

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ADMINISTRAÇÃO  
DEPARTAMENTO DE CIÊNCIAS ADMINISTRATIVAS

GABRIELA WILLE CLASEN

**ANÁLISE AUTOMÁTICA DE CRÉDITO:**  
DESEMPENHO DA MINERAÇÃO DE DADOS DE UMA ÁRVORE DE DECISÃO

PORTO ALEGRE  
2018

GABRIELA WILLE CLASEN

**ANÁLISE AUTOMÁTICA DE CRÉDITO:**  
DESEMPENHO DA MINERAÇÃO DE DADOS DE UMA ÁRVORE DE DECISÃO

Trabalho de Conclusão de Curso de graduação apresentado ao Departamento de Ciências Administrativas da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Bacharel em Administração.

Orientador: Prof. Dr. Marcelo Scherer Perlin

PORTO ALEGRE

2018

## CIP - Catalogação na Publicação

Clasen, Gabriela  
ANÁLISE AUTOMÁTICA DE CRÉDITO: DESEMPENHO DA  
MINERAÇÃO DE DADOS DE UMA ÁRVORE DE DECISÃO / Gabriela  
Clasen. -- 2018.  
72 f.  
Orientador: Marcelo Perlin.

Trabalho de conclusão de curso (Graduação) --  
Universidade Federal do Rio Grande do Sul, Escola de  
Administração, Curso de Administração, Porto Alegre,  
BR-RS, 2018.

1. Crédito. 2. Decisão Automática. 3. Árvore de  
Decisão. I. Perlin, Marcelo, orient. II. Título.

GABRIELA WILLE CLASEN

**ANÁLISE AUTOMÁTICA DE CRÉDITO:**  
DESEMPENHO DA MINERAÇÃO DE DADOS DE UMA ÁRVORE DE DECISÃO

Trabalho de Conclusão de Curso de graduação apresentado ao Departamento de Ciências Administrativas da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Bacharel em Administração.

Porto Alegre, 05 de dezembro de 2018.

Conceito Final: A

Aprovado em 05 de dezembro de 2018

BANCA EXAMINADORA

---

Prof. Marcelo Perlin (orientador)  
Universidade Federal do Rio Grande do Sul

---

Prof. Guilherme Kirch  
Universidade Federal do Rio Grande do Sul

## **AGRADECIMENTOS**

A Deus, por ter iluminado meu caminho todos os dias nessa trajetória acadêmica, desde o pré-vestibular, fazendo com que fosse concretizado o sonho de estudar nessa Universidade.

Mãe e Pai, por serem meu maior exemplo e nunca deixarem de me apoiar. Mãe, por me ouvir, por aconselhar, pela empolgação desde os lanchinhos até a escolha de cadernos, canetas e estojos. Pai, por vibrar junto comigo cada vitória, pela carona de todo dia de manhã, e às vezes, de noite.

Cecília, por me fazer companhia no quarto nas inúmeras e incontáveis horas passadas realizando esta monografia, seja no meu colo ou brincando. Por ser a alegria de chegar em casa e ter o melhor abraço do mundo.

Meu orientador, Prof. Dr. Marcelo Perlin, por desde o primeiro dia de orientação ter sido um mentor, desde a escolha do tema, direcionamentos do assunto, seleção do modelo, até ensinar a mexer no software no meu computador (que não é dos mais rápidos).

Paulo, Moll e Lisi, sem vocês esse trabalho não seria possível. Obrigada por auxiliar no acesso e interpretação dos dados, serem os guias deste estudo e sempre estarem dispostos a ajudar (até mesmo um dia antes de uma grande viagem).

William, por ter me ensinado a acessar e utilizar as bases de dados, pelas inúmeras vezes que eu precisei interromper tuas outras atividades para pedir ajuda (apesar de que tu nunca tinhas muita coisa para fazer).

Ao Presidente Romildo Bolzan e a Renato Portaluppi, por terem trazido meu Grêmio de volta e dado tantas alegrias nesse um ano e meio (período desde a idealização deste estudo) e tornarem a realização dele muito mais agradável e bem-humorada.

## RESUMO

Este estudo apresenta a análise de mineração de dados de uma árvore de decisão utilizada para aprovação ou negativa automática de crédito em uma instituição financeira. A partir do conteúdo inicial, é proposto que seja validada a efetividade da análise de crédito automática, objetivando verificar a assertividade das variáveis utilizadas no modelo estatístico que baseia esta análise, de forma a garantir ganhos de eficiência e redução de inadimplência para a instituição. Para tal, este estudo analisa o desempenho histórico dos clientes baseado nas variáveis definidas, de maneira a comprovar eficácia do modelo e verificar os casos de inadimplência. Inicialmente, é apresentada a metodologia de decisão automática e o contexto da instituição financeira que a utiliza, dissertando acerca do conceito dos principais produtos e fontes de recurso das operações de crédito. A base de dados da instituição é aplicada a um modelo linear generalizado com função *probit*, para que sejam encontradas as variáveis que possuem maior significância na ocorrência de *default*. A partir do resultado da regressão, o estudo propõe ajustes a serem aplicados na árvore de decisão, para confirmar os ganhos de eficiência que a análise automática possui quando comparada à análise manual.

Palavras-chave: Crédito. Decisão Automática. Árvore de Decisão.

## **ABSTRACT**

This study presents the data mining analysis of a decision tree used in automatic credit approval or declines in a financial institution. By the initial content, it is proposed to validate the effectiveness of automatic credit analysis, aiming to verify the assertiveness of the variables used in the statistic model that is the base of this analysis, in order to ensure efficiency gains and decrease in delinquency level for the institution. For this purpose, this study examines the historical performance of clients based in the variables defined in the current model, in order to prove the effectiveness of the model and check the default cases. At first, the automatic decision methodology is presented with the context of the financial institution which uses this approach, discussing on the concept of the main financial products and funding resources of the credit deals. The database of the institution is applied to a Linear Generalized Model with probit function, in order to find the variables that present the most significance level in default occurrences. By the regression result, this study purposes adjustments to be applied in the decision tree, in order to confirm the efficiency gains that automatic decision has when compared to manual decision.

Key-words: Credit. Automatic Decision. Decision Tree.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Organograma.....	23
Figura 2 – Decisões Automáticas.....	31
Figura 3 – Regularização dos Casos.....	33
Figura 4 – Classificação dos clientes.....	34
Quadro 1 – Empréstimos e Financiamentos.....	09
Quadro 2 – Programas de Financiamento do BNDES.....	11
Quadro 3 – 5 Cs do Crédito.....	15
Quadro 4 – Alçadas de Crédito.....	24
Quadro 5 – Critérios-base de Decisão Automática.....	28
Quadro 6 – Variáveis Utilizadas na Aplicação do Modelo Linear Generalizado.....	39

## LISTA DE TABELAS

Tabela 1 – Distribuição de casos de <i>Default</i> e regularização por faixa de atraso.....	32
Tabela 2 – Distribuição de Defaults por ano.....	33
Tabela 3 – Situação Financeira Atual dos Clientes.....	35
Tabela 4 – Endividamento dos Proponentes.....	36
Tabela 5 – Propriedades e Receitas x Valor do Financiamento.....	36
Tabela 6 – Inadimplência da decisão automática x decisão manual.....	38
Tabela 7 – Resultado da Regressão.....	42

## LISTA DE ABREVIATURAS E SIGLAS

BNDES	Banco Nacional de Desenvolvimento Social e Econômico
CDC	Crédito Direto ao Consumidor
CNDL	Confederação Nacional de Dirigentes Lojistas
IBGE	Instituto Brasileiro de Geografia e Estatística
IRB	<i>Internal Ratings-Based</i>
PRONAF	Programa Nacional de Fortalecimento da Agricultura Familiar
SFN	Sistema Financeiro Nacional
SPC	Serviço de Proteção ao Crédito

## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>2</b>
<b>2. JUSTIFICATIVA.....</b>	<b>7</b>
<b>3. REVISÃO TEÓRICA .....</b>	<b>8</b>
3.1. CONTEXTO DO ESTUDO.....	8
3.1.1 Bancos Múltiplos .....	8
3.2.2. Banco Nacional de Desenvolvimento Econômico e Social .....	9
3.2. REFERENCIAL TEÓRICO .....	12
3.3. REVISÃO DE ESTUDOS REALIZADOS .....	17
<b>4. OBJETIVOS.....</b>	<b>21</b>
4.1. OBJETIVO GERAL .....	21
4.2. OBJETIVOS ESPECÍFICOS.....	21
<b>5. INSTITUIÇÃO FINANCEIRA ALVO DO ESTUDO .....</b>	<b>22</b>
<b>6. PROCEDIMENTOS METODOLÓGICOS.....</b>	<b>25</b>
<b>7. CRITÉRIOS DA DECISÃO AUTOMÁTICA .....</b>	<b>28</b>
<b>8. ANÁLISE DOS DADOS DOS CLIENTES.....</b>	<b>30</b>
8.1. ANÁLISE DESCRITIVA DOS CASOS DE DEFAULT .....	30
8.2. APLICAÇÃO DO MODELO LINEAR GENERALIZADO.....	38
<b>9. CONSIDERAÇÕES FINAIS .....</b>	<b>46</b>
<b>REFERÊNCIAS.....</b>	<b>48</b>
<b>ANEXO I – RESULTADO DA REGRESSÃO PROBIT.....</b>	<b>60</b>

## 1. INTRODUÇÃO

O Sistema Financeiro Nacional (SFN) é definido por Fortuna (2015, p. 16) como o “conjunto de instituições que se dedicam, de alguma forma, ao trabalho de propiciar condições satisfatórias para a manutenção de um fluxo de recursos entre poupadores e investidores”. O mercado financeiro brasileiro integra o SFN, que, conforme o Relatório de Estabilidade Financeira publicado pelo BACEN em Abril de 2018 (Banco Central do Brasil, 2018), desde 2017 está em retomada do crescimento econômico, queda da inflação e redução da taxa básica de juros.

Esses fatores colaboraram, por um lado, para a melhoria nos indicadores de crédito, e, por outro, para que as instituições financeiras estivessem dispostas a assumir riscos maiores, dada a contrapartida de retorno e rentabilidade. O crédito é visto como um dos principais instrumentos para fomentar a economia do país, considerando que a disponibilidade de crédito seria um propulsor do crescimento econômico, por sustentar a demanda agregada e, por meio de investimentos no setor produtivo, propiciar um aumento da produtividade na economia (GALEANO; FEIJÓ, 2011).

O volume de crédito seguiu crescendo ao longo dos anos, porém o percentual de operações de crédito em prejuízo teve um salto de 2% para 3,6% de 2011 a 2016 (Banco Central do Brasil, 2017). A esse respeito, Magalhães (2017, p. 10) declara que:

O mercado de crédito brasileiro tem evoluído consideravelmente nos últimos anos em diversas vertentes, mas, principalmente, no que tange a quesitos relacionados à gestão de análise de crédito e seus riscos, visto que, com o aumento das concessões de crédito, também aumentaram os percentuais de inadimplência e prejuízo.

O relatório do BACEN (Banco Central do Brasil, 2018) também indica que, na percepção das instituições financeiras, os riscos relacionados à inadimplência e à recessão têm perdido força, associados à recuperação econômica no Brasil. Isso é possível de identificar notando a queda significativa na inadimplência do setor bancário em 2017 (0,46 p.p), retornando a um nível não atingido desde dezembro de 2015.

O principal momento de prevenir riscos de prospecção de clientes que poderão gerar inadimplência dá-se na análise de crédito. Segundo Steiner et al. (1999, p. 56), “a correta decisão de crédito é essencial para a sobrevivência das empresas bancárias”. Os autores ainda afirmam que “qualquer erro na decisão de concessão pode significar que em uma única operação haja a perda do ganho obtido em dezenas de outras bem-sucedidas”.

Acompanhando o constante avanço do conhecimento e tecnologias que promovem dinamismo em processos, a necessidade de retorno cada vez mais rápido da decisão de análise de crédito para empréstimos e financiamentos se tornou uma realidade no dia-a-dia das instituições financeiras. Denomina-se *credit scoring* a todo sistema de avaliação de crédito que permite valorar de forma automática o risco associado a cada solicitação de crédito (MEDINA, 2013). O *score* de crédito também se pode advir de diferentes abordagens de modelos estatísticos não-paramétricos e de inteligência artificial (THOMAS, 2000).

O modelo de *score* de crédito é definido por Chaia (2003, p. 23) como o uso de ferramental estatístico na identificação dos fatores determinantes da probabilidade de o cliente tornar-se inadimplente, e aponta como principal vantagem o fato de que “[...] decisões sobre a concessão são tomadas com base em procedimentos impessoais e padronizados, gerando um maior grau de confiabilidade”. Entre as diversas metodologias utilizadas para definir o modelo estatístico para formar a mineração de dados que embasa as análises de crédito das instituições, os mais utilizados são redes neurais, algoritmos genéricos e árvores de decisão (THOMAS, 2000). Os métodos quantitativos, nesse caso, originariam o *credit scoring* nos anos seguintes, que são a base da análise automática de crédito citada neste estudo.

Ao invés de visar minimizar a porcentagem de clientes que se tornam inadimplentes, Thomas (2000) ainda ressalta a necessidade de, nesse mesmo processo, também identificar quais são os clientes mais lucrativos. Parte da catálise para esse desenvolvimento é o aumento massivo de informações nas transações dos clientes que aconteceu na última década com o uso de ferramentas como *big data* e *open banking*. No estudo publicado pela McKinsey (2010), o fluxo de dados disponíveis duplicava a cada 18 meses, e as instituições financeiras estavam iniciando a

experimentação dessas ferramentas para segmentar os clientes em grupos de risco específicos. Em 2015, observa-se no Brasil o uso dessas ferramentas já em instituições financeiras como o Banco do Brasil, Itaú e Bradesco, mas ainda com a finalidade de organização e montagem de bases de dados que transitam diariamente nos canais eletrônicos (TESSAROLO; MAGALHÃES, 2015).

Diversos métodos de classificação de *score* de crédito em mineração de dados foram apresentados, sendo que, de acordo com Lemos, Steiner e Nievola (2005, p. 228):

O método de árvore de decisão é particularmente efetivo para construir um classificador da amostra de dados. [...] é conveniente usar árvore de decisão quando o objetivo for a categorização dos dados. Ela também é uma boa escolha quando o objetivo é gerar regras que podem ser facilmente entendidas, explicadas e traduzidas para linguagem natural.

A tarefa de conceder ou não crédito envolve diversas variáveis, com pesos que mudam de acordo com o contexto da instituição financeira e do tipo de mercado envolvido na análise. É necessário, além de preocupar-se em prever quais clientes possivelmente darão default, também analisar os motivos de default desses clientes e seu relacionamento com variáveis econômicas e sociais.

