

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E NEGÓCIOS
NÍVEL MESTRADO

CRISTIANO MARTINS PIRES

RISCO DE CRÉDITO:
UMA ANÁLISE COMPARATIVA ENTRE DIFERENTES MÉTODOS DE AVALIAÇÃO

Porto Alegre

2021

CRISTIANO MARTINS PIRES

RISCO DE CRÉDITO:

UMA ANÁLISE COMPARATIVA ENTRE DIFERENTES MÉTODOS DE AVALIAÇÃO

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Gestão e Negócios, pelo Programa de Pós-Graduação em Gestão e Negócios da Universidade do Vale do Rio dos Sinos - UNISINOS

Orientador: Prof. Dr. André Filipe Zago de Azevedo

Porto Alegre

2021

P667r

Pires, Cristiano Martins

Risco de crédito : uma análise comparativa entre diferentes métodos de avaliação / por Cristiano Martins Pires. – 2021.

54 f. : il. ; 30 cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Gestão e Negócios, 2021.

Orientação: Prof. Dr. André Filipe Zago de Azevedo.

1. Risco de crédito. 2. Regressão logística. 3. *Machine Learning*. 4. *Credit Scoring*. I. Título.

CDU 336.76

Catálogo na Fonte:

Bibliotecária Vanessa Borges Nunes - CRB 10/1556

CRISTIANO MARTINS PIRES

RISCO DE CRÉDITO:
UMA ANÁLISE COMPARATIVA ENTRE DIFERENTES MÉTODOS DE AVALIAÇÃO

Aprovado em (dia) (mês) (ano)

BANCA EXAMINADORA

Prof. Dr. André Filipe Zago de Azevedo – UNISINOS

Prof. Dr. João Zani – UNISINOS

Prof. Dr. Marcelo André Machado – UNISINOS

Dr. Pedro Lutz Ramos – Sicredi

*Aos meus pais, pela vida, educação e valores,
além do costumeiro apoio;*

Aos meus irmãos, pela fiel torcida e aprendizados;

*À Vanessa, pela paciência, compreensão e
incentivo incondicional.*

AGRADECIMENTOS

O presente trabalho não poderia ser realizado sem o incentivo, compreensão, parceria e paciência de várias pessoas.

Agradeço aos meus pais e irmãos, por entenderem os momentos ausentes, por incentivarem e torcerem, além do aprendizado prévio proporcionado.

Não poderia deixar de registrar um agradecimento especial ao meu caríssimo orientador, prof. Dr. André Azevedo, que sempre acreditou, incentivou, com toda paciência e perseverança, colaborando não só com a dissertação em si, mas também com minha evolução profissional. Ainda, aos demais professores, pelos ensinamentos e provocações, em aula ou no café (enquanto foi possível), pela compreensão, flexibilidade e adaptação diante das adversidades. Por falar em adversidades e adaptações, agradeço a Unisinos e seus colaboradores, pois diante da crise sanitária em função da Covid-19, rapidamente adaptaram-se para que as aulas continuassem da melhor maneira possível.

Cabe também agradecer aos colegas do mestrado, pelo convívio, pelos compartilhamentos, pelo companheirismo e parceria ao longo dessa jornada ímpar. Aqui, cabe um agradecimento especial ao nobre Gian Capelari, nosso representante perpétuo, sempre ativo, atento, ponderado, objetivo e empático, auxiliando a todos em diversas questões, estabelecendo uma relação de confiança mútua entre a turma e a coordenação do curso e professores.

Por fim, agradeço também aos colegas (Sicredi e Getnet) e amigos pela compreensão, flexibilidade e incentivo contribuindo também para que fosse possível a realização deste trabalho.

RESUMO

O presente trabalho tem como objetivo analisar diferentes métodos de avaliação de risco de crédito, considerando as técnicas de Regressão Logística, amplamente usada para avaliação de risco de crédito, e também de *Machine Learning*, que vem sendo explorada para diversas finalidades, inclusive no âmbito de *Credit Scoring*. A importância do tema é evidente considerando que o sistema econômico depende do crédito para promover a expansão do consumo e do crescimento econômico. Para atender ao objetivo, foi realizada uma pesquisa bibliográfica, de natureza qualitativa. Como constatações, ressalta-se que não há posição definitiva sobre qual método de avaliação de risco de crédito é o melhor a ser utilizado, tendo diferentes fatores para que a escolha seja feita. Fato é que, ambos demonstram nível de acurácia satisfatório, apresentando vantagens e desvantagens, dependendo do contexto que o problema está inserido. Outro ponto relevante, específico do crédito na economia brasileira, é de que o nível de concessão demonstra expansão ao longo do tempo, mantendo ou até reduzindo o nível de inadimplência, dependendo do período analisado. Isso demonstra que as técnicas e processos utilizados pelo sistema financeiro estão acompanhando essa evolução. Como contribuições, o trabalho faz uma revisão ampla sobre o tema, desde o papel do crédito na economia e sua evolução recente no contexto brasileiro, bem como as técnicas mais utilizadas para *Credit Scoring*. Esse levantamento contribui tanto para o meio acadêmico, estabelecendo um consolidado sintético sobre o tema, quanto para o meio profissional, para profissionais da área de risco de crédito e afins.

Palavras-chave: Risco de Crédito. Regressão Logística. *Machine Learning*. *Credit Scoring*.

ABSTRACT

This paper analyzes the different methods of credit risk assessment, considering the techniques of Logistic Regression, widely used for credit risk scoring, and also Machine Learning, which has been explored for various purposes, including in the context of credit scoring. The importance of the theme is evident considering that the economic system depends on credit to promote the expansion of consumption and economic growth. To reach the objective, bibliographic research of a qualitative nature was used in the study. As findings, it is noteworthy that there is no definitive position on which method of credit risk assessment is the best to be used, having different factors for the choice to be made. The fact is that both demonstrate a satisfactory level of accuracy, presenting advantages and disadvantages, depending on the context in which the problem is inserted. Another relevant point, specific to credit in the Brazilian economy, is the level of credit concession that shows expansion over time, maintaining or even reducing the level of default, depending on the period analyzed. This shows that the techniques and processes used by the financial system are following this evolution. As contributions, the paper provides a broad review of the subject, from the role of credit in the economy and its recent evolution in the Brazilian context, to the most used techniques for credit scoring. This research contributes both to the academic environment, by establishing a synthetic consolidation on the topic, and to the professional environment, for professionals in the business of credit risk and related areas.

Key-words: Credit Risk. Logistical Regression. Machine Learning. Credit Scoring.

LISTA DE FIGURAS

Figura 1 – Entidades e Instituições do SFN	17
--	----

LISTA DE GRÁFICOS

Gráfico 1 – Evolução do Saldo da Carteira de Crédito em relação ao PIB (%)	19
Gráfico 2 – Evolução do Saldo de Crédito PF e PJ em relação ao PIB (%)	20
Gráfico 3 – Evolução do Endividamento das Famílias em relação à Renda (%).....	21
Gráfico 4 – Evolução da Inadimplência da Carteira de Crédito PF (%)	22
Gráfico 5 – Evolução da Inadimplência da Carteira de Crédito PJ (%)	23
Gráfico 6 – Evolução do Endividamento Familiar – Peic (%)	24
Gráfico 7 – Evolução do Endividamento Familiar (até 10 salários mínimos) – Peic (%)	25
Gráfico 8 – Evolução do Endividamento Familiar (mais de 10 salários mínimos) – Peic (%)	26

LISTA DE QUADROS

Quadro 1 – Síntese dos Métodos de Avaliação.....	45
--	----

LISTA DE SIGLAS

Bacen	Banco Central do Brasil
Banrisul	Banco do Estado do Rio Grande do Sul
BNDES	Banco Nacional de Desenvolvimento Econômico e Social
CNC	Confederação Nacional do Comércio de Bens, Serviços e Turismo
Peic	Pesquisa de Endividamento e Inadimplência do Consumidor
PF	Pessoa Física
PIB	Produto Interno Bruto
SENAC	Serviço Nacional de Aprendizagem Comercial
SESC	Serviço Social do Comércio
SFN	Sistema Financeiro Nacional
Sicredi	Sistema de Crédito Cooperativo

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Problema	13
1.2 Objetivos	14
1.2.1 Objetivo Geral.....	14
1.2.2 Objetivos Específicos	14
1.3 Justificativa	15
2 CRÉDITO NO BRASIL: EVOLUÇÃO RECENTE E IMPORTÂNCIA ECONÔMICA	16
2.1 Sistema Financeiro Nacional	16
2.2 Evolução do Crédito e da Inadimplência no Brasil	18
2.3 O Papel do Crédito	27
3 METODOLOGIA	29
3.1 Classificação da Pesquisa	29
4 ANÁLISE DE RISCO DE CRÉDITO	31
4.1 <i>Credit Scoring</i>	34
4.1.1 Regressão Logística.....	36
4.1.2 <i>Machine Learning</i>	39
4.2 Análise Comparativa	43
5. CONCLUSÕES	47
REFERÊNCIAS	49

1 INTRODUÇÃO

A expansão de crédito no Brasil se deu de forma acentuada a partir de 2005. No final da década de 1990, a relação do saldo das operações de crédito sobre o Produto Interno Bruto (PIB) situava-se abaixo de 30%. Em março de 2004, atingiu 24% nesta métrica. Entretanto, a partir de 2005, por meio de estímulos do setor público, dentre outros aspectos, esse nível começou a aumentar de forma acentuada: no início de 2007, já havia rompido os 30% e na metade de 2009 já se situava acima de 40% do PIB. Considerando um período mais recente, esse patamar tem se mantido acima de 45% do PIB desde o final de 2011, atingindo 53% em maio de 2021 (BACEN, 2021).

Mesmo com essa expansão do crédito, tanto em termos absolutos quanto relativos, e o conseqüente aumento do endividamento das famílias, a inadimplência, em linhas gerais, apresenta declínio ao longo dos anos, considerando o período de janeiro de 2012 a maio de 2021, ainda que observadas oscilações. Diante disso, pode-se inferir que os critérios utilizados ao conceder crédito estão adequados, não gerando situações de risco aumentado. Para que isso ocorra, as instituições financeiras têm seu mérito nesse processo: eficiência tanto na concessão, quanto na recuperação e na cobrança dos empréstimos e financiamentos.

A literatura aponta diferentes métodos para avaliar o risco de concessão de crédito, baseados no *Credit Score*, com destaque para os Modelos Logísticos e *Machine Learning*. De acordo com Lessmann et al. (2015), a regressão logística é considerada a técnica padrão quando o tema é *Credit Scoring*. A variável dependente é a inadimplência, que é o evento que se quer explicar, enquanto as variáveis independentes, que determinam a inadimplência, são muitas e variam conforme o tipo de pessoa (física ou jurídica). O *Machine Learning*, por sua vez, consiste em um conjunto de procedimentos e regras que possibilitam que decisões sejam tomadas de forma automática com base em dados. Dessa forma, o computador deixa de ser um simples executor de algo programado anteriormente, de forma estática. Conforme Aniceto (2016), os algoritmos de *Machine Learning* não exigem um conhecimento tão aprofundado das variáveis a serem utilizadas e suas respectivas relações, o que difere das técnicas estatísticas tradicionais, como a regressão logística.

A principal contribuição deste trabalho é a realização de uma revisão da literatura sobre o tema, desde o papel do crédito na economia e sua evolução recente

no contexto brasileiro, mas principalmente das técnicas mais utilizadas para *Credit Scoring*. Esse levantamento contribui tanto para o meio acadêmico, estabelecendo um consolidado sintético sobre o tema, quanto para o meio profissional, para profissionais da área de risco de crédito e afins.

O presente trabalho está dividido em quatro capítulos, além desta introdução. No segundo capítulo, aborda-se a evolução recente do crédito no país, contemplando também a sua importância para a economia. As questões metodológicas estão descritas no terceiro capítulo. No quarto capítulo, os métodos de avaliação de risco de crédito são apresentados, explorando estudos anteriores sobre o objeto do presente trabalho. Por fim, as conclusões, finalizando o estudo, constam no quinto capítulo, onde também são abordadas oportunidades futuras de pesquisa.

1.1 Problema

O foco dado pelo Banco Central do Brasil (BACEN) tem como base a crescente população tomadora de crédito (os endividados), bem como uma atenção aos níveis de inadimplência (atraso nos pagamentos), geralmente motivo de preocupação às pessoas que se encontram nesta situação.

Após o Plano Real, o saldo da carteira de crédito em relação ao PIB caiu de 37% até aproximadamente 24% em março 2004. Depois, houve uma expansão muito acentuada e consistente até 2015, ultrapassando 50% do PIB, ainda que as pessoas jurídicas tenham perdido espaço nesse indicador, sendo compensado pelo aumento da participação de pessoas físicas no saldo de crédito. Mudanças na legislação contribuíram de forma significativa para a expansão do crédito após 2004, em virtude de evoluções nos processos de garantias, tais como a consignação em folha de pagamento e alienação fiduciária.

A Confederação Nacional do Comércio de Bens, Serviços e Turismo (CNC)¹ efetua, periodicamente, a Pesquisa de Endividamento e Inadimplência do Consumidor (Peic). O intuito da pesquisa é verificar, por meio de captura de percepção dos indivíduos, o nível de endividamento das famílias, o tipo de dívida, as dívidas em atraso, as condições e o prazo de pagamento deste atraso, o tempo médio do comprometimento das famílias endividadas, além do comprometimento da renda

¹ Integrante do Sistema CNC - SESC (Serviço Social do Comércio) - SENAC (Serviço Nacional de Aprendizagem Comercial).

familiar com dívida. Por meio das séries históricas da Peic, observa-se que nos últimos 5 anos houve um aumento no número de famílias endividadas, com destaque para o cartão de crédito, o que denota um possível problema, haja vista que o custo financeiro deste produto é um dos mais elevados, em caso de atraso ou prorrogação de valores. Quando se analisa as dívidas em atraso, também percebe-se uma elevação no período examinado, sendo o mesmo movimento de alta para famílias que acusam incapacidade de pagar estas dívidas em atraso.

Considerando o endividamento das famílias, sob a ótica do BACEN, demonstra crescimento, com participação significativa do crédito habitacional. Porém, mesmo diante da expansão do crédito e do endividamento, o nível de inadimplência se reduziu, de forma geral, ainda que com oscilações maiores entre 2015 e 2018.

Sendo assim, para compreender melhores mecanismos utilizados pelo sistema financeiro para administrar o crédito, eis a questão de pesquisa: quais são as principais técnicas de análise de risco de crédito e suas características?

1.2 Objetivos

1.2.1 Objetivo Geral

Este trabalho pretende analisar as principais técnicas de análise de risco de crédito, considerando aspectos como: dados disponíveis, implantação e manutenção, além de acurácia.

