

Análise Comparativa de Redes Neurais Artificiais e Algoritmos de *Ensemble* para a Predição de Risco de Crédito

Comparative Analysis of Artificial Neural Networks and Ensemble Algorithms for Credit Risk Prediction

Maria Letícia da Silva Manguinho¹

¹ Análise e Desenvolvimento de Sistemas, Instituto Federal de Pernambuco (IFPE), Paulista, PE, Brasil,
mlsm1@discente.ifpe.edu.br

Resumo. Neste estudo, é explorada a importância da análise de risco de crédito no sistema financeiro, destacando-se seus desafios e implicações econômicas. Diante da complexidade dos dados desbalanceados e da dinâmica do cenário econômico, o trabalho tem como objetivo empregar técnicas de Inteligência Artificial, como Redes Neurais Artificiais e modelos de ensembles, para aprimorar a detecção de casos de riscos de crédito. Seguindo os ideais do Cross Industry Standard Process for Data Mining, que por sua vez, é dividido em cinco fases iterativas, foi iniciada a pesquisa com a compreensão do contexto do problema e estabelecimento de objetivos claros. Em seguida, houve uma análise detalhada dos conjuntos de dados de crédito para compreender a relevância de cada atributo. Após o pré-processamento dos dados, incluindo limpeza e tratamento, foram então aplicadas as técnicas de modelagem. Finalmente, os modelos foram avaliados com métricas como acurácia, precisão e recall. Notavelmente, os modelos do tipo ensemble, especialmente o Gradient Boosting e o XGBoost, estiveram com desempenho consistente em todas as três bases de dados examinadas neste estudo. Ambos alcançaram uma notável acurácia de 94%, a mais alta registrada nesta pesquisa, enfatizando assim, a escolha desses modelos para o problema de identificação de casos de alto risco de crédito.

Palavras-chave: Previsão de Risco de Crédito; Aprendizado de Máquina; Modelos de Ensemble; Redes Neurais Artificiais; Avaliação Comparativa.

Abstract. In this study, the importance of credit risk analysis in the financial system is explored, highlighting its challenges and economic implications. Given the complexity of imbalanced data and the dynamics of the economic landscape, the aim of the work is to employ Artificial Intelligence techniques, such as Artificial Neural Networks and ensemble models, to enhance the detection of credit risk cases. Following the principles of the Cross Industry Standard Process for Data Mining, which is divided into five iterative phases, the research began with an understanding of the problem context and the establishment of clear objectives. Subsequently, a detailed analysis of credit datasets was conducted to understand the relevance of each attribute. After data preprocessing, including cleaning and treatment, modeling techniques were then applied. Finally, the models were evaluated using metrics such as accuracy, precision, and recall. Notably, ensemble models, especially Gradient Boosting and XGBoost, consistently performed well across all three datasets examined in this study. Both achieved a remarkable accuracy of 94%, the highest recorded in this research, emphasizing the effectiveness of these models.

Keywords: Credit Risk Prediction; Machine Learning; Ensemble Models; Artificial Neural Networks; Comparative Evaluation.

1. Introdução

A avaliação de risco de crédito desempenha um papel fundamental na garantia do funcionamento robusto e estável do sistema financeiro. Ela capacita instituições financeiras e empresas a tomarem decisões informadas sobre a concessão de empréstimos e investimentos, garantindo assim a alocação eficiente e segura de recursos financeiros. Ao avaliar a capacidade de pagamento de indivíduos e empresas, a análise de crédito ajuda a mitigar os riscos associados a transações financeiras, reduzindo, assim, a probabilidade de inadimplência e perdas financeiras (Schwarz, 2019).

Além disso, a análise de crédito desempenha um papel fundamental no estímulo ao crescimento econômico e ao desenvolvimento empresarial. Ao facilitar o acesso ao crédito para empresas e empreendedores, ela impulsiona investimentos, estimula a inovação e cria oportunidades para expansão. Essa análise permite que as empresas obtenham os fundos necessários para financiar projetos, expandir operações e desenvolver novos produtos ou serviços. Ao mesmo tempo, a análise de crédito contribui para a estabilidade financeira, prevenindo o acúmulo excessivo de dívidas e a formação de bolhas especulativas (Bussmann *et al.*, 2021).

Avanços tecnológicos impulsionaram ainda mais a evolução da análise de risco de crédito. Com o surgimento da aprendizagem de máquina (AM) (Raschka, 2015), análise de dados (AD) (Wickham; Wickham, 2016) e IA (Winston, 1984), as instituições financeiras adquiriram ferramentas poderosas para aprimorar suas capacidades de avaliação de risco. Essas tecnologias permitem a análise de conjuntos de dados vastos com notável velocidade e precisão, permitindo a identificação de padrões sutis e fatores de risco que anteriormente eram desafiadores de discernir. A integração de modelos sofisticados não apenas aprimora a precisão das avaliações de crédito, mas também abre caminhos para melhorias contínuas à medida que os modelos aprendem e se adaptam a novas informações (Bussmann *et al.*, 2021).

A análise de risco de crédito enfrenta uma série de desafios que podem dificultar a avaliação precisa da capacidade de pagamento de indivíduos e empresas. Um desafio primordial envolve lidar com dados desequilibrados, nos quais a maioria das observações pertence à classe de baixo risco de crédito, enquanto a classe de alto risco é significativamente menor. Esse desequilíbrio pode introduzir viés nos modelos de análise, tornando mais difícil a identificação correta de casos de alto risco. A representação inadequada da classe de alto risco pode comprometer a eficácia da análise e levar a previsões menos confiáveis (Bao; Lianju; Yue, 2019).

Por último, a natureza dinâmica do cenário econômico e financeiro introduz uma série de desafios adicionais na análise de risco de crédito. As condições econômicas têm o potencial de mudar rapidamente, influenciadas por fatores como eventos globais, mudanças políticas e tendências de mercado. Essas flutuações podem impactar significativamente a capacidade dos mutuários de cumprir seus compromissos financeiros, reconfigurando assim o panorama de riscos. Nesse contexto, a adaptabilidade dos modelos de análise de crédito se torna fundamental. Os modelos tradicionais de avaliação de risco podem ter dificuldade em acompanhar o ambiente econômico em rápida evolução, destacando a necessidade de estratégias que possam capturar e integrar rapidamente novos dados e *insights*. A capacidade de responder prontamente às mudanças nas condições econômicas garante que as avaliações de risco de crédito permaneçam relevantes e confiáveis, permitindo que os credores tomem decisões informadas mesmo diante da incerteza. Essa adaptabilidade serve como um elemento importante na proteção da integridade dos mercados de crédito, na promoção da estabilidade financeira e na viabilização de práticas de empréstimos responsáveis (Alonso; Carbo, 2020).

Esses desafios enfrentados na análise de risco de crédito servem como motivação para o pre-

sente estudo, uma vez que a superação dessas dificuldades é fundamental para aprimorar a precisão e eficácia das avaliações de crédito. Neste contexto, o objetivo deste trabalho é contribuir para a solução desses desafios por meio da inteligência artificial com o uso de técnicas RNAs (Yegnanarayana, 2009), e de *ensembles* (Zhang; Ma, 2012). Através da aplicação destas técnicas e da exploração de suas capacidades de lidar com dados desbalanceados, capturar não linearidades e interações complexas, busca-se fornecer qual a abordagem que apresenta o melhor desempenho baseado em métricas como: acurácia, precisão, *f1-score*, *recall* e no estudo da matriz de confusão, para a análise de risco de crédito, usando 3 bases de dados de crédito desbalanceadas, desta forma vendo sua aplicabilidade em diferentes cenários. Desta forma, essas contribuições permitirão melhorar a tomada de decisões financeiras, auxiliando instituições e empresas a realizar avaliações de crédito mais precisas e informadas.

Na seção 2, serão apresentados os trabalhos relacionados que desempenharam um papel importante na formação da base desta pesquisa. A seção 3 abordará detalhadamente a metodologia empregada na construção deste trabalho, com uma ênfase na apresentação do *Cross-Industry Standard Process for Data Mining* (CRISP-DM) e na explicação das técnicas específicas que serão aplicadas na pesquisa. A seção 4 fornecerá uma análise aprofundada dos dados, além de demonstrar como essa etapa se encaixa no contexto do CRISP-DM. Na seção 5, os detalhes referentes à etapa de preparação dos dados serão explorados, discutindo as técnicas utilizadas para garantir que os conjuntos de dados estejam prontos para serem empregados nos modelos de AM. A seção 6 concentra-se na modelagem, descrevendo o cenário específico desta pesquisa e as abordagens adotadas nessa fase do CRISP-DM. A seção 7 fornecerá uma análise detalhada da etapa de avaliação, destacando como os modelos construídos foram avaliados para determinar sua qualidade e eficácia. Finalmente, na seção 8, serão apresentadas as considerações finais deste estudo, consolidando as descobertas que possam fornecer *insights* para futuras pesquisas no domínio da análise de risco de crédito.

