

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CONTROLADORIA E CONTABILIDADE

ADRIANA BOITO JOHNSON

CREDIT SCORING EM INSTITUIÇÃO FINANCEIRA DIGITAL

Porto Alegre

2022

ADRIANA BOITO JOHNSON

CREDIT SCORING EM INSTITUIÇÃO FINANCEIRA DIGITAL

Dissertação submetida ao Programa de Pós-Graduação em Controladoria e Contabilidade da Faculdade de Ciências Econômicas da Universidade Federal do Rio Grande do Sul, como requisito para a obtenção do título de Mestre em Controladoria e Contabilidade.

Orientador: Prof. Dr. Marco Antônio Martins

Porto Alegre

2022

CIP - Catalogação na Publicação

Johnson, Adriana Lúcia Menegon
CREDIT SCORING EM INSTITUIÇÃO FINANCEIRA DIGITAL /
Adriana Lúcia Menegon Johnson. -- 2022.
90 f.
Orientador: Marco Antônio Martins.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Faculdade de Ciências Econômicas,
Programa de Pós-Graduação em Controladoria e
Contabilidade, Porto Alegre, BR-RS, 2022.

1. Risco de Crédito. 2. Credit Scoring. 3. Teoria
do Portfólio. 4. Modelo. 5. Instituição Financeira
Digital. I. Martins, Marco Antônio, orient. II.
Título.

ADRIANA BOITO JOHNSON

CREDIT SCORING EM INSTITUIÇÃO FINANCEIRA DIGITAL

Dissertação submetida ao Programa de Pós-Graduação em Controladoria e Contabilidade da Faculdade de Ciências Econômicas da Universidade Federal do Rio Grande do Sul, como requisito para a obtenção do título de Mestre em Controladoria e Contabilidade.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Marco Antônio Martins – Orientador

Universidade Federal do Rio Grande do Sul (UFRGS) – PPGCONT – UFRGS

Prof. Dr. Everton Farias

Universidade Federal do Rio Grande do Sul (UFRGS) – PPGCONT - UFRGS

Profa. Dra. Carla Marcolin

Universidade Federal do Rio Grande do Sul (UFRGS) – PPGCONT – UFRGS

Prof. Dra. Frederike Monika Butter Mette

Pontifícia Universidade Católica – PPGA – PUCRS

Para meu marido.
Meu amor e companheiro que me apoiou do início
ao fim.

RESUMO

Boito, A. L. M. (2022). *Credit Scoring em Instituição Financeira Digital*. [Dissertação de mestrado, Programa de Pós-Graduação em Controladoria e Contabilidade, Universidade Federal do Rio Grande do Sul].

O setor bancário tem função peculiar para a economia, principalmente pelo seu papel de instituição depositária e de fomento, e pela concessão de crédito. Os bancos digitais oferecem um relacionamento diferenciado e remoto, com custos mais baixos de serviços e concessões de crédito via internet e mobile banking. Assim, os bancos precisam gerenciar o risco de crédito inerente às suas operações. Para isto, o modelo de *credit scoring* é uma das ferramentas mais importantes para a sustentabilidade dos sistemas bancários. Considerando o exposto e a velocidade com que o crédito ao consumidor pode se deteriorar, o presente estudo analisa a aplicação de um modelo de *credit scoring* de um banco tradicional no banco digital X. Por meio de uma pesquisa documental, qualitativa e quantitativa em relação ao problema, e descritiva quanto aos objetivos, foram descritas as características de banco digital e risco de crédito, foram apresentadas as normas regulatórias e revisados os principais modelos de risco de crédito. Também, foi estimado um modelo de *credit scoring* para previsão de inadimplência do banco digital X. Os resultados demonstram que a regressão logística binária sobre a base de dados mais atual apresentou o maior acerto de classificação de adimplentes e inadimplentes, classificando corretamente 81,5% dos casos. As variáveis com maior poder preditivo foram Sexo, saldo em Operações não Consignadas, soma das parcelas das Operações não Consignadas e soma do Valor Bruto das Operações Consignadas e não Consignadas. Portanto, é necessário um acompanhamento contínuo das variáveis predictoras da inadimplência, pois a carteira de crédito passa por diversas alterações de perfil ao longo do tempo. Este estudo contribuiu trazendo dados atualizados sobre os temas estudados e demonstrou a importância da gestão contínua do risco de crédito para a liberação de capital, por meio de um provisionamento menor. Para pesquisas futuras, sugere-se a utilização de novas variáveis e datas base e a aplicação do modelo em outras carteiras da própria instituição e de outros bancos digitais. Como limitação, a dificuldade de acesso aos modelos de risco de crédito dos bancos digitais, por tratar-se de informação estratégica e sigilosa.

Palavras-chave: Risco de Crédito. *Credit Scoring*. Modelo. Teoria do Portfólio. Instituição Financeira Digital.

ABSTRACT

The banking sector has a peculiar function for the economy, mainly due to its role as a depository and development institution, and the credit granting. Digital banks offer a differentiated and remote relationship, with lower service costs and credit concessions via the internet. For that reason, banks need to manage the credit risk inherent in the whole portfolio and the risk of individual credits. Credit scoring model is one of the most important tools for the sustainability of the banking systems. Considering what was previously mentioned and the speed in which a credit to a consumer can deteriorate, this study analyses the application of a credit scoring model of a traditional bank in Digital Bank X. Through documental research, both qualitative and quantitative, the regulatory standards were presented, and the main risk credit models were reviewed. In addition, a credit scoring model to predict defaults for Digital Bank X was estimated. The results show that the application of binary logistic regression over the most up-to-date database presented the best classification of defaulters and non-defaulters. 81.5% of the cases were correctly classified. The variables with higher predictive power were Sex, balance in Non-Payroll Operations, sum of the instalments of Non-Consigned Operations, sum of the Gross Value of Consigned and Non-Consigned Operations. Ergo, a continuous monitoring of the predictive variables of default is necessary, since the credit portfolio undergoes several profile changes over time. This study contributed by bringing up-to-date data on the topics studied and demonstrated the importance of continuous management of credit risk, with the use of predictor variables of default, for the release of capital through a lower provisioning. For future research, we suggest the use of new variables and base dates and the application of the model in other portfolios of the institution itself and other digital banks. As a limitation, the difficulty of accessing the credit risk models of digital banks, as they are strategic and confidential information.

Keywords: Credit Risk. Credit Scoring. Model. Portfolio Theory. Digital Financial Institution.

LISTA DE SIGLAS

BACEN	Banco Central do Brasil
BCBS	Basel Committee on Banking Supervision
BIS	Bank for International Settlement
CMN	Conselho Monetário Nacional
CR	<i>Capital Requirement</i>
CSFP	Credit Suisse Financial Products
DD	<i>Distance to default</i>
DoS	Disk operating System
EAD	<i>Amount of exposure at default</i>
EDF	<i>Expected default risk frequency</i>
FEBRABAN	Federação Brasileira de Bancos
FMI	Fundo Monetário Internacional
GARCH	<i>Generalized AutoRegressive Conditional Heteroskedasticity</i>
IBM	International Business Machines
IFRS	International Financial Reporting Standards
IRB	<i>Internal Rating-Based</i>
IRRBB	<i>Interest Rate Risk in the Banking Book</i>
LGD	<i>Loss given default</i>
LOGIT	Regressão logística
M	<i>Maturity</i>
MMR	<i>Marginal mortality rate</i>
OCC	Office of the Comptroller of the Currency
PCLD	Provisão para Créditos de Liquidação Duvidosa
PD	<i>Probability of default</i>
PECLD	Perda Estimada com Créditos de Liquidação Duvidosa
PIB	Produto Interno Bruto
PMT	<i>Payment</i>
PR	Patrimônio de Referência
RAROC	<i>Risk-adjusted return on capital</i>

RAS	Declaração de Apetite por Riscos
ROA	<i>Return on Assets</i>
ROC	<i>Receiver operating characteristic</i>
RWA	<i>Risk-Weighted Assets</i>
S1	Segmento 1
S2	Segmento 2
S3	Segmento 3
S4	Segmento 4
S5	Segmento 5
SA	<i>Standardized Approach</i>
S&P 100	Índice com 100 empresas selecionadas pela Standard and Poor's
SFN	Sistema Financeiro Nacional
SPSS	<i>Statistical Package for the Social Science</i>
VaR	<i>Value at Risk</i>

LISTA DE TABELAS

Tabela 1 – Variável Dependente.....	43
Tabela 2 – Variáveis Independentes	43
Tabela 3 – Tabela de Classificação ^{a,b}	57
Tabela 4 – Variáveis não presentes na equação	58
Tabela 5 – Testes de Omnibus do Modelo de Coeficiente.....	59
Tabela 6 – Resumo do modelo	59
Tabela 7 – Teste de Hosmer e Lemeshow.....	60
Tabela 8 – Tabela de Classificação ^a	60
Tabela 9 – Variáveis na equação	61
Tabela 10 – Variáveis não presentes na equação.....	62
Tabela 11 – Testes de Omnibus do Modelo de Coeficiente.....	63
Tabela 12 – Resumo do modelo	63
Tabela 13 – Teste de Hosmer e Lemeshow.....	64
Tabela 14 – Tabela de Classificação ^a	64
Tabela 15 – Variáveis na equação	65
Tabela 16 – Variáveis não presentes na equação.....	66
Tabela 17 – Testes de Omnibus do Modelo de Coeficiente.....	67
Tabela 18 – Resumo do modelo	68
Tabela 19 – Teste de Hosmer e Lemeshow.....	68
Tabela 20 – Tabela de Classificação ^a	69
Tabela 21 – Variáveis na equação	70
Tabela 22 – Variáveis não presentes na equação.....	71
Tabela 23 – Testes de Omnibus do Modelo de Coeficiente.....	72
Tabela 24 – Resumo do modelo	72
Tabela 25 – Teste de Hosmer e Lemeshow.....	73
Tabela 26 – Tabela de Classificação ^a	73
Tabela 27 – Variáveis na equação	74
Tabela 28 – Resumo dos Resultados Principais	75

SUMÁRIO

1 INTRODUÇÃO.....	10
1.1 PROBLEMA DE PESQUISA	12
1.2 OBJETIVOS	13
1.2.1 Objetivo Geral.....	13
1.2.2 Objetivos Específicos.....	14
1.3 JUSTIFICATIVA DO ESTUDO.....	14
1.4 DELIMITAÇÃO DO ESTUDO	15
2 REFERENCIAL TEÓRICO	17
2.1 TEORIA DO PORTFÓLIO	17
2.2 INSTITUIÇÕES FINANCEIRAS	18
2.3 RISCO	20
2.3.1 Risco de Crédito	22
2.3.2 Modelos de Gestão de Risco de Crédito.....	23
2.3.2.1 Credit Scoring.....	32
2.4 ADEQUAÇÃO DE CAPITAL	35
2.5 ESTUDOS RELACIONADOS	37
3 PROCEDIMENTOS METODOLÓGICOS.....	41
3.1 CLASSIFICAÇÃO METODOLÓGICA DO ESTUDO.....	41
3.2 POPULAÇÃO E AMOSTRA	42
3.3 VARIÁVEIS DA PESQUISA	42
3.4 COLETA DOS DADOS.....	44
3.5 ANÁLISE DOS DADOS	44
3.6 DESCRIÇÃO DO MODELO	45
3.7 ESTIMAÇÃO DO MODELO	46
4 RESULTADOS	50
4.1 INSTITUIÇÕES FINANCEIRAS DIGITAIS	50
4.2 RISCO	51
4.2.1 Risco de Crédito	54
4.3 MODELO DE <i>CREDIT SCORING</i>	56
4.3.1 Apresentação dos Resultados das Regressões	57
4.3.2 Discussão.....	75
5 CONSIDERAÇÕES FINAIS.....	79
REFERÊNCIAS.....	82
APÊNDICE A - TERMOS DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA PARTICIPAÇÃO EM PESQUISA – GESTOR DE RISCOS.....	90

1 INTRODUÇÃO

O setor bancário desempenha função diferenciada no funcionamento da economia (Joaquim, Van Doornik, & Ornelas, 2019), principalmente pelo seu papel de instituição depositária e de fomento, e pela concessão de crédito (Yanenkova et al., 2021), tanto a empresas quanto a indivíduos (Lobo, 2017). As instituições financeiras, comparadas a outros setores, possuem peculiaridades como alto grau de alavancagem, robusta estrutura de governança, complexidade de seus produtos e operações, e alto nível de regulação por parte do governo (Lobo, 2017). Adicionalmente, Niemand et al. (2017) apontam que os bancos, em particular, têm avançado cada vez mais na geração de valor online e digitalizado.

A digitalização dos serviços bancários se deu pela necessidade de desburocratização dos grandes bancos e tem fornecido um aprimoramento da experiência do cliente, proporcionando acesso a mais segurança, agilidade e transparência nas operações (Banco Central do Brasil [BACEN], 2019). As tecnologias digitais permitem novos modelos de negócios, nos quais as pessoas buscam transações de empréstimos por meio de plataformas online, alterando o comportamento dos bancos e instituições de crédito (Niemand et al., 2017).

Isto é percebido pelo aumento expressivo das operações de contratação de crédito no *Mobile Banking*¹, que se consolidou como o principal meio de relacionamento dos bancos com seus clientes, representando mais da metade das transações bancárias, conforme Pesquisa da Federação Brasileira de Bancos (FEBRABAN) de Tecnologia Bancária (2021). Ainda, de acordo com a referida pesquisa, as prioridades estratégicas dos bancos são a expansão de novas tecnologias e o lançamento de novos produtos, tendo destaque para a expansão do crédito, dos meios de pagamento e abertura de novas contas.

Entretanto, deve-se considerar que o risco faz parte das transações das instituições financeiras, estando presente desde a intermediação até o processamento de suas rotinas operacionais (Borges, 2010). Conforme os referidos autores, com a integração dos serviços financeiros, as perdas podem alcançar todo o sistema bancário, pois o setor está suscetível a problemas de desvirtuamento de interesses, falhas e incorreções. Além disso, com o banco digital, o aumento da distância entre bancos e clientes pode levar a falta de confiança e maiores preocupações com segurança (Chaimaa, Najib, & Rachid, 2021).

¹ Aumento de 44% no *Mobile Banking* de 2019 para 2020 e diminuição de 11% no *Internet Banking* no mesmo período.

Com o avanço tecnológico, a intermediação financeira tem mudado consideravelmente ao longo das últimas duas décadas (Giudici, Hadji-Misheva, & Spelta, 2019). Os autores complementam que uma das mudanças mais significativas foi o surgimento das Fintechs, que juntamente com os bancos totalmente digitais, alteraram serviços e produtos financeiros, processos produtivos e a estrutura organizacional das instituições financeiras. Entretanto, apesar do surgimento de novas tecnologias, o modelo de pontuação de crédito, ou *credit scoring*, ainda é uma das ferramentas de tomada de decisão mais importantes para a sustentabilidade dos sistemas bancários (Kyeong, Kim, & Shin, 2022).

A grande quantidade de decisões envolvidas na concessão de crédito ao consumidor torna necessária a confiança em modelos e algoritmos ao invés da dependência somente no julgamento humano, baseando as decisões algorítmicas em informações consistentes e confiáveis. Por esse motivo, as instituições e agências de crédito têm criado seus próprios modelos de risco personalizados com base em informações privadas sobre os mutuários (Khandani, Kim, & Lo, 2010).

É válido mencionar que, tradicionalmente, o processo de gerenciamento de risco nas instituições financeiras é essencial para a sua sobrevivência de longo prazo. Os bancos precisam gerenciar o risco de mercado, de liquidez, risco operacional, legal e, principalmente, o risco de crédito. Isto ocorre pois, além das operações de crédito representarem a sua principal fonte de receita, eventuais perdas não esperadas em operações de crédito podem comprometer a credibilidade da instituição, podendo contaminar o sistema financeiro (Mishkin, 1992).

O autor prossegue evidenciando a importância do gerenciamento do risco de crédito, pois grande parte das crises financeiras começam com um aumento na expansão de liquidez e de concessão de crédito, com o conseqüente aumento de risco dos ativos bancários e quebras de bancos, contagiando todo o sistema. Dentro desta realidade, o sistema financeiro busca incessantemente aperfeiçoar os mecanismos de gerenciamento de risco e de previsão de inadimplência, em um trabalho conjunto entre reguladores e participantes do mercado.

Isto se torna ainda mais relevante quando considerado o saldo das operações de crédito do Sistema Financeiro Nacional (SFN), segundo dados do Banco Central do Brasil (BACEN), que totalizou R\$ 4,7 trilhões em janeiro de 2022 (BACEN, 2022). Ainda, enquanto as operações com pessoas jurídicas decresceram para 9,8%, em comparação à 10,5% no mês anterior, o crédito às pessoas físicas seguiu aumentando, com alta de 21,5% nos doze meses até janeiro, se comparado a 21% em dezembro de 2021 (BACEN, 2022).

A evolução do crédito é influenciada pelo cenário macroeconômico existente e depende do apetite ao risco das instituições financeiras e das suas expectativas quanto a ele no horizonte temporal. O BACEN busca influenciar essas expectativas, tomando medidas para manter a fluidez do crédito e a liquidez do mercado, beneficiando pessoas físicas e jurídicas com aumento nas concessões com taxas de juros mais baixas (BACEN, 2021). Vale ressaltar que, em 2020, para enfrentar os efeitos da pandemia de Covid-19, o BACEN e o Conselho Monetário Nacional (CMN) tomaram medidas adicionais para assegurar o bom nível de liquidez do sistema financeiro e dar vazão ao crédito (BACEN, 2021). O crédito para pessoas físicas, o qual é o foco deste estudo, teve crescimento de 11,2% de 2019 para 2020, com uma taxa de inadimplência de 2,9% (BACEN, 2020).

Dada a expressividade das concessões de crédito e sua importância para o desenvolvimento econômico, os bancos são obrigados a identificar as transações que os expõem ao risco de crédito da contraparte e calcular o seu encargo (Basel Committee on Banking Supervision [BCBS], 2020). Para isso, conforme Gutiérrez-Nieto, Serrano-Cinca e Camón-Cala (2016), as instituições financeiras usam a pontuação de crédito, ou *credit scoring*, para distinguir entre bons e maus tomadores, sendo que um bom tomador de crédito é considerado aquele que paga seus empréstimos.

Assim, os bancos precisam gerenciar o risco de crédito inerente a toda a carteira, bem como o risco em créditos ou transações individuais, considerando as relações com outros riscos (BCBS, 2020). Com base na Teoria do Portfólio, a qual fundamenta este estudo, se todo o risco é diversificável e os custos de monitoramento são fixos, a intermediação financeira só é viável para uma carteira bem diversificada de devedores, pois o intermediário é mais eficiente quanto maior ele for (Diamond, 1984). Assim, quanto mais eficiente for a alocação da carteira, menor tende a ser o risco de crédito.

1.1 PROBLEMA DE PESQUISA

Os problemas bancários têm diversas causas, sendo as principais relacionadas aos padrões de crédito menos conservadores para tomadores e contrapartes, gestão de risco de portfólio deficiente ou falta de atenção às mudanças na economia ou outras situações que podem conduzir a uma deterioração na qualidade de crédito das contrapartes de um banco (BCBS, 2000). Ainda, o uso do crédito sem planejamento ou desconsiderando a capacidade de pagamento do tomador, inclusive de forma excessiva, pode gerar consequências negativas para os cidadãos e para a economia (BACEN, 2019).

A crise financeira de 2007 a 2009 destacou a importância da gestão de risco nas instituições financeiras, sendo que muitas das suas políticas permanecem inacessíveis e sem eficácia comprovada. A implementação eficaz de estratégias de gerenciamento de risco requer que os bancos sejam capazes de identificar contas com probabilidade de inadimplência (Butaru, Chen, Clark, Das, Lo, & Siddique, 2016), o que leva a um maior controle sobre o risco de crédito.

Também, desde o início de 2020, os desafios na modelagem de crédito aumentaram consideravelmente com a pandemia de Covid-19, devido a intervenções de governos e bancos centrais em todo o mundo. As relações entre as medidas de risco de crédito e os fatores macroeconômicos que vinham sendo consideradas estáveis, sofreram alterações neste período (Engelmann, 2021).

Ademais, a possibilidade de que problemas em uma instituição possam se espalhar e perturbar o funcionamento normal de todo o sistema financeiro, remete à importância da regulamentação de capital, a qual foi reforçada pelo Acordo de Basileia II, quanto ao risco de crédito. Essa regulamentação fornece uma fonte de absorção de perdas para eventos inesperados e introduz incentivos para os bancos limitarem o risco de suas atividades. Porém, apresenta impacto no retorno sobre o patrimônio, influenciando a competitividade do setor financeiro (Antão, & Lacerda, 2011).

Considerando a contextualização apresentada, referente à crescente expansão do crédito e de operações bancárias digitais, juntamente com a constante necessidade de gestão do risco de crédito, este estudo apresenta a seguinte questão de pesquisa: *Qual a aderência de um modelo de credit scoring de um banco tradicional a um banco digital?*

1.2 OBJETIVOS

Levando em consideração a problemática apresentada, em um contexto de expansão da concessão de crédito, aumento de operações digitais dos bancos e da necessidade de gestão do risco de crédito dessas instituições financeiras digitais, essa seção apresenta os objetivos para responder à questão de pesquisa.

1.2.1 Objetivo Geral

O presente estudo tem como objetivo geral analisar a aplicação de um modelo de *credit scoring* de um banco tradicional no banco digital X.

1.2.2 Objetivos Específicos

Os objetivos específicos deste estudo, para atingir o objetivo geral e responder à questão problema, são:

- a) descrever as principais características de instituição financeira digital e de risco de crédito;
- b) descrever os principais modelos de risco de crédito existentes na literatura;
- c) apresentar as normas do Banco Central do Brasil quanto ao risco de crédito;
- d) aplicar o modelo de *credit scoring* utilizado para a avaliação de risco de crédito de um banco tradicional no banco digital X, considerando os parâmetros do BACEN.

1.3 JUSTIFICATIVA DO ESTUDO

Com a digitalização nos mercados financeiros, houve um desenvolvimento tecnológico sem precedentes e um aumento do número e variedade de novas metodologias estatísticas aplicadas ao setor (Bitetto et al., 2021). Por esse motivo, a gestão do risco de crédito visa reduzir os efeitos colaterais do empréstimo entre os seus recebedores. Um modelo de risco de crédito é um instrumento de gestão dos bancos para classificação objetiva das contrapartes com base em dados quantitativos e qualitativos (Zarei, 2016). O risco de crédito é, portanto, um indicador chave do desempenho e da estabilidade financeira dos bancos, pelo qual o Fundo Monetário Internacional (FMI) e o Banco Mundial avaliam a saúde do setor bancário (Saâda, & Gafsi, 2019).

Segundo Khandani, Kim e Lo (2010), considerando a velocidade com que o crédito ao consumidor pode se deteriorar, há uma clara necessidade de medidas mais oportunas de risco de crédito por parte dos bancos e reguladores. Com isso, é necessário que sejam realizados o acompanhamento e a gestão do risco de crédito para as instituições financeiras digitais. Assim, é importante que os bancos levem em consideração a necessidade de identificar, mensurar, monitorar e controlar o risco de crédito, bem como determinar se possuem capital adequado contra esses riscos e se são suficientemente compensados pelos riscos incorridos (BCBS, 2000).

A gestão eficaz do risco de crédito é um componente crítico de uma abordagem abrangente da gestão de risco e essencial para o sucesso a longo prazo de qualquer organização bancária (BCBS, 2000) e pode servir como motor de crescimento para a

organização, oferecendo uma vantagem competitiva na disputa pelo cliente bom pagador (Nehrebecka, 2021). Assim, a presente pesquisa contribui com estudos relacionados à gestão de risco de crédito de instituições financeiras digitais, complementando a literatura sobre o tema e dando suporte à aplicação de modelos de gestão de risco de crédito para uma alocação de capital mais eficiente nas referidas instituições.

O estudo aplica um modelo de *credit scoring* para dar suporte à gestão da inadimplência, referente à concessão de crédito pessoal, diminuindo o risco de crédito envolvido. Isso faz com que diminua a necessidade de provisionamento, com uma maior liberação de capital para os bancos digitais serem capazes de realizar a sua alocação com mais eficiência, proporcionando a possibilidade de maiores concessões de crédito, o que fomenta a economia e traz maiores benefícios à sociedade. Um aumento no volume de empréstimos contribui com o setor produtivo, com as necessidades crescentes dos consumidores e com o crescimento socioeconômico, além de ser uma fonte considerável de rentabilidade para os bancos (Yanenkova et al., 2021).

Adicionalmente, este estudo é relevante ao considerar a representatividade dos clientes pessoa física na margem de crédito líquida de despesas de Provisão para Créditos de Liquidação Duvidosa (PCLD) do resultado de intermediação financeira, a qual foi de 76%, levando em conta o total da carteira de crédito do sistema bancário em 2020. Ainda, nos produtos voltados a este público, as operações de empréstimo não consignado, cheque especial e de cartão de crédito representaram cerca de 35% da margem líquida de crédito (BACEN, 2020). A partir desses dados do Banco Central brasileiro, fica clara a necessidade da gestão de risco de crédito pessoal, por ser um dos produtos mais comercializados pelos bancos.

