

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ (UTFPR)
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO E
SISTEMAS

FLÁVIO FÜHR

PROPOSIÇÃO DE MODELOS DE PREVISÃO DE RISCO DE CRÉDITO PARA
PEQUENAS E MÉDIAS EMPRESAS POR MEIO DA REGRESSÃO LOGÍSTICA

Pato Branco
2018

FLÁVIO FÜHR

PROPOSIÇÃO DE MODELOS DE PREVISÃO DE RISCO DE CRÉDITO PARA
PEQUENAS E MÉDIAS EMPRESAS POR MEIO DA REGRESSÃO LOGÍSTICA

Dissertação apresentada como requisito parcial à
obtenção do grau de Mestre em Engenharia de
Produção, do Programa de Pós-Graduação em
Engenharia de Produção e Sistemas (PPGEPS), Área de
Concentração: Gestão de Sistemas Produtivos, da
Universidade Tecnológica Federal do Paraná.
Orientador: Prof. Dr. José Donizetti de Lima

Pato Branco

2018

F959p

Führ, Flávio.

Proposição de modelos de previsão de risco de crédito para pequenas e médias empresas por meio da regressão logística / Flávio Führ . -- 2018. 85 f. : il. ; 30 cm.

Orientador: Prof. Dr. José Donizetti de Lima
Dissertação (Mestrado) - Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Engenharia de Produção e Sistemas. Pato Branco, PR, 2018.
Bibliografia: f. 79 - 85.

1. Análise de crédito. 2. Sistema de avaliação de risco de crédito (Finanças). 3. Pequenas e médias empresas. I. Lima, José Donizetti de, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Engenharia de Produção e Sistemas. III. Título.

CDD 22. ed. 670.42

Ficha Catalográfica elaborada por
Suélem Belmudes Cardoso CRB9/1630
Biblioteca da UTFPR Campus Pato Branco



TERMO DE APROVAÇÃO DE DISSERTAÇÃO Nº 40

A Dissertação de Mestrado intitulada “**Proposição de modelos de previsão de risco de crédito para pequenas e médias empresas por meio da regressão logística**”, defendida em sessão pública pelo candidato **Flávio Führ**, no dia 20 de setembro de 2018, foi julgada para a obtenção do título de Mestre em Engenharia de Produção e Sistemas, área de concentração Gestão dos Sistemas Produtivos, e aprovada em sua forma final, pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas.

BANCA EXAMINADORA:

Prof. Dr. José Donizetti de Lima - Presidente – UTFPR

Prof. Dr. Gilson Ditzel Santos – UTFPR

Prof. Dr. Sady Mazzioni - Unochapecó

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Pato Branco, 04 de outubro de 2018.

Carimbo e assinatura do Coordenador do Programa.

Dedico este trabalho aos meus pais familiares e amigos que sempre me incentivaram.

AGRADECIMENTOS

Dou início aos agradecimentos, ao diretor de minha fé “DEUS”, por ter oportunizado esta experiência, que por muito tempo era sonho e hoje realidade.

Aos meus Pais: João Danilo Führ e Luiza Pinheiro Führ “*in memory*”, meu enorme agradecimento.

A minha esposa Denise Regina Acorsi e a meu filho Leonardo Augusto Führ, pela compreensão das ausências em momentos familiares e particulares, mas principalmente por seus apoios nas horas de dificuldades.

Aos meus colegas de curso pelas palavras e companheirismo que se formou no decorrer do curso.

Ao Professor Orientador Dr. José Donizetti de Lima, pela paciência, carisma e disposição que dispendeu a minha pessoa neste processo do conhecer, o qual foi permeado por seu poder de transmissão de seus vastos conhecimentos.

A todos os Professores do Colegiado do Curso de Mestrado.

Este agradecimento também é estendido a UTFPR e IFPR e seus funcionários que produzem ambiente e serviços a contento de seus usuários.

OBRIGADO A TODOS!

RESUMO

FÜHR, Flávio. PROPOSIÇÃO DE MODELOS DE PREVISÃO DE RISCO DE CRÉDITO PARA PEQUENAS E MÉDIAS EMPRESAS POR MEIO DA REGRESSÃO LOGÍSTICA. 2018. 86 folhas. Dissertação (Mestrado em Engenharia de Produção e Sistemas) – Universidade Tecnológica Federal do Paraná – Campus Pato Branco.

O presente trabalho, busca contribuir com o setor financeiro de fomento de crédito para pequenas e médias empresas (PMEs). Para isso propõe a elaboração de modelos de previsão de risco de crédito para PMEs por meio da Regressão Logística (RL). Utilizando de informações de cadastro e histórico de crédito, foram extraídas variáveis, com relevante significância, para a definição da probabilidade de ocorrência de inadimplência. O levantamento do dados e geração de informações, foram feitos por meio da pesquisa exploratória, com procedimentos experimentais, tendo como campo de exploração banco de dados de uma Cooperativa de Crédito. Metodologicamente, criou-se 4 classes de empresas: Microempreendedor Individual (MEI), Microempresa (ME), Pequena Empresa (PE) e Média Empresa (MédE), sendo que a base de dados geral, foi redistribuída conforme faixas de faturamento, surgindo assim as 4 novas bases de dados. Para melhoria dos modelos e redução das diferenças, dentro da base de dados, de cada classe de empresas, utilizou-se o processo de discretização e a criação de variáveis “*dummy*” ou artificiais. Como resultado da aplicação da técnica estatística na base de dados, nas 4 classes de empresas: MEI, ME, PE, MédE e nos Dados Gerais (DG), obteve-se uma confirmação da relevância da RL na elaboração dos modelos. As acurácias dos modelos apresentaram percentuais expressivos para base de dados com variáveis não contábeis e não auditáveis, atingindo percentuais satisfatórios. Para MEI, o percentual de acurácia foi de 83%, utilizando 2 variáveis na composição do modelo. Quanto a ME, apresentou um acurácia 84,9%, utilizando 5 variáveis na composição do modelo. Para PE a acurácia atingiu 88,5%, porém incluindo apenas 1 variável no modelo. Para MédE a acurácia foi de 83%, apresentando 3 variáveis no modelo e para os DG, a acurácia foi de 85%, apresentando 5 variáveis na composição do modelo. Ainda foi possível observar quais variáveis possuem maior relevância dentro da base de dados. Os modelos desenvolvidos são ferramentas que podem contribuir com o analista de crédito na identificação de possíveis adimplentes ou inadimplentes para instituições financeiras que possuem PMEs em seu portfólio.

Palavras-chave: Modelagem de *Credit Scoring*, Risco de Crédito, Pequenas e Médias Empresas, Pontuação de Crédito e Regressão Logística.

ABSTRACT

FÜHR, Flávio. PROPOSAL OF CREDIT RISK FORECAST MODELS FOR SMALL AND MEDIUM-SIZED ENTERPRISES THROUGH LOGISTIC REGRESSION. 2018. 86 leaves. Dissertation (Master in Engineering of Production and Systems) - Federal Technological University of Paraná - Pato Branco Campus.

The present work seeks to contribute to the financial sector of credit promotion for small and medium enterprises (SMEs). For this purpose, it proposes the elaboration of credit risk forecasting models for SMEs through Logistic Regression (RL). Using data from credit registry and credit history, variables with significant significance were extracted for the definition of probability of occurrence of default. Data collection and generation of information was done through the exploratory research, with experimental procedures, having as the field of exploitation a credit cooperative database. Methodologically, four classes of companies were created: Individual Microentrepreneur (MEI), Microenterprise (SM), Small Business (PE) and Medium Enterprise (MédE). The general database was redistributed according to the billing ranges. the 4 new databases. To improve the models and reduce the differences, within the database, of each class of companies, the process of discretization and the creation of dummy or artificial variables was used. As a result of the application of the statistical technique in the database, in the 4 classes of companies: MEI, ME, PE, MédE and in the General Data (DG), a confirmation of the relevance of the RL in the elaboration of the models was obtained. The accuracy of the models presented expressive percentages for the database with non-accounting and non-auditable variables, reaching satisfactory percentages. For MEI, the percentage of accuracy was 83%, using 2 variables in the composition of the model. As for ME, it presented an accuracy of 84.9%, using 5 variables in the model composition. For PE the accuracy reached 88.5%, however including only 1 variable in the model. For MEc the accuracy was 83%, presenting 3 variables in the model and for the DGs, the accuracy was of 85%, presenting 5 variables in the composition of the model. It was still possible to observe which variables have greater relevance within the database. The models developed are tools that can contribute to the credit analyst in the identification of possible good payer or defaulters for financial institutions that have SMEs in their portfolio.

Keywords: Modeling Credit Scoring, Credit Risk, Small, Medium-sized Enterprises, credit score and Logistic Regression.

LISTA DE ILUSTRAÇÕES

Figura 1 — Distribuição dos tópicos da Revisão de Literatura	22
Figura 2 — Estrutura Sistema SFN para bancos e cooperativas	24
Figura 3 — Estudos pioneiros na previsão de falências	28
Figura 4 — Curva da Logística	35
Figura 5 — Delimitação da pesquisa	39
Figura 6 — Predição e volume de técnicas encontradas nos modelos nos 20 artigos selecionados	41
Figura 7 — Acurácia das técnicas e o número de vezes que foram utilizadas dentro do portfólio dos 11 artigos direcionados para as PMEs	42
Figura 8 — Relação entre Objetivo Geral e Objetivos Específicos	43
Figura 9 — Ilustração da metodologia aplicada	43
Figura 10 — Discretização: Saldo Devedor	51
Figura 11 — Discretização: Tempo de Filiação	51
Figura 12 — Discretização: Renda anual	51
Figura 13 — Discretização: Idade	52
Figura 14 — Discretização: Número de Parcelas	52
Figura 15 — Discretização: Valor do Contrato	52
Figura 16 — Discretização: Saldo Devedor	53
Figura 17 — Discretização: Tempo Filiação	53
Figura 18 — Discretização: Renda Anual	53
Figura 19 — Discretização: Idade	54
Figura 20 — Discretização: Número de Parcela	54
Figura 21 — Valor Contrato	54
Figura 22 — Discretização: Saldo Devedor	55
Figura 23 — Discretização: Tempo Filiação	55
Figura 24 — Discretização: Renda anual	55
Figura 25 — Discretização: Idade	56
Figura 26 — Discretização: Número de Parcela	56
Figura 27 — Discretização: Valor do Contrato	56
Figura 28 — Discretização: Saldo Devedor	57
Figura 29 — Discretização: Tempo de Filiação	57

Figura 30 — Discretização: Renda Anual	57
Figura 31 — Discretização: Idade	58
Figura 32 — Discretização: Número de Parcelas	58
Figura 33 — Discretização: Valor do Contrato	58
Figura 34 — Discretização: Saldo Devedor	59
Figura 35 — Discretização: Tempo Filiação	59
Figura 36 — Discretização: Renda Anua	59
Figura 37 — Discretização: Idade	60
Figura 38 — Discretização: Número de Parcelas	60
Figura 39 — Discretização: Valor do Contrato	60
Figura 40 — Cálculo das variáveis utilizadas nos testes de RL	66
Figura 41 — Teste de verossimilhança, Cox & Snell e Nagelkerke	69
Figura 42 — Equação Logística	70
Figura 43 — Equação Logística para MEI	70
Figura 44 — Equação Logística para ME	71
Figura 45 — Equação Logística para PE	71
Figura 46 — Equação Logística para MédE	71
Figura 47 — Equação Logística para DG	72
Figura 48 — Resultados dos testes de acurácia para classe de empresas classificadas como MEI	72
Figura 49 — Resultados dos testes de acurácia para as empresas classificadas como ME	73
Figura 50 — Resultados dos testes de acurácia para classe de empresas classificadas como PE	73
Figura 51 — Resultados dos testes de acurácia para as empresas classificadas como MédE	74
Figura 52 — Resultados dos testes de acurácia para a base de DG	74
Figura 53 — Distribuição da Acurácia, em ordem crescente	75

LISTA DE EQUAÇÕES

Equação 1 — Razão de chance	34
Equação 2 — Constante matemática	34
Equação 3 — Probabilidade associada à ocorrência	34
Equação 4 — Equação Logística	35
Equação 5 — Z - score discretização	47

LISTA DE QUADROS

Quadro 1 — Técnicas e variáveis usadas para predição de PMEs	37
--	----

LISTA DE TABELAS

Tabela 1 — Variáveis primitivas extraídas da base de dados	45
Tabela 2 — Enquadramento das empresas conforme renda anual	47
Tabela 3 — Pressupostos para utilização das técnicas	49
Tabela 4 — Transformação das variáveis dummy para MEI	61
Tabela 5 — Transformação das variáveis dummy para ME	62
Tabela 6 — Transformação das variáveis dummy para PE	63
Tabela 7 — Transformação das variáveis dummy para MéDE	64
Tabela 8 — Transformação das variáveis dummy para DG	65
Tabela 9 — Valores assumidos pela variável dependente RL	66
Tabela 10 — Variáveis inclusas no modelo de RL	67
Tabela 11 — Número para amostra teste e principal de casos válidos	67
Tabela 12 — Teste dos coeficientes de Colinearidade	68
Tabela 13 — Teste de Hosmer e Lemeshow	70
Tabela 14 — Comparativo da acurácia das classes de empresas com a base de DG	75