2. Trabalhos Relacionados

No âmbito da avaliação de risco de crédito, estudos recentes têm explorado diversas abordagens utilizando técnicas de aprendizado de máquina para aprimorar a tomada de decisões financeiras. Nesta seção, será apresentado os estudos relevantes conduzidos entre os anos de 2019 a 2023. Esses trabalhos demonstram a contínua importância do domínio de avaliação de risco de crédito e enfatizam o escopo existente para exploração adicional. Ao destacar uma variedade de abordagens e metodologias, esses estudos ressaltam a relevância contínua do campo e as substanciais oportunidades para pesquisas e avanços futuros.

De acordo com (Shen *et al.*, 2019), pesquisas recentes têm explorado a avaliação de risco de crédito, empregando diversas técnicas de AM para a tomada de decisões financeiras. Embora RNAs tenham sido comumente utilizada para a avaliação de risco de crédito, estudos indicam que conjuntos de classificadores superam os classificadores individuais. Esse artigo introduz um modelo de conjunto inovador para a avaliação de risco de crédito pessoal, utilizando a *Synthetic Minority Over-sampling Technique* (SMOTE) (Chawla *et al.*, 2002) e otimização de classificadores. Para combater os efeitos de conjuntos de dados desequilibrados, o SMOTE reequilibra o conjunto de dados de treinamento, enquanto o algoritmo de otimização por enxame de partículas otimiza os pesos da RNAs. O modelo de conjunto combina o AdaBoost com classificadores otimizados de RNAs, sendo testado em conjuntos de dados reais e desequilibrados, demonstrando sua eficácia aprimorada no processamento de dados de crédito em comparação com outros modelos examinados.

Conforme (Bao; Lianju; Yue, 2019), a pontuação de crédito desempenha um papel fundamental na identificação de candidatos dignos de crédito para instituições financeiras. Embora métodos

de aprendizado de máquina supervisionados tenham sido amplamente utilizados para a pontuação de crédito, o potencial de integrar aprendizado não supervisionado ao lado do aprendizado supervisionado tem recebido atenção limitada. Esse estudo introduz uma estratégia inovadora que combina aprendizado não supervisionado com aprendizado supervisionado para a avaliação de risco de crédito. Ao contrário de abordagens anteriores, essa estratégia aplica técnicas de aprendizado não supervisionado em duas etapas distintas: a etapa de consenso e a etapa de agrupamento de dados. Uma análise comparativa é realizada entre várias combinações de modelos usando três conjuntos de dados de crédito. Os resultados demonstram que a integração do aprendizado não supervisionado em qualquer uma das etapas melhora o desempenho dos modelos de pontuação de crédito. Notavelmente, a combinação de ambas as etapas produz resultados ótimos, afirmando a superioridade dessa abordagem e sugerindo sua aplicabilidade potencial a diversos conjuntos de dados de crédito no domínio financeiro.

O estudo de (Teles *et al.*, 2020) afirma que o risco de crédito representa uma ameaça para as instituições financeiras, carregando o potencial de consequências graves. Ferramentas de previsão de risco prometem mitigar a insolvência bancária. Esse estudo realiza uma comparação entre redes Bayesianas e RNAs na previsão do valor recuperado dentro de um contexto de operação de crédito. Tradicionalmente, o desafio da pontuação de crédito tem sido tratado como um problema de classificação supervisionada no campo do aprendizado de máquina. O estudo aborda essa questão e conclui que as RNAs se mostram uma ferramenta mais eficaz para prever o risco de crédito comparado à abordagem ingênua de *Bayesian* (NB) (Zhang, 2004). O foco central gira em torno das decisões de empréstimo, no qual um conjunto de fatores influencia operações de crédito significativas, e probabilidades auxiliam na classificação de novos candidatos com base em suas características. Um bom resultado surgiu resultando em uma precisão média de 85%. Utilizando a abordagem NB, o algoritmo foi aplicado em quatro conjuntos de dados em um único processo antes de usar o conjunto de dados completo para construir uma matriz de confusão.

De acordo com (Caruso *et al.*, 2021), o risco de crédito continua sendo uma preocupação primordial para os bancos que oferecem serviços financeiros. Inúmeros métodos de pontuação surgiram, predominantemente baseados em métricas quantitativas, para avaliar o desempenho financeiro dos clientes. Esse estudo enfatiza a importância de combinar atributos de candidatos tanto quantitativos quanto qualitativos e introduz uma nova metodologia que integra técnicas de agrupamento de dados mistos. A abordagem reconhece que o agrupamento tradicional se concentra em dados quantitativos ou qualitativos, mas dado que os candidatos a crédito exibem características pessoais mistas, uma análise de cluster adaptada para dados mistos pode revelar padrões informativos. Essa técnica tem o potencial de estimar o risco de crédito de maneira mais precisa, considerando uma ampla gama de atributos de candidatos, melhorando assim a avaliação de risco na concessão de crédito.

O artigo de (Abedin *et al.*, 2022) diz que lidar com a natureza desequilibrada das classes de inadimplência e não inadimplência na avaliação de risco de crédito para pequenas empresas é um desafio. Para mitigar esse problema, essa pesquisa introduz uma abordagem de conjunto estendida chamada *WSMOTE-ensemble*, que combina a *Weighted Synthetic Minority Over-sampling Technique* (WSMOTE) com *Bagging* e misturas compostas de amostragem. Esse classificador de conjunto híbrido visa melhorar a robustez e a variabilidade na geração de instâncias sintéticas, aliviando assim as restrições de classes desequilibradas associadas às instâncias de pequenas empresas. O estudo utiliza um conjunto de dados de um banco comercial chinês e realiza experimentos abrangentes para avaliar a abordagem proposta sob vários cenários de dados desequilibrados. Através de uma avaliação minuciosa, o estudo sugere que o modelo de conjunto *WSMOTE-ensemble* com um classificador de

Random Forest alcança um equilíbrio favorável entre o desempenho das classes de inadimplência e não inadimplência. O método de conjunto proposto supera concorrentes ao melhorar a precisão da classe minoritária em 15,16%. Além disso, o estudo destaca a superioridade dos métodos de amostragem em relação a modelos sem amostragem, contribuindo com *insights* valiosos para a previsão de risco de crédito para pequenas empresas.

Segundo (Yang *et al.*, 2023), esse trabalho recente apresenta uma abordagem avançada para a previsão precisa de risco de crédito corporativo, que é vital para evitar falências proativas e ajustes estratégicos. Ao contrário da tendência comum de incorporar um número crescente de características em sistemas de previsão, esse estudo reconhece a desvantagem de introduzir informações redundantes e irrelevantes que podem prejudicar a eficácia dos modelos de previsão. Para abordar esse problema, os pesquisadores propõem um algoritmo inovador de *High-Dimensional Neural Network* (HDNN) (Yang *et al.*, 2023) como um aprimoramento aos modelos convencionais de *Deep Neural Network* (DNN) (Miikkulainen *et al.*, 2019), especificamente adaptado para a previsão de risco de crédito corporativo em alta dimensão. O estudo inclui provas teóricas que lançam luz sobre a interação entre técnicas de regularização e camadas de normalização em lote dentro de DNNs. As descobertas contribuem para resolver questões relacionadas à regularização em implementações industriais. A eficácia do algoritmo HDNN é demonstrada por meio de um estudo de caso envolvendo dados de crédito enriquecidos com informações de cadeia de suprimentos e rede, mostrando sua superioridade no tratamento de conjuntos de dados de alta dimensão dentro desses contextos.

Esse estudo em questão, motivado pela pesquisa anterior de (Teles *et al.*, 2020), concentra-se não apenas em enfrentar desafios críticos na análise de risco de crédito, mas também em demonstrar a aplicação de técnicas destinadas a aprimorar as avaliações de crédito. Conforme mencionado por (Abedin *et al.*, 2022), a importância de abordar o desequilíbrio entre as classes normalmente existentes em dados de créditos, levou à adoção de estratégias de tratamento de desequilíbrio, esse mesmo estudo destacou a superioridade das abordagens de amostragem em relação aos modelos sem amostragem, e com isso, o *Random Forest* (Rigatti, 2017) se sobressaiu como o algoritmo de escolha, abrindo assim portas para a exploração de algoritmos de *ensemble*.

Além disso, a pesquisa também se baseou nas descobertas de (Yang *et al.*, 2023), que demonstraram a eficácia das redes neurais profundas na previsão do risco de crédito. Como resultado, neste estudo, modelos de redes neurais profundas, como a *Convolutional Neural Network* (CNN) (Li *et al.*, 2021), foram empregados, apesar de seu uso mais comum ser em tarefas envolvendo imagens, experimentou-se a aplicabilidade da CNN a dados tabulares, estendendo sua utilização para além do domínio convencional, outra RNA utilizada foi o *Multilayer Perceptron* (MLP) (Taud; Mas, 2018) que no contexto deste estudo, optou-se por criá-lo manualmente, em vez de utilizar a implementação pronta disponível na biblioteca do *scikit-learn*. O objetivo principal desses esforços foi desenvolver modelos de RNAs que, juntamente com algoritmos de *ensemble* como *Random Forest*, *Gradient Boosting* (Liu; Fan; Xia, 2022) e *XGBoost* (Chen; Guestrin, 2016), contribuiriam para aprimorar as decisões de empréstimo e reduzir as potenciais perdas associadas à inadimplência, beneficiando assim instituições financeiras, empresas e o sistema financeiro, promovendo decisões mais informadas.

