INSPER

Gustavo de Souza Sampaio Moreira Piegas

Eficácia dos Modelos de Pontuação de Crédito na Predição de Risco de Inadimplência: Uma Revisão e Análise Comparativa de Modelos

Trabalho de Conclusão de Curso II

São Paulo – SP

2024

Gustavo de Souza Sampaio Moreira Piegas

Eficácia dos Modelos de Pontuação de Crédito na Predição de Risco de Inadimplência: Uma Revisão e Análise Comparativa de Modelos

Trabalho de Conclusão de Curso II

Orientador: Prof. Dr. Rinaldo Artes

São Paulo – SP

2024

RESUMO

Esta monografia investiga a eficácia de diferentes modelos de *credit scoring* aplicados a carteiras de crédito para pessoas físicas, destacando o impacto das reformas econômicas e avanços tecnológicos, como o uso de Big Data e inteligência artificial, no período pós-2000. O estudo analisa a evolução histórica do mercado de crédito, desde os sistemas informais baseados na confiança pessoal até os sofisticados algoritmos atuais utilizados por instituições financeiras. Três metodologias principais são abordadas: redes neurais, árvores de decisão e regressão logística, comparando suas capacidades preditivas e eficácia na gestão do risco de inadimplência. Utilizando a base de dados detalhada "*LendingClub Loan Data*", o trabalho revela que, embora modelos tradicionais ainda sejam robustos, técnicas avançadas, como redes neurais, oferecem melhorias incrementais na precisão das previsões. A pesquisa conclui com uma discussão sobre os desafios éticos e regulatórios associados ao uso de tecnologias emergentes na avaliação de crédito, propondo recomendações para futuras investigações e aplicações práticas no setor financeiro.

SUMÁRIO

1 Introdução 4

1.1 Contexto geral do mercado de crédito 5

1.2 Justificativas para o estudo 6

1.3 Objetivos do estudo 7

1.4 Estrutura do Trabalho 8

2.1 Credit Scoring 10

2.2 Histórico do mercado de crédito 11

2.2 Métodos de Credit Scoring 14

2.2.1 Modelos de Credit Scoring: metodologias, abordagens e eficiência 14

2.2.2 Credit Scoring com Redes Neurais 16

Credit Scoring com Árvores de decisão 17

Credit Scoring com Regressão Logística (LG) 18

2.3 Considerações finais 19

3. Metodologia 21

3.1 Base de dados utilizada 21

3.2 Processamento de dados 22

3.3 Modelos de Credit Scoring 22

3.4 Treinamento e validação dos modelos 24

3.5 Métricas de avaliação 24

4. Resultados 25

4.1 Análise descritiva 25

4.2 Desempenho dos modelos 27

4.3 Análise dos pontos de corte 29

4.3 Considerações finais 30

5. Conclusão 31

ANEXO A 33

REFERÊNCIAS 37

1 Introdução

1.1 Contexto geral do mercado de crédito

Desde tempos antigos, os mecanismos de concessão de crédito têm sido fundamentais para o desenvolvimento econômico e comercial das nações. Inicialmente baseados na confiança pessoal, esses sistemas evoluíram para estruturas mais sofisticadas e regulamentadas, especialmente com o advento do comércio internacional e a criação de bancos modernos. Exemplos históricos incluem os arranjos de crédito informais em sociedades antigas e a evolução para práticas mais formais, como as observadas na Europa e no Império Otomano.

Um marco significativo na história do crédito foi a fundação do Banco da Inglaterra em 1694, que impulsionou o crescimento substancial do mercado de crédito. Posteriormente, o desenvolvimento do cartão de crédito nos anos 1950 marcou uma nova era na especialização dos modelos de *credit scoring*. Este avanço foi crucial para a expansão do crédito, especialmente a pessoas físicas (BANCO DA INGLATERRA, 2024).

A evolução do mercado de crédito é caracterizada pela constante transformação, impulsionada por desenvolvimentos econômicos e regulatórios. Serviços financeiros fundamentais, como pagamentos, intermediação de operações de crédito e seguros, são fundamentais para a economia. As características do sistema financeiro no fornecimento desses serviços mudam de maneira contínua, especialmente com a integração da inteligência artifical.

No contexto de *credit scoring*, a modelagem associada a grandes bancos de dados é utilizada para avaliar a capacidade de pagamento de um tomador, permitindo uma análise mais completa do risco de crédito. Empregando algoritmos avançados e modelos preditivos, os dados coletados de diversas fontes, como histórico de transações, comportamento de pagamento, e até mesmo dados de redes sociais, são analisados para prever a probabilidade de inadimplência. Esta abordagem fornece às instituições financeiras uma ferramenta mais eficaz para tomar decisões de crédito, reduzindo o risco de inadimplência e possibilitando a oferta de produtos de crédito personalizados (SMITH e DOE, 2023).

Com a introdução do cartão de crédito e a ascensão da *Big Data*, a pontuação de crédito entrou em uma nova era. Modelos de *credit scoring*, agora aprimorados por algoritmos sofisticados e análise de grandes bases de dados, oferecem avaliações mais precisas e eficientes. Essas inovações são fundamentais para a gestão eficaz de carteiras de crédito e para a mitigação dos riscos associados a elas.

A introdução do cartão de crédito e a aplicação de tecnologias de big data nos sistemas de pontuação de crédito também têm um papel crucial na reforma da economia de mercado e no desenvolvimento econômico. Os sistemas de informação de crédito modernos, impulsionados pela tecnologia, ajudam a reduzir o risco de crimes de crédito e permitem que as instituições financeiras, como bancos, manejem negócios de crédito de forma mais eficiente e precisa para diferentes grupos de pessoas (HINDAWI, 2021).

1.2 Justificativas para o estudo

O estudo de modelos de *credit scoring*, ao longo dos últimos 41 anos, tem enfrentado desafios e oportunidades significativas. Este campo está passando por uma transformação quantitativa, que foca em abordagens centradas em dados e a utilização de fontes de dados não tradicionais, assim como seus aspectos regulatórios, são áreas promissoras para pesquisas futuras (ONAY e ÖZTÜRK, 2018).

A análise de 258 artigos acadêmicos revisados por pares, provenientes de 147 periódicos, demonstra que o *credit scoring* é uma área em evolução contínua, com temas de pesquisa diversificados e tendências emergentes. A introdução do Big Data nos modelos de *credit scoring* trouxe tanto riscos quanto benefícios, ressaltando a necessidade de desenvolvimento de novos modelos de crédito que identifiquem fontes de dados seguras e as convertam em insights úteis em conformidade com as regulamentações (ONAY e ÖZTÜRK, 2018).

A importância do *credit scoring* é destacada pela sua capacidade de transformar grandes bases de dados em informações valiosas para corporações, governos e entidades sem fins lucrativos que desenvolvem e utilizam novas tecnologias na avaliação de crédito (ONAY e ÖZTÜRK, 2018).

É notável uma grande expansão da concessão de crédito a pessoas físicas no Brasil e no Mundo. Este crescimento é impulsionado por fatores como avanços tecnológicos, mudanças na regulamentação e maior inclusão financeira. Dados do Banco Central do Brasil (Bacen) mostram um aumento contínuo no volume de crédito concedido a pessoas físicas. Esse crescimento é atribuído a fatores como a redução da taxa de juros e a implementação de políticas públicas de crédito. O aumento do crédito tem impacto direto no consumo das famílias. Conforme discutido por Silva e Santos (2021), o crédito facilita a aquisição de bens de maior valor, estimulando a economia.

Além disso, o crédito pessoal pode servir como uma ferramenta para investimentos em pequenos negócios, impulsionando o empreendedorismo. Apesar dos benefícios, a expansão do crédito também traz riscos. A literatura, incluindo Ferreira e Oliveira (2022), aponta para preocupações com o sobre-endividamento das famílias e possíveis impactos na estabilidade financeira.

O crédito desempenha um papel vital na economia brasileira, atuando como um importante motor de crescimento econômico. Durante os períodos eleitorais, o crédito adquire uma dimensão adicional, frequentemente sendo utilizado como instrumento de política econômica. O crédito influencia diretamente o consumo e o investimento. Segundo Pereira e Lima (2022), períodos de expansão de crédito estão associados a um aumento na atividade econômica, impulsionando o PIB.

1.3 Objetivos do estudo

O objetivo geral deste estudo é analisar a eficácia de diferentes modelos de *credit scoring* em carteiras de crédito a pessoas físicas, com ênfase no período pós-2000. Este período é marcado por significativas reformas microeconômicas que catalisaram um aumento substancial na concessão de crédito no Mundo, principalmente para indivíduos. A análise focará em avaliar como esses modelos se adaptaram às mudanças no mercado de crédito e sua efetividade em prever riscos. Desta forma, os objetivos específicos deste trabalho são:

- Realizar uma revisão histórica da evolução do *Credit Scoring;*
- Estimar modelos de Credit Scoring;
- Comparar e avaliar modelos de Credit Scoring.

Revisão Histórica e Evolução do *Credit Scoring*: Este objetivo busca traçar o desenvolvimento dos modelos de *credit scoring* no Brasil, particularmente após o ano 2000. Será dada atenção especial às mudanças legislativas e econômicas que influenciaram a prática de concessão de crédito, bem como às inovações nos modelos de *scoring* ao longo do tempo.

Avaliação da Eficácia dos Modelos de *Credit Scoring*: Aqui, o estudo se concentrará em avaliar a eficácia dos modelos de *credit scoring* no contexto das mudanças microeconômicas pós-2000. Será investigado como esses modelos responderam ao aumento da concessão de crédito e sua precisão em prever riscos de inadimplência, oferecendo insights sobre a adequação desses sistemas em um ambiente econômico em transformação.

Comparação de Modelos de *Credit Scoring*: Este objetivo envolve uma análise comparativa entre diferentes modelos de *credit scoring* utilizados no mundo, avaliando suas metodologias, abordagens e eficiência. O foco estará em como esses modelos se adaptam às necessidades das pessoas físicas.