1.4 DELIMITAÇÃO DO ESTUDO

A presente pesquisa delimita-se ao estudo do risco de crédito de um banco digital do segmento 4 (S4) e à análise da aplicação de um modelo de *credit scoring* empregado na avaliação de risco de crédito de um banco tradicional, no banco digital X deste segmento. Os segmentos S1 e S2 são compostos por bancos múltiplos, comerciais, de investimento, de câmbio e caixas econômicas, com porte superior a 1% do PIB brasileiro. O S3 é composto pelas demais instituições financeiras com PIB inferior a 1% e superior a 0,1%. O S4 e o S5 são compostos pelas instituições de porte inferior a 0,1%.

Entretanto, o S4 se diferencia do S5, conforme Resolução n. 4.553 de 2017, pois este último possui as instituições que utilizam metodologia facultativa simplificada para apuração dos requerimentos mínimos de Patrimônio de Referência (PR), de Nível I e de Capital Principal, exceto bancos múltiplos, bancos comerciais, de investimento, de câmbio e caixas econômicas e pelas instituições não sujeitas a apuração de PR. Portanto, este estudo engloba os bancos digitais em um segmento intermediário, os quais não são os maiores, em termos de percentual do PIB e nem os bancos com menos obrigatoriedade quanto aos requerimentos relativos à gestão dos níveis de risco.

2 REFERENCIAL TEÓRICO

Como estrutura deste referencial, são apresentadas a teoria de base, as características de instituições financeiras digitais e de risco de crédito, e os principais modelos de gestão de risco de crédito existentes na literatura, com o apoio de estudos relacionados ao tema.

2.1 TEORIA DO PORTFÓLIO

O retorno esperado de uma carteira é uma média ponderada dos retornos esperados sobre os títulos individuais, conforme Markowitz (1952). A variância do retorno da carteira, segundo o autor, é uma função particular das variâncias e das covariâncias entre títulos e seus pesos na carteira. Para diminuir a variação, deve-se evitar investir em títulos com covariâncias elevadas entre si. Além disso, de acordo com Markowitz (1952), deve-se diversificar entre os setores, pois os que têm características econômicas diferentes, têm covariâncias mais baixas do que as empresas de um mesmo setor.

O processo de seleção de um portfólio pode ser dividido em observação e experiência para obter crenças sobre o desempenho futuro dos títulos disponíveis e com as crenças relevantes, escolher o portfólio (Markowitz, 1952). Também, a aplicação de uma grande variedade de modelos de portfólio reduz a questão do risco de estimativa nos parâmetros de entrada do processo de seleção de portfólio (Platanakis, & Urquhart, 2020). Segundo Stiroh (2006), a visão do portfólio fornece uma estrutura familiar para interpretar os resultados de diferentes escolhas estratégicas das empresas. Além disso, um portfólio de crédito de baixa qualidade pode aumentar a probabilidade de inadimplência para os bancos (Sclip, Girardone, & Miani, 2019).

Assim, a Teoria do Portfólio é, em essência, uma teoria de investimento que considera os efeitos da diversificação quando os riscos são correlacionados, distinguindo entre portfólios eficientes e ineficientes e analisando *trade-offs* de risco e retorno na carteira como um todo (Markowitz, 1999). Adicionalmente, Markowitz (2014) afirma que, por mais de meio século, as pesquisas têm apoiado essa teoria. Também, por meio da diversificação, os bancos podem explorar receitas e economias de custo, de acordo com Calmès e Théoret (2021). A diversificação, portanto, com a combinação de diferentes ativos, permite obter uma variação menor, ou seja, um menor risco e uma maior eficiência de alocação de recursos. Em relação ao risco de crédito, um portfólio de crédito de alta qualidade pode diminuir a inadimplência para os concessionários.

2.2 INSTITUIÇÕES FINANCEIRAS

As instituições financeiras desempenham uma função fundamental para a economia, a de canalizar fundos provenientes de fornecedores, ou seja, daqueles com fundos excedentes, para os tomadores, aqueles com escassez de fundos. Devido aos riscos em geral, principalmente risco de liquidez e de precificação, e custos de monitoramento, os fornecedores de fundos preferem manter os direitos financeiros emitidos por instituições financeiras, ao invés dos emitidos por usuários finais (Saunders, & Cornett, 2014).

As instituições financeiras beneficiam a sociedade de diversas maneiras, fazendo a transmissão da oferta de moeda, com a veiculação das políticas monetárias, com o fornecimento de serviços de pagamento, de alocação de crédito, representando uma fonte de financiamento para diversos setores da economia, e pela transferência de riqueza entre gerações, pela qual os poupadores transferem suas economias aos seus sucessores. Estas instituições podem ser de diversos tipos, como bancos e empresas financeiras.

Os bancos comerciais são instituições depositárias, porque a maior parte de seus fundos vem de depósitos de clientes, representando o maior grupo dessas instituições pelo tamanho dos ativos. Seus principais ativos são empréstimos e seus principais passivos são depósitos. Desempenham funções semelhantes às das instituições de poupança e cooperativas de crédito, pois aceitam depósitos e fazem empréstimos, mas se distinguem delas pelo tamanho e composição dessas operações, além de serem regulamentados separadamente. Seus empréstimos são mais abrangentes, incluindo empréstimos ao consumidor, comerciais, internacionais e imobiliários (Saunders, & Cornett, 2014).

Os depósitos são um componente significativo da oferta de moeda, fazendo com que os bancos comerciais tenham um papel fundamental na transmissão da política monetária do banco central para o resto da economia. Os bancos comerciais são importantes, também, pela eficiência com que fornecem serviços de pagamento e serviços de intermediação de prazos, o que beneficia diretamente a economia (Saunders, & Cornett, 2014).

Enquanto o banco comercial tem as atividades de recepção de depósitos e empréstimos comerciais, o banco de investimento tem as atividades de subscrição, emissão e distribuição de títulos por meio de colocação pública ou privada (Saunders, & Cornett, 2014). Um banco múltiplo, por sua vez, pode operar com no mínimo duas carteiras operacionais, sendo uma delas comercial ou de investimento (BACEN, 2019).

As empresas financeiras são intermediárias financeiras que concedem empréstimos a pessoas físicas e jurídicas. Ao contrário dos bancos mencionados anteriormente, estas não

aceitam depósitos, mas dependem de dívidas de curto e longo prazo para financiamento. Entretanto, ambos são regulamentados para evitar falhas de mercado e os seus custos à economia e à sociedade em geral. Apesar da regulamentação trazer benefícios para as partes envolvidas, também impõe custos privados, os quais oneram todo o sistema.

As instituições financeiras estão expostas a diversos tipos de risco, estando sujeitas principalmente a inadimplência ou risco de crédito. Também, estão expostas ao risco cambial e ao risco-país ou soberano, à medida que expandem seus serviços para clientes estrangeiros ou com operações no exterior. Enfrentam, ainda, os riscos de mercado, liquidez, das taxas de juros, operacional e tecnológico (Saunders, & Cornett, 2014). A revolução digital tem sido relevante para o setor bancário, pois com o surgimento do banco digital, tanto online quanto móvel, este se tornou um dos canais mais estratégicos utilizados pelos clientes bancários para a realização de suas transações financeiras (Carbo-Valverde et al., 2019).

O banco digital teve origem como resultado da globalização, da concorrência e do rápido crescimento dos sistemas de TI (Chaimaa, Najib, & Rachid, 2021). Para os usuários, conforme Saputra e Chaerani, (2022), o banco digital torna mais rápida e fácil a realização de transações financeiras, sendo conveniente em termos de custo, tempo e lugar, pois os clientes não precisam mais se deslocar para as agências físicas. Ainda, segundo os autores, para as instituições financeiras, além de os serviços digitais reduzirem os custos operacionais se comparados aos custos dos bancos tradicionais, os quais precisam arcar com a manutenção das agências físicas, também aumentam as receitas com tarifas.

As tecnologias digitais, como inteligência artificial, *machine learning* e as interfaces de programação de aplicativos (APIs), estão permitindo que os bancos sejam mais ágeis na conexão e colaboração entre si para oferecer inovações ao mercado, atendendo às necessidades de seus clientes (Mishra, 2020). Assim, o banco digital permite aos clientes realizar transações bancárias, obter e comunicar informações por meio da Internet, representando a transformação das atividades e serviços bancários tradicionais em um ambiente digital (Saputra, & Chaerani, 2022).

A partir do exposto, fica evidenciada a importância do setor bancário para a economia e a transformação pela qual ele vem passando, se tornando cada vez mais digital e conectado ao dia a dia das pessoas. Com isso, o risco se faz cada vez mais presente nas operações bancárias. Tanto o risco de crédito, pelas concessões que podem ser feitas em qualquer lugar e em qualquer momento, como os demais riscos, havendo um grande destaque para o risco cibernético.

2.3 RISCO

O estudo de risco, o qual vem de muitos séculos, teve origem na economia, visando definir o conceito de risco e entender a sua aversão, pois faz parte de todo o empreendimento humano (Damodaran, 2007). O risco pode ser criado pelo ser humano, como nas guerras, nas mudanças nas políticas governamentais, com a inflação e em ciclos de negócios. O risco também ocorre como resultado de fenômenos naturais imprevistos ou de inovações tecnológicas, que podem tornar a tecnologia existente obsoleta e criar alterações nos níveis de emprego. Contudo, o risco e a vontade de correr riscos são essenciais para o crescimento da economia (Jorion, 2007).

As primeiras tentativas de mensurar o risco e expressar as perdas potenciais na carteira, são atribuídas a Francis Edgeworth e remontam a 1888. Ele fez contribuições importantes para a teoria estatística, defendendo o uso de dados de experiências anteriores como base para estimar as probabilidades futuras (Adamko et al., 2015). Define-se risco financeiro como a volatilidade de resultados inesperados, que podem representar ganhos, o valor dos ativos ou o patrimônio líquido (Damodaran; Jorion, 2007).

Partindo da simbologia chinesa, o risco é uma combinação de perigo e oportunidade. A minimização da exposição ao risco, reduz o potencial de oportunidade. Assim, para obter grandes retornos, deve-se estar disposto a uma exposição considerável aos riscos (Damodaran, 2007). As empresas estão expostas a vários tipos de riscos, que podem ser comerciais, como os riscos do negócio, e financeiros. Dentre os riscos financeiros, estão os riscos de mercado, riscos de liquidez, riscos operacionais e risco de crédito (Jorion, 2007).

Risco de mercado é o risco de perdas devido a movimentos no nível ou volatilidade dos preços de mercado. Pode representar um risco absoluto, medido em moeda relevante, o qual se concentra na volatilidade dos retornos totais, ou um risco relativo, em relação a um índice de referência, que mede o risco em termos de erro de rastreamento ou desvio do índice. Risco de liquidez é o risco de uma transação não ser realizada aos preços de mercado vigentes, devido ao tamanho da posição em relação aos lotes normais de negociação. Este risco varia entre categorias de ativos e ao longo do tempo em função das condições de mercado (Jorion, 2007).

O risco operacional é o risco de perda resultante de processos internos, sistemas inadequados ou pessoas e suas falhas, ou de eventos externos. Inclui o risco legal, decorrente da exposição a penalidades, multas ou danos punitivos de ações de fiscalização e de acordos privados (Jorion, 2007). O risco de tecnologia, o qual é um tipo de risco operacional, ocorre

quando os investimentos tecnológicos não produzem as economias de custo previstas na forma de escopo ou de escala, resultando em grandes perdas na eficiência competitiva e em fracasso a longo prazo. A inovação tecnológica tem sido uma grande preocupação das instituições financeiras, pois tem como objetivo a redução de custos operacionais, o aumento de lucros e a captação de novos mercados (Saunders, & Cornett, 2014).

O risco cibernético, o qual está intimamente ligado ao avanço da tecnologia, é o risco causado por uma ameaça cibernética. Como o ciberespaço tem um alcance bastante amplo, as possíveis fontes de ameaças podem residir em qualquer lugar do mundo e possuem um potencial de causar danos nas profundezas de qualquer sistema cibernético. Exemplos de riscos cibernéticos são violações de confidencialidade devido a ataques de vírus via ciberespaço e perda de disponibilidade devido a ataques DoS (Refsdal, Solhaug, & Stølen, 2015).

O risco de crédito é o risco de perdas devido ao fato de as contrapartes não quererem ou não conseguirem cumprir as suas obrigações contratuais, cujo efeito é medido pelo custo de reposição dos fluxos de caixa caso a outra parte decreta falência. Essa perda abrange a exposição, ou valor em risco, e a taxa de recuperação, que é a proporção paga de volta ao credor (Jorion, 2007). Devido à sua importância ao setor bancário e ser o foco deste estudo, o risco de crédito será explorado na próxima subseção.

A alta administração deve decidir qual o perfil de risco da empresa, levando em consideração a realidade dos negócios em que atua e, juntamente com o conselho, deve articular seu apetite pelo risco e as suas principais restrições e limites. A gestão de riscos deve ser eficaz, eficiente e deve contribuir de forma decisiva para o ambiente de controle dentro do negócio (Caouette, Altman, Narayanan, & Nimmo, 2008). Não é o processo de controlar e reduzir as perdas esperadas, mas o de compreensão, cálculo de custos e gerenciamento eficiente de níveis inesperados de variabilidade nos resultados financeiros de um negócio (Crouhy, Galai, & Mark, 2005).

A avaliação de risco bancário tradicional se concentra nos fatores quantitativos e qualitativos de um banco, conforme a sua adequação de capital, qualidade de ativos, gestão, lucratividade, liquidez e gestão de fundos (Liao, Chen, & Lu, 2009). Assim, para que os negócios sejam bem-sucedidos, precisam de uma visão completa do gerenciamento de riscos, considerando não apenas a proteção contra alguns riscos, mas também quais riscos explorar e como fazê-lo.

Medir a aversão ao risco é a primeira etapa para analisar e lidar com o risco nos contextos de portfólio e negócios. À medida que a aversão ao risco aumenta, o prêmio de

risco exigido para qualquer operação arriscada também aumenta. Assim, para que haja um gerenciamento de risco, é necessário, primeiramente, a sua mensuração (Damodaran, 2007). Para o risco de crédito, objeto deste estudo, essa mensuração pode ser feita por modelos de gestão do risco de crédito, como o *credit scoring*, que será explorado mais adiante.

2.3.1 Risco de Crédito

Segundo Jolevski (2017), Chen e Sivakumar (2021), os empréstimos são o núcleo da atividade bancária, com participação dominante nos ativos dos bancos, em que o acompanhamento da movimentação e do nível de inadimplência são fundamentais para a identificação de possíveis problemas na gestão do risco de crédito. Conforme Altman (1968), a avaliação de empréstimos é uma função importante na sociedade, principalmente para bancos comerciais e instituições de crédito. Mecanismos eficientes e rápidos para detectar riscos de crédito desfavoráveis podem evitar decisões prejudiciais à essas instituições.

O risco de inadimplência, o qual é um tipo de risco de crédito e representa o risco de uma empresa não cumprir as suas obrigações de dívida, tem atraído atenção significativa, principalmente no setor bancário, após a crise financeira de 2007 a 2009 (Kanas, 2014). A inadimplência, pela qual uma contraparte não deseja ou é incapaz de cumprir suas obrigações contratuais, é um caso extremo. Adicionalmente, há a exposição ao risco de que uma contraparte seja rebaixada por uma agência de classificação (Crouhy, Galai, & Mark, 2005).

O risco de crédito também inclui o risco soberano, o qual está associado aos países imporem controles cambiais que impossibilitam as contrapartes de honrar suas obrigações. Dessa forma, o risco de inadimplência é específico da empresa, enquanto o risco soberano é específico do país. Outra forma de risco de crédito é o risco de liquidação quando dois pagamentos são trocados no mesmo dia, surgindo se a contraparte entrar em *default* após a instituição já ter efetuado o seu pagamento (Jorion, 2007).

Ainda, vale mencionar o conceito de falência, pois está intimamente relacionado ao risco de crédito e à inadimplência. De acordo com Beaver (1966), falência é definida como a incapacidade da empresa de pagar suas obrigações financeiras no vencimento. Conforme Altman (1968), uma empresa com baixa rentabilidade ou solvência pode ser considerada como tendo um potencial para falência. Olson (1980) aponta quatro fatores como determinantes na previsão de falência, os quais são medidas da estrutura financeira (alavancagem), tamanho da empresa, mensuração da liquidez corrente e mensuração de performance.

Avaliar o risco de crédito significa reunir e interpretar informações. Porém, se torna mais difícil conforme aumenta a distância entre o tomador do empréstimo e o credor final. Dentre os Cs clássicos do crédito, caráter, capital e colateral continuam sendo elementos essenciais da decisão de crédito (Caouette, Altman, Narayanan, & Nimmo, 2008). O crédito pode aumentar rapidamente por motivos de ser concedido a quem não era elegível anteriormente, aumentos estruturais normais, com o crescimento da economia global e movimentos estruturais excessivos, e quando há um *boom* de crédito potencialmente desestabilizador. O crédito vem se expandindo em ritmo acelerado, em um momento em que as ferramentas para a sua gestão têm melhorado e as fontes de informação são significativamente melhores do que há alguns anos (Caouette, Altman, Narayanan, & Nimmo, 2008).

O risco de crédito possui três elementos, os quais são concentração, ou seja, a exposição a uma parte que pode entrar em *default*, a probabilidade de *default* e a taxa de recuperação. Embora a taxa de recuperação mais alta reduza o risco de crédito, os dois primeiros elementos aumentam com cada incremento deles. Os bancos costumam adotar um conjunto de modelos quantitativos padrão proposto pelas agências de risco para avaliar o risco de crédito (Rana, & Ghildiyal, 2016).

Ainda, existem dois tipos de abordagem estratégica principais nos bancos quanto à gestão do risco de crédito, que podem ser de diversificação ou concentração. A discussão sobre elas se transformou em um dos temas mais importantes levantados na questão da estabilidade econômica (Raei et al., 2016). O monitoramento, ou seja, a eficiência gerencial, e as estratégias de gerenciamento de risco de crédito afetam diretamente os retornos e os riscos da carteira de crédito. A diversificação entre ativos, como empréstimos expostos ao risco de crédito, reduz o risco de crédito global na carteira e aumenta a probabilidade de reembolso parcial ou total do principal e juros. Assim, reduz o risco de crédito específico da empresa, mas mantém a exposição ao risco de crédito sistêmico, o qual aumenta o risco de inadimplência de todas as empresas na economia (Saunders, & Cornett, 2017).

2.3.2 Modelos de Gestão de Risco de Crédito

Os indicadores tradicionais de risco são conceitos estatísticos de variância e desvio padrão, os quais são fundamentais para as teorias financeiras (Cisko, & Kliestik, 2013). Segundo Gavlakova e Kliestik (2014), durante as décadas de 1970 e 1980, foram criados

novos produtos financeiros, o que representou um novo desafio para a modelagem de risco, por não haver histórico, por meio do qual pudessem ser estimados seus riscos.

As principais instituições financeiras globais têm buscado técnicas eficazes para gerenciar o risco de crédito, desenvolvendo sistemas globais de informação que são atualizados continuamente para que a exposição e os preços possam ser monitorados em tempo real (Caouette, Altman, Narayanan, & Nimmo, 2008). Para se qualificar como um modelo interno de gestão de risco, a instituição precisa demonstrar ao regulador de que o risco de concentração, de rebaixamento, de inadimplência e de spread são apropriadamente capturados (Crouhy, Galai, & Mark, 2000).

Existem diversas abordagens com modelos que são aplicados para a mensuração do risco de crédito. De acordo com Saunders e Cornett (2017), esses modelos podem ser qualitativos ou quantitativos. Os modelos qualitativos, como os Sistemas Especialistas, contêm fatores específicos do tomador, como alavancagem e reputação, e fatores específicos do mercado, como o nível das taxas de juros e o ciclo de negócios.

Nos Sistemas Especialistas, as instituições financeiras precisam coletar informações sobre a qualidade dos tomadores de fontes privadas, como relatórios de crédito e depósito ou adquiri-las de fontes externas, como agências de classificação de crédito, quando não estão publicamente disponíveis. Essas informações auxiliam no julgamento da concessão de crédito, informando sobre a probabilidade de inadimplência do tomador, e na precificação correta do empréstimo ou dívida. Para isso, pode-se utilizar fatores específicos do tomador, tais como a sua reputação, a qual é possível de ser avaliada através do contrato implícito entre as partes, quando há um relacionamento de longo prazo, alavancagem, volatilidade dos rendimentos e o colateral, ou seja, os ativos utilizados como garantia na operação de crédito (Saunders, & Cornett, 2017).

Com isso, um dos sistemas especialistas mais comuns de crédito é o dos cinco Cs do crédito. Esse sistema considera características de caráter, capital, capacidade, colateral e ciclo econômico para avaliar a concessão de crédito. Além disso, um especialista também pode levar em consideração o nível das taxas de juros. Entretanto, esse sistema apresenta dois problemas principais, relacionados a consistência e subjetividade, pois é muito difícil para um indivíduo ser imparcial, fazer comparações sem julgamento de valor e analisar de forma equivalente grandes quantidades de informações (Saunders, & Allen, 2002).

Também, utiliza-se fatores específicos de mercado. Dentre eles, destaca-se o ciclo do negócio, em que a posição da economia é importante para avaliar a probabilidade de inadimplência do tomador, e o nível das taxas de juros. Taxas de juros mais altas significam

ações restritivas de política monetária por parte dos bancos centrais, fazendo com que os fundos para financiar decisões de empréstimo sejam mais escassos e caros, além de estarem correlacionadas a um maior risco de crédito (Saunders, & Cornett, 2017).

Os modelos quantitativos, como o modelo de probabilidade linear, o logit e o discriminante linear, fornecem pontuações de crédito, ou *credit scorings*, que classificam os empréstimos pelo risco de inadimplência esperado. O modelo de *credit scoring*, o qual é foco deste estudo, é apresentado com maiores detalhes na próxima subseção. Ainda, existem modelos quantitativos que utilizam a teoria financeira e dados financeiros, os quais são modelos de estrutura a termo e taxa de mortalidade, o RAROC (*risk-adjusted return on capital*) que considera o retorno de capital ajustado ao risco e modelos baseados em opções (Saunders, & Cornett, 2017).

O modelo de estrutura a termo de risco de crédito, que é um modelo de forma reduzida, baseado no mercado, analisa os prêmios de risco inerentes à estrutura corrente de rendimentos de dívidas corporativas ou empréstimos a mutuários com classificação de risco semelhante. Assim, o modelo demonstra que é possível extrair uma visão de mercado do risco de crédito, a probabilidade esperada de inadimplência de um tomador (Saunders, & Cornett, 2017).

O modelo relacionado à abordagem da taxa de mortalidade, analisa a experiência de risco de inadimplência histórica ou passada, considerando as taxas de mortalidade de títulos e empréstimos de qualidade semelhante. A taxa de mortalidade marginal é a probabilidade de inadimplência do empréstimo no primeiro ano de emissão. Para cada grau de qualidade do tomador corporativo, a curva da taxa de mortalidade marginal (MMR) pode mostrar a experiência histórica da taxa de inadimplência de títulos em qualquer classe de qualidade específica em cada ano após a emissão do título ou empréstimo. O principal problema deste modelo, assim como o *credit scoring*, é produzir medidas históricas ou retrospectivas (Saunders, & Cornett, 2017).

O RAROC é um modelo usado para avaliar e precificar o risco de crédito com base em dados de mercado. Foi lançado pelo Bankers Trust e foi adotado por praticamente todos os grandes bancos nos Estados Unidos e na Europa. O ponto principal do modelo é que, em vez de avaliar o ROA (*return on assets*) anual real ou contratualmente acordado em um empréstimo, são considerados os juros esperados e a receita de taxas menos o custo dos fundos contra o risco esperado do empréstimo. O numerador da equação RAROC é o lucro líquido do empréstimo. Além disso, a receita anual por ativos emprestados é dividida pelo chamado capital em risco, já que perdas inesperadas de empréstimos são baixadas contra o capital da instituição. Assim, um empréstimo deve ser feito apenas se o retorno ajustado ao

risco sobre o empréstimo adiciona ao valor patrimonial da instituição financeira, conforme ROE (*return on equity*) exigido pelos seus acionistas (Saunders, & Cornett, 2017).

Quanto aos modelos de opções, os quais são modelos estruturais, quando uma empresa levanta fundos emitindo títulos ou aumentando seus empréstimos bancários, ela possui uma opção de inadimplência ou reembolso valiosa. Isso foi identificado, a partir do trabalho pioneiro dos prêmios Nobel Merton e Black e Scholes. Assim, se os negócios vão bem, o mutuário pode manter a maior parte dos retornos dos investimentos em ativos depois do pagamento do principal e os juros da dívida. A KMV Corporation, comprada pela Moody's em 2002, transformou esse conceito simples em um modelo de monitoramento de crédito, o qual vai ser detalhado mais adiante. Esse modelo é utilizado para determinar a frequência de risco de inadimplência esperada (EDF) de grandes corporações (Saunders, & Cornett, 2017).

O sistema desenvolvido pelo Escritório do Controlador da Moeda dos Estados Unidos (OCC), conhecido como *Credit Rating*, é um dos sistemas de classificação mais antigos para empréstimos. Esse sistema classifica uma carteira de empréstimos em cinco categorias, das quais quatro são classificações de baixa qualidade e uma classificação de alta qualidade. Com o passar do tempo, foi ampliado pelos bancos para o desenvolvimento de sistemas de classificação internos, que subdividem de forma precisa a categoria de classificação de aprovação e desempenho (Saunders, & Allen, 2002).

As classificações de risco, ou *Credit Ratings*, utilizam vários critérios qualitativos, como pontos fortes e fracos das empresas e aspectos financeiros dos tomadores de empréstimos. O processo de classificação interna geralmente usa níveis de classificação, que listam os critérios usados para obter uma avaliação final da qualidade de crédito dos tomadores. Os ratings externos fornecem as informações necessárias para os investidores sobre a qualidade de crédito de uma emissão de dívida (Bessis, 2015).