À medida que transitamos para a próxima seção, nossa atenção se volta para esclarecer a metodologia utilizada neste estudo, em que o *framework* CRISP-DM (Martínez-Plumed *et al.*, 2019) foi empregado para desenvolver modelos capazes de prever a probabilidade de inadimplência de crédito, um aspecto importante da avaliação de risco de crédito.

3. Metodologia

A metodologia adotada neste estudo segue o CRISP-DM, um *framework* amplamente reconhecido que compreende seis fases para orientar o desenvolvimento de projetos de mineração de dados. As cinco fases iterativas e incrementais aplicadas são: Compreensão do Negócio, Compreensão dos Dados, Preparação dos Dados, Modelagem e Avaliação. Cada uma dessas fases desempenha um papel importante na moldagem da progressão do projeto e garante uma abordagem sistemática para a resolução de problemas. Vale destacar que, neste estudo, optou-se por não incluir a sexta e última etapa do CRISP-DM, que é a fase de Implantação, pois se trata de um estudo direcionado, e a aplicação dessa etapa não é necessária para os objetivos propostos. Esta seção fornecerá detalhes sobre como essas cinco fases foram aplicadas para desenvolver modelos de avaliação de risco de crédito.

3.1. Compreensão do Negócio

Na fase inicial, o foco foi compreender o contexto do problema e estabelecer os objetivos do projeto. Neste estudo, foi dada ênfase à avaliação de risco de crédito e ao seu impacto nas decisões financeiras. Os objetivos de negócios incluíram eficácia no tratamento de dados desbalanceados e a implementação de uma solução adaptativa capaz de lidar com as flutuações presentes no cenário econômico. O objetivo de mineração de dados consistia em identificar um modelo com boa capacidade de generalização, medida pela acurácia, e uma precisão relevante na detecção de casos de riscos. Isso visava encontrar os modelos que melhor se sairiam diante desse desafio.

3.2. Compreensão dos Dados

Nesta etapa, de acordo com o método CRISP-DM, a análise da base de dados de crédito foi conduzida com o objetivo de compreender o significado de cada um dos atributos e sua relevância para o processo de construção de modelos de previsão de concessão ou recusa de crédito. O foco principal dessa etapa foi proporcionar uma visão completa do perfil financeiro de um indivíduo, bem como dos detalhes relacionados ao empréstimo. O propósito fundamental da fase foi adquirir um entendimento abrangente da estrutura dos conjuntos de dados, de seus atributos e do potencial significado deles na tomada de decisão. Os detalhes referentes a essa etapa estão presentes na seção 4 deste estudo.

3.3. Preparação dos Dados

Neste ponto, conduziu-se o pré-processamento dos conjuntos de dados para assegurar sua preparação adequada para integração aos modelos de AM. Esse procedimento abrangeu a limpeza dos dados, o tratamento de valores ausentes, a normalização de variáveis, transformação de características e, quando necessário, a seleção das características mais relevantes para a análise de risco de crédito. Adicionalmente, estratégias de enfrentamento do desbalanceamento nas bases de dados foram aplicadas, sendo escolhida a técnica de *oversampling* SMOTE. Esta decisão baseou-se na ampla aceitação do SMOTE na literatura, como destacado nos artigos (Shen *et al.*, 2019) e (Abedin *et al.*, 2022). Essas etapas desempenham um papel importante na aprimoração da qualidade dos dados, visando otimizar o desempenho dos modelos de classificação.

3.4. Modelagem

Nesta fase, os modelos de análise de dados são selecionados e construídos, abrangendo uma variedade de técnicas de modelagem, incluindo modelos de aprendizado de máquina de RNAs e modelos de *ensemble*. Cada algoritmo é treinado utilizando os conjuntos de dados previamente preparados, e,

quando necessário, são realizados ajustes nos hiperparâmetros dos modelos para otimizar seu desempenho. A otimização de hiperparâmetros pode ser realizada por meio da técnica de busca bayesiana (Turner *et al.*, 2021), garantindo que os modelos atinjam seu melhor desempenho de acordo com os critérios estabelecidos na etapa de compreensão do problema.

3.5. Avaliação

Neste estágio, os modelos construídos são avaliados quanto à sua qualidade e eficácia, aplicando-os a conjuntos de dados de teste ou validação. As métricas de desempenho fundamentais para este estudo incluem acurácia, precisão, *recall*, *f1-score* e a matriz de confusão.

Embora a acurácia seja uma métrica amplamente utilizada, sua aplicação em conjuntos de dados desbalanceados pode ser enganadora. Mesmo assim, continua sendo relevante como uma medida geral de desempenho, oferecendo uma visão global da precisão do modelo.

A precisão ganha destaque em situações de desbalanceamento, concentrando-se na proporção de instâncias positivas classificadas corretamente em relação ao total previsto como positivo pelo modelo. No contexto de risco de crédito, essa métrica é valiosa para avaliar a capacidade do modelo de evitar falsos positivos, ou seja, classificar erroneamente casos não arriscados como arriscados.

O *recall* assume importância, destacando a proporção de instâncias positivas corretamente identificadas em relação ao total real de instâncias positivas. Em cenários desbalanceados, focar no *recall* é essencial para minimizar falsos negativos, ou seja, casos reais de risco de crédito erroneamente classificados como não arriscados.

A medida F1, como uma média harmônica entre precisão e *recall*, torna-se útil ao lidar com o compromisso entre essas duas métricas. Em análise de crédito, onde buscar um equilíbrio entre a identificação correta de casos arriscados e a minimização de falsos positivos é essencial, o *F1-score* se destaca como uma métrica valiosa.

A matriz de confusão fornece uma visão detalhada do desempenho do modelo, identificando verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. Essa visão pormenorizada é importante para entender como o modelo se comporta em diferentes aspectos e onde ocorrem as principais falhas ou sucessos.

Posteriormente, os modelos são comparados entre si e com os critérios de sucesso previamente estabelecidos, sendo estas etapas essenciais para garantir que os modelos escolhidos sejam eficazes no contexto da análise de risco de crédito.

As próximas seções abordarão detalhadamente as diversas etapas do processo de mineração de dados conforme delimitadas pelo CRISP-DM, explorando as abordagens adotadas e os resultados obtidos ao longo deste estudo. Essa análise proporcionará uma visão abrangente do processo de mineração de dados e das considerações derivadas do estudo.

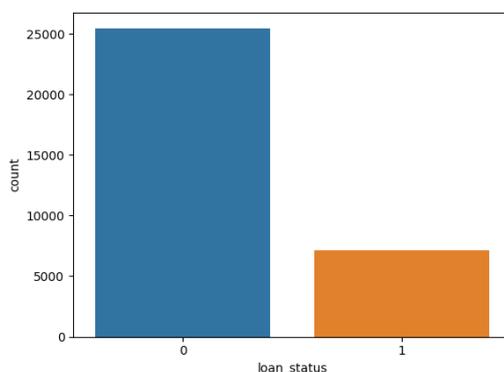
4. Análise das Bases

Nesta seção será descrito como foi implementado o método descrito na subseção 3.2. Inicialmente foram obtidos *insights* abrangentes sobre a primeira base de dados, composta por 32.581 registros contendo informações relacionadas a crédito de várias agências. Esse conjunto de dados é especialmente relevante para análise de risco de crédito (Credit..., 2020b). Ele engloba 12 características distintas, cada uma fornecendo informações valiosas. As características incluídas são as seguintes:

- **person_age**: Idade da pessoa;

- **person_income**: Renda anual da pessoa;
- **person_home_ownership**: Status de propriedade residencial da pessoa;
- **person_emp_length**: Tempo de emprego da pessoa em anos;
- **loan_intent**: Propósito do empréstimo;
- **loan_grade**: Classificação do empréstimo;
- **loan_amnt**: Valor do empréstimo;
- **loan_int_rate**: Taxa de juros do empréstimo;
- **loan_status**: Status do empréstimo (0 para pagantes e 1 para não pagantes), servindo como variável alvo;
- **loan_percent_income**: Porcentagem da renda relacionada ao empréstimo;
- **cb_person_default_on_file**: Histórico de inadimplência do indivíduo;
- **cb_person_cred_hist_length**: Duração do histórico de crédito do indivíduo.

Figura 1. Distribuição da coluna alvo "loan_status", sendo o valor 0 correspondente a classe de indivíduos adimplentes e o valor 1 aos inadimplentes.