A base de dados utilizada é a "*LendingClub Loan Data*". Esta base de dados é disponibilizada pela LendingClub, uma empresa de empréstimos peer-to-peer dos Estados Unidos. Contém informações detalhadas sobre empréstimos realizados através da plataforma, incluindo características dos mutuários (como pontuação de crédito, histórico de crédito, renda anual), detalhes do empréstimo (como montante, taxa de juros, finalidade do empréstimo) e o status do pagamento (incluindo informações sobre inadimplência). Esta base é particularmente valiosa para analisar o comportamento de crédito no contexto de empréstimos peer-to-peer, oferecendo insights sobre como variáveis individuais e financeiras influenciam a probabilidade de inadimplência.

1.4 Estrutura do Trabalho

Este TCC é estruturado para analisar de forma aprofundada os modelos de *credit scoring*, com foco especial no período pós-2000. A monografia é dividida em capítulos que se alinham sistematicamente aos objetivos do estudo, permitindo uma análise abrangente e coerente do tema.

Este capítulo 1 estabeleceu o contexto do estudo, apresentando o mercado de crédito e a importância dos modelos de *credit scoring*. Ele define claramente os objetivos gerais e específicos do trabalho, contextualizando a importância da pesquisa no cenário econômico atual.

O capítulo 2, de revisão da literatura, é focado em fornecer um alicerce teórico sólido, este capítulo aborda a evolução dos modelos de *credit scoring*, especialmente após as reformas microeconômicas do início dos anos 2000. Aqui, serão analisadas as mudanças legislativas, econômicas e as inovações nos modelos de *scoring*, atendendo ao primeiro objetivo específico do estudo.

O capítulo 3, de metodologia, detalhará a metodologia adotada para a análise comparativa dos modelos de *credit scoring* e para avaliar sua eficácia. Serão discutidas as abordagens metodológicas para comparar os modelos, assim como os critérios e ferramentas utilizadas para avaliar a eficácia desses modelos no período pós-2000.

O capítulo 4, de resultados, atende aos objetivos específicos de comparar os principais modelos de *credit scoring* e avaliar sua qualidade de ajuste após 2000, este capítulo apresenta os resultados obtidos através da metodologia aplicada. Serão discutidos os achados relativos às características, eficiência e adaptação dos modelos de credit scoring e como eles responderam às mudanças econômicas do período estudado.

O capítulo final, de conclusão, reunirá todas as descobertas, relacionando-as aos objetivos estabelecidos na introdução. Serão discutidas as implicações desses resultados para o campo da economia, a prática de concessão de crédito e as políticas econômicas. Também serão propostas recomendações para pesquisas futuras com base nas limitações e nos achados do estudo.

Cada capítulo será construído para contribuir de forma significativa para a compreensão holística da eficácia e da evolução dos modelos de credit scoring no mundo. Juntos, eles fornecem informações valiosas sobre a interação entre políticas econômicas, inovação em modelagem de risco e práticas de concessão de crédito no contexto mundial.

2. Revisão da Literatura

2.1 Credit Scoring

Os modelos de *credit scoring* atribuem pontuações a cada variável de decisão de crédito de um proponente, através de técnicas estatísticas. Tais modelos visam a definição de características que distingam os bons dos maus créditos (LEWIS, 1992).

Através de pesos pré-determinados para cada fator a ser considerado, os modelos de *credit scoring* podem retornar, por exemplo, uma pontuação do potencial devedor que indica o valor esperado da probabilidade de quitação de dívida, correspondente à linha de crédito por ele demandada, ou ainda, o valor esperado da probabilidade de pertencer a cada classificação de risco definida para a linha de crédito. A partir dessa análise da previsão do modelo, cabe ao gestor responsável optar pela concessão ou não de recursos a quem os demanda.

Como ressalta Saunders (2000), o *score* em geral é usado para classificação de créditos como adimplentes ou inadimplentes, bons ou maus, desejáveis ou não, de acordo com a pontuação obtida. Esta classificação serve como orientação à decisão do gestor em relação à liberação ou não do crédito solicitado.

A definição dos fatores a serem considerados em cada modelo se dá por meios estatísticos e é ela que difere um modelo de outro. Já Sicsu (1998a; 1998b) ratifica que a estrutura regular para o desenho de um modelo de c*redit scoring* apresenta os seguintes processos:

- Planejamento e definições: situações e linhas de crédito para os quais serão desenvolvidos o modelo; finalidades; tipos de proponentes; definição de inadimplência a ser adotada; horizonte de previsão do modelo;
- Identificação das variáveis: caracterização do potencial clientes do crédito; caracterização da linha; escolha das variáveis significativas; análise das restrições ao modelo;
- Coleta de dados e tratamento amostral: eleição e dimensionamento da amostra; coleta dos dados; montagem da base de dados;
- Determinação da fórmula de escoragem;
- Determinação do ponto de corte;

Modelos de c*redit scoring* se mostraram, ao longo do tempo, bastante eficazes em suas análises e foram fundamentais para possibilitar o aumento do crédito concedido ao

consumidor, seja ele pessoa física ou jurídica. Parkinson e Ochs (1998) e Caouette, Altman e Narayanan (1998) resumem os principais pontos positivos dos modelos *credit scoring*:

- Consistência: são modelos elaborados com base na carteira de crédito atualmente vigente na instituição e servem tanto para reavaliar constantemente essa carteira, quanto para analisar a concessão a novos proponentes;
- Facilidade: os modelos de avaliação de crédito devem ser simples e de fácil interpretação, com aplicação clara. Os modelos e suas metodologias são comuns e bem entendidos, assim como seus escores;
- Melhor organização das informações: a organização dos dados contribui para a melhoria do processo de análise e concessão de crédito;
- Redução da subjetividade: o uso de método quantitativo com regras claras e bem definidas contribui para a diminuição do subjetivismo na avaliação do risco de crédito;
- Maior eficiência: o uso de modelos *credit scoring* na concessão de crédito direciona os esforços dos analistas apenas ao processo de tomada de decisão;

2.2 Histórico do mercado de crédito

Mecanismos de concessão de crédito ao consumidor foram fundamentais para o desenvolvimento comercial e econômico das sociedades do longo do tempo. Se na antiguidade prazos eram dados com base na confiança entre credor e devedor – o que ainda hoje acontece em menor escala –, o desenvolvimento dos mercados tornou necessária a criação de modelos mais desenvolvidos e pulverizados de viabilização de consumo acima da renda média, em um determinado período, de trabalhadores e empresas.

O desenvolvimento dos mecanismos de concessão de crédito ao longo da história tem desempenhado um papel fundamental na evolução das atividades econômicas e comerciais globais. Desde sistemas informais baseados em relações interpessoais até complexos sistemas institucionais regulados por governos, o crédito tem sido uma alavanca para o crescimento econômico e o desenvolvimento comercial.

Inicialmente, os sistemas de crédito eram predominantemente informais, caracterizados por arranjos de crédito entre indivíduos ou grupos, com termos relativamente flexíveis e, muitas vezes, sem documentação legal (HISTORY OF FINANCE, 2023). Esses sistemas informais eram baseados em mecanismos de reputação, que desempenhavam um papel

fundamental na superação da assimetria de informações e reduzir o risco de seleção adversa e perigo moral (HISTORY OF FINANCE, 2023).

Com o desenvolvimento do comércio internacional, surgiu a necessidade de sistemas de transferência de recursos e de cessão de crédito, levando à formação de bancos modernos (HISTORY OF FINANCE, 2023). Os notários desempenharam um papel crucial nos negócios de crédito em diferentes partes da Europa, atuando como agentes que certificavam documentos legais, incluindo contratos de crédito e letras de câmbio (HISTORY OF FINANCE, 2023).

Os modelos de crédito evoluíram de maneira variada em diferentes partes do mundo. No Império Otomano, por exemplo, sociedades limitadas, chamadas "mufawadas", eram usadas para financiar o comércio de longa distância, enquanto na Inglaterra o surgimento de bancos de ações conjuntas e o uso de instrumentos de papel transferíveis, como letras de câmbio, ajudaram a expandir os mercados de crédito (HISTORY OF FINANCE, 2023).

A revolução do dinheiro de papel em 1968, quando o Presidente Johnson assinou a lei que acabou com a exigência de o Federal Reserve lastrear o dinheiro com ouro, permitiu um aumento significativo na criação de dinheiro e crédito (DUNCAN, 2018). Isso levou a uma expansão considerável e à consolidação do Sistema de Crédito Bancário ao longo das últimas cinco décadas, aumentando sua dependência de garantias governamentais explícitas e implícitas (OFFICE OF FINANCIAL RESEARCH, 2023).

O entendimento do desenvolvimento histórico dos mecanismos de concessão de crédito é fundamental para analisar as políticas e práticas de crédito atuais, revelando a importância crítica desses sistemas na promoção do comércio, comércio e crescimento econômico ao longo da história.

Um marco da criação do que hoje se pode chamar de mercado de crédito foi a fundação do Banco da Inglaterra, em 1694. A partir daí, e até hoje, o estoque de crédito cresceu de maneira substancial, sendo que a criação do cartão de crédito, nos anos 1950, foi fundamental para a especialização dos modelos de *credit scoring*.

Com o objetivo de chegar a uma compreensão completa dos ciclos financeiros e sua relevância ao longo do tempo, foram analisados os movimentos conjuntos em crédito, mercado imobiliário, de ações e taxas de juros em 17 economias avançadas ao longo de 130 anos. Os ciclos globais ganharam relevância ao longo do tempo, especialmente para os preços das ações, que agora constituem o principal motor de flutuações na maioria dos países. Os ciclos globais em crédito e habitação tornaram-se muito mais evidentes e longos desde os anos 1980, mas sua

relevância aumentou apenas para um subgrupo de economias financeiramente abertas e desenvolvidas (BANK OF ENGLAND, 2023).

2.3 Evolução do mercado de crédito desde o Banco da Inglaterra até os dias atuais.

O advento do cartão de crédito e a consequente evolução dos sistemas de pontuação de crédito (credit scoring) têm sido impulsionados significativamente pela tecnologia de big data. A pontuação de crédito é fundamental para a tomada de decisão das instituições financeiras. Com a penetração da inteligência artificial no campo financeiro, a avaliação de crédito pessoal entrou em uma nova era, tornando-se um dos tópicos de pesquisa mais discutidos atualmente (HINDAWI, 2021).