1.2.2 Objetivos Específicos

Para que o objetivo geral seja atingido, devem ser observados os seguintes objetivos específicos:

- a) Descrever o Sistema Financeiro Nacional (SFN), bem como suas atribuições;
- b) Analisar a evolução recente do saldo de operações de crédito em relação ao PIB e do endividamento das famílias e sua inadimplência no Brasil;
- c) Evidenciar a importância social e econômica do crédito;

- d) Analisar os métodos de Regressão Logística e *Machine Learning*, destacando suas principais vantagens e examinando as contribuições mais relevantes sobre o tema.

1.3 Justificativa

O presente trabalho aborda um tema relevante para a economia do país e, por conseguinte, para a sociedade como um todo. Conforme exposto na subseção anterior, e aprofundado ao longo do estudo, o crédito vem aumentando sua participação no SFN, considerando o nível de representatividade em relação ao PIB brasileiro. Por outro lado, apesar da expansão do volume das operações de crédito no Brasil, percebe-se uma relativa estabilidade da inadimplência. Ressalta-se também a importância do crédito para o sistema econômico, conforme *Schumpeter, Kalecki* e demais autores citados no segundo capítulo. Dessa forma, analisar as razões deste fenômeno torna-se relevante para verificar com maior clareza a evolução e as oportunidades nesse segmento.

Outro aspecto importante é que, quanto melhor avaliado o risco de crédito pelas instituições, menor tenderá a ser a inadimplência e as despesas decorrentes, auxiliando na redução do custo de crédito, impactando positivamente no poder de compra e/ou de investimentos de famílias/consumidores e de empresas.

Ainda, esta pesquisa, sob a ótica profissional, faz sentido frente à trajetória e posição ocupada pelo pesquisador. Com atuação predominante em crédito e riscos há mais de 15 anos, sendo responsável por uma das áreas de riscos financeiros de uma empresa relevante no sistema financeiro, investigar o tema tem potencial impacto na gestão dos riscos da companhia.

2 CRÉDITO NO BRASIL: EVOLUÇÃO RECENTE E IMPORTÂNCIA ECONÔMICA

O presente capítulo está dividido em três partes. A primeira apresenta, brevemente, o Sistema Financeiro Nacional (SFN), enquanto a segunda expõe a evolução recente do crédito, do endividamento das famílias e da inadimplência no Brasil. Por fim, a importância do crédito na dinâmica socioeconômica, por meio de uma revisão de literatura sobre o tema, é exibida ao final do capítulo.

2.1 Sistema Financeiro Nacional

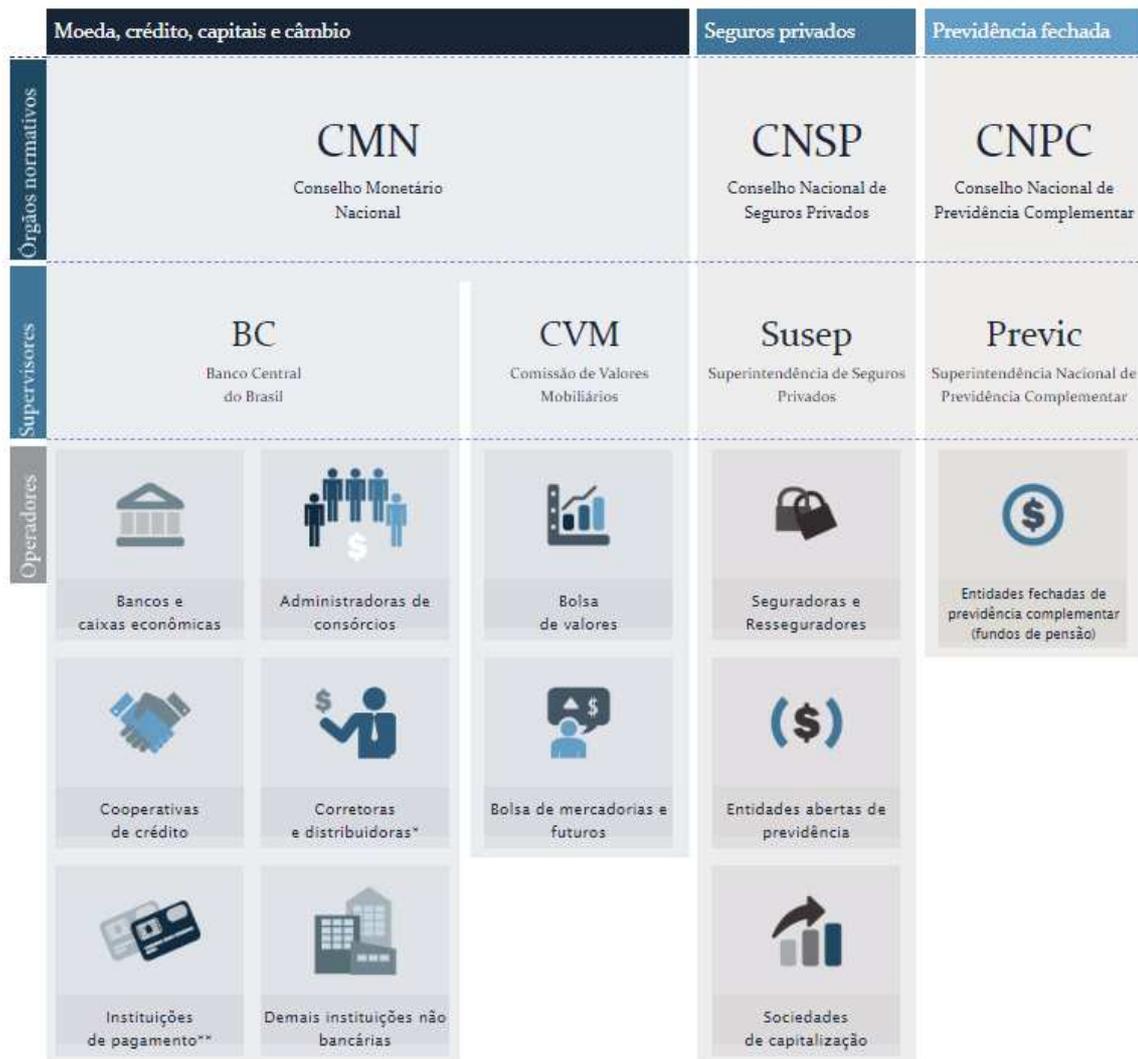
O SFN é formado por um conjunto de instituições e entidades que possuem o objetivo de promover a intermediação financeira, conceito que será abordado também na próxima seção. Os recursos financeiros das empresas, do governo e das pessoas circulam, de forma preponderante, por meio deste sistema, considerando, inclusive, pagamentos de dívidas, além de realização de investimentos. A principal finalidade do sistema financeiro é manter o fluxo de recursos entre agentes superavitários, detentores de capital excedente, e agentes deficitários, necessitados de capital de terceiros para desenvolver suas atividades. De acordo com Levine (1997), a existência de um sistema financeiro bem organizado é de suma importância para o desenvolvimento econômico de qualquer país. O tamanho do setor bancário e o nível de liquidez do mercado de capitais tem correlação direta com o crescimento do PIB *per capita* de um país.

Quanto à sua composição, destaca-se o Conselho Monetário Nacional (CMN), órgão normativo no setor de Moeda, Crédito, Capitais e Câmbio, o BACEN, agente supervisor, os bancos, caixas econômicas e cooperativas de crédito, agentes operadores do SFN. Há outros órgãos e entidades, conforme Figura 1.

De acordo com a CVM (2020), as Entidades Normativas são governamentais e têm a obrigação de definir as políticas de governo e transformá-las em regulamentos, sem função executiva. Esse papel é desempenhado pelo CMN, Conselho Nacional de Seguros Privado (CNSP) e Conselho Nacional de Previdência Complementar (CNPC). Cabe às Entidades Supervisoras assumirem as funções executivas de fiscalizar e orientar as demais instituições integrantes do SFN de forma que as políticas e as normas determinadas pelo governo sejam cumpridas. São consideradas Entidades Supervisoras no Brasil: o BACEN, a Comissão de Valores Mobiliários (CVM), a

Superintendência de Seguros Privados (SUSEP) e a Superintendência Nacional de Previdência Complementar (PREVIC). Existem também as Entidades Operadoras, que são todas as demais instituições financeiras integrantes do SFN, a quem reserva-se o papel de atuar no mercado financeiro com o objetivo de transferir recursos dos agentes superavitários para os deficitários, desde que respeitadas as normas emitidas pelas Entidades Normativas sob a supervisão das Entidades Supervisoras.

Figura 1 – Entidades e Instituições do SFN



* Dependendo de suas atividades corretoras e distribuidoras também são fiscalizadas pela CVM.

** As Instituições de Pagamento não compõem o SFN, mas são reguladas e fiscalizadas pelo BCB, conforme diretrizes estabelecidas pelo CMN.

Fonte: Banco Central do Brasil.

O CMN é um dos órgãos mais importantes de todo o SFN. Sua representatividade é tão grande que ele se responsabiliza, por exemplo, por ditar as políticas e diretrizes monetárias utilizadas por todo o mercado financeiro. Cabe a ele,

também, a tomada de decisões sobre as políticas da moeda e do crédito no país. Estão diretamente ligados a ele o BACEN, que opera como seu órgão executivo e a CVM, responsável pela regulação e estímulo do mercado de valores mobiliários (bolsa de valores e de balcão).

2.2 Evolução do Crédito e da Inadimplência no Brasil

Por meio das informações disponibilizadas pelo BACEN, é possível avaliar, por diversas óticas, a evolução recente do crédito, bem como do endividamento, da inadimplência² e do atraso³. Conforme percebe-se no Gráfico 1, houve uma expansão significativa do crédito no Brasil, no período recente, embora com alguma oscilação. A representatividade do crédito na economia aumentou nas últimas décadas, especialmente entre 2004 e 2015, inicialmente impulsionada pelos bancos privados, no tocante a recursos livres, mas em um segundo momento pelos bancos públicos, no âmbito de recursos direcionados, contribuindo com seu papel contracíclico, ante a crise financeira de 2008 (BACEN, 2013).

Essa expansão deu-se por diversos fatores, dentre eles: mudanças na legislação no que tange às garantias de crédito, como a alienação fiduciária (Lei n.º 9.514/1997), inclusive para financiamento imobiliário⁴, deixando de ser utilizada a hipoteca, gradativamente; a regulamentação do crédito consignado (Lei n.º 10.820/2003); e, ampliação dos prazos nas operações de empréstimo e financiamento.

Os dois primeiros fatores, utilização de alienação fiduciária e a consignação em folha de pagamento, proporcionaram redução no risco de crédito, aumentando assim o apetite das instituições financeiras nessas modalidades, bem como expandindo a demanda, gerando um aumento significativo da carteira de crédito, impulsionado pelo crédito habitacional e também consignado, tanto público, quanto privado. Quanto ao prazo aumentado, flexibiliza a capacidade de pagamento dos clientes, também favorecendo ao aumento da carteira de crédito brasileira (MORA, 2015).

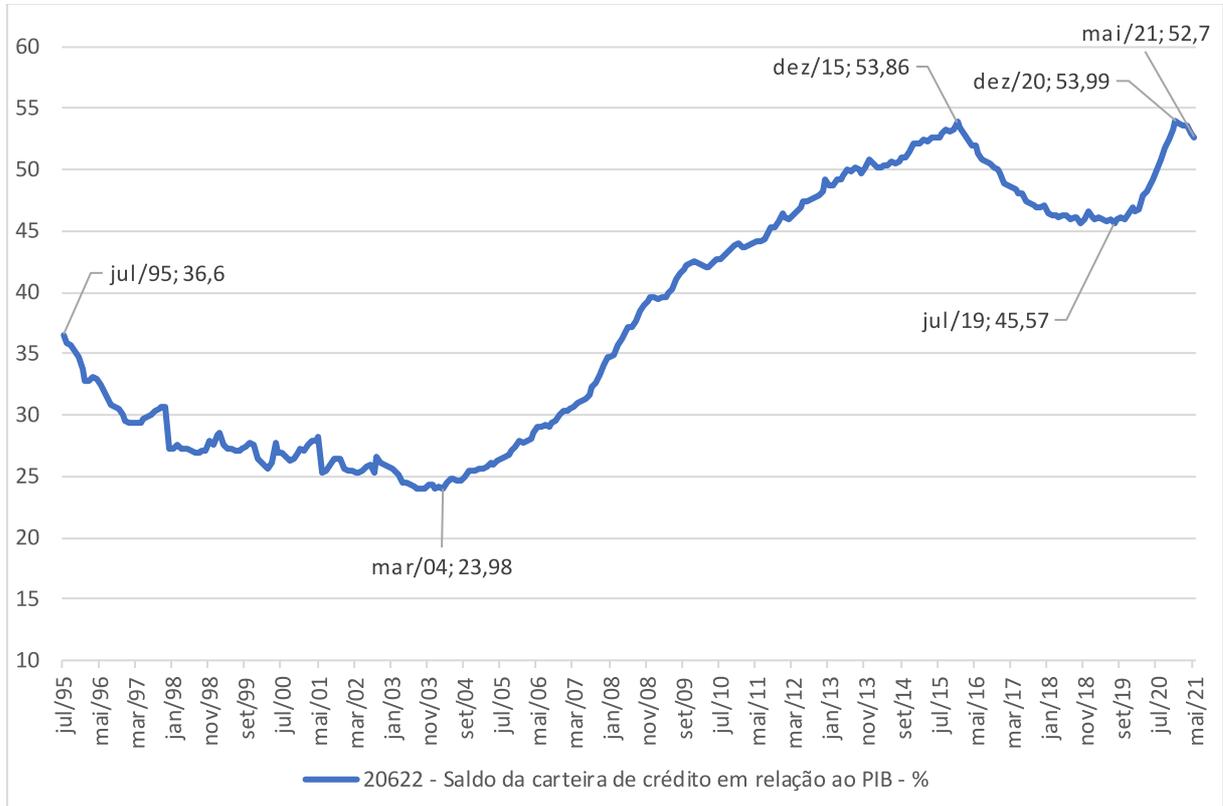
² Conceito estabelecido pelo BACEN, inadimplência se refere ao atraso no pagamento superior a 90 dias.

³ Conceito estabelecido pelo BACEN, se refere ao atraso no pagamento de 15 a 90 dias.

⁴ Observando o gráfico 3, nota-se que a diferença entre os percentuais de endividamento das famílias excluindo-se crédito habitacional e considerando o endividamento total, era de 10,9 p.p em janeiro de 2012, ante aos 22,3 p.p. em maio de 2021, denotando o aumento do crédito habitacional.

Importante ressaltar mais um aspecto: o fim do período hiper-inflacionário, que permitiu o sistema financeiro e suas instituições atuarem com taxas de juros pré-fixadas.