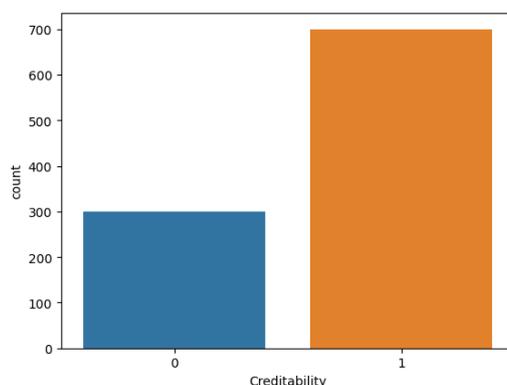
Os números de amostras retratados na Figura 1 destacam uma discrepância substancial entre as categorias de situação de empréstimo "pagante"(classe 0) e "não pagante"(classe 1). Especificamente, existem 25.473 registros classificados como "pagantes" e 7.108 registros categorizados como "não pagantes". Essa significativa desproporção de classes levanta preocupações sobre o possível viés nos resultados do modelo, uma vez que os modelos podem tender a prever com maior precisão a classe majoritária devido à sua maior representação.

Além disso, uma análise similar foi realizada em uma segunda base de dados (German..., 2019). Esse conjunto possui 21 características distintas e 1000 registros no total. Essa análise visa ampliar o entendimento sobre a aplicação das abordagens utilizadas e permite uma avaliação mais abrangente dos algoritmos de *ensemble* e RNAs utilizados para prever riscos de crédito. As características deste conjunto correspondem a:

- **Creditability**: Confiabilidade de Crédito (1 para confiável, 0 para não confiável)
- **Account Balance**: Saldo da Conta
- **Duration of Credit (month)**: Duração do Crédito (em meses)
- **Payment Status of Previous Credit**: Situação de Pagamento de Crédito Anterior
- **Purpose**: Finalidade do Crédito
- **Credit Amount**: Valor do Crédito
- **Value Savings/Stocks**: Valor de Poupança/Ações
- **Length of current employment**: Duração do Emprego Atual
- **Instalment per cent**: Percentual da Parcela

- **Sex & Marital Status:** Sexo e Estado Civil
- **Guarantors:** Fiadores
- **Duration in Current address:** Duração no Endereço Atual
- **Most valuable available asset:** Ativo mais Valioso Disponível
- **Age (years):** Idade (anos)
- **Concurrent Credits:** Créditos Concorrentes
- **Type of apartment:** Tipo de Apartamento
- **No of Credits at this Bank:** Número de Créditos neste Banco
- **Occupation:** Ocupação
- **No of dependents:** Número de Dependentes
- **Telephone:** Telefone
- **Foreign Worker:** Trabalhador Estrangeiro

Figura 2. Distribuição da coluna alvo "Creditability", sendo o valor 1 correspondente a classe de indivíduos adimplentes e o valor 0 aos inadimplentes.



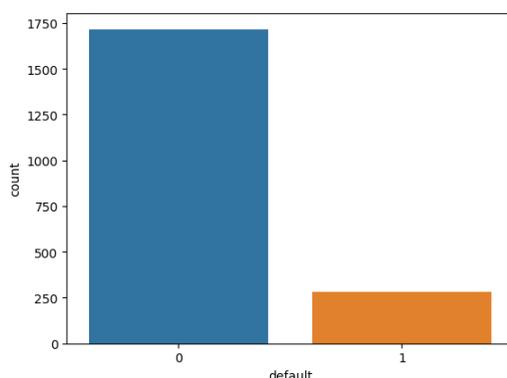
Os números de amostras representados na Figura 2 evidenciam uma discrepância substancial entre as categorias de *status* de crédito "confiável"(classe 1) e "não confiável"(classe 0). Especificamente, existem 300 registros classificados como "inadimplentes" e 700 registros categorizados como "adimplentes". Essa significativa desproporção pode afetar a capacidade do modelo de realizar previsões equilibradas para ambas as classes.

Além disso, estendendo essa investigação para uma terceira e última base de dados (Credit..., 2020a). Foram identificadas 5 características em um total de 2000 registros. A inclusão desse terceiro conjunto enriquece ainda mais a análise, permitindo uma avaliação completa das estratégias empregadas e uma compreensão mais profunda de como os modelos se comportam em diferentes cenários de dados. As características específicas desse conjunto são:

- **clientid:** Identificador do Cliente
- **income:** Renda do Cliente
- **age:** Idade do Cliente
- **loan:** Valor do Empréstimo
- **default:** Status de Inadimplência (1 para inadimplente, 0 para adimplente)

As situações anteriores se repetem na Figura 3 que revela uma discrepância significativa entre as categorias de *status* de empréstimo "adimplente"(classe 0) e "inadimplente"(classe 1). Precisamente, encontram-se 1717 registros classificados como "adimplentes" e 283 registros categorizados como "inadimplentes".

Figura 3. Distribuição da coluna alvo "default", sendo o valor 0 correspondente a classe de indivíduos adimplentes e o valor 1 aos inadimplentes.



Esses atributos, em conjunto, fornecem uma visão abrangente do perfil financeiro de um indivíduo e detalhes relacionados ao empréstimo. O objetivo desta fase era compreender completamente a estrutura dos conjuntos de dados, seus atributos e seu potencial significado na previsão de risco de crédito. Para compreensão do estudo cada uma das bases serão enumeradas conforme ordem da descrição feita nesta etapa, sendo assim base de dados 1, 2 e 3.

5. Pré-Processamento dos Dados

Descrevendo assim o que foi realizado na etapa de preparação dos dados descrita na subseção 3.3, com o objetivo de mitigar o desafio do desbalanceamento das bases e baseando-se na literatura, foi empregado o SMOTE. O SMOTE aborda esse desequilíbrio gerando instâncias sintéticas para a classe minoritária por meio da interpolação de instâncias existentes. Esse processo tem como objetivo equilibrar a distribuição das classes e melhorar o desempenho do modelo, especialmente na previsão precisa de instâncias da classe minoritária.

No contexto deste estudo, a utilização do SMOTE é justificada pela necessidade de evitar previsões enviesadas em direção à classe majoritária. Ao introduzir instâncias sintéticas para a classe minoritária, o SMOTE se esforça para proporcionar ao modelo uma representação mais equitativa de ambas as classes. Como resultado, o modelo fica mais bem preparado para compreender os padrões associados à classe "não pagante", aprimorando assim sua capacidade preditiva e promovendo uma generalização para dados não vistos.

Nas bases de dados 2 e 3, foi necessário realizar a utilização do método "cut", ou seja, a discretização de certos atributos contínuos, especificamente, a coluna "Age" de ambas as bases e a coluna "Duration of Credit (month)" da base de dados 2. Essa ação de discretização envolve o agrupamento de valores em intervalos, foi aplicada com a finalidade de mitigar o impacto de variações menores ou ruídos nos dados. A transformação em intervalos permite que padrões mais amplos e evidentes se destaquem, o que, por sua vez, facilita a identificação de tendências claras nos conjuntos de dados.

A fase de preparação dos dados envolveu uma série de etapas para garantir a adequação das bases de dados para a modelagem. Sendo assim, os dados foram normalizados usando a técnica *MinMaxScaler* (Brownlee, 2020) para garantir uma escala uniforme entre as características, evitando que certos atributos tenham influência desproporcional sobre o modelo devido a escalas diferentes.

Variáveis categóricas foram codificadas usando a codificação *one-hot* (Rodríguez et al., 2018)

para convertê-las em um formato numérico adequado para modelos de aprendizado de máquina. Esse processo transformou atributos categóricos como *"person_home_ownership"*, *"loan_intent"*, *"loan_grade"* e *"cb_person_default_on_file"* da base de dados 1, *"Age (years)"*, *"Duration of Credit (month)"* da base de dados 2 e *"age"* da base de dados 3 em colunas binárias, cada uma representando uma categoria distinta.

Ao tratar a questão dos dados faltantes, foram identificados valores ausentes em dois atributos na base de dados 1, especificamente *"person_emp_length"* e *"loan_int_rate"*. É importante destacar que na base de dados 2 não foram encontrados dados faltantes, enquanto na base de dados 3 foram identificados valores ausentes no atributo *"age"*. Para manter a integridade dos dados em todas as bases, optou-se por uma estratégia de imputação na qual os valores faltantes foram preenchidos pela média, quando se tratavam de atributos numéricos, e pela moda, quando se tratavam de valores categóricos, com base nos valores existentes em suas respectivas colunas. Essa abordagem foi escolhida com o propósito de preservar informações que de outra forma seriam perdidas, permitindo a realização de análises e modelagem de maneira sólida. Ao mesmo tempo, essa estratégia minimizou o impacto dos valores ausentes nos resultados, garantindo que os conjuntos de dados estivessem prontos para uma análise e uma modelagem eficaz.