As classificações são usadas para comunicar opiniões sobre a qualidade de crédito de emissores e obrigações, contendo diversas informações, como a probabilidade de inadimplência e a gravidade da perda nesses casos. Com isso, os reguladores têm interesse nos ratings pela sua alta qualidade, independência e aceitação generalizada das opiniões das agências sobre a qualidade do crédito. Entretanto, possuem um certo receio em confiar nas agências, pois não têm controle sobre suas atividades. (Caouette, Altman, Narayanan, & Nimmo, 2008).

As agências de classificação de risco de crédito categorizam os emissores de títulos corporativos em pelo menos sete classes principais de acordo com a qualidade de crédito percebida. Para a Standard and Poor's, por exemplo, as primeiras quatro categorias, as quais

são AAA, AA, A e BBB, indicam tomadores de empréstimo com qualidade de investimento. As classes BB, B e CCC são conhecidas como títulos de alto rendimento ou *junk bonds* (Saunders, & Cornett, 2017). Portanto, as instituições financeiras precisam medir a probabilidade de inadimplência do tomador, o que depende muito da quantidade das informações que elas possuem sobre o mutuário (Saunders, & Cornett, 2017). Também, há os modelos de monitoramento de crédito, os quais são o CreditMetrics, KMV Model, CreditRisk + e CreditPortfolioView.

O CreditMetrics do JP Morgan, publicado em 1997, tem sua abordagem baseada na análise de migração de crédito, ou seja, na probabilidade de mudança de uma qualidade de crédito para outra, incluindo inadimplência, dentro de um determinado horizonte de tempo. Modela a distribuição completa a termo dos valores de qualquer título ou carteira de empréstimos, normalmente para o período de um ano. As mudanças nos valores estão relacionadas apenas à migração de crédito, enquanto as taxas de juros são assumidas para evoluir de uma forma determinística (Crouhy, Galai, & Mark, 2000).

Paralelamente ao CreditMetrics, o Credit-VaR de uma carteira é derivado de forma semelhante ao risco de mercado, sendo o percentil da distribuição correspondente ao nível de confiança desejado (Crouhy, Galai, & Mark, 2000). O modelo de *Value at Risk* (VaR), por sua vez, vem de longa data. Segundo Adamko et al. (2015), a história do valor em risco (VaR) iniciou em 1945, quando Dickson H. Leavens fez a primeira menção ao VaR. Em 1952, houve destaque para Harry Markowitz, que recebeu o Prêmio Nobel de Economia, em 1990, por sua pesquisa pioneira na área de teoria do portfólio. De forma independente, Arthur D. Roy propôs indicadores de VaR que eram muito semelhantes. Ambos estavam tentando encontrar uma maneira de otimizar o lucro para um determinado nível de risco.

Valor em risco (VaR) pode ser definido como a pior perda esperada por manter um título ou um portfólio por um determinado período, dado um nível específico de probabilidade. É uma medida estatística que define um determinado nível de perda em termos de suas chances de ocorrência (Crouhy, Galai, & Mark, 2005). A partir da década de 1980, o VaR foi introduzido para a mensuração do risco de mercado das instituições financeiras, em suas carteiras de derivativos, de ações e demais valores mobiliários.

O Banco Americano J.P. Morgan, em 1994, tentou estabelecer uma metodologia padrão para o cálculo do VaR, criando e divulgando o *RiskMetrics*TM, que representa uma síntese técnica escrita por acadêmicos e profissionais do mercado financeiro, difundindo o uso do modelo. Outro movimento importante para a sua popularização ocorreu em 1995, quando o Comitê de Supervisão Bancária do Bank for International Settlements (BIS) propôs que os

bancos deveriam calcular a alocação e reserva de capital utilizando técnicas baseadas no VaR (Jorion, 2007).

O VaR de crédito de uma carteira é derivado de forma semelhante ao VaR de risco de mercado, sendo o percentil da distribuição correspondente ao nível de confiança desejado. Há duas etapas principais no cálculo do VAR: derivar uma distribuição ou retornos futuros de um portfólio em um determinado horizonte temporal e identificar o percentil necessário desta distribuição para que um valor de perda possa ser verificado. O VaR representa o capital econômico que os acionistas devem investir na empresa, para limitar a probabilidade de falência, dado um nível de confiança pré-determinado. Já o capital regulatório é a quantia mínima de capital imposta pelo regulador (Crouhy, Galai, & Mark, 2000).

Crouhy, Galai e Mark, (2005), complementam que, por ser uma medida agregada de risco por meio de todos os fatores de risco, tem a vantagem de ser calculado em cada nível de atividade da empresa. É uma boa forma de representar o apetite ao risco de curto prazo da empresa, pois mede a perda máxima que ela pode incorrer sob condições normais de mercado. Fornece, também, uma medida agregada de risco e risco ajustado pela performance, o qual pode ser traduzido em um requerimento de capital.

Dentre as abordagens de utilização do VaR, o cálculo por simulação histórica é simples e não requer que o usuário faça suposições analíticas sobre a distribuição, sendo suficientes apenas dois ou três anos de dados para resultados significativos. A simulação de Monte Carlo consiste em simular repetidamente processos aleatórios que governam os preços e taxas de mercado. Envolve três passos, os quais são especificar todos os fatores de risco relevantes, construir cenários de preços e valorar o portfólio para cada cenário. Essa simulação é uma forma potente e flexível de VaR, pois pode acomodar qualquer distribuição de fatores de riscos, inclusive com eventos extremos (Crouhy, Galai, & Mark, 2005).

Se comparado ao VaR de mercado, o VaR de crédito apresenta dois desafios: a distribuição da carteira tende a não ser normal e a mensuração do efeito da carteira devido à diversificação do crédito é mais complexo. O cálculo do VaR para risco de crédito requer a simulação da distribuição completa das variações no valor da carteira. Para isso, é necessário estimar as correlações nas mudanças na qualidade de crédito para todos os pares de devedores. O modelo não assume risco de mercado, pois os valores a termo e as exposições são derivados de curvas determinísticas a termo, com a única incerteza vinculada à migração de crédito. O modelo de VaR de crédito tem quatro etapas: especificar a matriz de transição, definir o horizonte de risco de crédito, especificar o modelo de preço futuro e derivar a distribuição futura das mudanças no valor da carteira (Crouhy, Galai, & Mark, 2000).

O VaR pode ser combinado com o Teste de Stress para produzir um reporte dos riscos mais relevantes. O objetivo do teste de stress e da análise de cenários é determinar o tamanho da potencial perda, relativa a cenários específicos. A sua principal vantagem é demonstrar quão vulnerável um portfólio pode ser a uma variedade de eventos extremos (Crouhy, Galai, & Mark, 2005).

A KMV Corporation, empresa especializada em análise de risco de crédito, desenvolveu uma metodologia e um extenso banco de dados para avaliar as probabilidades de inadimplência e a distribuição de perdas relacionadas aos riscos de inadimplência e migração. A KMV mostrou, por meio de simulação, que a taxa média histórica de inadimplência e as probabilidades de transição podem divergir significativamente das taxas reais, o que é uma fraqueza do modelo anterior. O KMV deriva a probabilidade real de inadimplência, a frequência de inadimplência esperada (EDF), para cada devedor com base no modelo de Merton (1974). A probabilidade de inadimplência é função da estrutura de capital da empresa, da volatilidade dos retornos do ativo e do valor atual do ativo (Crouhy, Galai, & Mark, 2000).

O modelo se aplica melhor a empresas de capital aberto para as quais o valor do patrimônio é determinado pelo mercado, em que as informações contidas no preço das suas ações e no balanço patrimonial podem ser traduzidas em um risco implícito de inadimplência. As empresas entram em *default* quando o valor do ativo atinge um nível entre o valor do passivo total e o valor da dívida de curto prazo. O modelo calcula a distância até o *default* (DD), a qual é o número de desvios padrão entre a média da distribuição do valor do ativo e um limite crítico, o ponto de inadimplência, definido pelo valor nominal do passivo circulante, incluindo dívida de curto prazo a ser servida ao longo do horizonte de tempo, mais metade da dívida de longo prazo (Crouhy, Galai, & Mark, 2000).

A EDF é um indicador importante da degradação da qualidade de crédito dos emissores, pois suas mudanças tendem a antecipar pelo menos em 1 ano o rebaixamento do emissor por agências de rating. A KMV construiu uma matriz de transição com base nas taxas de inadimplência, em vez de classes de classificação. A dependência da EDF é sua principal diferença para o CreditMetrics. Ambos contam com o modelo de valor do ativo proposto por Merton (1974), mas diferem nas suposições requeridas para facilitar sua implementação (Crouhy, Galai, & Mark, 2000).

O Credit Suisse Financial Products (CSFP) lançou, em 1997, o CreditRisk +, que foca apenas na inadimplência, assumindo que o *default* para títulos individuais, ou empréstimos, segue um processo de Poisson, o qual se aproxima da distribuição do número de eventos de *default*. Utiliza dados contábeis para derivar a distribuição de perdas de uma carteira de

empréstimos e o risco de migração de crédito não é explicitamente modelado nesta análise, mas permite taxas de inadimplência estocásticas que contabilizam parcialmente o risco de migração (Crouhy, Galai, & Mark, 2000).

A exposição para cada devedor é ajustada pela taxa de recuperação antecipada, para calcular a perda em caso de inadimplência, sendo exógena ao modelo e independente do risco de mercado e de rebaixamento. É o valor futuro da linha de crédito multiplicado pela perda, dada a taxa de inadimplência. Além disso, as contribuições de risco marginal por devedor podem ser facilmente calculadas. Concentra-se apenas na inadimplência, exigindo poucos dados para as estimativas, como a probabilidade de inadimplência e a exposição. Possui as mesmas limitações dos modelos anteriores, não assumindo o risco de mercado. Ignora o risco de migração de forma que a exposição de cada devedor seja fixa e não dependa de eventuais mudanças na qualidade de crédito do emissor e da variabilidade das taxas de juros futuras.

A McKinsey, por sua vez, propôs o modelo CreditPortfolioView que, como o CreditRisk +, mede apenas o risco de inadimplência. É um modelo de múltiplos períodos de tempo discreto, em que as probabilidades de inadimplência são função de macro variáveis como o desemprego, o nível das taxas de juros, a taxa de crescimento da economia, as despesas do governo, as taxas de câmbio e os ciclos de crédito (Crouhy, Galai, & Mark, 2000).

Esse modelo é baseado na observação casual de que as probabilidades de inadimplência e as de migração estão ligadas à economia. O CreditPortfolioView propõe uma metodologia para vincular esses fatores macroeconômicos às probabilidades de inadimplência e migração. As probabilidades de *default* são modeladas como uma função logit, em que a variável independente é um índice específico do grau especulativo do país que depende das variáveis macroeconômicas atuais e passadas (Crouhy, Galai, & Mark, 2000).

Dentre outros modelos citados na literatura, Sanvicente e Minardi (1998) desenvolveram um modelo para identificar os índices contábeis mais significativos para prever concordatas de empresas brasileiras. Fundamentados nas pesquisas de Edward Altman (1968 e 1977) e na dinâmica de *overtrading* de Assaf Neto e Tibúrcio Silva (1995), aplicaram a técnica estatística de análise discriminante para a análise de uma seleção de índices contábeis. Essa técnica é útil para atribuir escores associados ao risco de crédito. Com isso, a função discriminante que obteve a maior precisão foi $Z = -0,042 + 2,909X_1 - 0,875X_2 + 3,636X_3 + 0,172X_4 + 0,029X_8$, a qual classificou corretamente 81,8% dos casos. As variáveis do modelo são índices de liquidez (X_1), indicadores de lucros retidos (X_2), rentabilidade (X_3), endividamento (X_4) e cobertura de juros (X_8).

Em 1999, os referidos autores desenvolveram um sistema classificatório de risco para companhias abertas do Brasil, atribuindo ratings anuais a elas, com a utilização da análise de clusters. O modelo classificou 87% das empresas concordatárias como B e CCC no ano do evento da concordata e detectou piora ou alto risco para 81,8% das empresas que entraram em concordata no ano seguinte. Entretanto, a matriz de migração de ratings não teve o resultado esperado, falhando pela violação da ordem natural dos ratings e por casos em que o maior número de migrações de uma classe nem sempre foi para o estado inicial, como deveria ocorrer (Sanvicente, & Minardi, 1999).

O modelo desenvolvido em Vasicek (1987) é considerado a base para os requisitos de capital regulatório da Basileia e é muito utilizado no setor financeiro. Permite cálculos de risco de múltiplos períodos, apesar de ser mais utilizado em cálculos de risco anuais. O modelo afirma que o valor de uma contraparte é determinado por um fator macroeconômico próprio e um termo independente idiossincrático, possuindo como parâmetros a probabilidade de *default* (PD) e a sensibilidade macroeconômica. Tem a vantagem de fornecer uma representação fechada da distribuição de perdas, mas também, os cálculos de risco de mais de um período podem ser obtidos por meio da análise de cenários em que são estimadas as perdas da carteira em um cenário macroeconômico determinado (García-Céspedes, & Moreno, 2017).

Ainda, muitos sistemas estudam as decisões dos especialistas humanos, utilizando indução para inferir os processos de decisão. Com isso, as redes neurais artificiais têm sido propostas para esta finalidade, pois simulam o processo de aprendizagem humana. O sistema aprende a natureza da relação entre entradas e saídas por amostragem repetida de conjuntos dessas informações. Possuem a vantagem de fazer uma suposição fundamentada, como seria feito por um especialista humano, quando há ruídos nos dados ou são incompletos (Saunders, & Allen, 2002). Assim, esse tipo de sistema pode incorporar diversos tipos de informações, tanto quantitativas quanto qualitativas, para a análise de concessão de crédito.

Com o avanço do desenvolvimento de métodos que utilizam inteligência artificial, tem se destacado as técnicas de *machine learning*, ou aprendizado de máquina. Conforme Khandani, Kim e Lo (2010), *machine learning* se refere a um conjunto de algoritmos projetados com o objetivo de resolver problemas de reconhecimento de padrões intensivos em computação, em conjuntos de dados extremamente grandes. Por exemplo, a técnica conhecida como *Random Forest* é baseada na construção de diversas árvores de decisão, as quais produzem algumas previsões individualmente. O resultado de toda a floresta aleatória deve ser escolhido entre as previsões das diversas árvores individuais, fundamentado na frequência

de aparecimento de valores específicos. Sendo assim, o agrupamento de muitos classificadores fracos tem uma propriedade considerável na redução da variância da previsão (Biecek et al., 2021).

Entretanto, os modelos complexos de aprendizado de máquina, como redes neurais e modelos de árvore, fornecem uma alta precisão preditiva às custas de uma capacidade interpretativa limitada, se comparados aos modelos estatísticos de aprendizagem considerados simples, como modelos de regressão linear e logística. Estes últimos possuem alta capacidade de interpretação, mas podem ter uma precisão preditiva limitada (Financial Stability Board, 2017, World Bank Group, 2019, Bussmann et al., 2021). Vale ressaltar que, conforme os autores, a capacidade explicativa ou de interpretação significa que uma parte interessada pode compreender os principais impulsionadores de uma decisão baseada no modelo. Portanto, apesar de a inteligência artificial ser benéfica para a conveniência e a acessibilidade dos serviços financeiros, também traz novos riscos (Bussmann et al., 2021).

Adicionalmente, na literatura, encontram-se várias discussões relativas as técnicas de *machine learning* podendo introduzir vieses no comportamento de empréstimos, com o risco de exclusão financeira, e envolvendo questões relacionadas à ética, proteção do consumidor, privacidade e transparência, as quais requerem mais atenção dos órgãos reguladores (Bazarbash, 2019). Além disso, muitos métodos utilizam previsões baseadas em caixas pretas, consideradas processos obscuros de mineração de dados, as quais necessitam serem explicadas (Moscato, Picariello, & Sperlí, 2021).

2.3.2.1 Credit Scoring

O *credit scoring*, ou pontuação de crédito, é uma técnica difundida, não apenas no setor bancário, mas também em muitos outros setores, em que há a necessidade de verificar a posição de crédito de um cliente ou a probabilidade de registro de uma reclamação. É a principal ferramenta para medir o risco de crédito no varejo. O modelo usa um procedimento estatístico para converter informações sobre um tomador de crédito em números que são combinados para formar uma pontuação, a qual é considerada uma medida do seu risco de crédito. Quanto maior a pontuação, menor o risco (Crouhy, Galai, & Mark, 2005). Os métodos e ferramentas estatísticos para a sua elaboração, com a utilização de variáveis binárias, são considerados os mais eficazes, além de possuírem grande popularidade (Nehrebecka, 2021).

A pontuação de crédito permite que os bancos evitem os clientes de maior risco e os ajuda a avaliar se certos tipos de negócios têm probabilidade de serem lucrativos, comparando a margem de lucro que permanece depois que as despesas operacionais e inadimplências são subtraídas da receita bruta. Com esse modelo, os bancos podem automatizar o processo de adjudicação de pequenos créditos e cartões de crédito. O número de combinações possíveis de informações é relativamente alto, tornando impossível para um analista humano tratar as decisões de crédito de maneira idêntica ao longo do tempo (Crouhy, Galai, & Mark, 2005).

Os modelos de pontuação de crédito avaliam não apenas se um atributo é positivo ou negativo, mas também sua quantificação. A partir de uma pontuação de corte, o banco pode determinar, com base em sua experiência, a taxa de perda e a lucratividade do produto, sendo importante, além disso, para a determinação do risco que o banco está assumindo (Crouhy, Galai, & Mark, 2005).

Como o credor não pode observar as probabilidades de inadimplência dos tomadores, os modelos de *credit scoring* permitem que a instituição de crédito classifique os clientes potenciais de acordo com seu risco de inadimplência para melhorar a alocação de recursos. Para isso, é necessário estimar um modelo de pontuação de crédito imparcial, por meio da seleção de diferentes candidatos. Uma forma de medir o risco é pesar os riscos de inadimplência individuais por valor, por meio do cálculo do VaR, o qual permite a avaliação do impacto de diferentes políticas de empréstimo sobre a exposição ao risco e cria uma base de decisão sobre a taxa de perda implícita (Jacobson, & Roszbach, 2003).

Adicionalmente, o novo Acordo de Basileia, quanto ao requerimento de capital, exige que os bancos segmentem suas carteiras de varejo em sub carteiras com características de perda semelhantes. Assim, os bancos precisam estimar tanto a probabilidade de inadimplência (PD) quanto a perda por inadimplência (LGD) para essas carteiras. Isso pode ser alcançado segmentando cada portfólio de varejo por faixa de pontuação, correspondendo a um respectivo nível de risco (Crouhy, Galai, & Mark, 2005).

Ainda, os modelos de pontuação de histórico de crédito têm vantagens de baixo custo e rastreamento rápido e consistente e, por serem baseados nas informações contidas nos arquivos das agências de classificação de crédito, podem ser usados para rastrear a maioria dos clientes em potencial. Essas agências reúnem informações sobre as experiências dos indivíduos com crédito, leasing, letras não relacionadas a crédito, registros públicos monetários e consultas de crédito, e as compilam em um registro de crédito. Esses registros contêm diversos detalhes sobre cada conta, como sua identificação, data, saldos, descrição e desempenho de pagamento (Avery, Calem, & Canner, 2004).

Entretanto, os modelos de *credit scoring* podem apresentar erros. O modelo pode indicar baixo risco quando o risco é alto. Esse erro corresponde à atribuição de alta qualidade de crédito a devedores que entram ou quase entram em *default*. O custo do banco é a perda do valor do crédito ou os juros. Outro erro seria o modelo indicar alto risco quando o risco é baixo. Esse erro está relacionado a empresas com classificação baixa que deveriam receber classificação mais alta. As perdas deste segundo tipo de erro incluem a perda de retorno e taxas e uma queda na participação de mercado, quando os empréstimos são recusados ou perdidos devido a preços não competitivos (Blöchlinger, & Leippold, 2006).

Portanto, o *credit scoring* deve ser capaz de atribuir pontuações altas a créditos bons e pontuações baixas a créditos ruins, com o objetivo de minimizar a sobreposição da distribuição de créditos bons e ruins. Com a evolução tecnológica, os bancos progrediram de aplicativos de pontuação em um ponto no tempo para a pontuação de comportamento periódica. Com isso, são utilizadas informações sobre o comportamento de um cliente atual, como o uso da linha de crédito e informações sócio demográficas, para determinar o risco de inadimplência em um determinado período de tempo (Crouhy, Galai, & Mark, 2005).

Assim, o seu principal benefício é a previsão com mais precisão, pelos credores, do desempenho do mutuário sem necessitar recursos adicionais. Para sua utilização, deve-se identificar medidas econômicas e financeiras objetivas de risco para cada classe específica de tomador de empréstimo. Por exemplo, para dívidas do consumidor, as características objetivas podem incluir idade, ocupação, localização, renda e bens (Saunders, & Cornett, 2017).

Conforme Saunders e Cornett (2017), os modelos de pontuação de crédito podem ser de três tipos:

- a) modelos de probabilidade linear: esse modelo utiliza dados anteriores, como índices financeiros para explicar a experiência de reembolso de empréstimos passados. Os fatores que explicam o desempenho de reembolso passado podem ser usados para avaliar a probabilidade de reembolso e a probabilidade de inadimplência (PD). São divididos os empréstimos anteriores em dois grupos observacionais: os que entraram em *default* ($PD = 1$) e os que não entraram ($PD = 0$). Essas observações são relacionadas por regressão linear a um conjunto de variáveis causais que refletem informações quantitativas sobre o tomador de empréstimo, como alavancagem ou ganhos. Considerando a estimativa de importância das variáveis causais e a multiplicando pelas variáveis causais observadas para cada mutuário em potencial, é possível derivar um valor esperado de PD para o mutuário, o qual significa a

probabilidade de inadimplência do tomador. Sua principal fraqueza é que as probabilidades estimadas em caso de inadimplência podem situar-se fora do intervalo de 0 a 1.

- b) modelo logit: esse modelo supera a fraqueza do anterior, restringindo o intervalo estimado de probabilidades de inadimplência do modelo de regressão linear, limitado entre 0 e 1.
- c) análise discriminante linear: esse modelo divide os mutuários em classes de alto ou baixo risco de inadimplência, conforme suas características observadas e prevê em qual classe o empréstimo se enquadra. O modelo Z de Altman considera a variável Z como uma medida geral de classificação de risco de inadimplência de um mutuário comercial, a qual depende dos valores de vários índices financeiros dele e da importância ponderada desses índices com base na experiência anterior observada de mutuários inadimplentes versus inadimplentes derivados de um modelo de análise discriminante. O principal problema deste modelo é sua discriminação extrema, a qual classifica somente entre inadimplente ao não.

Portanto, o modelo de *credit scoring* traz bons resultados quanto à previsão de inadimplência para a gestão do risco de crédito. Possui um bom poder preditivo e é de fácil compreensão quanto ao seu poder explicativo. A pontuação de crédito auxilia na classificação de bons e maus pagadores, utilizando características objetivas e de fácil acesso dos tomadores, tanto pessoas físicas quanto jurídicas.

2.4 ADEQUAÇÃO DE CAPITAL

Uma instituição financeira precisa se proteger contra o risco de insolvência, para garantir a sua sobrevivência. Os reguladores também estão preocupados em garantir a solvência dessas instituições, pois o seu fracasso pode ter graves repercussões tanto na economia local quanto global. O principal meio de proteção contra o risco de insolvência e falência é o capital da instituição financeira. Segundo a teoria econômica, o capital deve ser medido com base no valor de mercado, como a diferença entre os valores de ativos e passivos. No entanto, os reguladores usam regras de valor contábil, pois consideram a exposição ao risco de crédito. Sendo assim, vale ressaltar que essas regras ignoram os efeitos das mudanças nas taxas de juros e da exposição a elas sobre o patrimônio líquido (Saunders, & Cornett, 2017).

As funções do capital são absorver perdas imprevistas com margem suficiente para transmitir confiança e permitir que as instituições financeiras continuem funcionando, proteger os fundos de seguro e os contribuintes e proteger depositantes não segurados, detentores de títulos e credores em caso de insolvência e liquidação. Também, tem a função de financiar investimentos reais necessários para fornecer serviços financeiros e proteger os proprietários contra aumentos nos prêmios de seguro. Então, quanto maior o patrimônio líquido da instituição financeira, em relação ao tamanho de seus ativos, maior a proteção contra insolvência ou seguro para os detentores de passivos e seus fiadores (Saunders, & Cornett, 2017).

Em 1993, o BIS estabeleceu índices de capital baseados em risco, sob o Acordo de Basileia I, o qual incorporou os diferentes riscos de crédito dos ativos nas medidas de adequação de capital. Em 1998, o risco de mercado foi incorporado ao capital com um acréscimo ao índice de 8% para a exposição ao risco de crédito. Em 2001, foi emitido o Novo Acordo de Capital de Basileia, que incluiu o risco operacional nos requisitos de capital e atualizou as avaliações de risco de crédito de 1993 (Saunders, & Cornett, 2017). Assim, os índices de capital sensíveis ao risco são a pedra angular da estrutura regulatória prudencial dos bancos, em que são obrigados a manter o capital regulamentar como uma porcentagem de seus ativos ponderados pelo risco (RWAs) (Cucinelli, Di Battista, Marchese, & Nieri, 2018).