Gráfico 1 – Evolução do Saldo da Carteira de Crédito em relação ao PIB (%)



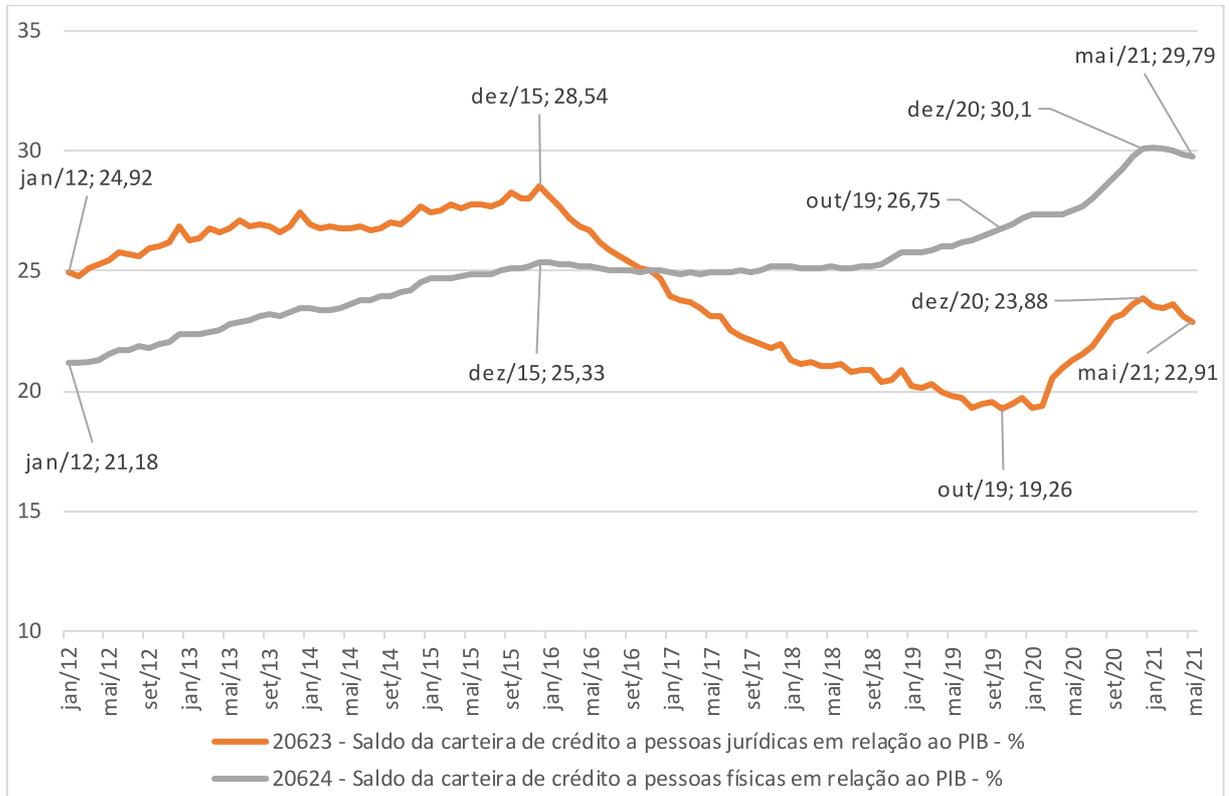
Fonte: BACEN – SGS - Sistema Gerenciador de Séries Temporais.

A sustentação do crédito durante 2020, frente à crise causada pela pandemia de Covid-19, deu-se, em grande parte, pelo papel do governo e regulação do BACEN, flexibilizando para alongamento de dívidas e carência de pagamentos (prorrogações), sem penalizar as instituições financeiras. Com isso, o nível de crédito em relação ao PIB ainda subiu, considerando que o crédito ainda crescia discretamente, e o PIB caía (BACEN, 2020).

Considerando o período de março de 2007 até dezembro de 2019, observa-se uma inversão na representatividade, sob a ótica da natureza de devedores, conforme ilustrado no Gráfico 2. No passado, pessoas jurídicas (PJ) possuíam maior participação na carteira de crédito em relação ao PIB. Entretanto, é possível perceber uma tendência de queda da participação de pessoas jurídicas nas operações de crédito entre o final de 2015 e o final de 2019. Como resultado, houve uma inversão da participação das pessoas físicas (PF) e jurídicas nas operações de crédito, com a

participação das pessoas físicas ultrapassando à das pessoas jurídicas, configurando uma diferença significativa até maio de 2021 (30% das PF *versus* 23% das PJ).

Gráfico 2 – Evolução do Saldo de Crédito PF e PJ em relação ao PIB (%)



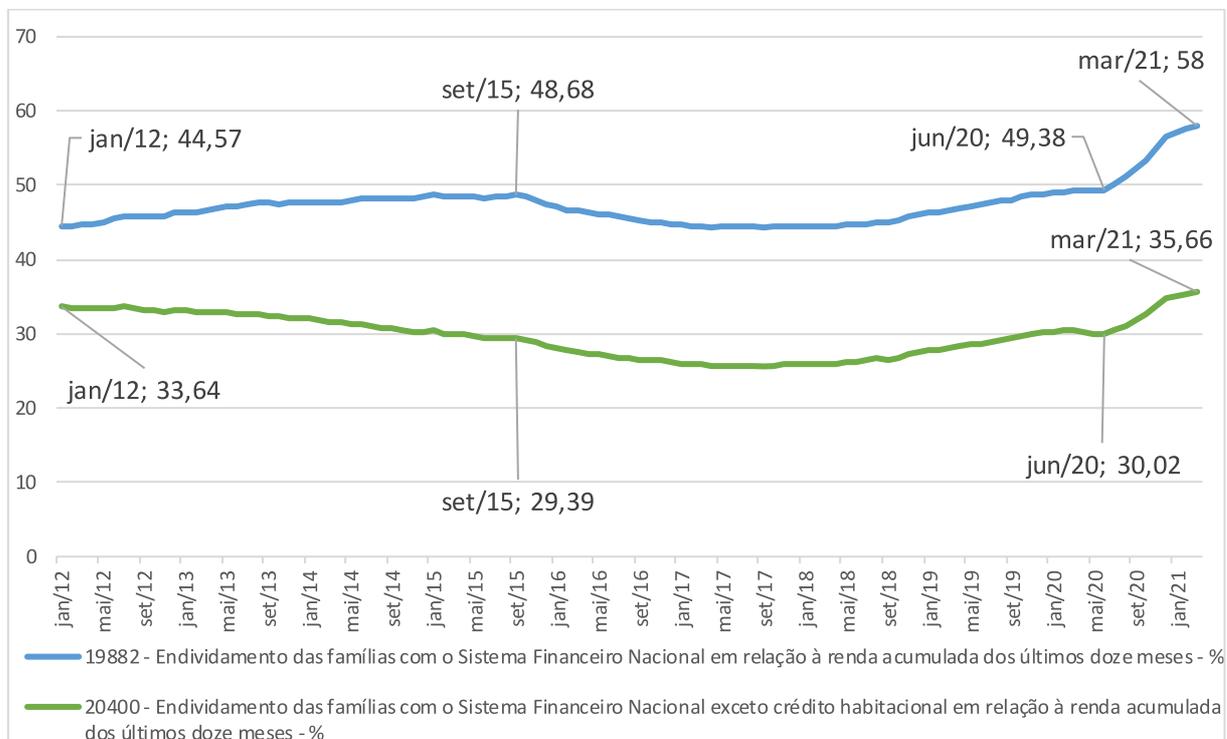
Fonte: Bacen – SGS - Sistema Gerenciador de Séries Temporais.

Essa redução da participação de pessoas jurídicas nas operações de crédito deu-se pelo aumento da incerteza em função da recessão no período 2015/2016, bem como redução dos desembolsos do Banco Nacional de Desenvolvimento Econômico e Social (BNDES), que vinha sendo o maior propulsor do crédito às empresas, através de recursos direcionados (BACEN, 2016).

Quando a ótica se detém mais especificamente sobre o endividamento das famílias, observa-se que, nos últimos anos, excetuando-se pelo nível e evolução nos anos de 2015 e 2016, apresenta-se uma curva ascendente. Em janeiro de 2012, como pode ser observado no Gráfico 3, o endividamento das famílias no SFN em relação à renda acumulada dos últimos 12 meses era de 45%, ao passo que, em março de 2021, chegou a 58%. Ressalta-se que, desconsiderando dívidas oriundas de crédito habitacional, a inclinação da curva é mais suave, além, obviamente, de variar em patamares inferiores: 34% em janeiro de 2012, para 36% em março de 2021. Ou seja,

como era de se esperar, o aumento do crédito na economia levou a uma ampliação do endividamento das famílias no período recente.

Gráfico 3 – Evolução do Endividamento das Famílias em relação à Renda (%)



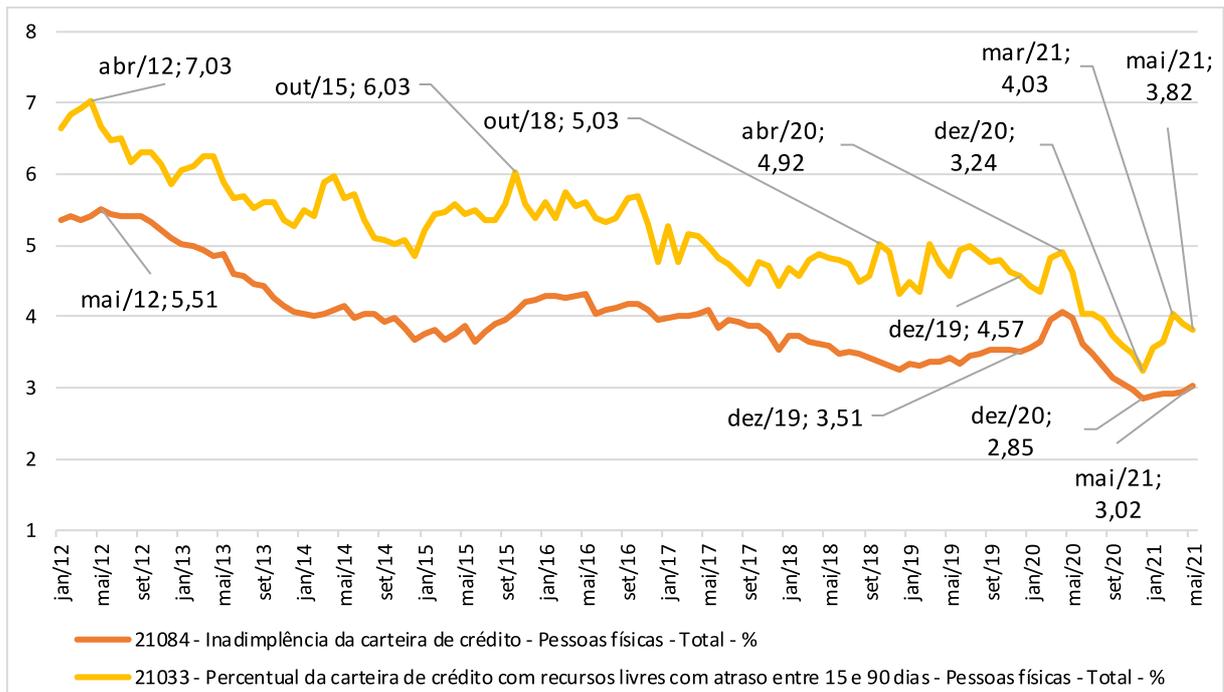
Fonte: Bacen – SGS - Sistema Gerenciador de Séries Temporais.

Ao analisar a evolução da inadimplência e do atraso de pessoas físicas, considerando o período de janeiro de 2012 em diante, percebe-se que os níveis mais elevados destes indicadores se situam nos primeiros meses de 2012.⁵ De acordo com o Gráfico 4, o auge da série, no mês de abril, atinge 7% para o nível de atraso e, no mês de maio, chega a 5% para a inadimplência. Entretanto, apresenta-se uma redução ao longo do período, tanto para o nível de inadimplência, quanto para o de atraso, atingindo, em dezembro de 2019, 3 % e 5%, respectivamente. A redução durante 2020, mesmo diante de um cenário de crise em virtude da pandemia (Covid-19) deve-se, basicamente, a dois fatores: manutenção do nível de concessão de crédito e prorrogações efetuadas pelas instituições financeiras ante a flexibilização

⁵ É importante que se tenha a distinção entre os significados de endividamento e de inadimplência. Há real necessidade de conceituar ambos os termos, pois não é rara a confusão feita pela população em geral, considerando de diferentes formas os conceitos de endividamento e inadimplência, conforme pesquisas efetuadas pela Confederação Nacional do Comércio de Bens, Serviços e Turismo (CNC). Enquanto o endividamento representa o quanto o indivíduo está comprometido em dívidas ante sua renda, a inadimplência significa que há atraso em pelo menos algum pagamento de dívida.

promovida pelo BACEN⁶. No primeiro trimestre de 2021, o nível de atraso volta a subir atingindo 4% e a inadimplência 3% no mês de maio de 2021. No entanto, apesar da recente elevação, os valores estão abaixo dos patamares observados na maior parte do período 2012-2021.

Gráfico 4 – Evolução da Inadimplência da Carteira de Crédito PF (%)



Fonte: Bacen – SGS - Sistema Gerenciador de Séries Temporais.

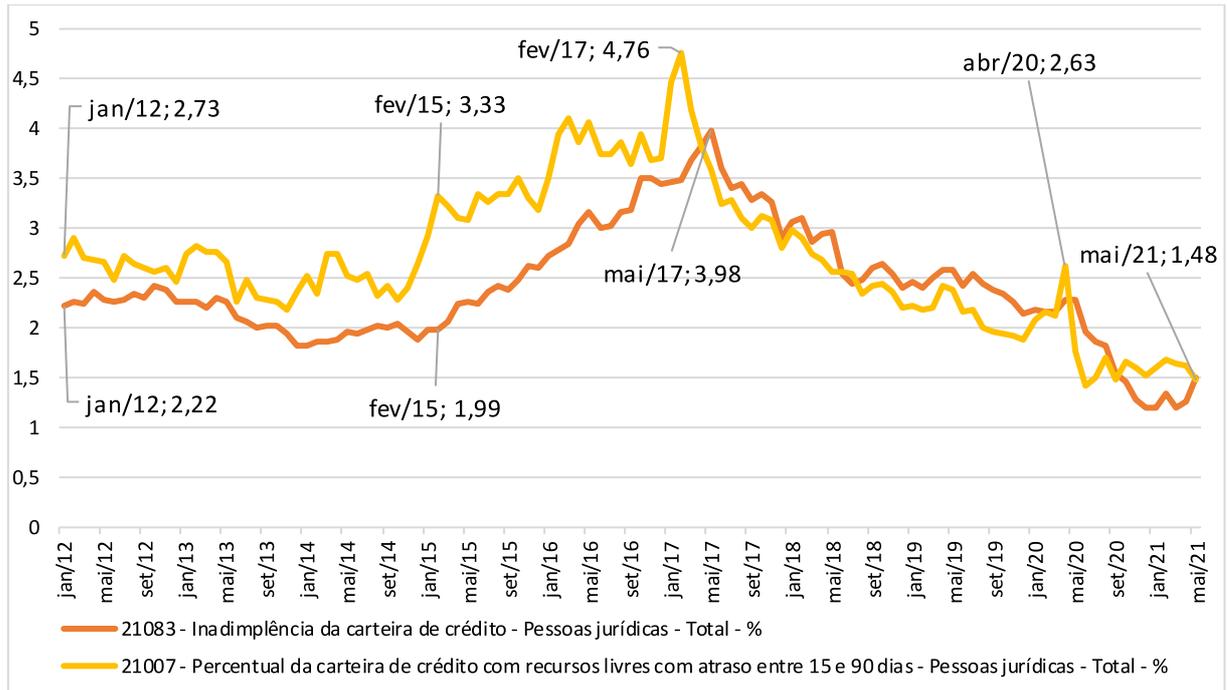
Considerando a evolução da inadimplência de pessoas jurídicas, a série histórica demonstra três tendências entre janeiro de 2012 e maio de 2021, como pode ser observado no Gráfico 5. Inicialmente, entre 2012 e o início de 2015, há certa estabilidade de ambos os indicadores. Mas, a partir do início de 2015 até o início de 2017, os indicadores mostram uma tendência de elevação, com a inadimplência atingindo o pico de 3,98% em maio de 2017 e o atraso chega a 4,76% em fevereiro de 2017. Já a partir de março de 2017, os índices de atraso e inadimplência caem substancialmente, atingindo seus pontos de mínimo no início de 2021.