LISTA DE ABREVIATURAS E SIGLAS

AD	Análise Discriminante
Ar	Artificiais
BACEN	Banco Central
DIEESE	Departamento Intersindical de Estatística e Estudos Socioeconômicos
DT	<i>Decision Tree</i>
DG	Dados Gerais
DRE	Demonstração de Resultados do Exercício
IC	Indicadores Contábeis
IE	Indicadores Econômicos
IF	Indicadores Financeiros
NF	Indicadores Não Financeiros
MPEs	Micro e Pequenas Empresas
MPME	Micro, Pequenas e Médias Empresas
ME	Microempresa
MédE	Média Empresa
MV	<i>Missing Values</i>
P	Privativas
PE	Pequenas Empresas
PMEs	Pequenas e Médias Empresas
PS	Primitivas Substituídas
PU	Primitivas Utilizadas
RL	Regressão Logística
RNA	Redes Neurais Artificiais
SEBRAE	Serviço Brasileiro de Apoio às Micro e Pequenas Empresas
SVM	<i>Support Vector Machine</i>

SUMÁRIO

1	INTRODUÇÃO	15
1.1	CONTEXTUALIZAÇÃO	15
1.2	PROBLEMÁTICA	17
1.3	OBJETIVOS	18
1.3.1	Objetivo Geral	18
1.3.2	Objetivos Específicos	18
1.4	TEMA E JUSTIFICATIVA	19
1.5	DELIMITAÇÃO DO TRABALHO	20
1.6	ESTRUTURA DA DISSERTAÇÃO	20
2	REVISÃO DE LITERATURA	22
2.1	CRÉDITO	23
2.1.1	Mercado Financeiro	23
2.1.2	Intermediários financeiros	23
2.1.3	Sistema Financeiro Nacional - SFN	23
2.1.4	Conselho Monetário Nacional – CMN	24
2.1.5	Banco Central do Brasil (BACEN)	24
2.1.6	Cooperativa de Crédito	25
2.2	RISCO DE CRÉDITO	25
2.3	ANÁLISE DE CRÉDITO	27
2.4	ESTUDOS PIONEIROS	28
2.5	REVISÃO DE ARTIGOS RELACIONADOS	30
2.6	PONTUAÇÃO DE CRÉDITO (<i>CREDIT SCORING</i>)	32
2.7	REGRESSÃO LOGÍSTICA	33
2.7.1	Modelo Matemático da Regressão Logística	34
2.8	VARIÁVEIS INDEPENDENTES E DEPENDENTES	35
2.8.1	Variáveis e técnicas utilizadas em trabalhos de modelagem	36
3	MATERIAIS E MÉTODOS	39
3.1	DELINEAMENTO DO TIPO DE PESQUISA	39
3.2	TÉCNICAS ESTATÍSTICAS MULTIVARIADAS	40
3.3	PROCEDIMENTOS METODOLÓGICOS	42
3.3.1	Análise da base de dados	44
3.3.2	Eliminação de informações incompletas (<i>missings values</i>)	44
3.3.3	Classificação das Variáveis	44

3.3.4	Divisão das classes de empresas	46
3.3.5	Discretização de variáveis	47
3.3.6	Variáveis <i>Dummies</i>	48
3.3.7	Seleção de variáveis – método <i>stepwise</i>	48
3.3.8	Validação e ajuste dos modelos	48
3.3.9	Pressupostos da técnica RL	49
3.3.10	Elaboração dos modelos	49
3.3.11	Teste da acurácia para a Técnicas da RL	49
4	RESULTADOS E DISCUSSÃO	50
4.1	TRATAMENTO DAS VARIÁVEIS INDEPENDENTES	50
4.1.1	Análise das variáveis	50
4.1.2	Discretização das variáveis	50
4.1.3	Transformação das variáveis em <i>dummies</i>	60
4.2	REGRESSÃO LOGÍSTICA	65
4.2.1	Teste de multicolinearidade	67
4.2.2	Teste da Verossimilhança	68
4.2.3	Modelos de pontuação de crédito com a técnica estatística de RL	70
4.2.4	Teste da acurácia para Regressão Logística	72
4.3	DISCUSSÃO DO PODER DISCRIMINATÓRIO DA TÉCNICA RL	76
5	CONSIDERAÇÕES FINAIS	77
	REFERÊNCIAS	79

1 INTRODUÇÃO

A introdução desta dissertação busca descrever a importância do crédito para o mercado produtivo, comercial e de serviços, evidenciando a complexidade no processo de evolução junto às instituições financeiras. Ainda, nesse capítulo, são apresentadas a contextualização, problemática, os objetivos (geral e específicos), o tema e justificativa, a delimitação do trabalho e a estrutura da dissertação.

1.1 CONTEXTUALIZAÇÃO

O crédito tem papel relevante na economia, uma vez que o financiamento é essencial para a evolução dos setores produtivos e para o consumo das famílias. Sua ampliação, no mercado do Brasil, é um dos principais responsáveis para o crescimento da economia, conforme destaca o Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE, 2014).

As Pequenas e Médias Empresas (PMEs), têm desempenho econômico relevante, conforme destaca o anuário do trabalho divulgado pelo Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE; DIEESE 2015). Em 2013, 67,40% dos empregos foram gerados nas PMEs, nos setores da indústria, construção, comércio e serviços, gerando mais de 22 milhões de empregos (SEBRAE; DIEESE, 2015). Somente no período acumulado de janeiro a julho de 2018, os pequenos negócios responderam pela criação de 395,3 mil postos de trabalho (SEBRAE, 2018). Essas empresas representam a espinha dorsal da economia de todos os países (GORDINI, 2014).

Por outro lado, o crédito é destacado pelo risco proporcionado pelo empréstimo (ou financiamento). A preocupação com o fornecimento de crédito e sua importância para sustentabilidade do sistema comercial remonta ao ano de 1938, quando a instituição “*National Bureau of Economic Research*” promoveu estudos em instituições de fomento de empréstimos nos quais buscou-se criar um histórico do crédito de bons e maus pagadores utilizando medidas estatísticas, a fim de identificar o risco de uma operação de crédito (DURAND, 1941).

Os modelos de predição de risco de crédito ganharam maior relevância em épocas de crises econômicas, como as ocorridas próximas e pós década de 1990 nos Estados Unidos, México, Brasil, Rússia, entre outros países. Conforme Pignata e Carvalho (2015), a eclosão de crises financeiras, econômicas e políticas, para os países que utilizam do sistema capitalista, é uma dura realidade, tendo algumas maiores abrangências, causando fortes impactos, afetando grande parte da população em escala mundial.

Na última década, mais especificamente no ano de 2008, o Brasil foi atingido novamente por mais uma crise internacional, chamada de crise do *subprime*, deflagrada nos

Estados Unidos, a qual aprofundou-se, havendo fuga de capitais e desvalorização das moedas das economias periféricas. O agravamento da crise financeira internacional, levou ao extremo a aversão aos riscos pelas instituições financeiras (PRATES; FARHI, 2009).

As instituições financeiras, visando melhorar a confiança no sistema financeiro, na tentativa de evitar ou mitigar crises econômico-financeiras, buscam emitir práticas e técnicas que visem aumentar a segurança ao mercado. Nesse contexto, em 1988, na Suíça foi lançado um acordo, a serem observados pelas organizações creditícias, conhecido como: *International Convergence of Capital Measurement and Capital Standards*, mas difundiu-se no Brasil como Basileia I. O 2º acordo, denominado ‘Basileia II’, foi lavrado em 2004. Recentemente, o 3º acordo (“Basileia III”) foi estabelecido em 2010, sendo fruto das crises de 2008 e 2009, o qual implantou regras mais rígidas (MENDES, 2013).

Os principais objetivos da Basileia III são: melhorar a capacidade dos bancos de absorver choques decorrentes de estresse financeiro e econômico, qualquer que seja a fonte causadora; e aprimorar as práticas de gestão e governança de riscos e fortalecer a transparência e as práticas de divulgação, conforme descreve a Empresa de Auditoria Price Water House Coopers (PWC, 2013).

O Brasil, por meio da Resolução 4.557 de 2017, em observação ao Relatório de Estabilidade Financeira (BACEN, 2017), demonstra seu alinhamento com as normas emanadas pelo Comitê da Basileia. Um desses alinhamentos, refere-se aos ativos problemáticos, os quais são definidos como as exposições em atraso há mais de 90 dias e aquelas que apresentem indicativos de que não serão honradas, sem que haja a necessidade de ampliação das garantias ou a colaterais.

Consta ainda desse relatório, a análise das operações entre 15 e 90 dias, que constitui um indicador antecedente de inadimplência, o qual registrou tendência de alta generalizada ao longo do primeiro semestre de 2016, sendo estabilizada no segundo semestre. Esse relatório destaca ainda que em outubro 2016, a inadimplência das PMEs atingiu o maior valor da série histórica, influenciado, no segundo semestre pela redução do estoque de crédito.

Assim, pode-se observar a importância das PMEs para o setor econômico e as preocupações quanto a “fragilidade financeira” junto as instituições fomentadoras de crédito. Para Galinari, Costa e Teixeira (2016), as MPMEs, são atores de relevância na economia brasileira, apresentam-se em todo o território nacional, sendo estratégicas regionalmente como fonte natural de empreendedorismo e geração de inovações.

Nesse contexto, a relevância da informação e mecanismos que auxiliem na tomada de decisão, a fim de mitigar o risco de crédito, atraem, cada vez mais, o interesse das instituições financeiras e de pesquisa em estudos relacionados com o sistema de pontuação de crédito. Martens *et al.* (2010), destacam que as instituições financeiras, nas últimas décadas, viram uma crescente necessidade de utilização de técnicas de análise quantitativa, visando a otimização e monitoramento das decisões relacionadas com o risco e o investimento em

gestão.

Esses sistemas trazem contribuição para o desenvolvimento de modelos de pontuação de crédito, tornando sua utilização mais prática e fácil. As Técnicas de pontuação de crédito ou escala de classificação, visam reduzir a probabilidade de inadimplência, oferecido pelo cliente, visando maximizar o lucro esperado pelas instituições financeiras (ABDOU; POINTON, 2011).

Como já exposto, a alteração do mercado financeiro e sua tentativa de prever a inadimplência, por meio de técnicas de pontuação de crédito, tem utilizado diferentes abordagens. Mário (2017), relatam os avanços nos estudos de modelos de ranqueamento de crédito denominado “*Credit Scoring Models*”, principalmente com a contribuição da técnica da Análise Discriminante (AD), a qual tem por objetivo, a classificação do risco do cliente no momento da concessão.

Outra técnica, descrita por Dias Filho e Corrar (2017), na elaboração da pontuação de crédito é a Regressão Logística (RL), a qual é recomendada em situações nos quais a variável dependente é de natureza dicotômica apresentando pequeno grau de complexidade operacional e também apresenta ausência de restrições mais rígidas.

Por outro lado, mais recentemente as técnicas computacionais, têm apresentado resultados promissores. Louzada, Ara e Fernandes (2016), por meio de uma revisão da literatura, expôs que as técnicas de Redes Neurais Artificiais (RNA) e *Support Vector Machine* (SVM) e outras combinações híbridas aparecem como as principais ferramentas mais recentes e as técnicas de Árvores de Decisão (DT) e RL são bastante utilizadas como meio comparativo para outras técnicas.

Diante do exposto, a formulação de modelos de crédito, para o suporte das decisões, quanto ao fornecimento ou não do crédito para as PMEs, por parte das instituições financeiras ou cooperativas de crédito, podem contribuir na melhoria da decisão técnica, reduzindo a subjetividade. Assim, este trabalho tem o intuito de desenvolver modelos de pontuação de crédito, para as 4 classes de empresas, sendo: MEI, ME, PE e MédE e uma para Dados Gerais (DG), utilizando uma base de dados de uma cooperativa de crédito, por meio da técnica estatística RL.

1.2 PROBLEMÁTICA

Uma das ferramentas essenciais na alavancagem das transações comerciais, destacadas na passagem do tempo das sociedades organizadas, chama-se “crédito”. Silva (2016) destaca que foram encontradas evidências de empréstimos contratados na antiga Babilônia (2.000 a.C.) no qual um fazendeiro captou recursos financeiros, os quais foram devolvidos após sua colheita, com o pagamento do capital e juros.

Na construção da sociedade organizada, entidades como bancos e posteriormente

cooperativas de crédito, realizam a intermediação financeira, captando recursos generalizados e fomentando pessoas e empresas para que estas adquiram bens e serviços necessários à consecução de seus objetivos. O fornecimento de valores a clientes, sob a condição creditícia, implica em problemas como o risco de inadimplência, isto é, o risco de não receber os valores emprestados (total ou parcial).

Deste risco, nasce a fundamental necessidade da avaliação do cliente e das variáveis que envolvem tal operação. Assim, torna-se crucial, para as instituições financeiras, avaliar o risco de crédito, devido aos elevados riscos associados à credibilidade de concessão inadequada (KAO; CHIU; CHIU, 2012).

O conseqüente crescimento e expansão do mercado internacional, caracterizado de globalização, aumenta a problematização da concessão do crédito. Altman, Caoquette e Narayanan (1998), destacam a posição vital que o crédito assume perante as intermediações financeiras junto ao mercado globalizado e o surgimento de corporações especializadas para avaliação do risco.