Ao realizar essas etapas de preparação de dados, os conjuntos de dados foram transformados em um formato adequado para treinar os modelos propostos neste trabalho, que incluem a CNN, MLP, *Random Forest*, *XGBoost* e *Gradient Boosting*. As seções subsequentes apresentarão detalhes sobre a arquitetura de cada um desses modelos, suas avaliações de desempenho e as percepções obtidas a partir do processo de avaliação.

6. Construção dos Modelos

Os conjuntos de dados foram divididos em subconjuntos de treinamento e teste, seguindo uma proporção de divisão de 85/15. Após essa divisão, a técnica SMOTE foi aplicada exclusivamente ao conjunto de treinamento. A escolha de aplicar o SMOTE apenas ao conjunto de treinamento decorre da necessidade de preservar a integridade do conjunto de teste, evitando influências de amostras sintéticas.

O subconjunto de treinamento foi utilizado como base para o treinamento e otimização do modelo, permitindo que este se ajustasse aos dados e aprendesse padrões. Por outro lado, o subconjunto de teste forneceu um conjunto separado de dados não vistos anteriormente pelo modelo, essencial para avaliar o desempenho do modelo em situações mais realistas e verificar sua capacidade de generalização. Essa abordagem de dividir os dados e aplicar o SMOTE de maneira seletiva visa garantir uma avaliação imparcial do modelo.

Nesta etapa, foram desenvolvidos modelos utilizando os algoritmos *XGBoost*, *Gradient Boosting* e *Random Forest*, que são amplamente reconhecidos por sua eficácia em tarefas de classificação. Para aprimorar o desempenho desses modelos, empregou-se um processo de busca otimizada, conhecido como busca bayesiana (Zhang, 2004), que se destaca por sua eficiência computacional, exigindo menos avaliações de modelos devido ao uso de informações anteriores. Adicionalmente, a busca bayesiana é eficaz na descoberta de melhores hiperparâmetros, utilizando um modelo probabilístico para direcionar a busca e alcançar configurações otimizadas de forma mais rápida.

Além disso, a validação cruzada (Berrar *et al.*, 2019) com 5 *splits* foi empregada para avaliar os modelos. A validação cruzada é uma técnica que divide a base de dados em partes e realiza rodadas de treinamento e teste. Em cada rodada, uma das partes é utilizada como conjunto de teste, enquanto

as outras quatro são usadas para treinar o modelo. Esse procedimento é repetido, de forma que cada parte da base de dados seja usada como teste uma vez. Isso permite uma avaliação mais confiável do desempenho do modelo, pois ele é testado em diferentes subconjuntos dos dados, ajudando a identificar se o modelo é consistente em sua precisão ou se varia significativamente entre diferentes divisões dos dados.

Os modelos foram treinados e otimizados considerando diferentes intervalos de hiperparâmetros. No caso do algoritmo *Random Forest*, os hiperparâmetros ajustados incluíram o número de estimadores (*n_estimators*) variando de 50 a 200, o critério de divisão (*criterion*) sendo escolhido entre "gini", "entropy", e "log_loss", e a profundidade máxima da árvore (*max_depth*) variando de 5 a 25.

Já para o algoritmo *Gradient Boosting*, foram explorados parâmetros como a função de perda (*loss*), com opções entre "log_loss" e "exponential", a taxa de aprendizado (*learning_rate*) variando de 0,001 a 0,5, o número de estimadores (*n_estimators*) de 50 a 200, e a profundidade máxima da árvore (*max_depth*) variando de 5 a 25.

No caso do *XGBoost*, os parâmetros ajustados incluíram a taxa de aprendizado (*learning_rate*) variando de 0,001 a 0,5, o número de estimadores (*n_estimators*) de 50 a 200, e a profundidade máxima da árvore (*max_depth*) variando de 5 a 25.

Essa abordagem sistemática de exploração de intervalos proporcionou uma compreensão abrangente do impacto desses parâmetros no desempenho dos modelos, contribuindo para uma análise aprofundada dos resultados obtidos.

Os melhores hiperparâmetros encontrados para os modelos de *ensemble* com a busca bayesiana estão presentes nas Tabelas 1, 2 e 3. Quando o parâmetro não é aplicável a determinado algoritmo, foi colocado o sinal "-", para representar a ausência.

Tabela 1. Configuração dos hiperparâmetros otimizados dos algoritmos de *ensemble* da base de dados 1

Algoritmo	Criterion	Max Depth	N. Estimators	Learning Rate	Loss
Random Forest	Entropy	25	190	-	-
Gradient Boosting	-	11	162	0,4773	Log Loss
XGBoost	-	12	177	0,1719	-

Tabela 2. Configuração dos hiperparâmetros otimizados dos algoritmos de *ensemble* da base de dados 2

Algoritmo	Criterion	Max Depth	N. Estimators	Learning Rate	Loss
Random Forest	Entropy	24	114	-	-
Gradient Boosting	-	5	183	0,0453	Exponential
XGBoost	-	22	96	0,2696	-

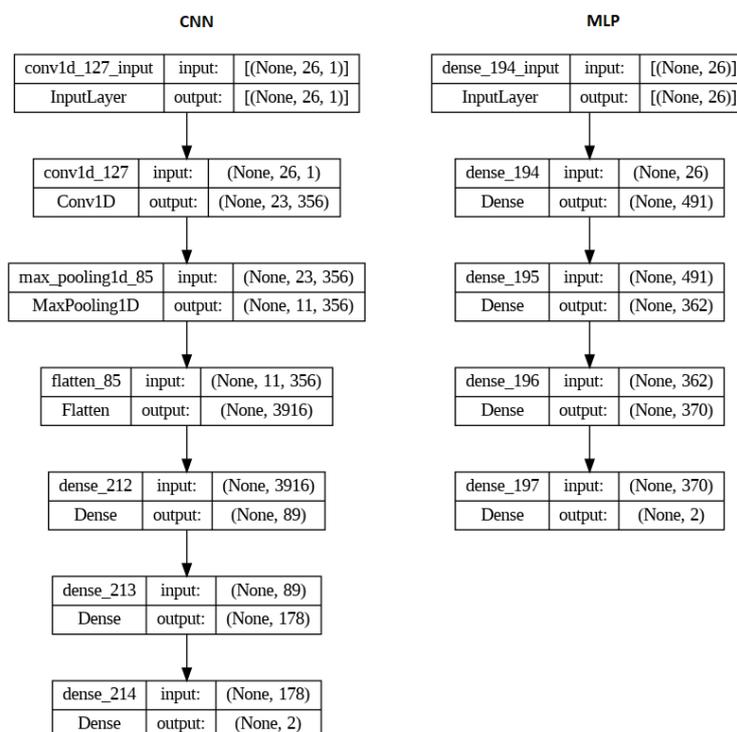
Um procedimento análogo foi conduzido para as RNAs, em que as arquiteturas da CNN e do MLP foram concebidas e implementadas com diversas configurações de hiperparâmetros. Os intervalos específicos estabelecidos para otimização dos hiperparâmetros foram, no caso do MLP, os parâmetros abordados incluíram o número de camadas (*num_layers*) de 1 a 3, a quantidade de

Tabela 3. Configuração dos hiperparâmetros otimizados dos algoritmos de *ensemble* da base de dados 3

Algoritmo	Criterion	Max Depth	N. Estimators	Learning Rate	Loss
Random Forest	Gini	13	173	-	-
Gradient Boosting	-	9	84	0,2068	Log Loss
XGBoost	-	13	199	0,3402	-

