

**APLICAÇÃO DA APRENDIZAGEM DE MÁQUINA NA ANÁLISE DE CRÉDITO  
NO SETOR DE COMÉRCIO DE MEDICAMENTOS**

***APPLICATION OF MACHINE LEARNING IN CREDIT ANALYSIS IN THE  
PHARMACEUTICAL RETAIL SECTOR***

**Emanuelle Mota Batista<sup>1</sup>  
João Paulo Calemba Batista Menezes<sup>2</sup>**

**RESUMO**

A análise de risco de crédito possui grande importância em diversas áreas. Por meio da análise, é possível classificar o risco de crédito envolvido na operação e determinar se um cliente será adimplente ou não, a fim de evitar futuros prejuízos. A análise de crédito é de grande importância para empresas que fornecem crédito aos seus clientes como forma estratégica de aumentar seus lucros. Entretanto, uma concessão errônea pode afetar o planejamento financeiro da organização. Assim, faz-se necessário o reconhecimento e a análise das melhores ferramentas e técnicas utilizadas para avaliar a capacidade de pagamento do devedor. A metodologia empregada neste estudo de caso possui uma abordagem quantitativa, bibliográfica e explicativa. Este estudo teve como foco verificar as formas de concessão de crédito em uma empresa varejista e quais são as ferramentas já utilizadas, sugerindo um novo método com o objetivo de melhorar a seletividade através da técnica de aprendizagem de máquina denominada Random Forest, combinada com o método SMOTE. Ao analisar o método aplicado, validou-se que a aprendizagem de máquina pode contribuir para o processo atual de análise. A combinação do fator humano com o fator "máquina" pode trazer maior eficiência e minimização dos riscos nas análises realizadas.

**PALAVRAS-CHAVE:** Análise de Crédito; Inadimplência; Varejo; Aprendizagem de máquina.

**ABSTRACT**

Credit risk analysis is of great importance in various fields. Through analysis, it is possible to classify the credit risk involved in the operation and determine whether a client will be compliant or not in order to avoid future losses. Credit analysis is highly significant for companies that provide credit to their clients as a strategic way to increase their profits.

---

<sup>1</sup> Bacharel em ciências contábeis pela UFVJM. R. Cruzeiro, 01 - Jardim São Paulo, Teófilo Otoni - MG, 39803-371.

<sup>2</sup> Doutor em Administração pela UFMG, professor adjunto no departamento de ciências contábeis (DCCO) da UFVJM. R. Cruzeiro, 01 - Jardim São Paulo, Teófilo Otoni - MG, 39803-371.

However, an erroneous concession may impact the organization's financial planning. Thus, it is necessary to recognize and analyze the best tools and techniques used to evaluate the debtor's payment capacity. The methodology employed in this case study is quantitative, bibliographic, and explanatory. This study focused on verifying the methods of credit concession in a retail company and the tools already used, suggesting a new method with the objective of improving selectivity through the machine learning technique called Random Forest, combined with the SMOTE method. It was validated by analyzing the applied method that machine learning can contribute to the current analysis process. Combining human and machine factors can bring greater efficiency and risk minimization in the analyses performed.

**KEYWORDS:** Credit Analysis, Default, Retail, Machine Learning.

## 1. INTRODUÇÃO

A concessão de crédito é uma prática que acompanha parte da evolução da sociedade, visto que, está ligada ao desenvolvimento econômico. Com o avanço da tecnologia houve o surgimento de ferramentas que auxiliam na implementação de recursos, garantindo uma maior segurança durante todo o processo.

Um dos meios mais importantes de financiamento no varejo é o crédito direto ao consumidor, ofertado pelas empresas no momento da venda. É realizado através de financeiras parceiras ou pela própria organização, que avaliam a condição financeira envolvida na operação, o crédito, podendo ser concedido ou negado sem que seja levado em conta seu efeito na rentabilidade do varejista.

É relevante que as organizações que vendem a prazo, possuam uma análise de crédito eficiente, pois é ela que irá garantir uma carteira de recebíveis notável. Sendo assim, Blatt (1999a, p. 39) ressalta que “Uma análise de crédito exige uma boa técnica no fluxo de caixa e nos endividamentos, fortalecendo as decisões de crédito”.

A análise de crédito realizada pelas empresas tem como objetivo auxiliar na sobrevivência da organização, minimizando o número de clientes inadimplentes. Ao conceder um crédito se tem a expectativa que ele retorne para a concedente e quando isso não ocorre pode gerar uma perda na organização.

“A análise de crédito compreende a aplicação de técnicas subjetivas, financeiras e estatísticas para avaliar a capacidade de pagamento do tomador de recursos, que é o proponente ao crédito.” (Castro Neto, Sérgio, 2009, p. 85). Através de uma concessão de crédito há o risco real de que quem o receba não venha a cumprir com suas obrigações, se tornando inadimplente.

A análise de crédito pode ser considerada como um instrumento essencial na organização, seja em qual for o cenário, se torna ainda mais relevante ao se considerar a situação econômica do país. Em um momento de crise, por exemplo, o número desemprego cresce, e conseqüentemente aumentar o índice de clientes inadimplentes.

As empresas concedem crédito aos clientes com intuito de aumentar as vendas e maximizar sua rentabilidade, porém é necessário ter cautela para que clientes

inadimplentes não gerem prejuízo a organização (Braga, 1995). De acordo com Silva (1997, p. 314), créditos inadimplentes são “aqueles que apresentam dificuldades de serem recebidos e conseqüentemente acarretam perdas para o credor”. As técnicas de avaliação de risco e de comportamento de crédito aplicadas de forma eficiente são instrumentos valiosos para as empresas avaliarem a concessão do crédito a seus clientes.

Segundo o Serasa Experian (2023), o processo de analisar as informações de crédito do cliente é uma das formas de identificar possíveis riscos envolvidos na negociação. Por isso, são utilizados diferentes critérios e pontuações que determinam o potencial risco de inadimplência do consumidor.

A concessão de crédito está cada dia mais cautelosa, segundo dados do Serasa Experian (2018) houve um aumento do número de inadimplência nas últimas década, impulsionado por fatores como: atraso de salários, desemprego, redução da renda.

Conforme enfatiza Blatt (1999a, p. 66) “É fundamental avaliar a situação financeira dos clientes e o risco associado aos seus negócios”.

Dessa forma, as empresas que optam por oferecer crédito como facilitador de vendas precisam também investir na gestão e controle do mesmo como forma de garantir sua saúde financeira (Silva, 2008).

Justifica-se a importância desse estudo sobre a relevância da análise de crédito antes que seja liberada uma compra, e balanceando os riscos envolvidos durante toda a operação. É preciso que haja uma cautela ao analisar as informações sobre o devedor, buscando informações suficientes que auxiliem na concessão do crédito.

Ao conceder a liberação de um crédito a um cliente, é necessário que haja cautela e uma avaliação rigorosa sobre o perfil do devedor, tendo em vista que, uma liberação errônea pode pôr em risco a situação financeira da empresa concessora.

Com base na delimitação do tema proposto, o presente trabalho tem como objetivo verificar a aplicação de aprendizagem de máquina durante o processo de análise de crédito por uma instituição varejista.