Conforme já mencionado, em virtude da pandemia de Covid-19, deflagrada a partir de março de 2020, o BACEN flexibilizou regras quanto ao provisionamento de

⁶ Resolução nº 4.782 (março de 2020), que dispõe sobre critérios para reestruturações de operações de crédito, resolução nº 4.803 (abril de 2020), que estabelece regras para mensuração da provisão para créditos de liquidação duvidosa das operações reestruturadas em função da pandemia da Covid-19, ambas emitidas pelo BACEN.

crédito para prorrogações de dívidas para que os devedores, e também credores, sofressem menos com a crise econômica causada pela crise sanitária. Esse é um dos motivos da redução dos patamares da inadimplência e do atraso, também para pessoas jurídicas a partir de 2020, mesmo em um cenário socioeconômico adverso (BACEN, 2020).

Gráfico 5 – Evolução da Inadimplência da Carteira de Crédito PJ (%)



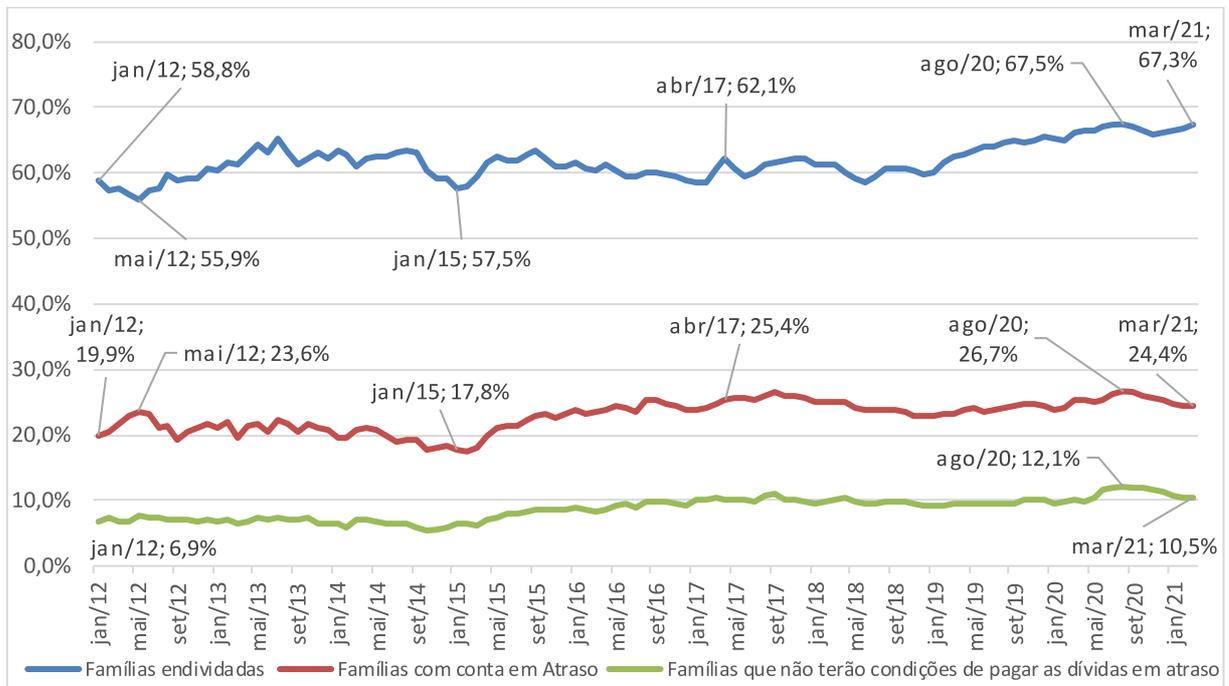
Fonte: Bacen – SGS - Sistema Gerenciador de Séries Temporais.

A seguir, aborda-se a Pesquisa de Endividamento e Inadimplência do Consumidor (Peic), como complemento às análises dos dados do BACEN apresentados anteriormente, no intuito de capturar a percepção dos indivíduos frente as dívidas. Ou seja, quando se analisam as informações extraídas do BACEN, são dados que o sistema financeiro (instituições supervisionadas pelo órgão regulador) informa. Quando se analisam as informações extraídas da Peic, são informações obtidas por meio de pesquisas junto à sociedade.

Considerando a série histórica, de janeiro de 2012 a maio de 2021, da Peic, pesquisa mencionada na seção introdutória do presente estudo, que visa coletar a percepção dos indivíduos sobre sua situação financeira, percebe-se que o percentual de famílias endividadas subiu de 58,8% para 67,3%, conforme demonstrado no gráfico 6. Entretanto, no mesmo período, sob a ótica das contas em atraso, observa-se uma elevação de 20% para 24%. Por fim, o percentual de famílias que acusa incapacidade

de honrar estes compromissos, oscilou entre de 7% e 12%, considerando o mesmo período.

Gráfico 6 – Evolução do Endividamento Familiar – Peic (%)

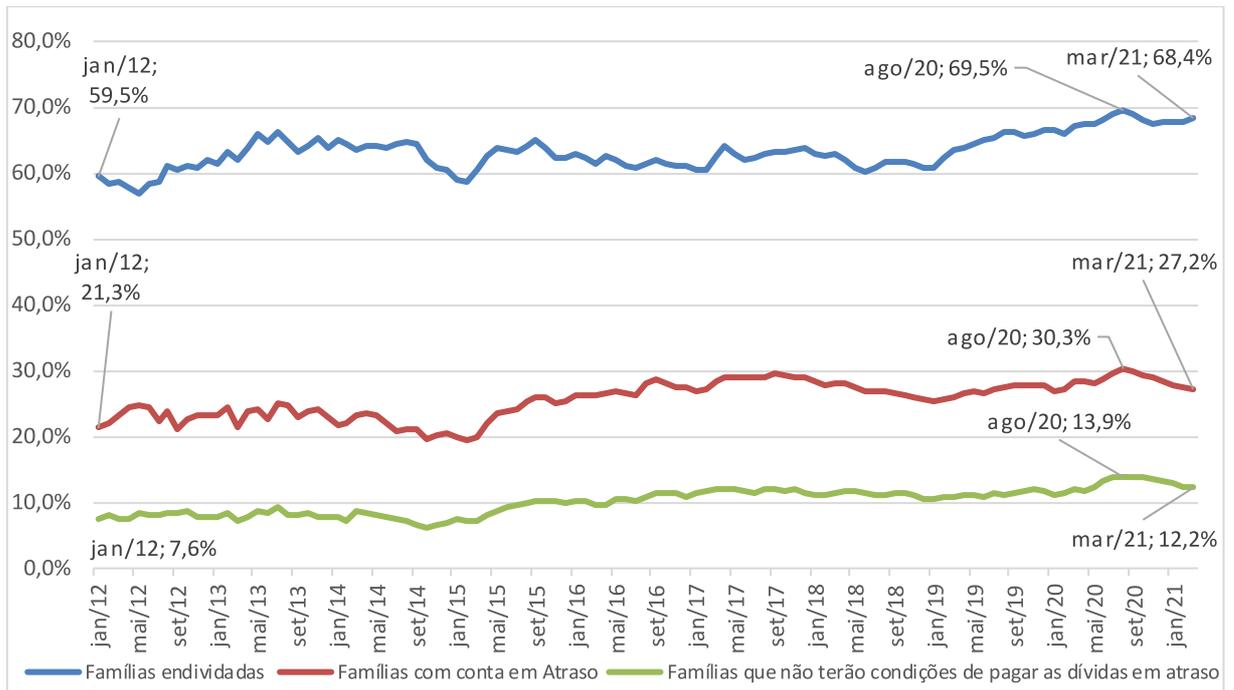


Fonte: CNC – Peic.

Considerando a mesma série histórica, entre janeiro de 2012 e março de 2021, mas apenas famílias com rendimentos de até 10 salários mínimos, percebe-se um aumento de 59% para 67%, no tocante ao percentual de famílias endividadas⁷ como pode ser observado no Gráfico 7. Sob a ótica de atraso, houve aumento: de 21% para 27%. Ainda, o percentual de famílias que declaram não ter condições de honrar estas dívidas estava em 8% em janeiro de 2012, passando para 12%, em março de 2021, refletindo a incapacidade de pagamento de suas dívidas em atraso, devido à pandemia.

⁷ Além do Cartão de Crédito, houve ligeiro incremento no percentual do Cheque Especial, ambos tipos de produto com taxas de juros bem elevadas). Cabe evidenciar a expansão de dívida oriunda de crédito habitacional.

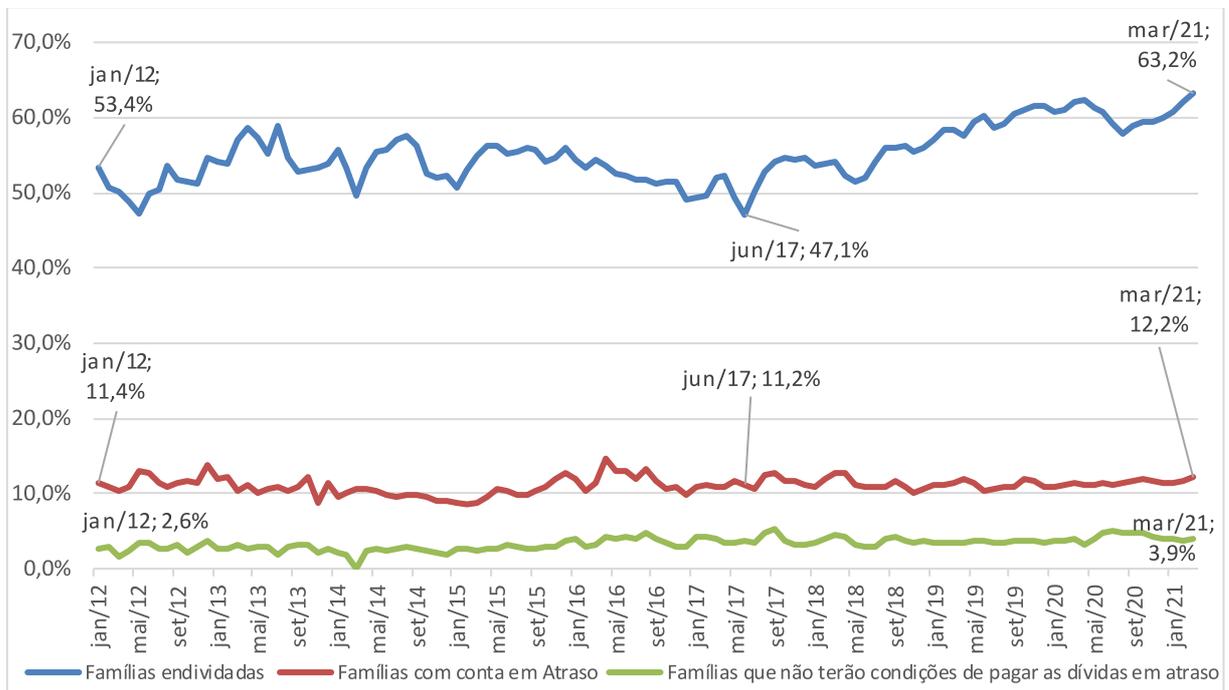
Gráfico 7 – Evolução do Endividamento Familiar (até 10 salários mínimos) – Peic (%)



Fonte: CNC – Peic.

Observando o Gráfico 8, os comportamentos das variáveis são semelhantes quando se avalia as famílias com rendimentos acima de 10 salários mínimos, exceto pelo aumento do percentual de dívida em Cheque Especial e aumento ainda mais significativo de financiamento imobiliário nesta classe. Entretanto, nota-se uma variação muito mais discreta nos níveis de inadimplência e incapacidade de pagamento nesse perfil de renda.

Gráfico 8 – Evolução do Endividamento Familiar (mais de 10 salários mínimos) – Peic (%)



Fonte: CNC – Peic.

Ainda que se observe um aumento das famílias endividadas, a inadimplência não demonstra expansão. Ao contrário, desde o início da série, em 2012, para a PF e desde o início de 2017 para a PJ, há uma tendência de queda da inadimplência e do atraso nos pagamentos. Isto é comprovado pelas informações disponibilizadas pelo BACEN, expostos no presente estudo, apesar da acentuada expansão do crédito ao longo do período examinado. Entretanto, quando observadas as séries históricas da Peic, os indicadores de atraso e incapacidade de pagamento das contas em atraso, demonstram elevação, principalmente nas famílias com renda de até 10 salários-mínimos, responsável pela elevação dos indicadores gerais (sem segmentação de renda). Importante ressaltar que a Peic reflete a percepção dos entrevistados, podendo conter vieses nas respostas, diferente dos dados obtidos junto ao BACEN.

Em suma, há elevação nos níveis de endividamento ao longo do tempo, seja na percepção das pessoas ante a sua situação financeira, seja através de uma análise nas séries históricas do BACEN. No entanto, houve declínio da inadimplência nesse período, considerando as informações obtidas junto ao BACEN, tendo a Peic demonstrando uma leve elevação nos níveis de atraso e incapacidade de pagamento.

2.3 O Papel do Crédito

A maioria das instituições financeiras, integrantes do SFN, tem como principal função orquestrar o processo de intermediação financeira. Esta atividade é de suma importância para o sistema capitalista, facilitando, inclusive, a circulação do dinheiro na economia. Em uma definição mais simples, nesse ambiente, há agentes superavitários, ou seja, possuem recurso excedente, e existem também agentes deficitários, ou seja, possuem insuficiência de recursos. A intermediação financeira se dá quando esse recurso excedente oriundo de agentes superavitários, que precisa ser remunerado, supre a insuficiência de agentes deficitários, que por sua vez, pagará juros. Daí têm-se duas modalidades de operação efetuadas pela instituição financeira: de captação de recursos e de crédito (CARVALHO, 2007).