A busca por modelos de previsão de insolvência, é relatado por Gepp, Kumar e Bhattacharya (2010), destacando que, torna-se imperativo questionar o longo prazo das avaliações de risco de crédito e que mesmo em tempos de dificuldades as instituições buscam ferramentas alternativas, funcionais e confiáveis para avaliar o risco de inadimplência de forma adequada.

Restringindo-se as atividades inerentes às PMEs e a responsabilidade que as instituições financeiras, na figura do analista de crédito, têm em fomentar o mercado, essa dissertação estabelece a seguinte situação problema: Com as variáveis disponíveis na base de dados da instituição financeira (cooperativa de crédito), é possível elaborar modelos de pontuação de crédito, para as 4 classes de empresas e DG, com o auxílio da técnica estatística Regressão Logística (RL), para contribuição na análise de crédito?

1.3 OBJETIVOS

1.3.1 **Objetivo Geral**

Elaborar modelos de pontuação de crédito destinados às entidades de fomento na análise das solicitações de crédito de pequenas e médias empresas, identificando riscos inerentes a operação e classificando-os como prováveis inadimplentes ou adimplentes.

1.3.2 **Objetivos Específicos**

Os objetivos específicos do trabalho são:

- a) Identificar as variáveis mais relevantes na identificação do risco dentro da construção dos modelos propostos;
- b) Desenvolver os modelos de pontuação de crédito, com a técnica da Regressão Logística e
- c) Verificar se a acurácia dos modelos por classe de empresas é superior ao modelo dos dados gerais.

1.4 TEMA E JUSTIFICATIVA

A atividade creditícia, em um país, fomenta os empreendimentos. Isso reflete indiretamente na melhoria das condições de vida de uma população, chegando, em 2014, a representar 58,9% do Produto Interno Bruto (PIB), conforme Relatório de Economia Bancária e Crédito, emitido pelo Banco Central do Brasil (BACEN).

Por meio de um ciclo econômico de redes, os recursos proporcionados pelo crédito impulsionam a movimentação da cadeia produtiva, consumindo matéria-prima, gerando produtos, utilizando mão de obra (serviços) e recolhendo tributos. Destaca Fonseca *et al.* (2013), que o processo da economia de produção passa necessariamente pelos meios monetários, sendo afetados diretamente pela atuação das Instituições vinculadas ao Sistema Financeiro Nacional (SFN).

A valoração dos meios monetários, encontra-se no Boletim Responsabilidade Social e Ambiental do SFN (2010), o qual destaca a importância do crédito, como uma mola propulsora de qualquer meio de desenvolvimento, a qual já se encontra consagrada nas escolas da teoria econômica.

A expansão dessa atividade traz consigo o fenômeno do risco de não recebimento do valor emprestado e os juros incorporados, os quais são designados como “lucro” da operação. Conforme destaca Magro, Mondini e Hein (2015), o risco do não recebimento (a inadimplência), força as instituições financeiras a buscarem ferramentas cada vez mais sofisticadas. A frequência da fraude de crédito está aumentando, o que gera prejuízos às instituições, sendo importante que os bancos avaliem com precisão o risco incorrido (ZHANG *et al.*, 2008).

No universo das instituições de crédito, também se encontram as cooperativas de crédito (SOARES; BALLIANA, 2009). Segundo esses autores, o cooperativismo de crédito apresenta-se como instrumento impulsionador dos setores econômicos estratégicos. Contudo, uma das maiores ameaças para a sobrevivência de uma instituição financeira, tipificada como cooperativa de crédito é o risco de inadimplência, pois essa instituição promove o acesso a serviços financeiros e assume os riscos correspondentes (SOARES; BALLIANA, 2009).

A realidade brasileira, quanto à informação contábil, assume outro panorama. Segundo Berti (2012), a análise por meio das demonstrações contábeis, das MPE's, apresenta o

inconveniente das estruturas contábeis não conseguirem refletir a realidade das transações. Nesse sentido, Berger, Cowan e Frame (2011), relatam a falta de auditoria das demonstrações financeiras, a fim de dar credibilidade em base regular.

Segundo Li, Sun e Wu (2010), prever o fracasso das empresas que tomam crédito, é crucial para as instituições financeiras. Na tentativa de evitar ou controlar o risco de inadimplência, as instituições buscam formas de fornecer crédito aos clientes ou cooperados, que proporcionem o menor risco, para tanto, os analistas buscam formas de reduzir a subjetividade.

As técnicas de predição de risco, tanto as estatísticas como computacionais, são ferramentas que podem auxiliar o especialista fornecendo informações no auxílio a tomada de decisão. Segundo Camargos, Camargos e Araújo (2012), há a necessidade de aprimoramento de mecanismos robustos de análise de crédito para redução dos níveis de inadimplência e a relevância do desenvolvimento de modelos para minimizar a inadimplência. Desta forma, construir um modelo para predição do risco, o qual forneça informações e auxilie na tomada de decisão, torna-se de vital importância para as instituições.

1.5 DELIMITAÇÃO DO TRABALHO

Neste estudo, serão construídos 4 modelos de pontuação de crédito, sendo: Microempreendedor Individual (MEI), Microempresa (ME), Pequena Empresa (PE), Média Empresa (MédE) e um modelo para a base de Dados Gerais (DG) com base em uma amostra de 1.491 contratos em vigor (até 10/10/2017), obtidos de um banco de dados fornecidos por uma instituição cooperativa de crédito. Ressalta-se que, neste trabalho, quando referido a abreviatura PMEs, compreender-se-á como inclusas as empresas caracterizadas como: MEI, ME, PE e MédE.

Outra delimitação é o uso de apenas técnicas estatísticas para a construção de modelos. Assim, técnicas computacionais ou de algoritmos genéticos, não são focadas neste estudo.

1.6 ESTRUTURA DA DISSERTAÇÃO

Este trabalho está estruturado em 5 capítulos. Após a introdução, o Capítulo 2 apresenta a Revisão de Literatura, conceituando: Crédito, Risco de Crédito, Análise de Crédito, Estudos Pioneiros, Revisão de artigos relacionados, Pontuação de Crédito (*Credit Scoring*), Regressão Logística, Variáveis Dependentes e Independentes. No Capítulo 3, é apresentado os Materiais e Métodos, conceituando: Delineamento do Tipo de Pesquisa, Técnicas Estatísticas Multivariadas, Procedimentos Metodológicos. Posteriormente, o Capítulo 4 apresenta, os Resultados e Discussão, com o conteúdo: Tratamento das Variáveis

Independentes, Regressão Logística, Discussão do poder discriminatório da técnica RL. No Capítulo 5 é descrito as Considerações finais e por último são apresentadas as referências bibliográficas utilizadas neste trabalho.

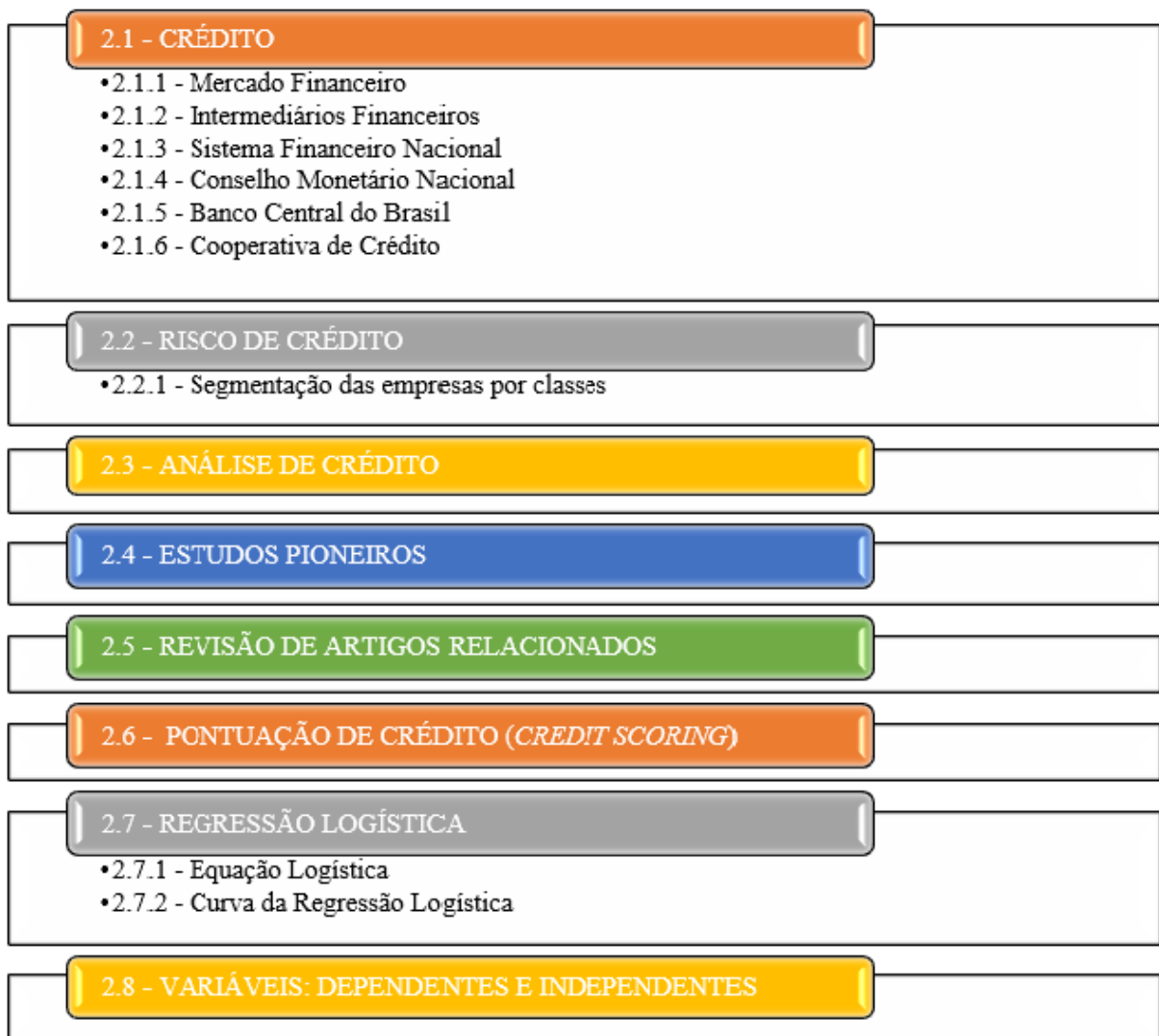
2 REVISÃO DE LITERATURA

Esta revisão de literatura buscou apresentar os referenciais teóricos necessários para embasamento desta pesquisa, relacionando-os com o objetivo geral.

Conforme ilustra o Figura 1, o Capítulo 2, está dividido em títulos e subtítulos, sendo que os títulos são: Crédito, Risco de Crédito, Análise de Crédito, Estudos Pioneiros, Revisão de Artigos Relacionados, Pontuação de Crédito (*Credit Scoring*), Regressão Logística e Variáveis Dependentes e Independentes.

Os títulos Crédito apresenta 6 subtítulos, Risco de Crédito apresenta 1 subtítulo e Regressão Logística 2 subtítulos.

Figura 1 - Distribuição dos tópicos da Revisão de Literatura



Fonte: Elaborada pelo autor

2.1 CRÉDITO

2.1.1 Mercado Financeiro

O mercado de crédito é o local destinado às negociações entre pessoas físicas e jurídicas com as instituições creditícias (sociedades de crédito). Essas instituições concedem crédito para suprimento das necessidades de capital de giro e financiamento de bens. Os principais produtos negociados nesse mercado, segundo Hoji (2010, p. 31) são: “empréstimos para capital de giro, descontos de títulos, conta garantida, assunção de dívidas, [...] financiamentos de importação e financiamento de serviços e bens de consumo duráveis”.

As instituições financeiras, além de captar o dinheiro excedente, emprestam para quem precisa. Essas instituições criaram o conceito de mesa de operações e de empréstimos, sendo que esta repassa os recursos captados aos agentes deficitários e aquela está ligada a captação de recursos como o Certificado de Depósito Bancário (CDB) e o Recibo de Depósito Bancário (RDB) (SALAZAR, 2012).

2.1.2 Intermediários financeiros

Os intermediários financeiros são compostos por instituições vocacionadas a trabalhar com recursos financeiros, contabilmente conhecidos como caixa ou equivalentes de caixa. Para Silva (2016), as instituições financeiras geram condições de acesso aos tomadores de recursos e aos emprestadores as condições para aplicarem seus recursos. Desta forma, essa satisfação financeira, gera aumento da economia e consequente produtividade.

De acordo com o Banco Central do Brasil (BACEN, 2017), o Brasil possui 1.761 instituições de fomento de crédito. Essas atuam no mercado financeiro, em 18 categorias, incluindo nessa categoria as Cooperativas de Crédito, as quais correspondem a 1.042 instituições. Ainda, segundo o BACEN (2017), essas instituições, na sua atividade captam recursos na forma de depósitos à vista e repassam a seus clientes em forma de contas de depósito, os quais utilizam para movimentação de recursos e pagamentos.

No mercado financeiro, os agentes econômicos que utilizam os recursos, em geral não tem a expertise necessária para avaliar os riscos decorrentes. Neste sentido, os intermediários financeiros, devem estar mais preparados para essa tarefa, ampliando a segurança dos clientes aplicadores de recursos (SILVA, 2016).