Com a evolução tecnológica, os modelos de credit scoring passaram a incorporar algoritmos sofisticados para a análise de dados pessoais de crédito. O uso de algoritmos como o pdC-RF tem otimizado a correlação das características dos dados, reduzindo a dimensionalidade e permitindo uma avaliação mais precisa e eficiente.

Já o pdC-RF (*Probability of Default Credit Risk Forecasting* é uma metodologia avançada de previsão de risco de crédito que utiliza modelos de aprendizado de máquina, especialmente Random Forests (Florestas Aleatórias), para prever a probabilidade de inadimplência de um mutuário. Esta técnica é reconhecida por sua capacidade de processar grandes volumes de dados e identificar padrões complexos, proporcionando uma avaliação de risco mais precisa e personalizada. Conforme destacado por Breiman (2001) em sua obra sobre *Random Forests*, esta abordagem supera muitas limitações dos métodos tradicionais de *scoring*, como a regressão logística, pela sua habilidade em lidar com interações não-lineares e dados faltantes sem a necessidade de pré-processamento extensivo. Adicionalmente, Lessmann et al. (2015) demonstraram a superioridade do uso de métodos de aprendizado de máquina, incluindo Random Forests, em comparação com técnicas estatísticas convencionais no contexto de credit scoring, enfatizando a importância desses métodos avançados na indústria financeira moderna.

A trajetória histórica do mercado de crédito, desde as primeiras formas de empréstimos baseados em confiança pessoal até a sofisticação trazida pela tecnologia de big data e os sistemas de *credit scoring*, refletem uma evolução notável em como as sociedades gerenciam e mitigam riscos financeiros. A introdução e aprimoramento dos cartões de crédito, especialmente no contexto pós-2000, marcaram uma era de transformações significativas nas práticas de concessão de crédito. Este desenvolvimento, juntamente com a integração de tecnologias avançadas e algoritmos complexos na avaliação de crédito, não só revolucionou a

maneira como as instituições financeiras operam, mas também desencadeou mudanças substanciais na economia global. Portanto, este estudo visa analisar a eficácia desses modelos de *credit scoring* avançados, enfocando sua aplicação em carteiras de crédito a pessoas físicas, um reflexo direto das inovações e desafios que emergiram neste campo ao longo da história.

2.2 Métodos de Credit Scoring

2.2.1 Modelos de Credit Scoring: metodologias, abordagens e eficiência

A literatura sobre modelos de credit scoring revela uma evolução significativa nas metodologias utilizadas para avaliar a solvência de indivíduos. Historicamente, a análise de crédito era realizada com base em abordagens tradicionais e subjetivas, como os 5 C's (caráter, capital, colateral, capacidade e condições). No entanto, com o advento da tecnologia e a disponibilidade de grandes volumes de dados, houve uma transição para modelos estatísticos e de aprendizado de máquina, que proporcionam decisões mais rápidas e precisas (Dastile, Celik e Potsane, 2020).

Os modelos estatísticos, como a Análise Discriminante Linear (LDA), a Regressão Logística (LR) e o Classificador Naïve Bayes (NB), são amplamente utilizados devido à sua simplicidade e capacidade de interpretação. A Regressão Logística, por exemplo, é preferida em instituições financeiras por sua transparência na tomada de decisões, sendo fácil de explicar para reguladores e clientes. Contudo, esses modelos assumem linearidade nas relações entre variáveis, o que nem sempre corresponde à realidade (Thomas, Crook e Edelman, 2002).

Com o avanço do aprendizado de máquina, métodos como k-Nearest Neighbors (k-NN), Árvores de Decisão (DT), Máquinas de Vetores de Suporte (SVM) e Redes Neurais Artificiais (ANN) passaram a ser utilizados. Esses modelos, embora mais complexos e capazes de capturar relações não lineares, enfrentam desafios em termos de transparência. As SVMs, por exemplo, oferecem alto desempenho, mas são sensíveis aos parâmetros de regularização e kernel, tornando sua interpretação difícil (Henley e Hand, 1997; Baesens et al., 2003; Jiang, 2006).

Recentemente, os modelos de aprendizado profundo, como Redes Neurais Profundas (DNN), Redes Neurais Convolucionais (CNN) e Máquinas de Boltzmann Restritas (RBM), começaram a ser explorados no credit scoring. Estudos indicam que esses modelos apresentam desempenho superior devido à sua capacidade de detectar características complexas e discriminativas entre os devedores. No entanto, a aplicação desses modelos ainda é limitada

devido à falta de interpretabilidade, um requisito crucial para o setor financeiro regulamentado (Hinton, Osindero e Teh, 2006; LeCun et al., 1989; Ribeiro e Lopes, 2011).

A literatura também destaca o uso de modelos ensemble, que combinam múltiplos classificadores para melhorar a precisão e a robustez das previsões. Técnicas como Boosting, Bagging e Random Forests são comumente usadas. Os modelos ensemble geralmente superam os classificadores individuais, mostrando-se particularmente eficazes em conjuntos de dados desbalanceados, onde as classes de devedores e bons pagadores estão desproporcionadas (BREIMAN, 2001; FREUND; SCHAPIRE, 1997).

A eficiência e a aplicabilidade dos modelos de credit scoring variam conforme o contexto e os requisitos da instituição financeira. Modelos estatísticos são rápidos e transparentes, adequados para situações em que a interpretabilidade é essencial. Em contrapartida, os modelos de aprendizado de máquina e profundo são mais adequados para grandes volumes de dados e ambientes dinâmicos, onde a precisão é prioritária, apesar dos desafios de transparência (THOMAS; CROOK; EDELMAN, 2002; HENLEY; HAND, 1997).

Embora não haja consenso sobre o melhor modelo de credit scoring, para Dastile *et al*. (2020), a tendência é combinar a simplicidade dos modelos estatísticos com a precisão dos modelos de aprendizado de máquina e profundo. Futuras pesquisas devem focar em técnicas que equilibrem classes de dados, incorporem variáveis macroeconômicas e melhorem a abertura à interpretação dos modelos complexos, garantindo que as instituições financeiras possam utilizar essas tecnologias avançadas de maneira eficaz e conforme as regulamentações.

Em relação aos métodos de *Credit Scoring*, Lapshin (2022) busca os top 150 papers sobre credit scoring publicados em revistas relevantes, entre 2015 e 2021, através das seguintes palavras-chave: “machine learning”, “data mining”, “classification”, “credit scoring”, “credit”, “deep learning” e “statistics”. Depois disso, seleciona as fontes para analisar apenas os artigos de “regular journals” e filtra os artigos que não estejam relacionados a credit scoring e credit risk assessment.

Para cada um dos 110 artigos remanescentes, selecionam título, autores, nome do jornal, data de publicação e preenchem em um questionário similar ao utilizado por Louzada (2016). Os aspectos analisados são: metadata, arquitetura do modelo, testes feitos a partir do modelo e sua performance.

Dos 110 artigos analisados, os modelos utilizados distribuem-se conforme a Figura 1 presente abaixo.

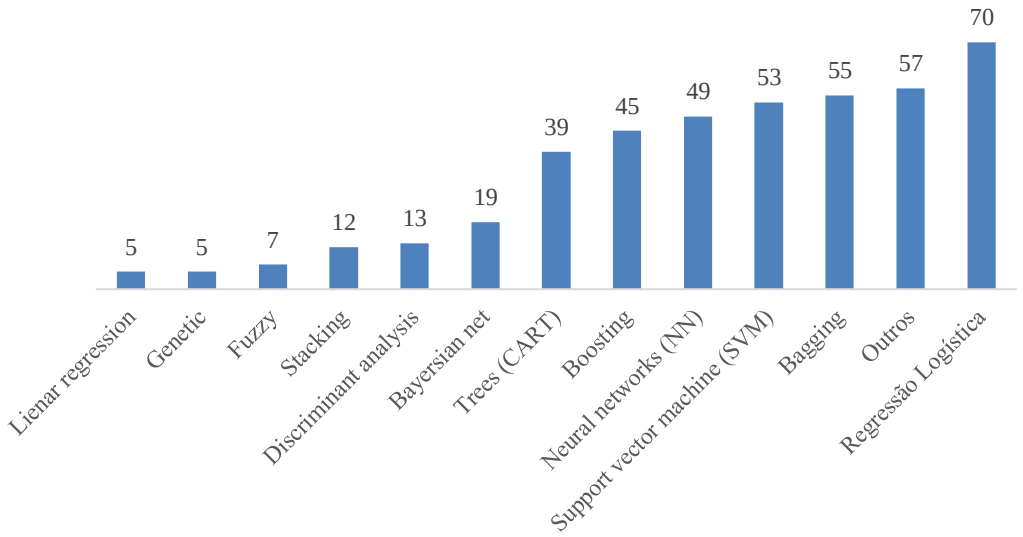

Figura 1: Distribuição dos modelos utilizados dos artigos analisados.

2.2.2 Credit Scoring com Redes Neurais

Uma rede neural (NN) é um sistema fundamentado em variáveis de entrada, também conhecidas como variáveis explanatórias, que são combinadas por meio de interações lineares e não lineares através de uma ou mais camadas ocultas, culminando em variáveis de saída, também denominadas variáveis de resposta.

As redes neurais foram concebidas na tentativa de simular o cérebro humano, pois se baseiam na transmissão de sinais eletrônicos entre muitos neurônios. A estrutura da NN é composta por elementos que recebem estímulos (as variáveis de entrada), estabelecem sinapses em diversos neurônios (ativação de neurônios em camadas ocultas) e resultam em respostas (variáveis de saída). As redes neurais variam conforme sua estrutura básica.

Em modelos de aprendizado de máquina, especialmente em redes neurais, uma consideração crítica é a configuração das camadas ocultas e a escolha das funções de ativação. As camadas ocultas, situadas entre a camada de entrada e a de saída, são fundamentais para o processamento não-linear e a capacidade de aprendizado do modelo. Cada neurônio em uma camada oculta aplica uma função de ativação aos seus inputs, permitindo que a rede neural capture relações complexas nos dados. As funções de ativação, como a sigmóide, ReLU ou tanh, introduzem não-linearidades e ajudam na convergência do modelo durante o treinamento.