## **2. REFERENCIAL TEÓRICO**

### **2.1 Análise de Crédito**

Segundo Schrickel (2000, p. 27), “a análise de crédito envolve a habilidade de fazer uma decisão de crédito, dentro de um cenário de incertezas e constantes mutações e informações incompletas”. Por meio de uma análise criteriosa é possível estabelecer estratégias e identificar risco assumido na concessão do crédito.

O termo crédito é originário do latim, *creditum*, faz referência à confiança e está baseado na segurança de que algo é realmente verdade, algo em que se possa confiar; numa reputação boa, ter consideração. Segundo Schrickel (2000), o principal objetivo da análise de crédito, é identificar os riscos, obter conclusões quanto à capacidade de pagamento do tomador e realizar recomendações à estruturação e tipo de concessão do limite.

Para Schrickel (2000), crédito é todo ato de vontade ou disposição de alguém de destacar ou ceder, temporariamente, parte do seu patrimônio a um terceiro, com a expectativa de que essa parcela volte a sua posse integralmente.

Ao liberar crédito a seus clientes, a empresa concessora precisa avaliar o perfil dos clientes, levando em consideração se terão capacidade de pagamento futuro. Rodrigues (2012), acrescenta dizendo que o crédito é uma ferramenta utilizada pelas empresas para alavancarem suas vendas, a partir do instante em que tenham confiança na quitação da dívida assumida pelos clientes.

De acordo Assaf e Lima (2023, p. 694), “A análise de crédito tem por objetivo selecionar os clientes a prazo, sua capacidade de pagamento, assim como os limites monetários de crédito que podem ser concedidos”.

Uma empresa ao optar por trabalhar com a concessão de crédito, obtém clientes com os mais variados perfis, portanto, é imprescindível que haja uma política de crédito que garanta a empresa uma perspectiva maior de no futuro receber o crédito concedido.

O crédito disponibilizado aos clientes representa uma ferramenta significativa afim de assegurar a continuação do empreendimento, motivando um aumento maior nas vendas e nos lucros da empresa. Segundo Ross, Westerfield e Jaffe (2007), ao conceder crédito aos seus clientes, a empresa realiza um investimento associado a alguma venda de produto ou serviço.

Ao conceder crédito a um cliente, as empresas devem analisar seu perfil e julgar se o mesmo possui a possibilidade de quitar futuramente com a dívida. Barros e Gomes (2006) afirmam que as empresas devem tanto analisar a chance de conceder ou não o crédito para seus clientes, como também necessitam considerar a quantia máxima de crédito que este pode vir a obter.

O setor responsável pela análise de crédito tem como finalidade buscar uma maior redução na inadimplência dos seus clientes, o departamento de contas a receber é responsável por assegurar o retorno do valor do produto vendido ao cliente. As empresas concedem crédito aos clientes com intuito de aumentar as vendas e maximizar sua rentabilidade, porém é necessário ter cautela para que clientes inadimplentes não gerem prejuízo a organização (Braga, 1995).

Quando empresas acreditam em um cliente, as chances de conceder o crédito são maiores, portanto, para uma análise de crédito eficiente é preciso saber identificar o perfil e as condições financeiras do cliente a partir das suas informações básicas e dos seus dados financeiros.

Para fazer uma análise de crédito rigorosa é preciso que a empresa compreenda o comportamento do cliente em relação a suas finanças e sua renda. De acordo com o Serasa Experian (2023) existem algumas variáveis que podem auxiliar no momento da análise, como: Ter um bom score de crédito; Dados cadastrais atualizados; Históricos de pagamentos em dia; Pouco comprometimento da renda; Não possuir restrições no CPF; Bom relacionamento financeiro com empresas, bancos, etc.

De acordo Ross, Westerfield e Jordan (2007) a concessão de crédito é dada pela necessidade da empresa de aumentar suas vendas, entretanto, isso ocasiona custos de imobilização do capital, assim como o risco de o cliente não quitar, desse modo é preciso definir como conceder e como cobrar.

Ao realizar uma análise de crédito é importante obter informações pessoais do cliente, de forma que seja possível incluí-lo em novas compras e inserir um limite de crédito. Blatt (1999a), enfatiza que o cadastro é uma ferramenta essencial utilizada pelo credor, através dos dados fornecidos é possível avaliar a situação financeira e dimensionar o potencial do cliente.

Blatt (1999a, p. 78) afirma que:

A ficha cadastral nada mais é que o resumo da vida do cliente, onde temos dados que nos permitem quantificar os fatores anteriormente identificados que, quando analisados com o

resultado da pesquisa cadastral junto aos órgãos restritivos e junto a credores que já tiveram experiência com o cliente (fator caráter), permitem-nos avaliar com segurança a liquidez do crédito.

A ficha cadastral pode ser definida como o conjunto de informações de um indivíduo, para a identificação do cliente que está recebendo o crédito e também a fim de avaliar a possibilidade de concessão de crédito. De acordo Schrickel (2000), a ficha cadastral permite agrupar informações sobre a identificação e qualificação do indivíduo como: nome, endereço, registros, profissão, etc.

Silva (2008) reforça que, não existe um modelo único de ficha cadastral, devendo a organização desenvolver seu próprio modelo, de acordo com as características de seus clientes e suas necessidades.

De acordo com Leoni e Leoni (1997), no cadastro de pessoa física deve ser apurado a maior quantidade de informações possíveis informadas pelo próprio cliente e os dados coletados deverão ficar arquivadas junto a empresa.

Fiorentini (2004), ressalta algumas informações a serem seguidos para a redução da inadimplência ao se realizar um cadastro, como: Registro Geral e cadastro de Pessoa física; empresa em que trabalha, cargo e renda; telefone para contato; comprovante de residência; consulta ao Serviço de Proteção ao Crédito; em caso de cliente antigo, manter os dados atualizados.

Com esses dados é possível que o analista realize uma consulta com mais precisão, e possa concluir com mais segurança se é possível conceder a liberação do crédito.

O limite de crédito é aquele concedido ao cliente com base em seu perfil, ele corresponde ao valor máximo que poderá ser utilizado em compras, o limite concedido varia a cada cliente, o consumidor que possui uma renda mais elevada consequentemente adquire um limite maior de crédito a ser liberado (Sá, 2004).

Silva (2002) define o limite de crédito à pessoa física como o valor total que o conessor compromete a disponibilizar ao cliente, seja por meio de produtos, serviços ou valor em espécie, diante da análise das suas características e avaliação do potencial de quitação de todo o valor concedido dentro do tempo estipulado. Nos dias atuais, uma das técnicas utilizadas por grandes empresas para definir um limite ao cliente é a mensuração a partir da renda presumida do cliente junto do seu limite de score.

Com o objetivo de se obter um limite condizente com a característica de cada cliente é necessário que as informações contidas nas fichas cadastrais estejam atualizadas frequentemente como: telefone, renda, emprego, endereço e entre outras variáveis que podem afetar a realização do pagamento.

Segundo Schrickel (2000), para minimizar o risco com inadimplência, a empresa precisa estabelecer um valor máximo de limite para venda, limite esse que varia de acordo com o perfil de cada cliente. Desse modo, quanto maior for a capacidade de pagamento, maior será o limite disponibilizado.