O Acordo de Basileia II, de 2006, consistiu em três pilares que se reforçam mutuamente e contribuem para a segurança e a solidez do sistema financeiro. O Pilar 1 cobriu as exigências regulatórias de capital mínimo para risco de crédito, mercado e operacional. O pilar 2 destacou a importância do processo de revisão da supervisão regulatória como um complemento crítico aos requisitos de capital mínimo. E o pilar 3, incentivou a disciplina de mercado, desenvolvendo um conjunto de requisitos sobre a divulgação da estrutura de capital, exposições de risco e adequação de capital (Saunders, & Cornett, 2017).

Ademias, conforme Cucinelli, Di Battista, Marchese e Nieri, (2018), em 2004, foi introduzida, pelo Acordo de Basileia II, a abordagem baseada em ratings internos (IRB – *Internal Ratings-based*) como uma alternativa à abordagem padronizada (SA – *Standard Approach*). Na SA, os bancos devem usar ponderações de risco fornecidas pelos reguladores. Já na IRB, podem usar seus próprios modelos de classificação interna para estimar os diferentes componentes da perda esperada.

No IRB de Fundação, os bancos determinam internamente a probabilidade de inadimplência (PD) apenas. Já no IRB Avançado, eles definem internamente a perda dada a inadimplência (LGD), a exposição à inadimplência (EAD) e a maturidade (M).

Consequentemente, os bancos que usam o IRB podem se beneficiar de uma medição de risco de crédito mais granular, que traz pesos de risco mais baixos e implica em encargos de capital menores. Entretanto, o uso de modelos IRB está condicionado à validação das autoridades de supervisão (Cucinelli, Di Battista, Marchese, & Nieri, 2018).

Durante o período de validação, que pode levar anos, a instituição financeira é obrigada a implementar em seus modelos internos as alterações solicitadas pelos supervisores, para fornecer evidências claras da robustez de tais modelos e demonstrar sua relevância na operação da gestão de risco diário e em sua própria cultura de risco. Vale ressaltar que a avaliação da eficácia da gestão do risco de crédito é uma tarefa complicada para um analista externo, pois pode ser afetada pela discricionariedade e requer dados que não estão disponíveis publicamente (Cucinelli, Di Battista, Marchese, & Nieri, 2018).

Portanto, o capital da instituição financeira é a principal fonte de proteção contra eventos adversos. Por esse motivo, os bancos precisam manter um capital adequado à sua exposição aos riscos, principalmente ao risco de crédito. Com o desenvolvimento de modelos internos de gestão do risco de crédito, é possível manter um capital regulatório mais adequado ao perfil do negócio, principalmente quando comprovado um risco mais baixo do que o padrão utilizado pelos reguladores. Com isso, é possível uma maior liberação de capital para o aumento das concessões de crédito pelos bancos.

2.5 ESTUDOS RELACIONADOS

Dentre os estudos relacionados, o estudo de Blöchlinger e Leippold (2006) investigou o impacto econômico do poder discriminatório dos modelos de *credit scoring*. Os autores demonstraram que os credores podem incorporar o modelo de pontuação e sua medida de desempenho, baseada na característica de operação do receptor (ROC), nas práticas tradicionais de empréstimo, com base em considerações de valor presente líquido. Foi relacionado o poder discriminatório de um modelo de pontuação de crédito à decisão de crédito ótima, derivando o corte de maximização do lucro e a curva de preços. Além disso, foi estudado um mercado de empréstimos estilizado com bancos que diferem na qualidade de seu modelo de *credit scoring*. Modelos mais poderosos levam a diferenças economicamente significativas no desempenho da carteira de crédito. Este estudo demonstra como um modelo de *credit scoring* bem aplicado pode gerar benefícios para a gestão do risco de crédito dos bancos, impactando positivamente a sua lucratividade.

Khandani, Kim e Lo (2010) utilizaram técnicas de *machine learning* para construir modelos não lineares de previsão não paramétrica de risco de crédito ao consumidor. Os autores propuseram uma medida cardinal de risco de crédito ao consumidor, combinando fatores de crédito tradicionais, como índices de dívida sobre a renda com transações bancárias ao consumidor. Ao analisar os padrões de gastos do consumidor, economias e pagamentos de dívidas, foi possível identificar relações não lineares sutis que são difíceis de detectar nesses conjuntos de dados usando modelos padrão de crédito ao consumidor, como logit, análise discriminante ou pontuação de crédito.

Uma medida do poder de previsão dessa abordagem foi comparar as pontuações previstas do modelo de *machine learning* dos clientes inadimplentes durante o período de previsão com as pontuações previstas dos adimplentes. Assim, uma boa compreensão da escolha do consumidor e dos sinais de alerta de superaquecimento no seu financiamento são essenciais para políticas eficazes de gestão de risco. Usando suposições conservadoras para os custos e benefícios de cortar linhas de crédito com base em previsões de *machine learning*, foi estimada uma economia de custos de 6% a 25% das perdas totais. Este estudo utilizou variáveis relacionadas aos tomadores pessoas físicas para estimar a inadimplência, o que está em linha com este estudo, porém diferentemente deste, utilizou técnicas de *machine learning* para a estimação do modelo de gestão de risco.

Giudici, Hadji-Misheva e Spelta (2019) apontam que as plataformas financeiras digitais introduziram diversas oportunidades aos tomadores de crédito, dentre as quais melhor experiência do cliente, maior velocidade e custos reduzidos, conforme já mencionado neste estudo. Entretanto, estão expostas a riscos mais elevados, como um maior risco de crédito e riscos sistêmicos, devido à grande interconexão entre os tomadores gerada pela plataforma, o que requer modelos de risco de crédito novos e mais precisos para preservar a estabilidade financeira e proteger os consumidores. Assim, os autores sugerem que uma maneira de aumentar a precisão da gestão do risco de crédito dessas plataformas seria utilizando informações topológicas disponíveis em redes de similaridade, como as redes sociais, derivadas de informações financeiras de tomadores de empréstimos. O uso desses coeficientes topológicos descreve a importância dos mutuários e as estruturas da comunidade, sendo empregados como variáveis explicativas adicionais para melhorar o desempenho preditivo dos modelos de *credit scoring*. Este estudo estima o risco de crédito de tomadores pessoas físicas, sendo os concessionários plataformas digitais. Para isso, utiliza variáveis adicionais, incluindo as provenientes de redes sociais, para melhorar a previsão dos modelos de *credit scoring*.

O estudo atual de Roa et al., (2021), apresenta o impacto do uso de dados alternativos originados em um mercado baseado em tecnologia, por meio de aplicativo, em contraste com os dados de bancos tradicionais, sobre os modelos de *credit scoring*. Esse tipo de dados possui um alto poder preditivo em relação ao comportamento do tomador em segmentos pouco atendidos pelas instituições financeiras tradicionais, como indivíduos jovens e de baixa renda. Com a utilização do método TreeSHAP para interpretação do Stochastic Gradient Boosting, o estudo identificou tendências não lineares nas variáveis originadas pelo aplicativo, as quais não estariam disponíveis para os bancos tradicionais, o que representa uma oportunidade para as empresas de tecnologia identificarem corretamente fontes de dados alternativos e lidar com essas informações adequadamente. Este estudo também estima o risco de crédito, com a aplicação de um modelo de *credit scoring* para tomadores pessoas físicas, com a concessão de crédito sendo realizada por meio de aplicativos e, com isso, gerando dados alternativos para análise.

O estudo de Kyeong, Kim e Shin (2022) utilizou dados de log registrados pelo sistema de aplicativos móveis de um banco digital líder na Coreia, o KakaoBank, juntamente com a técnica de regressão logística para criar um modelo de *credit scoring*, utilizando variáveis com altos valores de informação para as instituições financeiras. Os dados de log são relevantes, pois contêm todos os tipos de atividades online, incluindo ações do cliente e operações do sistema. Com isso, o sistema adquire informações de crédito externas durante os processos de empréstimo e, em seguida, os logs relacionados são acumulados, mesmo que não haja nenhuma modificação exibida em tela.

Foram utilizadas informações de transações de cartão de crédito, de empréstimos e seus tomadores e de inadimplência. Quanto mais variáveis baseadas em dados de log, mais transações ou atividades relacionadas às variáveis são feitas pelo cliente. Esses resultados indicam que quanto mais transações um cliente fizer, menor será a probabilidade de inadimplência. Assim, o poder de discriminação do modelo proposto, usando regressão logística, melhorou significativamente em relação a um modelo tradicional, com base nas estatísticas de Kolmogorov–Smirnov. Sendo assim, um banco pode utilizar dados de log acumulados internamente para melhorar os modelos de tomada de decisão, incluindo pontuação de crédito, a um baixo custo. Portanto, a adição de dados de log a um modelo de *credit scoring*, pode melhorar a sua performance. Neste estudo, também foi avaliada a inadimplência de tomadores pessoas físicas, com o uso de um modelo de *credit scoring*, porém com dados de log para as variáveis independentes.

A partir dos estudos relacionados, depreende-se que desde 2006 até os dias atuais ainda há diversos estudos utilizando o modelo de *credit scoring* para a gestão do risco de crédito, por meio da previsão de inadimplência. Além disso, fica claro um aumento no uso de dados alternativos gerados, principalmente, por novos meios de concessão de crédito, com o uso cada vez mais constante de novas tecnologias.

3 PROCEDIMENTOS METODOLÓGICOS

Com a finalidade de responder à questão problema e atingir os objetivos propostos, o presente estudo foi desenvolvido em duas etapas. A primeira etapa é direcionada a compreender os conceitos principais relacionados às instituições financeiras digitais, ao risco de crédito e seus modelos de gestão, incluindo a identificação das principais normas emitidas pelo Banco Central do Brasil que regulamentam o assunto. A segunda etapa refere-se à aplicação do modelo de *credit scoring* no banco digital X, sendo mais específica em termos da análise da inadimplência.

3.1 CLASSIFICAÇÃO METODOLÓGICA DO ESTUDO

A primeira etapa apresenta uma abordagem qualitativa para atender aos objetivos de descrever as principais características de instituição financeira digital e de risco de crédito, descrever os principais modelos de risco de crédito existentes na literatura e apresentar as normas do Banco Central do Brasil quanto ao risco de crédito. Dessa forma, por meio do procedimento técnico de pesquisa documental (Bowen, 2009), com a utilização de fontes primárias, como a legislação em vigor e documentos oficiais do BACEN, fontes secundárias, como o relatório de gestão de risco do banco digital X e suas demonstrações contábeis, e estudos seminais sobre o tema, foram revisadas as características referentes às instituições financeiras digitais, ao risco de crédito e seus modelos de gestão.

A segunda etapa buscou alcançar o objetivo de aplicar o modelo de *credit scoring* utilizado para a avaliação de risco de crédito de um banco tradicional em um banco digital, considerando os parâmetros do BACEN e tomando como base o banco digital X. Para isso, apresenta uma abordagem quantitativa (Raupp, & Beuren, 2008), pois foram utilizadas estatísticas que, juntamente com técnicas econométricas, como a Regressão Logística Binária, permitiram a estimação de um modelo de *credit scoring* para a previsão da inadimplência e, com isso, uma melhor gestão do risco de crédito.

Quanto aos objetivos, é uma pesquisa descritiva, de acordo com Singh (2019), pois buscou descrever os principais conceitos de instituições financeiras digitais e de risco de crédito, as principais características da regulamentação bancária no que diz respeito à risco de crédito, e os modelos de risco de crédito mais utilizados no setor bancário. Além disso, descreve as características mais importantes para a elaboração do modelo de *credit scoring*, demonstra a relação entre as variáveis e sua relevância para o modelo.

3.2 POPULAÇÃO E AMOSTRA

A população (Colauto, & Beuren, 2008) deste estudo é composta pelo segmento de crédito pessoal da carteira de crédito do banco totalmente digital X, com regulação quanto ao risco de crédito conforme o S4, o qual é composto pelas instituições de porte inferior a 0,1% do PIB brasileiro, de acordo com a Resolução n. 4.553 do Banco Central do Brasil, de 2017. A instituição selecionada possuía uma carteira de crédito bruta de R\$ 2,2 bilhões, em dezembro de 2020, dos quais 48% dizem respeito à crédito pessoal, 42% a crédito consignado e o restante à cartões de crédito.

Dessa forma, a população do presente estudo é composta pelo montante dos contratos referentes a carteira de crédito pessoal, a qual equivale a R\$ 1,06 bilhões no referido período. A escolha desta carteira está fundamentada pelo fato de ser a carteira mais exposta a fatores de risco e à concorrência no mercado. A sobrevivência no longo prazo dos bancos digitais dependerá fundamentalmente da sua expertise na área de crédito². A amostra (Colauto, & Beuren, 2008), por sua vez, é formada pela seleção aleatória de dois grupos de indivíduos da carteira de crédito pessoal do banco digital X, sendo:

Grupo A = clientes classificados como risco nível A, ou seja, os que não possuem atraso no pagamento de suas operações.

Grupo H = clientes inadimplentes.

A amostra foi selecionada aleatoriamente por uma quantidade de contratos de tomadores de crédito estatisticamente significativa, com um nível de 90% de confiança, admitindo um nível de erro de 5%. A amostra conta com 1000 contratos de clientes, segregados entre adimplentes e inadimplentes. As pessoas físicas inadimplentes são classificadas no nível de risco H, pois possuem atraso no pagamento das suas operações entre 180 e 360 dias, conforme Resolução n. 2.682 de 1999 do BACEN.

3.3 VARIÁVEIS DA PESQUISA

Em relação às variáveis, em um modelo de *credit scoring*, a variável dependente representa a qualidade do crédito. Para a classificação das observações em conformidade com a qualidade do crédito, foram selecionadas variáveis que possuem impacto na situação de inadimplência dos empréstimos.

² (<https://diariodocomercio.com.br/financas/novos-concorrentes-colocam-bancos-digitais-a-prova/>)

Para dívidas relacionadas aos empréstimos de crédito pessoal, que são concedidos às pessoas físicas, as variáveis relevantes podem incluir idade, ocupação, localização, renda e bens (Saunders, & Cornett, 2017). Adicionalmente, também foram utilizadas variáveis com base no estudo de Khandani, Kim e Lo (2010), mas com as adaptações necessárias para o negócio do banco digital X, conforme suporte prestado pelo Gerente de Riscos da referida instituição.

Sendo assim, as variáveis utilizadas no modelo são:

Variável dependente, conforme tabela 1:

Tabela 1

Variável Dependente

	Variável
Y	Classificação adimplente/inadimplente (grupo)

Variáveis independentes, de acordo com a tabela 2:

Tabela 2

Variáveis Independentes

	Variáveis
X ₁	Estado Civil do Cliente (ds_estadocivil)
X ₂	Sexo do Cliente (cdsexo)
X ₃	Região do Cliente (ds_regiao)
X ₄	Idade do Cliente (nr_idade)
X ₅	Renda Bruta do Cliente (vl_rendabruta)
X ₆	Saldo em Operações Consignadas (saldo_consignado)
X ₇	Saldo em Operações Não Consignadas (saldo_nconsignado)
X ₈	Soma das PMTs (parcelas) OP. Consignadas (pmt_consignado)
X ₉	Soma das PMTs (parcelas) OP. Não Consignadas (pmt_nconsignado)
X ₁₀	Soma do Valor Bruto OP. Consignadas (vl_emprest_consignado)
X ₁₁	Soma do Valor Bruto OP. Não Consignadas (vl_emprest_nconsignado)

Em suma, foram utilizadas variáveis com o objetivo de caracterização dos tomadores pessoas físicas, tais como estado civil, sexo, região, idade e renda bruta, e foram testadas a sua influência no evento de inadimplência. Também, foram utilizadas variáveis relativas às operações do cliente, incluindo valores, saldos e parcelas das operações consignadas e não consignadas de crédito pessoal, para mensurar o seu impacto na inadimplência.

3.4 COLETA DOS DADOS

O procedimento técnico para a coleta de dados é documental (Colauto, & Beuren, 2008). Foi utilizada documentação primária, coletada nos sítios eletrônicos do Banco Central do Brasil, do BIS e da instituição financeira digital em estudo, composta pela legislação de risco e de regulamentação de instituições financeiras digitais, incluindo resoluções e circulares emitidas pelo BACEN.

Foram realizadas consultas à documentação direta, como as demonstrações contábeis consolidadas publicadas pelo banco digital X, com data base de dezembro de 2020 e ao seu relatório de gestão de riscos do mesmo período. Com isto, foi possível fazer comparações entre os períodos analisados nas regressões logísticas binárias da etapa de análise dos dados. Além disso, após reunião com o gestor de risco do banco digital X, em junho de 2021, foram disponibilizados documentos, em meio digital, com os dados necessários para a estimação do modelo de *credit scoring*. É válido mencionar que a coleta dos dados se deu em conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD), Lei n. 13.709 de 2008, havendo a remoção de dados sensíveis anteriormente à disponibilização do documento.

3.5 ANÁLISE DOS DADOS

A análise dos dados se deu por análise documental (Bowen, 2009) e descritiva (Colauto, & Beuren, 2008). Os documentos coletados foram analisados para a descrição das principais características das instituições financeiras digitais, do risco de crédito e seus modelos, e para a apresentação das normas legais mais significativas sobre o tema. A análise e interpretação dos dados forneceu subsídios para a aplicação do modelo de *credit scoring* de um banco tradicional no banco digital X.

O modelo de *credit scoring* estimado para a gestão do risco de crédito possui a análise dos seus dados por meio da técnica econométrica de regressão logística binária. Esta técnica é usada para a mensuração do risco de crédito, referente à empréstimos e financiamentos, pois avalia a probabilidade de inadimplência dos clientes da carteira (Vicente, 2001). Assim, busca separar os clientes adimplentes dos inadimplentes, com a predição de presença ou ausência de uma característica, com base nas variáveis independentes, retornando valores 0 ou 1 para a variável dependente. Para este estudo, 1 representa a inadimplência, restando 0 para a adimplência.

A classificação dos clientes como inadimplentes ou adimplentes se dá pelo cálculo da probabilidade de inadimplência, baseada nas informações de cada cliente para as variáveis explicativas, conforme a equação gerada pelo modelo. Os clientes com probabilidade estimada de inadimplência inferior a 0,5, são classificados como adimplentes, sendo os demais classificados como inadimplentes. A técnica de regressão logística possui ponto de corte padrão de 0,5, o qual, conforme Hair Júnior et al. (1998), representa a probabilidade de ocorrência do evento de acordo com o critério de chances iguais ou aleatoriedade.

3.6 DESCRIÇÃO DO MODELO

A regressão logística (logit), conforme já mencionado neste estudo, é uma técnica estatística utilizada para a separação de dois grupos, com o objetivo de obter a probabilidade de que uma observação pertença a um grupo determinado, em função do comportamento das variáveis independentes (Hair Júnior et al., 1998). Esta técnica é muito utilizada para a análise de dados com resposta binária. Por meio de um modelo, relaciona a variável resposta, a qual corresponde a variável dependente, às variáveis independentes, que influenciam ou não a probabilidade de ocorrência de um evento específico.

A técnica de regressão logística, no que diz respeito à mensuração do risco relativo a operações de concessão de crédito, serve para a avaliação da inadimplência de determinado conjunto de clientes, assumindo que a probabilidade de inadimplência é logisticamente distribuída, apresentando resultado dicotômico de 0 ou 1. Como a variável dependente é considerada não-métrica, é inserida por meio do uso de variáveis *dummy*, as quais possuem efeito binário, assumindo valor 0 para indicar a ausência de um atributo e 1 para indicar a sua presença (Gujarati, 2000).

Para a aplicação da regressão logística, deve-se obter a informação sobre a ocorrência ou não de determinado evento, como a situação de inadimplência ou não de um cliente. Utiliza-se, então, um valor dicotômico ou binário para a variável dependente, com valores 0 ou 1, a depender da ocorrência ou não do evento considerado. A partir desse valor dicotômico, a regressão logística calcula a probabilidade de esse evento acontecer ou não (Hair Júnior et al., 1998). Sendo assim, conforme Gujarati (2000), a partir de um evento dependente, considerando a variável binária Y , e variáveis independentes, que neste estudo são X_1 a X_{11} , a função de distribuição logística, é dada por:

$$P_i = \Pr [Y_i = 1 | X = X_i] = \frac{1}{1 + e^{-Z}},$$

Em que:

P_i = representa a probabilidade de acontecer o evento;

$$Z = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_i X_i;$$

Para a interpretação desta equação para a mensuração do risco de crédito, considera-se que as variáveis independentes, constituídas por dados que caracterizam o tomador e suas operações de crédito, representam os fatores que influenciam a inadimplência. Assim, P é a probabilidade de inadimplência do cliente, ou seja, a probabilidade condicional de Y assumir o valor 1. Por fim, as medidas das variações na proporção das probabilidades são representadas pelos coeficientes estimados β s.

3.7 ESTIMAÇÃO DO MODELO

Para a estimação do modelo, foram executadas quatro regressões logísticas binárias, referentes a quatro datas base distintas, compreendendo o período de junho de 2019 a dezembro de 2020. Foi utilizado o software IBM SPSS Statistics, por meio do método Forward LR (Forward Stepwise - Razão de Verossimilhança), o qual é uma técnica com a função de encontrar a melhor combinação de variáveis predictoras (Tsai, 2009). Essa técnica encontra a melhor variável predictoras e adiciona, passo a passo, variáveis que atendam a algum critério específico, tendo como resultado final, a combinação de variáveis predictoras com coeficientes significativos, ou seja, que sejam relevantes para o modelo.

Para a execução das regressões logísticas, foram realizadas as seguintes etapas:

- a) A base de dados foi tratada para permanecer apenas dados das datas base 31/12/2020, 30/06/2020, 31/12/2019 e 30/06/2019;
- b) Foram importadas, no software IBM SPSS Statistics, as planilhas eletrônicas em Microsoft Excel ajustadas para as referidas datas base;
- c) No software IBM SPSS Statistics, foi selecionada a opção correspondente à Regressão Logística Binária, no menu Análise;
- d) A variável Classificação adimplente/inadimplente - Grupo (Y) foi definida como a variável dependente;

- e) As variáveis de X_1 a X_{11} , as quais são Estado Civil do Cliente, Sexo do Cliente, Região do Cliente, Idade do Cliente, Renda Bruta do Cliente, saldo em Operações Consignadas, saldo em Operações não Consignadas, soma das PMTs de Operações Consignadas, soma das PMTs de Operações não Consignadas, soma do Valor Bruto de Operações Consignadas e soma do Valor Bruto de Operações não Consignadas, foram definidas como variáveis independentes;
- f) Foi selecionado o método Forward LR (Forward Stepwise - Razão de Verossimilhança);
- g) Em seguida, no menu Opções, foram assinalados os seguintes campos: Gráficos de classificação, Qualidade do ajuste de Hosmer e Lemeshow, Listagem por caso de resíduos, Correlações de estimativas e Histórico de Iteração;
- h) Em relação à Probabilidade para Stepwise, foi definido o parâmetro de 0,05 para a Entrada e o parâmetro de 0,10 para a remoção;
- i) O corte de classificação foi definido em 0,5;
- j) E o máximo de iterações em 20;
- k) Por fim, foi selecionada a inclusão da constante no modelo.

Conforme mencionado anteriormente, os modelos de regressão logística são uma ferramenta para estimar a probabilidade de sucesso de uma variável de resposta binária, condicionando-se a diversas variáveis explicativas para construir um modelo que se adapte bem aos dados observados (Surjanovic, & Loughin, 2021). Sendo assim, é necessário aplicar alguns testes estatísticos para avaliar a qualidade e adequação do modelo.

O teste de ajuste de Omnibus serve para testar a especificação correta de meios condicionais paramétricos quando não há informações sobre a alternativa possível (Escanciano, 2009). Utiliza uma estatística de teste baseada na função geradora de probabilidade que é assintoticamente normalmente distribuída, permitindo sua implementação direta. O teste mostra um nível de significância empírica muito próximo do nominal para tamanhos de amostra moderados e possui poder empírico bastante satisfatório, se comparado ao teste qui-quadrado (Di Noia et al., 2021). Este teste é adequado ao modelo proposto, pois conforme Gijbels, Omelka e Veraverbeke (2021), a estatística de teste é projetada para ter poder contra muitas alternativas e pode ser usada para testar uma variedade de efeitos covariáveis.

O teste Log -2 de verossimilhança testa a viabilidade do modelo de regressão logística, incluindo a variável independente com base na diminuição do valor de probabilidade logarítmico inicial de -2, conforme são executadas as etapas do método Forward LR. De acordo com Azis (2021), com a diminuição do valor de Log -2 de verossimilhança, é demonstrado que o modelo utilizado é um bom modelo de regressão e é viável de usar.

O teste do R^2 de Nagelkerke indica a capacidade da variável independente em explicar a variável dependente, sendo uma versão ajustada do R^2 que adequa a escala da estatística para cobrir o intervalo de 0 a 1 (Dwiarti, Hazmi, Santosa, & Rahman, 2021). Assim, no modelo proposto, quanto mais próximo de 1 for o R^2 de Nagelkerke, mais as variáveis incluídas no modelo explicam a variação da inadimplência.

Para testar a adaptação do modelo, o teste de Hosmer-Lemeshow (HL), introduzido por Hosmer e Lemeshow (1980), examina a hipótese nula de que a estrutura do modelo ajustado está correta. A estatística de teste é uma estatística de Pearson que compara contagens de eventos observados e esperados de dados agrupados, de acordo com valores ajustados ordenados do modelo. Assim, a regra de decisão para o teste é baseada na comparação da estatística de teste com uma distribuição qui-quadrado com graus de liberdade que dependem do número de grupos usados para criar a estatística de teste (Surjanovic, & Loughin, 2021). Com isso, a estimativa do qui-quadrado destina-se a determinar o efeito das variáveis X_1 a X_{11} na previsão de inadimplência.