Além disso, a adoção de arquiteturas avançadas, como a mistura de especialistas e redes neurais de função de base radial (RBF), pode enriquecer significativamente a modelagem em contextos específicos como o de pontuação de crédito. A mistura de especialistas utiliza um

conjunto de modelos (especialistas) e um mecanismo de *gating* para determinar qual especialista é mais adequado para dar a resposta em um dado ponto, o que pode aumentar a precisão e a adaptabilidade do modelo. Por outro lado, as redes neurais RBF, que usam funções de base radial como funções de ativação, são eficazes em classificar padrões não-lineares e complexos, sendo particularmente úteis quando os dados apresentam características peculiares que são melhor capturadas por tais funções. Como destacado por West (2000), a incorporação dessas técnicas avançadas em modelos de pontuação de crédito pode oferecer melhorias significativas na precisão e na robustez das previsões de risco de crédito.

Lee et al. (2002) propuseram uma abordagem híbrida de modelagem, integrando a análise discriminante com a técnica de redes neurais artificiais. Mais recentemente, diversas redes neurais artificiais têm sido sugeridas para abordar o problema de pontuação de crédito, como a rede neural probabilística (PANG, 2005), a rede neural artificial logística parcial (LISBOA et al., 2009), a rede neural artificial de metaplasticidade (MARCANO-CEDENO et al., 2011) e redes neurais híbridas (CHUANG; HUANG, 2011).

Em alguns conjuntos de dados, as redes neurais demonstraram a mais alta taxa média de classificação correta quando comparadas a outras técnicas tradicionais, como análise discriminante e regressão logística, levando em consideração que os resultados eram muito próximos (ABDOU et al., 2008).

Rehfeld Schoen (2017) explora o uso de redes neurais na construção de modelos de credit scoring. O estudo foca em aprimorar a precisão desses modelos, analisando o uso de variáveis binárias e seleção stepwise de variáveis. O autor utilizou uma base de dados de um banco de Taiwan para desenvolver quatro modelos com variações no uso dessas técnicas. Os resultados mostraram que modelos com uso de variáveis binárias, independentemente do uso de seleção stepwise, apresentaram melhores resultados em comparação com a base padrão. A seleção stepwise, embora tenha trazido um pequeno ganho, implicou em um aumento do tempo necessário para sua aplicação, indicando que seu uso deve ser avaliado cuidadosamente (SCHOEN, 2017).

Credit Scoring com Árvores de decisão

O método Classification and Regression Trees (BREIMAN *et al*., 1984) é uma abordagem de classificação que utiliza dados históricos para criar regras de decisão organizadas em estruturas semelhantes a árvores. Em linhas gerais, o propósito desse método é estabelecer um conjunto de condições lógicas do tipo "if-then" que possibilitem a previsão ou classificação de casos.

Três algoritmos se destacavam, em 2011, para construir árvores: o detector automático de interação qui-quadrado (CHAID), a árvore de classificação e regressão (CART) e o C5, que diferem pelo critério de construção da árvore. O CART utiliza o índice Gini como critério de divisão, o C5 adota a entropia, enquanto o CHAID utiliza o teste qui-quadrado (Yap et al., 2011).

John et al. (1996) exemplificam uma implementação baseada em regras na seleção de ações. Bijak e Thomas (2012) empregaram CHAID e CART para avaliar o valor de segmentação na capacidade de desempenho. Kao et al. (2012) propõem uma combinação de um modelo de pontuação de comportamento bayesiano e um modelo de pontuação de crédito baseado em CART.

Okada *et al.* (2019) compararam os algoritmos de indução de árvores de decisão C4.5 e CART. Foi realizado um experimento quantitativo utilizando um conjunto de dados do "Mapa Global de Câncer", e os algoritmos foram testados no software WEKA. O algoritmo C4.5, também conhecido como J48 no WEKA, apresentou um tempo de execução de 2,65 segundos, 18 folhas, 35 nós e um percentual de acerto de 96,5278%. O algoritmo CART, conhecido como SimpleCART no WEKA, teve um tempo de execução de 5,73 segundos, 17 folhas, 33 nós e um percentual de acerto de 95,1389%. Concluiu-se que o algoritmo C4.5 é mais eficiente, sendo aproximadamente 46,24% mais rápido do que o CART.

Credit Scoring com Regressão Logística (LG)

Introduzida por Berkson (1944), o modelo logístico considera um conjunto de variáveis explicativas $X = \{X1, ..., Xp\}$ e uma variável de resposta com duas categorias $Y = \{y1, y2\}$. Na regressão logística, a abordagem envolve a estimativa de uma combinação linear entre X e a transformação logit de Y. Assumindo y1 como a categoria de interesse para análise, o modelo pode ser descrito como $\log \pi / (1-\pi) = X\beta$, onde $\pi = P(Y = y1)$ e β é o vetor que contém os coeficientes do modelo.

Em alternativa, o modelo pode ser representado por $\pi i = \exp\{Xi\beta\} / (1 - \exp\{Xi\beta\})$, (1) onde πi é a probabilidade do i-ésimo indivíduo pertencer à categoria y1, condicionada a Xi.

A regressão logística é um método clássico, frequentemente comparado com outras técnicas (LI; HAND, 2002; HAND, 2005a; LEE; CHEN, 2005; ABDOU et al., 2008; YAP et al., 2011; PAVLIDIS et al., 2012), ou utilizado em combinação com outras técnicas (LOUZADA et al., 2011). Outras abordagens específicas de regressão logística incluem a regressão logística regularizada e a regressão logística limitada.

Carmona (2009) avaliou a aplicabilidade de modelos de Credit Scoring em uma instituição de microcrédito, o Cred Cidadania. Utilizando análise discriminante e regressão logística, os autores desenvolveram dois modelos de Credit Scoring: um para aprovação de crédito e outro para escoragem comportamental. Os resultados mostraram que esses modelos atingiram um desempenho satisfatório, alcançando aproximadamente 80% de classificação correta dos clientes. A conclusão do estudo enfatizou a eficácia dos modelos Credit Scoring em auxiliar na prevenção e redução da inadimplência, bem como na diminuição dos custos operacionais da instituição.

Marques (2015) focou na construção de um modelo de Behavior Scoring utilizando a técnica de Regressão Logística. Baseado em dados reais de produtos de microfinanças de uma instituição financeira brasileira, o estudo envolveu análise descritiva de variáveis e validação do modelo através de critérios estatísticos. O objetivo era prever o comportamento dos clientes em relação aos compromissos financeiros assumidos, visando identificar se seriam bons ou maus pagadores após a obtenção do crédito. Este trabalho demonstrou a aplicabilidade e a eficácia da Regressão Logística na construção de modelos preditivos de comportamento de clientes em contextos financeiros.

2.3 Considerações finais

Ao longo deste capítulo, foi discutida a evolução e a importância dos modelos de *credit scoring* no sistema financeiro moderno. Esses modelos, utilizando técnicas estatísticas avançadas e algoritmos de aprendizado de máquina, têm desempenhado um papel crucial na avaliação de crédito e na gestão de riscos financeiros. A trajetória histórica desde as primeiras formas de empréstimos baseados em confiança pessoal até a sofisticação trazida pela tecnologia de big data e os sistemas de *credit scoring* reflete uma evolução notável em como as sociedades gerenciam e mitigam riscos financeiros.

Os desafios enfrentados pelos modelos de *credit scoring* são tanto tecnológicos quanto regulatórios. A necessidade de balancear a precisão dos modelos com questões éticas e de privacidade dos dados é uma preocupação crescente. Além disso, a integração de tecnologias

emergentes, como a inteligência artificial e a big data, oferece oportunidades para melhorar a eficácia desses modelos, mas também levanta questões sobre a interpretabilidade e a transparência das decisões automatizadas.

Este capítulo destaca a relevância contínua da pesquisa e desenvolvimento em *credit scoring*, enfatizando a necessidade de inovação contínua e avaliação crítica para enfrentar os desafios emergentes no campo do crédito e da avaliação de riscos financeiros.

3. Metodologia

3.1 Base de dados utilizada

No desenvolvimento deste estudo sobre credit scoring, optou-se exclusivamente pela utilização da base de dados *LendingClub Loan Data*, podendo ser encontrada em https://www.kaggle.com/datasets/wordsforthewise/lending-club ,que oferece uma rica coleção de variáveis específicas do mercado de crédito peer-to-peer dos Estados Unidos. Esta base inclui informações detalhadas sobre cada empréstimo, abrangendo desde características financeiras do mutuário até o desempenho do crédito ao longo do tempo. A seleção desta base deve-se à sua relevância e detalhamento, proporcionando um panorama abrangente necessário para avaliar a eficácia dos modelos de credit scoring.

O conjunto completo de variáveis e seus significados está presente no **anexo A**. Este tem um total de 917472 casos e conta com 107 colunas (variáveis). Como exemplo de variáveis, temos a proporção do saldo atual em relação ao limite de crédito total em contas de parcelamento ('il_util'), meses desde a abertura da conta de parcelamento mais recente ('mths_since_rcnt_il'), e proporção de saldo em relação ao limite de crédito em todas as transações ('all_util'). Informações sobre a atividade de crédito recente, tais como o número de novas contas abertas nos últimos seis meses ('open_acc_6m') e o número de consultas de crédito nos últimos doze meses ('inq_last_12m'), também são consideradas.

Além disso, há também dados sobre a estabilidade financeira e histórico de emprego do mutuário, como o cargo profissional ('emp_title') e o tempo de emprego ('emp_length'), são inclusos para enriquecer a análise. O conjunto também detalha o comportamento de pagamento, por exemplo, o número de contas com pagamentos atrasados em mais de 120 dias ('num_tl_120dpd_2m'), bem como informações sobre a utilização de linhas de crédito rotativo ('revol_util') e o comprometimento de renda mensal do mutuário frente às suas obrigações totais de dívida ('dti').

É importante notar que, enquanto a *LendingClub Loan Data* oferece um vislumbre detalhado das finanças dos mutuários e dos padrões de empréstimo, este estudo está limitado aos dados fornecidos pela plataforma. Assim, generalizações para outros mercados de crédito ou populações devem ser feitas com cautela. A análise deste estudo é conduzida em ambiente controlado, utilizando a linguagem de programação Python, garantindo a precisão dos testes estatísticos e a robustez dos modelos preditivos construídos.