Importante observar que a demanda de crédito não é exclusiva de agentes deficitários. Há diversas circunstâncias que agentes podem demandar crédito mesmo ainda possuindo recursos, sendo esta ação para otimização do seu fluxo de caixa, ou de antecipação de aquisições de bens ou serviços, sejam eles para investimento ou consumo.

Esse simples papel de intermediação, acima exposto, retrata a visão de *Knut Wicksell*, economista sueco. Na estrutura demonstrada por Wicksell (1935), o crédito é melhor organizado e desenvolvido pelos bancos e mercados de ações. Para crédito de longo prazo, há atuação preponderante das bolsas de valores, enquanto os empréstimos junto a bancos, são créditos com prazos relativamente curtos.

Para Keynes (1992), o papel dos bancos transcende a simples intermediação postulada por *Wicksell*, complementando com a questão de que os bancos são provedores de liquidez ao sistema. Isso porque, complementarmente a *Wicksell*, que se limitava a ideia de que bancos criam crédito, mas com lastro em quantidade de moeda poupada, Keynes explora a capacidade de criar ativos financeiros, que por sua vez, aumentam a liquidez da economia.

Além de *Wicksell* e *Keynes*, Schumpeter (1982) deixa explícita a importância do crédito para o desenvolvimento econômico, tendo em vista que o crescimento no sistema capitalista gera alto grau de mortalidade de empresas, em função da concentração dos ganhos em empresas líderes. Ainda, havendo crédito barato para micro e pequenas empresas, além de empresários individuais, pode-se obter uma boa

fonte de crescimento. Se esses recursos forem direcionados para inovações em seus processos, produtos e serviços, ter-se-ia o desenvolvimento schumpeteriano.

Ainda, segundo Kalecki (1978), para que ocorram investimentos, são duas as fontes possíveis: reservas financeiras e o crédito. Dessa forma, pode-se afirmar que não é necessário existir poupança prévia para que ocorram investimentos produtivos, por exemplo. Sob as óticas de *Kalecki* e *Schumpeter*, fica evidenciado o papel fundamental do crédito para a atividade econômica e seu desenvolvimento, demonstrando que a dinâmica de acumulação de capital é, em certo sentido, dependente do crédito, o que é ainda mais latente em sistemas financeiros que operam com uma moeda estritamente fiduciária, onde não há restrições para a expansão endógena do crédito.

Corroborando com os autores *Kalecki* e *Schumpeter*, Carvalho et al. (2007) expõem que o crédito é de grande importância em uma economia capitalista, por permitir investimento. Conforme Bernanke e Blinder (1988), o setor bancário é de suma importância, basicamente por duas razões: considerando a política monetária no sistema econômico, pois o segmento é o responsável pela disponibilização de crédito para consumo e investimento, estes, considerados fatores fundamentais para geração de renda; e em função da criação de moeda pela expansão do crédito, que resulta em efeitos positivos na economia. De acordo com Rochman (2014), o crédito é considerado um propulsor do crescimento econômico, expandindo a demanda no mercado interno, conforme sua oferta, atuando diretamente nos níveis de renda (PIB) e emprego.

Dessa forma, fica evidente a importância do crédito para o nível de atividade econômica, seja ele para empreendimentos (grandes ou pequenos), seja ele para pessoas físicas, que por sua vez, irão adquirir bens ou serviços, gerando demanda para os empreendimentos.

3 METODOLOGIA

Neste capítulo, apresenta-se a metodologia que será adotada no presente trabalho no intuito de atender aos objetivos geral e específicos. Crucial para qualquer estudo, a metodologia científica deve proporcionar melhor entendimento, dando suporte às análises, contribuindo para geração de conhecimento.

3.1 Classificação da Pesquisa

O presente trabalho apresenta uma pesquisa teórica, baseada em uma revisão de literatura, que busca gerar conhecimentos acerca dos principais métodos para avaliação de risco de crédito. Definiu-se também uma abordagem qualitativa, considerando análises, geralmente, indutivas.

A pesquisa bibliográfica é feita a partir do levantamento de referências teóricas já analisadas, e publicadas por meios escritos e eletrônicos, como livros, artigos científicos, páginas de web sites. Qualquer trabalho científico inicia-se com uma pesquisa bibliográfica, que permite ao pesquisador conhecer o que já se estudou sobre o assunto. Existem porém pesquisas científicas que se baseiam unicamente na pesquisa bibliográfica, procurando referências teóricas publicadas com o objetivo de recolher informações ou conhecimentos prévios sobre o problema a respeito do qual se procura a resposta. (FONSECA, 2002, p. 32).

Sendo assim, por meio de pesquisa bibliográfica, o trabalho reúne diversos estudos sobre o tema, trazendo seus destaques para o atendimento aos objetivos geral e específicos. Conforme Gil (2007), a pesquisa bibliográfica apresenta como principal exemplo ao processo de investigação dos diferentes posicionamentos sobre determinado tema e/ou problema.

Diversas fontes foram consultadas durante o processo de busca e revisão da literatura, principalmente para o segundo e quarto capítulos, no intuito de considerar diferentes percepções sobre o tema, buscando uma análise mais ampla e imparcial.

Inicialmente, buscou-se termos como inadimplência, risco de crédito, modelos para risco de crédito, previsão de inadimplência, além destes termos também em inglês. Em sua maioria, o número de ocorrências para estas buscas referia-se a *Credit Scoring*, passando a ser incluído também nas pesquisas. A partir daí, se buscou os principais métodos para avaliação de risco de crédito, percebendo-se a predominância de Regressão Logística e *Machine Learning* para estas pesquisas, que foram

efetuadas em diversos sites e bases de dados, a saber: Google acadêmico, Portal de Periódicos CAPES, Banco Central do Brasil, *Research Gate*, dentre outros.

Inúmeros materiais traziam apenas conceitos sobre essas técnicas, mas não efetuavam uma análise/teste quantitativo e interpretação de resultados. Neste caso, alguns foram selecionados para revisão de literatura, apenas. Porém, os que continham testes aplicados foram selecionados para estabelecer a análise comparativa proposta.

Em suma, os principais temas buscados foram: *Credit Scoring*, *Machine Learning*, Regressão Logística e Risco de Crédito. No intuito de contextualizar sobre o tema, além dos modelos de avaliação abordados, especificamente, pesquisou-se também sobre o papel do crédito na economia, a evolução do crédito na economia brasileira, além de metodologia de pesquisa.

Durante a pesquisa, priorizou-se materiais referenciados, com número significativo de citações, além do histórico (profissional, quando aplicável) e relevância dos autores, visando a objetividade e robustez do presente trabalho.

4 ANÁLISE DE RISCO DE CRÉDITO

Segundo o BACEN (2009), o risco de crédito significa a possibilidade de ocorrência de perdas associadas ao não cumprimento pelo tomador ou contraparte de suas respectivas obrigações financeiras nos termos pactuados, à desvalorização de contrato de crédito decorrente da deterioração na classificação de risco do tomador, à redução de ganhos ou remunerações, às vantagens concedidas na renegociação, aos custos de recuperação, entre outros.

No presente estudo, o foco é no tocante ao não cumprimento pelo tomador de suas respectivas obrigações financeiras, especificamente. O risco de crédito não deve ser visto como algo negativo e sim como oportunidade, desde que se tenha mecanismos e processos para mensuração e mitigação deste risco.

Importante conceituar a assimetria de informação que existe no mercado, onde o cliente sabe mais de si do que seu possível credor. Dessa forma, em virtude da informação assimétrica, há um risco inerente para o credor. Como mitigador dessa assimetria, a análise de crédito, ferramentas e técnicas derivadas são aperfeiçoadas ao longo da história. Um fenômeno de assimetria de informação é o da seleção adversa, que pode ser exemplificado justamente no setor bancário. Ocorre que, como a taxa de juros dos empréstimos são muito maiores que a taxa básica de juros (Selic), os potenciais bons pagadores tentam a contrair menos dívidas, ao passo que os potenciais maus pagadores contratam os empréstimos independente da taxa de juros cobrada, pois a intenção de honrar seus compromissos é mais fraca (PINDYCK, 2006).

Conforme Silva (2016), como mitigadores, devem ser avaliados os chamados Cs do crédito, a saber:

- **Caráter:** visa avaliar de forma indireta a índole e a intenção de honrar seus compromissos. Essa avaliação é realizada pelo histórico de pagamentos do proponente;
- **Capacidade:** como o próprio termo reflete, indica a capacidade de pagamento dos compromissos financeiros, tanto existentes, quanto pretendidos;
- **Capital:** diz respeito à solidez financeira do indivíduo, que reflete indiretamente a forma de lidar com seu orçamento e também sua resiliência

frente a infortúnios. Em suma, se há lastro patrimonial frente às suas dívidas, atuais e pretendidas;

- Condições: refere-se a possíveis impactos externos a sua renda principal. Há de se avaliar o contexto socioeconômico da pessoa;
- Colateral: quando o empréstimo ou financiamento possui garantia(s) real(is) vinculadas ao negócio;
- Conglomerado: considera a situação financeira do grupo econômico, ou seja, participações societárias, empresas coligadas, controladoras.

Ainda segundo o autor, a avaliação desses aspectos pode ser feita de forma individual, julgamental, com certa subjetividade, ou de forma quantitativa, massificada, por meio de técnicas estatísticas, matemáticas e/ou computacionais.

Cabe ressaltar que análise individual não significa ausência de padrão e de critérios. Cada instituição pode adotar critérios e exigências conforme seu apetite ao risco, estratégia, considerando também diferentes modalidades, como empréstimos ou financiamentos, com garantias estabelecidas ou não. Para isso, definem políticas e normativos em geral, padronizando seus critérios, seja para análise individual dos clientes, seja para análise massificada (SILVA, 2016).

O que se pretende, geralmente, é a expansão da carteira de crédito, com eficiência operacional e rentabilidade. Para tanto, faz-se necessário conhecer os riscos incorridos e, por sua vez, mitigá-los, filtrando clientes, ajustando valores, condições, modalidades disponíveis, ou até mesmo, exigindo garantias para a operação em questão.

Conforme Caouette et al. (1999), o risco de crédito é um dos mais antigos riscos existentes no mercado financeiro, definindo crédito como a expectativa de recebimento de uma quantia em dinheiro, considerando um prazo estabelecido. Assim, define-se o risco de crédito como a chance de que essa expectativa não se concretize.

Ressalta-se que, para gestão do negócio crédito, considerando aspectos como eficiência e rentabilidade, não basta que se tenha uma análise de crédito, de qualquer natureza, bem executada. Há outras disciplinas que compõem o negócio como um todo que são de suma importância para seu sucesso e perpetuidade, como por exemplo, políticas de recuperação e cobrança, estratégia comercial, formatos de produtos aderentes às necessidades dos clientes, além de adequada precificação sobre as operações a serem concedidas (MANFIO, 2007).

Importante recapitular a origem do *Credit Scoring*, em 1945, quando foi desenvolvido o primeiro modelo estatístico de análise de crédito, com foco no público consumidor final, no intuito de empresas de varejo e financeiras mitigarem a assimetria de informação. O crescimento no desenvolvimento e utilização desse tipo de modelos está diretamente relacionada ao aumento do mercado de crédito (varejo, pulverizado) e o aprimoramento dos sistemas computacionais, proporcionando tratamento estatístico em grandes quantidades de registros e informações (LEWIS, 1992).

Segundo Carmona e Amorin Neto (2002), muitos modelos se limitam a prover classificações úteis para a concessão propriamente dita e respectivo valor, entretanto, modelos baseados no comportamento dos clientes podem ser utilizados, não somente na concessão inicial de crédito, mas também na manutenção dos limites de crédito expostos e até mesmo em formas preventivas de cobrança, possibilitando traçar diversas estratégias para o ciclo de crédito.

Ainda que existam diversas técnicas utilizadas, sejam elas mais simplificadas ou rudimentares, ou até as mais avançadas e sofisticadas, não é possível determinar a melhor técnica. Isso depende do objetivo do estudo, da estrutura de dados e das ferramentas disponíveis (HAND; HENLEY, 1997). Outro aspecto importante é que os resultados de alguns estudos demonstram que as capacidades de predição não são tão distintas, sendo as eventuais diferenças insignificantes para a determinação de melhor técnica a ser utilizada (ABDOU; POINTON, 2011).

Conforme Camargos et al. (2012), no Brasil, as instituições financeiras passaram a utilizar de forma mais significativa os modelos de *Credit Scoring* em meados dos anos 1990, após estabilidade alcançada com a implantação do Plano Real. Ainda, a Regressão Logística é a técnica mais utilizada no mercado para modelos de *Credit Scoring* (CROOK; EDELMAN; THOMAS, 2007).

Apesar de alguns credores utilizarem modelos julgamentais, a grande maioria dos players do mercado de crédito utilizam modelos desenvolvidos com técnicas de análise estatística como análise discriminante⁸ e Regressão Logística, ou até mesmo em modelos de inteligência artificial (ANDRADE, 2004). Estes dois últimos são os objetos de análise deste trabalho: Regressão Logística e *Machine Learning*.

⁸ Análise discriminante: técnica estatística multivariada utilizada para classificação e discriminação de objetos da população estudada, tendo a discriminação sua primeira etapa (KHATREE; NAIK, 2000).

Nas próximas subseções, são descritos os métodos de Regressão Logística, sendo este mais antigo e tradicional, conforme descrito anteriormente, e *Machine Learning*, que vem sendo explorado de forma significativa, surgindo como possível alternativa ao método anterior. Além da descrição de cada método, realizou-se também a análise alguns artigos relevantes que examinaram o risco de crédito por meio desses dois métodos. O critério de escolha desses estudos se baseou principalmente no número de citações na literatura, obtido no estudo de Assef e Steiner (2020).

4.1 Credit Scoring

Utilizado no Brasil de forma mais significativa a partir da década de 1990, *Credit Scoring* consiste em métodos estatísticos adotados para classificar em grupos de risco, possíveis proponentes, por meio de análise de informações passadas, verificando o ocorrido em momento futuro. De acordo com Thomas (2000), o principal uso do *Credit Scoring* é a previsão de inadimplência. Dessa forma, quanto melhor for a pontuação de determinado proponente, menos propenso à inadimplência ele será.

Esta classificação pode ser realizada por meio de modelos estáticos ou dinâmicos. Um modelo estático, denominado de *Application Score*, baseia-se em informações em determinado momento, predominantemente oriundas de cadastro, podendo conter algumas informações de operações existentes. Já um modelo dinâmico, denominado *Behaviour Score*, compreende uma análise contínua dos clientes, como se fosse um filme. Trata-se de um modelo comportamental. Importante citar também o *Collection Score* também caracterizado como modelo dinâmico (SICSÚ, 2010; MANFIO, 2007).