Com um modelo estatístico aderente para a árvore de decisão, é possível minimizar as ameaças de inadimplência e reduzir consideravelmente o tempo de retorno da análise, gerando maior volume de negócios tanto para o financiador quanto para o financiado. Mester (1997) indica que a exatidão do modelo, a atualização dos dados e a avaliação e readequação dos modelos são alguns fatores críticos do *credit scoring*. Falhas nesses fatores limitam o uso de tal modelo.

O fenômeno de automatização de processos tem levado as instituições financeiras a investir massivamente em modelos de análise de crédito via *credit scoring* e *rating*, objetivando o maior número de decisões automáticas possível. Porém, Sousa (2014) nota que a análise de crédito utilizando mineração de dados ainda é rara no Brasil. Ainda não há um resultado conclusivo sobre qual metodologia é a melhor a ser utilizada, dado que sua validação dá-se no desenrolar do desempenho de pagamento dos clientes que receberam o empréstimo ou financiamento, o qual depende fatores relativos de acordo com o macroambiente.

É necessário que clientes que estão no mesmo tipo de mercado sejam avaliados com base nos mesmos critérios (TOM, 2005). Dado que o panorama de mercado econômico está em constante mudança, as variáveis abrangidas pelo modelo e seus pesos devam ser sempre atualizados. Isso indica que o trabalho a ser desenvolvido também é contínuo, porém uma vez que o comportamento das variáveis é identificado, a aplicação da metodologia para validar a árvore de decisão pode ser feita regularmente.

Um sistema de análise automática de crédito pode rapidamente unir a informação necessária, avaliar e determinar se a solicitação de crédito deve ou não ser aprovada (SAKPRASAT; SINCLAIR, 2007). Isso reduz o número de solicitações de crédito que precisam de uma revisão e análise mais aprofundada por parte do analista, diminuindo também o desperdício de tempo e permitindo que os analistas possam focar apenas nas solicitações que são realmente mais difíceis ou importantes.

Medina (2013) propõe a classificação de clientes utilizando a seguinte régua: acima de 70% de probabilidade de pagamento devem ser aprovados automaticamente; menos que 30% de probabilidade de pagamento devem ser negados automaticamente, e entre essa porcentagem os casos devem ser considerados como duvidosos.

Este estudo busca validar a efetividade da análise de crédito automática proposta por uma instituição financeira de uma rede global, com presença em mais de trinta países, incluindo o Brasil. O nome da instituição financeira analisada neste estudo foi omitido, evitando exposição de imagem e divulgação do que pode ser considerado como vantagem competitiva.

Ferramentas quantitativas aliadas à experiência do analista de crédito são imprescindíveis, pois, incompativelmente, o rápido retorno da análise de crédito se depara com financiamentos com períodos de carência que podem ir, por exemplo, a até 18 meses, conforme permitido pelo Banco Nacional de Desenvolvimento Econômico e Social (BNDES).

Considerando a tendência de aumento de solicitações de crédito e necessidade de rápido retorno, a instituição financeira analisada nesse trabalho vem aplicando a análise de crédito automática para clientes que atendem a determinados quesitos, organizados estatisticamente de acordo com sua relevância para o modelo. Levando

em conta os investimentos necessários para implementação dessa ferramenta, este estudo almeja validar as variáveis selecionadas para minerar os dados dos clientes na análise de crédito automática, de forma que agreguem assertividade ao processo, sejam confiáveis e justifiquem o uso desse instrumento para a organização.

## 2. JUSTIFICATIVA

O crédito é visto como um dos principais propulsores do desenvolvimento do país, influenciando na alavancagem do investimento e acumulação de capital; no financiamento da inovação; e, para a sociedade, na redução da concentração de recursos e quebra da dinâmica centro-periferia (ROMERO; ÁVILA, 2010). Além de socioeconomicamente, o resultado do estudo poderá contribuir de maneira direta para o aumento de retorno sobre o capital investido dos acionistas da instituição financeira abordada.

Devido à elaboração ainda recente de análise automática de crédito baseada em modelos estatísticos, existe ainda certa dificuldade em determinar quais variáveis (e seus respectivos pesos) devem ser utilizadas e abordadas pela árvore de decisão. O presente estudo argumenta que esse obstáculo pode ser superado com o estudo aprofundado de comportamento das variáveis, considerando, como métrica de sucesso ou fracasso, o desempenho de pagamento dos clientes que receberam o crédito.

Espera-se que, com o resultado do estudo, seja comprovada a aderência e confiabilidade do modelo de árvore de decisão utilizado na aprovação automática, fazendo com que as solicitações de crédito possam retornar positivamente ou negativamente de maneira instantânea ao cliente. Considerando a relevância de confiabilidade na decisão automática, os impactos positivos são vistos em diferentes áreas das instituições financeiras - no aumento de volume de negócios de áreas comerciais, na redução de clientes inadimplentes para as áreas de cobrança e eficiência para áreas de crédito como um todo.

Tendo em vista que a aplicação do modelo estatístico atualmente utilizado para embasar a análise automática de crédito foi realizada recentemente (desde Dezembro de 2014), o estudo torna-se pertinente visto que o resultado financeiro de um banco está diretamente atrelado ao sucesso da decisão correta de crédito, sendo uma importante ferramenta de suporte ao processo analítico dos clientes da instituição.

### 3. REVISÃO TEÓRICA

Na mesma vertente, o capítulo apresenta os principais conceitos que embasam a análise de crédito, bem como as definições do modelo estatístico utilizado na análise automática de crédito, o qual será validado no presente estudo. Igualmente, nessa seção, os estudos e pesquisas já existentes sobre a temática da monografia serão sucintamente discutidos.

#### 3.1. CONTEXTO DO ESTUDO

Para que seja possível contextualizar o assunto a ser estudado, este capítulo busca apresentar brevemente o panorama da instituição financeira que é abordada na pesquisa.

##### 3.1.1 Bancos Múltiplos

A instituição financeira que é abordada nessa pesquisa, em termos de estudo de seus dados e procedimentos metodológicos para análise de crédito automático, é considerada como um Banco Múltiplo. O Banco Central Brasileiro define os Bancos Múltiplos como

“Instituições financeiras privadas ou públicas que realizam as operações ativas, passivas e acessórias das diversas instituições financeiras, por intermédio das seguintes carteiras: comercial, de investimento e/ou de desenvolvimento, de crédito imobiliário, de arrendamento mercantil e de crédito, financiamento e investimento. Essas operações estão sujeitas às mesmas normas legais e regulamentares aplicáveis às instituições singulares correspondentes às suas carteiras. A carteira de desenvolvimento somente poderá ser operada por banco público. O banco múltiplo deve ser constituído com, no mínimo, duas

carteiras, sendo uma delas, obrigatoriamente, comercial ou de investimento, e ser organizado sob a forma de sociedade anônima.” (RESOLUÇÃO Nº 2.099, 17/8/1994, 1994, p. 6)

Dentre as modalidades de operações disponíveis para bancos múltiplos, o banco abordado nesse estudo transaciona apenas com financiamentos atrelados (ou não) à limites de crédito, optando por não operar com empréstimos de capital de giro (Quadro 1).

**Quadro 1 – Empréstimos e Financiamentos**

	EMPRÉSTIMOS	FINANCIAMENTOS
VINCULAÇÃO A UM BEM	Não	Sim
TAXAS DE JUROS	Maiores	Menores
PRAZO DE PAGAMENTO	Menores	Maiores
VALORES	Mais baixos	Mais altos
BEM COMO GARANTIA	Não	Sim

Fonte: Elaborada pelo autor

Essa escolha dá-se principalmente pela formação de parcerias com Sociedades de Crédito, Financiamento e Investimento (SFI), conhecidas popularmente como “bancos de fábrica”, que concentram suas operações no financiamento de seus próprios produtos.

### **3.2.2. Banco Nacional de Desenvolvimento Econômico e Social**

A Instituição Financeira em questão utiliza-se principalmente de recursos advindos do Banco Nacional de Desenvolvimento Econômico e Social (BNDES), porém também possui uma carteira de financiamentos oriunda de recursos próprios. De acordo com Filho e Costa (2012), o BNDES foi criado objetivando promover a expansão da infraestrutura e da indústria moderna, com a canalização de recursos internacionais e nacionais, em longo prazo. Para isso, deveriam contar com corpo funcional capaz de

avaliar projetos com base em técnicas mais avançadas de análise e atuar conforme as melhores práticas de contratação financeira.

Considerando a retomada do crescimento ocorrida em 2004, Filho e Costa (2012, p. 981) observam que “a economia brasileira atravessou o mais longo e intenso ciclo de investimento desde os anos 1970”, com projetos que necessitam de grandes estruturas de financiamento que contem com longos valores e alto prazo de repagamento. Os autores notam que a demanda por recursos do BNDES é tanta no atual período que as fontes próprias de financiamento de longo prazo chegaram ao ponto de esgotamento. Tais fatores explicam o aumento do uso cada vez mais frequente desse tipo de fomento.

Dentre as diversas modalidades oferecidas pelo BNDES, a Instituição Financeira citada opta por produtos específicos que atendem ao seu modelo de negócios, conforme o Quadro 2. Os programas são ajustados de acordo com a política governamental vigente, sofrendo alterações cíclicas de taxa, prazo e equipamentos abrangidos a cada novo Plano Safra.

A cada produto do BNDES aplicam-se linhas de financiamento. Elas seguem as condições do respectivo produto. As linhas se destinam a beneficiários, setores e empreendimentos específicos, e podem trazer regras adicionais, mais adequadas aos objetivos dos clientes (ROSS et al., 2013).

Utilizando o exemplo do Programa Nacional de Fortalecimento da Agricultura Familiar (PRONAF), um dos programas mais populares do BNDES, conforme Heredia e Cintrão (2006) foi criado com o objetivo de manter as pessoas ocupadas nos estabelecimentos familiares, gerar renda para remunerar estes postos de trabalho, agregar novos empregos em atividades agrícolas e não agrícolas nos estabelecimentos ao longo da cadeia produtiva e no mercado local.

Atualmente, segundo Coelho e Negri (2010), o BNDES é o principal instrumento de política industrial do governo brasileiro. Os autores concluíram que os programas impactam positivamente na produtividade do trabalho, do número de empregados e da receita líquida de vendas dos clientes financiados.

**Quadro 2 – Programas de Financiamento do BNDES**

PROGRAMA	DEFINIÇÃO	PRAZO/CARÊNCIA	TAXA BNDES
AGRO	Apoio ao aumento da capacidade de armazenagem das agroindústrias e à aquisição de pulverizadores aéreos agrícolas	Até 12 anos, com até 03 anos de carência	1,5% a.a. (BNDES) + 2,2% a.a. (Agente Financeiro)
FINAME	Financiamento de Máquinas e Equipamentos	Até 10 anos, com até 02 anos de carência	1,5% a.a.
MODERINFRA	Financiamento para o desenvolvimento da agropecuária irrigada sustentável	Até 10 anos, com até 03 anos de carência	7,5% a.a.
MODERFROTA	Financiamento para aquisição de diversos equipamentos agrícolas	Itens novos - 07 anos, com até 11 meses de carência Itens usados - 04 anos, com até 11 meses de carência	De 7,5% a.a. até 10,5% a.a., dependendo do faturamento anual do cliente
PCA	Programa para Construção e Ampliação de Armazéns	Até 15 anos, com até 03 anos de carência	6,5% a.a.
PRONAF	Programa Nacional de Fortalecimento da Agricultura Familiar	Até 10 anos, com até 03 anos de carência	5,5% a.a.
PRONAMP	Financiamento para investimentos dos médios produtores rurais em atividades agropecuárias	Até 08 anos, com até 03 anos de carência	7,5% a.a.

Fonte: BNDES (Adaptado pelo autor)

### 3.2. REFERENCIAL TEÓRICO

De acordo com o dicionário Cambridge English (2018), o crédito possui, em seu sentido econômico, a definição de ser um método de pagamento por bens e serviços de maneira mais tardia, normalmente pagando juros assim como o valor original. No inglês americano, é definido como um método de comprar bens e serviços que permite que você pague por eles no futuro.

Especificamente falando de instituições financeiras, o dicionário Cambridge English (2018) expõe o crédito como sendo o dinheiro emprestado por uma instituição financeira para empresas, governos, pessoas, etc. e a quantidade de risco quando se empresta dinheiro para uma pessoa ou organização específica, baseado em quão provável eles irão pagar de volta. Isto posto, o crédito é considerado como uma maneira de permitir que os clientes possam organizar seus fluxos de caixa e realizar compras de valores mais elevados, acreditando que, com um prazo maior e mais parcelamentos, poderão realizar o pagamento dos bens ou serviços.

Em uma visão mais ampla, o crédito também tem como objetivo possibilitar a melhoria das condições sociais e financeiras do cliente (BANCO POPULAR DO BRASIL, 2018). Para a empresa, o crédito serve como um método de incentivar as vendas e reduzir os estoques. (ROSS et al., 2013). Em países com economias menos desenvolvidas, tal prática torna-se extremamente importante como alternativa de aumento de capacidade de pagamento, esperando que o investimento realizado com o crédito se transforme em retorno positivo para o cliente. De acordo com Tiryaki et al. (2017, p. 3),

O crédito é essencial para o financiamento das atividades produtivas, por parte das firmas, e do consumo, por parte das unidades familiares. Exerce, por isso, papel fundamental tanto para a promoção do crescimento econômico quanto para as flutuações de curto prazo.

Uma das caracterizações mais clássicas referentes à análise de crédito e seu processo decisório vem de Hale (1983), que observa a decisão de crédito como sendo uma reflexão de um julgamento pessoal sobre a habilidade de repagamento de um tomador.

Santos (2010) trata a análise de crédito como uma técnica de previsão, que permite empresa vendedora estimar a capacidade de pagamento em curto prazo do pretendente do crédito. A análise de crédito permite que a empresa vendedora tenha alguma estimativa sobre a situação financeira futura de um pretendente com um cadastro recomendável (COBRA, 1997).

Segundo Silva (2014), a função do crédito é bastante antiga e alguns estudiosos destacam que foram encontradas evidências escritas nas ruínas da antiga Babilônia sobre empréstimos feitos a um fazendeiro, o qual teria se comprometido a pagar os juros e o principal após sua colheita.

Schricket (1997) traz uma das principais definições de análise de crédito dentro de um panorama de incertezas e constantes mudanças e informações superficiais. O autor observa que esta habilidade está ligada a capacidade de considerar situações, e por isso, não é incomum ser difícil se chegar a uma conclusão clara, prática e factível.

O conceito dos “C’s do Crédito” é abordado por Schrickel (1997) e Blatt (1999), sendo estes conceitos considerados como os princípios básicos das análises praticadas por empresas e instituições financeiras.

Inicialmente, busca-se analisar o caráter do cliente, em termos de o quanto disposto está para pagar o crédito no momento de sua concessão. Schrickel (1997) acredita este C ser o principal a ser analisado, independente do valor da operação. Para esta característica, é possível buscar dados dos clientes nos *bureaus* de crédito, obtendo seu histórico de pagamento junto ao Sistema de Informações de Crédito do Banco Central (SCR), bem como suas restrições em exposição na mídia (por exemplo, políticos envolvidos em alguma suspeita de fraude) e em processos abertos nos Tribunais de Justiça.