3.2 Processamento de dados

A metodologia empregada neste estudo se concentra na aplicação e comparação detalhada de três modelos estatísticos distintos, usando a base de dados LendingClub Loan Data. Cada modelo é escolhido por sua capacidade em abordar aspectos diferentes da predição de inadimplência em empréstimos, e são integrados por meio de técnicas pré-processamento de dados e validação.

Para a análise, os modelos de *credit scoring* foram aplicados utilizando uma combinação de técnicas estatísticas e de aprendizado de máquina para prever a probabilidade de inadimplência. Foi criada a variável binária de inadimplência (‘is_default’) com base na variável loan_status.

Os status "Default" (inadimplente) e "Charged Off" (cancelado por falta de pagamento) são a categoria 1 e os outros status receberam valor 0 ('Fully Paid', 'Current', 'In Grace Period' e ‘Late’). As linhas com valores vazios foram removidas e as variáveis categóricas foram transformadas em dummy.

3.3 Modelos de Credit Scoring

Regressão Logística: Este modelo clássico de estatística é amplamente usado para estimar a probabilidade de ocorrência de um evento, baseando-se na relação entre a variável dependente binária e uma ou mais variáveis independentes. No contexto deste estudo, a regressão logística é aplicada para prever a probabilidade de inadimplência com base em características financeiras e pessoais dos mutuários. O modelo inclui variáveis como renda anual, razão dívida/renda, histórico de pagamentos atrasados, utilização de crédito rotativo e muitas outras. O ajuste do modelo é feito através do método de máxima verossimilhança, e sua performance é avaliada por métricas como AUC (Área Sob a Curva ROC) e o índice Gini.

Random Forests: Este modelo de aprendizado ensemble utiliza múltiplas árvores de decisão para fazer previsões, oferecendo robustez especialmente em casos com muitas variáveis explicativas e complexidade nos dados. Random Forests é conhecido por sua capacidade de manejar overfitting melhor do que árvores de decisão únicas. No estudo, este modelo é utilizado para identificar os principais preditores de risco de inadimplência, com cada árvore do "floresta" votando para uma previsão final. A importância das variáveis é avaliada, e a performance do modelo é quantificada usando as mesmas métricas de avaliação aplicadas à regressão logística.

Redes Neurais: Como uma forma de aprendizado profundo, as redes neurais são empregadas para capturar relações não-lineares complexas entre as variáveis. No estudo, uma rede com várias camadas ocultas é treinada, incluindo técnicas como dropout para evitar overfitting. Este modelo é particularmente útil para modelar interações complexas entre características dos mutuários que podem não ser prontamente aparentes ou capturadas por métodos mais tradicionais.

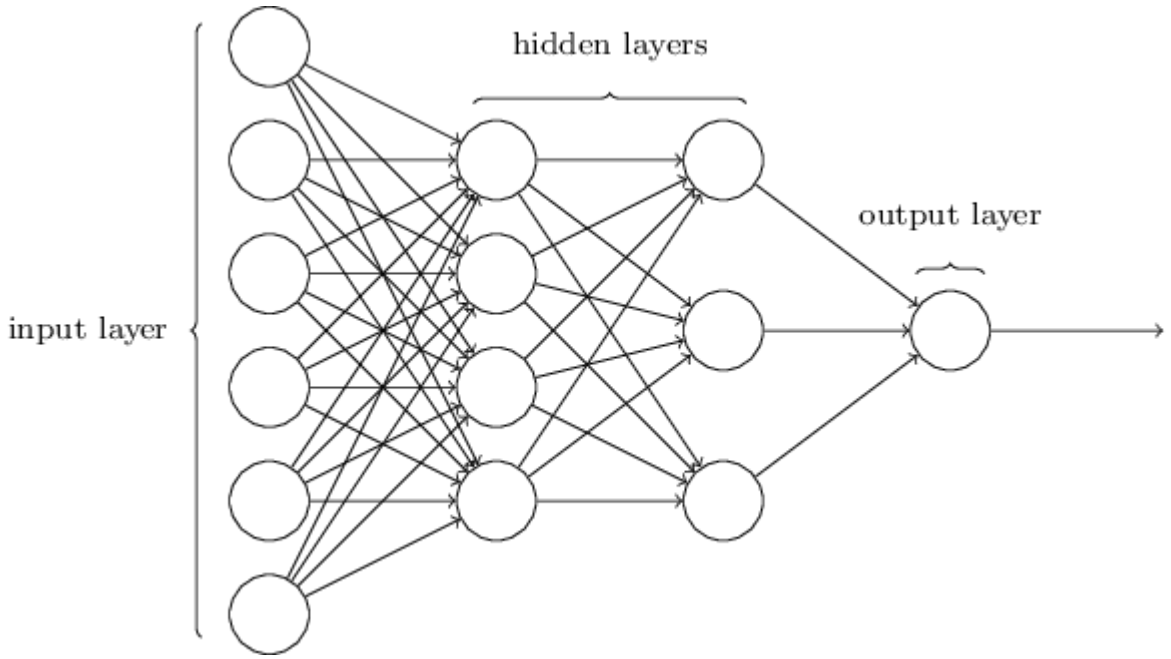

Figura 2: Modelo gráfico representativo de uma rede neural artificial

Uma Rede Neural Artificial é composta por várias camadas, incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada camada contém neurônios, que são unidades computacionais que recebem entradas, aplicam uma função de ativação e produzem uma saída. Durante o treinamento, as conexões entre os neurônios são ajustadas iterativamente por meio de algoritmos de otimização, como o algoritmo de retropropagação, para minimizar uma função de perda e melhorar o desempenho do modelo.

A arquitetura de uma ANN é determinada pelo número de camadas, o número de neurônios em cada camada e as funções de ativação usadas. As camadas ocultas permitem que a rede aprenda representações hierárquicas dos dados, enquanto a camada de saída produz as previsões ou classificações desejadas. As funções de ativação introduzem não-linearidades na rede, permitindo que ela aprenda relações não-lineares nos dados.

3.4 Treinamento e validação dos modelos

Cada modelo usado foi treinado utilizando uma divisão de 60% para treino e 40% para teste, assegurando que haja uma avaliação rigorosa e imparcial. O pré-processamento dos dados inclui a exclusão de casos com valores faltantes e codificação one-hot para variáveis categóricas, garantindo que os modelos operem em um espaço de características consistentemente formatado.

Para proporcionar uma comparação justa entre os modelos, são construídas as curvas ROC para cada um, permitindo visualizar a eficácia de cada modelo em distinguir entre as classes de resposta. A análise dos resultados é profundamente integrada com uma discussão sobre as implicações práticas das descobertas, especialmente em termos de identificar mutuários com maior risco de inadimplência e ajustar estratégias de concessão de crédito correspondentes.

3.5 Métricas de avaliação

A precisão dos modelos foi medida por métricas como a Área Sob a Curva ROC (AUC) e o Índice Gini, fornecendo uma comparação quantitativa da capacidade de cada modelo em diferenciar entre mutuários adimplentes e inadimplentes. A interpretação dos resultados contou com a aplicação de análises estatísticas para validar a significância dos achados.

Cada modelo exporta uma probabilidade de um indivíduo da base de teste ser inadimplente. Para cada indivíduo a seguinte classificação prevista é realizada: se a probabilidade prevista é maior que a nota de corte (cutoff) então o indivíduo é classificado como inadimplente (1), e caso contrário, adimplente (0). Então foram avaliadas métricas da matriz de confusão para diferentes notas de corte e assim, poder analisar a sensibilidade de cada modelo à classificação. Foram escolhidas as notas de corte 0,5; 0,75 e 0,25, por serem bem distribuídas em relação aos extremos 0 e 1. E posteriormente foram calculadas as quantidades de falsos e verdadeiros positivos (1) e negativos (0).

4. Resultados

4.1 Análise descritiva

A utilização de modelos de predição de inadimplência baseada em dados do Lending Club revela provas importantes sobre a capacidade de diferentes abordagens estatísticas em prever riscos financeiros. Utilizando uma base de dados robusta do site Kaggle, foram implementados modelos de Regressão Logística, Random Forest e uma Rede Neural, cada um ajustado para avaliar a probabilidade de inadimplência entre os tomadores de empréstimo.

Após a seleção das variáveis mais relevantes usando como critério a plausibilidade (BAESENS *et al*., 2021), foi realizada a análise descritiva disposta na tabela abaixo. Todas as variáveis com valor máximo igual a 1 são binárias. Quanto à variável binária 'is_default', sua média é de 0,084985, ou seja, 8,4985% da base de dados é de inadimplentes. Havendo 839501 pessoas adimplentes e 77971 inadimplentes.

A análise descritiva das variáveis selecionadas da base de dados do Lending Club informa a respeito do perfil dos mutuários e suas características financeiras. Primeiramente, as variáveis relacionadas à duração do emprego (emp_length) indicam que a maioria dos mutuários possui uma trajetória profissional significativa. Por exemplo, 35,77% dos mutuários têm mais de 10 anos de emprego, enquanto as categorias intermediárias, como 5 e 4 anos de emprego, representam aproximadamente 6,70% cada. As categorias com menor tempo de emprego, como menos de 1 ano, têm uma representação média de 8,50%. A variabilidade nessas categorias, conforme indicado pelos altos desvios padrão, reflete uma distribuição diversa na estabilidade de emprego entre os mutuários.

A variável num_tl_90g_dpd_24m, que mede o número de contas com 90 ou mais dias de atraso nos últimos 24 meses, apresenta uma média baixa de 0,078 e um desvio padrão de 0,500. Isso sugere que a maioria dos mutuários não tem registros significativos de atrasos graves, embora existam casos extremos com até 35 ocorrências. Esse padrão é consistente com a variável pub_rec, que contabiliza o número de registros públicos desfavoráveis, com uma média de 0,208 e um desvio padrão de 0,577. Novamente, a maioria dos mutuários não possui registros desfavoráveis, mas há exceções notáveis, com até 61 registros.

Outra variável de interesse é term_60 months, que indica se o prazo do empréstimo é de 60 meses. Aproximadamente 29,05% dos empréstimos têm esse prazo, com um desvio padrão de 45,40%, refletindo a natureza binária dessa variável. Isso sugere que uma proporção significativa dos mutuários opta por prazos mais longos para pagamento, o que pode influenciar a análise de risco e a política de concessão de crédito.