Adicionalmente, foi analisada a razão de chances, ou *odds ratio*, que significa a razão entre a probabilidade de ocorrência de eventos e a probabilidade de não ocorrência. A regressão logística mapeia o número de intervalos de números reais para a faixa de 0 a 1 e fornece a probabilidade de ocorrência do evento (Li, 2019). Como neste caso foi definido o valor crítico de 0,5, o teste julga se o evento de inadimplência ocorre ou não, em relação a esse limite.

Também, foi aplicado o teste de Wald, pois é comumente utilizado como teste de significância para os coeficientes de regressão logística (Demidenko, 2007). Este teste verifica se as variáveis independentes apresentam uma relação estatisticamente significativa com a variável dependente, pois testa se cada coeficiente é significativamente diferente de zero. Assim, esse teste é importante para testar se todas as variáveis independentes do modelo são relevantes para a previsão da inadimplência.

Portanto, o modelo foi estimado com a utilização de regressão logística binária para as quatro datas base mencionadas. Além disso, foram aplicados os testes apresentados nesta subseção para validar a adequação e qualidade do modelo. Com isso, os resultados foram

interpretados para apresentar as principais características da previsão de inadimplência para o gerenciamento de risco de crédito de uma instituição financeira digital, por meio da aplicação do modelo de *credit scoring*.

4 RESULTADOS

Os resultados da pesquisa estão apresentados em três seções: (4.1) Instituições Financeiras Digitais, (4.2) Risco, com a subseção (4.2.1) Risco de Crédito e (4.3) Modelo de *Credit Scoring*, com as subseções (4.3.1) Apresentação dos Resultados das Regressões e (4.3.2) Discussão.

4.1 INSTITUIÇÕES FINANCEIRAS DIGITAIS

As instituições financeiras, conforme veiculado na Lei n. 4.595 de 1964, em seu artigo 17, são as pessoas jurídicas públicas ou privadas, cuja atividade principal ou acessória, seja a coleta, intermediação ou aplicação de recursos financeiros próprios ou de terceiros, em moeda nacional ou estrangeira, e a custódia de valor de propriedade de terceiros. Dentre essas instituições, desponta a modalidade de bancos digitais, para a qual o arcabouço regulatório não determina regulamentação específica. Assim, para a constituição de um banco digital, conforme o seu modelo de negócio, é necessária a autorização para funcionamento como uma instituição bancária, podendo ser um banco comercial, de investimento ou múltiplo (BACEN, 2019).

Conforme mencionado, devido à falta de regulamentação específica para os bancos digitais, eles se denominam como tais por estratégia operacional e de mercado. Essas instituições têm a peculiaridade de um relacionamento diferenciado e totalmente remoto com os seus clientes, oferecendo-lhes vantagens e experiências como facilidade de acesso e integração com outras conveniências e custos mais baixos de serviços. Porém, enfrentam alguns desafios como desenvolver análises mais complexas, agir de maneira mais proativa e informativa e ter uma visão mais abrangente do relacionamento com os clientes. Contudo, tanto esses bancos, como os tradicionais, buscam uma maior inserção no mercado de negócio digital (BACEN, 2019).

Ainda, com o avanço de novas tecnologias, o BACEN, juntamente com o CMN, autorizou a nova modalidade de *Fintechs*, as quais estão regulamentadas, desde abril de 2018, por meio das Resoluções n. 4.656 e 4.657. As *Fintechs* são um modelo de negócio exclusivamente digital, baseado em tecnologias inovadoras. Elas possuem normas de funcionamento diferenciadas, como um menor limite de patrimônio líquido e capital social, processo de autorização mais simples, opção pelo enquadramento no segmento 5 (S5), o qual

permite uma gestão de risco mais flexível, e a participação direta de fundos de investimento no seu controle (BACEN, 2019).

Portanto, ao solicitar uma licença bancária, as entidades com um modelo de negócios habilitado para tecnologia devem cumprir os mesmos procedimentos e requisitos de licenciamento que os modelos de negócios tradicionais. Assim, os bancos digitais podem oferecer uma gama completa de produtos e serviços bancários aos seus clientes, pois são licenciados para receber depósitos e utilizar estes fundos para realizar suas atividades bancárias. Incorrem em riscos financeiros semelhantes aos bancos tradicionais, incluindo risco de crédito, risco de mercado e risco de liquidez. Além disso, incorrem em riscos específicos, como risco operacional e cibernético, os quais podem ser acentuados devido à natureza de sua operação (Ehrentraud, Ocampo, & Vega, 2020).

4.2 RISCO

A Resolução 4.557 de 2017, do Banco Central do Brasil, dispõe sobre as estruturas de gerenciamento de risco e de capital para as instituições financeiras, veiculando as principais definições sobre risco de crédito e o seu gerenciamento, incluindo a necessidade de gestão de capital para suportar possíveis perdas.

As instituições autorizadas a funcionar pelo Banco Central do Brasil, financeiras ou não, enquadradas nos Segmentos de 1 (S1) a 4 (S4), devem ter uma estrutura de gerenciamento contínuo e integrado de riscos e de capital. Essa estrutura deve ser compatível com o modelo de negócio, com a natureza e com a complexidade das operações, produtos e serviços da instituição, deve ser adequada ao seu perfil de riscos e à sua importância sistêmica, proporcional à sua dimensão e relevância de exposição aos riscos, e com condições de avaliar os riscos das condições macroeconômicas e dos mercados em que atua.

A Resolução 4.553 de 2017, do Banco Central do Brasil, estabelece a segmentação do conjunto das instituições financeiras e das instituições autorizadas a funcionar pelo BACEN. Os segmentos vão do Segmento 1 (S1) ao Segmento 5 (S5), considerando o porte e a atividade internacional das instituições que compõem cada segmento. O Segmento 4 (S4), o qual está sendo estudado nesta pesquisa, é composto pelas instituições de porte inferior a 0,1% do PIB. O porte é atribuído com base na razão entre o valor da Exposição Total da instituição e o valor do PIB do Brasil.

As referidas instituições devem documentar os níveis de apetite por riscos na Declaração de Apetite por Riscos (RAS), considerando os tipos de riscos e os níveis que a

instituição está disposta a assumir de cada um deles, os seus objetivos estratégicos, a sua capacidade de gerenciamento prudente e efetivo desses riscos, e o ambiente regulatório no qual está inserida e a competitividade do mercado.

Especificamente em relação à estrutura de gerenciamento de riscos, ela deve identificar, mensurar, avaliar, monitorar, reportar, controlar e mitigar todos os riscos da instituição, dentre eles o risco de crédito, de mercado, o risco de variação das taxas de juros para instrumentos classificados na carteira bancária (IRRBB), risco operacional, de liquidez, os riscos social, ambiental e climático, incluídos pela Resolução 4.943 de 2021, e demais riscos relevantes, mesmo que não estejam na apuração da quantia dos ativos ponderados pelo risco (RWA). O referido gerenciamento deve ser integrado com o objetivo de minimizar os efeitos adversos das interações entre os riscos mencionados.

A estrutura de gerenciamento de riscos deve ser composta por políticas e estratégias de gestão de risco que estejam de acordo com a Declaração de Apetite por Riscos (RAS), sistemas e procedimentos de gerenciamento de riscos e sua avaliação periódica, processos efetivos de rastreamento e reporte tempestivo de exceções às regras firmadas, e políticas e controles capazes de identificar previamente o surgimento de novos riscos. Ainda, deve prever programa de testes de stress, papéis e responsabilidades de gerenciamento de riscos documentados, avaliação contínua das estratégias de gestão de riscos e a elaboração de relatórios gerenciais tempestivos para a diretoria da instituição.

Adicionalmente, as políticas e estratégias de gestão de riscos devem ser compartilhadas com o pessoal da instituição, em todos os níveis, incluindo prestadores de serviços terceirizados relevantes e deve ser objetiva e de fácil compreensão por todas as áreas envolvidas. Para que a referida estrutura funcione adequadamente, a instituição precisa manter quantidade suficiente de profissionais tecnicamente qualificados nas áreas sujeitas à assunção de riscos.

O programa de testes de estresse, segundo a referida Resolução 4.557 de 2017, por exemplo, é o conjunto coordenado de processos e rotinas, dotado de metodologias, documentação e governança próprias, com o objetivo principal de identificar potenciais vulnerabilidades da instituição. O teste de estresse é o exercício, com finalidade definida, de avaliação prospectiva dos potenciais impactos de eventos e circunstâncias adversos na instituição ou em um portfólio específico. A análise de sensibilidade, por sua vez, é a metodologia de teste de estresse que permite avaliar o impacto decorrente de variações em um parâmetro relevante específico no capital da instituição, em sua liquidez ou no valor de um portfólio. Já, a análise de cenários é uma metodologia de teste de estresse que permite avaliar,

ao longo de um determinado período, o impacto decorrente de variações simultâneas e coerentes em um conjunto de parâmetros relevantes no capital da instituição, em sua liquidez ou no valor de um portfólio.

Os requisitos para o teste de estresse são abranger os riscos relevantes, avaliar o impacto de concentrações significativas de riscos, utilizar a análise de sensibilidade, a análise de cenários, o teste de estresse reverso, prever a utilização de premissas e parâmetros adversos adequadamente severos e documentar os principais processos.

A instituição deve assegurar o uso dos resultados do teste de estresse na identificação, na mensuração, na avaliação, no monitoramento, no controle e na mitigação dos riscos da instituição. Esse teste deve ser realizado de forma integrada para os diversos riscos e unidades de negócios da instituição e considerar os efeitos adversos resultantes das interações entre os riscos, prevendo a utilização de cenário comum.

Os resultados do programa de testes de estresse devem ser incorporados nas decisões estratégicas da instituição, como na revisão dos níveis de apetite por riscos e na revisão das políticas, das estratégias e dos limites estabelecidos para fins do gerenciamento de riscos e do gerenciamento de capital. Devem ser utilizados na avaliação dos níveis de capital e de liquidez da instituição e na elaboração dos respectivos planos de contingência e de recuperação, no processo estruturado de comunicação e na avaliação da adequação de capital.

Por fim, as políticas para a gestão de continuidade de negócios devem identificar, classificar e documentar os processos críticos de negócio, e avaliar os potenciais efeitos da interrupção desses processos. Ainda, é fundamental a elaboração de estratégias para assegurar a continuidade das atividades da instituição e limitar perdas causadas pela interrupção dos processos críticos de negócio, e de planos que estabeleçam procedimentos e prazos estimados para reinício e recuperação das atividades no caso de interrupções. Por fim, é necessário testar e revisar os planos de continuidade de negócios com periodicidade adequada.

Em suma, as instituições financeiras devem ter uma estrutura específica para o gerenciamento de riscos, sendo capazes de identificar, mensurar, avaliar, monitorar, reportar, controlar e mitigar todos os seus riscos. Devem documentar os níveis de apetite por riscos na Declaração de Apetite por Riscos (RAS), abrangendo todos os riscos envolvidos no negócio e de acordo com os seus objetivos estratégicos. Devem revisar periodicamente os níveis de riscos, aplicando testes para a sua mensuração, desenvolver estratégias para a sua mitigação e manter tudo documentado e disponível aos órgãos reguladores.

4.2.1 Risco de Crédito

Uma das atividades principais dos bancos é a intermediação financeira. A operação de crédito pode ter características diversas, como sendo um compromisso financeiro assumido para realização de mútuo, de emissão e aceite de título, pela abertura de crédito, aquisição financiada de bens, recebimento antecipado de valores provenientes da venda a termo de bens e serviços, arrendamento mercantil e outras operações semelhantes, incluindo a utilização de derivativos financeiros, conforme veiculado na Lei de Responsabilidade Fiscal (Lei Complementar n. 101, 2000).

Para muitas instituições financeiras, as operações de crédito são a maior e mais clara fonte de risco de crédito. O risco de crédito é definido como a possibilidade de um tomador de empréstimo ou contraparte deixar de cumprir suas obrigações de acordo com os termos acordados. Por esse motivo, de acordo com o Comitê de Basileia, a gestão de risco de crédito tem o objetivo de maximizar a taxa de retorno ajustada ao risco de um banco, mantendo a exposição ao risco de crédito dentro de parâmetros aceitáveis (BCBS, 2000).

A Resolução 4.557 de 2017, define risco de crédito como a possibilidade de ocorrência de perdas associadas ao não cumprimento pela contraparte de suas obrigações nos termos acordados. Também, pela desvalorização, redução de remunerações e ganhos esperados em instrumento financeiro decorrentes da deterioração da qualidade creditícia da contraparte, do interveniente ou do instrumento mitigador, pelos custos de recuperação de exposições caracterizadas como ativos problemáticos e pela reestruturação de instrumentos financeiros.

Essa definição ainda considera o risco de crédito da contraparte, o qual é a possibilidade de perdas decorrentes do não cumprimento de obrigações relativas à liquidação de operações que envolvam fluxos bilaterais, juntamente com a negociação de ativos financeiros ou de derivativos, e o risco país, ou seja, a possibilidade de perdas relativas ao não cumprimento de obrigações associadas a contraparte ou instrumento mitigador localizados fora do País, o qual inclui o risco soberano, cuja exposição é assumida perante governo central de jurisdição estrangeira. As contrapartes conectadas são as que compartilhem o risco de crédito perante a instituição, inclusive por meio de relação de controle e, por este motivo, devem constituir uma única contraparte para fins de gerenciamento de riscos.

Portanto, seguindo os parâmetros internacionais, o Banco Central do Brasil, por meio da Resolução n. 4.557 de 2017, determina que as instituições financeiras devem ter uma estrutura de gerenciamento contínuo e integrado de riscos, as quais devem ser compatíveis

com o modelo de negócio, adequadas ao perfil de riscos, proporcionais à dimensão e à relevância da exposição da instituição aos riscos e capaz de avaliar os riscos decorrentes das condições macroeconômicas e dos mercados em que atua. Dessa forma, essa estrutura deve identificar, mensurar, avaliar, monitorar, reportar, controlar e mitigar o risco de crédito.

A referida resolução ressalta que para o gerenciamento do risco de crédito, a exposição deve ser classificada como ativo problemático quando estiver atrasada em mais de noventa dias ou houver de que não será cumprida sem o acionamento de garantias. Além disso, os bancos devem documentar os seus níveis de apetite por riscos na Declaração de Apetite por Riscos (RAS), considerando os seus objetivos estratégicos, os tipos de riscos e os respectivos níveis que estão dispostos a assumir, as condições de competitividade e o seu ambiente regulatório, e a sua capacidade de gerenciar riscos de forma prudente e efetiva.

A Resolução n. 2.682/99 do Banco Central do Brasil estipula que os bancos e as demais instituições financeiras autorizadas a funcionar pelo BACEN devem classificar as suas operações de crédito em ordem crescente de risco, conforme o prazo de inadimplência, e avaliar e ponderar cada um dos seus elementos com base em critérios consistentes e verificáveis. Para isso, utiliza a constituição de provisão, a qual possui um percentual para cada nível de classificação de risco, que deve ser multiplicado pelo montante da operação. Também, dá poderes ao BACEN de fiscalização e de emissão de normas complementares, podendo requerer a reclassificação de operações, a exigência de provisionamento adicional, de providências saneadoras e demais informações, dentre outros.

Ademais, por meio da Resolução n. 4.193/2013 e da Circular n. 3.648/2013, o BACEN regulamenta que as instituições financeiras devem seguir as orientações de Basileia para o cálculo do requerimento de capital para as exposições ao risco de crédito. A Resolução n. 4.783/20 reduziu temporariamente a exigência de capital para amenizar os impactos da pandemia de Covid-19 no Brasil e garantir a continuidade do fluxo de crédito no país.

A IFRS 9, e o CPC 48 no Brasil, tratam do reconhecimento da PECLD – Perda Estimada com Créditos de Liquidação Duvidosa, por meio do cálculo de perdas futuras estimadas. Os seus requisitos de redução ao valor recuperável objetivam o reconhecimento de perdas de crédito esperadas para todos os instrumentos financeiros. Para isso, utiliza-se de variáveis estatísticas para estimar o montante de perdas de créditos que irão ocorrer no futuro e irão impactar os fluxos de caixa. Dessa forma, os bancos devem provisionar um valor adequado para fazer frente a essas perdas. Assim, essa norma introduzida em 2018, alterou o formato de reconhecimento de perdas incorridas, por perdas estimadas que impactam os fluxos de caixa, também conhecido como *forward-looking*.

Portanto, para o gerenciamento do risco de crédito, as instituições financeiras devem seguir, além da regulamentação interna do Banco Central, também as orientações do comitê de Basileia e desejável cumprimento das novas especificações advindas da IFRS 9, CPC 48 no Brasil, quanto ao reconhecimento da PECLD (Engelmann, 2021). Vale mencionar que como a Resolução n. 2.682/99 permanece em vigor, muitos bancos já utilizam as duas formas de provisionamento, tanto por esta quanto pela IFRS 9.

4.3 MODELO DE *CREDIT SCORING*

O banco digital X possuía uma carteira de crédito bruta de 2,2 bilhões em 31/12/2020, representando um aumento de 35% em relação ao mesmo período do ano anterior e um aumento de 36,9% se comparado a junho do mesmo ano. A carteira de crédito pessoal para correntistas teve um aumento expressivo de 124,6%, em linha com a estratégia do banco de estreitar o relacionamento com seus clientes. Isso representa mais que o dobro do total de não-correntistas. Outro ponto relevante a ser mencionado, é que cerca de 95% da carteira de crédito possui contraparte de âmbito federal, o que traz uma maior segurança, melhorando o seu perfil de risco. A PECLD, para a qual o banco segue as orientações, além da Resolução n. 2.682/99, da IFRS 9, representou 8,5% da carteira de crédito bruta, com uma queda de 17,9% quando comparada ao mesmo período do ano anterior.

Quanto à qualidade da carteira de crédito, 85% possui classificação de risco de nível A, exigindo um provisionamento de apenas 0,5%, conforme a Resolução n. 2.682 de 1999 do BACEN. Apenas 6% da carteira está classificada em nível de risco H, requerendo provisionamento de 100%. Sendo assim, os 9% restantes estão distribuídos nos demais níveis de risco. Com isso, percebe-se uma melhora expressiva na qualidade da carteira de crédito se comparada com as datas base de 30/06/2020, a qual era de 76% com nível de risco A, e 31/12/2019, que possuía 57% da carteira classificada no referido nível. Em relação à gestão de capital, o banco digital X possuía, em dezembro de 2020, índice de Basileia de 12,5% e índice de Basileia amplo, de 12,4%, o qual além de ser a razão entre o Patrimônio de Referência e os ativos ponderados pelo risco, considera no denominador o risco bancário.

Em relação à gestão do risco de crédito, o banco digital X utiliza a abordagem padronizada, cumprindo os requisitos de metodologia estipulados pelo banco central. Isto significa que no momento, conforme informado em seu relatório de gestão de riscos, o banco não desenvolveu uma abordagem baseada em *ratings* internos. Para complementar, no seu relatório de Demonstrações Financeiras, o banco afirma estar evoluindo em questões

relacionadas à modelagem de crédito, com a utilização de novos *bureaus* e novas variáveis internas, que permitem a segmentação de grupo de riscos e uma oferta de crédito mais customizada a cada classe de tomadores, com taxas mais apropriadas para cada nível de risco.

Para analisar a aplicação de um modelo de *credit scoring* de um banco tradicional no banco digital X, foi estimado o modelo por meio de regressão logística binária, com a utilização do método Forward LR, por meio do software estatístico IBM SPSS Statistics em quatro datas base, 31/12/2020, 30/06/2020, 31/12/2019 e 30/06/2019. Conforme explicado na seção de metodologia, foram utilizados 1000 contratos distribuídos entre adimplentes e inadimplentes, em ambas as regressões. Sendo assim, são apresentados os resultados com data base de dezembro de 2020 e em seguida os resultados das demais datas base. Por fim, ambos os resultados são analisados e comparados entre si.

4.3.1 Apresentação dos Resultados das Regressões

Primeiramente, vale mencionar que em ambas as regressões logísticas foram incluídos 992 casos, ou seja 99,2% da amostra, restando 8 (0,8%) casos omissos. Esses casos omissos referem-se a variável X_1 - Estado Civil do Cliente, a qual não estava informada na base de dados para estes casos. A codificação da variável dependente se dá com 1 para a inadimplência e 0 para a adimplência. No bloco inicial, no qual o modelo está sem variável previsora, isto é, variável independente, em ambas as regressões, o Log-2 de verossimilhança inicial é de 1375,200 e coeficientes constante de -0,04. Com isso, a estimação foi finalizada no número de iteração 2, porque as estimativas de parâmetro mudaram e foram alteradas para menos de 0,001. Ainda, no bloco inicial, a tabela de classificação apresenta 497 contratos como adimplentes e 495 como inadimplentes. Como o bloco inicial ainda não possui variável independente, o modelo classificou todos os contratos como adimplentes, apresentando um acerto de 50,1%, conforme demonstrado na tabela 3 a seguir:

Tabela 3
Tabela de Classificação^{a,b}

Observado	Grupo	Previsto		
		0	1	
Etapa 0 grupo	0	497	0	100
	1	495	0	0
Porcentagem global				50,1

Nota. a. A constante está incluída no modelo.

b. O valor de recorte é ,500

Em relação à regressão com data base em 31/12/2020, no bloco inicial, é apresentada a relação de variáveis não incluídas no modelo, composta pelas 11 variáveis independentes citadas na seção de metodologia. Este resultado demonstra que existem duas variáveis irrelevantes para o modelo, as quais são Região e Idade. Entretanto, o modelo pode seguir com a inclusão das nove variáveis restantes: Estado Civil, Sexo, Renda Bruta, saldo em Operações Consignadas, saldo em Operações não Consignadas, soma das parcelas (PMT) das Operações Consignadas, soma das parcelas (PMT) das Operações não Consignadas, soma do Valor Bruto das Operações Consignadas e soma do Valor Bruto das Operações não Consignadas. Assim, pode-se seguir com a construção de um modelo mais complexo, utilizando as variáveis relevantes que possuem significância abaixo de 5%. As variáveis não incluídas no modelo são apresentadas na tabela 4:

Tabela 4
Variáveis não presentes na equação

		Escore	df	Sig.	
Etapa 0	Variáveis	ds_estadocivil	6.864	1	0,009
		cdsexo	9.469	1	0,002
		ds_regiao	0,595	1	0,441
		Zscore(nr_idade)	0,863	1	0,353
		Zscore(vl_rendabruta)	5.177	1	0,023
		Zscore(saldo_consignado)	11.856	1	<,001
		Zscore(saldo_nconsignado)	8.883	1	0,003
		Zscore(pmt_consignado)	16.578	1	<,001
		Zscore(pmt_nconsignado)	73.855	1	<,001
		Zscore(vl_emprest_consignado)	8.228	1	0,004
		Zscore(vl_emprest_nconsignado)	84.169	1	<,001
Estatísticas globais		311.417	11	<,001	

No bloco 1, é apresentado o histórico de iteração das variáveis, no qual a estimação total é finalizada na iteração de número 6, porque as estimativas de parâmetro mudam e são alteradas para menos de 0,001. Os testes de Omnibus do Modelo de Coeficientes indicam a qualidade do modelo. Dessa forma, fica evidenciada que a inclusão das variáveis independentes melhora o modelo, se comparado ao modelo sem a inclusão de previsores, pois o grau de significância, nas cinco etapas testadas, ficou abaixo de 0,05, conforme tabela 5:

Tabela 5
Testes de Omnibus do Modelo de Coeficiente

		Qui-quadrado	df	Sig.
Etapa 1	Etapa	126.189	1	<,001
	Bloco	126.189	1	<,001
	Modelo	126.189	1	<,001
Etapa 2	Etapa	391.034	1	<,001
	Bloco	517.223	2	<,001
	Modelo	517.223	2	<,001
Etapa 3	Etapa	19.521	1	<,001
	Bloco	536.744	3	<,001
	Modelo	536.744	3	<,001
Etapa 4	Etapa	7.322	1	0,007
	Bloco	544.066	4	<,001
	Modelo	544.066	4	<,001
Etapa 5	Etapa	5.173	1	0,023
	Bloco	549.239	5	<,001
	Modelo	549.239	5	<,001

Em relação ao resumo do modelo, o melhor Log -2 de verossimilhança se deu na etapa 5, com o valor de 825,961, o qual foi o mais baixo, representando a maior quantidade de informação explicada pelo modelo. Também, foi a etapa 5 que obteve o maior valor de R quadrado de Cox & Snell (0,425) e R quadrado de Nagelkerke (0,567). Então, considerando o R quadrado de Nagelkerke, o modelo explica aproximadamente 57% da variabilidade da variável dependente. Isto é demonstrado na tabela 6:

Tabela 6
Resumo do modelo

Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	1249,011a	0,119	0,159
2	857,977b	0,406	0,542
3	838,456b	0,418	0,557
4	831,134b	0,422	0,563
5	825,961b	0,425	0,567

Nota. a, Estimação finalizada no número de iteração 5 porque as estimativas de parâmetro mudaram foram alteradas para menos de ,001,

b, Estimação finalizada no número de iteração 6 porque as estimativas de parâmetro mudaram foram alteradas para menos de ,001,

O Teste de Hosmer e Lemeshow está demonstrando nas etapas 1, 3, 4, e 5 que o modelo deve ser rejeitado, ou seja, que o modelo não é o mais adequado para esta previsão. Entretanto, a etapa 2 demonstra que o modelo deve ser aceito, pois seu nível de significância

ficou acima de 0,05, aceitando a hipótese nula, de que o modelo é adequado, de acordo com a tabela 7:

Tabela 7
Teste de Hosmer e Lemeshow

Etapa	Qui-quadrado	df	Sig.
1	64.435	8	<,001
2	8.650	8	0,373
3	16.307	8	0,038
4	18.634	8	0,017
5	19.654	8	0,012

No que diz respeito à Tabela de Classificação, o modelo classificou corretamente 81,5% dos casos na etapa 5, sendo a etapa com o melhor resultado. Assim, percebe-se que o modelo com a inclusão das variáveis independentes Sexo, saldo em Operações não Consignadas, soma das parcelas (PMT) das Operações não Consignadas, soma do Valor Bruto das Operações Consignadas e soma do Valor Bruto das Operações não Consignadas classificou melhor do que o modelo sem previsores. Neste caso, o percentual de classificação correta de inadimplência é de 80,2%. A tabela 8 a seguir demonstra a classificação:

Tabela 8
Tabela de Classificação^a

	Observado	Grupo		Previsto	
		0	1	Porcentagem correta	
Etapa 1	grupo	0	315	182	63,4
		1	115	380	76,8
	Porcentagem global				70,1
Etapa 2	grupo	0	392	105	78,9
		1	98	397	80,2
	Porcentagem global				79,5
Etapa 3	grupo	0	410	87	82,5
		1	101	394	79,6
	Porcentagem global				81
Etapa 4	grupo	0	408	89	82,1
		1	97	398	80,4
	Porcentagem global				81,3
Etapa 5	grupo	0	411	86	82,7
		1	98	397	80,2
	Porcentagem global				81,5

Nota. a, O valor de recorte é ,500

Referente às variáveis incluídas no modelo, a estatística de Wald demonstra que as 5 variáveis utilizadas, as quais são Sexo, saldo em Operações não Consignadas, soma das parcelas (PMT) das Operações não Consignadas, soma do Valor Bruto das Operações Consignadas e soma do Valor Bruto das Operações não Consignadas, possuem coeficientes diferentes de zero, tendo peso relevante para o modelo. Quando testada a significância, a 5%, todas as variáveis são significantes.