## **2.2 Política de Crédito**

A política de crédito é método utilizado pelas empresas com a finalidade de minimizar os riscos e balancear seus recebimentos, trata-se de um conjunto de regras determinadas com a finalidade de subsidiar recursos aos clientes através da concessão de crédito, estabelecer a quem disponibilizar o crédito.

Segundo Silva (2008), as políticas de crédito são modelos que determinam padrões na decisão ou até mesmo auxiliam na resolução de problemas. A principal finalidade da política de crédito é orientar nas decisões de crédito, tendo foco nas metas estabelecidas.

De acordo com Ross, Jaffe e Werterfield (2007), para uma política de crédito bem elaborada é preciso que haja um equilíbrio entre os benefícios gerados por meio do aumento de vendas e os custos com a concessão de crédito.

Conforme Silva (2008, p.104) a política de crédito é definida como ‘padrões de crédito’ e seu objetivo é direcionar nas decisões desejadas. Silva ainda complementa descrevendo a política de crédito como: Uma orientação para a decisão de crédito, porém não é a decisão; Guia a concessão de crédito, entretanto não concede o crédito; Orienta a concessão de crédito para finalidade pretendida, mas não é o objetivo em si.

Schrickel (2000, p. 33) afirma que, “a política de crédito deve ser lúcida, flexível e pragmática, ou seja, realista em relação ao mercado, mutável ao longo do tempo e auxiliadora da concretização dos negócios”

A escolha para a formação desta política deve levar em consideração a associação entre os benefícios oriundos do aumento de venda com o controle do custo para a concessão do crédito. Adotando-se uma política restritiva, ela terá uma escassez de crédito. Adotando uma política de crédito rigorosa, o volume de possíveis inadimplentes será baixo, todavia, o volume de suas vendas também (Ross; Westerfield; Jaffe, 2007).

Em outras palavras a política de crédito busca equilibrar a intenção de lucrar com as necessidades do cliente. O objetivo é lucrar, reduzir os riscos envolvidos, e satisfazer os clientes, ao mesmo tempo em que se mantém um portfólio de crédito sólido.

### **2.3 Inadimplência**

A inadimplência é uma realidade que afeta grande parte da população, em consequência da falta do cumprimento de uma obrigação dentro do prazo estabelecido muitas pessoas acabam se endividando por não conseguirem arcar com o pagamento de suas dívidas.

“A inadimplência é caracterizada pelo não cumprimento de uma obrigação, no prazo pactuado. Este descumprimento, parcial ou total, ocorre em desacordo com as normas” (Trento, 2009, p. 53).

A inexistência de uma gestão eficaz no orçamento atrelada ao crescente número de desempregados e a instabilidade econômica do País são algumas das muitas razões que contribuem diretamente para a inadimplência. Nas últimas décadas as empresas conessoras de crédito vêm aumentando sua preocupação em decorrência do não pagamento da dívida por parte dos clientes, o que pode gerar um impacto financeiro negativo no caixa.

Trento (2009) evidencia que a inadimplência sucede a concessão de crédito ao cliente que já provém de instabilidade financeira e que não possui capacidade de quitar suas dívidas vencidas e a vencer. Portanto, as empresas são as responsáveis pela decisão de conceder ou não o crédito, e, muitas vezes, por não possuírem ferramentas corretas para uma análise eficaz, acabam disponibilizando por confiança.

De acordo Teixeira e Paixão (2019), as ações dos homens em relação às decisões financeiras vêm sendo estudadas de maneira mais intensiva, visto que os seres humanos

possuem necessidades, entretanto, o modo ao qual recorrem para supri-las apresenta características específicas voltadas a condições como idade, gênero, renda, etc.

As empresas buscam reduzir os números de inadimplentes em sua carteira de clientes, em consequência disso, é indispensável conhecer o perfil dos clientes para que se entenda os reais motivos que fomentam a inadimplência e assim evitar ou reduzir tal acontecimento. Fornecer crédito aos clientes que possuem condições para quitar futuramente com o compromisso é uma forma de amenizar o risco assumido e livrar-se de uma possível inadimplência.

Segundo a definição do dicionário Dicio, a inadimplência significa “Ação ou efeito de inadimplir, de não cumprir; incumprimento.” Conforme Ferreira (2006), endividados podem ser descritos como aqueles que contraem dívidas e comprometem uma parcela significativa de suas rendas e rendimentos para honrá-las.

Inadimplência significa o “descumprimento de uma obrigação previamente acordada”, conforme Ferreira (2006), endividados podem ser descritos como aqueles que contraem dívidas e comprometem uma parcela significativa de suas rendas e rendimentos para honrá-las.

Para Bessis (1998, p. 82) a inadimplência traz o seguinte conceito: “[...] deixar de pagar uma obrigação, quebrar um acordo, entrar em um procedimento legal ou default econômico”. Em consequência da inadimplência, muitos consumidores começam ter o crédito negado na hora de fazer suas compras.

O Serasa Experian (2023) aponta um estudo na área de análises de informações onde mostra o perfil do inadimplente no Brasil, sendo ele: Homens são maioria, sendo 50,9% dos inadimplentes; Pessoas que ganham de 1 a 2 salários mínimos (39,1%); Pessoas que ganham até 1 salário mínimo (38,8%); Inadimplentes que possuem apenas uma dívida (37,3%); Consumidores que possuem quatro dívidas ou mais (30,7%); Inadimplentes com idade entre 41 e 50 anos (19,4%); Inadimplentes que possuem idade entre 18 a 25 anos (14,9%).

## **2.4 Riscos na Concessão de Crédito**

Risco de crédito, segundo Caouette, Altman e Narayanan (1999, p. 1), “[...] é a mais antiga forma de risco nos mercados financeiros”, tendo sua origem há pelo menos 1.800 anos a.C.

O risco de crédito é geralmente definido como o risco de um devedor não liquidar integralmente seus compromissos. De acordo com Duarte Júnior (1996, p. 5), o risco está associado ao "recebimento de um valor contratado, a ser pago por um tomador de empréstimo, contraparte de um contrato ou emissor de um título, descontadas as expectativas de recuperação e realização de garantias".

Em outras palavras o risco de crédito é a probabilidade de que o tomador de recursos não queira ou não consiga pagar tanto o valor principal da dívida, como seus respectivos juros.

Santos (2009, p. 2), reforça que:

O risco de inadimplência constitui-se em umas das principais preocupações dos credores, tendo em vista relacionar-se com a ocorrência de perdas financeiras que poderão prejudicar a liquidez (capacidade de honrar dívidas com os supridores de capital) e a captação de recursos nos mercados financeiros e de capitais.

De acordo Blatt (1999a, p. 66) “É fundamental avaliar a situação financeira dos clientes e o risco associado aos seus negócios, através de uma avaliação criteriosa a empresa consegue minimizar os riscos envolvidos durante uma transação”.

Diferente de compras à vista, onde o dinheiro passa imediatamente para conta de quem realizou a venda, nas operações que envolvem crédito, a empresa precisa assumir o risco que o cliente irá honrar com aquela dívida. Entretanto, existe a possibilidade da não quitação da dívida em qualquer transação financeira, o que é chamado de risco de crédito.