A Capacidade de pagamento é considerada imprescindível no momento de análise, pois determina, além de sua intenção, a real disponibilidade que o cliente possui para repagar a Instituição Financeira. Nesse ponto, difere a análise de pessoa física e pessoa jurídica, dado que um forte indicativo da capacidade da pessoa física é a Declaração do Imposto de Renda, e da pessoa jurídica são os Balanços Patrimoniais. Blatt (1999) observa a necessidade de, em casos de pessoa jurídica, avaliar também os administradores da empresa, pois eles são os beneficiários finais do crédito e, em caso

de a empresa não estar rentável, eles supostamente irão honrar com os pagamentos da dívida.

Capital aparece como uma terceira característica a ser analisada, sendo definida como a análise do patrimônio de clientes, em termos de sua capacidade de obter maneiras alternativas de disponibilidade de recursos. Da mesma forma que a capacidade de pagamento, os bens de pessoa física são declarados no Imposto de Renda e os de pessoa jurídica encontram-se no Patrimônio Líquido que consta nos Balanços Patrimoniais. Nesse ponto, Blatt (1999) sugere a importância de analisar também o ramo de atividade do cliente, dado que as informações que constam nos demonstrativos financeiros não podem ser analisadas separadamente. Um indicador que é considerado bom para determinada atividade pode não ter a mesma relevância, se comparado a outro mercado. O mais correto, nesses casos, é utilizar-se de algum comparativo entre clientes do mesmo setor.

Outra característica avaliada no momento de concessão de crédito é a situação do macroambiente em que se encontra o cliente, chamada de Condições. Esse fator é relevante especialmente em países de instabilidade econômica, como é o caso do Brasil, em que setores sofrem com quedas inesperadas, como aconteceu com a construção civil em meados de 2015. Em caso de clientes do agronegócio, é relevante conhecer a principal cultura que o cliente trabalha e seu período de colheita, em que irá monetizar a plantação. Nem sempre a inadimplência do cliente virá apenas por suas próprias características, mas cabe ao analista de crédito identificar se o mercado em que o cliente está inserido é favorável para aquele tipo de investimento. Schrickel (1997, p. 53) afirma que

condições dizem a respeito ao micro e macrocenário em que o tomador de empréstimos está inserido. No caso de empresas, tal cenário é o ramo de atividade e a economia como um todo. Vale dizer, é muito importante saber avaliar o momento do empréstimo (timing).

Ao longo do tempo, foram incluídos outros “Cs do Crédito”. Santos (2015) aborda o conceito de Colateral, que é definido pelo valor de garantia que o cliente apresenta na proposta de crédito. De acordo com Ross et al. (2013), as garantias podem ser subdivididas entre reais (hipoteca, penhor, alienação fiduciária de bens móveis e imóveis e caução de títulos) e fidejussórias (aval e a fiança). A maioria das operações

financiadas na Instituição Financeira analisada nesse estudo possuem garantia fidejussória de aval dos sócios, nos casos de pessoa jurídica, e busca contar com o máximo possível de inclusão de garantia real no formato de hipoteca, para casos de pessoa física. Em ambos os casos, a alienação fiduciária do bem que está sendo financiado é requisito mínimo para o crédito, tanto com recursos oriundos do BNDES como para os recursos próprios do banco.

**Quadro 3 – 5 Cs do Crédito**

Caráter	Idoneidade no mercado de crédito (situação na Serasa, Equifax, SPC etc.)
Capacidade	Habilidade / conhecimento em converter negócios em renda
Capital	Situação financeira e capacidade de pagamento
Colateral	Disponibilidade de bens móveis, imóveis e financeiros
Condições	Impacto de fatores externos na geração de fluxos de caixa

Fonte: Santos (2015, p. 44)

Quando se aborda crédito, inicialmente pensa-se no cartão de crédito emitido pelos principais bancos de varejo e o crédito lojista, em geral. Essa é apenas uma das modalidades de crédito existentes no país. De acordo com a Circular 1273 do Plano Contábil das Instituições do Sistema Financeiro Nacional - COSIF,

As operações de crédito distribuem-se segundo as seguintes modalidades: (Circ 1273):

- a) empréstimos - são as operações realizadas sem destinação específica ou vínculo à comprovação da aplicação dos recursos. São exemplos os empréstimos para capital de giro, os empréstimos pessoais e os adiantamentos a depositantes;
- b) títulos descontados - são as operações de desconto de títulos;
- c) financiamentos - são as operações realizadas com destinação específica, vinculadas à comprovação da aplicação dos recursos. São exemplos os financiamentos de parques industriais, máquinas e equipamentos, bens de consumo durável, rurais e imobiliários

No momento de análise de crédito e formulação da decisão de conceder (ou não) o recurso, deve ser levada em conta a possibilidade de fatores internos e externos levarem o cliente a não ser capaz de repagar o que lhe foi recebido. Tal situação é

considerada como “risco de crédito”. O Banco Central do Brasil (2009, p. 1) define risco de crédito como

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

Caouette et al. (2008) afirma que, sempre que algo não é pago imediatamente no momento de compra, existe risco de crédito. Desta forma, o risco de crédito torna-se inerente à operações financeiras que envolvam pagamentos futuros. Se o cliente não tiver condições de honrar com seus pagamentos, irá se tornar inadimplente.

No Brasil, de acordo com pesquisa realizada em Agosto de 2017 pela CNDL em conjunto com o SPC Brasil, aproximadamente 60 milhões de pessoas estão em situação de inadimplência, representando praticamente 40% de todos os consumidores acima de 18 anos no país (CONFEDERAÇÃO NACIONAL DE DIRIGENTES LOJISTAS, 2017). Dessa porção, quase metade dos inadimplentes admite não ter condições de adimplir essa dívida nos próximos três meses, fazendo com que a maturação da carteira inadimplente das instituições seja razoavelmente prolongada.

Para mitigar os riscos de crédito, as instituições financeiras utilizam-se de técnicas variadas que possuem o objetivo de minimizar as perdas resultantes da inadimplência. Essas técnicas vão desde o treinamento de especialistas, até o uso de modelagens estatísticas que lhe confirmam maior robustez e eficácia na análise dos clientes (CAOQUETTE et al., 2008)

Dentre os modelos estatísticos existentes para este propósito, “a mineração de dados surge como uma aplicação específica de algoritmos para a extração de padrões e modelos dos dados” (BERRY; LINOFF, 2004). A mineração de dados aplicada à análise de crédito de uma instituição financeira, nesse sentido, cria um algoritmo que analisa os dados disponíveis referente aos clientes, com o objetivo de identificar um comportamento que permita inferir o nível de risco de crédito envolvido na solicitação de empréstimo ou financiamento.

A mineração de dados é aplicada com diversas técnicas, sendo a árvore de decisão uma das principais e mais populares (WANG et al., 2012). A árvore permite que

seja possível organizar os dados disponíveis de acordo com sucessivas regras de decisão, de maneira que os indivíduos sejam classificados em determinados grupos de similaridade.

Esses grupos são formados de acordo com seus padrões de comportamento dentro da árvore de decisão. Para a instituição financeira, é possível que o modelo estatístico em questão determine o risco de crédito de um cliente e suas chances de ser um bom pagador ou tornar-se inadimplente, conforme supra as regras de decisão.

A árvore de decisão torna-se, nesse estudo, a base da negativa ou aprovação automática para o crédito dos clientes que atendam aos critérios determinados pela instituição financeira, de acordo com as regras selecionadas.

### 3.3. REVISÃO DE ESTUDOS REALIZADOS

O uso de técnicas estatísticas na análise de crédito foi estudado por autores ao longo do tempo. Assim como notado por Hand e Henley (1997), a literatura referente ao assunto da pesquisa é limitada por questões de confidencialidade comercial e possível receio de compartilhamento de vantagem competitiva, por tratar-se de modelos estatísticos que são, em grande parte, responsáveis pela qualidade da carteira de clientes das instituições financeiras.

A maior parte da literatura existente sobre análise de crédito automática preocupa-se em estimar qual método estatístico de mineração de dados deve ser utilizado para construir a decisão de crédito, dentre regressão logística, análise discriminante, árvores de decisão e redes neurais. Hand e Henley (1997) inferem que não existe um método “melhor” que o outro, mas dependem da estrutura dos dados disponíveis. Esta conclusão é similar à que foi inferida por Koh, Tan e Goh (2006) e Crone e Finlay (2012).

Em casos de decisão automática, as árvores de decisão seriam mais adequadas pela facilidade de entendimento de como o modelo chegou naquela decisão específica, e igual facilidade em revisá-lo, quando necessário (HAND; HENLEY, 1997). As redes

neurais seriam mais adequadas em dados não tão estruturados como a vasta base disponível nas instituições financeiras.

A regressão linear foi considerada como o modelo de mineração de dados mais estático (KOH; TAN; GOH, 2006). Em contrapartida, quando aplicada à uma base de dados de um banco chinês (NIE et al, 2011) e um banco egípcio (YAP; ONG; HUSAIN, 2011), ambos estudos afirmam que as redes neurais apresentam um desempenho levemente melhor que a árvore de decisão e a regressão logística. Importante destacar que, quanto maior a amostra da base de dados, maior a assertividade dos modelos (CRONE; FINLAY, 2012).

Faria (2006) utilizou-se de análise fatorial e lógica *fuzzy* para determinar a inadimplência, percebendo que o porte do cliente e o setor de atividade são variáveis que necessitam ser observadas no momento de escolha do modelo estatístico. Complementarmente, Camargos et al. (2010), apesar de utilizarem um modelo estatístico diferente (regressão logística), confirmaram as conclusões de Faria (2006), adicionando ainda variáveis como oportunidades de mercado, tempo de atividade, informatização, bens do avalista, experiência dos sócios, uso de capital de giro e faturamento da empresa.

Em 2006, Pailhé apresenta um estudo aplicado em um banco argentino em que se demarcava um ponto de corte de *score*, e os solicitantes que apresentassem uma pontuação menor do que a zona de corte eram automaticamente negados. Medina (2013) propõe que o modelo de *scoring* também crie um ponto de aprovação automática, baseado na probabilidade de os clientes tornarem-se inadimplentes, deixando apenas os casos “duvidosos” para análise manual. Se a maior parte dos clientes se localiza na faixa duvidosa, a análise ganha custos de transação consideravelmente superiores e o tempo de retorno da análise aumenta da mesma forma.

No Brasil, merece destaque a obra de Minussi, Damacena e Junior (2002), que utilizaram a regressão logística para avaliar o risco de crédito, obtendo uma previsão correta em mais de 94% dos casos. Aprofundando o estudo, Chen e Huang (2011) combinaram os modelos de mineração de dados de redes neurais e árvore de decisão,

chegando à assertividade de 96.5%. De todos os estudos analisados nessa pesquisa, este foi o que obteve o maior nível de assertividade.

Ainda há espaço para evolução na literatura referente à aplicação da metodologia de árvore de decisão em instituições financeiras brasileiras, porém cita-se a pesquisa de Lemos, Steiner e Niévola (2005), que abordou a aplicação desse modelo estatístico em uma agência do Banco do Brasil, e Sousa e Figueiredo (2014), que a aplicaram em uma Cooperativa de crédito, ambas as pesquisas indicando o sucesso dessas técnicas em termos de melhoria de desempenho.

Com os avanços em *machine learning*, Khandani, Kim e Lo (2010) aplicaram algoritmos que permitem prever a inadimplência de 03 a 12 meses antes de ela acontecer, resultando em uma redução de custos de 6% a 23% em perdas. Chen e Huang (2011) reconhecem a importância do *credit scoring*, porém notam que as políticas atuais de crédito que se baseiam apenas nesse modelo não chegam em decisões precisas e negligenciam as mudanças ambientais. A principal vantagem da base utilizando algoritmos de *machine learning* é a adaptação constante do modelo, considerando um cenário econômico que também está em frequentes mudanças.

Seguindo nessa linha, Wang et al. (2012) realizaram um estudo utilizando base de dados reais de duas instituições financeiras que, na árvore de decisão, utilizou classificadores para agrupar os dados baseados em *machine learning*, técnica chamada de método de *bagging* (MACHOVÁ; BARČÁK; BEDNÁR, 2006). O resultado demonstra que essa técnica é a que apresenta os melhores resultados para chegar no *score* de crédito. Abellan e Mantas (2014) tiveram conclusões similares, realizando um estudo experimental.

Sakprasat e Sinclair (2007) abordaram a análise automática de crédito, porém utilizando apenas a metodologia de *credit scoring*. Esse modelo, conforme as limitações citadas pelos autores, não leva em consideração as regras e valores permitidas pelo algoritmo de árvore de decisão, categorizando o risco de crédito em apenas dois níveis: “bom” ou “ruim”. Os casos ruins não seriam negados automaticamente, mas passariam para análise manual, perdendo a vantagem de eficiência da análise automática mais assertiva. Chitra e Subashini (2013), na mesma linha de estudo, aplicaram os modelos estatísticos de mineração de dados na análise automática, preocupando-se também

com mitigar os riscos de fraude no momento da solicitação do crédito, concluindo que, para este fim, a árvore de decisão possui o melhor desempenho.

Em termos de resultados práticos, Curnow et al. (1997) realizaram a análise automática de crédito para clientes considerados *small-ticket* na AT&T Capital Corporation, chegando a um aumento de competitividade, na época, de redução anual de \$3,1 milhões em custos operacionais, aumento de volume de negócios anuais em \$86 milhões e redução de perdas anuais em \$1,1 milhão.

Importante destacar a revisão literária feita por Marqués, García e Sánchez (2013), comprovando a evolução percebida em publicações de 2000 a 2012 e as conclusões aqui discutidas, e inferindo que “o futuro da análise de risco de crédito é uma confiança aprimorada em modelos computadorizados de *credit scoring*.” (MARQUÉS; GARCÍA; SÁNCHEZ, 2013, p. 1395). Os autores ainda observam as limitações existentes nos estudos analisados no período, em conformidade com Hand e Henley (1997), notando a falta de bases de dados de crédito que estejam publicamente disponíveis. A escolha de variáveis foi uma das poucas limitações que recebeu grandes esforços ao longo do período para ser resolvida.

Nota-se também que os artigos foram, em sua maior parte, publicados em periódicos que são ligados à área computacional, e não à área de finanças (*Expert Systems with Applications, International Journal of Scientific & Engineering Research, International Journal of Forecasting, JISTEM - Journal of Information Systems and Technology Management, Knowledge-Based Systems*, entre outros).

## 4. OBJETIVOS

Levando em consideração o panorama contextual e a relevância do assunto abordado, juntamente com a base teórica apresentada, o estudo elenca objetivos a serem atingidos.