Conforme Sicsú (2010), a variável resposta para o modelo comportamental (*Behaviour Score*), usualmente será o evento inadimplência do cliente, podendo variar o número de dias em atraso (30 dias, 60 dias etc.). Quando o modelo em questão for o *Collection Score*, a variável resposta é o pagamento da dívida em atraso. A seguir, os modelos são vinculados às respectivas etapas do ciclo de crédito, tornando mais tangível a aplicação e diferenças entre eles.

Elencados os três modelos, *Credit Score*, *Behaviour Score* e *Collection Score*, é preciso definir em que etapa do ciclo de crédito cada um deles se insere. O *Credit Score* é utilizado como apoio a decisão para tomada inicial de risco, ou seja, na

iniciação em crédito. Isso ocorre pois até então, não há dados e informações suficientes para estabelecer um modelo comportamental, dada a maior assimetria de informação e, por conseguinte, um risco de crédito maior. O *Behaviour Score* é utilizado na manutenção em crédito, onde já é possível analisar o comportamento do cliente, tendo em vista o tempo de relacionamento e utilização de produtos e serviços. E, por fim, o *Collection Score* é utilizado na recuperação e cobrança, onde visa estimar a chance de recuperar o crédito concedido anteriormente, conforme Anderson (2007).

Em suma, conforme Crouhy et al. (2008, p. 184) e corroborando com Manfio (2007, p.31):

Contudo, a classificação de crédito também é importante por razões de custo e consistência. Os grandes bancos tradicionalmente detêm milhões de clientes e executam bilhões de transações a cada ano. Utilizando o modelo de *Credit Scoring*, os bancos podem automatizar o máximo possível o processo de concessão de pequenos créditos e cartões de crédito. Antes da ampla adoção da escoragem do crédito, um executivo de crédito teria de rever uma solicitação e usar uma combinação de experiência, conhecimento do setor e técnica pessoal para chegar a uma decisão de crédito baseada na grande quantidade de informação de uma típica solicitação de crédito.

Portanto, trata-se de um ciclo que pode e deve ser retroalimentado, aprimorando a qualidade da concessão, manutenção e recuperação de crédito.

No segmento de instituições financeiras, empresas que efetuam venda a crédito (crediário próprio) ou oferecem produtos e serviços financeiros como cartão de crédito, é comum a utilização de *Credit Score*, seja ela de *bureau* externo, ou desenvolvimento interno. A escolha dependerá da estrutura da empresa, bem como expertise no tema, pois modelos dessa natureza requerem análises específicas e manutenção. Dessa forma, dependendo da atividade principal da empresa, pode ser mais vantajoso contratar esse serviço de um *bureau* externo.⁹

Não há um consenso na literatura sobre o melhor método para obter um bom modelo de *Credit Scoring*. De acordo com Altman e Saunders (1997), só a partir da década de 1970 que as técnicas estatísticas passaram a ser empregadas para modelagem, com predomínio da Regressão Logística. Mais recentemente, a partir dos anos 1990, tecnologias de inteligência artificial passaram a ser usadas para a análise

⁹ Uma alternativa é o uso combinado: existência de modelo interno, mas que uma das variáveis é o escore de crédito de *bureau* externo. Esse formato é interessante principalmente quando o cliente (ou potencial) é novo ou sem muito relacionamento com a instituição, o que limitaria a quantidade de informações disponíveis. Todavia, pode ser utilizado dessa forma para simplesmente enriquecer o modelo, aumentando sua estabilidade e/ou poder de discriminação (MANFIO, 2007).

de crédito, especialmente o Aprendizado de Máquina (*Machine Learning*). Assim, as próximas subseções examinam esses métodos de avaliação.

4.1.1 Regressão Logística

A técnica de regressão logística teve sua relevância revelada devido a análises na área de saúde com o objetivo de prever fatos antecedentes a doenças coronárias, segundo Bittencourt (2003). Conforme Hosmer e Lemeshow (1989) e Cox e Snell (1989), o modelo de regressão logística teve seu maior reconhecimento após o projeto *Framingham Heart Study*, de Truett et al. (1967), sendo considerado o marco inicial dos estudos envolvendo esse método nas áreas da saúde.

Cabe ressaltar que a referida técnica permanece com muitas aplicações até hoje, como observa Mesquita (2014), tendo diversas áreas do conhecimento utilizando-se desta, a saber: estudo de mercado, ciências médicas, intenção de voto e, inclusive, avaliação de crédito, tema do presente estudo. Segundo Lessmann et al. (2015), a Regressão Logística é considerada a técnica padrão quando o tema é *Credit Scoring*, corroborando com demais autores citados anteriormente (CROOK; EDELMAN; THOMAS, 2007).

A regressão logística, assim como a regressão linear, busca explicar a relação entre duas ou mais variáveis. Porém, sua variável dependente é binária (0 ou 1). Considerando a variável dependente Y assumir apenas dois valores, 0 (quando o cliente é adimplente) ou 1 (quando ele é inadimplente), e um conjunto de p variáveis independentes $X_1, X_2, X_3, \dots, X_p$, o modelo de regressão pode ser representado pela equação 1:

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}}$$

$$\text{Onde: } g(x) = B_0 + B_1X_1 + \dots + B_pX_p$$

No caso, a variável dependente seria a inadimplência, que é o evento que se quer explicar. As variáveis independentes, que determinam a inadimplência, são muitas e variam conforme o tipo de pessoa (física ou jurídica). No caso de pessoas físicas, segundo Sicsú (2010), podem ser variáveis explicativas: local de nascimento, idade, profissão, estado civil, tipo de residência, renda mensal, tipos de dívida que

possui, tempo de relacionamento com a instituição, se possui histórico de atraso, além de restritivos em geral (protestos, negativas em serviços de proteção ao crédito, cheques sem fundo). No que se refere a pessoas jurídicas, tem-se: quantidade de sócios, data de constituição, setor de atividade, concentração societária, participações em outras empresas, faturamento, endividamento, liquidez, patrimônio, histórico de atraso, além de restritivos (semelhantes aos citados para pessoas físicas).

Ferreira et al. (2012) propõem um modelo estatístico para análise da carteira de crédito de uma agência bancária. Por meio de regressão logística (*logit binomial*), a análise efetuada considerou informações de uma carteira de crédito de uma instituição do município de Viçosa (MG), clientes pessoa física. As variáveis explicativas selecionadas, inicialmente, foram: estado civil, grau de escolaridade, idade, renda, tempo de relacionamento com a instituição, saldo médio em conta, limite de cheque especial e existência de cadastro de cheque sem fundo. A variável resposta definida foi a inadimplência, nesse caso representada por atraso superior a 14 dias.

O modelo se mostrou eficiente, contribuindo positivamente com as instituições na decisão do crédito, apresentando taxa de acurácia de 92%. Quanto à expectativa sobre as variáveis explicativas, ressalta-se que algumas se comportaram de forma oposta: quanto maior o tempo de relacionamento, maior a inadimplência observada; quanto maior a renda, maior a inadimplência. Entretanto, as que se comportaram de acordo com a expectativa descrita no artigo foram: idade do proponente, quanto maior, menor a inadimplência; limite de cheque especial, quanto maior o valor, menor a inadimplência; quanto ao estado civil, solteiros apresentam menor inadimplência quando comparados aos tomadores casados. A variável saldo médio não foi utilizada, pois demonstrou baixa significância.

Outro estudo analisado que, em busca de mecanismos mais robustos de análise de risco de crédito, especificamente para o público de pessoas físicas com o objetivo de minimizar a inadimplência em uma instituição privada de ensino superior, propôs um modelo de risco de crédito utilizando a técnica de regressão logística foi de Ribeiro et al. (2009). Os autores obtiveram resultados satisfatórios, atingindo um percentual de classificação correta na ordem de 82% dos casos, proporcionando uma gestão mais eficiente na concessão de crédito, reduzindo a inadimplência na universidade.

As variáveis explicativas iniciais foram diversas, somando 59, dentre elas: gênero, idade, estado civil, histórico curricular, histórico de atraso, renda, escolaridade

dos pais, relacionamento bancário, dentre outras similares. Porém, em função dos testes realizados (teste não-paramétrico do qui-quadrado e correlação), considerando os índices de significância inferiores a 5%, somente três variáveis compuseram o modelo: se o aluno já teve reprovação em seu histórico, se já teve que negociar débitos e se é detentor de cartão de crédito. Quanto à interpretação do modelo e o comportamento das variáveis explicativas, observa-se o seguinte: o aluno que já foi reprovado, já negociou débitos e não possui cartão de crédito tem probabilidade maior de inadimplência.

Pereira e Ness (2003) elaboraram um estudo propondo um modelo de regressão logística capaz de estimar a probabilidade do evento de falência em empresas. Análogo a um modelo de *Credit Score*, em virtude de que se há capacidade preditiva para falências, essa variável resposta também é útil para tomada de decisão de concessão de crédito, esse modelo demonstrou ser bastante eficiente ao estimar a probabilidade de falência ou concordada de empresas, perfazendo um percentual de acertos na ordem de 97%.

Para tal, foram testadas inicialmente 63 variáveis, dentre elas: ativo circulante sobre passivo circulante, ativo circulante acrescido de ativo realizável a longo prazo sobre o passivo circulante, resultado financeiro líquido sobre o passivo total, caixa sobre passivo circulante, caixa sobre o ativo total, dentre outras baseadas nas informações constantes nas demonstrações financeiras das empresas. Após aplicação de testes (*Levene* e teste *t*), restaram 26 variáveis, que ainda foram avaliadas as de maior significância, permanecendo 18 delas. Ainda, após testes adicionais (*Kolmogorov-Smirnov* e avaliação do índice de *Spearman*), restando três variáveis compondo o modelo: pesquisa e desenvolvimento sobre o número de funcionários, despesas de juros sobre o passivo total e fluxo de caixa operacional sobre o passivo circulante.

Dessa forma, o modelo apresentou grande utilidade para instituições financeiras no intuito de estabelecer o risco de crédito vis a vis rentabilidade esperada, selecionando melhor os clientes para o portfólio.

Em outro artigo, Camargos et al. (2012) analisaram aproximadamente dez mil empresas do segmento micro e pequenas, em um universo de 25 mil processos de financiamento. No estudo, foram utilizadas 22 variáveis explicativas, sendo cinco (valor do financiamento, relação entre bens do avalista/sócio e o valor do financiamento, valor em investimentos fixos do projeto, tempo de atividade da

empresa e relação entre faturamento e valor do financiamento) significativas para predição da inadimplência com nível de acerto de 88,5%.

Diante dos estudos abordados, a regressão logística demonstra ser uma técnica adequada para auxiliar na tomada de decisão no que tange ao risco de crédito, proporcionando maior eficiência e rentabilidade nos processos de análise, tanto para pessoas físicas, quanto para pessoas jurídicas.

4.1.2 *Machine Learning*

Machine Learning consiste em um conjunto de procedimentos e regras que possibilitam que decisões sejam tomadas de forma automática com base em dados. Dessa forma, o computador deixa de ser um simples executor de algo programado anteriormente, de forma estática. O uso desta tecnologia tem como objetivo prever resultados, conforme os dados existentes, valendo-se de fundamentos estatísticos, matemáticos e computacionais. Considerando a definição de Han et al. (2011), trata-se da área de pesquisa que utiliza “[...] programas de computador que automaticamente aprendem a reconhecer padrões complexos e a tomar decisões inteligentes com base em dados”.

Há diversas aplicações possíveis, em variados segmentos do mercado, a saber: indústrias, no intuito de identificar falhas ou oportunidades de melhoria no processo produtivo; vendas no varejo, verificando melhores ofertas conforme perfis de pessoas, aumentando a aderência dos produtos e vendas; para o setor público, para entender melhor o comportamento da população, além de prevenir fraudes; seguros, indo além da prevenção à fraudes, mas também identificando pacientes com maior risco de doenças, bem como intervenções mais eficazes; e, por fim, serviços financeiros, desde prevenção a fraudes, melhor oferta, *cross-sell*¹⁰, redução de *churn*¹¹, e sobretudo, mensurar o risco de crédito (ALPAYDIN, 2010).

Importante elucidar os métodos e algoritmos no âmbito de *Machine Learning*. Para tal, tem-se os métodos *ensemble*, que nada mais são do que a utilização conjunta de modelos simples, até mesmo de baixo poder preditivo, buscando um resultado mais robusto, aumentando a acurácia como um todo (GROVER, 2017).

¹⁰ Considerando a necessidade do cliente, trata-se de técnica utilizada com o objetivo de que o cliente aceite ofertas complementares ao que já está em primeiro plano (KOTLER; SETIAWAN, 2017).

¹¹ Conforme Glady et al. (2009), a definição é quando o ciclo de vida do cliente é decrescente, incluindo seu encerramento.

Dentre os mais utilizados, pode-se elencar os algoritmos *random forest* e *extreme gradient boosting (XGBoost)*. O primeiro consiste num classificador considerando diversas árvores de decisão estruturadas com vetores distribuídos de forma idêntica, aleatórios e independentes. Cada árvore gera um voto, onde o resultado final do algoritmo se dá pela soma dos votos. Como vantagem, o método é capaz de minimizar a correlação entre os erros gerados nas árvores de decisão, com certo controle sobre a variância (BREIMAN, 2001; HASTIE et al., 2001).

O *XGBoost*, desenvolvido por Chen e Guestrin (2016), segundo Grover (2017), pode ser compreendido sem notações matemáticas, estabelecendo um paralelo com a regressão linear, onde a premissa básica é que a soma dos resíduos é zero. Trata-se de um modelo de classificação e regressão baseado em árvores de decisão agregadas que tem como objetivo combinar resultados de classificação relativamente fracos, mas que quando combinados, se tornam fortes, auxiliando na tomada de decisão.

Segundo Belloti e Crook (2009), o *support vector machine (SVM)* é um algoritmo usado em modelo de classificação onde dados binários são separados num hiperplano, maximizando a distância entre os exemplos. Primeiro, ocorre a classificação das informações e, na sequência, maximiza as distâncias.

Outro algoritmo citado é a rede neural, que possui um campo de aplicação significativo como física, psicologia, biologia, engenharia, além do que se propõe o presente trabalho (MASSON; WANG, 1990). Análogo ao funcionamento do cérebro humano, possui pequenas unidades de processamento interligadas por conexões diretas ponderadas. Essas unidades representam os neurônios e as conexões, as sinapses. No que tange ao *Credit Scoring*, Hand e Henley (1997) observam que o método se torna mais adequado quando a estrutura de dados não é de fácil compreensão.

Segundo Marra (2019), o uso de *Machine Learning* tem contribuído de forma significativa no aprimoramento de processos de avaliação e decisão acerca do risco de crédito. No estudo, onde analisou-se índices financeiros de empresas da América Latina, o algoritmo *XGBoost* supera de forma substancial a regressão logística. O critério utilizado para esse diagnóstico foi um indicador largamente aplicado, segundo Zhang et al. (2018), para avaliação de performance de modelos, denominado *Kolmogorov-Smirnov (KS)*, que consiste em verificar o poder de discriminação do modelo em questão. Dessa forma, pode-se dizer que a evolução do poder preditivo

se utilizando de *Machine Learning* tem proporcionado melhorias para a análise do risco de crédito.