A análise da incidência de inadimplência (delinq_2yrs) revela uma média de 0,307, com um desvio padrão de 0,879. Isso significa que a maioria dos mutuários não teve incidências de inadimplência nos últimos dois anos, mas há uma minoria com múltiplas ocorrências, chegando até 35 em casos extremos. Esse comportamento é crítico para a avaliação de risco, pois incidências passadas são indicativos importantes do comportamento futuro dos mutuários.

Variáveis	n	Média	Desvio Padrão	Min	P25%	P50%	P75%	Max
emp_length_9 years	917472	0,035681	0,185493	0	0	0	0	1
emp_length_8 years	917472	0,038212	0,191707	0	0	0	0	1
emp_length_7 years	917472	0,038787	0,193087	0	0	0	0	1
emp_length_6 years	917472	0,048229	0,21425	0	0	0	0	1
emp_length_5 years	917472	0,066957	0,249947	0	0	0	0	1
emp_length_4 years	917472	0,067951	0,251662	0	0	0	0	1
num_tl_90g_dpd_24m	917472	0,077946	0,499539	0	0	0	0	35
emp_length_< 1 year	917472	0,084985	0,278859	0	0	0	0	1
emp_length_3 years	917472	0,089008	0,284755	0	0	0	0	1
emp_length_2 years	917472	0,100503	0,30067	0	0	0	0	1
pub_rec	917472	0,207873	0,576732	0	0	0	0	61
term_ 60 months	917472	0,290516	0,454001	0	0	0	1	1
delinq_2yrs	917472	0,307045	0,878634	0	0	0	0	35
emp_length_10+ years	917472	0,357664	0,479313	0	0	0	1	1
num_accts_ever_120_pd	917472	0,565819	1	0	0	0	0	51
total_rec_late_fee	917472	2	13	0	0	0	0	1484
num_tl_op_past_12m	917472	2	2	0	1	2	3	31
open_acc	917472	13	6	2	8	11	15	101
int_rate	917472	13	5	5	9	12	15	30
dti	917472	20	13	-1	13	18	25	999
total_acc	917472	25	12	3	16	23	32	176
revol_util	917472	46	24	0	27	45	64	183
pct_tl_nvr_dlq	917472	94	9	5	91	98	100	100
installment	917472	455	280	7	250	380	610	1719
fico_range_low	917472	701	34	660	675	695	720	845
fico_range_high	917472	705	34	664	679	699	724	850
loan_amnt	917472	15.355	9.612	1000	8000	13000	20000	40000
revol_bal	917472	16.149	21.295	0	5784	11048	19719	2358150
annual_inc	917472	83.881	85.180	20	50000	70000	100000	10999200

Tabela 1: Estatística descritiva das variáveis selecionadas

Os valores médios das taxas de juros (int_rate) e da relação dívida/renda (dti) são, respectivamente, 13% e 20%, com desvios padrão de 5% e 13%. Esses valores sugerem que, em média, os mutuários pagam taxas de juros moderadas, enquanto a relação dívida/renda indica um nível de endividamento relativamente alto. Esse aspecto é corroborado pela variável revol_util, que mede a taxa de utilização de crédito rotativo e apresenta uma média de 46%, com um desvio padrão de 24%. Essa alta taxa de utilização pode indicar uma dependência significativa de crédito rotativo pelos mutuários.

As variáveis fico_range_low e fico_range_high mostram que os scores de crédito dos mutuários estão, em média, entre 701 e 705, com desvios padrão de 34 pontos. Isso posiciona a maioria dos mutuários na faixa de crédito boa a excelente, sugerindo um perfil de risco relativamente baixo. O montante médio dos empréstimos solicitados (loan_amnt) é de $15.355, com um desvio padrão de $9.612, indicando uma ampla faixa de valores solicitados, desde pequenas quantias até empréstimos significativos.

A renda anual autorrelatada (annual_inc) tem uma média de $83.881, com um desvio padrão considerável de $85.180, refletindo a diversidade de rendas entre os mutuários. Esse dado é crucial para a análise de capacidade de pagamento e a determinação de condições de empréstimo adequadas. As estatísticas descritivas fornecem uma visão abrangente do perfil dos mutuários, destacando variações significativas em termos de estabilidade de emprego, histórico de crédito, e capacidade financeira, todos fatores essenciais para a gestão eficaz do risco de crédito.

4.2 Desempenho dos modelos

Os resultados apresentados na ROC Curve e os valores de AUC e Gini para cada modelo indicam uma performance relativamente similar entre os modelos de Regressão Logística e Random Forest, ambos com AUC de 0.70 e Gini de 0.40. Isto sugere que ambos os modelos são moderadamente eficazes na distinção entre os tomadores que possuem maior probabilidade de inadimplir e aqueles que não.

No entanto, a Rede Neural apresenta uma leve superioridade com AUC de 0.72 e Gini de 0.43, indicando uma eficiência marginalmente melhor na predição de inadimplência. Este aumento na performance pode ser atribuído à capacidade das redes neurais de capturar relações

não-lineares complexas nos dados que podem ser menos perceptíveis para modelos mais tradicionais.

Modelo	AUC	Índice Gini
Regressão Logística	0.70	0.40
Random Forest	0.70	0.40
Rede Neural	0.72	0.43

Tabela 2: Avaliação da qualidade dos modelos utilizados

A curva ROC (Figura 3), que traça a taxa de verdadeiros positivos contra a taxa de falsos positivos para cada modelo, mostra que a Rede Neural, representada pela linha verde, se mantém consistentemente acima das outras duas abordagens ao longo do espectro de *threshold*. Isso implica que a Rede Neural consegue manter uma taxa de detecção de inadimplentes mais alta para o mesmo nível de falsos positivos, o que é crucial em contextos financeiros onde falsos negativos (inadimplentes não detectados) podem ter consequências econômicas significativas.

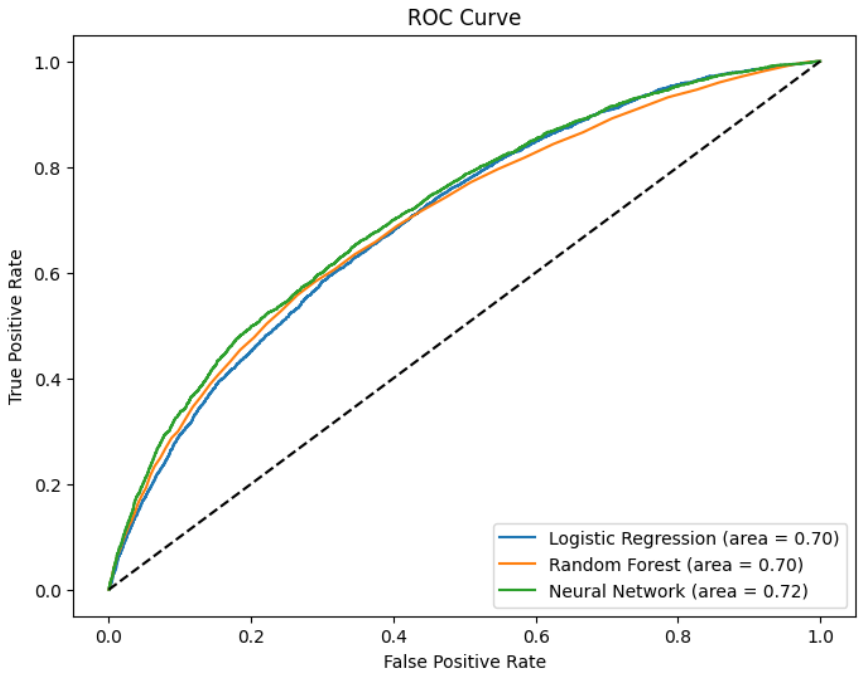

Figura 3: Curva ROC dos 3 modelos estimados

Esses achados são instrutivos para instituições financeiras que dependem de modelagem preditiva para minimizar riscos. A pequena vantagem da Rede Neural sugere que

tecnologias de aprendizado de máquina mais avançadas podem ser mais adequadas para este tipo de análise, especialmente quando é possível treinar os modelos em grandes conjuntos de dados com uma vasta gama de características preditoras. No entanto, a escolha do modelo também deve considerar outros fatores como a abertura à interpretação do modelo, a facilidade de implementação e a necessidade de recursos computacionais.

Desse modo, enquanto a Regressão Logística e o Random Forest oferecem soluções viáveis e robustas para a avaliação do risco de crédito, a inclusão de técnicas de aprendizado profundo, como redes neurais, pode proporcionar melhorias incrementais na performance preditiva. Este estudo destaca a importância de continuar explorando e integrando novas abordagens analíticas no campo da análise de crédito para aprimorar as estratégias de decisão e gestão de risco no setor financeiro.

4.3 Análise dos pontos de corte

A tabela 3 apresenta a avaliação dos modelos com diferentes pontos de corte (cutoffs) de 0.25, 0.5 e 0.75. A análise é baseada nas métricas de Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN). Primeiramente, ao observar a Regressão Logística, percebe-se que um ponto de corte mais baixo (0.25) resulta em um maior número de Verdadeiros Positivos (2830), mas também em um número significativo de Falsos Positivos (6732) e Falsos Negativos (28359). À medida que o ponto de corte aumenta para 0.5 e 0.75, o número de Verdadeiros Positivos diminui drasticamente para 168 e 27, respectivamente, indicando uma maior seletividade do modelo. Entretanto, isso também resulta em um aumento dos Falsos Negativos, que alcançam 31021 e 31162, respectivamente. Isso sugere que, embora um ponto de corte maior reduza os Falsos Positivos, ele também aumenta a chance de inadimplentes não serem identificados.

Modelo	Cutoff	Verdadeiros Positivos	Verdadeiros Negativos	Falsos Positivos	Falsos Negativos
Regressão Logística	0.25	2830	329068	6732	28359
Regressão Logística	0.5	168	335373	427	31021
Regressão Logística	0.75	27	335663	137	31162
Random Forest	0.25	6996	320428	15372	24193
Random Forest	0.5	452	335285	515	30737
Random Forest	0.75	0	335800	0	31189
Rede Neural	0.25	5126	326427	9373	26063
Rede Neural	0.5	0	335800	0	31189
Rede Neural	0.75	0	335800	0	31189

Tabela 3: Avaliação dos erros e acertos dos modelos sob o uso de diferentes notas de corte.