### 4.1. OBJETIVO GERAL

Validar a assertividade das variáveis utilizadas no modelo estatístico de análise automática de crédito da instituição financeira abordada no estudo de forma a garantir ganhos de eficiência e redução de inadimplência.

### 4.2. OBJETIVOS ESPECÍFICOS

Para que seja possível realizar o objetivo geral, os seguintes objetivos específicos são endereçados para este estudo:

- a) Realizar um levantamento das propostas de crédito analisadas pela árvore de decisão da instituição financeira estudada, em um determinado período histórico;
- b) Delimitar as principais variáveis utilizadas para agrupar esses dados em sua mineração, baseado no impacto que elas possuem na árvore de decisão;
- c) Analisar o desempenho histórico dos clientes baseado nas variáveis definidas, de maneira a comprovar eficácia do modelo.

## 5. INSTITUIÇÃO FINANCEIRA ALVO DO ESTUDO

A instituição financeira que forneceu a base de clientes para que fosse possível a realização deste estudo é subsidiária de multinacional e está estabelecida no mercado brasileiro desde 2002. As operações de crédito foram sempre realizadas em linhas de financiamento atreladas a algum bem, com o financiamento de capital de giro não estando em seu escopo de produtos.

Com escritórios presentes em duas das principais capitais brasileiras, a instituição é considerada uma empresa de médio porte por seu quadro de funcionários de acordo com o SEBRAE (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2013).

O segmento de financiamentos com recursos oriundos do BNDES ocupa a maior parte da carteira de clientes da instituição desde sua fundação, porém está proposto nos objetivos estratégicos da instituição que sejam disponibilizadas mais linhas de recursos próprios em formato de CDC (Crédito Direto ao Consumidor), em que os juros vão aumentando conforme o prazo da operação.

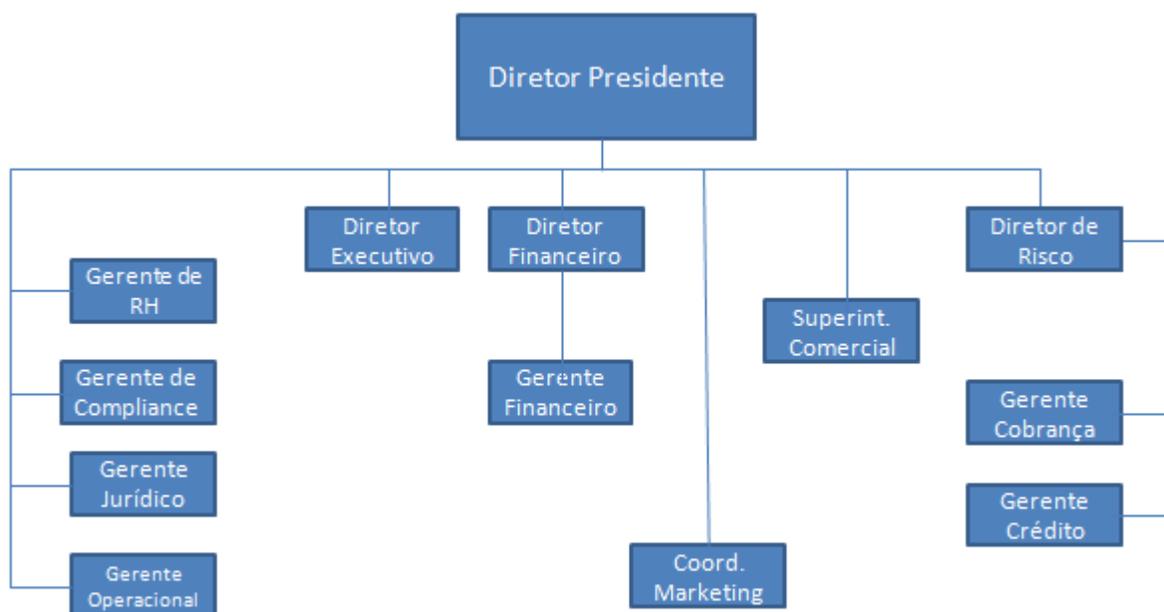
A margem de lucro para os bancos das operações com recursos do BNDES estão cada vez menos competitivas, pois são fixadas pelo governo para todas as instituições que as utilizam. Desta forma, o que diferencia a escolha de um determinado banco para financiar com condições similares é a qualidade de atendimento e a velocidade em que as propostas são analisadas e o recurso é desembolsado.

Para atender melhor os clientes, a instituição financeira atua de maneira próxima aos responsáveis por fazer com que os beneficiários tenham acesso ao financiamento: os fabricantes dos bens e as revendas dos equipamentos. Dessa maneira, o risco que o banco assume é dividido entre os agentes envolvidos, e é possível aprovar o crédito em condições mais atrativas.

A estrutura interna divide-se nos dois escritórios de maneira que uma trabalha mais com os casos do segmento agrícola e a outra trabalha com as unidades de negócio não-agrícola. Essa divisão foi realizada considerando a proximidade física com os clientes dos outros segmentos, dado que há representantes comerciais externos

espalhados por todas as regiões do Brasil para tratar dos clientes agrícolas. O organograma é apresentado conforme a Figura 1.

**Figura 1 – Organograma**



Fonte: Elaborada pelo autor

As operações de crédito que entram na “esteira” do departamento para análise manual são divididas em dois segmentos: Agrícola e Não-Agrícola. Há analistas Júnior, Pleno, Sênior e um Coordenador em cada um dos segmentos, com um Gerente Executivo geral na área, que responde ao Diretor de Risco da instituição.

A alçada decisória depende do valor e do caráter de exceção da proposta de crédito. Quando um caso está fora da política de Crédito, a operação automaticamente é colocada na alçada decisória superior a que seria se não existisse a exceção.

Conforme o Quadro 4, o país pode aprovar casos até R\$ 12.500.000,00 (doze milhões e quinhentos mil reais). A alçada superior à permitida no país torna-se decidida pelo Diretor de Risco regional.

**Quadro 4 – Alçadas de Crédito**

<b>Alçada</b>	<b>Valor Máximo</b>
1 – Analista de Crédito Júnior	R\$ 350.000,00
2 – Analista de Crédito Pleno	R\$ 500.000,00
3 – Analista de Crédito Sênior	R\$ 1.000.000,00
4 – Especialista de Crédito	R\$ 1.500.000,00
5 – Coordenador de Crédito	R\$ 2.000.000,00
6 – Gerente de Crédito ou 2 Coordenadores de Crédito	R\$ 3.125.000,00
7 – Diretor de Risco ou Gerente de Crédito + Coordenador de Crédito	R\$ 6.000.000,00
8 – Diretor de Risco + Gerente de Crédito	R\$ 6.500.000,00
9 – Comitê Local de Crédito	R\$ 12.500.000,00

Fonte: Elaborado pelo autor

## 6. PROCEDIMENTOS METODOLÓGICOS

Para que seja possível determinar se as variáveis do modelo estatístico estão adequadas à realidade de mercado em um nível satisfatório, inicialmente será preciso delimitar quais critérios são utilizados atualmente para minerar os dados dos clientes e de quais os pesos que são aplicados no modelo estatístico. É imprescindível coletar a quantidade de operações que foram aprovadas com base na árvore de decisão, que resulta em um retorno imediato das propostas de crédito para os clientes.

Levando em consideração a indispensabilidade de medir a eficiência da análise de crédito automática da instituição financeira abordada neste estudo, deve-se entender como está ocorrendo, ao longo dos anos, o desempenho de pagamento dos clientes-alvo da pesquisa. Baseando-se em dados bancários e informações sobre o mercado, é possível estimar e simular os fluxos de caixa futuros dos clientes. Criar uma base para armazenar e analisar os dados bancários facilita o processo de estimação e simulação de fluxos de caixa futuros dos clientes, beneficiando as duas partes envolvidas nas transações (PERLIN; RIGHI; FILOMENA, 2016).

Tendo em mente que o estudo aborda uma validação de um modelo estatístico, o acesso aos sistemas de armazenamento de dados da instituição financeira estudada será crucial, para que sejam obtidos os detalhes da análise de cada cliente, como esses dados se comportaram dentro da árvore de decisão e como ocorreram ou estão ocorrendo os pagamentos dos clientes a cada parcela.

A partir do acesso às bases de informações da organização, será necessário estruturar os dados e filtrar apenas os clientes que foram aprovados de maneira automática, construindo uma base de dados à parte. A coleta deve ser feita levando em consideração o início da aplicação da árvore de decisão atual. Os critérios encontrados devem ser combinados de acordo com certas faixas de agrupamento, que correspondam às utilizadas na mineração de dados do modelo estatístico, separando as variáveis que são vistas de maneira prioritária na regra de decisão.

Após a listagem de clientes que tiveram aprovação automática, seus dados, o valor de suas operações, e seu desempenho de pagamento, será necessário criar uma

segunda base de dados, selecionando os clientes que, ao longo do financiamento, estão em situação de *default*. Conforme recomendado pelas abordagens de crédito internas (IRB) de Basileia II (SCHECHTMAN et al., 2004), a instituição financeira analisada considera como *default* os atrasos de pagamento que ultrapassam 90 dias após a data de vencimento da parcela.

Chegando à amostra de clientes inadimplentes, será possível realizar uma regressão linear generalizada utilizando a função Probit, considerando que a variável de *default* possui valores binários, para identificar as variáveis que possuíram maior peso em sua aprovação automática e observar como elas desempenharam em clientes que não apresentaram *default*, possuindo um bom histórico de pagamento, de forma a identificar sua relevância. Essa regressão será rodada com auxílio do *software* “R”, de forma que seja possível chegar às variáveis mais significativas no que tange ao *default*. Da mesma maneira, será possível observar se algum critério está recebendo prioridade inadequada na regra da árvore de decisão, podendo alterá-la para garantir uma assertividade ainda maior.

Para que seja possível realizar a regressão que determinará a assertividade das variáveis que estão atualmente incluídas nas regras de decisão automática, é utilizado um modelo linear generalizado (GLM), adequado quando há variáveis binárias, dado que a distribuição dos resíduos seria afetada se fosse utilizado um modelo linear (PERLIN, 2018). o modelo GLM pode ser expressado da seguinte forma:

$$E(y_t) = G\left(\alpha + \sum_{i=1}^N \beta_i x_{i,t}\right)$$

Wooldridge (2013) nota que o modelo Probit é uma das funções de ligação não lineares mais utilizadas em modelos de resposta binária, e generaliza a classe de modelos de resposta binária da seguinte forma:

$$P(y = 1 \parallel x) = G(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)$$

Nessas equações, “ $y$ ” é a variável binária dependente; “ $x$ ” é utilizado para denotar o vetor de variáveis explicativas; “ $G$ ” é a função que assume os valores estritamente entre “zero” e “um” ( $0 < G(z) < 1$ ), “ $\alpha$ ” é a intersecção, “ $\beta$ ” é o vetor de coeficientes da regressão. Esse modelo é utilizado principalmente pelo fato de as probabilidades ajustadas poderem ser menores que “zero” ou maiores que “um”, e o efeito parcial de qualquer variável explicativa é constante (WOOLDRIDGE, 2013).

O modelo GLM utilizado nesse estudo apresenta estimadores que foram obtidos pelo método da máxima verossimilhança, em que os parâmetros desconhecidos de modelos econométricos podem ser estimados, com a realização de testes de hipóteses sobre restrições lineares e não-lineares ao vetor dos parâmetros (PORTUGAL, 1995).

Trabalhando com a hipótese nula de que o default não é explicado pelas variáveis selecionadas para realizar a regressão, “utilizam-se os resultados para decidir no sentido de a rejeitar (assumindo que é falsa) ou não rejeitar (assumindo que é verdadeira)” (LOUREIRO; GAMEIRO, 2011, p. 154). Desta forma, é esperado que o resultado da regressão linear generalizada, considerando um nível de significância como 5%, encontre variáveis que apresentem evidências de que é possível rejeitar a hipótese nula de que tais variáveis não possuem relevância na variável dependente, ou seja, ocorrência de *default*.

Com as bases de dados abordando as informações necessárias para o desenvolvimento do estudo, será realizado um *backtesting*, em que as previsões de medição de risco do modelo serão comparadas com as reais perdas financeiras que ocorreram ao longo do tempo, com um teste estatístico de correção do procedimento (NOLDE; ZIEGEL, 2017).

## 7. CRITÉRIOS DA DECISÃO AUTOMÁTICA

Para que a decisão automática ocorra na instituição financeira alvo deste estudo, é necessário que certos critérios sejam atendidos. Conforme o cadastro dos clientes, realizado por meio do representante comercial (seja ele diretamente do fabricante ou da revendedora), são fornecidas informações para que seja possível alimentar a base de dados do sistema e o cliente passa pela árvore de decisão.

No momento de cadastramento, é solicitado ao cliente Pessoa Física que envie o termo de autorização para consulta ao banco de dados do Banco Central do Brasil e o Imposto de Renda, juntamente com documentos que comprovem os bens do cliente. Para Pessoa Jurídica, é solicitado o mesmo termo de autorização de consulta, juntamente com os Balanços Patrimoniais (e balancete do ano atual) das empresas.

Considerando que muitos dos dados são cadastrados manualmente, há diversos casos de inconsistência das informações que o cliente manifesta ter, e do que ele realmente possui. Isto ocorre especialmente em casos de clientes agrícolas informais de pequeno porte, não declarando imposto de renda, ou empresas que optam pelo Simples Nacional, não necessitando apresentar balanços. Nesses casos, a decisão automática não se aplica e o caso é avaliado manualmente pelos analistas de crédito.

Além do citado acima, existem outros critérios que previnem a negação automática, como o histórico sem dívidas vencidas com outras instituições, de acordo com o sistema de dados do Banco Central. Para assegurar a aprovação automática, é necessário que o cliente atenda os seguintes critérios:

**Quadro 5 – Critérios-base de decisão automática**

<b>Variável</b>	<b>Limite Aceitável</b>
Risco da proposta	Até R\$ 200.000,00 (duzentos mil reais)
Restrições conforme consultorias de crédito	Zero
Atrasos de pagamento	45 dias
Idade do proponente	Maior que 25 anos e até 75 anos
Tipo de equipamento	Apenas novos, não usados

Prazo e carência	Dentro da política, sem exceções
Histórico de renegociação com a instituição	Zero
Classificação do cliente	Até "F"

Fonte: Base de dados da Instituição. Elaborado pelo autor

Todos esses critérios são utilizados para que seja possível criar uma pontuação de crédito para o cliente, que é a base para a decisão automática. Essa pontuação, além de embasar os casos avaliados automaticamente, também auxilia a guiar os analistas na decisão manual, dando um indicativo da qualidade do cliente e buscando ser o mais assertivo possível. Em um segundo momento, são analisados os dados patrimoniais e financeiros dos proponentes.