O risco de crédito é dividido em duas categorias, a primeira é quando o cliente tem grandes chances de não quitar o crédito concedido, nesse caso, a empresa pode optar por não conceder. Já na segunda categoria, a empresa entende que sofre menos riscos ao oferecer crédito para determinado cliente e permite a concessão do crédito.

De acordo Bessis (1998, p. 81), o risco de crédito pode ser avaliado por meio dos seus componentes, que são o risco de default, o risco de exposição e o risco de recuperação. O risco de default está associado à probabilidade de ocorrer um episódio de default com o tomador em um certo intervalo de tempo, o risco de exposição sucede da incerteza em relação ao valor do crédito no momento do default, já o risco de recuperação está atrelado à incerteza quanto ao valor que pode ser recuperado pelo credor no caso de um default do tomador.

## 2.5 Scoring de Crédito

Conforme Barbieri (2011) o scoring de crédito é um tipo de análise baseada em estatística no qual busca prever a possibilidade de quitação de um crédito. É muito utilizada para avaliação de concessões de crédito pelas empresas para mensurar a probabilidade de o consumidor cumprir com suas obrigações financeiras, por meio de uma pontuação de crédito.

O *credit scoring* é um método de pontuação que surgiu em meados dos anos 1970, sendo uma ferramenta de auxílio durante o grande fluxo de análises de crédito (Santos, 2015). Silva (2008, p. 348) ressalta que “O sistema de credit scoring possibilita resposta rápida para decisão de crédito massificado.

Goncalves (2005) retrata que os modelos de análise de concessão de crédito, são tidos a partir de dados históricos da base de clientes já existentes a fim de analisar se um possível cliente terá mais probabilidade de ser um bom ou mau pagador, medindo a tendência de tornar-se inadimplente.

Para Saunders (2000), por meio da pontuação obtida em cada crédito, é possível classificar créditos como adimplentes ou inadimplentes, bons ou maus, desejáveis ou não. Através do resultado obtido, o analista pode tomar uma decisão se vai conceder o crédito ou não.

Blatt (1999a) enfatiza que para o desenvolvimento do Credit Scoring deve-se seguir os seguintes passos:

- Planejamento e definições: mercados e produtos de crédito para os quais será desenvolvido o sistema, finalidades de uso, tipos de clientes, conceito de inadimplência a ser adotado e horizonte de previsão do modelo;
- Identificação das variáveis potenciais: caracterização do proponente ao crédito, caracterização da operação, seleção das variáveis significativas para o modelo e análise das restrições a serem consideradas em relação às variáveis;



- Planejamento amostral e coleta de dados: seleção e dimensionamento da amostra, coleta dos dados e montagem da base de dados;
- Determinação da fórmula de escoragem através de técnicas estatísticas, como por exemplo, a análise discriminante ou regressão logística;
- Determinação do ponto de corte, a partir do qual o cliente é classificado como adimplente ou bom pagador; em outras palavras, é o ponto a partir do qual a instituição financeira pode aprovar a liberação do crédito.

A utilização de um modelo de score de crédito não é o suficiente para uma gestão bem sucedida, é indispensável que seja avaliada a situação financeira atual do cliente. Blatt (1999b) enfatiza que os modelos de score de crédito evoluíram a partir de um modelo do C's do crédito, no qual determinadas características estão relacionadas aos maus e bons pagadores.

Segundo dados do Serasa Experian (2023) o cliente que possui um cadastro positivo, com um score considerado relativamente bom, teria o perfil de cliente que paga pontualmente suas dívidas.

Através da pontuação do cliente as empresas conseguem avaliar os riscos na concessão do crédito. A pontuação varia de 1 a 1000, quanto maior a pontuação, menor será o risco do cliente se tornar inadimplente, o consumidor que possui uma alta pontuação possui uma melhor chance de conseguir um crédito com condições melhores.

Caouette, Altman e Narayanan (1999) ressaltam que no uso do Score de crédito, embora sejam utilizados sistemas com ênfase na avaliação do risco de crédito para tomada de decisão sobre a concessão de crédito, algumas empresas também utilizam do sistema para delimitar o crédito a ser concedido

Conforme o portal Serasa Score (2022), para um score de até 299, há um grande risco de o consumidor atrasar o pagamento de suas contas. Para uma pontuação entre 300 e 699 o risco se torna moderado. Em contrapartida, consumidores que possuem uma pontuação de score igual ou acima de 700 são mais bem avaliados por possuírem uma maior capacidade de arcar com suas contas.

De acordo com o Serasa Experian (2023), algumas das variáveis utilizadas para cálculo do score são:

- Histórico de pagamentos: Clientes que possuem pagamentos em dia adquirem um score maior que aqueles que estão em situação de atraso.
- Registro de dívidas e pendências: Pendências prejudicam a pontuação.
- Consultas ao CPF para ofertas de crédito: Muitas consultas em um prazo curto tendem a prejudicar o score.
- Evolução financeira: Clientes que possuem um cadastro antigo costumam obter uma pontuação mais alta.

Em conformidade com as variáveis utilizadas para cálculo do score, as mesmas também são utilizadas como parâmetro para analisar o perfil do cliente na empresa estudada. Como exemplo, clientes que possuem registro SPC por outras empresas não conseguem obter crédito para compras.

### **3. PROCEDIMENTOS METODOLÓGICOS**

De acordo Minayo (2010, p. 46), a metodologia “mais que uma descrição formal dos métodos e técnicas a serem utilizados, indica as conexões e a leitura operacional que o pesquisador fez do quadro teórico e de seus objetos de estudo”.

O estudo desse trabalho busca analisar é o processo de concessão de crédito em uma empresa que atua no comércio varejista farmacêutico, visando conhecê-lo no âmbito de sua política de crédito, análise e concessão de crédito, e os riscos presentes em cada operação e assim desenvolver uma ferramenta que auxilie na tomada de decisão.

A pesquisa consiste em uma análise das variáveis realizadas envolvidas no processo de concessão de crédito, vale ressaltar que a empresa estudada não terá seu nome mencionado, sendo mantido em sigilo.

No decorrer deste procedimento iremos retratar o modo que será realizado o presente estudo, buscando uma melhor maneira de se alcançar os objetivos propostos na elaboração deste trabalho.

Há diversas maneiras de classificar as pesquisas. Para Gil (2019) as maneiras mais comuns de classificação estão relacionadas a forma que o problema é abordado, aos objetivos do estudo e também em relação aos procedimentos técnicos utilizados.

Desta forma, a começar do ponto de vista que o problema é abordado, o presente estudo pode ser considerado como uma pesquisa quantitativa, uma vez que resultará em análises e classificações com base nos dados numéricos.

Em relação aos objetivos, a pesquisa pode ser classificada como exploratória, explicativa e descritiva. Conforme Silva (2008), a pesquisa exploratória busca gerar uma proximidade com um problema de pouco conhecimento, buscando torná-lo claro ou idealizar hipóteses.

A pesquisa explicativa objetiva detectar os fenômenos que contribuem para a ocorrência de determinado acontecimento. Já a pesquisa descritiva expõe as particularidades de determinado fenômeno, através do uso de técnicas padronizadas de coleta de dados, como questionário.