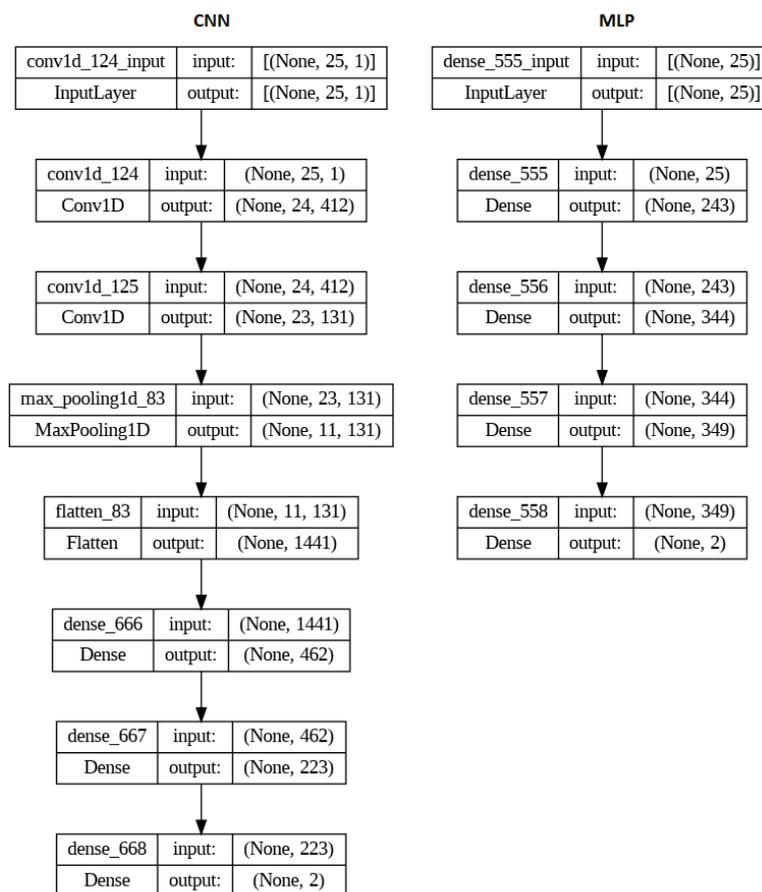
neurônios em cada camada (*neurons1*, *neurons2* e *neurons3*) variando de 32 a 512, a função de ativação (*activation*) podendo ser "relu", "tanh" ou "sigmoid" e o otimizador (*optimizer*) possuindo as variações "adam", "sgd", "rmsprop", "adadelta", "adamax", "nadam". Já para a CNN, os intervalos englobaram o número de camadas convolucionais e densas (*num_layers_conv* e *num_layers_dense*) de 1 a 2, a quantidade de neurônios em cada camada convolucional e densa (*neurons_conv1*, *neurons_conv2*, *neurons_dense1* e *neurons_dense2*) variando de 32 a 512, o tamanho do kernel (*kernel_size*) de 2 a 3, o tamanho do pool (*pool_size*) de 2 a 4, a função de ativação e o otimizador, com as mesmas opções mencionadas anteriormente. Essa abordagem abrangente permitiu uma exploração ampla do espaço de hiperparâmetros para encontrar configurações eficazes em ambas as arquiteturas de redes neurais.

Figura 4. Configuração dos hiperparâmetros otimizados das RNAs da base de dados 1



Desenvolveu-se uma função personalizada para otimizar os hiperparâmetros das RNAs. Dado que as redes neurais exigem a definição de quantidades específicas de camadas e neurônios por camada, essa função foi projetada para receber esses hiperparâmetros como entrada. Posteriormente, ela construía a rede neural com base nos hiperparâmetros fornecidos e realizava o ajuste com base nas amostras do conjunto de treinamento. Além disso, a validação cruzada de 5 *splits* foi empregada,

Figura 5. Configuração dos hiperparâmetros otimizados das RNAs da base de dados 2



reservando uma proporção de 30% para os testes dentro do treinamento da rede. Ao final de cada iteração, a função retornava a acurácia, proporcionando uma avaliação do desempenho dos modelos a fim de realizar a melhor escolha dos hiperparâmetros.

Os melhores hiperparâmetros encontrados após a otimização para os modelos de RNAs foram os indicados nas Figuras 4, 5 e 6. Esse processo realizado se refere a etapa de modelagem do CRISP-DM explicada no ponto 3.4 deste artigo.

7. Resultados Obtidos

Esse estudo se concentra na avaliação do desempenho de diferentes modelos na previsão de risco de crédito, com um enfoque na identificação de inadimplentes. A importância dessa tarefa reside na necessidade de evitar perdas financeiras, destacando-se o objetivo central de identificar com precisão os casos de não pagamento. Para isso, foram utilizadas métricas como acurácia, precisão, *recall*, *f1-score* e matriz de confusão para avaliar o desempenho dos modelos.

Os resultados obtidos na base de dados da Tabela 4, revelam que todos os modelos alcançaram desempenhos sólidos na previsão de risco de crédito, com acurácias gerais superiores a 87%.

O *Gradient Boosting* e o *XGBoost* emergiram como os melhores classificadores, alcançando uma acurácia de 94%. Esses modelos possuem um ponto de atenção quanto a identificação de risco, apresentando uma precisão de 94%, e *recalls* de 76% ambos, a análise dessas métricas indicam que,

Figura 6. Configuração dos hiperparâmetros otimizados das RNAs da base de dados 3

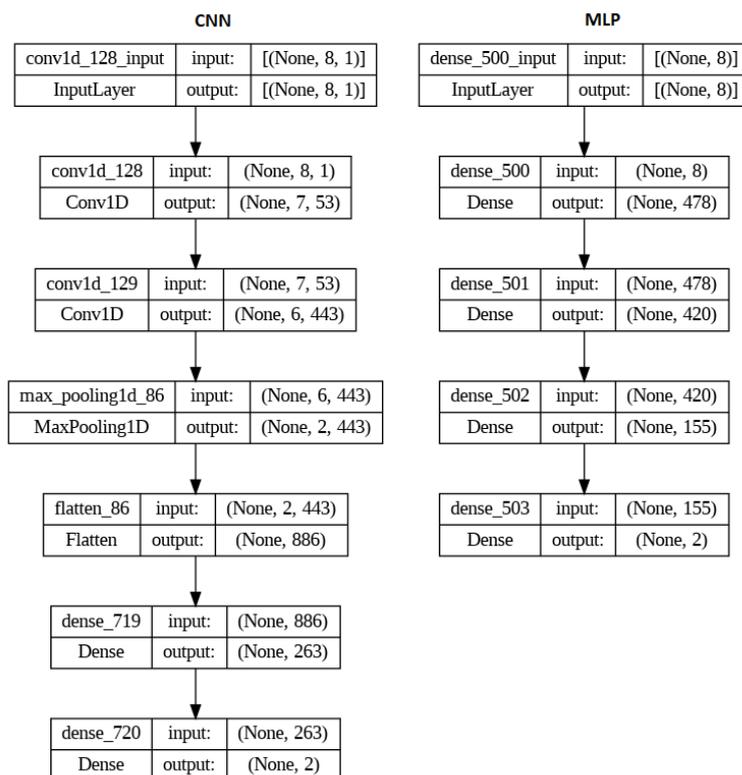


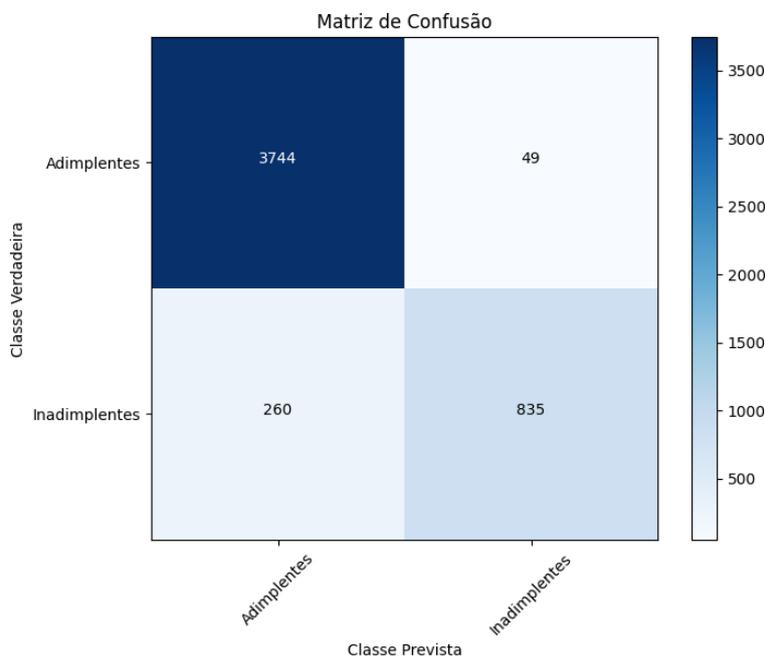
Tabela 4. Resultados dos modelos no conjunto de teste da base de dados 1

Algoritmo	Acurácia	Precisão	Recall	F1-Score
Gradient Boosting	0,94	0,94	0,76	0,84
XGBoost	0,94	0,94	0,76	0,84
Random Forest	0,93	0,90	0,75	0,82
MLP	0,90	0,81	0,72	0,77
CNN	0,88	0,71	0,75	0,73

embora o modelo seja preciso na identificação dos casos de risco, ele não é tão bom em identificar todos os casos arriscados disponíveis no conjunto de dados, conforme evidenciado pelo *recall* relativamente baixo. Isso sugere que há espaço para melhorias no modelo, especialmente na sensibilidade em relação aos casos de inadimplentes. O *f1-score* apresentou o valor de 0,84, sugerindo uma média equilibrada entre precisão e *recall*, sendo essa métrica útil para avaliar o desempenho do modelo quando há um cenário como o da presente base, de desequilíbrio entre as classes.

A análise da matriz de confusão da Figura 7 revela que o *Gradient Boosting* identificou corretamente 835 casos de inadimplência (verdadeiros negativos), demonstrando sua eficácia na detecção de riscos financeiros. Os falsos negativos, totalizando 49, indicam erros ao classificar casos adimplentes como inadimplentes. Os verdadeiros positivos, com um total de 3.744, mostram que o modelo acertou ao classificar casos adimplentes. Os falsos positivos, contabilizando 260, são erros do modelo ao classificar casos de inadimplência como adimplentes, podendo resultar em perdas financeiras.

