ao risco.
days_since_oldest_board_member (não financeira)	5,24	Mensura a quantidade de dias desde a última atualização do status financeiro dos membros do conselho . Frequentemente, empresas com atualizações regulares tendem a ser mais transparentes, demonstrando gestão adequada, governança diligente e bons controles internos, reduzindo o risco de inadimplência.
v001r_social_capital	5,04	Este indicador refere-se ao capital social da empresa , sendo que um capital social elevado pode indicar uma base financeira sólida, logo, tendem a ter maior capacidade de enfrentar adversidades financeiras, reduzindo o risco de default.
chk_esp_cred_rot_a_vencer_pct_risco	5,02	Cheque especial + crédito rotativo como % do risco total.
min_lj_ij3_prev_6m	2,84	Taxa de juros mínima de longo prazo estimada para empréstimos de mais de 360 dias (últimos 10 meses). Taxas de juros mais baixas podem indicar condições de crédito mais favoráveis, sinaliza a capacidade da empresa de acessar crédito a custos mais baixos.
general_limite_do_cred_sum_pct_risco	2,62	Refere-se à soma dos limites de crédito até 365 dias , tomados em outras instituições. Este índice reflete a capacidade de acesso a crédito, influenciando a liquidez e o risco financeiro. Limites de crédito elevados podem sinalizar comprometimento de capacidade financeira quanto exposição ao risco.
v254r_cnt_board_members (não financeira)	2,42	Refere-se ao número de membros do conselho . Um conselho maior pode indicar uma estrutura de governança mais robusta. E boa governança está associada a menor risco de <i>default</i> devido, à melhor tomada de decisões, melhor supervisão e mitigar riscos de agência.
avg_final_number_prediction_prev_6m	2,38	Previsão média dos últimos 10 meses.
days_since_stopped_simples	2,35	Dias desde que o status SIMPLES foi interrompido.
general_limite_de_cred_sum	2,17	Limite total de crédito.
general_limite_trend_cred_sum	1,90	Crescimento do limite de crédito nos últimos 10 meses.
v204r_is_popular_domain	1,80	Indicador de domínio de e-mail de receita popular (Gmail, Yahoo)
cnt_late_not_to_calc_interest	1,49	Contagem de prestações irregulares nos últimos 10 meses.
special_interest	1,37	Taxa de juros média em todos os tipos de empréstimo.
primary_median_ij3_prev_6m	1,01	Taxa de juros da maior exposição nos últimos 10 meses.

4.2 Treinamento / validação cruzada / Teste do GBDT

A definição clara de inadimplência (*default*) ajuda a garantir consistência na rotulagem dos dados. A tabela 7 descreve os dados utilizados para treinar o modelo de análise de crédito, destacando a distribuição das operações e das inadimplências. Sendo assim, a divisão dos dados foi realizada em subconjuntos de treino, validação cruzada (CV) e teste permite a avaliação do modelo em diferentes fases de desenvolvimento, garantindo sua força e capacidade de generalização. Cada subconjunto contém informações sobre o número de operações e a taxa de

inadimplência. A diminuição das operações de treinamento para validação cruzada e teste é um reflexo de práticas comuns de divisão de dados em aprendizado de máquina. Durante o pré-processamento, algumas operações podem ser filtradas ou excluídas devido à falta de dados, inconsistências ou outliers. Esse processo de limpeza pode reduzir o número total de operações disponíveis para validação e teste.

Tabela 7

Divisão em Treinamento / Teste dos dados

2020 - 09	Operações	Defaults	Taxa de Default
Dados de Treinamento	5469	374	6,8%
Validação Cruzada (CV)	2344	160	6,8%
Dados de Teste	1908	161	8,4%

Para garantir que a distribuição de classes (por exemplo, taxa de *default*) seja representativa em cada subconjunto, a estratificação pode ser usada. Isso garante que a proporção de defaults seja mantida consistente entre os conjuntos de treinamento, validação e teste, o que pode resultar na exclusão de algumas operações para manter essa consistência. O conjunto de treinamento tem mais dados do que os de validação e teste combinados. Isso é uma prática comum, pois o treinamento exige mais dados para que o modelo aprenda padrões robustos.

Conforme a Tabela 7, os pontos de análise dos dados são os relacionados a seguir: 1. Taxa de default consistente no treinamento e validação: a taxa de inadimplência no conjunto de treinamento e validação cruzada é exatamente igual (6,8%), o que é bom para garantir que o modelo está sendo ajustado em dados com características semelhantes; 2. Taxa de default nos dados de teste: a taxa de default no conjunto de teste (8,4%) é maior do que nos outros conjuntos. Isso pode sugerir que o conjunto de teste representa um período diferente ou um segmento com maior risco de crédito, ou seja, uma maior variabilidade ou dificuldade nos dados não vistos. De certa forma, isto é esperado e desejável para uma avaliação robusta do modelo. Também pode sinalizar uma limitação do modelo em generalizar perfeitamente para novos dados.

4.3 Sistema de Pontuações de scores utilizados pelo BS2

O BS2 utiliza 02 modelos de pontuação para análise das amostras elegíveis: Aplicação, com dados fornecidos pelo bureau Serasa e o Comportamental do BS2. Os sistemas de pontuações Aplicação e Comportamental sustentam o modelo para o monitoramento contínuo e ajustes de linha de crédito ao longo da duração da concessão do crédito para o cliente.

A pontuação Aplicação (Tabela 8) é utilizada na pré-triagem para identificar clientes com potenciais atrasos superiores a 60 dias com dados fornecidos pelo bureau de crédito-Serasa. A pontuação Comportamental (Tabela 8) é utilizada para identificar clientes com tendência a acumular atrasos significativos maior ou igual a 10% de atraso em todas as linhas de crédito bancária, para esta análise o foco é no bureau de crédito do Bacen – Sistema de Informação de Crédito - SCR. Esta pontuação é utilizada no monitoramento do cliente, na tomada de decisão em linha e final. A avaliação das duas pontuações é baseada em 4 indicadores: F1-score, indicador KS, indicador KS-Analog e a área sobre a curva ROC (AUC ROC). A avaliação do sistema é baseada no índice F1 para o treinamento, validação e no monitoramento da performance mensal e o modelo também recebe validação externa.