Este projeto de pesquisa, levando em consideração os seus objetivos abordados, pode ser classificado como uma pesquisa explicativa, em razão de que ele busca identificar e analisar as variáveis utilizadas na análise de crédito por uma instituição privada do ramo farmacêutico.

Para dar início ao estudo, o presente projeto de pesquisa é pautado de natureza bibliográfica, através de material já publicado, às informações e o embasamento teórico fundamental para o desenvolvimento da questão central do projeto.

Gil (2019) afirma que a pesquisa bibliográfica é realizada com base em material já desenvolvido, constituído especialmente de artigos científicos e livros. O levantamento bibliográfico foi realizado a partir da leitura de livros, artigos, revistas, sites e publicações relacionadas ao tema.

### **3.1 Aprendizado da máquina**

#### **3.1.1 Modelo *Machine Learning***

O presente trabalho está focado no uso de técnicas de Machine Learning, de forma que auxilie na decisão de crédito, tornando o processo de concessão de crédito mais seguro e conseqüentemente a minimização da inadimplência.

A área de Machine Learning ou aprendizagem de máquina é um subcampo dentro da Inteligência Artificial (IA) que pesquisa algoritmos de computador e os usa para

aprender. Seja com o objetivo de comportar de uma maneira inteligente ou de finalizar uma tarefa através de previsões claras (Mitchell, 1997).

O conceito por trás do machine learning é a elaboração de algoritmos que melhoram automaticamente com a experiência. O presente estudo busca utilizar dos métodos de machine learning para *credit scoring*, apontando seu desempenho ao estimar a probabilidade de concessão ou não do crédito utilizando como base de dados uma amostra de clientes. Segundo Matos (2015), programas de Machine Learning também são programados para aprender e aperfeiçoar ao longo do tempo quando expostos a novos dados.

Desse modo, seu uso nos modelos de *credit scoring* aumentam o grau de eficiência e assertividade nos retornos. Quanto aos métodos de machine learning, considerou-se o algoritmo floresta aleatória (*random forest*) combinado com o método SMOTE.

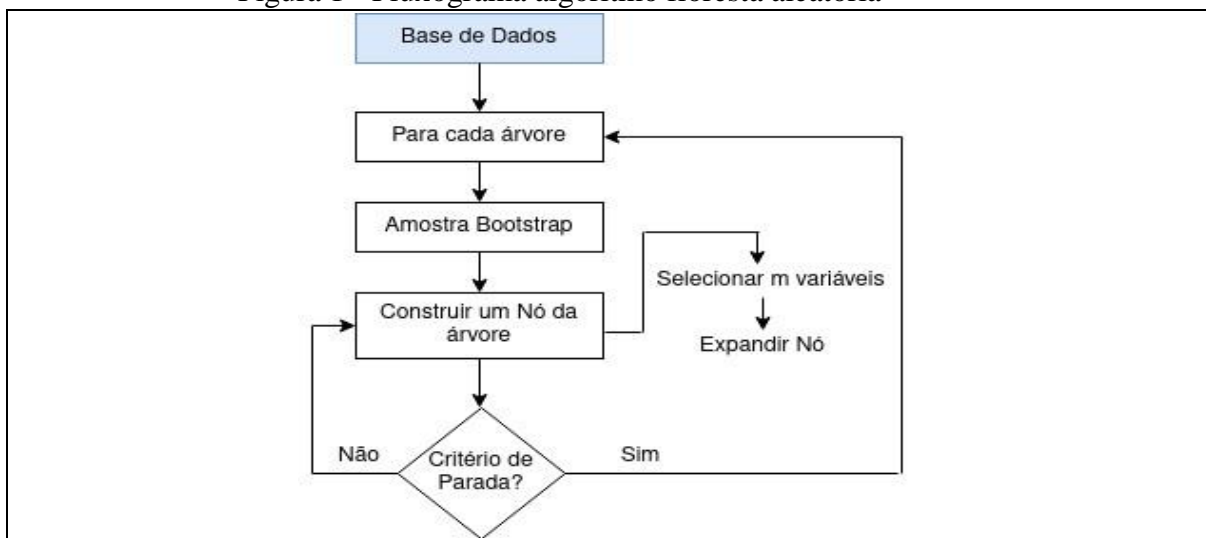
O *random forest* (floresta aleatória em português) é um algoritmo de machine learning baseado em árvores de decisão, proposto por Breiman em 2001, trata-se de um aperfeiçoamento do método *bagging*, empregando uma seleção aleatória de algumas das variáveis na construção das árvores de decisão. Esse algoritmo é capaz de reduzir o número de erros gerados em cada árvore.

Conforme Machado e Karray (2022) os modelos de floresta aleatória são uma espécie de modelos aditivos que realizam previsões misturando as saídas do modelo árvore de decisão, que vêm de uma sequência de modelos básicos. Cada árvore é gerada através do desenvolvimento de duas fases, sendo elas:

1. Um número aceitável de árvores de decisão é originado através de um conjunto de dados. Cada árvore possui um subconjunto aleatório de variáveis  $m$  (preditores), tal que  $m < M$  (sendo  $M$  = total de preditores).
2. Cada árvore cresce ao máximo.

A maior vantagem de se utilizar o Random Forest para o processo de análise de crédito é o fato de ser uma técnica eficiente, capaz de detectar fraudes e classificar clientes para concessão de crédito, figura 1.

Figura 1 - Fluxograma algoritmo floresta aleatória



Fonte: Oliveira Neto (2021)

O algoritmo SMOTE é uma técnica estatística utilizada para expandir número de casos em um conjunto de forma equilibrada. Atualmente é o algoritmo de amostragem mais utilizado para o balanceamento dos conjuntos de dados, através dessa técnica é possível produzir dados sintéticos por meio da inserção de exemplos reais.

Para desenvolvimento do presente estudo, os dados coletados apresentaram um desbalanceamento, tendo em vista que, em uma base de 155 análises estudadas apenas 33 eram casos de crédito negado. Sendo assim, para equilibrar as classes utilizou-se da técnica SMOTE.

Para a execução do experimento, o modelo foi testado utilizando a linguagem Python. A linguagem Python é aplicada conforme seu propósito, contudo, este instrumento não possui um objetivo específico, ou seja, não existe um trabalho definido que o idioma deva completar, portanto, seu objetivo é geral.

O Python é uma linguagem rápida e de fácil entendimento, possui código aberto, e abriga milhares de módulos de terceiros. Tanto os módulos quanto as bibliotecas padrões do Python permitem inúmeras possibilidades para sua utilização. Para o desenvolvimento desse estudo a linguagem Python foi utilizada para rodar os códigos, garantindo que a aplicação pudesse ser testada com maior eficiência.

### 3.1.2 Operacionalização das variáveis da amostra

Por meio da análise estatística dos dados foi possível observar o efeito das variáveis selecionadas durante uma análise de crédito e seus respectivos riscos. A seguir, são apresentadas as variáveis utilizadas para desenvolvimento do modelo algoritmo, as mesmas são utilizadas no dia a dia da empresa estudada, a fim de determinar se o crédito será concedido ou não ao cliente.