Pela coluna Exp (B), razão de chances, é possível identificar as chances que cada contrato de determinada variável tem de pertencer à inadimplência. Assim, a variável com maior chance de possuir contratos que levam à inadimplência são as variáveis saldo em Operações não Consignadas, soma das parcelas (PMT) das Operações não Consignadas e Sexo na etapa 5. Inicialmente, alguns coeficientes (B) apresentados na tabela estavam zerados, pelo fato de haver variáveis com escalas muito diferentes. Esse problema foi corrigido com o cálculo dos valores normalizados das variáveis. Assim, tornou-se mais adequada a comparação. Os resultados são apresentados na tabela 9:

Tabela 9
Variáveis na equação

		B	S,E	Wald	df	Sig,	Exp(B)	95% C,I, para EXP(B)	
								Inferior	Superior
Etapa 1 ^a	Zscore(vl_emprest_nconsignado)	-1.109	0,117	89.523	1	<,001	0,33	0,262	0,415
	Constante	-0,091	0,07	1.729	1	0,188	0,913		
Etapa 2 ^b	Zscore(saldo_nconsignado)	3.437	0,253	184.809	1	<,001	31.104	18.949	51.055
	Zscore(vl_emprest_nconsignado)	-4.555	0,301	228.895	1	<,001	0,011	0,006	0,019
Etapa 3 ^c	Constante	-0,086	0,092	0,879	1	0,348	0,918		
	Zscore(saldo_nconsignado)	2.908	0,277	109.935	1	<,001	18.315	10.635	31.540
	Zscore(pmt_nconsignado)	0,811	0,204	15.857	1	<,001	2.251	1.510	3.355
Etapa 4 ^d	Zscore(vl_emprest_nconsignado)	-4.399	0,308	204.330	1	<,001	0,012	0,007	0,022
	Constante	0,038	0,101	0,14	1	0,708	1,039		
	Zscore(saldo_nconsignado)	2.961	0,283	109.831	1	<,001	19.315	11.102	33.603
	Zscore(pmt_nconsignado)	0,833	0,205	16.425	1	<,001	2.300	1.537	3.440
	Zscore(vl_emprest_consignado)	-0,251	0,093	7.297	1	0,007	0,778	0,649	0,934
Etapa 5 ^e	Zscore(vl_emprest_nconsignado)	-4.440	0,312	202.294	1	<,001	0,012	0,006	0,022
	Constante	0,04	0,101	0,153	1	0,696	1,040		
	cdsexo	0,401	0,177	5.141	1	0,023	1.494	1.056	2.113
	Zscore(saldo_nconsignado)	2.981	0,283	110.625	1	<,001	19.705	11.307	34.341
	Zscore(pmt_nconsignado)	0,824	0,207	15.891	1	<,001	2.280	1.520	3.418
Etapa 5 ^e	Zscore(vl_emprest_consignado)	-0,262	0,092	8.057	1	0,005	0,77	0,642	0,922
	Zscore(vl_emprest_nconsignado)	-4.459	0,314	201.452	1	<,001	0,012	0,006	0,021
	Constante	-0,12	0,123	0,942	1	0,332	0,887		

Nota. a, Variável(is) inserida(s) no passo 1: Zscore(vl_emprest_nconsignado),

b, Variável(is) inserida(s) no passo 2: Zscore(saldo_nconsignado),

c, Variável(is) inserida(s) no passo 3: Zscore(pmt_nconsignado),

d, Variável(is) inserida(s) no passo 4: Zscore(vl_emprest_consignado),

e, Variável(is) inserida(s) no passo 5: cdsexo,

Foi realizada uma regressão logística binária para verificar se as variáveis X_1 a X_{11} são previsores da inadimplência. O modelo estimado com data base em 31/12/2020, com o melhor resultado foi [$X^2(1) = 549,239$; $p <,001$, R^2 Nagelkerke = 0,567]. A equação que melhor descreve essa relação é: $P(\text{inadimplência}) = e^{(-0,120 + 0,401X_2 + 2,981X_7 + 0,824X_9 - 0,262X_{10} - 4,459X_{11})} / 1 + e^{(-0,120 + 0,401X_2 + 2,981X_7 + 0,824X_9 - 0,262X_{10} - 4,459X_{11})}$. Assim, a equação do modelo utiliza as variáveis independentes Sexo, saldo em Operações não Consignadas, soma das parcelas (PMT) das Operações não Consignadas, soma do Valor Bruto das Operações Consignadas e soma do Valor Bruto das Operações não Consignadas.

Sobre a regressão com data base em 30/06/2020, o bloco inicial apresenta a relação de variáveis não incluídas no modelo, dentre as 11 variáveis independentes citadas na seção de metodologia. Este resultado demonstra que existem 2 variáveis irrelevantes para o modelo, as quais também são Região e Idade. Sendo assim, pode-se seguir com a estimação de um modelo mais complexo, utilizando as demais variáveis, consideradas relevantes, conforme apresentado na tabela 10:

Tabela 10
Variáveis não presentes na equação

		Escore	df	Sig.	
Etapa 0	Variáveis	ds_estadocivil	6.864	1	0,009
		cdsexo	9.469	1	0,002
		ds_regiao	0,595	1	0,441
		Zscore(nr_idade)	0,863	1	0,353
		Zscore(vl_rendabruta)	5.177	1	0,023
		Zscore(saldo_consignado)	4.422	1	0,035
		Zscore(saldo_nconsignado)	41.703	1	<,001
		Zscore(pmt_consignado)	5.415	1	0,02
		Zscore(pmt_nconsignado)	95.040	1	<,001
		Zscore(vl_emprest_consignado)	4.177	1	0,041
		Zscore(vl_emprest_nconsignado)	23.770	1	<,001
Estatísticas globais		313.369	11	<,001	

O histórico de iteração das variáveis, apresentado no bloco 1, indica que a estimação total é finalizada com 6 iterações, porque as estimativas de parâmetro mudam e são alteradas para menos de 0,001. Pelos testes de Omnibus do Modelo de Coeficientes, fica evidenciada que a inclusão das variáveis independentes melhora o modelo, se comparado ao modelo sem a inclusão de previsores, pois o grau de significância, nas sete etapas testadas, ficou abaixo de 0,05, conforme tabela 11:

Tabela 11
Testes de Omnibus do Modelo de Coeficientes

		Qui-quadrado	df	Sig.
Etapa 1	Etapa	122.401	1	<,001
	Bloco	122.401	1	<,001
	Modelo	122.401	1	<,001
Etapa 2	Etapa	186.705	1	<,001
	Bloco	309.106	2	<,001
	Modelo	309.106	2	<,001
Etapa 3	Etapa	168.033	1	<,001
	Bloco	477.139	3	<,001
	Modelo	477.139	3	<,001
Etapa 4	Etapa	10.142	1	0,001
	Bloco	487.281	4	<,001
	Modelo	487.281	4	<,001
Etapa 5	Etapa	4.924	1	0,026
	Bloco	492.205	5	<,001
	Modelo	492.205	5	<,001
Etapa 6	Etapa	6.104	1	0,013
	Bloco	498.309	6	<,001
	Modelo	498.309	6	<,001
Etapa 7	Etapa	7.496	1	0,006
	Bloco	505.805	7	<,001
	Modelo	505.805	7	<,001

Referente ao resumo do modelo, o melhor Log -2 de verossimilhança se deu na etapa 7, com o valor de 869,395, o qual foi o mais baixo, representando a maior quantidade de informação explicada pelo modelo. Também, foi a etapa 7 que obteve o maior valor de R quadrado de Cox & Snell (0,399) e R quadrado de Nagelkerke (0,533). Então, considerando o R quadrado de Nagelkerke, o modelo explica aproximadamente 53% da variabilidade da variável dependente. Os resultados estão demonstrados na tabela 12:

Tabela 12
Resumo do modelo

Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	1252,799a	0,116	0,155
2	1066,094b	0,268	0,357
3	898,061b	0,382	0,509
4	887,919b	0,388	0,517
5	882,995b	0,391	0,522
6	876,891b	0,395	0,527
7	869,395b	0,399	0,533

Nota. a, Estimação finalizada no número de iteração 5 porque as estimativas de parâmetro mudaram foram alteradas para menos de ,001,

b, Estimação finalizada no número de iteração 6 porque as estimativas de parâmetro mudaram foram alteradas para menos de ,001,

O Teste de Hosmer e Lemeshow demonstra, em todas as etapas, que o modelo deve ser rejeitado, ou seja, que o modelo não é o mais adequado para esta previsão. Todos os níveis de significância ficaram abaixo de 0,05, rejeitando a hipótese nula de que o modelo é adequado, de acordo com os resultados da tabela 13. Entretanto, como os demais testes apresentaram um resultado favorável ao modelo, foi dado seguimento à sua análise.

Tabela 13
Teste de Hosmer e Lemeshow

Etapa	Qui-quadrado	df	Sig.
1	79.128	8	<,001
2	37.364	8	<,001
3	29.924	8	<,001
4	25.872	8	0,001
5	21.396	8	0,006
6	28.884	8	<,001
7	22.328	8	0,004

Em relação à Tabela de Classificação, o modelo classificou corretamente 80,6% dos casos na etapa 7, sendo a etapa com o melhor resultado. Neste caso, o percentual de classificação correta de inadimplência é de 78,6%. A tabela 14 apresenta os resultados das 7 etapas:

Tabela 14
Tabela de Classificação^a

	Observado	Grupo		Previsto	
		0	1	Porcentagem correta	
Etapa 1	grupo	0	403	94	81,1
		1	246	249	50,3
	Porcentagem global				65,7
Etapa 2	grupo	0	418	79	84,1
		1	170	325	65,7
	Porcentagem global				74,9
Etapa 3	grupo	0	410	87	82,5
		1	110	385	77,8
	Porcentagem global				80,1
Etapa 4	grupo	0	411	86	82,7
		1	111	384	77,6
	Porcentagem global				80,1
Etapa 5	grupo	0	411	86	82,7
		1	109	386	78
	Porcentagem global				80,3
Etapa 6	grupo	0	411	86	82,7
		1	106	389	78,6
	Porcentagem global				80,6
Etapa 7	grupo	0	411	86	82,7
		1	106	389	78,6
	Porcentagem global				80,6

Nota. a, O valor de recorte é ,500

Quanto às variáveis incluídas no modelo, a estatística de Wald demonstra que as sete variáveis utilizadas, conforme citadas anteriormente, possuem coeficientes diferentes de zero, tendo peso relevante para o modelo. Quando testada a significância a 5%, todas as variáveis são significantes.

Pela coluna Exp (B), ou seja, *odds ratio*, é possível identificar as chances que cada contrato de determinada variável tem de pertencer à inadimplência. Assim, a variável com maior chance de possuir contratos que levam à inadimplência são as variáveis saldo em Operações não Consignadas, soma das parcelas (PMT) em Operações não Consignadas e Sexo, na etapa 7. Para tornar mais adequada a comparação, as variáveis tiveram os seus valores normalizados. Resultados apresentados na tabela 15:

Tabela 15
Variáveis na equação

		B	S,E,	Wald	df	Sig.	Exp(B)	95% C,I, para EXP(B)	
								Inferior	Superior
Etapa 1 ^a	Zscore(pmt_nconsignado)	1.020	0,118	75.024	1	<,001	2.773	2.202	3.493
	Constante	0,093	0,071	1.722	1	0,189	1.097		
Etapa 2 ^b	Zscore(pmt_nconsignado)	2.222	0,19	136.836	1	<,001	9.226	6.358	13.388
	Zscore(vl_emprest_nconsignado)	-1.457	0,129	126.925	1	<,001	0,233	0,181	0,3
Etapa 3 ^c	Constante	0,278	0,087	10.178	1	0,001	1.320		
	Zscore(saldo_nconsignado)	3.818	0,365	109.553	1	<,001	45.536	22.275	93.087
Etapa 4 ^d	Zscore(pmt_nconsignado)	0,413	0,238	3.022	1	0,082	1.511	0,949	2.408
	Zscore(vl_emprest_nconsignado)	-3.792	0,279	184.694	1	<,001	0,023	0,013	0,039
Etapa 5 ^e	Constante	0,298	0,096	9.687	1	0,002	1.347		
	Zscore(vl_rendabruta)	-0,536	0,133	16.198	1	<,001	0,585	0,45	0,759
Etapa 6 ^f	Zscore(saldo_nconsignado)	4.099	0,379	117.144	1	<,001	60.307	28.705	126.697
	Zscore(pmt_nconsignado)	0,465	0,239	3.802	1	0,051	1.592	0,998	2.542
Etapa 7 ^g	Zscore(vl_emprest_nconsignado)	-3.831	0,279	187.952	1	<,001	0,022	0,013	0,038
	Constante	0,308	0,097	10.130	1	0,001	1.361		
Etapa 8 ^h	Zscore(vl_rendabruta)	-0,519	0,138	14.119	1	<,001	0,595	0,454	0,78
	Zscore(saldo_nconsignado)	4.073	0,378	115.865	1	<,001	58.756	27.986	123.358
Etapa 9 ⁱ	Zscore(pmt_consignado)	-0,209	0,093	5.062	1	0,024	0,812	0,677	0,973
	Zscore(pmt_nconsignado)	0,493	0,238	4.272	1	0,039	1.637	1.026	2.612
Etapa 10 ^j	Zscore(vl_emprest_nconsignado)	-3.847	0,281	187.297	1	<,001	0,021	0,012	0,037
	Constante	0,296	0,096	9.429	1	0,002	1.344		
Etapa 11 ^k	cdsexo	0,421	0,171	6.065	1	0,014	1.523	1.090	2.130
	Zscore(vl_rendabruta)	-0,524	0,133	15.522	1	<,001	0,592	0,456	0,768
Etapa 12 ^l	Zscore(saldo_nconsignado)	4.094	0,381	115.552	1	<,001	60.003	28.442	126.587
	Zscore(pmt_consignado)	-0,23	0,094	6.025	1	0,014	0,794	0,661	0,955
Etapa 13 ^m	Zscore(pmt_nconsignado)	0,493	0,238	4.288	1	0,038	1.638	1.027	2.613
	Zscore(vl_emprest_nconsignado)	-3.873	0,283	186.790	1	<,001	0,021	0,012	0,036
Etapa 14 ⁿ	Constante	0,124	0,119	1.090	1	0,296	1.132		
	cdsexo	0,502	0,175	8.280	1	0,004	1.653	1.174	2.327
Etapa 15 ^o	Zscore(nr_idade)	-0,234	0,086	7.404	1	0,007	0,791	0,668	0,937
	Zscore(vl_rendabruta)	-0,532	0,126	17.869	1	<,001	0,587	0,459	0,752
Etapa 16 ^p	Zscore(saldo_nconsignado)	4.141	0,385	115.670	1	<,001	62.854	29.553	133.676

Continua

	B	S,E	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
							Inferior	Superior
Zscore(pmt_consignado)	-0,229	0,093	6.093	1	0,014	0,795	0,663	0,954
Zscore(pmt_nconsignado)	0,524	0,241	4.737	1	0,03	1.688	1.053	2.706
Zscore(vl_emprest_nconsignado)	-3.940	0,289	185.689	1	<,001	0,019	0,011	0,034
Constante	0,088	0,12	0,534	1	0,465	1.092		

Nota. a, Variável(is) inserida(s) no passo 1: Zscore(pmt_nconsignado),
 b, Variável(is) inserida(s) no passo 2: Zscore(vl_emprest_nconsignado),
 c, Variável(is) inserida(s) no passo 3: Zscore(saldo_nconsignado),
 d, Variável(is) inserida(s) no passo 4: Zscore(vl_rendabruta),
 e, Variável(is) inserida(s) no passo 5: Zscore(pmt_consignado),
 f, Variável(is) inserida(s) no passo 6: cd_sex0,
 g, Variável(is) inserida(s) no passo 7: Zscore(nr_idade),

Com a realização da regressão logística binária para verificar se as variáveis X_1 a X_{11} são previsores da inadimplência, o modelo estimado com o melhor resultado, considerando a data base de 30/06/2020, foi [$X^2(1) = 505,805$; $p <,001$, R^2 Nagelkerke = 0,533]. A equação que melhor descreve essa relação é: $P(\text{inadimplência}) = e^{(0,088 + 0,502X_2 - 0,234X_4 - 0,532X_5 + 4,141X_7 - 0,229X_8 + 0,524X_9 - 3,940X_{11})} / 1 + e^{(0,088 + 0,502X_2 - 0,234X_4 - 0,532X_5 + 4,141X_7 - 0,229X_8 + 0,524X_9 - 3,940X_{11})}$. Nesta equação, foram utilizadas as variáveis Sexo, Idade, Renda Bruta, saldo em Operações não Consignadas, soma das parcelas (PMT) em Operações Consignadas e não Consignadas e soma do Valor Bruto das Operações não Consignadas.

Referente à regressão com data base em 31/12/2019, no bloco inicial, é apresentada a relação de variáveis não incluídas no modelo, considerando as 11 variáveis independentes citadas na seção de metodologia. Este resultado demonstra que existem seis variáveis irrelevantes para o modelo, as quais são Região, Idade, saldo em Operações Consignadas, soma das parcelas (PMT) das Operações Consignadas, soma do Valor Bruto das Operações Consignadas e soma do Valor Bruto das Operações não Consignadas. Entretanto, o modelo pode seguir com a inclusão das cinco variáveis restantes: Estado Civil, Sexo, Renda Bruta, saldo em Operações não Consignadas e soma das parcelas (PMT) das Operações não Consignadas. Pode-se seguir com a construção de um modelo mais complexo, utilizando as cinco variáveis relevantes, conforme demonstrado na tabela 16:

Tabela 16
Variáveis não presentes na equação

		Escore	df	Sig.	
Etapa 0	Variáveis	ds_estadocivil	6.864	1	0,009
		cd_sex0	9.469	1	0,002
		ds_regiao	0,595	1	0,441
		Zscore(nr_idade)	0,863	1	0,353
		Zscore(vl_rendabruta)	5.177	1	0,023

	Escore	df	Sig.
Zscore(saldo_consignado)	0,024	1	0,877
Zscore(saldo_nconsignado)	49.170	1	<,001
Zscore(pmt_consignado)	0,246	1	0,62
Zscore(pmt_nconsignado)	91.305	1	<,001
Zscore(vl_emprest_consignado)	0,4	1	0,527
Zscore(vl_emprest_nconsignado)	0,085	1	0,771
Estatísticas globais	187.926	11	<,001

A apresentação do histórico de iteração das variáveis, no bloco 1, indica que a estimação total é finalizada na quinta iteração, porque as estimativas de parâmetro mudam e são alteradas para menos de 0,001. Pelos testes de Omnibus do Modelo de Coeficientes, depreende-se que a inclusão das variáveis independentes melhora o modelo, se comparado ao modelo sem a inclusão de previsores, pois o grau de significância, nas sete etapas testadas, ficou abaixo de 0,05, conforme resultados da tabela 17:

Tabela 17
Testes de Omnibus do Modelo de Coeficientes

		Qui-quadrado	df	Sig.
Etapa 1	Etapa	117.269	1	<,001
	Bloco	117.269	1	<,001
	Modelo	117.269	1	<,001
Etapa 2	Etapa	71.233	1	<,001
	Bloco	188.502	2	<,001
	Modelo	188.502	2	<,001
Etapa 3	Etapa	46.748	1	<,001
	Bloco	235.250	3	<,001
	Modelo	235.250	3	<,001
Etapa 4	Etapa	9.308	1	0,002
	Bloco	244.558	4	<,001
	Modelo	244.558	4	<,001
Etapa 5	Etapa	5.861	1	0,015
	Bloco	250.419	5	<,001
	Modelo	250.419	5	<,001
Etapa 6	Etapa	4.663	1	0,031
	Bloco	255.082	6	<,001
	Modelo	255.082	6	<,001
Etapa 7	Etapa	3.975	1	0,046
	Bloco	259.056	7	<,001
	Modelo	259.056	7	<,001

Quanto ao resumo do modelo, o melhor Log -2 de verossimilhança se deu na etapa 7, com o valor de 1116,144, o qual foi o mais baixo, representando a maior quantidade de

informação explicada pelo modelo. Adicionalmente, foi a etapa 7 que obteve o maior valor de R quadrado de Cox & Snell (0,230) e R quadrado de Nagelkerke (0,306). Então, considerando o R quadrado de Nagelkerke, o modelo explica aproximadamente 30% da variabilidade da variável dependente. O resumo do modelo é apresentado na tabela 18:

Tabela 18
Resumo do modelo

Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	1257,931a	0,111	0,149
2	1186,698a	0,173	0,231
3	1139,950a	0,211	0,282
4	1130,642a	0,218	0,291
5	1124,781a	0,223	0,297
6	1120,118a	0,227	0,302
7	1116,144a	0,23	0,306

Nota. a, Estimação finalizada no número de iteração 5 porque as estimativas de parâmetro mudaram foram alteradas para menos de ,001,

Segundo o Teste de Hosmer e Lemeshow, as etapas 1, 2, 3, 4, e 5 demonstram que o modelo deve ser rejeitado, ou seja, que o modelo não é o mais adequado para esta previsão. Entretanto, as etapas 6 e 7 indicam que o modelo deve ser aceito, pois seu nível de significância ficou acima de 0,05, aceitando a hipótese nula, de que o modelo é adequado, conforme a tabela 19 a seguir:

Tabela 19
Teste de Hosmer e Lemeshow

Etapa	Qui-quadrado	df	Sig,
1	28.362	7	<,001
2	31.909	8	<,001
3	35.226	8	<,001
4	24.623	8	0,002
5	23.612	8	0,003
6	12.090	8	0,147
7	15.244	8	0,055

Pela Tabela de Classificação, depreende-se que o modelo classificou corretamente 71,3% dos casos na etapa 7, etapa com o melhor resultado. Assim, o modelo com a inclusão das variáveis independentes Estado Civil, Sexo, Idade, Renda Bruta do Cliente, saldo em Operações não Consignadas, soma das parcelas (PMT) das Operações não Consignadas e soma do Valor Bruto das Operações não Consignadas classificou melhor do que o modelo

sem previsores. Neste caso, o percentual de classificação correta de inadimplência é de 61,8%. A referida tabela de classificação é apresentada na tabela 20:

Tabela 20
Tabela de Classificação^a

	Observado	Grupo		Previsto	
		0	1	Porcentagem correta	
Etapa 1	grupo	0	406	91	81,7
		1	251	244	49,3
	Porcentagem global				65,5
Etapa 2	grupo	0	411	86	82,7
		1	234	261	52,7
	Porcentagem global				67,7
Etapa 3	grupo	0	416	81	83,7
		1	190	305	61,6
	Porcentagem global				72,7
Etapa 4	grupo	0	412	85	82,9
		1	185	310	62,6
	Porcentagem global				72,8
Etapa 5	grupo	0	414	83	83,3
		1	186	309	62,4
	Porcentagem global				72,9
Etapa 6	grupo	0	401	96	80,7
		1	192	303	61,2
	Porcentagem global				71
Etapa 7	grupo	0	401	96	80,7
		1	189	306	61,8
	Porcentagem global				71,3

Nota. a, O valor de recorte é ,500

Consoante às variáveis incluídas no modelo, a estatística de Wald demonstra que as sete variáveis utilizadas, conforme citadas anteriormente, possuem coeficientes diferentes de zero, tendo peso relevante para o modelo. Quando testada a significância, a 5%, todas as variáveis são significantes.