No caso do modelo Random Forest, um comportamento semelhante é observado, embora com variações significativas. Com um ponto de corte de 0.25, o número de Verdadeiros Positivos é o mais alto entre todos os modelos analisados (6996), porém com um elevado número de Falsos Positivos (15372) e Falsos Negativos (24193). Quando o ponto de corte é ajustado para 0.5, os Verdadeiros Positivos caem para 452, e os Falsos Positivos reduzem-se significativamente para 515, mas os Falsos Negativos aumentam para 30737. Curiosamente, um ponto de corte de 0.75 resulta em nenhum Verdadeiro Positivo, o que sugere uma falha em identificar qualquer inadimplente, embora elimine os Falsos Positivos e mantenha um alto número de Falsos Negativos (31189).

Para o modelo de Rede Neural, os resultados mostram que um ponto de corte de 0.25 proporciona um número relativamente alto de Verdadeiros Positivos (5126) e um número considerável de Falsos Positivos (9373) e Falsos Negativos (26063). No entanto, ao ajustar o ponto de corte para 0.5 e 0.75, o modelo não identifica nenhum Verdadeiro Positivo e não gera Falsos Positivos, mas mantém um número elevado de Falsos Negativos (31189), indicando uma total ausência de previsão de inadimplentes.

4.3 Considerações finais

Comparando os três modelos, verifica-se que a Regressão Logística e o Random Forest mostram um equilíbrio mais ajustável entre Verdadeiros Positivos e Falsos Positivos ao variar os pontos de corte, enquanto a Rede Neural apresenta uma rigidez maior, especialmente nos pontos de corte mais altos. É evidente que a escolha do ponto de corte é crítica e deve ser ajustada conforme o contexto e a tolerância ao risco de Falsos Positivos e Falsos Negativos.

Portanto, para aplicações práticas na economia, especialmente na avaliação de inadimplência, é crucial considerar o trade-off entre diferentes métricas de desempenho. Os tomadores de decisão devem ponderar entre a necessidade de identificar o maior número possível de inadimplentes (maximizando Verdadeiros Positivos) e a necessidade de evitar falsas acusações de inadimplência (minimizando Falsos Positivos). Cada modelo oferece vantagens e desvantagens distintas, que devem ser cuidadosamente analisadas para a implementação mais eficaz em cenários reais.

5. Conclusão

O estudo realizado buscou avaliar a eficácia dos modelos de *credit scoring* no mercado de crédito contemporâneo, enfocando particularmente as inovações e adaptações desses modelos no período pós-2000. Com a incorporação de avanços tecnológicos, como o uso de Big Data e inteligência artificial, os modelos de *credit scoring* têm se transformado, oferecendo ferramentas mais precisas e eficazes para a gestão de riscos de crédito.

Os resultados alcançados ao longo da pesquisa reforçam a importância e a robustez dos modelos tradicionais, como a regressão logística e as árvores de decisão, mas também destacam o potencial incrementado pelos métodos mais recentes, como as redes neurais, que mostram uma superioridade na captura de relações complexas e não-lineares entre os dados. Esta capacidade é crucial em um ambiente econômico que se mostra cada vez mais volátil e data-driven.

É fundamental reconhecer que, apesar dos avanços tecnológicos que potencializam os modelos de *credit scoring*, existem desafios persistentes, especialmente relacionados à ética na utilização de dados e à transparência das decisões automatizadas. As implicações regulatórias e a necessidade de manter um equilíbrio entre inovação e privacidade dos dados são questões que necessitarão de atenção contínua dos desenvolvedores de modelos, reguladores e instituições financeiras.

Para o futuro, antecipa-se que os modelos de *credit scoring* continuarão a evoluir, integrando novas tecnologias e adaptando-se às mudanças nas dinâmicas do mercado. O desenvolvimento contínuo dessas ferramentas é essencial para assegurar um sistema de crédito eficiente, justo e acessível, reforçando o papel do *credit scoring* como pilar fundamental na estrutura financeira global.

Este trabalho também destaca a necessidade de pesquisas futuras que explorem os impactos a longo prazo dos modelos de *credit scoring* na estabilidade econômica e na inclusão financeira. Investigações adicionais poderiam focar em como esses modelos afetam o acesso ao crédito em diferentes segmentos da população e como podem ser aprimorados para fomentar uma economia mais inclusiva.

Este estudo não apenas contribui para o corpo acadêmico sobre modelos de *credit scoring*, mas também serve como referência para profissionais da área financeira que buscam

entender e aplicar as técnicas mais eficazes e éticas na avaliação de crédito. A colaboração entre academia, indústria e reguladores será crucial para moldar o futuro dos modelos de *credit scoring*, assegurando que eles se mantenham relevantes e eficientes em um mundo financeiro em constante mudança.

ANEXO A

Variável	Descrição
il_util	Proporção do saldo atual total em relação ao crédito/limite de crédito máximo em todas as contas parceladas
mths_since_rcnt_il	Meses desde a abertura da conta de parcelamento mais recente
all_util	Saldo em relação ao limite de crédito em todas as transações
open_acc_6m	Número de transações abertas nos últimos 6 meses
inq_last_12m	Número de consultas de crédito nos últimos 12 meses
total_cu_tl	Número de transações financeiras
open_rv_12m	Número de transações rotativas abertas nos últimos 12 meses
max_bal_bc	Saldo atual máximo devido em todas as contas rotativas
total_bal_il	Saldo atual total de todas as contas parceladas
inq_fi	Número de consultas de financiamento pessoal
open_rv_24m	Número de transações rotativas abertas nos últimos 24 meses
open_il_24m	Número de contas parceladas abertas nos últimos 24 meses
open_il_12m	Número de contas parceladas abertas nos últimos 12 meses
mths_since_recent_inq	Meses desde a consulta mais recente
emp_title	O cargo fornecido pelo Mutuário ao solicitar o empréstimo
num_tl_120dpd_2m	Número de contas atualmente com 120 dias de atraso (atualizado nos últimos 2 meses)
emp_length	Duração do emprego em anos. Os valores possíveis variam de 0 a 10, onde 0 significa menos de um ano e 10 significa dez anos ou mais
mo_sin_old_il_acct	Meses desde a abertura da conta de parcelamento bancário mais antiga
bc_util	Proporção do saldo atual total em relação ao crédito/limite de crédito máximo para todas as contas de cartão de banco
percent_bc_gt_75	Percentual de todas as contas de cartão de banco > 75% do limite
bc_open_to_buy	Total disponível para compra em cartões de banco rotativos
mths_since_recent_bc	Meses desde a abertura da conta de cartão de banco mais recente
pct_tl_nvr_dlq	Percentual de transações nunca em atraso
avg_cur_bal	Saldo médio atual de todas as contas
mo_sin_old_rev_tl_op	Meses desde a abertura da conta rotativa mais antiga
mo_sin_rcnt_rev_tl_op	Meses desde a abertura da conta rotativa mais recente
num_rev_accts	Número de contas rotativas
mo_sin_rcnt_tl	Meses desde a abertura da conta mais recente
num_tl_op_past_12m	Número de contas abertas nos últimos 12 meses
num_op_rev_tl	Número de contas rotativas abertas
num_il_tl	Número de contas parceladas

num_bc_tl	Número de contas de cartão de banco
num_actv_rev_tl	Número de transações rotativas atualmente ativas
num_actv_bc_tl	Número de contas de cartão de banco atualmente ativas
num_accts_ever_120_pd	Número de contas com 120 ou mais dias de atraso
num_tl_30dpd	Número de contas atualmente com 30 dias de atraso (atualizado nos últimos 2 meses)
num_tl_90g_dpd_24m	Número de contas com 90 ou mais dias de atraso nos últimos 24 meses
total_il_high_credit_limit	Limite de crédito/alta parcela total
num_rev_tl_bal_gt_0	Número de transações rotativas com saldo >0
tot_cur_bal	Saldo atual total de todas as contas
tot_coll_amt	Total de valores de cobrança já devidos
tot_hi_cred_lim	Total de crédito/limite de crédito máximo
num_bc_sats	Número de contas de cartão de banco satisfatórias
num_sats	Número de contas satisfatórias
mort_acc	Número de contas de hipoteca
total_bc_limit	Limite de crédito/alta de cartão de banco total
acc_open_past_24mths	Número de transações abertas nos últimos 24 meses
total_bal_ex_mort	Saldo total de crédito excluindo hipoteca
title	O título do empréstimo fornecido pelo mutuário
last_pymnt_d	Último mês em que o pagamento foi recebido
revol_util	Taxa de utilização de linha rotativa, ou a quantidade de crédito que o mutuário está usando em relação a todo o crédito rotativo disponível
dti	Uma proporção calculada usando os pagamentos mensais totais da dívida do mutuário em todas as obrigações de dívida, excluindo hipoteca e o empréstimo LC solicitado, dividido pela renda mensal autorrelatada do mutuário
pub_rec_bankruptcies	Número de falências registradas publicamente
collections_12_mths_ex_med	Número de cobranças em 12 meses, excluindo cobranças médicas
chargeoff_within_12_mths	Número de cancelamentos de dívida em 12 meses
tax_liens	Número de penhoras fiscais
last_credit_pull_d	O mês mais recente em que o LC verificou o crédito para este empréstimo
inq_last_6mths	O número de consultas nos últimos 6 meses (excluindo consultas de automóveis e hipotecas)
pub_rec	Número de registros públicos desfavoráveis
delinq_amnt	O valor em atraso devido nas contas em que o mutuário está inadimplente
delinq_2yrs	O número de incidências de inadimplência de mais de 30 dias no arquivo de crédito do mutuário nos últimos 2 anos