**Valor da compra (VC):** Variável que representa o valor no qual o cliente deseja realizar compras.

**Idade (ID):** A variável idade foi transformada entre jovens, adultos e idosos para ser analisada. Espera-se que indivíduos com idades avançadas, assim como aqueles recém maiores de dezoito anos sejam caracterizados um risco maior.

**Limite Disponível (LD):** Após verificar se o cliente está apto para obter algum tipo de crédito é concedido um limite fixo inicial de acordo com o perfil de cada cliente.

**Limite temporário concedido (LTC):** Variável indica o limite temporário concedido ao cliente para efetuar compras quando o valor da sua compra for maior que o limite fixo já disponível. Limite só é concedido após aprovação do crédito.

**Status do cliente (SC):** Variável define situação do cliente, definindo se está apto ou não a realizar compras.

Para desenvolvimento do modelo foi realizada exclusão de algumas variáveis utilizadas pela empresa estudada, sendo elas:

**Profissão (P):** Através dessa variável é avaliada a profissão, visto que, o limite concedido pode sofrer influência da profissão exercida pelo cliente.

**Estado (E):** A variável estado busca analisar a região em que o cliente se encontra, considerando que algumas regiões da empresa estudada há uma maior incidência de clientes inadimplentes.

**Produto (SC):** Produtos considerados de uso contínuo, durante uma análise de crédito possuem uma maior flexibilidade, considerando a necessidade do cliente.

As variáveis descritas acima foram excluídas por não serem numéricas, portanto, sua conversão em classes poderia gerar erros na leitura do algoritmo.

## 4. ANÁLISE DOS RESULTADOS

Um modelo de escore é considerado eficaz através da avaliação da sua capacidade em distinguir a amostra testada em “maus” e “bons” e “maus” cliente. O modelo de Random Forest utilizado nesse estudo confirma que os algoritmos de aprendizado de máquina podem auxiliar na tomada de decisão.

Para o desenvolvimento do modelo foi utilizado uma base com 155 casos reais fornecida pela empresa objeto de estudo, os dados obtidos forma gerados por clientes de operações de crédito ao consumidor no varejo. Na elaboração do modelo utilizou-se de variáveis já utilizadas na empresa estudada, de modo que foi possível confrontar se o resultado obtido pelo modelo seria compatível com os casos reais do dia a dia.

Após aplicação do Random Forest foi confirmado por meio dos testes realizados se tratar de um modelo de grande eficiência na construção de modelos de escore de crédito, sendo o mesmo capaz de validar as variáveis e identificar em quais casos o crédito poderá ou não ser concedido.

De acordo com Diniz e Louzada (2012), para modelos de *Credit Scoring* estima-se que a estatística do Teste KS seja maior ou igual a 30%, de modo que o modelo seja considerado como um bom discriminante e valores inferiores a 20% representem baixa resultado. Diante disto, o algoritmo de Random Forest apresentou um desempenho satisfatório, obtendo uma acurácia de até 81% em relação a amostra original.

### 4.1 Ferramentas

#### 4.1.1 Base de dados

A base de dados utilizada neste trabalho foi formada por clientes de uma operação de crédito, foram coletados dados a partir de uma amostragem de 155 casos reais contendo informações pessoais dos clientes, e o retorno de aprovação ou não do crédito. Conforme a figura 3, das 155 vendas analisadas em um período de 1 mês, 79% representam análises aprovadas, enquanto 21% correspondem as vendas negadas.

A amostragem coletada procede de uma amostra desbalanceada, visto que, foi apurado poucos casos de crédito negado, por isso utilizou-se da técnica SMOTE para equilibrar a realização dos testes. A base de dados da amostra é formada por 8 variáveis. A tabela 1 apresenta a estatística descritiva da amostragem com os 155 casos reais analisados.

Conforme os dados apresentados na tabela 1, em uma contagem com 155 dados é possível identificar a média, o desvio padrão, o mínimo, 25%, 50%, 75% e o máximo das variáveis analisadas. O retorno apurado se trata de uma distribuição desbalanceada apresentando um desvio padrão de 0.41068.

O mínimo do valor de venda corresponde a 15, em contrapartida o máximo é de 950. A variável limite fixo possui uma média de 368.69, já o limite utilizado apresenta uma média de 254.56, chegando a um máximo de 2.974.

Conforme a tabela 1, o limite disponível médio obtido foi de 114.13, tendo seu máximo de 1026.

Tabela 1 - Estatística descritiva da amostragem

	Retorno	Faixa etária	Valor da venda	SPC	Limite fixo	Limite disponível	Limite utilizado
Contagem	155	155	155	155	155	155	155
Média	0.21290	0.23871	184.43	0.406	368.69	114.13	254.56
Desvio padrão	0.41068	0.44260	130.81	0.492	390.49	164.22	371.53
Mínimo	0	0	15	0	0	-420	0
25%	0	0	100	0	200	48.50	62
50%	0	0	163	0	265	120	160
75%	0	0	238.50	1	400	195	313
Máximo	1	2	950	1	4.000	1.026	2.974

Fonte: Dados da pesquisa

#### 4.1.2 Análise e tratamento dos dados

A primeira fase é a de consolidação da base de dados, logo, busca-se os preparar os dados antes de serem testados pelo algoritmo. Após essa fase é iniciado a etapa de processar os dados obtidos e tratá-los através do google Colab (ferramenta com acesso gratuito) de forma que seja possível rodar os códigos por meio da linguagem python.

Para desenvolvimento do presente trabalho foram realizados dois testes, sendo um em base original de 155 casos e outro com dados simulados contendo 16.000 ocorrências, no qual 8.000 observações são casos negados e os outros 8.000 representam os aprovados.

O algoritmo utilizado realiza uma divisão entre treino e teste, sendo assim, por default ele usa 70% para treinar o algoritmo e 30% para testar. Após a importação dos dados, é realizado uma verificação para validar se a amostra está balanceada, também é avaliado se existem dados faltantes na amostra.

Em situações em que existem um considerado desequilíbrio entre ocorrências na variável resposta, a possibilidade de que as amostras iniciais possuam poucas observações da classe minoritária é significativo, resultando em um menor desempenho de previsão da classe inferior. Após avaliação, foi comprovado tratar-se de uma amostragem desbalanceada, tendo em vista que, das 155 ocorrências reais analisadas, 122 amostras tratava-se de análises aprovadas, e apenas 33 negadas.

Tendo em vista proceder-se de uma amostra desbalanceada, conforme figura 2, é realizado o balanceamento por meio da aplicação do método SMOTE. Os dados são classificados como desbalanceados quando a base de dados não possui as classes uniformemente representadas, pois há ocorrência de uma ou mais classes minoritárias em comparação com as demais classes majoritárias.

Figura 2 - Amostra com os 155 casos desbalanceados

```

0    122
1     33
Name: retorno, dtype: int64
0     78.7%
1     21.3%
Name: retorno, dtype: object
    
```

Fonte: Dados da pesquisa

A aplicação da técnica SMOTE na amostragem de 155 casos gerou uma nova amostra com 244 casos balanceados, obtendo 85 amostras aprovadas para 85 negadas, gerando um total de 170 casos para treinamento do modelo, já no segundo bloco obteve-se 37 amostras aprovadas para 37 negadas, resultando em um total de 74 amostras para testes.