Tabela 8
Métricas das Pontuações Aplicação e Comportamental extraídas pelo modelo

Pontuação Aplicação		Pontuação Comportamental	
Métrica	Valor	Métrica	Valor
F1 Score	0,66	F1 Score	0,77
KS	0,783	KS	0,691
KS Analog	0,766	KS Analog	0,742
AUC ROC	0,94	AUC ROC	0,919

Com base na Tabela 8, a explicação das métricas extraídas do modelo Aplicação é:

- Índice F1, uma métrica de desempenho que combina precisão e *recall*, indicando um equilíbrio entre essas duas medidas (0,66), quanto mais próximo de 1 maior a precisão;
- Indicador KS (Kolmogorov-Smirnov), para medir a diferença máxima entre as distribuições acumuladas de duas populações, neste caso os bons e maus pagadores, quanto mais próximo de 1 menor é a diferença entre as classes (0,783);
- KS Analog: análoga ao KS, utilizada para validações internas (0,766), além de validar o KS, consegue diminuir as diferenças entre as classes;

AUC ROC: A área sob a curva ROC (*Receiver Operating Characteristic*) indica a capacidade do modelo de distinguir entre classes positivas (verdadeiros positivos) e (falsos positivos), com valores próximos de 1.0 representando alta eficácia do modelo. A curva ROC e a AUC são utilizadas para avaliar a capacidade do modelo de discriminar as classes. Uma AUC ROC de 0,94 indica excelente capacidade de discriminação.

A Tabela 9 apresenta as métricas e as características mais importantes utilizadas pelo modelo Aplicação, junto com suas respectivas importâncias e o valor acumulado. O modelo apresenta as métricas de desempenho: precisão, *recall* (ou sensibilidade), pontuação F1 e a quantidade de ocorrências para as classes “Falso” e “Verdadeiro”. A precisão é muito alta para a classe “Falso” (0,98), o que significa que o modelo é excelente em identificar corretamente os casos negativos (não inadimplentes). O *recall* ou sensibilidade da classe “Verdadeiro” (0,71) indica que o modelo também é bom em identificar os casos positivos (inadimplentes).

Tabela 9
Métricas do Modelo Aplicação

Classificação	Precisão	<i>Recall</i>	F1-Score
False	0,98	0,97	0,97
True	0,61	0,71	0,66
<i>Macro Avg</i>	0,80	0,84	0,82

Com relação aos resultados da modelagem Aplicação para as métricas de desempenho: precisão, *recall* (ou sensibilidade), pontuação F1 e a quantidade de ocorrências para as classes “Falso” e “Verdadeiro”, demonstradas na Tabela 9, pode-se verificar que:

- A precisão é muito alta para a classe “Falso” (0,98), o que significa que o modelo é excelente em identificar corretamente os casos negativos (não inadimplentes);
- O *recall* ou sensibilidade da classe “Verdadeiro” (0,71) indica que o modelo também é bom em identificar os casos positivos (inadimplentes);
- *F1-Score macro average* é utilizada para avaliar a capacidade do modelo de discriminar as classes. O resultado *macro average* foi de 0,82 indicando alta capacidade de discriminação.

4.3.1 Avaliação da Pontuação Comportamental (BS2)

A pontuação Comportamental permite o monitoramento contínuo dos clientes durante o relacionamento com o banco, ao contrário da pontuação Aplicação, que é baseada apenas nas

informações iniciais. Isso ajuda a captar mudanças no perfil de risco ao longo do tempo. Adicionalmente o modelo Comportamental utiliza dados atualizados do cliente, como movimentações de conta, uso de crédito rotativo e histórico de pagamentos. Esses dados são mais abrangentes e refletem de forma mais clara a situação financeira atual do cliente.

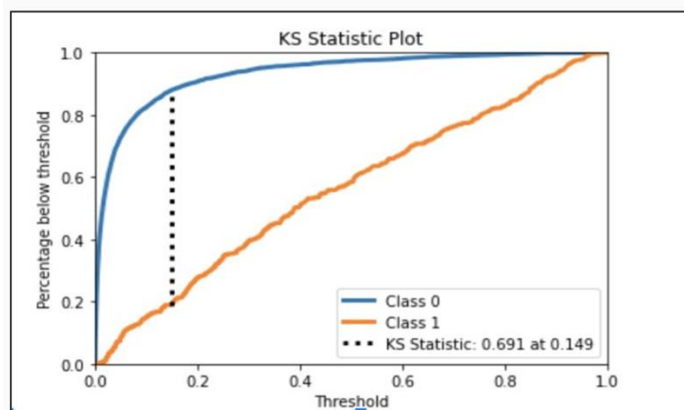
A explicação das métricas extraídas do modelo Comportamental (Tabela 8) é:

- F1: 0.77 - um índice F1 mais alto que o da pontuação de aplicação, indicando melhor equilíbrio entre precisão e recall para este tipo de pontuação;
- KS (Kolmogorov-Smirnov): 0.691, ainda uma boa métrica, mas ligeiramente inferior ao KS da pontuação de aplicação;
- KS Analog: 0.742, similar ao KS, utilizado para validações internas;
- AUC ROC: 0.919, indica alta eficácia na distinção entre bons e maus comportamentos financeiros, embora ligeiramente inferior ao AUC ROC da pontuação de aplicação.

A pontuação Comportamental é baseada no histórico de comportamento do cliente, como pagamentos, inadimplência e movimentações financeiras, enquanto a pontuação Aplicação é focada em informações fornecidas no momento da solicitação de crédito. O BS2 decidiu pelo modelo Comportamental, com melhor F1 score (0,77), indicando que ele é mais apurado para prever corretamente tanto os clientes adimplentes quanto os inadimplentes.