Em relação ao desempenho superior de algoritmos de *Machine Learning* quando comparado com a regressão logística, no que tange a modelos para concessão de crédito, Becker (2018) expõe um aspecto relevante sobre qual método deve ser utilizado: sua facilidade (ou dificuldade) de implementação. Isso sugere que, apesar de pior poder preditivo, a tradicional regressão logística pode ainda ser mais vantajosa em virtude do processo de implantação em si.

Porém, segundo Aniceto (2016), os algoritmos de *Machine Learning* não exigem um conhecimento tão aprofundado das variáveis a serem utilizadas e suas respectivas relações, o que difere das técnicas estatísticas tradicionais, como a regressão logística. Outro ponto de reflexão é abordado por Munkhdalai et al. (2019), que enfatiza a contribuição proporcionada pelo *Machine Learning*, reduzindo processos e trabalhos de especialistas quando comparado com técnicas mais tradicionais.

Esse ramo da inteligência artificial, *Machine Learning*, pode ser dividido em três métodos primários: supervisionado, não supervisionado e semi-supervisionado. O aprendizado supervisionado consiste em utilizar dados rotulados¹², treinando algoritmos para classificar ou prever resultados. O aprendizado não supervisionado utiliza algoritmos para analisar e agrupar dados não rotulados. Nesse caso, são identificados agrupamentos de dados e/ou padrões ocultos sem qualquer intervenção humana. Quanto ao aprendizado semi-supervisionado, este consiste na utilização de um grupo de dados rotulados que servirá para orientação de classificação de dados não rotulados.

Para o funcionamento desta técnica, necessita-se de dados, variáveis e algoritmos, basicamente. No quesito dados, faz-se necessário uma massa significativa de dados, além de diversificação. A obtenção dos dados pode ser classificada de duas formas: manual e automática. Na primeira, obtêm-se dados de maior qualidade, porém, com elevado tempo para coleta, o que requer maior

¹² Dados rotulados são aqueles onde o resultado que se busca já é conhecido (REZENDE, 2003). Exemplo: pretende-se identificar, num conjunto de clientes, quais são os inadimplentes. Para o método de aprendizado supervisionado, inicialmente se utiliza dados rotulados, ou seja, revela-se quais são os clientes inadimplentes, para que o algoritmo aprenda e seja capaz de identificar outros clientes sem que seja informada a situação de inadimplência.

investimento. Na segunda, de forma automática, requer menos investimento. Porém, pode apresentar mais erros, impactando no resultado final (ZHU, 2005).

Sobre as variáveis, que são de fato o que será analisado, estas podem estar estruturadas (como em tabelas, por exemplo), ou não estruturadas, como um conjunto de imagens. Por fim, os algoritmos: um conjunto de operações lógicas, sequenciais que possibilitam encontrar a solução para um problema específico.

Importante ressaltar que, independente da qualidade do algoritmo, se a qualidade dos dados for muito ruim, o resultado acompanhará. Não há como melhorar, via algoritmo, a qualidade dos dados, prejudicando a resposta do modelo.

No estudo de Xia et al. (2017), os autores propõem um modelo de *Credit Scoring* baseado em *XGBoost*, utilizando bases de dados da Alemanha, Austrália, Taiwan e mais dois oriundos de serviços de crédito *peer-to-peer*. Todos são bancos de dados reais. Segundo o trabalho, as variáveis utilizadas são compatíveis com bancos de dados tradicionais de *Credit Scoring* (informações demográficas, solvência, credibilidade). O modelo proposto teve desempenho satisfatório em todos os bancos de dados testados, comparado inclusive a outros modelos de classificação.

Em outro trabalho, de Barboza et al. (2017), modelos baseados em *Machine Learning* são testados para prever falência com um ano de antecedência. Foi utilizado banco de dados contendo informações de empresas norte-americanas do período de 1985-2013. Em média, a técnica demonstrou 10% maior acurácia que técnicas tradicionais. As variáveis utilizadas contemplam diversos indicadores econômico-financeiros, considerando capital de giro, total do ativo, valor de mercado por ação, ganhos antes dos juros e impostos, número de empregados, retorno sobre o patrimônio, etc.

Segundo Sun et al. (2018), em testes realizados para avaliação de crédito para empresas, o método baseado em *Machine Learning* apresentou desempenho satisfatório quando comparados com outros métodos, tendo como objetivo classificar as companhias em alto risco e baixo risco (de crédito). A base de dados utilizada continha informações financeiras de 552 empresas chinesas.

No estudo de Abreu (2020), o objetivo é analisar o desempenho de modelos de *Machine Learning*. Para tal, a autora utilizou-se de duas bases de dados: a primeira, composta por pessoas físicas, analisando-se uma carteira de cartões de crédito em Taiwan; e a segunda, formada por pedidos de empréstimos na Alemanha. Foram testados diferentes algoritmos, a saber: árvores de decisão, *support vector machine*,

random forest e redes neurais. Como conclusão, tem-se que o modelo *random forest* se destaca nas classificações tanto para a base de Taiwan, quanto para da Alemanha.

Lobato e Carvalho (2021) propõem construir um modelo capaz de classificar clientes entre bons e maus pagadores, utilizando variáveis que descrevem o perfil dos indivíduos. A base de dados utilizada continha 1.000 registros, com 20 variáveis, como: histórico de crédito, justificativa para o crédito, tempo de emprego, gênero, estado civil, se trabalhador estrangeiro ou não, se possui imóvel próprio, se possui telefone em seu nome, dentre outras similares.

O modelo proposto foi comparado com diversos algoritmos como árvore de decisão, *support vector machine*, *random forest*, dentre outros. Como conclusão, os autores concluem que por meio de *random forest* a sensibilidade e *F-measure* foram os melhores índices. Porém, o modelo proposto apresentou melhor acurácia, especificidade e menor índice de erro de falso positivo.

Diante dos estudos citados sobre *Machine Learning*, pode-se dizer que os modelos oriundos desta técnica apresentam bom desempenho para suporte ao processo de análise e decisão de risco de crédito.

4.2 Análise Comparativa

A análise dessa subseção se baseia nos artigos selecionados para a revisão bibliográfica, apresentados na subseção anterior. Conforme já mencionado, os estudos foram escolhidos de acordo, principalmente, com o número de citações na literatura, obtido no estudo de Assef e Steiner (2020).

Considerando os métodos analisados, tanto Regressão Logística quanto *Machine Learning* têm apresentado desempenho satisfatório para o processo de análise de risco de crédito e tomada de decisão. Como mais tradicional e difundido, o método de Regressão Logística apresenta algumas vantagens de implantação em função da relativa facilidade de interpretação ante ao *Machine Learning*. No caso de *Machine Learning*, pode haver, inclusive, resistência cultural em função da dificuldade de interpretação dos parâmetros, tornando seu uso mais restrito, ainda que seu poder preditivo, em alguns casos, possa ser maior do que o modelo com uso de Regressão Logística. Ainda, observa-se que um único algoritmo no âmbito de *Machine Learning* não necessariamente terá o melhor desempenho em todas as situações. Isso

depende dos dados que estão sendo trabalhados para obter-se a predição (FORTI, 2018).

Por outro lado, para modelos baseados em Regressão Logística, é desejável que os especialistas conheçam, além da técnica de regressão em si, também aspectos técnicos do negócio, o que pode tornar mais interessante a adoção de *Machine Learning*, que não necessita de inserções de variáveis pelo especialista, para combinações entre si, sendo esta função executada pelo algoritmo. No caso da regressão logística, existe a possibilidade de o especialista verificar as fontes de erro, permitindo que o modelo seja otimizado. Porém, isso gera esforço no desenvolvimento, o que não ocorre no caso de *Machine Learning* (FENSTERSTOCK, 2005).

Após a análise comparativa dos dois métodos, ao examinar os estudos mencionados na subseção anterior, é possível estabelecer alguns aspectos em comum entre esses métodos de avaliação de risco de crédito: as principais variáveis explicativas. No caso de pessoa física, a renda, o gênero e o histórico de crédito se repetem e apresentam poder explicativo muito relevante nos estudos examinados. No caso de pessoa jurídica, o faturamento da empresa, bem como o tempo de sua constituição são variáveis que também se repetem com maior frequência. Isso ocorre em função de serem, de fato, relevantes para o processo de análise de risco de crédito: o que poderá variar é o peso de cada variável, dependendo das demais variáveis disponíveis e/ou utilizadas *vis a vis* o método a ser adotado.

Todavia, a maioria das variáveis está diretamente ligada a um ou mais Cs do crédito, citados anteriormente, o que corrobora seu uso, indiferente do método utilizado (Regressão Logística ou *Machine Learning*) e segmento examinado (PF ou PJ). Cabe ressaltar também que, inclusive, em diferentes modelos específicos de Regressão Logística, essas variáveis explicativas estão presentes de alguma forma. O mesmo ocorre para *Machine Learning*. No Quadro 1, evidencia-se uma síntese dos métodos de avaliação.

Quadro 1 – Síntese dos Métodos de Avaliação

Método	Referências	Ênfase	Principais variáveis	Acurácia
Regressão Logística	Ferreira et al. (2012)	Pessoa Física	Estado civil, grau de escolaridade, idade, renda, tempo de relacionamento com a instituição, limite de cheque especial e existência de cadastro de cheque sem fundo.	91,9% de classificações corretas
	Ribeiro et al. (2009)	Pessoa Física	Histórico de reprovação, histórico de negociação de débitos, possui cartão de crédito.	82,0% de classificações corretas
	Pereira e Ness (2003)	Pessoa Jurídica	Razão entre Investimentos em Pesquisa e desenvolvimento e o número de funcionários, despesas de juros sobre o passivo total e fluxo de caixa operacional sobre o passivo circulante.	88,1% de classificações corretas
	Camargos et al. (2012)	Pessoa Jurídica	Valor do financiamento, relação entre bens do avalista/sócio e o valor do financiamento, valor em investimentos fixos do projeto, tempo de atividade da empresa e relação entre faturamento e valor do financiamento.	88,5% de classificações corretas
Machine Learning	Xia et al. (2017)	Pessoa Jurídica	Informações demográficas, solvência, credibilidade.	Classificações corretas para base de dados da Alemanha: 77,18%; Áustria: 87,82%; Taiwan:69,41%.
	Barboza et al. (2017)	Pessoa Jurídica	Capital de giro, total do ativo, valor de mercado por ação, ganhos antes dos juros e impostos, número de empregados, retorno sobre o patrimônio.	87,0% de classificações corretas
	Sun et al. (2018)	Pessoa Jurídica	Indicadores econômico-financeiros em geral	81,1% de classificações corretas
	Abreu (2020)	Pessoa Física	Estado civil, gênero, histórico de crédito, tempo de relacionamento com a instituição, patrimônio, motivo do crédito, montante do crédito.	Não informado
	Lobato e Carvalho (2021)	Pessoa Física	Histórico de crédito, justificativa para o crédito, tempo de emprego, gênero, estado civil, trabalhador estrangeiro, imóvel próprio, telefone próprio.	Não informado

Fonte: Elaboração do autor.

Com isso, nota-se a importância do conceitual precedente da literatura e prática no negócio crédito, capaz de contribuir com ambos os métodos de avaliação, dando mais segurança ao processo decisório e também de manutenção dos modelos, sejam eles quais forem. Portanto, a escolha do método mais adequado para análise de risco de crédito depende da percepção da empresa em relação aos benefícios gerados por cada um deles, considerando aqueles aspectos mencionados no início desta subseção.

5. CONCLUSÕES

O objetivo geral deste trabalho foi realizar um exame de diferentes técnicas de análise de risco de crédito, estabelecendo um comparativo entre elas. Ainda, considerando os objetivos específicos, descreveu-se o Sistema Financeiro Nacional, examinou-se a evolução recente do crédito no contexto brasileiro, tanto pela ótica do saldo, quanto da inadimplência, contemplando segmentações para melhor compreensão do tema, e por fim, analisou-se os métodos de Regressão Logística e *Machine Learning* no âmbito da análise de risco de crédito, mais especificamente *Credit Scoring*. Para isto, a metodologia utilizada foi qualitativa, baseada em pesquisa bibliográfica, conforme visto no terceiro capítulo.

Sobre a evolução recente do crédito no Brasil, destaca-se o aumento de sua participação na economia, mantendo o nível de inadimplência sem crescimento expressivo e até com períodos de declínio. Isso ocorre, de forma preponderante, em função da evolução das técnicas de análise e concessão de crédito, objeto do presente trabalho, justificando, ainda, a relevância do tema para a academia e para o mundo dos negócios.

Considerando os métodos analisados por meio da revisão bibliográfica, tanto a Regressão Logística quanto *Machine Learning* mostraram desempenho satisfatório para o processo de análise de risco de crédito e tomada de decisão. Como mais tradicional e difundido, o método de Regressão Logística apresenta algumas vantagens de implantação, em função da relativa facilidade de interpretação ante ao *Machine Learning*. No caso de *Machine Learning*, pode haver inclusive resistência cultural, em função da dificuldade de interpretação dos parâmetros, tornando seu uso questionável, ainda que seu poder preditivo, em alguns casos, possa ser maior do que modelo com uso de Regressão Logística. Ainda, observa-se que um único algoritmo no âmbito de *Machine Learning* não necessariamente terá o melhor desempenho em todas as situações. Isso depende dos dados que estão sendo trabalhados para obter-se a predição.

Em relação aos modelos baseados em Regressão Logística, é desejável que os especialistas conheçam, além da técnica de regressão em si, mas também aspectos técnicos do negócio, o que pode tornar mais interessante a adoção de *Machine Learning*, que não necessita de inserções de variáveis pelo especialista, para combinações entre si, sendo esta função executada por algoritmo. No caso da

regressão logística, existe a possibilidade de o especialista verificar as fontes de erro, permitindo que o modelo seja otimizado. Porém, isso gera esforço no desenvolvimento, o que não ocorre no caso de *Machine Learning*.

Ressalta-se que, em ambos os métodos, as principais variáveis explicativas são comuns. Em relação a pessoas físicas, destacam-se: estado civil, gênero, histórico de crédito e presença de restritivos. Considerando pessoas jurídicas, também são identificadas variáveis comuns entre os métodos, sendo estas baseadas principalmente em índices econômico-financeiros, como total do ativo, retorno sobre o patrimônio, capital de giro, além de informações demográficas.

A identificação das variáveis nos trabalhos que abordam especificamente a técnica de *Machine Learning* não se dá de forma tão objetiva, possivelmente em função de que o próprio método contempla etapas e algoritmos que definem conforme significâncias apresentadas, sem calibragem ou interferência humana.