Conforme a figura 3, é possível validar a nova amostra com os dados balanceados por meio da técnica SMOTE.

Figura 3 - Balanceamento da amostra com 244 casos

1	85
0	85
Name: retorno, dtype: int64	
1	37
0	37
Name: retorno, dtype: int64	

Fonte: Dados da pesquisa

Na amostra com 16.000 casos após a aplicação da mesma técnica, obteve-se em um balanceamento de 5.600 amostras aprovadas para 5.600 negadas, constituindo um total de 11.200 casos para treinamento do modelo, enquanto no segundo bloco obteve-se 2.400 amostras aprovadas para 2.400 negadas, ocasionando em um total de 4.800 amostras para testes. Na figura 4 é possível observar o balanceamento da amostra com as 16.000 ocorrências.

Figura 4- Balanceamento da amostra com 16.000 casos

1	5600
0	5600
Name: retorno, dtype: int64	
1	2400
0	2400
Name: retorno, dtype: int64	

Fonte: Dados da pesquisa

#### 4.1.3 Aplicação do Random Forest

Após a fase de balanceamento de dados é realizado a aplicação do Random Forest em ambas as amostragens. Para mensuração dos resultados foram avaliadas as métricas do modelo, sua Acurácia e matriz de confusão e assim apurar o nível de eficiências em acertos do modelo

Através do indicador de desempenho, definido como acurácia, é possível avaliar a taxa de acertos gerada pelo modelo. A acurácia é responsável por indicar a taxa de classificação correta total do modelo e assim informar qual o número de predições corretas sobre o total de predições, sejam negativas positivas.

Na matriz de confusão, as linhas simbolizam valores citados pelo modelo e as colunas retratam os valores conhecidos e anotados. A diagonal principal apresenta os valores corretos classificados pelo algoritmo, sendo eles verdadeiros positivos (VP) e verdadeiros negativos (VN). As outras células da matriz retratam os valores classificados incorretamente, os falsos positivos (FP) e os falsos negativos (FN).

Para a amostragem de 244 casos, conforme a figura 5, validou-se que o modelo acertou 81% do que foi testado, ou seja, de 100 casos testados o modelo é capaz de obter 81 acertos.

Figura 5 - Classificação do modelo com 244 amostras

```

Classification metrics:
      precision    recall  f1-score   support

   0           0.85     0.76     0.80        37
   1           0.78     0.86     0.82        37

 accuracy          0.81
 macro avg         0.81
 weighted avg      0.81

Acurácia:
0.8108108108108109
Confusion Matrix:
[[28  9]
 [ 5 32]]
Curva ROC:
0.8108108108108109
    
```

Fonte: Dados da pesquisa

Comparando os resultados da matriz de confusão, observamos que o sistema conseguiu analisar corretamente 28 (VP) e 32 (VN) análises. Em relação aos erros, o sistema errou em identificar 5 (FP) e 9 (FN) retornos.

Analisando a amostra com 16.000 casos fictícios, conforme a figura 6, verificou-se que o modelo acertou 99% do que foi testado, logo, em uma amostra de 100 ocorrências o modelo consegue obter uma média de 99 acertos.

Figura 6 - Classificação do modelo com 16.000 amostras

```

Classification metrics:
      precision    recall  f1-score   support

   0           1.00     0.99     0.99    2400
   1           0.99     1.00     0.99    2400

 accuracy          0.99
 macro avg         0.99
 weighted avg      0.99

Acurácia:
0.9922916666666667
Confusion Matrix:
[[2370  30]
 [  7 2393]]
Curva ROC:
0.9922916666666668
    
```

Fonte: Dados da pesquisa

Confrontando os resultados da matriz de confusão na amostra com 16.000 ocorrências, constatou-se que o sistema conseguiu analisar corretamente 2.370 (VP) e 2.393 (VN) análises. Em relação aos erros, o sistema errou em identificar 7 (FP) e 30 (FN) retornos.



#### 4.1.4 Validação das variáveis importantes - Random Forest

Após a aplicação do modelo Random Forest na amostragem com 244 ocorrências é verificado quais são as variáveis com maior importância no processo decisório. A Figura 7 apresenta a relevância das variáveis na elaboração do modelo.

Figura 7 – Demonstração das variáveis importantes com 244 amostras

	feature	importance
5	limiteutilizado	0.322617
1	valorvenda	0.226418
4	limitedisp	0.194545
3	limitefixo	0.187781
2	spc	0.056382
0	faixaet	0.012258

Fonte: Dados da pesquisa

Destacam-se o limite utilizado com uma importância de 32,26%, seguido do valor de venda com 22,64% e o limite disponível apresentando 19,45%. Em contrapartida é possível verificar as variáveis com menores relevâncias como: limite fixo apresentando 18,79 % de relevância, SPC com 6,63% e pôr fim a faixa etária com 1,22%.

Com esses dados pode-se concluir que o limite utilizado possui uma maior relevância em relação as outras variáveis apresentadas, confrontando o resultado obtido com o dia a dia da empresa, valida-se que a quantidade de débitos em aberto do cliente é um ponto que sempre é ponderado ao conceder um crédito.

Assim como nas análises reais, em conformidade com o resultado do estudo que gerou a faixa etária como a variável de menor importância, a idade do cliente não possui tanta relevância em relação as outras estudadas ao avaliar se deve ou não conceder um crédito.

Posteriormente a aplicação do modelo na amostra com os 16.000 casos fictícios, os resultados obtidos conforme a figura 8 foram: limite disponível, possuindo uma maior relevância de 24,79%, seguido do valor de venda com 22,77% e o limite utilizado com 20,61%.

Figura 8 – Demonstração das variáveis importantes com 16.000 amostras

	feature	importance
4	limitedisp	0.247857
1	valorvenda	0.227677
5	limiteutilizado	0.206137
3	limitefixo	0.162066
2	spc	0.151559
0	faixaet	0.004704

Fonte: Dados da pesquisa

Por outro lado, as variáveis com a menor relevância são: limite fixo com 16,20%, acompanhado do SPC com 15,15% e a faixa etária possuindo a menor relevância com 0,47%.

Em conformidade com os resultados obtidos na amostra com as 244 ocorrências, a amostra com os 16.000 casos manteve o limite disponível, o valor da venda e o limite utilizado como as variáveis mais importantes durante uma análise de crédito.

Sendo assim, concluir-se que os resultados obtidos em ambas as amostras estão compatíveis com o processo de análise realizado na empresa estudada.

## 5. CONSIDERAÇÕES FINAIS

A competitividade no comércio está a cada dia mais acirrada, empresas buscam maneiras de atrair clientes e fidelizá-los, de forma a garantir uma vantagem competitiva, desse modo, uma das estratégias adotadas por grandes empresas é a oferta de políticas de crédito.

A pesquisa apresentada teve como objetivo principal a criação de um modelo de previsão capaz de auxiliar na liberação de crédito, de modo que traga mais segurança e ao mesmo tempo e reduza o número de clientes inadimplentes.