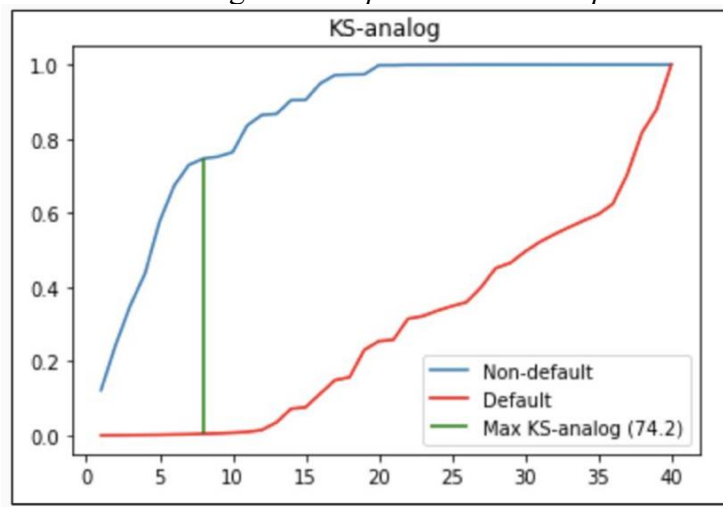
Para comparar a distribuição acumulada das pontuações de duas classes, o BS2 utiliza a estatística KS (Figura 1), sendo: “Classe 0” provavelmente bons pagadores e “Classe 1” maus pagadores. A curva azul representa a distribuição cumulativa da “Classe 0”. A curva laranja representa a distribuição cumulativa da “Classe 1” (Figura 1). A linha tracejada vertical representa a máxima diferença entre as duas distribuições, que é a estatística KS. Conforme demonstrado na Figura 1, a estatística KS é 0,691, indicando uma diferença significativa entre as distribuições das duas classes, sugerindo que o modelo é eficaz em distinguir entre bons e maus pagadores, pois há uma clara separação entre as distribuições das pontuações das duas classes.

Figura 1
Métrica KS extraída pelo modelo Comportamental



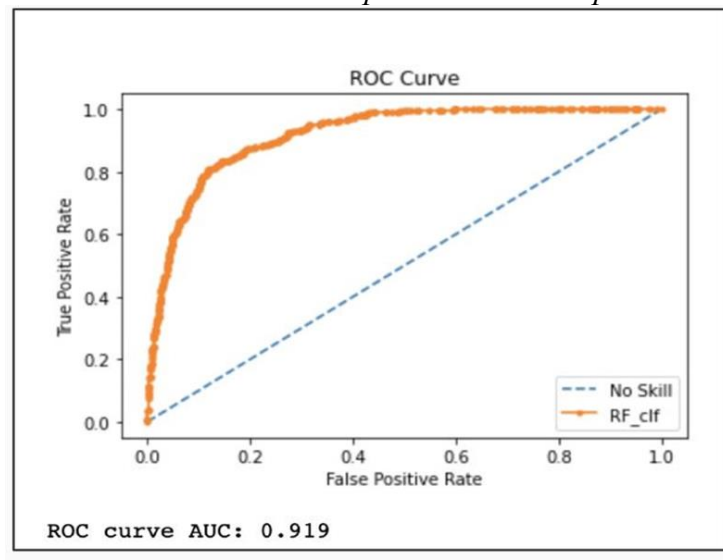
A linha verde vertical na Figura 2 representa o máximo KS análogo, que é 0,742. Assim como o gráfico KS tradicional, o KS análogo mostra a máxima diferença entre as duas distribuições. O valor alto 0,742 sugere uma boa separação entre inadimplentes e adimplentes, indicando que o modelo é eficaz em discriminar esses dois estados.

Figura 2
Métrica KS Analog extraída pelo modelo Comportamental



A Figura 3 descreve a performance do modelo de classificação ao plotar a taxa de verdadeiros positivos (TPR) contra a taxa de falsos positivos (FPR) para diferentes limiares de decisão. A AUC de 0,919 indica que o modelo tem uma excelente capacidade de discriminar entre as classes (bons e maus pagadores). Quanto mais próximo de 1 for o valor da AUC, melhor é a performance do modelo.

Figura 3
Métrica AUC ROC extraída pelo Modelo Comportamental



Em suma, os resultados na Tabela 10 indicam que, apesar da Pontuação Comportamental apresentar índices KS e AUC ROC mais baixos do que a Pontuação Aplicação, o modelo de crédito registra um bom desempenho em distinguir entre bons e maus pagadores, com alta separação entre as distribuições de pontuação (alta estatística KS e KS análogo), bem como boa capacidade discriminativa, conforme indicado pela AUC da curva ROC.

Tabela 10

Métricas das Pontuações Aplicação e Comportamental do BS2

	KS	KS Analog	F1	AUC (área sobre curva ROC)
Aplicação	0,783	0,766	0,66	0,94
Comportamental	0,691	0,742	0,77	0,919

Com relação **aos resultados mais significativos obtidos através das métricas**, o entrevistado salientou que:

Ao realizar a comparação dos modelos, mesmo que a Pontuação Comportamental tenha apresentado índices menores que a Pontuação Aplicação, os resultados demonstraram que o modelo GBDT alcançou uma boa precisão e recall, indicando que foi eficaz tanto na identificação de clientes adimplentes quanto inadimplentes. O F1-score forneceu uma medida balanceada do desempenho do modelo, demonstrando um bom equilíbrio entre precisão e recall. A AUC-ROC também foi alta, sugerindo que o modelo possui uma forte capacidade discriminativa das classes. Esses resultados significativos indicam que o modelo é robusto e confiável para a tomada de decisões de crédito.

Estes resultados vêm ao encontro com o descrito na literatura, que conforme Zhou et al. (2019), evidenciam que a métrica F1 é um critério comum na recuperação de informação para avaliar o desempenho de um modelo de classificação. Tanto esta métrica como também a AUC-ROC, que evidencia um valor alto. A predileção de se basear nestas métricas é que apesar de muitos estudos utilizarem a medida da acurácia para avaliar o desempenho do modelo, Khemakhem e Boujelbene (2018) apontam que esta medida pode conduzir ao comportamento enviesado se utilizar dados desbalanceados, implicando a uma má seleção de um modelo de previsão.

4.4 Inadimplência do Banco BS2 de acordo com a regra BCB 2682

A Resolução n.º 2682 do BCB estabelece diretrizes para a classificação de risco de crédito e o correspondente provisionamento para perdas associadas (BCB, 1999). Esta medida visa garantir que as instituições financeiras mantenham uma reserva adequada de capital para cobrir potenciais perdas, promovendo assim a estabilidade do sistema financeiro. A classificação de risco é categorizada em oito níveis de A a H, com cada nível representando de forma crescente o risco de inadimplência. Ou seja, a categoria A é considerada de baixo risco, exigindo um provisionamento de apenas 0,5%, enquanto a categoria H, que representa o maior risco, exige provisionamento integral de 100%.