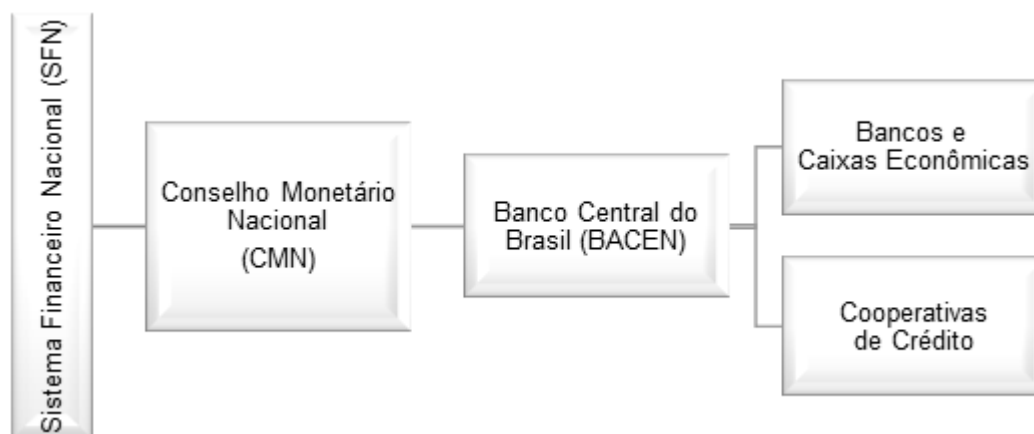
2.1.3 Sistema Financeiro Nacional - SFN

No Brasil, a estrutura responsável pelo fluxo financeiro da moeda e do crédito,

chama-se Sistema Financeiro Nacional (SFN). O SFN foi estruturado em 1964, sendo os seus principais órgãos o Conselho Monetário Nacional (CMN) e Banco Central do Brasil (BACEN, 2017).

O SFN é formado por um conjunto de instituições, as quais têm como finalidade intermediar o fluxo de recursos entre os poupadores e investidores e os tomadores de recursos, em condições satisfatórias para o mercado (HOJI, 2010). A Figura 2 destaca parte da estrutura do SFN, relacionado com a estrutura normativa dos bancos e cooperativas de crédito.

Figura 2 - Estrutura Sistema SFN para bancos e cooperativas



Fonte: Elaborada pelo autor

2.1.4 Conselho Monetário Nacional – CMN

O CMN é o órgão superior do SFN e tem a responsabilidade de formular a política da moeda e do crédito, objetivando a estabilidade da moeda e o desenvolvimento econômico e social do País. O CMN foi criado pela Lei nº 4.595, de 31 de dezembro de 1964, sendo efetivamente instituído em 31 de março de 1965, uma vez que o art. 65 dessa Lei estabeleceu a entrada em vigor 90 dias após sua publicação. A atual composição do CMN é formada por: (i) Ministro da Fazenda, como Presidente do Conselho; (ii) Ministro do Planejamento, Desenvolvimento e Gestão; e (iii) Presidente do Banco Central do Brasil (BACEN, 2017).

2.1.5 Banco Central do Brasil (BACEN)

O BACEN é uma autarquia federal vinculada ao Ministério da Fazenda e responsável pelo controle da inflação no país. Ele atua de forma a regular a quantidade de moeda na economia buscando a estabilidade de preços. Além disso, as atividades do BACEN incluem a preocupação com a estabilidade financeira. Para isso, o BACEN regula e supervisiona as instituições financeiras. (BACEN 2018).

O BACEN também conduz as políticas monetária, cambial, de crédito e de relações financeiras com o exterior; a regulação e da supervisão do SFN; a administração do Sistema de Pagamentos Brasileiro (SPB); e os serviços do meio circulante. Dentre as suas principais atribuições do BACEN estão: (i) emitir papel-moeda e moeda metálica; (ii) executar os serviços do meio circulante; (iii) receber recolhimentos compulsórios e voluntários das instituições financeiras; (iv) realizar operações de redesconto e empréstimo às instituições financeiras; (v) regular a execução dos serviços de compensação de cheques e outros papéis; (vi) efetuar operações de compra e venda de títulos públicos federais; (vii) exercer o controle de crédito; (viii) exercer a fiscalização das instituições financeiras; (ix) autorizar o funcionamento das instituições financeiras; (x) estabelecer as condições para o exercício de quaisquer cargos de direção nas instituições financeiras; e (xi) vigiar a interferência de outras empresas nos mercados financeiros e de capitais e controlar o fluxo de capitais estrangeiros no país.

2.1.6 Cooperativa de Crédito

Conforme disposto pelo BACEN, a Cooperativa de Crédito é uma instituição financeira formada pela associação de pessoas para prestar serviços financeiros exclusivamente aos seus associados. Os cooperados são ao mesmo tempo proprietários e usuários da cooperativa, participando de sua gestão e usufruindo de seus produtos e serviços.

As cooperativas de crédito ofertam aos seus associados, serviços iguais aos de outras instituições financeiras, diferenciando-se somente na forma de relacionamento. A finalidade do cooperativismo, não é o lucro, mas a inclusão financeira das pessoas no mundo, independentemente dos fatores de origem ou classe social (SANTOS; SANTOS; SANTOS, 2016).

Por meio de uma cooperativa de crédito, os meios de produção mais desfavoráveis, tem uma oportunidade de obter atendimento personalizado para suas necessidades. O cooperativismo surge como a “resposta de mercado” para o financiamento das pequenas e médias empresas, mobilizando um arranjo institucional (FONSECA *et al.*, 2013).

No tocante a essas cooperativas, é importante ressaltar que a relação existente entre ela e seus associados, causam responsabilidades sociais e financeiras, participando o associado das sobras e dos prejuízos. Essa responsabilidade pode ser limitada ou ilimitada, ou seja, extensiva a suas cotas ou não, conforme dispõe a Lei 5.764 de 1971 e os dispositivos do Código Civil Brasileiro.

2.2 RISCO DE CRÉDITO

De acordo com a Resolução do BACEN, nº 3.721/09, risco de crédito é definido como

“a possibilidade de ocorrência de perdas associadas ao não cumprimento pelo tomador ou contraparte de suas respectivas obrigações”. Conforme Brito e Assaf Neto (2008), o crédito gera a expectativa do recebimento de um determinado valor em um espaço de tempo, sendo o risco entendido como a possibilidade de incursão em perdas pelo credor devido a não serem cumpridas as obrigações por parte do tomador.

Gestão de risco de crédito compreende o controle do risco dentro de parâmetros aceitáveis, usando para esta finalidade uma análise quantitativa do portfólio de seus clientes, predizendo a capacidade de reembolso futuro pelos clientes (VAHID; AHMADI, 2016). Ainda Salazar (2012), descreve que o risco é a probabilidade do credor receber como retorno sobre o investimento um resultado aquém das expectativas esperadas.

Cabe a análise de crédito avaliar o volume de informações pertinentes a solicitação do crédito pela área de vendas, determinar parâmetros, medir riscos e emitir parecer formalizado sobre a solicitação. A análise deve ser periódica, pois o mercado é dinâmico e em alguns casos pode causar movimentos relevantes de massa, quanto aos aspectos de pontualidade, capacidade de pagamento e situação financeira (HOJI, 2003). As percepções baseadas em intuição, experiência, desconfiança, não são suficientes para orientar decisões que envolvam risco e retorno (SILVA, 2004).

Quanto à questão informacional, Menezes e Riccio (2005), reclassificam a mesma incluindo-a como um dos itens que compõe o conhecimento e destacando que as informações estão disponíveis em abundância e quando esta é editada surge o conhecimento. Para as entidades, uma das principais fontes de entrada de dados que gera a informação é o cadastro, o qual concentra um volume de informações relevantes, reunindo fontes pessoais, comerciais, institucionais, contábil, entre outras. Assim, a ficha cadastral deve conter informações importantes para efetiva contribuição na análise e concessão do crédito (SILVA, 2008).

A contabilidade é outra fonte essencial de geração de dados e informações, a qual possibilita a extração de uma visão a curto e longo prazo. Destaca o Comitê de Pronunciamento Contábil (CPC 00), que “a Informação contábil-financeira relevante é aquela capaz de fazer diferença nas decisões que possam ser tomadas pelos usuários, [...] valor preditivo, valor confirmatório ou ambos”. Estas informações são utilizadas para efetuar a análise de balanço por meio de indicadores patrimoniais e financeiros e ainda realizar a análise dinâmica (BERTI, 2012).

Devido as características específicas de cada nível ou setor de atuação das empresas, voltadas a realidades específicas da área, se faz necessário informação adequada. Segundo Alvim (1998), é importante a gestão ter informação adequada à determinada necessidade, em tempo e custo compatíveis. Da mesma forma, Rodrigues e Blattmann (2014), destacam a importância de conhecer a origem das informações que envolvem o ambiente que a organização se encontra inserida, ainda destacando as variações em formatos, conteúdos e natureza que influenciam no processo de uso da informação.

2.3 ANÁLISE DE CRÉDITO

A avaliação do risco tem por objetivo melhorar a qualidade de uma carteira de clientes, favorecendo uma venda saudável e evitando ao máximo a perda de valores, sobre créditos fornecidos de forma equivocada ou a clientes que geram prejuízos aos negócios. Empresas que possuem boa avaliação, levam vantagens sobre seus concorrentes (GONÇALVES, 2005).

Ao longo dos anos, muitos administradores de crédito buscaram uma forma de reduzir o processo de análise de crédito a uma fórmula numérica. Entretanto, até o desenvolvimento dos computadores, poucos avanços foram feitos na análise de grandes massas de dados (GONÇALVES; GOUVÊA; MANTOVANI, 2013).

A base informacional em sua grande maioria utiliza-se de fontes contábeis, exigindo confiabilidade da informação. Destaca o CPC “00”, que o provimento de informação fidedigna e relevante, por parte da contabilidade, pode melhorar a confiança do usuário e assim contribuir para a promoção da estabilidade econômica.

Conforme já descrito no item 1.4, a realidade brasileira, quanto à informação contábil, assume outro panorama. Segundo Berti (2012), a análise por meio das demonstrações contábeis, das MPE’s, apresenta o inconveniente das estruturas contábeis não conseguirem refletir a realidade das transações. Nesse sentido, Berger, Cowan e Frame (2011), relatam a falta de auditoria das demonstrações financeiras, a fim de dar credibilidade em base regular.

O crescimento das organizações, que tem o recurso financeiro como produto, é dependente de uma análise de crédito eficiente. A correta tomada de decisão, ao decidir sobre concessão de crédito, é essencial para a sobrevivência das instituições (STEINER *et al.*, 1999).

Um dos meios utilizados para analisar o cliente, solicitante do crédito, são os C’s do crédito. Os 5 principais C’s do crédito são: Caráter, Capacidade, Capital, Colateral e Condições. Para Hoji (2010), os 5 C’s do crédito devem ser analisados conjuntamente, conforme descrito:

- **Caráter** - "É o item mais importante na análise de crédito. A avaliação da cultura da empresa e do caráter de seus administradores, apesar de alto grau de subjetividade, por se referir a aspectos morais e éticos [...]".
- **Capacidade** - "A firme determinação de pagar (caráter) não terá validade se o cliente não tiver capacidade de saldar seus compromissos financeiros. O potencial de o cliente saldar os compromissos financeiros pode ser obtido por meio da análise das demonstrações financeiras [...]".
- **Capital** - "A análise da estrutura econômica e financeira evidencia o nível de solidez financeira da empresa. Analisa-se, por exemplo, o nível de imobilização em

relação a patrimônio líquido, o nível de investimentos em máquinas, equipamentos e tecnologias[...]"

- **Colateral** - "É uma palavra em inglês que significa garantia. O colateral é utilizado para contrabalançar ou reforçar a fragilidade de um ou mais "C" dos quatro outros itens. Pode ser representado por ativos tangíveis ou ativos financeiros, deste que cubram o período de crédito concedido".
- **Condições** - " As condições econômicas atuais e o cenário econômico em que a empresa estará inserida devem ser avaliados em conjunto com o ramo de atividade em que ela atua. Se for esperada forte recessão, mas o ramo de atividade em que a empresa atua for substancialmente promissor, o risco de inadimplência será minimizado". (HOJI, 2010, p. 128–129).

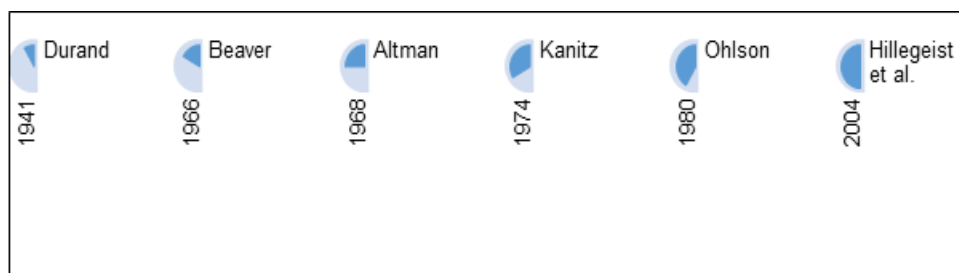
Junto aos C's do crédito, Stell (2009), incorpora a questão da "mudança" a qual descreve o desafio de enfrentar a mudança e a capacidade de transição do ponto A para o ponto B. Ainda segundo Stell (2009), a mudança é uma oportunidade ou um obstáculo, pois cada indivíduo faz sua escolha. Refere-se a atitude e é uma opção pessoal que pode expor vulnerabilidade ou perda de controle.