O crediário próprio ofertado pela empresa varejista objeto de estudo é um importante instrumento utilizado afim de fazer que os clientes continuem fazendo compras, principalmente aqueles que não possuem outros meios de pagamento.

Considerando que o modelo utilizado apresentou um expressivo nível de acerto nas classificações, a junção da análise humana com a análise de crédito realizada por meio de um modelo algoritmo pode ser benéfica, pois permite que os resultados da análise de crédito sejam sempre consistentes com as variáveis, tornando o processo mais imparcial e com parâmetros claros para a decisão de aprovação do crédito.

Em contrapartida aos consideráveis resultados obtidos, a pesquisa possuiu algumas limitações, sendo elas, utilizou-se apenas de um algoritmo de aprendizagem para o estudo, foi coletada uma amostra reduzida de apenas 155 análises a fim de manter o sigilo da empresa.

O modelo algoritmo utilizado nesse estudo pode servir como base para o desenvolvimento de métodos alternativos de avaliação de crédito, possibilitando uma análise mais segura. Para estudos futuros recomenda-se a utilização de outras técnicas de machine learning a fim de comparar quais modelos registram melhor um melhor resultado.

Considerando que a empresa objeto de estudo possui uma média de 16.400 análises de crédito mensal, após analisar os resultados obtidos na amostragem com os 16.000 casos simulados, que obteve uma eficiência de 99% de acertos, seria viável a criação de um modelo de inteligência artificial capaz de realizar o trabalho humano com uma estimativa de acerto de 99% dos casos, considerando as variáveis apresentadas.

Por fim, para uma maior eficácia e aceitabilidade do modelo, é necessário aumentar o número de registros na base de dados, principalmente para as concorrências minoritárias. Dessa forma, será possível construir um modelo mais aprofundado e com um maior desempenho dos resultados.

## REFERÊNCIAS

ASSAF, N.; LIMA, G. Curso de administração financeira. 4. ed. São Paulo: Editora Atlas, 2023.

BARBIERI, C. BI2 - business intelligence: modelagem e qualidade. Rio de Janeiro: Editora Elsevier, 2011.

- BARROS, C.; GOMES, L. Análise das características de clientes adimplentes e inadimplentes de uma empresa do setor têxtil. *Revista Acadêmica Senac On-line*, v. 1, p. 1-10, 2006.
- BESSIS, J. *Risk Management in Banking*. Chichester, England: John Wiley, 1998.
- BLATT, A. *Avaliação de risco e decisão de crédito: um enfoque prático*. São Paulo: Editora Nobel, 1999a.
- BLATT, A. *Dicas para conceder crédito com menos risco*. São Paulo: Editora Nobel, 1999b.
- BRAGA, R. *Fundamentos e técnicas de administração financeira*. 4. ed. São Paulo: Editora Atlas, 1995.
- CAOUCETTE, J.; ALTMAN, E.; NARAYANAM, P. *Gestão do Risco de Crédito: o próximo grande desafio financeiro*. Rio de Janeiro: Editora Qualitymark, 1999.
- CASTRO NETO, J.; SÉRGIO, R. *Análise de risco e crédito*. Curitiba: Editora IESDE Brasil, 2009.
- DINIZ, C.; LOUZADA, F. *Modelagem estatística para risco de crédito*. São Paulo: Editora ABE, 2012.
- DUARTE JÚNIOR, A. *Risco: Definições, Tipos Medição e Recomendações para o seu Gerenciamento*. São Paulo, 1996.
- FERREIRA, R. *Como planejar, organizar e controlar seu dinheiro*. São Paulo: Editora Thomson IOB, 2006.
- FIORENTINI, S. *Inadimplência: como evitar e resolver*. Sebrae, 2004.
- Gil, A. C.. *Métodos e Técnicas de Pesquisa Social*. Brasil: Atlas, 2019.
- GONÇALVES, E. *Análise de risco de crédito com o uso de modelos de regressão logística, redes neurais e algoritmos genéticos*. 2005. Dissertação (Mestrado em Administração) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2005.
- LEONI, G.; LEONI, E. *Cadastro, crédito e cobrança*. 2. ed. São Paulo: Editora Atlas, 1997.
- MACHADO, M.; KARRAY, S. *Assessing credit risk of commercial customers using hybrid machine learning algorithms*. *Expert Systems with Applications*, 2022.
- MATOS, D. *Conceitos Fundamentais de Machine Learning*. *Ciência e Dados*. 22 set. 2015. Disponível em: <http://www.cienciaedados.com/conceitosfundamentais-de-machine-learning/>. Acesso em: 22 nov. 2023.
- MINAYO, M. *Pesquisa social: teoria, método e criatividade*. 29. ed. Petrópolis: Editora Vozes, 2010.
- MITCHELL, T. M. *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997.
- OLIVEIRA NETO, R. *Ciência dos dados pelo processo de Kdd*. Petrolina, 2021.
- RODRIGUES, C. *Análise de crédito e risco*. Curitiba: Editora InterSaberes, 2012.
- ROSS, S.; WESTERFIELD, R.; JAFFE, J. *Administração financeira*. 2. ed. São Paulo: Editora Atlas, 2007.
- SÁ, C. *Estabelecimento de limite de crédito: uma nova abordagem para um velho problema*. Rio de Janeiro: Editora Qualitymark, 2004.
- SANTOS, J. *Análise de crédito: Empresas, pessoas físicas, agronegócio e pecuária*. 3. ed. São Paulo: Editora Atlas, 2009.

- SANTOS, J. *Análise de Crédito: Empresas, pessoas físicas, varejo, agronegócio e pecuária*. 6. ed. São Paulo: Editora Atlas, 2015.
- SAUNDERS, A. *Medindo o risco de crédito: novas abordagens para o value at risk e outros paradigmas*. Rio de Janeiro: Editora Qualitymark, 2000.
- SCHRICKEL, W. *Análise de Crédito: Concessão e gerência de empréstimos*. 4. ed. São Paulo: Editora Atlas, 2000.
- SERASA EXPERIAN. *Análise de crédito: um guia completo para desenvolver análises precisas*. São Paulo: Serasa Experian, 2023.
- SERASA EXPERIAN. *Conheça as 7 principais causas de inadimplência no Brasil hoje*. São Paulo, 2018.
- SERASA SCORE. *Qual score é bom para crédito? Entenda o sistema de pontuação*. São Paulo, 2022.
- SILVA, J. *Análise e decisão de crédito*. São Paulo: Editora Atlas, 1997.
- SILVA, J. *Análise do estabelecimento do limite de crédito: um estudo de caso*. 2002. Dissertação (Mestrado em Administração) - Faculdade de Economia, Contabilidade, Administração e Secretariado, Universidade de Taubaté, São Paulo, 2002.
- SILVA, J. *Gestão e análise de risco de crédito*. 6. ed. São Paulo: Editora Atlas, 2008.
- TEXEIRA, S.; PAIXÃO, W. *Impacto financeiro causado por inadimplência: um estudo em uma empresa de crédito*. Revista Innovare, 27. ed. Paraná, 2019.
- TRENTO, A. *Crédito e cobrança para micro e pequenas empresas*. Guarapuava: Unicentro, 2009.