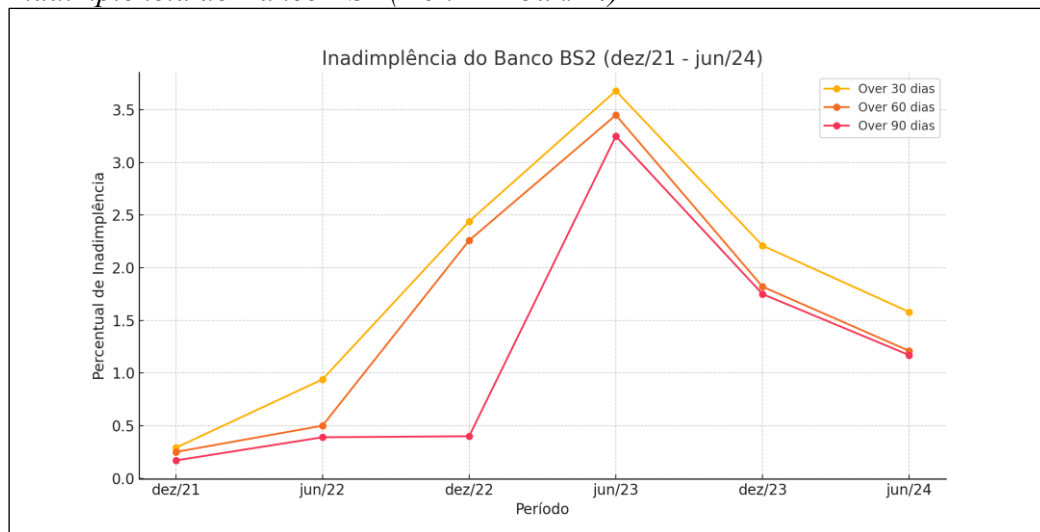
As instituições financeiras são obrigadas a revisar periodicamente suas carteiras de crédito e ajustar as provisões conforme necessário, com base na reavaliação do risco de inadimplência dos devedores. Esta prática é essencial para assegurar que os balanços das instituições reflitam de forma precisa o risco de crédito e estejam preparados para absorver eventuais perdas. A Tabela 11 apresenta os dados de Inadimplência do Banco BS2 de Dez/21 a Jun/24, representados na Figura 4, que demonstram uma queda na taxa de inadimplência (default), após Jun/23, quando já havia passado 6 meses da data da adoção do novo modelo de análise de crédito com algoritmo GBDT, apresentando os impactos positivos do modelo preditivo de crédito.

Tabela 11

Inadimplência do Banco BS2 (Dez/21 – Jun/24)

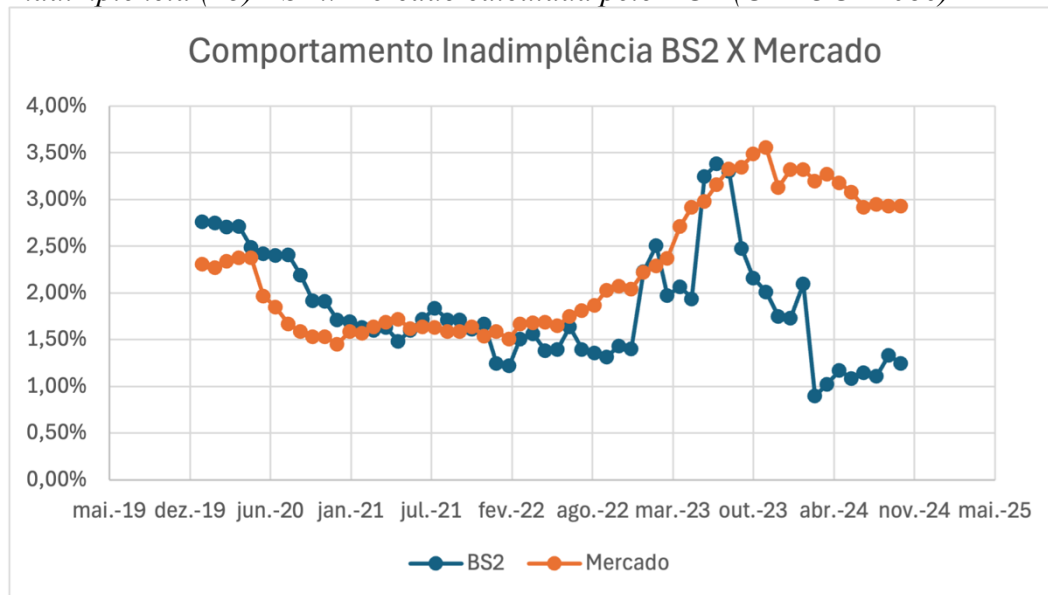
Inadimplência	Dez/21	Jun/22	Dez/22	Jun/23	Dez/23	Jun/24
Over 30	0,29%	0,94%	2,44%	3,68%	2,21%	1,58%
Over 60	0,25%	0,50%	2,26%	3,45%	1,82%	1,21%
Over 90	0,17%	0,39%	0,40%	3,25%	1,75%	1,17%

Figura 4
Inadimplência do Banco BS2 (Dez/21 – Jun/24)



Na Figura 5 é apresentada a taxa de inadimplência PJ (Pessoa Jurídica) do BS2 versus a de mercado, publicada no CADOC- 21086 do BCB (BCB, 2024). O CADOC, ou Catálogo de Documentos, é um conjunto de documentos que as instituições financeiras devem apresentar ao BCB.

Figura 5
Inadimplência (PJ) BS2 x Mercado calculada pelo BCB (CADOC 21086)



Fonte: BCB-CADOC (2024)

O que se pode constatar é um histórico de registros da taxa de inadimplência que descrevem índices em linha com o mercado, apresentando uma queda decrescente a partir de julho de 2023. Pode-se dividir a inadimplência em 3 fases: 2020 – pandemia: a inadimplência rodou acima da taxa do mercado; 2021-2022-houve uma convergência em linha com a taxa de mercado; e 2023-2024: taxa decrescente após a implementação do novo modelo de análise de crédito. Ou seja, os dados refletem considerável melhora e maior eficiência na identificação de riscos e na tomada de decisões de crédito após julho de 2023, quando já havia passado 6 meses da data da adoção do

novo modelo de análise de crédito com o algoritmo GBDT. Para se afirmar com toda a certeza que estes impactos nos índices de inadimplência foram fruto da adoção do novo modelo, seria necessário mais um tempo de acompanhamento.