Para Stell (2009), a concessão do crédito está ligada ao objetivo e a condição da aplicação dos recursos emprestados, pois, a empresa deve estar ciente do estado da arte de onde queira utilizar o empréstimo. O bom julgamento é um trunfo do credor o qual é aumentado quando possui ferramentas analíticas quantitativas e qualitativas como os C's do crédito, incluindo a mudança.

2.4 ESTUDOS PIONEIROS

A busca por instrumentos que possam sinalizar alterações ou destacar desvios de padrão dentro do universo creditício não é de hoje, mas pode-se observar a intensificação dos estudos a partir do século XX. A Figura 3 destaca os principais autores sobre o assunto de previsão de falências.

Figura 3 - Estudos pioneiros na previsão de falências



Fonte: Elaborada pelo autor

A preocupação com o fornecimento de crédito e sua importância para sustentabilidade do sistema comercial já era presente em 1938. Nessa data, a instituição “*National Bureau of Economic Research*“, promoveu estudos em instituições de fomento de empréstimos nos quais buscou-se criar um histórico do crédito de pagamentos bons e ruins utilizando medidas estatísticas a fim de identificar o risco do crédito (DURAND, 1941).

O estudo de Beaver (1966), denominado “*Financial ratios as predictors of failure*“, objetivou quantificar quais índices poderiam identificar com melhor clareza a insolvência de instituições jurídicas (empresas). Beaver utilizou um volume de 79 empresas identificadas com o critério de não-pagamento de dividendos (falência) e comparou suas demonstrações contábeis com outras 79 empresas solventes, as quais possuíam situação financeira saudável. Seus estudos compuseram inicialmente 30 indicadores, sendo que destes foram selecionados 14. No Grupo I (índices de fluxo de caixa), como fluxo de caixa para vendas e para dívida total etc. No Grupo II (Rendimento Líquido), como lucro líquido das vendas e para dívida total etc. No grupo III (Dívidas ao ativo total), Passivo circulante para ativos totais, passivos de longo prazo para ativos totais etc. Grupo IV (liquidez), dinheiro no ativo total, recursos rápidos para ativos totais, ativo circulante para ativos totais e capital de giro. As variáveis foram dicotômicas. A análise foi baseada na observação dos índices projetados sobre as demonstrações contábeis. Para seu desenvolvimento Beaver calculou as médias das razões para as empresas com falhas (falência ou dificuldades financeiras) e para as empresas sem falhas em cada um dos anos anteriores à falha. Essa comparação dos valores médios, foi chamado de análise de perfil.

Em 1968, o trabalho de Altman denominado “*Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy*“, utilizou técnicas estatísticas multivariadas na elaboração de seu modelo. A amostra inicial foi composta por 66 empresas, com 33 empresas em cada um dos dois grupos: falidos e não falidos. As empresas falidas eram fabricantes que apresentaram uma petição de falência. Foram utilizadas no modelo vinte e duas variáveis, potencialmente úteis, oriundas de cinco categorias: liquidez, rentabilidade, alavancagem, solvência e índices de atividade. A escolha dos índices teve como base a popularidade na literatura e a potencialidade da elaboração de novos índices. Altman, com a utilização do método atingiu uma acurácia superior a 90%. Em suas observações destaca que a principal preocupação não é com relações em si, mas com os dados contábeis que compõem os índices.

Em 1972, Deakin, propôs um modelo alternativo para prever falhas, utilizando como base os trabalhos de Beaver, com um teste de classificação dicotômica e Altman, 1968 com a técnica AD. Concluiu que a aplicação de técnicas estatísticas, particularmente a análise discriminante, pode ser usada para prever a falha de negócios com dados contábeis com até três anos de antecedência. Deakin, com a aplicação da função discriminante, atingiu uma previsão de 90%. Para composição do modelo utilizou indicadores elaborados com dados retirados de relatórios financeiros.

No Brasil, Stephen Charles Kanitz, em 1974 publicou o artigo denominado de “Como Prever Falências”, na Revista Exame, o qual introduziu no Brasil o conceito de Pontuação de Crédito, internacionalmente conhecido como: “*Credit Scoring*”. Por meio de seus estudos em um universo de aproximadamente 5.000 empresas, selecionou, de forma aleatória, 21 empresas que havia falido entre 1972 e 1974, fazendo a análise de suas demonstrações contábeis. Desenvolveu uma fórmula que ficou conhecida como Termômetro de Insolvência, utilizando como variáveis os índices de Rentabilidade do Patrimônio, Liquidez Geral, Liquidez Seca, Liquidez Corrente e Grau de Endividamento. Usou a técnica da AD para estimar os pesos multiplicadores de sua fórmula.

Na década de 1980, Ohlson, no artigo “*Financial ratios and the probabilistic prediction*”, trabalhou com a técnica estatística denominada de RL, com uma amostra de 105 empresas industriais com situação de falência. Utilizou como variáveis, os indicadores: passivo dividido pelo ativo total, capital de giro dividido pelo ativo total, lucro líquido dividido pelo ativo total, geração bruta de caixa dividido pelo passivo entre outros. Este modelo também conseguiu uma acurácia de previsão superior a 90%. Ohlson relata a fácil aplicação do modelo. O trabalho de Ohlson foi um dos primeiros a utilizar a RL para modelo previsão de risco.

Em 2004, Hillegeist, Cram, Lundstedt, utilizaram a Teoria de Precificação de Opções, para o desenvolvimento do artigo “*Assessing the Probability of Bankruptcy*”, para fazer a previsibilidade de falência. No artigo de Hillegeist, Cram, Lundstedt, avaliaram medidas baseadas em contabilidade, o Z-Score de Altman (1968) e o O-Score de Ohlson (1980), efetuando comparações com uma medida da probabilidade de falência, baseada no mercado, desenvolvida por eles próprios, com base no modelo de precificação de opções “*Black-Scholes-Merton*” (BSM-Prob). Utilizaram uma amostra de 756 empresas falidas no período 1980-2000. Os resultados demonstraram que este modelo fornece mais informações sobre a probabilidade de falência do que outras medidas baseadas apenas em índices de contabilidade.

A análise dessas publicações permite observar que as principais variáveis para composição dos modelos desenvolvidos entre 1941 a 2004, foram as informações registradas pela contabilidade para entidades tidas como pessoas jurídicas. Além disso, as principais técnicas estatísticas utilizadas foram AD e RL.

2.5 REVISÃO DE ARTIGOS RELACIONADOS

Os autores: García, Marqués e Sánchez (2014), apresentaram relevante trabalho de revisão, publicado no “*Journal of Intelligent Information Systems*”, no qual foi efetuado uma pesquisa nos documentos publicados entre 2000 a 2013, com volume superior a 140 artigos. O objetivo da pesquisa foi estudar como os experimentos foram projetados e os resultados validados no campo do risco de crédito e da previsão de falência corporativa. Esses autores

destacam a importância de ao menos ter quatro fases de forma bem definida para extrair conclusões com boa fundamentação dos resultados, sendo: (i) dados experimentais, (ii) técnicas de divisão de dados, (iii) avaliação de desempenho e (iv) testes estatísticos de significância.

Quanto aos dados experimentais, os autores identificam algumas deficiências, a saber: (i) limitação da base pública de dados; (ii) número reduzido de dados da amostra, o que pode aumentar a variação dos resultados; e (iii) utilização de um banco de dados exclusivo e único, aumentando a cautela quando das conclusões. Referente a divisão dos dados, o fato de maior relevância encontra-se na falta de especificação da técnica de divisão de dados empregados, dificultando a reprodução dos experimentos. (GARCÍA; MARQUÉS; SÁNCHEZ, 2014)

No que se refere a avaliação de desempenho, a maioria dos artigos utilizaram a precisão ou a taxa de erro, mesmo com dados desequilibrados de classes e diferentes custos de classificação errada. Esses autores, descrevem ainda, que a preocupação do comportamento tendencioso da precisão pode induzir os pesquisadores a conclusões enganosas do modelo de predição.

Por fim, relativo aos testes estatísticos de significância, esses autores descrevem que os mesmos não são muito frequentes. Assim, no desenvolvimento do trabalho verificaram que alguns estudos apresentaram apenas o desvio-padrão e ausência do teste de hipóteses, sendo que outros aplicaram apenas um teste paramétrico (principalmente o teste t) sem verificar a normalidade dos dados.

Em outro trabalho relevante, Louzada, Ara e Fernandes (2016), intitulado “*Classification methods applied to credit scoring: Systematic review and overall comparison*”, publicado no periódico “*Surveys in Operations Research and Management Science*”, realizaram uma revisão histórica de mais de 20 anos de pesquisa, incluindo 187 artigos entre os anos de 1992 a 2015. O objetivo principal desse estudo foi realizar uma revisão sistemática da literatura relativa à teoria e aplicação de técnicas de classificação binária para análise financeira de pontuação de crédito.

Quanto a revisão, esses autores observaram que o mais comum quanto aos objetivos foi propor um novo modelo de classificação de pontuação de crédito, principalmente com técnicas híbridas buscando melhor o desempenho. De acordo com os resultados de Louzada, Ara e Fernandes (2016), as técnicas mais utilizadas foram: Redes Neurais Artificiais (RNA), Máquina de Suporte Vetor (*Support Vector Machines* – SVM), sistemas híbridos e técnicas combinadas. Louzada, Ara e Fernandes (2016), também observaram que as técnicas de Regressão Logística (RL) e Árvores de Decisão (DT), servem de base para comparar a potencialidade de outras técnicas, sendo que por vezes, o objetivo destas é superar a eficiência da RL e DT.

Como destaque, a técnica SVM apresentou alto percentual preditivo e baixo esforço computacional, quando comparado a outras técnicas. Louzada, Ara e Fernandes (2016)

destacam ainda a preocupação dos cuidados com a interpretação dos resultados *versus* a validação dos dados e os tipos de erros que estão sujeitos.

O estudo de Prado *et al.* (2016), sob o título “*Multivariate analysis of credit risk and bankruptcy research data: a bibliometric study involving different knowledge fields (1968–2014)*”, identificou e descreveu o uso de técnicas multivariadas de análise de dados em pesquisas sobre risco de crédito e falência. A pesquisa foi realizada por meio de estudo bibliométrico, com dados coletados em publicações indexadas à base de dados da *Web of Science da Thomson Reuters* entre 1968 e 2014. Os resultados, desde o início da década de 1990, apontaram um crescimento de técnicas computacionais como RNA, transformando-se em técnica relevante, superando as técnicas estatísticas.

Ainda Prado *et al.* (2016), descrevem que as técnicas estatísticas de AD e RL, ainda são frequentemente usadas em pesquisas sobre o tema análise de crédito. Esses autores identificaram um elevado número de citações envolvendo as técnicas: RNA, AD e RL, sendo destacado o crescimento de modelos híbridos. A análise de crédito é um campo multidisciplinar, envolvendo diversas áreas do conhecimento como economia, negócios, gestão, engenharia e estatística, entre outras. Por fim, esses autores identificaram um número crescente de publicações sobre esse assunto após a crise de 2008, tendo praticamente dobrado o número de publicações em 2009.

Mais recentemente, Fuhr, Lima e Schenatto (2017), no estudo denominado “Uma Revisão Sistemática da literatura sobre *Credit Scoring*”, realizou uma revisão sistemática da literatura sobre as técnicas utilizadas para esse fim em PMEs cobrindo o período 2008 a 2016, diretamente em artigos, principalmente os voltados a área de finanças utilizando o *Proknow-C*, como metodologia de pesquisa junto as bases *Scopus e Web of Science*. Como resultado foi identificado que as técnicas mais utilizadas para análise de crédito de PMEs foram: RL, AD, RNA, SVM e DT. Além disso, as medidas de acurácias dessas técnicas, aplicadas sobre as PMEs, quando comparadas às aplicações em bases gerais, foram mais modestas, o que pode estar ligado a questão da qualidade das informações. Segundo Bertti (2012), em uma boa parte das empresas brasileiras, principalmente as MPEs a contabilidade não reflete a realidade das transações, usando a contabilidade apenas para fins fiscais. Ainda Ciampi e Gordini (2013) ressaltam as dificuldades das MPEs em apresentar documentos comprobatórios das demonstrações financeiras. Uma vez que as mesmas não são auditadas e por sua vez não possuem capital social negociado em mercado de preços, deixando de ter classificação pública que possa sugerir qualidade.

2.6 PONTUAÇÃO DE CRÉDITO (*CREDIT SCORING*)

As informações, sejam cadastrais, creditícias ou contábeis, são fatores relevantes nos modelos de *rating* ou “classificação”. A palavra *rating*, bastante utilizada na construção de

modelos matemáticos e estatísticos para enquadramento de crédito. Silva (2004), destaca que *rating* é uma avaliação da informação feita por meio da mensuração e ponderação de variáveis determinantes, fornecendo uma graduação. Essa classificação é utilizada na modelagem para previsão de risco de inadimplência, chamados de pontuação de crédito.