A coluna do *odds ratio* ($\text{Exp}(B)$), identifica as chances que cada contrato de determinada variável tem de pertencer à inadimplência. Assim, as variáveis com maior chance de possuir contratos que levam à inadimplência são as variáveis saldo em Operações não Consignadas, soma das parcelas (PMT) das Operações não Consignadas e Sexo na etapa 7. Para tornar mais adequada a comparação, as variáveis tiveram os seus valores normalizados. Os resultados são apresentados na tabela 21:

Tabela 21
Variáveis na equação

		B	S,E	Wald	df	Sig.	Exp(B)	95% C,I, para EXP(B)	
								Inferior	Superior
Etapa 1 ^a	Zscore(pmt_nconsignado)	0,992	0,116	72.488	1	<,001	2.696	2.146	3.387
	Constante	0,089	0,071	1.576	1	0,209	1.093		
Etapa 2 ^b	Zscore(pmt_nconsignado)	1.659	0,16	107.616	1	<,001	5.256	3.841	7.191
	Zscore(vl_emprest_nconsignado)	-0,852	0,107	63.472	1	<,001	0,427	0,346	0,526
	Constante	0,164	0,076	4.616	1	0,032	1.178		
	Zscore(saldo_nconsignado)	2.042	0,327	39.007	1	<,001	7.708	4.061	14.632
Etapa 3 ^c	Zscore(pmt_nconsignado)	0,581	0,216	7.222	1	0,007	1.788	1.170	2.733
	Zscore(vl_emprest_nconsignado)	-1.876	0,197	91.175	1	<,001	0,153	0,104	0,225
	Constante	0,198	0,078	6.404	1	0,011	1.219		
	Zscore(vl_rendabruta)	-0,347	0,099	12.174	1	<,001	0,707	0,582	0,859
	Zscore(saldo_nconsignado)	2.231	0,314	50.562	1	<,001	9.312	5.034	17.224
	Zscore(pmt_nconsignado)	0,629	0,203	9.640	1	0,002	1.877	1.261	2.792
Etapa 4 ^d	Zscore(vl_emprest_nconsignado)	-1.881	0,192	95.935	1	<,001	0,152	0,105	0,222
	Constante	0,208	0,079	6.921	1	0,009	1.232		
	Zscore(nr_idade)	-0,174	0,072	5.825	1	0,016	0,84	0,73	0,968
	Zscore(vl_rendabruta)	-0,362	0,1	13.215	1	<,001	0,696	0,573	0,846
Etapa 5 ^e	Zscore(saldo_nconsignado)	2.293	0,317	52.321	1	<,001	9.900	5.319	18.426
	Zscore(pmt_nconsignado)	0,636	0,203	9.822	1	0,002	1.889	1.269	2.811
	Zscore(vl_emprest_nconsignado)	-1.910	0,194	97.061	1	<,001	0,148	0,101	0,217
	Constante	0,214	0,08	7.267	1	0,007	1.239		
	cdsexo	0,322	0,149	4.652	1	0,031	1.380	1.030	1.849
	Zscore(nr_idade)	-0,202	0,073	7.543	1	0,006	0,817	0,708	0,944
Etapa 6 ^f	Zscore(vl_rendabruta)	-0,361	0,1	13.139	1	<,001	0,697	0,573	0,847
	Zscore(saldo_nconsignado)	2.295	0,318	52.109	1	<,001	9.922	5.321	18.501
	Zscore(pmt_nconsignado)	0,629	0,203	9.595	1	0,002	1.877	1.260	2.795
	Zscore(vl_emprest_nconsignado)	-1.905	0,195	95.588	1	<,001	0,149	0,102	0,218
	Constante	0,082	0,1	0,675	1	0,411	1.086		
	ds_estadocivil	0,133	0,067	3.942	1	0,047	1.142	1.002	1.301
	cdsexo	0,365	0,151	5.826	1	0,016	1.441	1.071	1.938
	Zscore(nr_idade)	-0,22	0,074	8.786	1	0,003	0,802	0,693	0,928
Etapa 7 ^g	Zscore(vl_rendabruta)	-0,368	0,1	13.597	1	<,001	0,692	0,569	0,842
	Zscore(saldo_nconsignado)	2.298	0,319	51.973	1	<,001	9.957	5.331	18.599
	Zscore(pmt_nconsignado)	0,619	0,203	9.302	1	0,002	1.857	1.247	2.763
	Zscore(vl_emprest_nconsignado)	-1.878	0,196	91.981	1	<,001	0,153	0,104	0,224
	Constante	-0,126	0,145	0,754	1	0,385	0,882		

Nota. a, Variável(is) inserida(s) no passo 1: Zscore(pmt_nconsignado),

b, Variável(is) inserida(s) no passo 2: Zscore(vl_emprest_nconsignado),

c, Variável(is) inserida(s) no passo 3: Zscore(saldo_nconsignado),

d, Variável(is) inserida(s) no passo 4: Zscore(vl_rendabruta),

e, Variável(is) inserida(s) no passo 5: Zscore(nr_idade),

f, Variável(is) inserida(s) no passo 6: cdsexo,

g, Variável(is) inserida(s) no passo 7: ds_estadocivil,

Com a realização da regressão logística binária, com data base em 31/12/2019, para verificar se as variáveis X_1 a X_{11} são previsores da inadimplência, o modelo estimado com o

melhor resultado foi [$X^2(1) = 259,056$; $p <,001$, R^2 Nagelkerke = 0,306]. A equação que melhor descreve essa relação é: $P(\text{inadimplência}) = e^{(-0,126 + 0,133X_1 + 0,365X_2 - 0,220X_4 - 0,368X_5 + 2,298X_7 + 0,619X_9 - 1,878X_{11})} / 1 + e^{(-0,126 + 0,133X_1 + 0,365X_2 - 0,220X_4 - 0,368X_5 + 2,298X_7 + 0,619X_9 - 1,878X_{11})}$. As variáveis que compõem a equação são Estado Civil, Sexo, Idade, Renda Bruta, saldo em Operações não Consignadas, soma das parcelas (PMT) das Operações não Consignadas e soma do Valor Bruto das Operações não Consignadas.

E a regressão com data base de 30/06/2019 apresenta, no bloco inicial, a relação de variáveis não incluídas no modelo. Este resultado demonstra que existem cinco variáveis irrelevantes para o modelo, as quais são Região, Idade, saldo em Operações Consignadas, soma das parcelas (PMT) das Operações Consignadas e soma do Valor Bruto das Operações Consignadas. Entretanto, o modelo pode seguir com a inclusão das seis variáveis restantes para a estimação de um modelo mais complexo, ou seja, com as variáveis relevantes que possuem significância abaixo de 5%, de acordo com os resultados da tabela 22:

Tabela 22
Variáveis não presentes na equação

		Escore	df	Sig.	
Etapa 0	Variáveis	ds_estadocivil	6.864	1	0,009
		cdsexo	9.469	1	0,002
		ds_regiao	0,595	1	0,441
		Zscore(nr_idade)	0,863	1	0,353
		Zscore(vl_rendabruta)	5.177	1	0,023
		Zscore(saldo_consignado)	0,159	1	0,69
		Zscore(saldo_nconsignado)	34.081	1	<,001
		Zscore(pmt_consignado)	0,113	1	0,737
		Zscore(pmt_nconsignado)	33.836	1	<,001
		Zscore(vl_emprest_consignado)	0,194	1	0,66
		Zscore(vl_emprest_nconsignado)	11.345	1	<,001
	Estatísticas globais		75.949	11	<,001

No bloco 1, é apresentado o histórico de iteração das variáveis, no qual a estimação total é finalizada na iteração de número 4, porque as estimativas de parâmetro mudam e são alteradas para menos de 0,001. Os testes de Omnibus do Modelo de Coeficientes indicam a qualidade do modelo. Dessa forma, se torna claro que a inclusão das variáveis independentes melhora o modelo, se comparado ao modelo sem a inclusão de previsores, pois o grau de significância, nas cinco etapas testadas, ficou abaixo de 0,05, segundo os resultados apresentados na tabela 23:

Tabela 23
Testes de Omnibus do Modelo de Coeficientes

		Qui-quadrado	df	Sig,
Etapa 1	Etapa	38.213	1	<,001
	Bloco	38.213	1	<,001
	Modelo	38.213	1	<,001
Etapa 2	Etapa	20.166	1	<,001
	Bloco	58.379	2	<,001
	Modelo	58.379	2	<,001
Etapa 3	Etapa	7.344	1	0,007
	Bloco	65.722	3	<,001
	Modelo	65.722	3	<,001
Etapa 4	Etapa	5.310	1	0,021
	Bloco	71.032	4	<,001
	Modelo	71.032	4	<,001
Etapa 5	Etapa	6.053	1	0,014
	Bloco	77.085	5	<,001
	Modelo	77.085	5	<,001

Em relação ao resumo do modelo, o melhor Log -2 de verossimilhança se deu na etapa 5, com o valor de 1298,115, o qual foi o mais baixo, representando a maior quantidade de informação explicada pelo modelo. Também, foi a etapa 5 que obteve o maior valor de R quadrado de Cox & Snell (0,075) e R quadrado de Nagelkerke (0,100). Então, considerando o R quadrado de Nagelkerke, o modelo explica aproximadamente 10% da variabilidade da variável dependente. O referido resumo do modelo é apresentado na tabela 24:

Tabela 24
Resumo do modelo

Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	1336,987a	0,038	0,05
2	1316,821a	0,057	0,076
3	1309,477a	0,064	0,085
4	1304,168a	0,069	0,092
5	1298,115a	0,075	0,1

Nota. a, Estimação finalizada no número de iteração 4 porque as estimativas de parâmetro mudaram foram alteradas para menos de ,001,

Quanto ao Teste de Hosmer e Lemeshow, está demonstrando nas etapas 1 e 2 que o modelo deve ser rejeitado, ou seja, que o modelo não é o mais adequado para esta previsão. Entretanto, as etapas 3, 4 e 5 demonstram que o modelo deve ser aceito, pois seu nível de significância ficou acima de 0,05, aceitando a hipótese nula, de que o modelo é adequado. Vide tabela 25:

Tabela 25
Teste de Hosmer e Lemeshow

Etapa	Qui-quadrado	df	Sig.
1	20.664	8	0,008
2	15.588	8	0,049
3	12.243	8	0,141
4	6.223	8	0,622
5	2.123	8	0,977

Referente à Tabela de Classificação, o modelo classificou corretamente 61,4% dos casos na etapa 5, sendo a etapa com o melhor resultado. Assim, percebe-se que o modelo com a inclusão das variáveis independentes Estado Civil, Sexo, Idade, saldo em Operações não Consignadas, e soma do Valor Bruto das Operações não Consignadas classificou melhor do que o modelo sem previsores. Neste caso, o percentual de classificação correta de inadimplência é de 52,9%, conforme tabela 26:

Tabela 26
Tabela de Classificação^a

	Observado	Grupo		Previsto	
		0	1	Porcentagem correta	
Etapa 1	grupo	0	381	116	76,7
		1	303	192	38,8
	Porcentagem global				57,8
Etapa 2	grupo	0	374	123	75,3
		1	249	246	49,7
	Porcentagem global				62,5
Etapa 3	grupo	0	349	148	70,2
		1	240	255	51,5
	Porcentagem global				60,9
Etapa 4	grupo	0	343	154	69
		1	234	261	52,7
	Porcentagem global				60,9
Etapa 5	grupo	0	347	150	69,8
		1	233	262	52,9
	Porcentagem global				61,4

Nota. a, O valor de recorte é ,500

A estatística de Wald demonstra que, em relação às variáveis incluídas no modelo, as 5 variáveis utilizadas, as quais são Estado Civil, Sexo, Idade, saldo em Operações não Consignadas, e soma do Valor Bruto das Operações não Consignadas, possuem coeficientes diferentes de zero, tendo peso relevante para o modelo. Quando testada a significância, a 5%, todas as variáveis são significantes.

Pela razão de chances, apresentada na coluna Exp (B), é possível identificar as chances que cada contrato de determinada variável tem de pertencer à inadimplência. Assim, as variáveis com maior chance de possuir contratos que levam à inadimplência são Saldo em Operações não Consignadas, Sexo e Estado Civil do Cliente na etapa 5. Para tornar mais adequada a comparação, as variáveis tiveram os seus valores normalizados. Os resultados são apresentados na tabela 27:

Tabela 27
Variáveis na equação

		B	S,E	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
								Inferior	Superior
Etapa 1 ^a	Zscore(saldo_nconsignado)	0,459	0,083	30.888	1	<,001	1.583	1.346	1.861
	Constante	0,01	0,065	0,022	1	0,881	1.010		
Etapa 2 ^b	Zscore(saldo_nconsignado)	1.151	0,189	36.919	1	<,001	3.160	2.180	4.580
	Zscore(vl_emprest_nconsignado)	-0,763	0,18	17.893	1	<,001	0,466	0,327	0,664
Etapa 3 ^c	cdsexo	0,36	0,133	7.322	1	0,007	1.433	1.104	1.860
	Zscore(saldo_nconsignado)	1.129	0,19	35.359	1	<,001	3.091	2.131	4.484
Etapa 4 ^d	Zscore(vl_emprest_nconsignado)	-0,748	0,181	17.084	1	<,001	0,473	0,332	0,675
	Constante	-0,135	0,086	2.459	1	0,117	0,873		
Etapa 5 ^e	ds_estadocivil	0,135	0,059	5.233	1	0,022	1.145	1.020	1.286
	cdsexo	0,398	0,134	8.784	1	0,003	1.489	1.144	1.938
Etapa 5 ^e	Zscore(saldo_nconsignado)	1.060	0,19	31.164	1	<,001	2.885	1.989	4.185
	Zscore(vl_emprest_nconsignado)	-0,669	0,182	13.517	1	<,001	0,512	0,359	0,732
Etapa 5 ^e	Constante	-0,35	0,127	7.540	1	0,006	0,705		
	ds_estadocivil	0,153	0,06	6.486	1	0,011	1.165	1.036	1.310
Etapa 5 ^e	cdsexo	0,458	0,137	11.127	1	<,001	1.581	1.208	2.069
	Zscore(nr_idade)	-0,166	0,068	6.006	1	0,014	0,847	0,742	0,967
Etapa 5 ^e	Zscore(saldo_nconsignado)	1.085	0,191	32.172	1	<,001	2.960	2.035	4.307
	Zscore(vl_emprest_nconsignado)	-0,678	0,183	13.707	1	<,001	0,508	0,355	0,727
Etapa 5 ^e	Constante	-0,398	0,129	9.443	1	0,002	0,672		

Nota. a, Variável(is) inserida(s) no passo 1: Zscore(saldo_nconsignado),

b, Variável(is) inserida(s) no passo 2: Zscore(vl_emprest_nconsignado),

c, Variável(is) inserida(s) no passo 3: cdsexo,

d, Variável(is) inserida(s) no passo 4: ds_estadocivil,

e, Variável(is) inserida(s) no passo 5: Zscore(nr_idade),

Portanto, foi realizada uma regressão logística binária para verificar se as variáveis X_1 a X_{11} são previsores da inadimplência. O modelo estimado com data base em 30/06/2019, com o melhor resultado foi [$X^2(1) = 77,085$; $p <,001$, R^2 Nagelkerke = 0,100]. A equação que melhor descreve essa relação é: $P(\text{inadimplência}) = e^{(-0,398 + 0,153X_1 + 0,458X_2 - 0,166X_4 + 1,085X_7 - 0,678X_{11})} / 1 + e^{(-0,398 + 0,153X_1 + 0,458X_2 - 0,166X_4 + 1,085X_7 - 0,678X_{11})}$. Assim, a equação do modelo utiliza

as variáveis independentes Estado Civil, Sexo, Idade, saldo em Operações não Consignadas, e soma do Valor Bruto das Operações não Consignadas.

4.3.2 Discussão

Ao compararmos os resultados das quatro regressões logísticas binárias executadas, cada uma com uma data base distinta, podemos identificar diferenças significativas entre elas, o que nos leva a perceber a importância de um acompanhamento contínuo das variáveis relevantes para a gestão de risco de crédito. Em todas as regressões foram utilizadas as mesmas variáveis como possíveis previsoras da inadimplência do banco digital X, porém com dados com pontos de corte temporal diferentes. Isto significa que o modelo de *credit scoring* deve ser aplicado constantemente para identificar tendências e alterações no comportamento dos previsores, podendo indicar mudanças sociais e econômicas que influenciam os tomadores de crédito pessoal.

Na Tabela 28, a seguir, é apresentado um resumo dos principais resultados:

Tabela 28
Resumo dos Resultados Principais

Data Base	Testes de Omnibus	Verossimilhança de log -2	R quadrado Nagelkerke	Classificação
31/12/2020	549.239	825,961	0,567	81,5
30/06/2020	505.805	869,395	0,533	80,6
31/12/2019	259.056	1.116,144	0,306	71,3
30/06/2019	77.085	1.298,115	0,1	61,4

O modelo com a data mais atual foi o que trouxe o melhor resultado, tanto em termos de R^2 , quanto em capacidade de classificação correta dos contratos de concessão de crédito de pessoas físicas em adimplentes ou inadimplentes. Esse modelo de *credit scoring*, com data base de 31/12/2020, classificou corretamente 81,5% dos casos, o que representa um percentual de acerto mais de 30% maior do que o modelo com data base em junho de 2019. As variáveis com maior poder preditivo neste modelo foram Sexo, saldo em Operações não Consignadas, soma das parcelas (PMT) das Operações não Consignadas, soma do Valor Bruto das Operações Consignadas e soma do Valor Bruto das Operações não Consignadas, as quais conseguiram explicar 57% da variação da inadimplência, conforme o R^2 de Nagelkerke.

Ainda, ao analisar a equação deste modelo, depreende-se que a variável Saldo em Operações Não Consignadas foi a que apresentou maior coeficiente, sendo a variável com

maior impacto na inadimplência, ou seja, quanto mais operações não consignadas, mais probabilidade de ter inadimplência. Em seguida, as variáveis soma das parcelas (PMT) das Operações não Consignadas e Sexo tiveram o segundo e terceiro maior coeficiente, demonstrando maior impacto na inadimplência. As variáveis soma do Valor Bruto das Operações Consignadas e não Consignadas, entretanto, por terem coeficientes negativos, seriam as menos propensas a explicar a inadimplência.

Na análise deste primeiro modelo, é interessante mencionar que as variáveis Idade, Estado Civil, Renda Bruta do tomador e valores na modalidade de consignado não foram relevantes para a mensuração da inadimplência. Isto gera um contraponto ao senso comum de que pessoas jovens, solteiras e de baixa renda teriam uma tendência maior a inadimplência, por motivos de terem piores condições financeiras ou não terem uma disciplina financeira, por muitas vezes ainda não terem constituído uma família, não tendo obrigações com o sustento de filhos e custeio e manutenção de moradia.

O modelo com data de 30/06/2019, conseguiu explicar apenas 10% da variabilidade da inadimplência, segundo o mesmo parâmetro (R^2 de Nagelkerke). Mesmo assim, o modelo atingiu uma assertividade de classificação de 61,4% dos casos, o que é maior do que o modelo inicial, sem variáveis independentes, que classificou de forma correta apenas 50,1% dos casos analisados. O modelo em questão, diferentemente do modelo anterior, utilizou as variáveis Estado Civil, Sexo, Idade, saldo em Operações não Consignadas, e soma do Valor Bruto das Operações não Consignadas.

Ao comparar os resultados de dezembro de 2020 com os resultados de dezembro de 2019, depreende-se que o modelo com data base em 31/12/2020 obteve um percentual de acerto de classificação 14% acima. O modelo de 31/12/2019 utilizou duas variáveis a mais que o modelo comparado, incluindo as variáveis Estado Civil, Idade e Renda Bruta e removendo a variável soma do Valor Bruto das Operações Consignadas. Com isso, fica evidenciado que as variáveis incluídas neste modelo não são as mais relevantes para a previsão da inadimplência.

Por fim, ao comparar o modelo mais atual (31/12/2020) com o modelo que possui data base em junho do mesmo ano, foi percebido um percentual de acerto de 98,9% do percentual de acerto do modelo de dezembro de 2020. Isso pode acontecer pela proximidade entre as datas e, conseqüentemente, pelo perfil mais semelhante da carteira neste período. O R^2 de Nagelkerke também apresentou um resultado próximo, explicando 53% da variabilidade da inadimplência, 6% abaixo do modelo em comparação. O modelo de junho de 2020 incluiu as variáveis Idade, Renda Bruta e soma das parcelas (PMT) das Operações Consignadas, as

quais não estão presentes no modelo de dezembro de 2020 e também eliminou a variável soma do Valor Bruto das Operações Consignadas, a qual possuía relevância para a previsão da inadimplência em dezembro do mesmo ano.

As variáveis que foram relevantes para todos os modelos são Sexo, saldo em Operações não Consignadas e soma do Valor Bruto das Operações não Consignadas, ou seja, essas variáveis permaneceram significantes para a previsão da inadimplência em quatro períodos de tempo distintos. Quanto a variável Sexo, não foi feita uma distinção entre os gêneros para estimar se existe algum com maior probabilidade para a inadimplência, permanecendo como sugestão para estudos futuros. Em relação as operações não consignadas, possuem uma natureza mais arriscada, pois o banco que concede o empréstimo não possui garantia quanto ao seu pagamento, podendo apresentar um impacto maior na inadimplência.

Já as variáveis Região e saldo em Operações Consignadas não foram incluídas nos modelos, indicando que não são boas previsoras da inadimplência. Tanto a região de localização do tomador quanto o saldo de empréstimos consignados não têm poder de identificar a inadimplência. A região provavelmente não representa um diferencial no comportamento do tomador, não influenciando no seu perfil de bom ou mau pagador do empréstimo. O empréstimo consignado também não é um causador de inadimplência, mas o oposto, por ser uma operação mais segura para o concessor de crédito, pois o pagamento das parcelas é descontado de forma automática do tomador.

Os modelos com datas base mais distantes, ambos do ano de 2019, apresentaram um R^2 de Nagelkerke muito baixo, representando uma explicação da variabilidade da inadimplência de 31% em dezembro e apenas de 10% em junho. Sob este ângulo, estes dois modelos não podem ser considerados os mais adequados para a previsão da inadimplência, pois para possuírem um bom poder explicativo, precisam apresentar um R^2 de Nagelkerke de pelo menos 50%. Este resultado pode se dar devido a variáveis exógenas ao modelo, relacionados a fatores macroeconômicos do período.

Os resultados deste estudo se comparados aos estudos relacionados, confirmam a boa performance dos modelos de *credit scoring* em prever a inadimplência. Todas as pesquisas utilizaram modelos de *credit scoring* para a gestão do risco de crédito, exceto a pesquisa de Khandani, Kim e Lo (2010), a qual utilizou técnicas de *machine learning*. No entanto, o estudo de Khandani, Kim e Lo (2010) foi relevante por analisar dados tradicionais de transações de pessoas físicas para a previsão da inadimplência. Os demais estudos, diferentemente deste, utilizaram dados alternativos, gerados por novos meios de concessão de crédito originados pela adesão a novas tecnologias, para a previsão da inadimplência.

Portanto, a partir dos resultados apresentados para alcançar o objetivo de analisar a aplicação de um modelo de *credit scoring* de um banco tradicional no banco digital X, foi demonstrado que as bases de dados mais atuais trouxeram melhores resultados em termos de classificação correta de adimplentes e inadimplentes. Foi possível evidenciar a aderência do modelo de *credit scoring* de um banco tradicional ao banco digital X, aplicado neste estudo, pois todas as classificações apresentam assertividade acima do modelo sem previsores. Ademais, foi possível verificar a necessidade de um acompanhamento contínuo das variáveis previsoras da inadimplência, o qual é um processo dinâmico, pois a carteira de crédito pode passar por diversas alterações de perfil, conforme os impactos de mudanças econômicas e sociais na população.

5 CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo analisar a aplicação de um modelo de *credit scoring* de um banco tradicional no banco digital X. Para isso, foi realizada uma pesquisa qualitativa e quantitativa quanto à abordagem do problema, e descritiva no que diz respeito ao objetivo. O procedimento técnico se deu por meio de pesquisa documental. A análise dos dados foi documental, pela análise dos documentos coletados, e descritiva, com a utilização da técnica estatística de regressão logística binária, por meio do software estatístico IBM SPSS Statistics.

A partir da análise dos documentos coletados, foi possível descrever as principais características dos bancos digitais e de risco de crédito, descrever os principais modelos de risco de crédito existentes na literatura e apresentar as normas do Banco Central do Brasil quanto ao risco de crédito. Também, a partir da base de dados coletada por meio do banco digital X, foi aplicado o modelo de pontuação de crédito para a previsão da inadimplência.

Destaca-se que os bancos digitais se denominam dessa forma por estratégia de mercado e da própria operação, devido à falta de regulamentação específica para este tipo de negócio. Esses bancos oferecem um relacionamento diferenciado e remoto aos seus clientes, com vantagens e experiências como facilidade de acesso, custos mais baixos de serviços e integração com outras conveniências. Ao solicitar uma licença bancária, as instituições financeiras digitais devem cumprir os mesmos requisitos de licenciamento que os bancos tradicionais. Os bancos digitais incorrem em riscos financeiros semelhantes aos dos demais bancos, incluindo risco de crédito, mas enfrentam riscos específicos, como risco operacional e cibernético, devido à natureza de sua operação.