Figura 7. Matriz de confusão do conjunto de teste do Gradient Boosting da base de dados 1



Em resumo, o *Gradient Boosting* e o *XGBoost* emergiram como os modelos mais eficazes na identificação de não pagamento, destacando-se desta forma o grupo de algoritmos de *ensemble*. As RNAs, MLP e a CNN, embora tenham alcançado resultados competitivos, ainda possuem espaço para aprimoramentos na detecção de risco. Assim, nesta base de dados específica, as redes neurais demonstraram um desempenho um pouco inferior em comparação com os modelos baseados em *ensemble*.

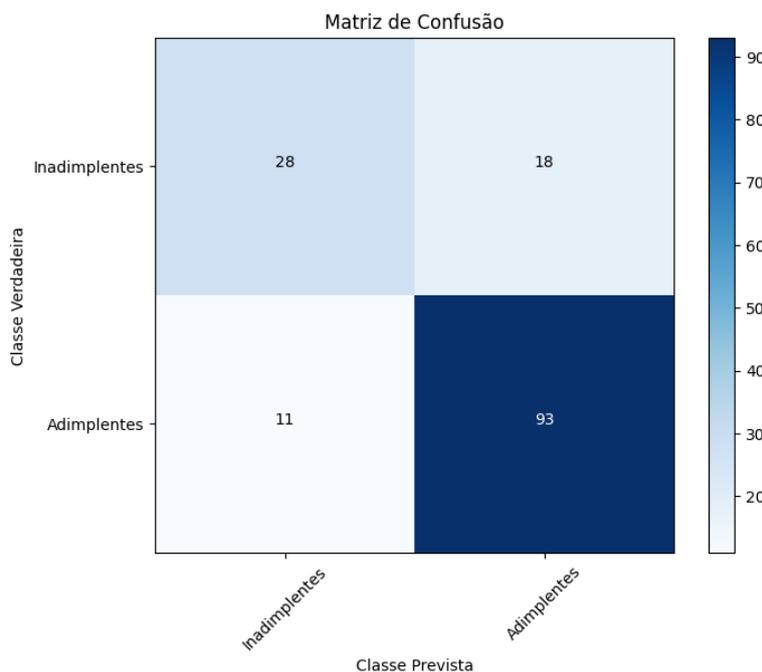
Tabela 5. Resultados dos modelos no conjunto de teste da base de dados 2

Algoritmo	Acurácia	Precisão	Recall	F1-Score
Random Forest	0,81	0,72	0,61	0,66
MLP	0,79	0,66	0,67	0,67
Gradient Boosting	0,79	0,67	0,61	0,64
CNN	0,77	0,64	0,54	0,59
XGBoost	0,72	0,55	0,46	0,50

Na análise dos resultados na base de dados da Tabela 5, que se concentra na identificação dos não pagantes, é possível perceber que todos os modelos avaliados apresentaram um desempenho mediano em comparação com os resultados obtidos na base de dados 1.

O desempenho dos modelos foi avaliado considerando as métricas citadas anteriormente. O *Random Forest* demonstrou uma acurácia razoável com 81%, indicando uma taxa geral de classificações corretas. Sua precisão de 72% sugere uma capacidade de evitar falsos positivos, enquanto o *recall* de 61% mostra sua capacidade de identificar verdadeiros positivos. O equilíbrio entre precisão e *recall* é refletido no *f1-score* de 66%, consolidando o *Random Forest* como o melhor modelo na comparação com os outros presentes neste estudo para essa determinada base.

Figura 8. Matriz de confusão do conjunto de teste do Random Forest da base de dados 2



Os verdadeiros positivos, com um total de 28, indicam que o modelo acertou ao classificar corretamente casos inadimplentes como demonstrado na Figura 8. Porém, os falsos positivos, contabilizando 11, representam erros do modelo ao classificar casos de adimplentes como inadimplência. Portanto, a interpretação da matriz de confusão ressalta a importância de minimizar tanto os falsos positivos quanto os falsos negativos para aprimorar a precisão das previsões em contextos financeiros.

De forma geral, os resultados desta base de dados indicam que nenhum dos modelos conseguiu alcançar uma *performance* excepcional na identificação dos não pagantes. Todos os modelos apresentaram um desempenho mediano, com uma acurácia variando de 72% a 81%. Embora cada algoritmo tenha suas próprias características, nenhum deles se destacou significativamente nesta tarefa específica. Portanto, para melhorar a identificação dos não pagantes, seria necessário explorar estratégias adicionais, como considerar abordagens diferentes de modelagem.

Tabela 6. Resultados dos modelos no conjunto de teste da base de dados 3

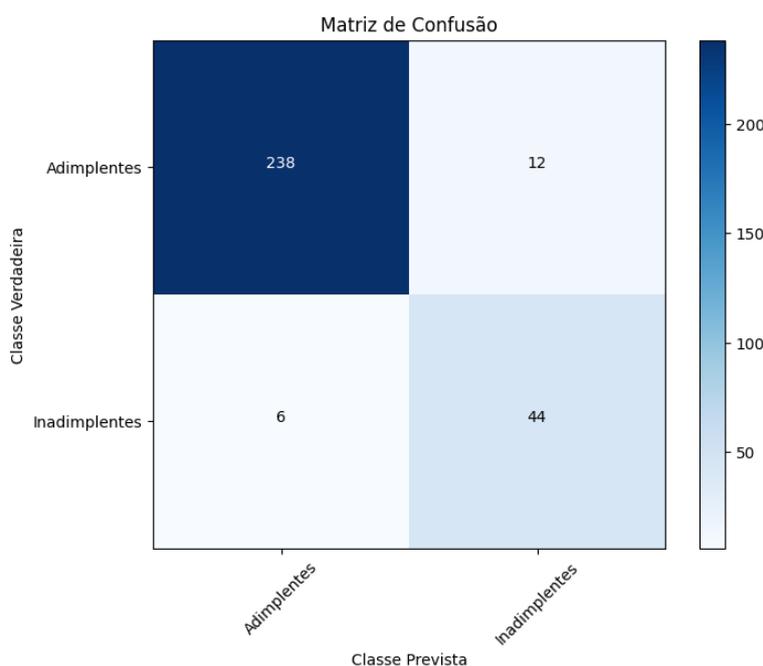
Algoritmo	Acurácia	Precisão	Recall	F1-Score
CNN	0,94	0,79	0,88	0,83
MLP	0,94	0,75	0,94	0,83
XGBoost	0,93	0,76	0,88	0,81
Gradient Boosting	0,92	0,71	0,84	0,77
Random Forest	0,91	0,68	0,88	0,77

Na avaliação dos resultados na base de dados da Tabela 6, focada na identificação dos não pagantes, é evidente que todos os modelos apresentaram um desempenho satisfatório.

A CNN demonstrou uma acurácia sólida de 94%, indicando um desempenho consistente em classificações corretas. No entanto, uma análise detalhada revela áreas passíveis de aprimoramento,

especialmente na minimização de falsos positivos, evidenciados pela precisão de 79% e *recall* de 88%. Apesar dessas considerações, o *f1-score* de 83% destaca um equilíbrio razoável entre precisão e *recall*. Ao comparar com o MLP, a CNN obteve um desempenho um pouco superior, com 282 amostras classificadas corretamente em comparação com as 281 do MLP, conforme evidenciado pela matriz de confusão. Embora a diferença entre os modelos seja sutil, vale destacar que, mesmo com essa diferença modesta, a inovação de aplicar uma CNN a dados tabulares ressalta a possibilidade de explorar arquiteturas de redes neurais não convencionais para problemas dessa natureza, abrindo caminho para contribuições interessantes no campo da análise de risco de crédito.

Figura 9. Matriz de confusão do conjunto de teste da CNN da base de dados 3



A matriz de confusão da Figura 9, reflete o desempenho da CNN na classificação de classes como inadimplentes e adimplentes. Com 44 verdadeiros negativos e 238 verdadeiros positivos, o modelo demonstra eficácia em identificar ambas as categorias. No entanto, 12 falsos negativos e 6 falsos positivos destacam a necessidade de melhorar a precisão na classificação.

Em síntese, os resultados demonstram um desempenho global satisfatório da base, porém, há uma necessidade de aprimorar a precisão na identificação dos não pagantes. É importante ressaltar que as RNAs mostraram-se mais eficazes nesse contexto, especialmente a CNN, abrindo caminho para uma nova perspectiva de modelo a ser considerada como solução no problema de risco de crédito. Aperfeiçoamentos são importantes para reduzir classificações incorretas, especialmente os falsos positivos, como nesse cenário apresentado.

8. Considerações Finais

Em conclusão, a análise de risco de crédito é um componente essencial do sistema financeiro, desempenhando um papel crítico na promoção da estabilidade, no estímulo ao crescimento econômico e na prevenção de inadimplência. O avanço tecnológico, particularmente com o uso de RNAs e técnicas de *ensembles*, oferece oportunidades significativas para melhorar essas análises, especialmente em cenários com dados desequilibrados e complexidades não lineares.