Dessa forma, a pesquisa contribui positivamente agrupando diferentes estudos antecedentes sobre o tema de forma sintética, trazendo uma análise comparativa entre os métodos avaliados, além da contextualização no tema crédito e risco de crédito. Essa síntese pode servir de insumos para futuras pesquisas no tema, ou até mesmo para profissionais do segmento que queiram conhecer o histórico e debates que permeiam esses assuntos relacionados.

Apesar das contribuições apresentadas, algumas limitações deste estudo podem ser destacadas. Ao focar na revisão bibliográfica, o estudo não trouxe um conhecimento novo à literatura, que poderia ocorrer caso se utilizasse um dos métodos examinados para a realização de um estudo empírico em relação a um tipo de crédito específico. Outro ponto relevante e que poderia ampliar e dar mais consistência aos resultados da pesquisa bibliográfica é a utilização de uma variedade maior de estudos examinados.

REFERÊNCIAS

ABREU, M. C. F. **Modelos de Avaliação de Risco de Crédito**: Aplicação de *Machine Learning*. Universidade de Coimbra, 2020.

ALPAYDIN, E. **Introduction to Machine Learning**. MIT Press, 2010.

ANDERSON, Raymond. **The Credit Scoring Toolkit**: Theory and Practice for Retail Credit Risk Management and Decision Automation. Oxford: Oxford University Press, 2007.

ANDRADE, F. W. M. **Desenvolvimento de Modelo de Risco de Portfólio para Carteiras de Crédito a Pessoas Físicas**. Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, São Paulo, 2004.

ANICETO, M. C. **Estudo comparativo entre técnicas de aprendizado de máquina para estimação de risco de crédito**. 2016.

ASSEF, F. M.; STEINER, M. T. A. Ten-year evolution on credit risk research: a Systematic Literature Review approach and discussion. **Ingeniería e Investigación**, v. 40, n. 2, p. 50-71, 2020.

BANCO CENTRAL DO BRASIL. **Boletim do Banco Central do Brasil – Relatório Anual 2015**. Disponível em: <<https://www.bcb.gov.br/pec/boletim/banual2015/rel2015p.pdf>>. Acesso em junho 2020.

BANCO CENTRAL DO BRASIL. **Relatório de Estabilidade Financeira**. Brasília, 2013. Disponível em: <<file:///Users/cristianopires/Downloads/RELESTAB201303-refP.pdf>>. Acesso em junho 2020.

BANCO CENTRAL DO BRASIL. **SGS – Sistema Gerenciador de Séries Temporais**. Disponível em: <<https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>>. Acesso em maio 2020.

BANCO CENTRAL DO BRASIL. **Sistema Financeiro Nacional (SFN)**. Disponível em: <<https://www.bcb.gov.br/estabilidadefinanceira/sfn>>. Acesso em maio 2020.

BANCO CENTRAL DO BRASIL. **Resolução nº 3.721**. Dispõe sobre a implementação de estrutura de gerenciamento de risco de crédito. Disponível em: <https://www.bcb.gov.br/pre/normativos/res/2009/pdf/res_3721_v1_O.pdf>. Acesso em agosto 2021.

BANCO CENTRAL DO BRASIL. **Resolução nº 4.782**. Estabelece critérios temporários para a caracterização das reestruturações de operações de crédito, para fins de gerenciamento de risco de crédito. Disponível em: <https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/50937/Res_4782_v2_P.pdf>. Acesso em agosto 2021.

BARBOZA F.; KIMURA H.; ALTMAN E. Machine learning models and bankruptcy prediction. **Expert Systems with Applications**, 2017.

BECKER, C. **Estudo comparativo entre abordagens de aprendizado de máquina em modelos de credit scoring**. Monografia. Universidade Federal do Rio Grande do Sul, 2018. Disponível em <<http://hdl.handle.net/10183/201492>>. Acesso em julho 2021.

BELLOTI, T.; CROOK, J. Support vector machines for credit scoring and discovery of significant features. **Expert systems with applications**, 2009.

BERNANKE, Ben S.; BLINDER, Alan S. Credit, money, and aggregate demand. **The American Economic Review: Papers and Proceedings**, v. 78, n. 2, p.435-439, 1988. Disponível em: <https://www.nber.org/papers/w2534>. Acesso em: 12 set. 2019.

BITTENCOURT, H. R. Regressão logística politômica: revisão teórica e aplicações. **Acta Scientiae**, Canoas, v. 5, 2003.

BREIMAN, L. Random forests. Machine learning, **Springer**, v. 45, n. 1, 2001.

BROWN, I.; MUES, C. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. **Expert Systems with Applications**, Elsevier, v. 39, n. 3, p. 3446–3453, 2012.

CAMARGOS, M. A.; CAMARGOS, M. C. S.; ARAÚJO, E. A. A inadimplência em um programa de crédito de uma instituição financeira pública de Minas Gerais: uma análise utilizando regressão logística. **Revista de Gestão**, v. 19, n. 3, p. 467-486, 2012.

CAMARGOS, M. A.; SOARES, G. O. G.; COUTINHO, E. S. Determinantes do rating de crédito de companhias brasileiras. **Revista Contabilidade Vista & Revista**, v. 23, p. 109-143, 2012.

CAOQUETTE, John B.; ALTMAN, Edward I.; NARAYANAN, Paul. **Gestão de Risco de Crédito: O próximo grande desafio financeiro**. Rio de Janeiro: Qualitmark, 1999.

CARMONA, C. U; AMORIN NETO, A. Modelagem do Risco de Crédito: Um Estudo do Segmento de Pessoas Físicas em um Banco de Varejo. **Revista Eletrônica de Administração da UFRGS - REAd**. 40^a Edição, Porto Alegre, v. 10, Jul/ago, 2004.

CARVALHO, F. J. C. de et al.. **Economia monetária e financeira: teoria e política**. Rio de Janeiro: Elsevier-Campus, 2007.

CHEN, T.; GUESTRIN, C. XGBoost. Proceedings of the 22nd ACM SIGKDD **International Conference on Knowledge Discovery and Data Mining – KDD**, 2016.

CONFEDERAÇÃO NACIONAL DO COMÉRCIO DE BENS, SERVIÇOS E TURISMO (CNC). **Pesquisa de Endividamento e Inadimplência do Consumidor (Peic)**. Disponível em: <<http://stage.cnc.org.br/editorias/economia/pesquisas/pesquisa-de-endividamento-e-inadimplencia-do-consumidor-peic-marco-1>>. Acesso em junho 2021.

COX, D. R.; SNELL, E. J. **Analysis of Binary Data**. 2nd ed. London: Chapman & Hall, 1989.

CROOK, J. N.; EDELMAN, D. B.; THOMAS, L. C. Recent developments in consumer credit risk assessment. **European Journal of Operational Research**, v. 183, 2007.

CROUHY, Michel; GALAI, Dan; MARK, Robert. **Fundamentos da Gestão de Risco**. Editora Qualitymark, SERASA, 2008.

CVM, C. DE V. M. **Estrutura do Sistema Financeiro Nacional - SFN**. Disponível em: <https://www.investidor.gov.br/menu/Menu_Academico/O_Mercado_de_valores_mobiliarios_brasileiro/Estrutura_Funcionamento.html>. Acesso em: 30 ago. 2021.

FENSTERSTOCK, F. Credit scoring and the next step. **Business Credit**, New York, v. 107, n. 3, p. 46-49. 2005.

FERREIRA, M. A. M.; CELSO, A. S.; NETO, J. E. B. Aplicação do modelo logit binomial na análise de risco de crédito em uma instituição bancária. **Revista de Negócios**, Blumenau, v. 17, p. 41-59, 2012.

FORTI, M. **Técnicas de machine learning aplicadas na recuperação de crédito do mercado brasileiro**. Dissertação de Mestrado em Economia. Escola de Economia de São Paulo da Fundação Getúlio Vargas, 2018.

GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 2021.

GLADY, N; et al.; Modeling churn using customer lifetime value. **European Journal of Operational Research**, v. 197, n. 1, p. 402-411, 2009.

GOUVÊA, M. A.; GONÇALVES, E. B.; MANTOVANI, D. M. N. Análise de risco de crédito com aplicação de regressão logística e redes neurais. **Revista Contabilidade Vista & Revista**, Belo Horizonte, v. 24, p. 96-123, 2013.

GUÉGAN, D.; HASSANI, B. Regulatory learning: How to supervise machine learning models? An application to credit scoring. **The Journal of Finance and Data Science**, 2018. Disponível em <<https://doi.org/10.1016/j.jfds.2018.04.001>>. Acesso em maio 2021.

GROVER, P. Gradient Boosting from scratch. **Media**, 2017. <<https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d>>. Acesso em agosto 2021.

GUJARATI, Damodar N. **Econometria Básica**. Rio de Janeiro: Elsevier, 2006.

HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques**. Elsevier, 2011.

HAND, D. J.; HENLEY, W. E. Statistical classification methods in consumer credit scoring: a review. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, Wiley Online Library, 1997.

- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning**. New York: Springer, 2001.
- HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**. 2st ed. New York: John Wiley & Sons, 2000.
- KALECKI, Michal. **Teoria da dinâmica econômica**. São Paulo: Abril Cultural, 1978.
- KEYNES, J. M. **A Teoria Geral do Emprego, do Juro e da Moeda**. São Paulo: Atlas, 1992.
- KHATTREE, R. & NAIK, D.N. **Multivariate data reduction and discrimination with SAS software**. Cary, NC, USA: SAS Institute Inc., p. 558, 2000.
- KOTLER, Philip; SETIAWAN, Iwan. **Marketing 4.0: do tradicional ao digital**. Rio de Janeiro: Sextante, 2017.
- LAKATOS, E. M. de A.; MARCONI, M. de A. **Fundamentos da metodologia científica**. São Paulo: Atlas, 2017.
- LESSMANN, S. et al.; Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. **European Journal of Operational Research**, Elsevier, v. 247, n. 1, p. 124-136, 2015.
- LEVINE, R. Financial development and economic growth: views and agenda. **Journal of Economic Literature**, v. 35, p. 688–726, 1997.
- LEWIS, E. **An Introduction to Credit scoring**. Fair Isaac: San Rafael, California, 1992.
- LOBATO T.C.; CARVALHO B. N. Proposta de um modelo ensemble para credit scoring. **Brazilian Journal of Development**, v. 7, n. 3, p. 24280-24297. Curitiba, 2021.
- MANFIO, Fernando. **O risco nosso de cada dia: uma orientação objetiva para os profissionais da área**. Estação das Letras Editora, 2007.
- MARRA, V. N. **Previsão de dificuldades financeiras em empresas latino-americanas via aprendizagem de máquina**. 2019.
- MASSON, E.; WANG, Y.-J. Introduction to computation and learning in artificial neural networks. **European Journal of Operational Research**, v. 47, n.1, p. 1-28, 1990.
- MESQUITA, P. S. B. **Um Modelo de Regressão Logística para Avaliação dos Programas de Pós-Graduação no Brasil**. Universidade Estadual do Norte Fluminense, Campo dos Goytacazes, 2014.
- MITCHELL, T. M. **Machine learning**. New York: McGraw-Hill, 1997.
- MORA, Mônica. A Evolução do Crédito no Brasil entre 2003 e 2010. **Texto para discussão. Ipea**, 2015. Disponível em:

<<http://repositorio.ipea.gov.br/bitstream/11058/3537/1/td2022.pdf>>. Acesso em abril 2020.

MUNKHDALAI, L.; MUNKHDALAI, T.; NAMSRAI, O.-E.; LEE, J., & RYU, K. **An empirical comparison of machine-learning methods on bank client credit assessments**. Sustainability, 2019.

PEREIRA, O. M.; NESS JR., W. L. O modelo e-score de previsão de falências para empresas de internet. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 27., 2003, Atibaia. Anais... Atibaia: **ANPAD**, 2003.

PINDYCK, Robert S. e RUBINFELD, Daniel L. **Microeconomia**. São Paulo: Pearson, 2006.

REZENDE, Solange Oliveira. **Sistemas inteligentes: fundamentos e aplicações**. Barueri, SP: Manole, 2003.

RIBEIRO, C. F.; ZANI, J.; ZANINI, F. A. M. Estimação da probabilidade de inadimplência: uma verificação empírica na Universidade Católica de Pelotas. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 33., 2009, São Paulo. Anais... Rio de Janeiro: **ANPAD**, 2009.

ROCHMAN, R. R. **Crédito, Serviços e Captações**. In: Clovis de Faro. (Org.). Administração Bancária: Uma Visão Aplicada. Rio de Janeiro: FGV, 2014. p. 81-106.

ROESCH, Sylvia Maria Azevedo. **Projeto de estágio e de pesquisa em administração: guia para estágios, trabalhos de conclusão, dissertações e estudo de caso**. São Paulo: Atlas, 2005.

SANTOS, José Odálio dos. **Análise de crédito – empresas e pessoas físicas**. São Paulo: Atlas, 2003.

SANT'ANNA, A. A.; BORÇA JÚNIOR, G. R.; ARAUJO, P. Q. Mercado de crédito no Brasil: evolução recente e o papel do BNDES (2004-2008). **Revista do BNDES**, v. 16, n. 31, p. 41-60, jun. 2009.

SCHUMPETER, Joseph Alois. **Teoria do Desenvolvimento Econômico: uma investigação sobre lucros, capital, crédito juro e o ciclo econômico**. São Paulo: Abril Cultural, 1982.

SICSÚ, A. L. **Credit Scoring**. Desenvolvimento, implantação e acompanhamento. Editora Blucher, 2010.

SILVA, José Pereira da. **Gestão e análise de risco de crédito**. Editora São Paulo: Cengage Learning, 2016.

SUN J. et al.. Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates. **Information Sciences**, 2018.

TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introdução ao Data Mining: Mineração de Dados**. 1. Ed. Rio de Janeiro, Ciência Moderna, 2009.

THOMAS, L.; CROOK, J.; EDELMAN, D. **Credit Scoring and Its Applications**. [S.l.]: SIAM, 2017.

TRUETT, J.; CORNFIELD, J.; KANNEL, W. A multivariate analysis of the risk of coronary heart disease in framinghan. **Journal of Chronic Diseases**, v. 20, p. 511-524, 1967.

WEST, D. Neural network credit scoring models. **Computers & Operations Research**, v. 27, p. 1131-1152, 2000.

WICKSELL, K. **Lectures on political economy** – Vol. II: money. London: George Routledge & Sons, Ltd., 1935.

WOOLDRIDGE, Jeffrey M. **Introdução à Econometria: uma abordagem moderna**. São Paulo: Cengage Learning, 2008.

XIA Y. et al. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit risk. **Expert Systems with Applications**, v. 78, p. 225-241, 2017.

ZHANG, L.; PRIESTLEY, J.; NI, X. Influence of the event rate on discrimination abilities of bankruptcy prediction models. **International Journal of Database Management Systems**, v.10, p. 1-14, 2018.

ZHU, X. Semi-supervised learning literature survey. **Citeseer**, 2005.