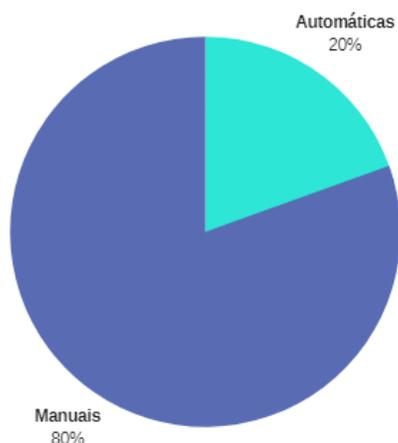
## 8. ANÁLISE DOS DADOS DOS CLIENTES

### 8.1. ANÁLISE DESCRITIVA DOS CASOS DE DEFAULT

Para a satisfação dos objetivos propostos por este estudo, inicialmente buscou-se organizar a base de dados dos clientes do banco que é alvo desta análise, de maneira que fosse possível separar os casos que foram aprovados automaticamente pela árvore de decisão atual. Até setembro de 2018, conforme é possível observar na Figura 2, as decisões automáticas representavam 20% de todas as análises que são realizadas no setor de Crédito. Os clientes da população de aprovações automáticas são, em sua totalidade, formados por trabalhadores da atividade agrícola, tomando os recursos financeiros na figura de Pessoa Física.

A instituição ainda não permite que seja realizada decisão automática para solicitações de crédito de Pessoa Jurídica. Tal fato ocorre devido à política interna do banco e a demanda do Banco Central, em que é sempre necessário identificar o beneficiário final do financiamento (em termos de Pessoa Física) de acordo com o quadro societário da empresa, e isso não é possível em casos que o quadro societário é formado por outras Pessoas Jurídicas (subsidiárias e empresas de participação, por exemplo). Há projetos em andamento para encontrar maneiras de realizar essa identificação e iniciar a oferecer decisão automática para clientes de Pessoa Jurídica, que são os principais proponentes do ramo Não-Agrícola. Todos os casos avaliados são ligados à financiamentos do BNDES para proponentes do ramo Agrícola, cuja principal atividade está ligada a produção rural.

Considerando a complexidade dos casos de valores mais expressivos, a aprovação automática, no momento, é permitida para operações de crédito no valor limite de R\$ 200.000,00 (duzentos mil reais), o que abrange uma gama relevante de propostas, dado que o ticket médio das operações de crédito da instituição financeira é de R\$ 130.000,00 (cento e trinta mil reais). Conforme verificada sua eficácia, é possível que sejam permitidas para valores maiores, de acordo com os objetivos da instituição de chegar a uma porcentagem de 50% de decisões automáticas até o final de 2018.

**Figura 2 – Decisões Automáticas**

Fonte: Elaborada pelo autor

A partir da base de dados de clientes filtrada em decisões automáticas, foram selecionadas as aprovações que ocorreram conforme os critérios atuais levados em consideração na análise automática, inicialmente aplicada em 2015 e sendo revalidada em 2018. Foi analisado o desempenho de pagamento dos clientes que estão inclusos nessa base, levando em consideração o conceito de *default* conforme recomendado pelas notações de crédito internas (IRB) de Basileia II (SCHECHTMAN et al., 2004): atrasos de pagamento que ultrapassam 90 dias após a data de vencimento da parcela.

Em compatibilidade com a característica do perfil inadimplente do brasileiro, conforme o que é apresentado na Tabela 1, é possível observar que a distribuição é dividida de maneira igualitária entre as faixas de atraso. A maior concentração de casos (29%) está alocada na faixa de atrasos de 91 a 120 dias, em que é mais provável em que ainda se possa existir recuperação. Por outro lado, a faixa acima de 360 dias é a maior probabilidade de perda real, em que os valores possivelmente serão baixados a prejuízo.

Levando em consideração que a decisão automática começou a ser colocada em prática em 2015 e o prazo de carência para operações que são financiadas via

recursos do BNDES costuma ser de 12 a 18 meses, era esperado que a quantidade de casos de *default* em 2015 fossem menores do que os outros anos.

**Tabela 1 – Distribuição de casos de *Default* e regularização por faixa de atraso**

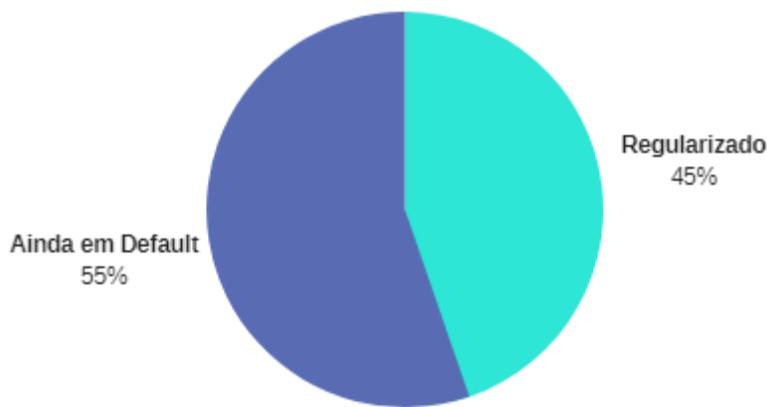
<b>Faixa de Atraso</b>	<b>Número de Casos de <i>Default</i></b>	<b>% de Casos Regularizados</b>
De 91 a 120 dias	11	54%
De 121 a 150 dias	8	25%
De 151 a 180 dias	2	50%
De 181 a 360 dias	9	78%
Acima de 360 dias	8	13%

Fonte: Base de dados da Instituição. Elaborado pelo autor

A faixa de atraso que apresentou maior porcentagem de recuperação de crédito foi a faixa de 181 a 360 dias, em que se pode esperar que os clientes tiveram efetividade em sua reestruturação para angariar recursos e colocar suas obrigações em dia com a instituição financeira. Isso é possível quando a inadimplência dos clientes foi causada por fatores que ocorreram em certo período (como, por exemplo, uma safra ruim) e que é possível a recuperação de sua situação financeira. A baixa regularização de casos em atraso acima de 360 dias está em compatibilidade com a menor chance de recuperação a partir desse período, em que os valores são usualmente notificados ao Banco Central como baixados à prejuízo.

De acordo com a Figura 3, pode-se observar que, em 17 casos (45%), apesar de ainda existirem situações desfavoráveis com outras instituições e credores, os clientes utilizaram seus recursos para regularizar as pendências de pagamento com a instituição financeira analisada neste estudo. Tal situação ocorre normalmente devido à boas negociações realizadas pelo setor de Cobrança da instituição, levando em consideração que certas dívidas possuem taxas maiores que as outras.

Figura 3 – Regularização dos Casos



Fonte: Base de dados da Instituição. Elaborado pelo autor

De acordo com a Tabela 3, é possível ver que a maior concentração de *defaults*, tanto em ocorrências quanto em valor financiado, ocorreu de no ano de 2017. Como 2018 apresenta resultados parciais, considerando apenas até o mês Outubro, é possível que ainda supere 2017 nesses quesitos.

Tabela 2 – Distribuição de *Defaults* por ano

Ano do <i>Default</i>	Quantidade de Ocorrências	Total Financiado em <i>Default</i>
2015	1	R\$ 99.000,00
2016	5	R\$ 437.380,00
2017	18	R\$ 1.385.010,00
2018	14	R\$ 1.093.042,30

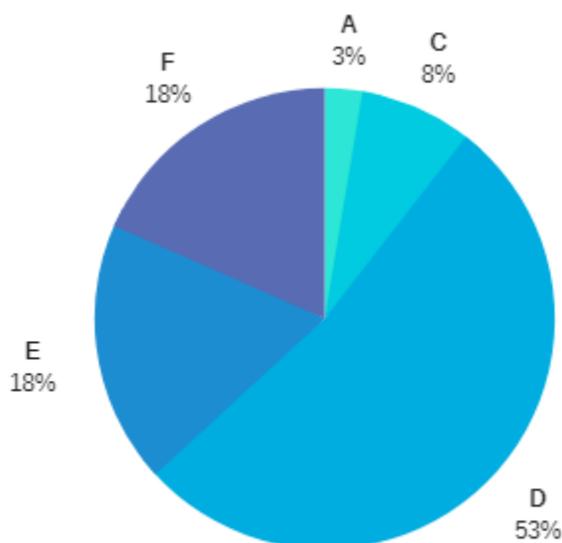
Fonte: Base de dados da Instituição. Elaborado pelo autor

Os clientes da instituição financeira deste estudo são ranqueados por uma pontuação de crédito de acordo com a árvore de decisão e a metodologia de *scorecard*, sendo classificados de “A” a “J”, em que “A” significa a maior pontuação possível e “J” a

piores pontuação possível, conforme a Figura 4. Todos os clientes que apresentaram *default* possuíam, no momento da aprovação automática, um Rating considerado satisfatório (até a classificação “F”), e é possível observar que a principal concentração encontra-se no Rating “D” (20 dos 38 casos).

A disponibilidade de o banco assumir risco e ceder crédito ao cliente está diretamente relacionada à sua pontuação de crédito, com cada classificação possuindo um nível diferente. Busca-se assumir mais riscos com clientes “A”, e apenas em casos de exceção justificável assumir risco com os clientes “J”. No panorama brasileiro, o banco classifica como boa a classificação de “A” a “E”, sendo que o comportamento dos clientes “F” e o risco atrelado a eles acaba ainda sendo indefinido precisamente.

**Figura 4 – Classificação dos clientes**



Fonte: Base de dados da Instituição. Elaborado pelo autor

Nos 38 casos analisados nesse capítulo do estudo, foram avaliadas as condições dos clientes *a priori*, no momento que a decisão automática ocorreu, e como as situações se desenrolaram ao passar do tempo, como tentativa de observar semelhanças que indiquem o desempenho de alguma variável que poderia ter sido determinante e evitar a aprovação. A quantidade de clientes que apresentaram

condições desfavoráveis e o tipo dessa condição é apresentado na Tabela 3. Os valores não foram divulgados em conformidade com o teor sigiloso dos dados.

**Tabela 3 – Situação Financeira Atual dos Clientes**

<b>Categoria</b>	<b>Quantidade de Ocorrências</b>
Pendências Financeiras	21
Restrições Financeiras	15
Ações Judiciais	09
Títulos Protestados	11
Cheques sem Fundo	17

Fonte: Serasa (2018), em acesso concedido para a instituição. Elaborado pelo autor

Dos 38 casos de *default*, 11 (29%) não apresentam nenhuma situação de inadimplência com outros credores além da instituição financeira deste estudo. Os outros 27 casos possuem mais de uma categoria de condição desfavorável e pendências no mercado financeiro. No momento de aprovação, nenhuma dessas restrições estava vigente para os proponentes, sendo possível inferir que a situação financeira deteriorou-se no período de tempo desde a aprovação do crédito.

Na análise de crédito manual, sempre é observado o comportamento do endividamento bancário dos proponentes, motivo pelo qual é solicitado o termo de autorização para consultas ao Banco Central como documento obrigatório para solicitação das propostas de crédito. Nesse momento da análise, é observado o valor de endividamento já existente com outras instituições, seu desempenho de pagamento (dívidas vencidas ou baixadas à prejuízo) e suas modalidades de dívida, sejam elas empréstimos ou financiamentos. Conforme é possível observar na Tabela 4, 11 dos 14 casos apresentavam particularidades no momento da aprovação automática, as quais normalmente são questionadas ao longo da análise manual.

Tabela 4 – Endividamento dos Proponentes

Quantidade de Casos	Modalidades de Endividamento
8	Desconto de Títulos
24	Conta Garantida e Cheque Especial
22	Parcelamento de Cartão de Crédito
38	Financiamentos

Fonte: Base de dados da Instituição. Elaborado pelo autor

Com exceção da modalidade de Financiamento, que aparece em 100% dos casos de default, as outras modalidades de dívida citadas na tabela acima são consideradas como um indício negativo de Risco de Crédito, considerando que são linhas com taxas mais altas, que os proponentes solicitam tipicamente quando já se encontram em dificuldades financeiras. Isso pode ser verificado especialmente no P-valor da modalidade de Desconto de Títulos, descrito posteriormente na Tabela 7, em que o impacto na ocorrência de *default* é confirmado. Esses critérios poderiam ser abordados de maneira mais restritiva na árvore de decisão desses casos.

Considerando que a totalidade dos casos é do ramo Agrícola, a análise de crédito é baseada nas propriedades que os clientes possuem (bens) e receitas advindas da atividade rural. As características dos proponentes nesses critérios são na Tabela 5.

Tabela 5 – Propriedades e Receitas x Valor do Financiamento

Indicadores Patrimoniais	Quantidade de Clientes (%)
Propriedade de Terras	67%
Arrendamentos	31%
Cobertura de Renda do Endividamento Total	21%
Cobertura de Renda do Valor Financiado	53%

Fonte: Base de dados da Instituição. Elaborado pelo autor

É possível observar a maioria dos clientes possuem propriedade de terras, não necessitando arcar com o custo recorrente dos arrendamentos. Apesar da cobertura de renda ser insuficiente em grande parte dos casos (79% para endividamento total e 47% para valor financiado), a cobertura poderia considerar patrimônio oferecido como garantia hipotecária, o que não foi acessado nos casos desses clientes – nenhum apresentou bens como garantia.

Ainda observando a Tabela 5, é possível observar que apenas 53% dos casos possuíam receita suficiente para cobrir o valor de financiamento proposto, sem considerar a quantidade de endividamento que o cliente já possui. Tal número baixa para 21% de casos quando se considera, além do valor financiado, as dívidas já existentes. Dados de fluxo de caixa e cobertura de serviço da dívida não são checados atualmente na árvore de decisão.

Conforme a regra de declaração do Imposto de Renda para 2015 (PORTAL BRASIL, 2015), era necessário realizar a declaração quem tivesse a posse ou a propriedade, em 31 de dezembro de 2014, de bens ou direitos, inclusive terra nua, de valor total superior a R\$ 300,000.00 (trezentos mil reais) e quem teve, no ano anterior, receita bruta em valor superior a R\$ 134.082,75 (cento e trinta e quatro mil, oitenta e dois reais e setenta e cinco centavos) oriunda de atividade rural. Para os declarantes de 2016, esse valor subiu para receita superior à R\$ 140.619,55 (cento e quarenta mil, seiscentos e dezenove reais e cinquenta e cinco centavos) (SUBSECRETARIA DE ARRECAÇÃO E ATENDIMENTO, 2016), e em 2017 para R\$ 142.798,50 (cento e quarenta e dois mil, setecentos e noventa e oito reais e cinquenta centavos) (PORTAL BRASIL, 2017). Na população dos casos de default, 34% dos clientes não realizam a declaração, característica típica de produtores menores e informais. Dos 66% clientes que declaram, foi detectado casos de divergência entre a receita declarada e a receita calculada a partir do cadastro do cliente em 54% dos casos. Isso se torna um indicativo de comportamento, levando em conta que se declara menos do que se realmente possui a fim de redução em imposto de renda.