4.5. Benefícios advindos com a implementação do modelo baseado no algoritmo GBDT

Com relação a este tópico, o entrevistado enfatizou que:

Os principais benefícios foram aqueles que conduziram a uma maior precisão na avaliação de risco de crédito, permitindo a uma redução nas taxas de inadimplência sobre 60 dias e 90 dias, desde sua implementação em dezembro de 2022 e, conseqüentemente, a uma diminuição nas perdas financeiras. Ainda, segundo suas palavras: “A capacidade de identificar precocemente riscos de crédito permitiu a implementação de medidas preventivas mais eficazes e decisões de crédito mais rápidas e baseadas em dados”.

Ou seja, o entrevistado reforça que, a partir da adoção do GBDT, a capacidade do banco foi aperfeiçoada na identificação de padrões complexos e na previsão de comportamentos de risco. Ao possibilitar a detecção precoce de inadimplência, ações preventivas, como ajustes de limite de crédito, ofertas de renegociação de dívidas ou ofertas de programas de educação financeira, podem ser viabilizadas.

Com relação à análise em tempo real, o entrevistado informou “que o banco pode ajustar rapidamente suas políticas de crédito em resposta a flutuações econômicas ou mudanças no comportamento do cliente. Este foi o caso durante o período da pandemia”.

4.6 Feedback dos analistas de crédito / Perspectivas futuras

O entrevistado reforçou que o feedback dos analistas de crédito foi amplamente positivo, destacando a melhoria na precisão das previsões e a facilidade de interpretação dos resultados, com “o reconhecimento do impacto positivo na eficiência operacional e na redução de perdas financeiras”. Quanto às melhorias ou ajustes adicionais, “há planos para otimizar ainda mais o modelo de crédito, a partir da integração de novas fontes de dados, tanto internas quanto externas, para enriquecer o conjunto de dados utilizado pelo modelo”. Isso pode incluir dados alternativos, como histórico de pagamento de aluguel e utilidades, que podem fornecer insights adicionais sobre o comportamento de crédito dos clientes.

4.7 Considerações Éticas e Regulatórias

Conforme o entrevistado, o Banco BS2 adota uma abordagem rigorosa para garantir a conformidade com as regulamentações financeiras ao utilizar modelos de IA e aprendizado de máquina. Isso inclui “a implementação de políticas de governança de dados robustas, a realização de auditorias regulares de conformidade e a manutenção de uma equipe de conformidade dedicada”. Os modelos são continuamente monitorados e avaliados para garantir que estão em conformidade com as diretrizes regulatórias, como a Lei Geral de Proteção de Dados (LGPD) e outras normas internacionais relevantes. Para garantir a privacidade, o BS2 implementou várias camadas de proteção de dados, incluindo criptografia de dados em repouso e em trânsito, políticas rigorosas de controle de acesso e monitoramento contínuo de atividades suspeitas. Adicionalmente, foi criado um comitê de privacidade de dados para revisar e aprovar o uso de dados no desenvolvimento de modelos de aprendizado de máquina, “garantindo que todas as práticas estejam alinhadas com as regulamentações de privacidade vigentes”.

5 CONCLUSÃO

Este estudo de caso teve como objetivo analisar a utilização de uma ferramenta de aprendizado de máquina, especificamente o algoritmo *Gradient Boosting Decision Tree* (GBDT), na gestão de risco de crédito pelo Banco BS2. GBDT é um algoritmo de árvore de decisão iterativo que consiste em múltiplas árvores de decisão; a decisão final é baseada em conclusões combinadas de todas as árvores (Zhang et al., 2018). A triangulação dos dados foi a alternativa metodológica adotada para redução de ruídos no entendimento, tendo em vista a complexidade do tema. A implementação do modelo preditivo no banco em questão, por meio do algoritmo GBDT, representa um avanço significativo na identificação de bons pagadores, permitindo uma melhor avaliação do perfil de crédito dos clientes de forma preditiva. O intuito é aperfeiçoar o gerenciamento de crédito, como também contribuir para a redução da inadimplência e a otimização das operações financeiras, impactando de forma direta nos resultados, uma vez que melhora os níveis de provisão e inadimplência (*default*). A utilização do GBDT na análise de crédito também permite ao Banco BS2 se adaptar rapidamente às mudanças no comportamento dos clientes e nas condições de mercado, garantindo melhor dinamismo nas políticas de crédito.

O BS2 optou por este algoritmo após analisar os resultados das métricas de desempenho de algoritmos de classificação diferenciados, tais como: redes neurais, árvore de decisão, floresta aleatória, GBDT e regressão logística. Isto possibilitou uma análise comparativa entre as abordagens adotadas, como também relativas às suas vantagens específicas, que incluem a simplicidade, a facilidade de interpretação e a habilidade de detectar padrões complexos. Adicionalmente, conforme relatado pelo *Head of Decision Science & Analytics*, após testes realizados, este algoritmo se destaca na análise de grandes volumes de dados, na resolução de problemas de desbalanceamento de amostras e na melhoria da seleção de variáveis financeiras e não financeiras, com base em sua relevância direta para a avaliação do risco de crédito. O GBDT é conhecido por sua alta precisão e robustez, especialmente em tarefas de classificação e regressão.

A concessão de crédito desempenha um papel importante nas operações das instituições financeiras para garantir a saúde financeira. Conforme a teoria do ciclo de vida de Modigliani e Brumberg (1954), a capacidade de pagamento dos indivíduos varia ao longo de suas vidas. Jovens tendem a tomar mais crédito para financiar consumo e investimentos, enquanto indivíduos de meia-idade tendem a poupar mais, e aposentados consomem suas economias. A precisão na avaliação de risco é um fator determinante para evitar inadimplências e perdas financeiras.

A aplicação de tecnologias vem conduzindo uma análise mais precisa e robusta dos dados, mas também uma mitigação eficaz dos riscos associados à concessão de crédito, conforme sugerido por Hand e Henley (1997), e estas modelagens têm revolucionado a forma como as instituições financeiras avaliam o risco de crédito. A integração dessas tecnologias pode transformar a forma como os dados são analisados e as decisões são tomadas, resultando em um sistema de crédito mais eficiente, preciso e seguro. A capacidade de prever comportamentos de inadimplência com maior precisão e a otimização dos processos operacionais são apenas algumas das muitas vantagens oferecidas pela inteligência artificial (IA) e pelo aprendizado de máquina.