Rogers, Mendes-da-Silva e Rogers (2016), relatam que o *rating* de crédito e estrutura de capital, são pouco estudados no ambiente institucional da América Latina, atribuindo, em parte, a iminência de reclassificações do *rating*. Mais especificamente, no Brasil a utilização do *rating* e a criação das pontuações de crédito, de grande interesse das instituições financeiras, tiveram maior visibilidade a partir da década de 1990 (GONÇALVES; GOUVÊA; MANTOVANI, 2013). A ideia central é criar um modelo que transcreva informações quantitativas e qualitativas em uma escala de classificação (ou pontuação), refletindo a capacidade financeira do indivíduo (ou organização) honrar com o compromisso assumido (MILERIS, 2012).

2.7 REGRESSÃO LOGÍSTICA

Ao selecionar uma técnica, visando construir modelos, busca-se que a mesma apresente condição de identificar a influência entre a variável dependente e o volume das variáveis independentes, de forma parcimoniosa e funcional.

A técnica Regressão Logística (RL) tornou-se uma ferramenta para análise de dados, estabelecendo uma relação entre uma variável resposta e uma ou mais variáveis explicativas e o conjunto de variáveis independentes (preditora ou explicativa) (HOSMER; LEMESHOW, 1989).

A RL é uma técnica estatística que tem como objetivo produzir um modelo de previsão a partir de um conjunto de valores tomados por uma variável de resposta binária. Para Dias Filho e Corrar (2017), a RL também explica ou prediz valores de uma variável em função de valores conhecidos de outras variáveis. Para Hair *et al.* (2009), a variável estatística, “é uma combinação linear de variáveis formada na técnica multivariada determinando-se pesos empíricos aplicados a um conjunto de variáveis especificado pelo pesquisador”.

A RL tem os seguintes objetivos: (i) estimar a probabilidade de enquadramento de certos fenômenos, em uma ou outra categoria; esses fenômenos podem ser objetos ou pessoas (DIAS FILHO; CORRAR 2017); ou (ii) se caracteriza como uma regressão, de forma especializada, objetivando prever e explicar uma variável categórica binária (dois grupos) e sua variável estatística representa uma relação multivariada, indicando o impacto relativo de cada variável preditora. (HAIR *et al.*, 2009).

A Regressão Logística, apresenta a vantagem de ser menos afetada do que a análise discriminante quando as suposições básicas, particularmente a normalidade das variáveis, não

são satisfeitas (HAIR *et al.*, 2009).

Regressão logística é uma modelagem matemática com abordagem que pode ser usada para descrever o relacionamento de variável X para uma variável dicotômica dependente. Outras abordagens de modelagem também são possíveis, mas a regressão logística é de longe o procedimento mais popular (KLEIN; KLEINBAUM, 2010).

2.7.1 Modelo Matemático da Regressão Logística

Uma das diferenças entre a técnica da regressão linear múltipla com a RL é que aquela pode assumir valores menores que zero e maiores que um. A RL apresenta uma base logarítmica, assim evitando valores menores que zero e maiores que um. Com base no Capítulo de Dias Filho e Corrar (2017), na sequência, são descritos os 5 passos para construção do modelo logístico.

1º - Efetua-se a conversão da probabilidade associada a cada observação em razão de chance (odds ratio), que representa a probabilidade de sucesso comparada com a de fracasso.

2º - Com objetivo de melhorar a ordem operacional e para facilitar a interpretação dos resultados, a fórmula busca obter um logaritmo natural de razão de chance.

Equação 1 - Razão de chance

$$\text{Razão de Chance} = \frac{P(\text{sucesso})}{1 - P(\text{sucesso})}$$

Fonte: Adaptado de Dias Filho e Corrar (2017)

3º - Consiste em elevar a constante matemática "e" ao expoente composto dos coeficientes estimados.

Equação 2 - Constante matemática

$$\left(\frac{P(\text{sucesso})}{1 - P(\text{sucesso})} \right) = e^{(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_{ki})}$$

Fonte: Adaptado de Dias Filho e Corrar (2017)

4º - Desta forma, chega-se ao objetivo de estimar a probabilidade associada à ocorrência de determinado evento.

Equação 3 - Probabilidade associada à ocorrência

$$P(\text{evento}) = \frac{e^{(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_{ki})}}{1 + e^{(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_{ki})}}$$

Fonte: Adaptado de Dias Filho e Corrar (2017)

5º - A equação logística pode assumir a seguinte forma.

Equação 4 - Equação Logística

$$P(\text{evento}) = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_{ki})}}$$

Fonte: Adaptado de Dias Filho e Corrar (2017)

A curva logística é utilizada para representar a relação existente entre as variáveis dependentes e independentes, sendo que, nos níveis mais baixos da variável independente aproxima-se de 0. Por outro lado, quando os valores previstos crescem, a curva se aproxima de 1, sem, no entanto, exceder tal valor (HAIR *et al.*, 2009). A Figura 4 apresenta o gráfico da curva da RL.

Figura 4 - Curva da Logística



Fonte: Adaptado de Hair et al. (2009, p. 284)

Gouvêa, Gonçalves e Mantovani (2013); Li *et al.* (2016); Zhu *et al.* (2016), Lima e Marques (2002), utilizaram a técnica de RL, entre outras, em modelos de predição de pontuação de crédito, a qual mostrou-se um instrumento promissor no levantamento e classificação de informações para análise de crédito.

2.8 VARIÁVEIS INDEPENDENTES E DEPENDENTES

Primeiramente, vale definir o termo variável, o qual pode ser entendido como um substantivo, representativo de uma classe de objetos, podendo ou sendo a mesma manipulada pelo pesquisador (DUARTE; FURTADO, 2014). A variável independente, causa influência sobre outra variável, afetando-a, torna-se condição ou causa para o resultado. Para Marconi e Lakatos (2017), é geralmente manipulada pelo pesquisador, efetuando tentativas para

assegurar a relação de fator com um fenômeno.

A variável dependente é o que se pretende descobrir, em virtude de serem afetadas ou influenciadas pela variável independente. Marconi e Lakatos (2017), descrevem que à medida que o investigador avança nos testes, introduzindo, extraindo ou modificando a variável independente, causa impacto na variável dependente.

Ainda para Hair *et al.* (2009, p.22) as variáveis são assim descritas:

- I. Variável dependente é o efeito presumido, ou resposta, a uma mudança nas variáveis independentes.
- II. Variável dicotômica: Variável não-métrica transformada em uma variável métrica designando-se 1 ou 0 a um objeto, dependendo se este possui ou não uma característica particular.
- III. Variável estatística: Combinação linear de variáveis formada na técnica multivariada determinando-se pesos empíricos aplicados a um conjunto de variáveis especificado pelo pesquisador.
- IV. Variável independente: causa presumida de qualquer mudança na variável dependente.

2.8.1 Variáveis e técnicas utilizadas em trabalhos de modelagem

O Quadro 1 - ilustra trabalhos de modelagens de crédito utilizando várias técnicas, entre elas, a RL, sobre dados de empresas elaborados a partir das informações geradas pelas demonstrações contábeis como Balanço Patrimonial (BP) e Demonstração de Resultados do Exercício (DRE). Estes dados são retirados de empresas de porte médio a grande, os quais possuem uma contabilidade estruturada de informações, passíveis de auditoria, capazes de gerar indicadores financeiros (IF), indicadores contábeis (IC) e indicadores econômicos (IE).

Estes indicadores como índices de liquidez, estrutura patrimonial e de resultados, geram informações, tratadas posteriormente como variáveis na aplicação de um modelo de pontuação de crédito.

Para empresas que não possuem obrigatoriedade de publicações da informação e demonstrações contábeis para a tomada de decisão, utilizam nos modelos de pontuação de crédito, dados oriundos da ficha cadastral do cliente, dados de crédito e histórico interno, chamados de dados não financeiros (NF). O Quadro 1, apresenta estudos relacionados com pontuação de crédito para PMEs e suas acurácias.

Quadro 1 - Técnicas e variáveis usadas para predição de PMEs

Autores	Objetivo do Trabalho	Variáveis	Técnicas	Acurácia	País
Mselmi, Lahiani e Hamza (2017)	Examinar a capacidade dos índices financeiros para sinalizar financiamento	IF	SVM RNA PLS-DA Logit	88,57 87,14 84,28 85,71	França
Li <i>et al.</i> (2016)	Elaborar modelo híbrido RN e RL	IC	RL RNA Híbrido	70,21 79,18 83,11	China
Zhu <i>et al.</i> (2016)	Previsão de risco de crédito para pequenas empresas	IF e NF	RL RNA Híbrido I Híbrido II Híbrido III	61,3 68,8 70,2 88,5 87,4	China
Smaranda C. (2014)	Testar os modelos de previsão de falências	IF	RL Outros	87,2	Romênia
Ciampi F.; Gordini N. (2013)	Testar um modelo de previsão padrão para PMEs	IC	AD RL RNA	68,9 69,5 77,4	Itália
Wang e Zhou (2011)	Verificar a eficiência tradicional das variáveis financeiras para prever a inadimplência das PMEs	IF	RL Model II RL Model III	88,6 100	China
Lugovskaya, L. (2009)	Desenvolver modelos baseados em dados russos adequados para uso no contexto russo	IF e NF	LDA Model 1 LDA Model 2	68,1 79,0	Rússia
Ciampi F.; Gordini N. (2008)	Desenvolver modelos de previsão padrão para pequena empresa	IE e IF	AD RL	75,5 80	Itália

Fonte: Elaborada pelo autor

Pode-se observar que os modelos construídos, devido a questão estrutural econômico-financeira, utilizam dados contábeis, em praticamente todas as modelagens, sendo que em alguns casos incorporam variáveis não financeiras para melhorar os resultados.

A falta de confiabilidade na informação contábil (BERGER; COWAN; FRAME, 2011), ou mesmo a ausência dessa informação, levam as instituições financeiras a buscarem outras fontes de dados para geração de informações. Conforme a Lei Complementar 123 de 2006, aplicável às MPes, em seu Artigo 27, dispõe que “as microempresas e empresas de pequeno porte optantes pelo Simples Nacional poderão, opcionalmente, adotar contabilidade simplificada”.

Reforça Ribeiro, Freire e Barella (2013), que um dos importantes motivos que levam as MPes ao descontrole financeiro está relacionado a falta de conhecimento contábil. As

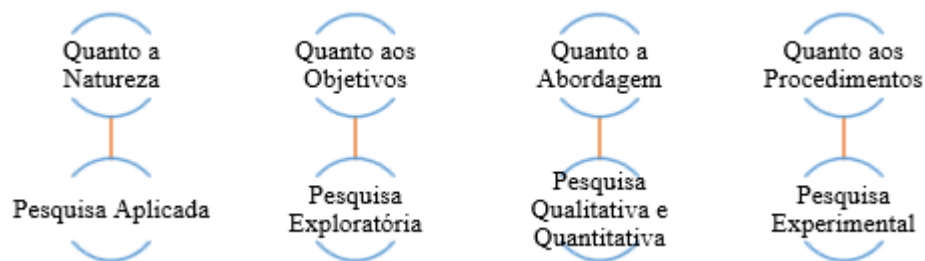
MPEs, devido as dificuldades financeiras, deixam em segundo plano o investimento em contabilidade, fazendo o caminho contrário do gerenciamento.

As dificuldades específicas do mercado brasileiro, e das Leis que norteiam esse mercado para as MPEs, às deixam fora do universo informacional gerencial, culminada com a ausência de auditoria, coloca em descrédito a informação contábil para as PMEs. Conforme Manual de Procedimentos Contábeis para Micro e Pequenas Empresas – CFC; SEBRAE (2002), reforça que a “falta de escrituração contábil é uma das principais dificuldades para se avaliar a economia informal, o que distorce as estatísticas no Brasil”. Ainda Santos, Dorow e Beuren (2016), descrevem que a tomada de decisão necessita de instrumentos que forneçam informações confiáveis e fidedignas, não podendo ser baseada apenas na intuição e experiência.

3 MATERIAIS E MÉTODOS

Ao descrever a metodologia utilizada neste trabalho, buscou-se estabelecer os caminhos que o pesquisador percorreu para chegar aos resultados. O método é o conjunto das atividades sistemáticas e racionais que permite alcançar o objetivo com maior segurança e economia, obtendo conhecimentos válidos (LAKATOS; MARCONI, 2010). A Figura 5 ilustra o delineamento da pesquisa, destacando a natureza, objetivos, abordagem e procedimentos da pesquisa adotados neste trabalho.

Figura 5 - Delimitação da pesquisa



Fonte: Elaborada pelo autor

3.1 DELINEAMENTO DO TIPO DE PESQUISA.

Esta pesquisa é de natureza aplicada, devido ao interesse e aplicação prática do conhecimento científico junto a instituição de crédito. Segundo Gil (2008), a principal característica da pesquisa aplicada é o interesse na aplicação prática de conhecimentos. A pesquisa aplicada busca a solução imediata de problemas concretos do cotidiano por meio da orientação prática visando a orientação, a gestão e tomada de decisão (DUARTE; FURTADO, 2014).

Quanto aos objetivos de pesquisa, é caracterizada como exploratória. Segundo Beuren (2014), essa forma de pesquisa busca conhecer o assunto em profundidade, visando elaborar questões importantes para a condução da pesquisa. Segundo Gil (2008), a pesquisa exploratória tem como principal finalidade, desenvolver, modificar ou esclarecer ideias, conceitos, voltado a formulação dos problemas, ainda com vistas a formulação de hipóteses pesquisáveis para estudos posteriores.