Em todos os casos, é projetada a receita agrícola do cliente de acordo com o tipo de atividade rural em conjunto com a quantidade de hectares que o proponente

comprova possuir, sendo utilizado o que for maior (receita calculada x receita declarada). Foram identificadas rendas advindas além da produção agrícola, em formato de participação societária (dividendos e juros sobre o capital próprio) em 34% dos casos que entraram em default. Essa receita é adicionada à receita projetada dos clientes.

Como esperado, levando em consideração os fatores estatísticos envolvidos, nota-se a eficácia da árvore de decisão automática nos casos de aprovação, considerando o nível baixo de inadimplência, se comparado ao nível das decisões manuais de crédito. No período avaliado, com solicitações de crédito que foram aprovadas de Janeiro de 2015 a Setembro de 2018, a inadimplência apresenta-se de acordo com a Tabela 6.

**Tabela 6 – Inadimplência da decisão automática x decisão manual**

<b>Decisão</b>	<b>Em inadimplência</b>	<b>Total de Casos</b>	<b>%</b>
Automática	38	3.666	1,04
Manual	437	22.629	1,93

Fonte: Base de dados da Instituição. Elaborado pelo autor

## 8.2. APLICAÇÃO DO MODELO LINEAR GENERALIZADO

Após coletados os dados dos clientes que tiveram sua aprovação automática de Janeiro de 2015 a Outubro de 2018, foi necessário elencar as variáveis que seriam utilizadas no modelo, levando em consideração as principais informações que baseiam as regras da árvore de decisão. Conforme é possível observar no Quadro 6, foram encontradas 12 variáveis possíveis de serem determinantes da variável dependente, ou seja, a existência de *default*.

**Quadro 6 – Variáveis Utilizadas na Aplicação do Modelo Linear Generalizado**

<b>Variável</b>	<b>Tipo</b>	<b>Significado</b>
<i>Default</i>	Binária	Ocorrência de atrasos superiores a 90 dias
Orientação Sexual	2 fatores	Feminino ou Masculino
Alavancagem Financeira	Binária	Existência de crescimento rápido e expressivo da dívida do cliente
Linhas de Endividamento	4 fatores	Existência de concentração nas modalidades de dívida de Empréstimos (Conta Garantida e Cheque Especial), Financiamentos, Títulos Descontados e Parcelamento de Cartão de Crédito
Propriedade de Terra	Binária	Comprovação de terreno próprio para exercer a atividade rural
Cobertura do Endividamento	Binária	Razão (Receita/Endividamento) superior a 1,0; identificando capacidade de pagamento das obrigações a partir da renda
Cobertura do Valor Financiado Solicitado	Binária	Razão (Receita/Valor do Financiamento) superior a 1,0; identificando capacidade de pagamento do valor proposto a partir da renda

Região	5 fatores	Local onde ocorre a atividade do cliente: Norte, Nordeste, Centro-Oeste, Sudeste e Sul
Classificação do Revendedor	5 fatores	Nota que o revendedor recebe anualmente, baseando-se no portfólio de clientes que ele traz para a instituição financeira: A, B, C ou D
Idade	7 fatores	Idade do comprador: de 25 a 34 anos, de 35 a 44 anos, de 45 a 54 anos, de 55 a 64 anos, 65 a 74 anos
Comprovação de Receita	Binária	Existência de receita agrícola declarada em Imposto de Renda
Divergência de Receita	Binária	Quando a receita que o cliente declara em Imposto de Renda é menor do que a receita que é cadastrada à ele pela revenda
Participação Societária	Binária	Casos em que o cliente possui receita advinda de participação em sociedades, além da receita rural em Pessoa Física

Fonte: Elaborado pelo autor

As variáveis apresentadas no Quadro 6 foram aplicadas na função *probit* por meio do *software* “R”, de maneira que fossem obtidos os resultados da regressão com valor estimado, erro padrão, valor Z e probabilidade de significância (Valor-P). Assume-se por hipótese nula que tais variáveis não são relevantes para a variável dependente, considerando um nível de significância de 5%.

A regressão encontrada pode ser descrita na seguinte fórmula:

$$\begin{aligned}
 P(y_{\text{default}} = 1 | x) = & G(\beta_0 + \beta_1 x_{\text{sexo}} + \beta_2 x_{\text{idade}} + \beta_3 x_{\text{endividamento}} + \beta_4 x_{\text{propr.terras}} + \beta_5 x_{\text{cob.endividamento}} \\
 & + \beta_6 x_{\text{cob.valorfin}} + \beta_7 x_{\text{regiao}} + \beta_8 x_{\text{clas.revenda}} + \beta_8 x_{\text{clas.revenda}} + \beta_9 x_{\text{alavancagem}} + \beta_{10} x_{\text{compr.receita}} + \beta_{10} x_{\text{compr.receita}} \\
 & + \beta_{11} x_{\text{div.receita}} + \beta_{12} x_{\text{part.societaria}}
 \end{aligned}$$

De acordo com Malacrida e Yamamoto (2006), se o Valor-P calculado for menor que o nível de significância escolhido (5%), a hipótese nula é rejeitada. As variáveis que possuíram probabilidade de significância abaixo ou próxima desse valor, ou seja, evidenciando que é possível rejeitar a hipótese nula, são apresentadas na Tabela 7, ordenadas do menor Valor-P ao maior. Tais variáveis apresentam relevância perante a variável dependente, ou seja, existência de *default*. O resultado completo da regressão pode ser encontrado no Anexo I.

Considerando que o cadastramento da proposta de financiamento dos clientes é feita pelo revendedor, juntamente com o cadastro dos dados de bens, direitos, receitas e obrigações dos clientes, era esperado que o impacto da classificação da variável de “Classificação do Revendedor” seja determinante. Caso o cadastro não seja realizado corretamente, a receita do cliente e seus bens podem ser menores do que o estimado, fazendo com que sua capacidade de pagamento seja afetada, levando ao possível *default*. Levando em conta que a decisão automática obtém um retorno instantâneo ao cliente, é possível que a revenda cadastre as informações do cliente de maneira enviesada, tentando impedir que a proposta seja direcionada para análise manual interna.

Considerando que, na decisão automática, não há uma revisão manual de cadastro por parte dos analistas de crédito, detectando e corrigindo possíveis erros, é imprescindível que a revenda possua uma boa qualificação. Isso é demonstrado por meio do resultado da regressão (Tabela 7), em que as classificações “B” e “C” obtiveram Valor-P abaixo do nível de significância.

Tabela 7 – Resultado da Regressão

Variável	Coefficiente	Erro-Padrão	Teste Z	Valor-P
Classificação do Revendedor - B	1,7839	0,4114	4,336	0,00000145
Linhas de Endividamento – Títulos Descontados	2,0844	0,4953	4,208	0,00000258
Classificação do Revendedor – C	1,3758	0,5363	2,565	0,0103
Cobertura do Valor Financiado Solicitado	-0,9080	0,3921	-2,316	0,0206
Propriedade de Terra	-0,6687	0,3792	-1,764	0,0778

Fonte: *Software R*. Elaborado pelo autor

Não seria possível propor que todas as propostas de crédito enviadas por revendas que não possuam classificação “A” sejam enviadas para análise interna, considerando a quantidade de casos em que isso ocorre (16% do total de operações de crédito); perder-se-iam os ganhos de tempo e eficiência que a análise automática propõe. Desta forma, é sugerido que existam ações punitivas para as revendas de classificações “B” e “C” que apresentarem casos de inadimplência (por exemplo, aumento de taxas), e ações de treinamento para realização do cadastro das propostas de crédito.

É interessante notar que a classificação “B” obteve menor Valor-P do que a classificação “C” que, racionalmente, deveria possuir um valor maior. Tal característica

pode ser explicada pelo impacto que a classificação traz nas vendas, em que, de acordo com seu *ranking*, há taxas maiores incorridas no momento de solicitação das propostas. Isso pode fazer com que exista maior preocupação nas vendas “C” de tentar melhorar sua classificação na próxima revisão de *ranking*, realizando cadastros de maneira mais correta.

As linhas de títulos descontados são definidas como concessão de um empréstimo mediante a garantia de um título representativo de um crédito futuro, como uma forma de antecipar o recebimento do crédito (NETO, 2011). Essa linha, quando tomada na figura de Pessoa Física, aplica-se usualmente em agricultores que recebem ou realizam as vendas a prazo, e é utilizada quando é necessário antecipar o recebimento dessas vendas.

A taxa cobrada na linha de desconto de títulos faz com que o trabalhador receba menos, porém antecipadamente. Essas taxas variam de 1,4% ao mês a 3,47% (BANCO CENTRAL DO BRASIL, 2018) nos principais bancos comerciais nacionais, enquanto a linha de empréstimo de capital de giro pode ser captada por 0,62% ao mês nos mesmos bancos. Racionalmente, esse endividamento é tomado quando o cliente não possui fluxo de caixa suficiente para arcar com suas obrigações atuais e necessita antecipar seus recebimentos, logo, já está passando por situações financeiras desfavoráveis. É proposto que clientes que possuam concentração nessas linhas entrem para análise interna, de forma que o analista possa observar o desempenho dessa dívida (em termos de prazo, taxas, indexadores, distribuição entre longo e curto prazo) perante as outras linhas existentes.

A variável de “Cobertura do Valor Financiado Solicitado” compreende a razão entre a receita anual cadastrada para o cliente e o valor da solicitação de financiamento. Essa variável não pretende identificar acuradamente o fluxo de caixa, pois seria necessário ter acesso a todas as obrigações do cliente, seus prazos de pagamento e o incremento da receita a partir da nova aquisição com o financiamento. Por outro lado, a variável objetiva estimar minimamente se o cliente teria capacidade de arcar com o valor de financiamento apenas com sua receita atual, não existindo qualquer outra dívida em suas obrigações.

Com exceção de casos que entram em *default* por mau comportamento dos clientes, ou seja, casos em que eles possuem condições de pagar, mas não o fazem, é possível inferir que o cliente deixe de pagar a instituição financeira porque não possui recursos que permitam que essa dívida seja honrada. Tal situação ocorre especialmente no segmento agrícola, em que a receita dos clientes é altamente dependente do sucesso da safra e da venda dos insumos obtidos com a colheita. A tecnologia busca mitigar cada vez mais os efeitos de condições climáticas, que não podem ser previstas pelo cliente, porém ainda é necessário que exista certa folga de caixa para produtores agrícolas conseguirem arcar com situações atípicas sem prejudicar a continuidade do seu negócio.

A propriedade de terras torna-se fator determinante na mesma linha de raciocínio da estimação de fluxo de caixa do cliente. De acordo com o Instituto Brasileiro de Geografia e Estatística (2012), um estabelecimento agropecuário pode ser constituído de terras próprias, de terras de terceiros (sendo utilizadas via contrato de arrendamento) ou de ambos (parcialmente de terras próprias e parcialmente de terras de terceiros). Isso significa que o cliente pode ter um custo fixo mensal para utilizar as terras em que é realizada a atividade agrícola, caso não sejam de sua propriedade. Esse custo fará com que um cliente que possua arrendamentos tenha uma obrigação a arcar maior do que o que possui propriedade de terra, podendo impactar seu fluxo de caixa e, de acordo com o resultado desse estudo, afetando a existência de *default*.

A árvore de decisão completa não será apresentada neste estudo por sigilo, porém, atualmente, não existe nenhuma variável nela que estime o fluxo de caixa do cliente, mas, por outro lado, os dados referentes a bens, direitos e obrigações são conhecidos. Levando isso em consideração, é proposto que seja estimado um cálculo de cobertura com as informações disponíveis no cadastro do cliente de bens e direitos, face ao endividamento total do cliente para os próximos doze meses (conforme sistema do Banco Central), já ponderando os custos de produção e de arrendamento, se existentes. Se a cobertura for insuficiente, o caso deve ser encaminhado à análise manual. A fórmula de Cobertura do Serviço da Dívida proposta é descrita abaixo:

$$DSC = \frac{NP - RC}{CPLTD + EXP}$$

*DSC = Debt Service Coverage* (Cobertura do Serviço da Dívida);

*NP = Net Profit* (Lucro Líquido);

*RC = Rental Costs* (Custo de Arrendamento);

*CPLTD = Current Portion of Long Term Debt* (Porção de Curto Prazo de Dívidas de Longo Prazo)

*EXP = Expenses* (Despesas)

## 9. CONSIDERAÇÕES FINAIS

Este estudo tem como objetivo principal encontrar evidências da assertividade e adequação das variáveis presentes em uma árvore de decisão automática de crédito, contribuindo com a meta institucional de evitar a ocorrência de inadimplência. A decisão automática de crédito, desta forma, superaria a maneira tradicional de decisão feita manualmente por analistas de crédito.

Para tal, são apresentadas definições básicas de crédito e sistema financeiro, juntamente com metodologias de decisão automática e mineração de dados para árvore de decisão. Utilizando a base de dados de uma instituição financeira que implementou essa metodologia inicialmente em 2015, demonstra-se a aplicação das principais variáveis usadas na árvore de decisão atual e como foi o desempenho delas ao longo dos períodos analisados. O resultado é apresentado na forma de modelo linear generalizado, como uso da função Probit, para entender o impacto que cada variável possui na ocorrência de *default*.

A contribuição deste estudo é vista principalmente pela conclusão de que, no modelo de árvore de decisão atual, as variáveis que possuem maior impacto nos clientes não estão recebendo a priorização correta. Não há restrições quanto à classificação da revenda em que é cadastrada a proposta de financiamento dos clientes, bem como não é realizado qualquer cálculo estimado de fluxo de caixa para os proponentes dessas operações de crédito. Tais critérios seriam verificados mais profundamente se a análise fosse realizada de maneira tradicional, por um analista de crédito.

A eficiência em termos de ganho de tempo já é vista na análise automática, considerando que a decisão é instantânea face ao tempo médio de dois dias para cada análise manual. A redução em inadimplência também já é comprovada, observando que, no mesmo período, a inadimplência da análise automática apresenta índice 46% menor do que a inadimplência manual, conforme visto nos resultados da carteira da instituição financeira.

Com a proposição de uma nova variável que permita estimar o fluxo de caixa dos proponentes e com maiores restrições quanto ao cadastramento das propostas dos

clientes pelas revendedoras, é esperado que a inadimplência tenha índice ainda menor, mesmo que a instituição almeje chegar em 50% de decisões automáticas do total de propostas de crédito. Isto é relevante especialmente no segmento agrícola, em que é necessário um nível de provisionamento maior para lidar com as variáveis imprevisíveis do mercado.

No momento de definição do modelo de decisão automática, cada organização deve utilizar variáveis condizentes à sua realidade e à realidade do mercado em que está inserido. A metodologia utilizada neste estudo possui a vantagem de não ser estática e ser facilmente aplicável a novas variáveis, à medida que a carteira de clientes que passaram por decisão automática aumenta.

Considerando que o mercado está em constante mudança, é necessário que o exercício realizado neste estudo seja revisitado com certa frequência, pois as variáveis que possuem impacto relevante atualmente podem ser substituídas por novas variáveis. Desta forma, é possível controlar a inadimplência de maneira mais efetiva e melhorar a acurácia do modelo, com maior quantidade histórica de base de dados.