A adaptabilidade é a chave para enfrentar a dinâmica do cenário econômico, e a capacidade de incorporar rapidamente novos dados e *insights* é fundamental. Portanto, ao adotar abordagens inovadoras baseadas em tecnologias avançadas, é possível fortalecer a integridade dos mercados de crédito, promover a estabilidade financeira e facilitar práticas de empréstimos responsáveis.

A análise dos resultados dos modelos de previsão de risco de crédito em três bases de dados diferentes revela uma variedade de desempenhos, cada um com suas próprias forças e fraquezas. Esses modelos desempenham um papel importante na tomada de decisões financeiras informadas, permitindo que instituições e empresas avaliem a capacidade de pagamento de mutuário.

No entanto, é importante observar que o contexto e os objetivos específicos podem influenciar a escolha do modelo mais adequado. Dado que a ênfase deste estudo está na minimização de riscos e na identificação precisa de casos de alto risco, destaca-se que o *Gradient Boosting* e o *XGBoost* se destacaram ao acertarem significativamente mais na classificação do número de amostras de inadimplentes. Suas altas precisões na identificação desse grupo específico ressaltam a solidez desses modelos e sugerem que podem ser escolhas satisfatórias para cenários em que a identificação precisa de inadimplentes é prioritária.

Em resumo, a escolha do modelo de previsão de risco de crédito depende de considerações específicas, como a importância relativa de precisão e *recall*, a disponibilidade de dados, os recursos computacionais e os objetivos do projeto. No entanto, os modelos avaliados neste estudo podem fornecer *insights* de melhoria na tomada de decisões financeiras. A aplicação de técnicas de aprendizado de máquina demonstram potencial para aprimorar a análise de risco de crédito e contribuir para um sistema financeiro mais robusto e resiliente.

É importante reconhecer que esse trabalho possui algumas limitações inerentes. Em primeiro lugar, a utilização de dados fictícios pode não refletir completamente a complexidade e a variabilidade dos dados reais de crédito. Embora os dados fictícios sejam úteis para fins de experimentação, os resultados obtidos podem não ser diretamente transferíveis para cenários do mundo real. Esse artigo não teve acesso a dados reais pelo motivo de dados de crediário se tratarem de dados sensíveis, sendo assim indisponíveis publicamente. Além disso, a não realização de testes em ambiente real é uma limitação importante. Embora os experimentos tenham sido conduzidos em condições controladas, a eficácia das abordagens propostas pode ser influenciada por fatores não considerados nos dados fictícios, como flutuações econômicas e mudanças nas políticas financeiras. Por fim, a limitação das bases de dados pequenas também é relevante, pois pode limitar a generalização dos resultados. Bases de dados menores podem não capturar completamente a diversidade de cenários de análise de risco de crédito, o que pode impactar a robustez e a aplicabilidade das abordagens propostas em situações mais complexas. Portanto, essas limitações devem ser consideradas ao interpretar os resultados e ao aplicar as descobertas deste estudo em contextos do mundo real.

Em trabalhos futuros, a aplicação deste estudo em bases de dados reais de instituições financeiras seria de extrema relevância. Isso permitiria avaliar a eficácia do modelo em um ambiente de operação financeira do mundo real, onde as decisões de crédito têm um impacto direto nas operações comerciais e na economia em geral. Além disso, a utilização de dados reais proporcionaria *insights* valiosos sobre como o modelo se comporta em cenários complexos e dinâmicos, onde as condições econômicas podem variar rapidamente. Também possibilitaria uma análise mais abrangente das implicações práticas das previsões do modelo, ajudando a instituições financeiras a tomar decisões informadas e aprimorar suas práticas de concessão de crédito. Em última análise, a aplicação em bases de dados reais abriria caminho para uma validação mais sólida e uma implementação prática

que poderia beneficiar tanto as instituições financeiras quanto os mutuários.

Referências

- ABEDIN, M. Z. *et al.* Combining weighted smote with ensemble learning for the class-imbalanced prediction of small business credit risk. *Complex & Intelligent Systems*, Springer, p. 1–21, 2022. 4, 5, 6
- ALONSO, A.; CARBO, J. M. Machine learning in credit risk: Measuring the dilemma between prediction and supervisory cost. Banco de Espana Working Paper, 2020. 2
- BAO, W.; LIANJU, N.; YUE, K. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Systems with Applications*, Elsevier, v. 128, p. 301–315, 2019. 2, 3
- BERRAR, D. *et al.* *Cross-Validation*. 2019. 11
- BROWNLEE, J. How to use standardscaler and minmaxscaler transforms in python. *Machine Learning Mastery*, v. 10, 2020. 10
- BUSSMANN, N. *et al.* Explainable machine learning in credit risk management. *Computational Economics*, Springer, v. 57, p. 203–216, 2021. 2
- CARUSO, G. *et al.* Cluster analysis for mixed data: An application to credit risk evaluation. *Socio-Economic Planning Sciences*, Elsevier, v. 73, p. 100850, 2021. 4
- CHAWLA, N. V. *et al.* Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, v. 16, p. 321–357, 2002. 3
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. [S.l.: s.n.], 2016. p. 785–794. 5
- CREDIT Risk. 2020. (<https://www.kaggle.com/datasets/upadorprofzs/credit-risk>). Acesso em: 21 de jun de 2023. 9
- CREDIT Risk Dataset. 2020. (<https://www.kaggle.com/datasets/laotse/credit-risk-dataset>). Acesso em: 21 de jun de 2023. 7
- GERMAN Credit. 2019. (<https://online.stat.psu.edu/stat508/book/export/html/796>). Acesso em: 21 de jun de 2023. 8
- LI, Z. *et al.* A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, IEEE, 2021. 5
- LIU, W.; FAN, H.; XIA, M. Credit scoring based on tree-enhanced gradient boosting decision trees. *Expert Systems with Applications*, Elsevier, v. 189, p. 116034, 2022. 5
- MARTÍNEZ-PLUMED, F. *et al.* Crisp-dm twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 33, n. 8, p. 3048–3061, 2019. 5
- MIKKULAINEN, R. *et al.* Evolving deep neural networks. In: *Artificial intelligence in the age of neural networks and brain computing*. [S.l.]: Elsevier, 2019. p. 293–312. 5
- RASCHKA, S. *Python machine learning*. [S.l.]: Packt publishing ltd, 2015. 2
- Instituto Federal de Educação, Ciências e Tecnologia de Pernambuco. *Campus Paulista*. Curso de 20 Análise e Desenvolvimento de Sistemas. 22 de dezembro de 2023.

- RIGATTI, S. J. Random forest. *Journal of Insurance Medicine*, American Academy of Insurance Medicine 1700 Magnavox Way, Fort Wayne, IN 46804, v. 47, n. 1, p. 31–39, 2017. [5](#)
- RODRÍGUEZ, P. *et al.* Beyond one-hot encoding: Lower dimensional target embedding. *Image and Vision Computing*, Elsevier, v. 75, p. 21–31, 2018. [10](#)
- SCHWARZ, K. Mind the gap: Disentangling credit and liquidity in risk spreads. *Review of Finance*, Oxford University Press, v. 23, n. 3, p. 557–597, 2019. [2](#)
- SHEN, F. *et al.* A novel ensemble classification model based on neural networks and a classifier optimisation technique for imbalanced credit risk evaluation. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 526, p. 121073, 2019. [3](#), [6](#)
- TAUD, H.; MAS, J. Multilayer perceptron (mlp). *Geomatic approaches for modeling land change scenarios*, Springer, p. 451–455, 2018. [5](#)
- TELES, G. *et al.* Artificial neural network and bayesian network models for credit risk prediction. *Journal of Artificial Intelligence and Systems*, Institute of Electronics and Computer, v. 2, n. 1, p. 118–132, 2020. [4](#), [5](#)
- TURNER, R. *et al.* Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020. In: PMLR. *NeurIPS 2020 Competition and Demonstration Track*. [S.l.], 2021. p. 3–26. [7](#)
- WICKHAM, H.; WICKHAM, H. *Data analysis*. [S.l.]: Springer, 2016. [2](#)
- WINSTON, P. H. *Artificial intelligence*. [S.l.]: Addison-Wesley Longman Publishing Co., Inc., 1984. [2](#)
- YANG, M. *et al.* Deep neural networks with l1 and l2 regularization for high dimensional corporate credit risk prediction. *Expert Systems with Applications*, Elsevier, v. 213, p. 118873, 2023. [5](#)
- YEGNANARAYANA, B. *Artificial neural networks*. [S.l.]: PHI Learning Pvt. Ltd., 2009. [3](#)
- ZHANG, C.; MA, Y. *Ensemble machine learning: methods and applications*. [S.l.]: Springer, 2012. [3](#)
- ZHANG, H. The optimality of naive bayes. *Aa*, v. 1, n. 2, p. 3, 2004. [4](#), [11](#)