A pesquisa exploratória, familiariza o problema, auxiliando na formulação de hipóteses, ela tem caráter de reconhecimento, envolvendo conversas informais com os especialistas, ainda podendo o pesquisador participar de reuniões como ouvinte (DUARTE; FURTADO, 2014). A parte exploratória, relativa a este trabalho, concentra-se mais especificamente, em entrevista com especialista quanto ao tratamento das variáveis independentes utilizadas no estudo de forma qualitativa.

Quanto a abordagem a pesquisa caracteriza-se como Qualitativa e Quantitativa, ou seja, Quali-Quant. A pesquisa com abordagem qualitativa é aquela que envolve dados qualitativos e formas de análise qualitativa; como exemplo, tem-se dados obtidos em entrevista, correspondendo a forma, não numérica, de agrupar tal informação (DUARTE; FURTADO, 2014). Por outro lado, a pesquisa com abordagem quantitativa, são aquelas apoiadas em métodos estatísticos, os quais tem seu foco na comprovação de hipóteses. Aceita-se ou refuta-se tal hipótese, esse tipo de pesquisa tem objetivos mais bem definidos e é de forma mais específica e analítica (OLIVEIRA, 2011).

Contudo existem pesquisas que utilizam as duas abordagens para atingimento do objetivo. Segundo Oliveira (2011), os métodos de pesquisa qualitativos e quantitativos não são excludentes, podendo ser utilizados conjuntamente de forma complementar. Dentro da abordagem conjunta (qualitativa e quantitativa), Duarte e Furtado (2014), destacam que independente do caráter quantitativo ou qualitativo, quanto maior as fontes de informação e mais evidências o pesquisador possuir, maior será a segurança das afirmativas, com redução da margem de erro e da possibilidade de contestação.

Quanto aos procedimentos, esta pesquisa é experimental, uma vez que consiste em uma investigação empírica, voltada para a manipulação de variáveis. Duarte e Furtado (2014), relatam que este método avalia os efeitos ocorridos nas variáveis dependentes (ou investigadas), em decorrência da manipulação das variáveis independentes. A relação experimental, torna-se uma situação de pesquisa voltada para a validade dos resultados. Esta validade de resultados exige uma maior preocupação com o ambiente que envolve a pesquisa, como a melhoria dos controles e as delimitações da pesquisa. (DUARTE; FURTADO, 2014).

Ainda conforme Marconi e Lakatos (2003), pontuam que o "método por excelência da ciência é o experimental: ela caminha apoiada nos fatos reais e concretos, afirmando somente aquilo que é autorizado pela experimentação".

3.2 TÉCNICAS ESTATÍSTICAS MULTIVARIADAS

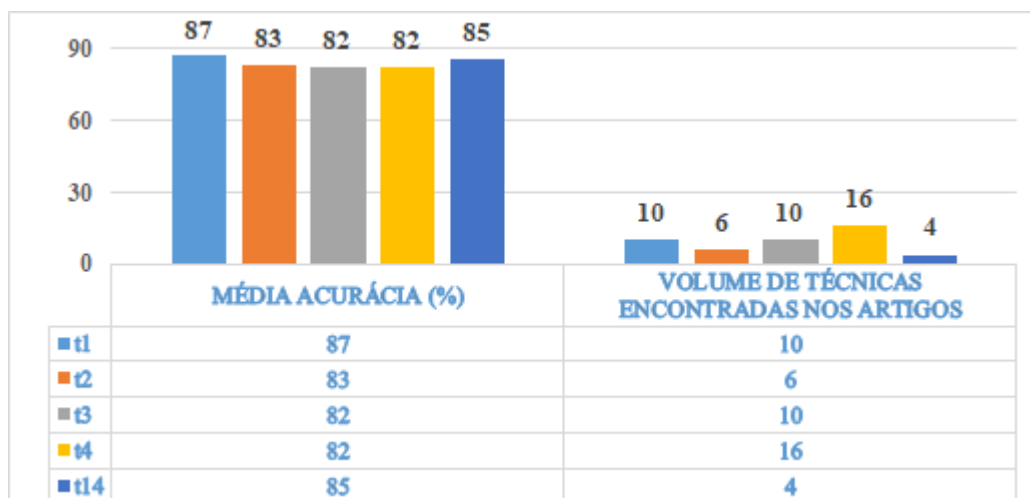
A técnica utilizada, sobre a base de DG, bem como nas 4 bases das classes de empresas, descritas na seção 3.3.4, para desenvolver os modelos foi a RL. Já técnica de AD exige a atenção à vários pressupostos, os quais nem sempre são atendidos. Para Hair *et al.* (2009), a técnica da AD deve atender as suposições de normalidade multivariada e de igualdade entre as matrizes de variância-covariância nos grupos. Reforça Mário (2017), a importância do atendimento as premissas, antes de efetuar as segregações de amostras e adotar outros procedimentos para o desenvolvimento da AD.

Destacam Fuhr, Lima e Schenatto (2017), em estudo de revisão bibliográfica da literatura, o qual cobriu o período de 2008 a 2016, selecionando 20 artigos dentro de um portfólio de 1277 artigos, sendo possível verificar a importância das técnicas escolhidas.

Dentro das 17 técnicas encontradas na revisão bibliográfica de Fuhr, Lima e Schenatto (2017), verifica-se que a mais utilizada foi RL que aparece em 16 trabalhos, tanto como técnica principal preditiva, como técnica comparativa para medir a predição de outras técnicas.

A RL, nas 16 vezes que foram utilizadas nos trabalhos, apontou uma média de acurácia de 82% entre os 20 artigos do portfólio, a qual é identificada na Figura 6 como “t4”. As demais técnicas que apresentaram forte poder de acurácia foram: t1 = SVM; t2 = RNA; t3 = AD e t14 = *Decision tree* (DT) e variações. São ilustradas na Figura 6 apenas as técnicas que apareceram 4 vezes ou mais dentro do portfólio dos 20 artigos.

Figura 6 - Predição e volume de técnicas encontradas nos modelos nos 20 artigos selecionados

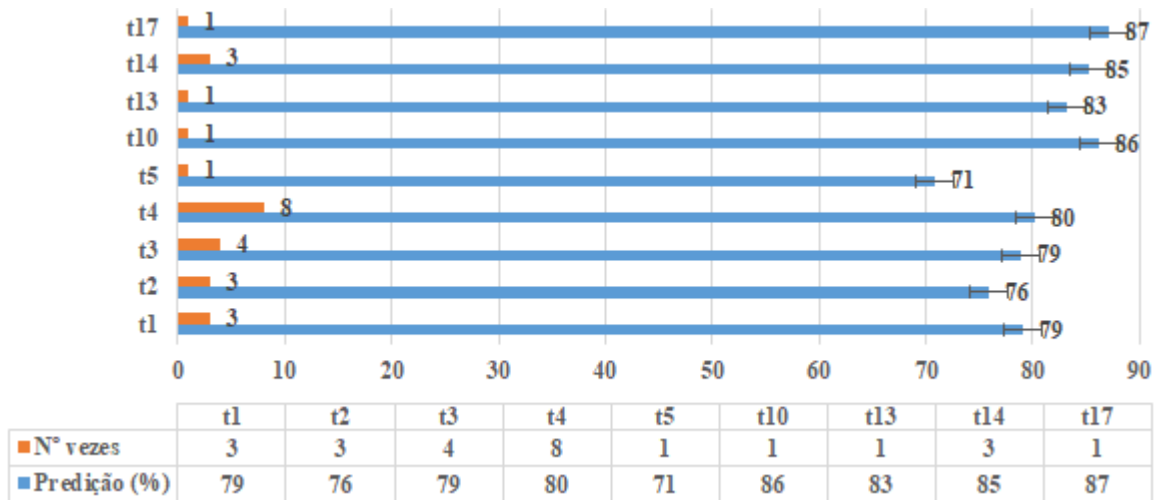


Fonte: Adaptada de Fuhr, Lima e Schenatto (2017)

Ainda, no trabalho de Fuhr, Lima e Schenatto (2017), foi identificado que dos 20 artigos selecionados, 11 deles são direcionados às PMEs. Ao analisar, a técnica RL, conforme Figura 7, identificada como “t4”, aparece em 8 trabalhos, demonstrando uma média de acurácia de 80%.

É importante observar que, no portfólio direcionado às PMEs, 11 trabalhos, apresentaram um poder de predição das técnicas menor do que a acurácia do portfólio geral, composto por 20 trabalhos. Este fato pode estar relacionado ao tipo de variáveis utilizadas para as PMEs, geralmente não contábeis e não auditadas, enquanto que as empresas de porte maior possuem uma estrutura contábil informacional e por vezes auditadas. A Figura 7 ilustra as técnicas que foram utilizadas nos artigos, que direcionaram esforços para predição ou falências, extraídas dos 11 artigos. Foram identificadas 9 técnicas nos 11 artigos analisados.

Figura 7 - Acurácia das técnicas e o número de vezes que foram utilizadas dentro do portfólio dos 11 artigos direcionados para as PMEs



Fonte: Adaptada de Fuhr, Lima e Schenatto (2017)

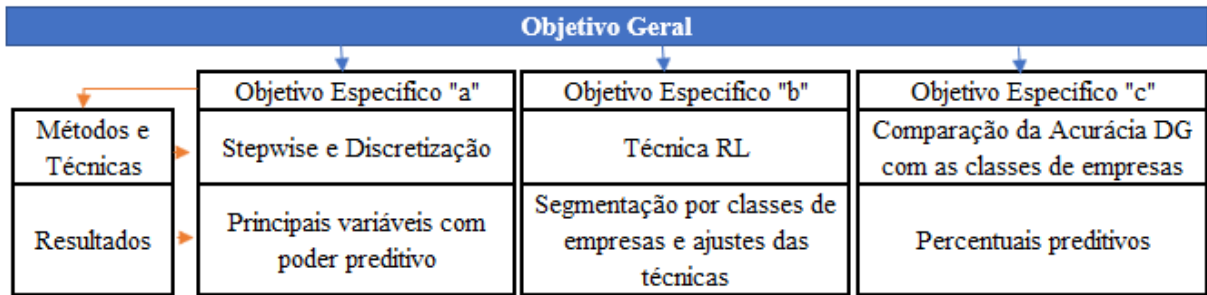
Observa-se que as técnicas t5, t10, t13 e t17 aparecem apenas uma vez dentro do portfólio, o que dificulta a comparabilidade entre trabalhos. Nos trabalhos de Prado *et al.*, (2016) e Louzada, Ara e Fernandes (2016), também é possível verificar a robustez da técnica RL e suas variações híbridas. Os trabalhos abordam a utilização da técnica na predição de risco e pontuação de crédito, suas potencialidades como base comparativa para medir a eficiência de outras técnicas, o que justifica a utilização da mesma neste estudo.

Mais especificamente, o trabalho de Prado *et al.* (2016) teve como objetivo identificar e descrever a aplicação de técnicas de análise multivariada de dados para cenários de risco de crédito e falência. O trabalho de Louzada, Ara e Fernandes (2016), teve como objetivo, apresentar uma revisão sistemática da literatura relativas a teoria e aplicação de técnicas de classificação utilizados para discriminar mutuários entre bons e maus pagadores.

3.3 PROCEDIMENTOS METODOLÓGICOS

A Figura 8 apresenta a relação entre os objetivos (Geral e Específicos), os Métodos e Técnicas utilizadas para atingi-los e os resultados esperados. O Objetivo Geral está ligado a uma visão global e abrangente do tema. Possui relacionamento com o conteúdo intrínseco, podendo ser dos fenômenos e eventos, ou das idéias estudadas. Por sua vez o Objetivo Específico têm função intermediária e instrumental, possibilitando atingir o objetivo geral e também aplicá-lo a situações particulares (MARCONI; LAKATOS, 2003).