Para estudos futuros, é sugerido que seja realizada a análise do efeito de redução de inadimplência das sugestões propostas nesse estudo. É igualmente recomendado que aplique-se o mesmo modelo para mineração de dados que são baseadas em redes neurais e inteligência artificial, comparando o resultado apresentado por variáveis similares.

## REFERÊNCIAS

ABELLAN, Joaquín; MANTAS, Carlos. Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. **Expert Systems with Applications**, [S.l.], v. 41, n. 8, p. 3825-3830, jun. 2014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417413009676>>. Acesso em: 07 jun. 2018.

BANCO CENTRAL DO BRASIL. Brasil. **Circular N° 1.273**: Plano Contábil das Instituições do Sistema Financeiro Nacional – COSIF. [S.l.: s.n.], 1987. 1294 p.

BANCO CENTRAL DO BRASIL. Portal do Fomento. **Taxas de desconto de duplicatas nos bancos**: Veja o quanto os bancos estão cobrando. 2018. Disponível em: <<http://portaldofomento.com.br/noticia.php?id=4264>>. Acesso em: 15 nov. 2018.

BANCO CENTRAL DO BRASIL. **Relatório de Estabilidade Financeira**. Brasília, 2018, 80 p. Disponível em: <[http://www.bcb.gov.br/htms/estabilidade/2018\\_04/refPub.pdf](http://www.bcb.gov.br/htms/estabilidade/2018_04/refPub.pdf)>. Acesso em: 17 abr. 2018.

BANCO CENTRAL DO BRASIL. **Relatório de Estabilidade Financeira**. Brasília, 2017, 69 p. Disponível em: <[http://www.bcb.gov.br/htms/estabilidade/2017\\_04/](http://www.bcb.gov.br/htms/estabilidade/2017_04/)>. Acesso em: 23 jun. 2018.

BANCO CENTRAL DO BRASIL. **Relatório de Estabilidade Financeira**. Brasília, 2016, 79 p. Disponível em: <[http://www.bcb.gov.br/htms/estabilidade/2016\\_04/refPub.pdf](http://www.bcb.gov.br/htms/estabilidade/2016_04/refPub.pdf)>. Acesso em: 30 jun. 2018.

BANCO CENTRAL DO BRASIL. Resolução n. 2.099, de 17 de ago. de 1994. **Aprova Regulamentos que dispõem sobre as condições relativamente ao acesso ao Sistema Financeiro Nacional**. [S.l.], p. 1-33, ago. 1994. Disponível em:

<[http://www.bcb.gov.br/pre/normativos/res/1994/pdf/res\\_2099\\_v1\\_O.pdf](http://www.bcb.gov.br/pre/normativos/res/1994/pdf/res_2099_v1_O.pdf)>. Acesso em: 02 jun. 2018.

BANCO CENTRAL DO BRASIL. Resolução n. 3.721, de 30 de abr. de 2009. **Dispõe sobre a implementação de estrutura de gerenciamento do risco de crédito.** [S.l.], p. 1-6, abr. 2009. Disponível em: <[https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/47611/Res\\_3721\\_v1\\_O.pdf](https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/47611/Res_3721_v1_O.pdf)>. Acesso em: 04 jun. 2018.

BANCO POPULAR DO BRASIL. Banco do Brasil. **O que é crédito?**. Disponível em: <<http://www.bb.com.br/portalbb/page251,8900,8923,0,0,1,0.bb?codigoMenu=5415&codigoNoticia=8133>>. Acesso em: 20 maio 2018.

BERRY, Michael; LINOFF, Gordon. **Data Mining Techniques: For marketing, sales and customer relationship management.** 2. ed. Indianapolis: Wiley Publishing, 2004. 672 p. Disponível em: <<http://197.14.51.10:81/pmb/GESTION2/MARKETING/Data%20Mining%20Techniques%20For%20Marketing%20Sales%20And%20Customer%20Relationship%20Management%20Ed.pdf>>. Acesso em: 29 maio 2018.

BLATT, Adriano. **Avaliação de Risco e Decisão de Crédito: um enfoque prático.** 1. ed. São Paulo: Nobel, 1999. 240 p.

CAMARGOS, Marcos Antônio de et al. Fatores Condicionantes de Inadimplência em Processos de Concessão de Crédito a Micro e Pequenas Empresas do Estado de Minas Gerais. **Revista de Administração Contemporânea**, Curitiba, v. 14, n. 2, p. 333-352, mar. 2010. Disponível em: <<http://www.scielo.br/pdf/rac/v14n2/v14n2a09.pdf>>. Acesso em: 29 maio 2018.

CAMBRIDGE ENGLISH. (Inglaterra). **Cambridge Dictionary**. [S.l.]: Cambridge University Press, 2018. Disponível em: <<https://dictionary.cambridge.org/pt/>>. Acesso em: 31 maio 2018.

CAOQUETTE, John et al. **Managing Credit Risk: The Great Challenge for the Global Financial Markets**. 2. ed. Hoboken: John Wiley & Sons, 2008. 655 p. Disponível em: <[http://www.untag-smd.ac.id/files/Perpustakaan\\_Digital\\_1/CREDIT%20RISK%20Managing%20credit%20risk%20%20the%20great%20challenge%20for%20global%20financial%20markets.pdf](http://www.untag-smd.ac.id/files/Perpustakaan_Digital_1/CREDIT%20RISK%20Managing%20credit%20risk%20%20the%20great%20challenge%20for%20global%20financial%20markets.pdf)>. Acesso em: 24 maio 2018.

CHAIA, Alexandre Jorge. **Modelos de Gestão do Risco de Crédito e sua Aplicabilidade ao Mercado Brasileiro**. 2003. 121 p. Dissertação (Mestrado em Administração) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2003. Disponível em: <[http://www.teses.usp.br/teses/disponiveis/12/12139/tde-21012004-224716/publico/Modelos\\_de\\_Credito.pdf](http://www.teses.usp.br/teses/disponiveis/12/12139/tde-21012004-224716/publico/Modelos_de_Credito.pdf)>. Acesso em: 18 abr. 2018.

CHEN, S. C.; HUANG, M. Y. Constructing credit auditing and control & management model with data mining technique. **Expert Systems with Applications**, [S.l.], v. 38, n. 5, p. 5359-5365, maio. 2011. Disponível em: <<http://ir.lib.ncut.edu.tw/bitstream/987654321/6854/2/323.pdf>>. Acesso em: 07 jun. 2018.

CHITRA, K.; SUBASHINI, B. Automatic Credit Approval using Classification Method. **International Journal of Scientific & Engineering Research**, [S.l.], v. 4, n. 7, p. 2026-2029, jul. 2013. Disponível em: <<https://www.ijser.org/researchpaper/Automatic-Credit-Approval-using-Classification-Method.pdf>>. Acesso em: 10 jun. 2018.

COBRA, Marcos Henrique Nogueira. **Marketing Básico: Uma Perspectiva Brasileira**. 4. ed. São Paulo: Atlas, 1997. 552 p.

COELHO, Danilo; NEGRI, João Alberto De. Impacto do Financiamento do BNDES sobre a Produtividade das Empresas: Uma Aplicação do Efeito Quantílico de Tratamento. In: Encontro Nacional de Economia, XXXVIII., 2010, Salvador. **Trabalhos...** [S.l.: s.n.], 2010. p. 1-15. Disponível em: <<http://www.anpec.org.br/encontro2010/inscricao/arquivos/000-62cf725608576526071dfbbed02385cf.pdf>>. Acesso em: 02 jun. 2018.

CONFEDERAÇÃO NACIONAL DE DIRIGENTES LOJISTAS. **Inadimplentes no Brasil 2017**: Perfil e Comportamento frente às Dívidas. São Paulo, 2017, 14 p. Disponível em: <[https://www.spcbrasil.org.br/wpimprensa/wp-content/uploads/2017/08/Analise\\_perfil\\_inadimplente\\_2017.pdf](https://www.spcbrasil.org.br/wpimprensa/wp-content/uploads/2017/08/Analise_perfil_inadimplente_2017.pdf)>. Acesso em: 20 abr. 2018.

CRONE, Sven; FINLAY, Steven. Instance sampling in credit scoring: An empirical study of sample size and balancing. **International Journal of Forecasting**, [S.l.], v. 28, n. 1, p. 224-238, jan. 2012. Disponível em: <<https://pdfs.semanticscholar.org/5512/e1506be31093a2d8c03599038210e0f40592.pdf>>. Acesso em: 09 jun. 2018.

CURNOW, George et al. Automating Credit and Collections Decisions at AT&T Capital Corporation. **Interfaces**, [S.l.], v. 27, n. 1, p. 29-52, jan. 1997. Disponível em: <<https://pubsonline.informs.org/doi/citedby/10.1287/inte.27.1.29>>. Acesso em: 10 jun. 2018.

FARIA, Maurício Piragibe de Carvalho. **Análise de crédito à pequena empresa**: Um modelo de escoragem baseado nas metodologias estatísticas: análise fatorial e lógica fuzzy. 2006. 123 p. Dissertação (Mestrado Profissionalizante em Economia) - Faculdade de Economia e Finanças, Ibmec, Rio de Janeiro, 2006. Disponível em: <<http://livros01.livrosgratis.com.br/cp012808.pdf>>. Acesso em: 29 maio 2018.

FILHO, Ernani Teixeira Torres; COSTA, Fernando Nogueira da. BNDES e o financiamento do desenvolvimento. **Economia e Sociedade**, Campinas, v. 21, p. 979-1009, dez. 2012. Disponível em: <<http://www.scielo.br/pdf/ecos/v21nspe/v21nspea11.pdf>>. Acesso em: 31 maio 2018.

FORTUNA, Eduardo. **Mercado Financeiro: Produtos e Serviços**. 20. ed. Rio de Janeiro: Qualitymark, 2015. 904 p.

GALEANO, Edileuza Vital; FEIJÓ, Carmem. Crédito e crescimento econômico: evidências a partir de um painel de dados regionais para a economia brasileira nos anos 2000. **Revista Economica do Nordeste**, [S.l.], v. 43, n. 02, p. 202-219, abr. 2011. Disponível em: <<https://ren.emnuvens.com.br/ren/article/view/213/191>>. Acesso em: 03 nov. 2018.

HALE, Roger. **Credit Analysis: A Complete Guide**. 1. ed. Singapura: John Wiley & Sons, 1983. 312 p.

HAND, David John; HENLEY, William. Statistical Classification Methods in Consumer Credit Scoring: a Review. **Journal of the Royal Statistical Society Series A**, [S.l.], v. 160, n. 3, p. 523-541, set. 1997. Disponível em: <<https://pdfs.semanticscholar.org/fa58/5ac49b37a801ccd1b2e49118518414c810e2.pdf>>. Acesso em: 09 jun. 2018.

HEREDIA, Beatriz Maria Alásia de; CINTRÃO, Rosângela Pezza. Gênero e acesso a políticas públicas no meio rural brasileiro. **Revista NERA**, Presidente Prudente, n. 8, p. 1-28, jan. 2006. Disponível em: <<http://www.mstemdados.org/sites/default/files/1443-4191-1-PB.PDF>>. Acesso em: 02 jun. 2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Pesquisa Nacional da Atividade Agropecuária. **Manual do Entrevistador: Pré-teste**. 2012. Disponível em:

<[https://ww2.ibge.gov.br/home/estatistica/indicadores/prpa/Manual\\_da\\_PNAG\\_Modulo\\_1%20\\_pre\\_teste\\_2012.pdf](https://ww2.ibge.gov.br/home/estatistica/indicadores/prpa/Manual_da_PNAG_Modulo_1%20_pre_teste_2012.pdf)>. Acesso em: 17 nov. 2018.

JÚNIOR, Paulo de Tarso Dutra Lima. **O Processo de Crédito para Empresas em uma Instituição Financeira**. 2011. 53 p. Trabalho de Conclusão de Curso (Especialização em Gestão de Negócios Financeiros) - Escola de Administração, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/77632/000894757.pdf?sequence=1>>. Acesso em: 01 mai. 2018.

KHANDANI, Amir; KIM, Adlar Jeewook; LO, Andrew Wen-Chuan. Consumer Credit Risk Models via Machine-Learning Algorithms. **Journal of Banking & Finance**, [S.l.], v. 34, n. 11, p. 2767-2787, nov. 2010. Disponível em: <<https://dspace.mit.edu/openaccess-disseminate/1721.1/66301>>. Acesso em: 09 jun. 2018.

KOH, Hian Chye; TAN, Wei Chin; GOH, Chwee Peng. A Two-step Method to Construct Credit Scoring Models with Data Mining Techniques. **International Journal of Business and Information**, [S.l.], v. 1, n. 1, p. 96-118, jan. 2006. Disponível em: <<https://ijbi.org/ijbi/article/view/5/6>>. Acesso em: 09 jun. 2018.

LEMOS, Eliane Prezepiorski; STEINER, Maria Teresinha Arns; NIEVOLA, Julio César. Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining. **Revista de Administração**, São Paulo, v. 40, n. 3, p. 225-234, jul. 2005. Disponível em: <<http://200.232.30.99/download.asp?file=V4003225.pdf>>. Acesso em: 17 abr. 2018.

LOUREIRO, Luís Manuel de Jesus; GAMEIRO, Manuel Gonçalves Henriques. Interpretação crítica dos resultados estatísticos: para lá da significância estatística. **Revista de Enfermagem Referência**, Coimbra, v. 3, n. 3, p. 151-162, mar. 2011. Disponível em: <<http://www.scielo.mec.pt/pdf/ref/vserIIIIn3/serIIIIn3a16.pdf>>. Acesso em: 15 nov. 2018.

MACHOVÁ, Kristína; BARČÁK, František; BEDNÁR, Peter. A Bagging Method using Decision Trees in the Role of Base Classifiers. **Acta Polytechnica Hungarica**, Budapeste, v. 3, n. 2, p. 121-132, jan. 2006. Disponível em: <[https://www.uni-obuda.hu/journal/Machova\\_Barca\\_Bednar\\_6.pdf](https://www.uni-obuda.hu/journal/Machova_Barca_Bednar_6.pdf)>. Acesso em: 10 jun. 2018.

MAGALHÃES, Matheus Fraga. **Gestão da análise e concessão de crédito**: proposta de política de crédito para a Cooperativa Alfa. 2017. 101 p. Trabalho de Conclusão de Curso (Bacharelado em Administração) - Escola de Administração, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Disponível em: <<https://www.lume.ufrgs.br/bitstream/handle/10183/169940/001051565.pdf?sequence=1>>. Acesso em: 17 abr. 2018.

MALACRIDA, Mara Jane Contrera; YAMAMOTO, Marina Mitiyo. Governança Corporativa: Nível de Evidenciação das Informações e sua Relação com a Volatilidade das Ações do Ibovespa. **Revista Contabilidade & Finanças**, São Paulo, p. 65-79, set. 2006. Disponível em: <<http://www.scielo.br/pdf/rcf/v17nspe/v17nspea06>>. Acesso em: 15 nov. 2018.