A implementação do algoritmo GBDT, combinada com as consultas aos *bureaus* de crédito na gestão de risco de crédito pelo Banco BS2, permitiu à instituição criar um modelo preditivo que melhorou a eficiência operacional, possibilitando decisões de crédito mais rápidas e baseadas em dados. Adicionalmente, a taxa de inadimplência do Banco BS2, registrada no BCB-CADOC, tem se mostrado decrescente após a implementação do novo modelo de análise de crédito, baseado no algoritmo GBDT. No sistema de avaliação tradicional de análise de crédito, as informações sociodemográficas e de solicitação de empréstimos são projetadas como entradas nos modelos de análise de crédito. Entretanto, o histórico dinâmico de transações dos candidatos, o qual é uma métrica importante do comportamento de retorno dos candidatos, normalmente não é incluída no sistema de avaliação tradicional.

Sendo assim, o modelo Comportamental introduzido no BS2 pretende resolver este problema, apresentando convergência ao estudo de Zhang et al. (2018), que propôs um método de avaliação abrangente que inclui dados tradicionais, sociodemográficos individuais e sobre solicitações de empréstimo, bem como dados comportamentais de transações dinâmicas dos solicitantes. Os resultados do teste deste estudo em particular mostraram que este método melhorou significativamente o desempenho da previsão com base nos critérios de avaliação de modelo mais comuns (Zhang et al., 2018). Isto também é demonstrado na métrica F1, que o Banco BS2 utiliza como parâmetro, para demonstrar a eficácia e a precisão do modelo, que na comparação com o modelo Aplicação da Serasa apresenta um índice superior, de 0,77. A métrica F1 é um critério comum na recuperação de informação para avaliar o desempenho de um modelo de classificação (Zhou et al., 2019). Adicionalmente, este estudo apresenta dados desbalanceados, nos quais as empresas adimplentes dominam, de forma que o melhor modelo, neste caso, é aquele que usa o maior índice de F1. A medida F1 representa a média ponderada da sensibilidade e precisão. Muitos estudos utilizam também a medida da acurácia para avaliar o desempenho do modelo, porém, conforme sinalizado por Khemakhem & Boujelbene (2018), o problema com este parâmetro é que o comportamento enviesado de uma boa taxa de precisão com dados desbalanceados pode levar a uma má seleção de um modelo de previsão.

Com o intuito de buscar o aperfeiçoamento do desempenho do modelo Comportamental, conforme reforçado pelo *Head of Decision Science & Analytics*, o BS2 poderá utilizar as seguintes estratégias:

- Aprimoramento dos Dados de Entrada com a introdução de *LLMs (Large Language Models)* no modelo de análise de crédito para trazer melhorias: *incorporar dados alternativos*, além dos dados comportamentais financeiros, incluir informações adicionais, como dados de mídias sociais ou feedback on-line (com consentimento do cliente), histórico de interações com suporte ao cliente (e-mails, mensagens de chat, transcrições de chamadas), análise de comportamento de consumo e uso de cartões;
- Treinamento do Modelo: usar um conjunto maior de dados históricos para treinar o modelo pode melhorar a capacidade preditiva com atualização frequente do modelo, ou seja, retreinar o modelo regularmente para capturar mudanças no comportamento dos clientes e nas condições econômicas;
- Combinação de Pontuações: criar um modelo híbrido, que combine o modelo Aplicação e o modelo Comportamento, pode melhorar os resultados gerais. Por exemplo: modelo Aplicação poderia ser usado para pré-seleção e o modelo Comportamento seria aplicado em uma segunda etapa, com foco em clientes aprovados na primeira triagem.

Baseado nos achados deste estudo, futuras pesquisas deveriam se concentrar nos seguintes tópicos:

- uma maior aplicação de métodos *ensemble*, tais como, os de *boosting*, dentre eles o GBDT, para predizer risco de crédito, tendo em vista que a aplicação dos métodos *ensemble* tem sido limitada para estes modelos conforme informado previamente neste estudo;
- seria recomendado que pesquisadores realizassem a comparação dos algoritmos destas técnicas para mensurar o desempenho dos modelos no futuro;
- na coleta da amostra de dados para futuros estudos é recomendado manter o foco em grupos homogêneos de inadimplência/adimplência;
- uma recomendação final seria a seleção das variáveis, porque uma boa combinação e métodos de seleção de variáveis podem permitir uma melhora no modelo preditivo, comparando o desempenho entre os algoritmos individualmente e suas várias combinações no desenvolvimento dos modelos preditivos.

Este estudo pretende contribuir com o conhecimento advindo da experiência do Banco BS2, o qual pode servir de modelo para outras instituições financeiras que buscam melhorar suas

práticas de gerenciamento de risco implementando um modelo preditivo por meio de aprendizado de máquina. Adicionalmente, em virtude das consequências sociais e econômicas advindas do gerenciamento de risco de crédito, o desenvolvimento de um modelo preditivo mais preciso poderá melhorar a acurácia das decisões, tornando-se um diferencial estratégico ao promover a inovação. E, por último, o conhecimento disseminado pelo estudo do caso do Banco BS2 pretende auxiliar na direção de se testar outros algoritmos que venham oferecer desempenhos melhores, exemplificando como a aplicação dessas tecnologias pode contribuir para um sistema financeiro mais resiliente e estável, com maior capacidade de mitigar crises sistêmicas e promover a sustentabilidade financeira a longo prazo. Contudo, é importante destacar a limitação relativa ao período de tempo no qual o modelo preditivo está em funcionamento, pelo fato de que será necessário um volume maior de dados para que venha de fato solidificar a efetiva melhora na inadimplência com a utilização de aprendizado de máquina, e em específico, o algoritmo GBTD.