earliest_cr_line	O mês em que a linha de crédito mais antiga do mutuário foi aberta
acc_now_delinq	O número de contas nas quais o mutuário está atualmente inadimplente
total_acc	O número total de linhas de crédito atualmente no arquivo de crédito do mutuário
open_acc	O número de linhas de crédito abertas no arquivo de crédito do mutuário
annual_inc	A renda anual autorrelatada fornecida pelo mutuário durante o registro
zip_code	Os primeiros 3 números do código postal fornecido pelo mutuário na solicitação de empréstimo
last_fico_range_high	A faixa limite superior à qual o último FICO do mutuário pertence
loan_status	Status atual do empréstimo
issue_d	O mês em que o empréstimo foi financiado
verification_status	Indica se a renda foi verificada pelo LC, não verificada ou se a fonte de renda foi verificada
home_ownership	O status de propriedade da casa fornecido pelo mutuário durante o registro. Nossos valores são: ALUGUEL, PROPRIEDADE, HIPOTECA, OUTRO
last_fico_range_low	A faixa limite inferior à qual o último FICO do mutuário pertence
last_pymnt_amnt	Último valor total do pagamento recebido
url	URL para a página do LC com dados de listagem
sub_grade	Subclassificação do empréstimo atribuída pelo LC
grade	Classificação do empréstimo atribuída pelo LC
installment	O pagamento mensal devido pelo mutuário se o empréstimo for originado
int_rate	Taxa de Juros do empréstimo
term	O número de pagamentos do empréstimo. Os valores estão em meses e podem ser 36 ou 60
funded_amnt_inv	O valor total comprometido pelos investidores para esse empréstimo naquele momento
funded_amnt	O valor total comprometido para esse empréstimo naquele momento
pymnt_plan	Indica se um plano de pagamento foi estabelecido para o empréstimo
addr_state	O estado fornecido pelo mutuário na solicitação de empréstimo
purpose	Uma categoria fornecida pelo mutuário para a solicitação do empréstimo
recoveries	Recuperação bruta após baixa

out_prncp_inv	Principal restante para a porção do valor total financiado pelos investidores
total_pymnt	Pagamentos recebidos até a data para o valor total financiado
loan_amnt	O valor listado do empréstimo solicitado pelo mutuário. Se em algum momento o departamento de crédito reduzir o valor do empréstimo, isso será refletido neste valor
total_pymnt_inv	Pagamentos recebidos até a data para a porção do valor total financiado pelos investidores
total_rec_prncp	Principal recebido até a data
total_rec_int	Juros recebidos até a data
total_rec_late_fee	Taxas de atraso recebidas até a data
initial_list_status	O status de listagem inicial do empréstimo. Os valores possíveis são – W, F
policy_code	"policy_code disponível publicamente = 1 novos produtos não disponíveis publicamente policy_code = 2"
collection_recovery_fee	Taxa de cobrança após baixa
revol_bal	Saldo total de crédito rotativo
fico_range_high	A faixa limite superior à qual o FICO do mutuário na origem do empréstimo pertence
fico_range_low	A faixa limite inferior à qual o FICO do mutuário na origem do empréstimo pertence
application_type	Indica se o empréstimo é uma solicitação individual ou uma solicitação conjunta com dois co-mutuários
out_prncp	Principal restante para o valor total financiado
id	Um ID único atribuído pelo LC para a listagem do empréstimo

REFERÊNCIAS

ARAÚJO, E. A.; CARMONA, C. U. D. M. Construção de modelos credit scoring com análise discriminante e regressão logística para a gestão do risco de inadimplência de uma instituição de microcrédito. 2009.

BAESENS, B.; SETIONO, R.; MUES, C.; VANTHIENEN, J. Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science, v. 49, n. 3, p. 312-329, 2003.

BAESENS, B. et al. Credit Risk Modelling and Benchmarking: Theory and Applications. Cambridge: Cambridge University Press, 2021.

BANCO CENTRAL DO BRASIL. Relatório de Economia Bancária. Brasília, DF: Bacen, 2023.

BANCO DA INGLATERRA. History. Disponível em: https://www.bankofengland.co.uk. Acesso em: 26 mai. 2024.

BANK OF ENGLAND. Evolution of the UK banking system. 2023. Disponível em: https://www.bankofengland.co.uk/. Acesso em: 26 nov. 2023.

BANK OF ENGLAND. 867 - Global financial cycles since 1880. 2023. Disponível em: https://www.bankofengland.co.uk/. Acesso em: 26 nov. 2023.

BREIMAN, L. Random forests. Machine Learning, v. 45, n. 1, p. 5-32, 2001.

CAOUETTE, J.; ALTMAN, E.; NARAYANAM, P. Gestão do Risco de Crédito: o Próximo Grande Desafio Financeiro. Rio de Janeiro: Qualitymark, 1999.

DASTILE, X.; CELIK, T.; POTSANE, M. Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing Journal, v. 91, p. 106263, 2020. https://doi.org/10.1016/j.asoc.2020.106263

DUNCAN, R. A Concise History Of Creditism. 2018. Disponível em: https://www.richardduncaneconomics.com/. Acesso em: 26 nov. 2023.

FAZENDA, E. Risco de Crédito: Desenvolvimento de modelo Credit Scoring para a gestão da inadimplência de uma instituição de microcrédito. 2016.

FERREIRA, J. M.; OLIVEIRA, L. R. Crédito e sobreendividamento no Brasil: uma análise crítica. Revista de Administração Financeira, v. 19, n. 2, p. 213-230, 2022.

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, v. 55, n. 1, p. 119-139, 1997.

HENLEY, W. E.; HAND, D. J. Construction of a k-nearest-neighbour credit-scoring system. IMA Journal of Management Mathematics, v. 8, n. 4, p. 305-321, 1997.

HINDAWI. Application Analysis of Credit Scoring of Financial Institutions Based on Machine Learning Model. 2021. Disponível em: https://www.hindawi.com/journals/complexity/2021/9222617/. Acesso em: 26 nov. 2023.

HINTON, G. E.; OSINDERO, S.; TEH, Y. W. A fast learning algorithm for deep belief nets. Neural Computation, v. 18, n. 7, p. 1527-1554, 2006.

HISTORY OF FINANCE. The Evolution of Credit Systems Throughout History: A Comprehensive Overview. 2023. Disponível em: https://www.history-of-finance.com/. Acesso em: 26 nov. 2023.

JIANG, Y. A novel credit scoring model based on simulating annealing algorithm. Journal of Systems Science and Complexity, v. 19, n. 2, p. 195-201, 2006.

LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. Neural Computation, v. 1, n. 4, p. 541-551, 1989.

LEWIS, E. An Introduction to Credit Scoring. Fair Isaac: San Rafael, California, 1992.

LESSMANN, S.; BAESENS, B.; SEOW, H.-V.; THOMAS, L. C. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, v. 247, n. 1, p. 124-136, 2015.

LOUSADA, F.; ARA, A.; FERNANDES, G. B. Classification methods applied to credit scoring: Systematic review and overall comparison. In: GORMAN, M. F. Surveys in operation research and management science. USA: Science Direct, v. 21, p. 117-134, dezembro, 2016.

MARQUES, D. O. C. Credit scoring: regressão logística aplicada à dados de uma instituição financeira brasileira. Universidade de Brasília, 2015.

MARKOV, A.; SELEZNYOVA, Z.; LAPSHIN, V. Credit scoring methods: Latest trends and points to consider. In: The Journal of Finance and Data Science. USA: KeAi, v. 9, p. 180-201, novembro, 2023.

MENEZES, T. S.; ROCHA, E. M. Implicações do crédito na estabilidade macroeconômica: perspectivas para o Brasil. Revista de Análise Econômica, v. 69, n. 2, p. 234-250, 2021.

OFFICE OF FINANCIAL RESEARCH. Banking Credit System, 1970-2020. 2023. Disponível em: https://www.financialresearch.gov/. Acesso em: 26 nov. 2023.

OKADA, H. K. R.; NEVES, A. R. N. das; SHITSUKA, R. Análise de Algoritmos de Indução de Árvores de Decisão. Research, Society and Development, v. 8, n. 11, p. 01-15, 2019.

ONAY, C.; ÖZTÜRK, E. A review of credit scoring research in the age of Big Data. Journal of Financial Regulation and Compliance, v. 26, n. 3, p. 382-405, 2018. Disponível em: https://doi.org/10.1108/JFRC-06-2017-0054. Acesso em: 26 nov. 2023.

PARKINSON, K. L.; OCHS, J. R. Using credit screening to manage credit risk. Business Credit, p. 23-27, março, 1998.

PEREIRA, C. A.; LIMA, M. F. Crédito e crescimento econômico no Brasil. Revista de Economia Contemporânea, v. 76, n. 1, p. 58-77, 2022.

SAUNDERS, A. Medindo o Risco de Crédito – Novas Abordagens para o Value at Risk e Outros Paradigmas. Rio de Janeiro: Qualitymark, 2000.

SCHOEN, P. R. Análise de credit scoring utilizando redes neurais. 2017. Disponível em: http://hdl.handle.net/10183/179763.

SELEZNYOVA, Z.; LAPSHIN, V. Credit scoring methods: Latest trends and points to consider. In: The Journal of Finance and Data Science. USA: KeAi, Volume 9. Novembro, 2023.

SICSU, A. L. Desenvolvimento de um sistema de credit scoring – Parte I. São Paulo, Tecnologia de Crédito, n. 4, p. 63-76, janeiro, 1998a.

SICSU, A. L. Desenvolvimento de um sistema de credit scoring – Parte II. São Paulo, Tecnologia de Crédito, n. 5, p. 57-68, março, 1998b.

SILVA, A. L.; SANTOS, M. R. Impacto do crédito no consumo das famílias brasileiras. Revista Brasileira de Economia, v. 75, n. 3, p. 345-360, 2021.

SMITH, J.; DOE, A. Big Data in Credit Scoring: A Comprehensive Analysis. Journal of Financial Technology, v. 35, n. 4, p. 567-589, 2023.

SOUZA, R. B.; MENDES, F. J. Políticas de crédito e eleições no Brasil: uma análise histórica. Revista Brasileira de Política Econômica, v. 80, n. 4, p. 660-685, 2023.

THOMAS, L. C. Consumer credit models: Pricing, profit and portfolios. OUP Oxford, 2009.

THOMAS, L. C.; CROOK, J. N.; EDELMAN, D. B. Credit Scoring and Its Applications. SIAM, 2002.