MARQUÉS, Ana Isabel; GARCÍA, Vicente; SÁNCHEZ, Javier Salvador. A literature review on the application of evolutionary computing to credit scoring. **Journal of the Operational Research Society**, [S.l.], v. 64, n. 9, p. 1384-1399, set. 2013. Disponível em: <<https://link.springer.com/content/pdf/10.1057%2Fjors.2012.145.pdf>>. Acesso em: 10 jun. 2018.

MCKINSEY QUARTERLY. **Clouds, big data, and smart assets**: Ten tech-enabled business trends to watch. Nova Iorque, 2010, 14 p. Disponível em: <[http://www.itglobal-services.de/files/100810\\_McK\\_Clouds\\_big\\_data\\_and%20smart%20assets.pdf](http://www.itglobal-services.de/files/100810_McK_Clouds_big_data_and%20smart%20assets.pdf)>. Acesso em: 21 abr. 2018.

MEDINA, Rosa Puertas; SELVA, Maria Luisa Martí. Análisis del Credit Scoring. **RAE - Revista de Administração de Empresas**, São Paulo, v. 53, n. 3, p. 303-315, maio. 2013. Disponível em: <<http://www.scielo.br/pdf/rae/v53n3/07.pdf>>. Acesso em: 21 abr. 2018.

MINUSSI, João Alberto; DAMACENA, Cláudio; JUNIOR, Walter Lee Ness. Um Modelo de Previsão de Solvência Utilizando Regressão Logística. **Revista de Administração Contemporânea**, Curitiba, v. 6, n. 3, p. 109-128, set. 2002. Disponível em: <<http://www.scielo.br/pdf/rac/v6n3/v6n3a07.pdf>>. Acesso em: 07 jun. 2018.

NETO, Alexandre Assaf. **Mercado Financeiro**. 10. ed. São Paulo: Atlas, 2011. 339 p.

NIE, Guangli et al. Credit card churn forecasting by logistic regression and decision tree. **Expert Systems with Applications**, [S.l.], v. 38, n. 12, p. 15273-15285, nov. 2011. Disponível em: <<https://www.semanticscholar.org/paper/Credit-card-churn-forecasting-by-logistic-and-tree-Nie-Rowe/74f601ec8f54fd818140714a8b11cba90b07b41f>>. Acesso em: 09 jun. 2018.

NOLDE, Natalia; ZIEGEL, Johanna. Elicitability and Backtesting: Perspectives for Banking Regulation. **The Annals of Applied Statistics**, Bethesda, v. 11, n. 04, p. 1833-1874, dez. 2017. Disponível em: <[https://projecteuclid.org/download/pdfview\\_1/euclid.aos/1514430265](https://projecteuclid.org/download/pdfview_1/euclid.aos/1514430265)>. Acesso em: 01 jul. 2018.

PAILHÉ, Cristina. Information systems for the management of credit risk: survey in the Argentine financial system. **Munich Personal RePEc Archive**. Buenos Aires, 2006, 33 p. Disponível em: <[https://mpira.ub.uni-muenchen.de/1170/1/MPRA\\_paper\\_1170.pdf](https://mpira.ub.uni-muenchen.de/1170/1/MPRA_paper_1170.pdf)>. Acesso em: 21 abr. 2018.

PERLIN, Marcelo; RIGHI, Marcelo; FILOMENA, Tiago. **A Consumer Credit Risk Structural Model Based on Affordability**: Balance at Risk. 2016. Disponível em:

<[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2754375](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2754375)>. Acesso em: 30 jun. 2018.

PERLIN, Marcelo Scherer. **Processamento e Análise de Dados Financeiros e Econômicos com o R**. 2. ed. Porto Alegre: [s.n.], 2018. 432 p.

PORTAL BRASIL. Governo do Brasil. **Confira as regras do Imposto de Renda 2015**. 2015. Disponível em: <<http://www.brasil.gov.br/economia-e-emprego/2015/02/confira-as-regras-do-imposto-de-renda-2015>>. Acesso em: 12 out. 2018.

PORTAL BRASIL. Governo do Brasil. **IRPF 2017**: Saiba quem deve apresentar a declaração de Imposto de Renda 2017. 2017. Disponível em: <<http://www.brasil.gov.br/economia-e-emprego/2017/03/saiba-quem-deve-apresentar-a-declaracao-de-imposto-de-renda-2017>>. Acesso em: 12 out. 2018.

PORTUGAL, Marcelo. **Notas Introdutórias Sobre o Princípio de Máxima Verossimilhança**: Estimação e Teste de Hipóteses. 1995. Disponível em: <<http://www.agg.ufba.br/maximavrossi.pdf>>. Acesso em: 24 nov. 2018.

ROMERO, João Prates; ÁVILA, Jorge Luís Teixeira. Sistema Financeiro e Desenvolvimento Regional: Um Estudo sobre o Financiamento Bancário da Atividade Industrial no Brasil. **Economia**, Brasília, v. 11, n. 01, p. 217-263, jan. 2010. Disponível em: <[http://www.anpec.org.br/revista/vol11/vol11n1p217\\_263.pdf](http://www.anpec.org.br/revista/vol11/vol11n1p217_263.pdf)>. Acesso em: 03 nov. 2018.

ROSENBERG, Eric; GLEIT, Alan. Quantitative Methods in Credit Management: A Survey. **Operations Research**, Maryland, v. 42, n. 4, p. 589-613, jul. 1994. Disponível em: <<https://pubsonline.informs.org/doi/pdf/10.1287/opre.42.4.589>>. Acesso em: 18 abr. 2018.

ROSS, Stephen Alan et al. **Fundamentos de Administração Financeira**. 9. ed. Porto Alegre: AMGH, 2013. 806 p. Disponível em: <<https://pt.scribd.com/document/362911820/Administracao-financeira-Stephen-A-Ross-pdf>>. Acesso em: 20 maio 2018.

SAKPRASAT, Sam; SINCLAIR, Mark C. . Classification Rule Mining for Automatic Credit Approval using Genetic Programming. In: IEEE Congress on Evolutionary Computation, 9., 2007, Singapura. **Conference proceedings...** [S.l.: s.n.], 2007. p. 548-555. v. 9. Disponível em: <<http://sci2s.ugr.es/keel/pdf/specific/congreso/scopus66.pdf>>. Acesso em: 21 abr. 2018.

SANTOS, Edno Oliveira dos. **Administração Financeira da Pequena e Média Empresa**. 2. ed. São Paulo: Atlas, 2010. 280 p.

SANTOS, José Odalio dos. **Análise de Crédito: Empresas e Pessoas Físicas**. 6. ed. São Paulo: Atlas, 2015. 352 p.

SCHECHTMAN, Ricardo et al. Credit Risk Measurement and the Regulation of Bank Capital and Provision Requirements in Brazil: A Corporate Analysis. **Working Paper Series**, Brasília, n. 91, p. 1-46, dez. 2004. Disponível em: <<https://www.bcb.gov.br/pec/wps/ingl/wps91.pdf>>. Acesso em: 01 jul. 2018.

SCHRICKEL, Wolfgang Kurt. **Análise de Crédito: Concessão e Gerência de Empréstimos**. 3. ed. São Paulo: Atlas, 1997. 353 p.

SEBRAE - SERVIÇO BRASILEIRO DE APOIO ÀS MICRO E PEQUENAS EMPRESAS. **Anuário do Trabalho na Micro e Pequena Empresa**. Brasília, 2013, 288 p. Disponível em: <[http://www.sebrae.com.br/Sebrae/Portal%20Sebrae/Anexos/Anuario%20do%20Trabalho%20Na%20Micro%20e%20Pequena%20Empresa\\_2013.pdf](http://www.sebrae.com.br/Sebrae/Portal%20Sebrae/Anexos/Anuario%20do%20Trabalho%20Na%20Micro%20e%20Pequena%20Empresa_2013.pdf)>. Acesso em: 12 nov. 2018.

SILVA, José Pereira da. **Gestão e Análise de Risco de Crédito**. 8. ed. São Paulo: Atlas, 2014. 487 p.

SOUSA, Marcos de Moraes; FIGUEIREDO, Reginaldo Santana. Análise de Crédito por meio de Mineração de Dados: Aplicação em Cooperativa de Crédito. **JISTEM - Journal of Information Systems and Technology Management**, São Paulo, v. 11, n. 2, p. 379-396, maio. 2014. Disponível em: <[http://www.scielo.br/pdf/jjstm/v11n2/pt\\_1807-1775-jistem-11-2-0379.pdf](http://www.scielo.br/pdf/jjstm/v11n2/pt_1807-1775-jistem-11-2-0379.pdf)>. Acesso em: 18 abr. 2018.

STEINER, Maria Terezinha Arns et al. Sistemas especialistas probabilísticos e redes neurais na análise do crédito bancário. **Revista de Administração**, São Paulo, v. 34, n. 3, p. 56-67, jul. 1999. Disponível em: <[http://200.232.30.99/busca/artigo.asp?num\\_artigo=98](http://200.232.30.99/busca/artigo.asp?num_artigo=98)>. Acesso em: 18 abr. 2018.

SUBSECRETARIA DE ARRECAÇÃO E ATENDIMENTO. Receita Federal. **Obrigatoriedade**: Pessoas obrigadas a apresentar a Declaração de Ajuste Anual do IRPF 2016. 2016. Disponível em: <<http://idg.receita.fazenda.gov.br/interface/cidadao/irpf/2016/declaracao/obrigatoriedade>>. Acesso em: 12 out. 2018.

TESSAROLO, Pedro Henrique; MAGALHÃES, Willian Barbosa. A Era do Big Data no Conteúdo Digital: os Dados Estruturados e não Estruturados. In: SEINPAR - Semana de Informática de Paranavaí, 17., 2015, Paranavaí. **Publicações...** Paranavaí: [s.n.], 2015. p. 1-5. Disponível em: <[http://web.unipar.br/~seinpar/2015/\\_include/artigos/Pedro\\_Henrique\\_Tessarolo.pdf](http://web.unipar.br/~seinpar/2015/_include/artigos/Pedro_Henrique_Tessarolo.pdf)>. Acesso em: 21 abr. 2018.

THOMAS, Lyn C. A survey of credit and behavioral scoring: forecasting financial risk of lending to consumers. **International Journal of Forecasting**, [S.l.], v. 16, n. 2, p. 149-172, abr. 2000. Disponível em:

<<http://pages.ucsd.edu/~aronatas/project/academic/A%20survey%20of%20credit%20and%20behavioural%20scoring%20Forecasting%20fina.pdf>>. Acesso em: 17 abr. 2018.

TIRYAKI, Gisele Ferreira et al. Ciclos de Crédito, Inadimplência e as Flutuações Econômicas no Brasil. **Revista de Economia Contemporânea**, Rio de Janeiro, v. 21, n. 1, p. 1-33, fev. 2017. Disponível em: <<http://www.scielo.br/pdf/rec/v21n1/1415-9848-rec-21-01-e172112.pdf>>. Acesso em: 21 maio 2018.

TOM, Diana. Credit Risk Analysis and Credit Scoring: Now and in the Future. **Business Credit**, [S.l.], v. 107, n. 3, p. 1-3, mar. 2005. Disponível em: <<https://www.questia.com/magazine/1G1-130569338/credit-risk-analysis-and-credit-scoring-now-and-in>>. Acesso em: 21 abr. 2018.

WANG, Gang et al. Two credit scoring models based on dual strategy ensemble trees. **Knowledge-Based Systems**, [S.l.], v. 26, p. 61-68, fev. 2012. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950705111001353>>. Acesso em: 29 maio 2018.

WOOLDRIDGE, Jeffrey M. **Introductory Econometrics: A Modern Approach**. 5. ed. Mason: Cengage Learning, 2013. 910 p. Disponível em: <[http://economics.ut.ac.ir/documents/3030266/14100645/Jeffrey\\_M.\\_Wooldridge\\_Introductory\\_Econometrics\\_A\\_Modern\\_Approach\\_\\_2012.pdf](http://economics.ut.ac.ir/documents/3030266/14100645/Jeffrey_M._Wooldridge_Introductory_Econometrics_A_Modern_Approach__2012.pdf)>. Acesso em: 03 nov. 2018.

YAP, Bee Wah; ONG, Seng Huat; HUSAIN, Nor Huselina Mohamed. Using data mining to improve assessment of credit worthiness via credit scoring models. **Expert Systems with Applications**, [S.l.], v. 38, n. 10, p. 13274-13283, set. 2011. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/5713481/>>. Acesso em: 09 jun. 2018.

## ANEXO I – RESULTADO DA REGRESSÃO PROBIT

Variável	Coefficiente	Erro-Padrão	Teste Z	Valor-P
Orientação Sexual	13,3380	4303,3680	0,003	0,9975
Alavancagem Financeira	-17,6316	2453,5964	-0,007	0,9943
Classificação do Revendedor - B	1,7839	0,4114	4,336	0,00000145
Linhas de Endividamento – Empréstimos	0,4998	0,3932	1,271	0,2037
Linhas de Endividamento – Financiamentos	16,0082	1080,1554	0,015	0,9882
Linhas de Endividamento – Títulos Descontados	2,0844	0,4953	4,208	0,00000258
Linhas de Endividamento – Parcelamento de Cartão de Crédito	-0,2342	0,4053	-0,578	0,5634
Classificação do Revendedor – C	1,3758	0,5363	2,565	0,0103
Classificação do Revendedor - A	-15,1428	1953,5421	-0,008	0,9938

Cobertura do Valor Financiado Solicitado	-0,9080	0,3921	-2,316	0,0206
Propriedade de Terra	-0,6687	0,3792	-1,764	0,0778
Cobertura do Endividamento	0,3379	0,4645	0,727	0,4670
Região – Nordeste	-17,0712	1751,6637	-0,010	0,9922
Região – Centro- Oeste	-16,3780	2596,0653	-0,006	0,9950
Região – Sudeste	-0,5559	0,4949	-1,123	0,2614
Região – Sul	-0,8215	0,5111	-1,607	0,1080
Idade – 18 a 24 anos	15,4740	10816,3914	0,001	0,9989
Idade – 25 a 34 anos	15,1581	10816,3914	0,001	0,9989
Idade – 35 a 44 anos	14,0357	10816,3914	0,001	0,9990
Idade – 45 a 54 anos	13,5238	10816,3914	0,001	0,9990
Idade – 55 a 64 anos	14,6153	10816,3914	0,001	0,9989
Idade – 65 a 74 anos	-0,6733	13557,6602	0,000	1,0000
Comprovação de Receita	0,2520	0,4236	0,595	0,5520
Divergência de	-0,2748	0,4058	-0,677	0,4983

Receita				
Participação Societária	-0,4328	0,3989	-1,085	0,2780

Fonte: *Software R.* Elaborado pelo autor