REFERÊNCIAS

- Agarwal, M., & Vandana, T.R. (2022). Exchange rate crises in Latin America, East Asia and Russia. *Brazilian Journal of Political Economy*, 42(2), 263-282, <http://dx.doi.org/10.1590/0101-31572022-3299>
- Altman, E.I., Sabato, G., & Wilson, N. (2010). The value of non-financial information in small and medium-sized enterprise risk management. *The Journal of Credit Risk*, 6(2), 95-127.
- Ambavat, P. P. (2021). Credit Bureaus Must Adopt AI-ML, Data Analytics for Holistic Credit Scores. *CRIF Highmark*. <https://www.crifhighmark.com>
- Avelar, E. A., Leocádio, V. A., Campos, O. V., Ferreira, P. O. & Orefici, J. B. P. (2022). Algoritmo Random Forest para Previsão de Comportamento de Preços de Ativos. *Revista FSA*, 19(10). <http://www4.unifsa.com.br/revista/index.php/fsa/article/view/2592>
- Banco BS2 (2024). Somos o BS2. <https://www.bancobs2.com.br/somos-o-bs2/>
- Banco Central do Brasil (1999). *Resolução CMN nº 2.682 de 21/12/1999*. <https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o&numero=2682>
- Banco Central do Brasil (2022). *Resolução CMN nº 5.037 de 29/9/2022*. <https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20CMN&numero=5037>
- Banco Central do Brasil (2023). *BC e bureaus de crédito assinam acordo para compartilhamento de informações*. <https://www.bcb.gov.br/detalhenoticia/668/notici>
- Banco Central do Brasil (BCB-CADOC) (2024). *Estatísticas monetárias e de crédito*. <https://www.bcb.gov.br/estatisticas/estatisticasmonetariascredito>
- Bank for International Settlements (BIS) (2024, julho). *Basel Committee on Bank Supervision: International Convergence of Capital Measurement and Capital Standards*. <https://www.bis.org/publ/bcbs238.htm>

- Berrar, D. (2019). Cross-Validation, Bootstrap, and ROC Analysis. *Encyclopedia of Bioinformatics and Computational Biology*, 542-560. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Botelho, D., & Tostes, F. D. (2011). Modelagem de probabilidade de churn. *Revista de Administração de Empresas*, 4(396).
- Calvo, G. A. (2008). Crises in Emerging Markets Economies: A Global Perspective," Central Banking, Analysis, and Economic Policies Book Series. In Kevin Cowan & Sebastián Edwards & Rodrigo O. Valdés & Norman Loayza (Series Editor) & Klaus Schmidt- (ed.), *Current Account and External Financing* (ed. 1, 12, chapter 3, 085-115), Central Bank of Chile.
- Castro, C. L. de & Braga, A. P. (2011), Aprendizado supervisionado com conjunto de dados desbalanceados. *Revista Controle & Automação*, 22(5),441-466.
- Creswell, J. W. & Poth, C. N. (2016), *Qualitative inquiry and research design: choosing among five approaches*. Sage.
- Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Dabrowski, M. (2023). Thirty years of economic transition in the former Soviet Union: Microeconomic and institutional dimensions. *Russian Journal of Economics*, 9,1-32. <https://doi.org/10.32609/j.ruje.9.104761>
- Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business Press.
- Denzin, N. (1978) *The research act: a theoretical introduction to sociological methods*. (2a ed). Mc Graw-Hill.
- Demirgüç-Kunt, A., & Singer, D. (2017). Financial inclusion and inclusive growth: A review of recent empirical evidence. *World bank policy research working paper*, (8040). <http://documents.worldbank.org/curated/en/403611493134249446/pdf/WPS8040.pdf>
- Dewasiri, N. J., Dharmarathna, D. G; Choudhary,M.,(2024). Leveraging Artificial Intelligence for Enhanced Risk Management in Banking: A Systematic Literature Review. In Singh et al (Eds). *Artificial Intelligence Enabled Management: An Emerging Economy Perspective*, Chapter 13, 197-213. <https://doi.org/10.1515/9783111172408013>
- Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 297(3), 1178-1192.
- Flick, U. (2004). *Introdução à pesquisa qualitativa*. Bookman.

- Fosu, S., Boapeah, H. A. & Ciftci, N. (2023). Credit information sharing and cost of debt: Evidence from the introduction of credit bureaus in developing countries. *Financial Review*, 58(4), 653-930.
- Freeman, R. E., & Dmytriyev, S. D. (2017). Corporate social responsibility and stakeholder theory: Learning from each other. *Symphonya. Emerging Issues in Management*, (1), 7-15. <https://symphonya.unicusano.it/index.php/symphonya/article/view/2017.1.02freeman.dmytriyev>
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- Gil, A. C. (2009). *Estudo de Caso*. Editora Atlas.
- Godoy, A. S. (2006). Estudo de caso qualitativo. In C. K. Godoi, R. Bandeira-de-Mello, & A. B. Silva. *Pesquisa qualitativa em estudos organizacionais: paradigmas, estratégias e métodos*. Saraiva.
- Godoi, C. K., Bandeira-De-Melo, R., & Silva, A. B. (Orgs.). (2010). *Pesquisa qualitativa em estudos organizacionais: paradigmas, estratégias e métodos* (2a ed.). Saraiva.
- Grunert, J., Norden, L., & Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of banking & finance*, 29(2), 509-531.
- Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523-541.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine Learning: Trends, Perspectives, and Prospects. *Science*, 349(6245). <https://www.cs.cmu.edu/~tom/pubs/Science-ML-2015.pdf>
- Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767-2787.
- Khemakhem, S., & Boujelbene, Y. (2018). Predicting credit risk on the basis of financial and non-financial variables and data mining. *Review of accounting and finance*, 17(3), 316-340.
- Lassance, L. C. B. K., & Ternoski, S. (2021). Score ia cresol: Utilizando inteligência artificial para estimar viabilidade de crédito. *Revista Aproximação*, 3(06). <https://revistas.unicentro.br/index.php/aproximacao/article/view/6923>
- Lee, I., & Shin, Y. J., (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63, 150-170. <https://doi.org/10.1016/j.bushor.2019.10.005>
- Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1), 29.
- Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: A ten-year update. *European Journal of*

Operational Research, 247(1), 124-136.
<https://www.sciencedirect.com/science/article/pii/S0377221715007692>