O risco está presente em todas as atividades dos seres humanos, sendo que a predisposição para tomá-lo é essencial ao desenvolvimento econômico (Jorion, 2007). O risco de crédito é um tipo de risco financeiro pelo qual as empresas estão expostas. Conforme Damodaran (2007), a primeira etapa para analisar e lidar com o risco nos contextos de portfólio e negócios é medir a sua aversão. Assim, para que haja um gerenciamento de risco, é necessário, primeiramente, a sua mensuração. Essa mensuração pode ser feita por modelos de gestão do risco de crédito, como o *credit scoring*.

De acordo com o Comitê de Basileia, a gestão de risco de crédito tem o objetivo de maximizar a taxa de retorno ajustada ao risco de um banco, mantendo a exposição ao risco de crédito em níveis aceitáveis (BCBS, 2000). No Brasil, a Resolução 4.557 de 2017, do Banco Central do Brasil, dispõe sobre as estruturas de gerenciamento de risco e de capital para as

instituições financeiras, as quais devem documentar os níveis de apetite por riscos na Declaração de Apetite por Riscos (RAS). A referida resolução ressalta que a exposição deve ser classificada como ativo problemático quando estiver atrasada em mais de noventa dias ou houver evidências de que não será cumprida sem a utilização de garantias, no que diz respeito ao gerenciamento do risco de crédito.

A Resolução n. 2.682/99 do Banco Central do Brasil dispõe que os bancos devem classificar as suas operações de crédito em ordem crescente de risco, conforme o prazo de inadimplência, e avaliar e ponderar cada um dos seus elementos com base em critérios verificáveis e consistentes. Para isso, a entidade deve provisionar um percentual do valor da operação para cada nível de classificação de risco. Ainda, por meio da Resolução n. 4.193/2013 e da Circular n. 3.648/2013, o BACEN regulamenta que as instituições financeiras devem seguir as orientações de Basileia para o cálculo do requerimento de capital para as exposições ao risco de crédito.

Os modelos para a mensuração e gestão do risco de crédito podem ser qualitativos ou quantitativos. Os modelos qualitativos, como os Sistemas Especialistas, contêm fatores específicos do tomador, como alavancagem e reputação, e fatores específicos do mercado, como o nível das taxas de juros e o ciclo de negócios. Um dos sistemas especialistas de crédito mais comum é o dos cinco Cs do crédito, o qual considera características de caráter, capital, capacidade, colateral e ciclo econômico para avaliar a concessão de crédito. Dentre os modelos quantitativos, há os modelos de probabilidade linear, o logit e o discriminante linear, que fornecem pontuações de crédito, ou *credit scoring*, classificando os empréstimos pelo risco de inadimplência esperado.

O modelo de *credit scoring* utiliza técnicas estatísticas para converter informações sobre um tomador de crédito em números que são combinados para formar uma pontuação, a qual é considerada uma medida do risco de crédito. Este modelo não avalia apenas se um atributo é positivo ou negativo, mas também a sua quantificação. É considerado a principal ferramenta para medir o risco de crédito no varejo, pois permite que os bancos evitem os clientes de maior risco e selecionem os que são considerados bons pagadores.

Ademais, a partir de 2004 foi introduzida, pelo Acordo de Basileia II, a abordagem baseada em ratings internos como uma alternativa à abordagem padronizada. Pela IRB, os bancos podem usar seus próprios modelos de classificação interna para estimar os diferentes componentes da perda esperada. Assim, os bancos que usam o IRB, após a autorização dos órgãos reguladores, podem se beneficiar de uma mensuração de risco de crédito mais

adequada ao seu perfil, muitas vezes com pesos de risco e encargos de capital menores, o que permite uma liberação maior de capital para a concessão de crédito.

A primeira etapa do estudo conceituou e contextualizou os principais assuntos discutidos neste estudo e criou um vínculo teórico com a segunda etapa. A partir das quatro regressões logísticas binárias executadas para analisar a aplicação do modelo de *credit scoring* de um banco tradicional no banco digital X, foi possível identificar que as bases de dados mais atuais, trouxeram melhores resultados em termos de classificação correta de adimplentes e inadimplentes.

O modelo com dados de dezembro de 2020, classificou corretamente 81,5% dos casos. As variáveis com maior poder preditivo foram Sexo, saldo em Operações não Consignadas, soma das parcelas das Operações não Consignadas, soma do Valor Bruto das Operações Consignadas e soma do Valor Bruto das Operações não Consignadas. Com isso, o estudo evidenciou a necessidade de um acompanhamento contínuo das variáveis previsoras da inadimplência, pois a carteira de crédito pode passar por diversas alterações de perfil, conforme ocorrem mudanças econômicas e sociais na população.

Portanto, os objetivos foram atingidos com a ampla descrição das características de instituição financeira digital e de risco de crédito, e com a apresentação das normas do Banco Central do Brasil quanto ao risco de crédito. Foram revisados os principais modelos de gestão de risco de crédito presentes na literatura dedicada ao tema e foi aplicado o modelo de *credit scoring* utilizado para a avaliação de risco de crédito de um banco tradicional no banco digital X, considerando os parâmetros do BACEN, cujos resultados confirmaram a sua aderência.

Este estudo contribuiu com a revisão da literatura relacionada ao assunto, com a análise de documentos atuais, trazendo dados atualizados sobre os temas estudados e demonstrou a importância da gestão contínua do risco de crédito, com a utilização de variáveis previsoras da inadimplência, para a liberação de capital por meio de um provisionamento menor. Como limitação do estudo, pode-se citar a dificuldade de obter acesso ao detalhamento dos modelos de risco de crédito utilizados pelos bancos digitais em suas operações, devido ao teor estratégico e sigiloso da informação. Para pesquisas futuras, sugere-se uma análise mais aprofundada das diferenças de modelos de gestão de risco de crédito e a inclusão de novas variáveis, inclusive exógenas, de cunho macroeconômico e social, e data base, bem como a aplicação do modelo em outras carteiras da própria instituição e de outros bancos digitais.

REFERÊNCIAS

- Adamko, P., Spuchl'áková, E., & Valášková, K. (2015). The History and Ideas Behind VaR. *Procedia Economics and Finance*, 24, 18–24. [https://doi.org/10.1016/S2212-5671\(15\)00607-3](https://doi.org/10.1016/S2212-5671(15)00607-3)
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4), 589-609.
- Antão, P., & Lacerda, A. (2011). Capital requirements under the credit risk-based framework. *Journal of Banking & Finance*, 35(6), 1380-1390.
- Avery, R. B., Calem, P. S., & Canner, G. B. (2004). Consumer credit scoring: do situational circumstances matter?. *Journal of Banking & Finance*, 28(4), 835-856.
- Azis, Y. M. (2021). Financial Ratio Analysis In Predicting Financial Distress Of Food And Beverage Companies With Logistic Regression. *Sinergi: Jurnal Ilmiah Ilmu Manajemen*, 11(2), 43-49.
- Banco Central do Brasil (2019). *Relatório de Economia Bancária*. https://www.bcb.gov.br/content/publicacoes/relatorioeconomiabancaria/REB_2019.pdf
- Banco Central do Brasil (2020). *Relatório de Economia Bancária*. https://www.bcb.gov.br/content/publicacoes/relatorioeconomiabancaria/REB_2020.pdf
- Banco Central do Brasil (2021). *Evolução Recente do Crédito no SFN*. https://www.bcb.gov.br/content/acessoinformacao/covid19_docs/Evolucao_Recente_do_Credito.pdf
- Banco Central do Brasil (2022). *Estatísticas. Notas Econômico-financeiras. Estatísticas Monetárias e de Crédito*. <https://www.bcb.gov.br/estatisticas/estatisticasmonetariascredito>
- Basel Committee on Banking Supervision (2000). *Principles for the management of credit risk*. Bank for International Settlements.
- Basel Committee on Banking Supervision (2020). *The Basel Framework*. Bank for International Settlements.
- Bazarbash, M. (2019). *Fintech in financial inclusion: machine learning applications in assessing credit risk*. International Monetary Fund.
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, 4, 71-111.
- Bessis, J. (2015). *Risk management in banking*. John Wiley & Sons.

- Biecek, P., Chlebus, M., Gajda, J., Gosiewska, A., Kozak, A., Ogonowski, D., Sztachelski, J., & Wojewnik, P. (2021). Enabling Machine Learning Algorithms for Credit Scoring-- Explainable Artificial Intelligence (XAI) methods for clear understanding complex predictive models. *arXiv preprint arXiv:2104.06735*.
- Bitetto, A., Cerchiello, P., Filomeni, S., Tanda, A., & Tarantino, B. (2021). *Machine Learning and Credit Risk: Empirical Evidence from SMEs* (No. 201). University of Pavia, Department of Economics and Management.
- Blöchlinger, A., & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking & Finance*, 30(3), 851-873.
- Borges, J. F. (2010). *Integração entre Abordagens Qualitativa e Quantitativa para a Mitigação do Risco Operacional: estudo no banco central do brasil*. [Dissertação de Mestrado, Fundação Edson Queiroz, Universidade de Fortaleza].
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, 57(1), 203-216.
- Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A. W., & Siddique, A. (2016). Risk and risk management in the credit card industry. *Journal of Banking & Finance*, 72, 218-239.
- Carbo-Valverde, S., Cuadros-Solas, P., & Rodríguez-Fernández, F. (2020). A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests. *Plos one*, 15(10), e0240362.
- Calmès, C., & Théoret, R. (2021). Portfolio analysis of big US banks' performance: the fee business lines factor. *Journal of Banking Regulation*, 22(2), 112-132.
- Caouette, J. B., Altman, E. I., Narayanan, P., & Nimmo, R. (2008). *Managing credit risk: the Great Challenge for the Global Financial Markets*, John Wiley & Sons.
- Chaimaa, B., Najib, E., & Rachid, H. (2021). E-banking Overview: Concepts, Challenges and Solutions. *Wireless Personal Communications*, 117(2), 1059-1078.
- Chen, Y., & Sivakumar, V. (2021). Investigation of finance industry on risk awareness model and digital economic growth. *Annals of Operations Research*, 1-22. <https://doi.org/10.1007/s10479-021-04287-7>
- Circular BACEN n. 3.648, de 04 de março de 2013. Estabelece os requisitos mínimos para o cálculo da parcela relativa às exposições ao risco de crédito sujeitas ao cálculo do requerimento de capital mediante sistemas internos de classificação do risco de crédito (abordagens IRB) (RWACIRB), de que trata a Resolução nº 4.193, de 1º de março de 2013. <https://www.bcb.gov.br/htms/Normativ/CIRCULAR3648.pdf>

- Cisko, S., Kliestik, T. (2013). *Financny manazment podniku II*, EDIS Publishers, University of Zilina, Zilina, pp. 775
- Colauto, R. D., & Beuren, I. M. (2008). Coleta, Análise e Interpretação de Dados. In Beuren, I. M. (Org.). *Como elaborar trabalhos monográficos em contabilidade: teoria e prática*. Atlas, 117-144.
- Crouhy, M., Galai, D., & Mark, R. (2000). A comparative analysis of current credit risk models. *Journal of Banking & Finance*, 24(1-2), 59-117.
- Crouhy, M., Galai, D., & Mark, R. (2005). *The Essentials of Risk Management*. McGraw Hill Professional.
- Cucinelli, D., Di Battista, M. L., Marchese, M., & Nieri, L. (2018). Credit risk in European banks: The bright side of the internal ratings based approach. *Journal of Banking & Finance*, 93, 213-229.
- Damodaran, A. (2007). *Strategic risk taking: a framework for risk management*. Pearson Prentice Hall.
- Demidenko, E. (2007). Sample size determination for logistic regression revisited. *Statistics in medicine*, 26(18), 3385-3397.
- Di Noia, A., Barabesi, L., Marcheselli, M., Pisani, C., & Pratelli, L. (2021). Omnibus goodness-of-fit tests for count distributions. *arXiv preprint arXiv:2112.11861*.
- Diamond, D.W. (1984). Financial intermediation and delegated monitoring. *Rev. Econ. Stud.* 51, 393–414.
- Drenovak, M., Ranković, V., Ivanović, M., Urošević, B., & Jelic, R. (2017). Market risk management in a post-Basel II regulatory environment. *European Journal of Operational Research*, 257(3), 1030-1044.
- Dwiarti, R., Hazmi, S., Santosa, A., & Rahman, Z. (2021). Does bankruptcy matter in non-banking financial sector companies?: evidence from Indonesia. *The Journal of Asian Finance, Economics and Business*, 8(3), 441-449.
- Ehrentraud, J., Ocampo, D. G., & Vega, C. Q. (2020). Regulating fintech financing: digital banks and fintech platforms. FSI Insights on policy implementation. No 27. August 2020. *Financial Stability Institute, Bank of International Settlements*.
- Engelmann, B. (2021). Credit risk modeling in the presence of central bank and government intervention (preprint).
- Federação Brasileira de Bancos (2021). *Pesquisa FEBRABAN de Tecnologia Bancária 2021*. <https://cmsarquivos.febraban.org.br/Arquivos/documentos/PDF/pesquisa-febraban-relatorio.pdf>

- Escanciano, J. C. (2009). On the lack of power of omnibus specification tests. *Econometric Theory*, 25(1), 162-194.
- Forcadell, F. J., Aracil, E., & Úbeda, F. (2020). The Impact of Corporate Sustainability and Digitalization on International Banks' Performance. *Global Policy*, 11, 18-27.
- García-Céspedes, R., & Moreno, M. (2017). An approximate multi-period Vasicek credit risk model. *Journal of Banking & Finance*, 81, 105-113.
- Gavlakova, P., Kliestik, T., 2014. Credit Risk Models and Valuation. In: 4th International Conference on Applied Social Science (ICASS 2014), *Information Engineering Research Institute, Singapore, Advances in Education Research*, 51, 139-143.
- Gijbels, I., Omelka, M., & Veraverbeke, N. (2021). Omnibus test for covariate effects in conditional copula models. *Journal of Multivariate Analysis*, 186, 104804.
- Gil, A. C. (2002). *Como elaborar projetos de pesquisa*. Atlas.
- Giudici, P., Hadji-Misheva, B., & Spelta, A. (2019). Network based scoring models to improve credit risk management in peer to peer lending platforms. *Frontiers in artificial intelligence*, 2, 3.
- Gujarati, D. N. (2000). *Econometria básica*. Makron Books.
- Gutiérrez-Nieto, B., Serrano-Cinca, C., & Camón-Cala, J. (2016). A credit score system for socially responsible lending. *Journal of Business Ethics*, 133(4), 691-701.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis* (Vol. 5, No. 3, pp. 207-219). Upper Saddle River, NJ: Prentice hall.
- Jacobson, T., & Roszbach, K. (2003). Bank lending policy, credit scoring and value-at-risk. *Journal of banking & finance*, 27(4), 615-633.
- Joaquim, G., van Doornik, B. F. N., & Ornelas, J. R. (2019). *Bank competition, cost of credit and economic activity: evidence from Brazil*. Banco Central do Brasil.
- Jolevski, L. (2017). The impact of the non-performing loans to banks' performance: the case of the Republic of Macedonia. *Emc Review-Časopis Za Ekonomiju*, 14(2).
- Jorion, P. (2007). *Value at risk: the new benchmark for managing financial risk*. The McGraw-Hill Companies, Inc.
- Kanas, A. (2014). Default risk and equity prices in the US banking sector: Regime switching effects of regulatory changes. *Journal of International Financial Markets, Institutions and Money*, 33, 244-258.
- Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767-2787.

- Kyeong, S., Kim, D., & Shin, J. (2022). Can System Log Data Enhance the Performance of Credit Scoring?—Evidence from an Internet Bank in Korea. *Sustainability*, 14(1), 130.
- Lei Complementar n. 101, de 4 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências.
http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm
- Lei n. 4.595, de 31 de dezembro de 1964. Dispõe sobre a Política e as Instituições Monetárias, Bancárias e Creditícias, Cria o Conselho Monetário Nacional e dá outras providências.
http://www.planalto.gov.br/ccivil_03/leis/14595.htm
- Lei n. 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD).
http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm
- Li, Y. (2019). Credit risk prediction based on machine learning methods. In 2019 *14th International Conference on Computer Science & Education (ICCSE)* (pp. 1011-1013). IEEE.
- Liao, H. H., Chen, T. K., & Lu, C. W. (2009). Bank credit risk and structural credit models: Agency and information asymmetry perspectives. *Journal of Banking & Finance*, 33(8), 1520-1530.
- Lobo, G. J. (2017). Accounting research in banking: A review. *China Journal of Accounting Research*, 10(1), 1-7.
- Marconi, M. de A., & Lakatos, E. M. (2003). *Fundamentos da Metodologia Científica* (5o ed). Atlas.
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77-91.
- Markowitz, H. M. (1999). The early history of portfolio theory: 1600–1960. *Financial analysts journal*, 55(4), 5-16.
- Markowitz, H. (2014). Mean–variance approximations to expected utility. *European Journal of Operational Research*, 234(2), 346-355.
- Mishkin, F. S. (1992). Anatomy of a financial crisis. *Journal of evolutionary Economics*, 2(2), 115-130.
- Mishra, M. (2020). Evolution of the Invisible bank: How Partnerships with FinTechs Are Driving Digital Innovation. *Journal of Digital Banking*, 5: 36–40.
<https://hstalks.com/article/5804/evolution-of-the-invisible-bank-how-partnershipsw/>
- Moscato, V., Picariello, A., & Sperlí, G. (2021). A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications*, 165, 113986.
- Nehrebecka, N. (2021). Internal credit risk models and digital transformation: what to prepare for? an application to Poland. *European Research Studies Journal*, 24(Special 3), 719-736.

- Niemand, T., Rigtering, C., Kallmünzer, A., Kraus, S., & Matijas, S. (2017). Entrepreneurial orientation and digitalization in the financial service industry: A contingency approach. In Proceedings of the 25th *European Conference on Information Systems (ECIS)*, 1081-1096.
- Novos concorrentes colocam bancos digitais a prova (2019, dezembro 18). Diário do Comércio. <https://diariodocomercio.com.br/financas/novos-concorrentes-colocam-bancos-digitais-a-prova/>
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 18(1), 109-131.
- Platanakis, E., & Urquhart, A. (2020). Should investors include bitcoin in their portfolios? A portfolio theory approach. *The British accounting review*, 52(4), 100837.
- Raei, R., Farhangzadeh, B., Safizadeh, M., & Raei, F. (2016). Study of the Relationship between Credit Diversification Strategy and Banks' Credit Risk and Return: Evidence from Tehran Stock Exchange (TSE). *Procedia Economics and Finance*, 36, 62–69. [https://doi.org/10.1016/S2212-5671\(16\)30016-8](https://doi.org/10.1016/S2212-5671(16)30016-8)
- Rana, R., & Ghildiyal, V. (2016). A Comparative Analysis of Credit Risk in Investment Banks: A Case Study of JP Morgan, Merrill Lynch and Bank Of America.
- Raupp, F. M., & Beuren, I. M. (2008). Metodologia da Pesquisa Aplicável às Ciências. In Beuren, I. M. (Org.). *Como elaborar trabalhos monográficos em contabilidade: teoria e prática*. Atlas, 76-97.
- Refsdal, A., Solhaug, B., & Stølen, K. (2015). Cyber-risk management. In *Cyber-Risk Management* (pp. 33-47). Springer, Cham.
- Resolução BACEN n. 2.682, de 21 de dezembro de 1999. Dispõe sobre critérios de classificação das operações de crédito e regras para constituição de provisão para créditos de liquidação duvidosa. https://www.bcb.gov.br/pre/normativos/res/1999/pdf/res_2682_v2_L.pdf
- Resolução BACEN n. 4.193, de 01 de março de 2013. Dispõe sobre apuração dos requerimentos mínimos de Patrimônio de Referência (PR), de Nível I e de Capital Principal e institui o Adicional de Capital Principal. https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/49006/Res_4193_v1_O.pdf
- Resolução BACEN n. 4.553, de 30 de janeiro de 2017. Estabelece a segmentação do conjunto das instituições financeiras e demais instituições autorizadas a funcionar pelo Banco Central do Brasil para fins de aplicação proporcional da regulação prudencial. https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/50335/Res_4553_v1_O.pdf
- Resolução BACEN n. 4.557, de 23 de fevereiro de 2017. Dispõe sobre a estrutura de gerenciamento de riscos e a estrutura de gerenciamento de capital.

https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/50344/Res_4557_v1_O.pdf

Resolução BACEN n. 4.783, de 16 de março de 2020. Estabelece, por prazos determinados, percentuais a serem aplicados ao montante RWA, para fins de apuração da parcela ACPConservação de que trata a Resolução nº 4.193, de 1º de março de 2013.
https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/50938/Res_4783_v1_O.pdf

Roa, L., Correa-Bahnsen, A., Suarez, G., Cortés-Tejada, F., Luque, M. A., & Bravo, C. (2021). Super-app behavioral patterns in credit risk models: Financial, statistical and regulatory implications. *Expert Systems with Applications*, 169, 114486.

Saâda, M. B., & Gafsi, Y. (2019). Does disclosure of internal control system of credit risk improve banks' performance? Evidence from Tunisian listed banks. *International Journal of Financial Engineering*, 6(04), 1950031.

Sanvicente, A. Z., & Minardi, A. M. A. F. (1998). Identificação de indicadores contábeis significativos para a previsão de concordata de empresas. *Instituto Brasileiro de Mercado de Capitais, Working Paper*, (1968), 1-12.

Sanvicente, A. Z., & Minardi, A. M. A. F. (1999). Migração de Risco de Empresas Brasileiras: *Uma Aplicação de Análise de Clusters na Área de Crédito* (No. flwp_16). Finance Lab, Insper Instituto de Ensino e Pesquisa.

Saputra, M. P. A., & Chaerani, D. (2022). Estimation of Maximum Potential Losses for Digital Banking Transaction Risks Using the Extreme Value-at-Risks Method. *Risks*, 10(1), 10.

Saunders, A., & Allen, L. (2002). *Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms*. John Wiley & Sons.

Saunders, A., & Cornett, M. M. (2014). *Financial Markets and Institutions*. McGraw-Hill Education.

Saunders, A., & Cornett, M. M. (2017). *Financial institutions management: a risk management approach*. McGraw-Hill Education.

Sclip, A., Girardone, C., & Miani, S. (2019). Large EU banks' capital and liquidity: Relationship and impact on credit default swap spreads. *The British Accounting Review*, 51(4), 438-461.

Singh, S. (2019). Purpose and Process of Research. In *Methodological Issues in Management Research: Advances, Challenges, and the Way Ahead*. Emerald Publishing Limited, 27-36.

Stiroh, K. J. (2006). A portfolio view of banking with interest and noninterest activities. *Journal of Money, Credit and Banking*, 1351-1361.

- Surjanovic, N., & Loughin, T. M. (2021). Improving the Hosmer-Lemeshow Goodness-of-Fit Test in Large Models with Replicated Trials. *arXiv preprint arXiv:2102.12698*.
- Tsai, C. F. (2009). Feature selection in bankruptcy prediction. *Knowledge-Based Systems*, 22(2), 120-127.
- Vicente, E. F. R. (2001). *A estimativa do risco na constituição da PDD* [Tese de Doutorado, Universidade de São Paulo].
- World Bank Group (2019). *Credit Scoring Approaches Guidelines*. Technical report.
- Yanenkova, I., Nehoda, Y., Drobyazko, S., Zavorodnii, A., & Berezovska, L. (2021). Modeling of Bank Credit Risk Management Using the Cost Risk Model. *Journal of Risk and Financial Management*, 14(5), 211.
- Zarei, A. (2016). Measurement of credit risk portfolio of bank by method of coefficient variation bound. *Social Sciences (Pakistan)*, 11(16), 3908–3913.
<https://doi.org/10.3923/sscience.2016.3908.3913>

APÊNDICE A - TERMOS DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA PARTICIPAÇÃO EM PESQUISA – GESTOR DE RISCOS

Título do Projeto: *CREDIT SCORING* EM INSTITUIÇÃO FINANCEIRA DIGITAL

Prezado(a) Gestor(a) de Riscos,

Você está sendo convidado(a) a participar desta pesquisa sobre gestão de risco de crédito em instituições financeiras digitais, uma iniciativa que tem o objetivo de analisar a aplicação dos modelos empregados na avaliação de risco de crédito de um banco tradicional em um banco digital. As suas informações serão relevantes para a formulação do modelo.

Trata-se de um projeto de Mestrado, sob a orientação do Prof. Dr. Marco Antônio Martins, vinculado ao Programa de Pós-Graduação em Controladoria e Contabilidade (PPGCONT) da Universidade Federal do Rio Grande do Sul.

Para participar, é preciso ser o responsável pela gestão de risco de crédito, mesmo que formalmente designado em outro cargo na instituição.

Sua participação consiste em disponibilizar os relatórios e dados solicitados.

Sua identificação será preservada, sendo levantados os dados do Banco Digital e apresentados de forma agregada.

Em caso de dúvidas ou esclarecimentos que se fizerem necessários, você poderá entrar em contato com Adriana Boito Johnson (mestranda), através do e-mail e telefone: adriana.boito@gmail.com e (51) 99993 1625.

Destacamos que:

- Você não receberá nenhuma compensação financeira relacionada à sua participação;
- Se decidir interromper sua participação, isso não lhe causará nenhum prejuízo;
- Ressalta-se que, uma vez participando da pesquisa, o respondente autoriza a divulgação dos resultados em sua totalidade, de forma que não será identificado individualmente e sendo esses utilizados exclusivamente para fins acadêmicos.

Você aceita participar desta pesquisa?

Sim

Não

Nome do participante da pesquisa

Assinatura

Adriana Boito Johnson _____

Nome do pesquisador que aplicou o Termo

Assinatura

Local e Data: _____