- Louzada, F., Ara, A., Fernandes, G. B. (2016), Classification methods applied to credit scoring: Systematic review and overall comparison, *Surveys in Operations Research and Management Science*, 21, 117-134. <http://dx.doi.org/10.1016/j.sorms.2016.10.001>
- Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216-231. <https://doi.org/10.1016/j.patcog.2019.02.023>
- Marini, J. M. & Manfrim, L. F. (2020). Metodologia de análise de crédito aplicada na redução do risco de inadimplência. *REGRAD - Revista Eletrônica de Graduação do UNIVEM*, 13(1), 76-91. <https://revista.univem.edu.br/REGRAD/article/view/3105>
- Mashrur, A., Luo, W., Zaidi, N. A., & Robles-Kelly, A. (2020). Machine learning for financial risk management: a survey. *Ieee Access*, 8, 203203-203223.
- Mendonça, A. R. R. D., & Deos, S. (2020). Regulação bancária: uma análise de sua dinâmica por ocasião dos dez anos da crise financeira global. *Revista de Economia Contemporânea*, 24, e202427. <https://doi.org/10.1590/198055272427>
- Mian, A., & Sufi, A. (2009). The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. *The Quarterly Journal of Economics*, 124(4), 1449-1496. <https://doi.org/10.1162/qjec.2009.124.4.1449>
- Modigliani, F., & Brumberg, R. (1954). Utility analysis and the consumption function: An interpretation of cross-section data. In K. K. Kurihara (Ed.). *Post-Keynesian Economics*. Rutgers University Press.
- Montevechi, A. A., Carvalho Miranda, R., Medeiros, A. L., & Montevechi, J. A. B. (2024). Advancing credit risk modelling with Machine Learning: A comprehensive review of the state-of-the-art. *Engineering Applications of Artificial Intelligence*, 137, 109082. <https://doi.org/10.1016/j.engappai.2024.109082>
- Moodys Local (2024). Relatório do Emissor: Banco BS2 S.A. 2024-05-13. <https://moodyslocal.com.br/reporte/issuer-report/relatorio-do-emissor-banco-bs2-s-a/>
- Oliveira, R., & Santos, P. (2022). Avaliação de scores de crédito e práticas de bureau no Brasil. *Journal of Credit Analysis*, 28(3), 123-138. <http://doi.org/10.1000/j.joca.2022.03.014>
- O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Books.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830. https://www.researchgate.net/publication/51969319_Scikit-learn_Machine_Learning_in_Python

- Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine learning*, 42, 203-231. <https://doi.org/10.1023/A:1007601015854>
- Rahman, M., Ming, T. H., Baigh, T. A., & Sarker, M. (2021). Adoption of artificial intelligence in banking services: an empirical analysis. *International Journal of Emerging Markets*, 18(10), 4270-4300. <https://doi.org/10.1108/IJOEM-06-2020-0724>
- Rahmani, A. M., Rezazadeh, B., Haghparast, M., Chang, W. C., & Ting, S. G. (2023). Applications of artificial intelligence in the economy, including applications in stock trading, market analysis, and risk management. *IEEE Access*.
- Reinhart, C. M., & Rogoff, K. S. (2020). This time is different: A panoramic view of eight centuries of financial crises. *Journal of Economic Perspectives*, 34(3), 3-24. <https://www.aeaweb.org/articles?id=10.1257/jep.34.3.3>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). <https://doi.org/10.1145/2939672.2939778>
- Rizvi, S. A. R., Arshad, S., & Alam, N. (2015). Crises and contagion in Asia Pacific—Islamic v/s conventional markets. *Pacific-Basin Finance Journal*, 34, 315-326. <https://doi.org/10.1016/j.pacfin.2015.04.002>
- Serasa Experian. (2023). *Importância do Bureau de crédito para análise de crédito*. <https://www.serasaexperian.com.br/conteudos/credito/bureau-de-credito-conceito-e-importancia-na-analise-de-credito/>
- Sfeir, E. (2023). *Inovação aplicada ao mercado de crédito e ao setor de birôs*. <https://anbc.org.br/inovacao-aplicada-ao-mercado-de-credito-e-ao-setor-de-biros/>
- Stake, R. E. (1995). *The art of case study research*. Sage Publications.
- Stake, R. E. (2000). Case studies. In N. K. Denzin, & Y. S. Lincoln, Y. S. *Handbook of qualitative research*, (2^a ed.), 435-454, Thousand Oaks.
- Sun, Y., Kamel, M. S., Wong, A. K., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern recognition*, 40(12), 3358-3378.
- Timotio, J. G. M., Vieira, V. E. L., Oliveira, R. A. de, & Silva, R. C. F. e. (2024). Inteligência Artificial no campo de finanças. *Revista de Gestão e Secretariado*, 15(6), e3935. <https://doi.org/10.7769/gesec.v15i6.3935>
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International data privacy law*, 7(2), 76-99. <https://doi.org/10.1093/idpl/ix005>
- Wanzeller, W. F., Alves, C. M. O., & Cota, M. P. (2023). Sistema de apoio à decisão integrando cadastro negativo, scoring, análise qualitativa de crédito com inteligência artificial e

criação de contratos: Protocolo para revisão de escopo. *Research, Society and Development*, 12(7), e18012742680. <https://doi.org/10.33448/rsd-v12i7.42680>

Zhang, T., Zhang, W., Wei, X. U., & Haijing, H. A. O. (2018). Multiple instance learning for credit risk assessment with transaction data. *Knowledge-Based Systems*, 161, 65-77. <http://dx.doi.org/10.1016/j.knosys.2018.07.030>

Zhang, L., & Song, Q. (2022). Credit Evaluation of SMEs Based on GBDT-CNN-LR Hybrid Integrated Model. *Wireless Communications and Mobile Computing*, 2022. <https://doi.org/10.1155/2022/5251228>

Zhou, J., Li, W., Wang, J., Ding, S., & Xia, C. (2019). Default prediction in P2P lending from high-dimensional data based on machine learning. *Physica A: Statistical Mechanics and its Applications*, 534, 122370. <https://doi.org/10.1016/j.physa.2019.122370>

Zöllner, M. A., & Huber, M. F. (2021). Benchmark and survey of automated machine learning frameworks. *Journal of artificial intelligence research*, 70, 409-472. <https://arxiv.org/pdf/1904.12054>

Yazan, B. (2016). Três abordagens do método de estudo de caso em educação: Yin, Merriam e Stake. *Meta: Avaliação*, 8(22), 149-182. <http://dx.doi.org/10.22347/2175-2753v8i22.1038>

Yin, R. K. (2009). *Case Study Research: Design and Methods* (4a ed.). Sage.

CONFLITO DE INTERESSES

Os autores afirmam não haver conflito de interesses com relação a este trabalho submetido.

CONTRIBUIÇÕES DOS AUTORES

Funções	1 ^a autor	2 ^o autor
Conceituação	♦	♦
Curadoria de dados	♦	
Análise Formal		♦
Obtenção de financiamento		
Investigação	♦	
Metodologia		♦
Administração do projeto	♦	♦
Recursos	♦	
Software	♦	
Supervisão		♦
Validação	♦	♦
Visualização	♦	♦
Escrita – primeira redação	♦	
Escrita – revisão e edição		♦