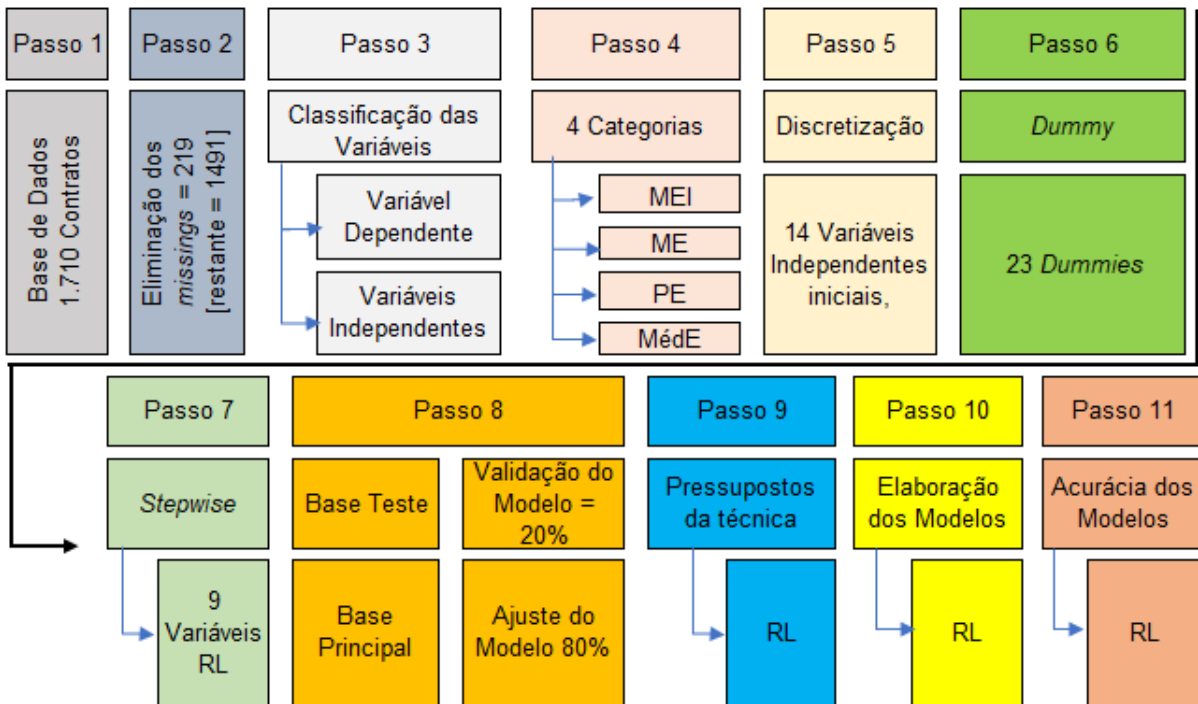
Figura 8 - Relação entre Objetivo Geral e Objetivos Específicos



Fonte: Elaborada pelo autor

Quanto aos procedimentos para desenvolvimento de um modelo de pontuação de crédito, foram aplicados 11 passos: sendo Passo 1: análise da base de dados, Passo 2: eliminação de informações ausentes (*missing values*), Passo 3: classificação das variáveis, Passo 4 divisão das classes de empresas, Passo 5: discretização, Passo 6: criação de variáveis artificiais (*dummy*), Passo 7: aplicação do método *stepwise*, Passo 8: divisão da base geral em base teste e base principal, Passo 9: aplicação dos pressupostos, Passo 10: elaboração dos modelos para RL e Passo 11: acurácia da técnica RL. Estes passos estão descritos no Figura 9, sintetizando a metodologia aplicada.

Figura 9 - Ilustração da metodologia aplicada



Fonte: Elaborada pelo autor

3.3.1 Análise da base de dados

As observações e dados utilizados nesta pesquisa, população (universo), foram coletados de forma secundária, por meio dos relatórios em planilhas eletrônicas MS-Excel®, junto a uma instituição financeira com sede na região Sudoeste do Paraná, vocacionada para o crédito comercial e industrial. Contudo, o nome da instituição não foi autorizado para divulgação, desta forma, o nome e outras informações de identificação dos clientes foram substituídos por numerações aleatórias. Importante observar a existência de ao menos uma característica em comum entre os indivíduos pertencentes a população (MARCONI; LAKATOS, 2003). Neste estudo, as características que envolveram os clientes ou cooperados ficaram a cargo de todas serem pessoas jurídicas ou equiparadas em busca de crédito no mercado financeiro para alavancagem dos negócios ou parcelamento das dívidas.

Essas planilhas compuseram um total de 1.710 linhas, correspondendo a cada linha um cliente classificado como Pessoa Jurídica ou equiparada a Pessoa Jurídica, detentor de um contrato de empréstimo ou financiamento.

3.3.2 Eliminação de informações incompletas (*missings values*)

Nessa etapa foi efetuada uma análise do banco de dados, sendo eliminados 219 contratos da totalidade de 1.710, restando uma base de contratos de 1.491. A análise minuciosa da base de dados, focou-se na busca de informações faltantes nas variáveis independentes para todos os contratos. O período de levantamento dos dados corresponde aos créditos ativos (não liquidados), até 10 de outubro de 2017. Esse processo de eliminação dos dados ausentes, também foi adotado por (CAMARGOS; CAMARGOS; ARAUJO, 2012). O processo de “dados perdidos”, para Rodrigues e Paulo (2017), corresponde a qualquer evento sistemático externo ao respondente, levando primariamente o pesquisador a buscar as razões inerentes a esses. A justificativa para eliminação dos dados, encontra-se na necessidade da informação para classificação do porte das empresas como renda (faturamento). Destaca-se que as informações ausentes não foram captadas devido a fazer parte de outro sistema (base de dados), não sendo possível acessar na data do levantamento dos dados.

3.3.3 Classificação das Variáveis

Definição da variável resposta: a variável resposta, também conhecida como dependente, direcionada ao setor de concessão de crédito, refere-se à qualidade do crédito, sendo identificado como bom cliente (adimplente) ou mau cliente (inadimplente). Conforme política da instituição de estudo, são considerados maus clientes (inadimplentes), aqueles que apresentaram atraso superior a 30 dias ininterruptos dentro do ano fiscal. Por outro lado, os

clientes que foram identificados com pagamentos sem atrasos ou com atrasos iguais ou inferiores a 30 dias, foram classificados como bons clientes (adimplentes).

Preparação dos dados: as variáveis utilizadas no modelo são qualitativas e quantitativas. As variáveis quantitativas podem ser definidas como as que expressam mensuração, geralmente numérica (ou escalonar) e as variáveis qualitativas definidas como aquelas que expressam uma característica (ou atributo). Segundo Rodrigues e Paulo (2017), as variáveis quantitativas podem ser medidas em escalas apresentando da forma discreta ou contínua, enquanto que as variáveis qualitativas não possuem valores quantitativos, também conhecidas como categóricas, podendo ser dividida em nominal e ordinal. A Tabela 1, apresenta as variáveis que foram utilizadas nos modelos, suas ordens e codificações como qualitativa ou quantitativa, oriundas da base de dados fornecidas pela Instituição Financeira, denominadas de variáveis primitivas.

Tabela 1 - Variáveis primitivas extraídas da base de dados

Código	Variável	Referência	Ordem	Codificação
V1	Código Município	C_Município	Nominal	Qualitativa
V2	Cnae	Cnae	Nominal	Qualitativa
V3	Renda Mensal	R_Mensal	Contínua	Quantitativa
V4	Valor das Garantia	V_Garantia	Discreta	Quantitativa
V5	Risco	Risco	Ordinal	Qualitativa
V6	Valor do Contrato	V_Contrato	Contínua	Quantitativa
V7	Nº Parcela	N_Parcela	Contínua	Quantitativa
V8	Dias Atraso	D_Atraso	Contínua	Qualitativa
V9	Nº de Produtos	N_Produto	Contínua	Quantitativa
V10	Valor Cota Capital	V_C.Capital	Contínua	Quantitativa
V11	Idade	Idade	Ordinal	Qualitativa
V12	Renda Anual	R_Anual	Contínua	Quantitativa
V13	Tempo de Filiação	T_Filiação	Contínua	Quantitativa
V14	Saldo Devedor	S_Devedor	Contínua	Quantitativa
V15	Classificação	Porte da empresa	Nominal	Qualitativa

Fonte: Elaborada pelo autor

No tratamento das variáveis independentes, Camargos, Camargos e Araújo (2012), utilizaram variáveis como: setor da atividade, valor dos bens do avalista, valor do financiamento, tempo de atividade da empresa, valor do faturamento anual da empresa, entre outras. Essas variáveis também foram identificadas no universo dos dados deste estudo como (V2), (V4), (V6) e (V12).

No trabalho de Gonçalves, Gouvêa e Mantovani (2013), utilizaram variáveis cadastrais como idade, salário (compatível com renda), entre outras e variáveis de utilização e

restrição como: valor do empréstimo, número de parcelas, entre outras. Essas variáveis também foram identificadas no universo dos dados deste estudo como: (V9), (V11), (V12) e (V16). Ainda Gonçalves, Gouvêa e Mantovani (2013), complementam que as variáveis quantitativas foram categorizadas, ou seja, transformadas em variáveis categóricas.

Da mesma forma, Silva, Ribeiro e Matias (2016), utilizaram variáveis como: montante do crédito requerido, duração do empréstimo requerido, montante do crédito requerido, idade, entre outras. O foco desse estudo foi a pessoa física, porém, a forma de tratamento para pessoa jurídica é semelhante. Essas variáveis também foram identificadas no universo dos dados deste estudo como: (V6), (V7) e (V11).

No trabalho de Araújo e Carmona (2009), foram utilizadas variáveis como: natureza da atividade econômica do negócio, número de parcelas do último empréstimo, tempo de funcionamento (idade), receita bruta do negócio, tempo, entre outras. Essas variáveis também foram identificadas no universo dos dados deste estudo como: (V2), (V7), (V9), (V11) e (V12). Descrevem Araújo e Carmona (2009), que as variáveis qualitativas foram inseridas por meio das variáveis *dummy*.

Ainda no trabalho de Lemos, Steiner e Nievola (2005), foram utilizadas variáveis como: setor de atividade, risco atribuído, tempo de atividade, faturamento bruto anual, entre outras. Essas variáveis também foram identificadas no universo dos dados deste estudo como: (V2), (V5), (V11) e (V12).

3.3.4 Divisão das classes de empresas

A importância de reconhecer as diferenças existentes entre as classes de empresas, entendendo suas dificuldades de buscar recursos, apresentar informações concisas e confiáveis e ofertar garantias, possibilita analisá-las, mais especificamente, dentro de suas formas de atuação e gestão. Também entendendo que, quando em situações de recessão de mercado, geralmente são as primeiras a apresentar dificuldades e as que levam um tempo maior para sair (SILVA 2016).

As Pequenas Empresas, caracterizadas como: MEI, ME, PE, diferem das MédE, assim como diferem entre si, nas suas formas estruturais, barreiras, benefícios fiscais entre outros, apresentando realidades diferenciadas, gerando informações próprias dentro do seu universo de atuação. Algumas diferenças como: faturamento e número de funcionários, encontram-se amparados pela Lei Complementar 123 de 2006 e instituições de apoio como Sebrae respectivamente.

Devido as peculiaridades da fonte de informação próprias de cada formato de empresa, conforme descreve Alvim (1998), quanto a informação adequada à determinada necessidade, os modelos propostos foram elaborados sobre as características individuais, de cada categoria de empresas, sendo: MEI, ME, PE e MédE, as quais estão distribuídas conforme faixas (ou

classes) do Banco Nacional de Desenvolvimento Estadual (BNDES, 2010), na qual a faixa da MEI, foi incluída conforme Lei 128/2008, que alterou a Lei 123/2006, Lei da Microempresa. A classificação adotada é apresentada na Tabela 2.

Tabela 2 - Enquadramento das empresas conforme renda anual

Categoria	MEI	ME	PE	MédE
Faixas	Até R\$ 60 mil	De R\$ 60 mil Até R\$ 2,4 milhões	De R\$ 2,4 milhões Até R\$ 16 milhões	De R\$ 16 milhões Até R\$ 90 milhões

Fonte: Dados obtidos dos sites do BNDES e Lei 128/2008

Ainda, quanto as peculiaridades da informação, Rodrigues e Blattmann (2014), contribuem dizendo que é importante conhecer a fonte da informação e o ambiente que a organização está inserida. Essas fontes variam em formatos e natureza, podendo influenciar no processo de uso de forma ótima.

Para a modelagem da pontuação de crédito, com a técnica de RL, foi utilizado o *Software SPSS Statistics®*. Foram elaborados os modelos de pontuação de crédito com a técnica da RL para as 4 classes de empresas e também para os DG, conforme ilustram as Figuras de 43 a 47.

3.3.5 Discretização de variáveis

As variáveis idade, renda anual, tempo de filiação, saldo devedor, número de parcelas e valor do contrato, foram discretizadas em intervalos distintos, sendo que o ponto de corte na média e desvio padrão calculados foi (+/- 1 desvio padrão) para cada categoria de empresa ou seja: MEI, ME, PE, MédE e DG. Segundo Pirolla (2012), a discretização de dados, caracteriza-se por dividir um conjunto contínuo de valor em intervalos. Segundo Lunet, Severo e Barros (2006), o valor do desvio padrão reflete a variabilidade das observações em relação à média, caracterizado como uma medida de dispersão. Conforme apresenta o trabalho de Pirolla (2012), a discretização normal ou Z-score pode ser apresentada de acordo com a equação 6:

Equação 5 - Z - score discretização

$$z = \frac{x - \mu}{\sigma}$$

Fonte: Adaptado de Pirolla (2012),

Na qual:

z = score bruto a ser normalizado,

x = variável a ser padronizada,

μ = média aritmética da população e

σ = desvio padrão da população.

3.3.6 Variáveis *Dummies*

As variáveis: idade, renda anual, tempo de filiação, saldo devedor, número de parcelas e valor do contrato, nas suas formas naturais ou primitivas, foram discretizadas, com a finalidade de simplificar os dados e escalonar faixas de dados coincidentes. Essa estratégia reduz as distâncias entre o mínimo e o máximo dentro de cada faixa discretizada.

As variáveis idade, renda anual, tempo de filiação, saldo devedor, número de parcelas e valor do contrato, foram transformadas em variáveis *dummy*, surgindo uma nova variável para cada faixa/classe discretizada, o que ocasionou o surgimento de 23 novas variáveis, em substituição as variáveis primitivas. Segundo Missio e Jacobi (2007), a variável *dummy* é uma variável artificial a qual assume valor igual a 0 ou 1, indicando a ausência ou presença de algum atributo transformando o modelo de regressão em uma ferramenta flexível para lidar com problemas encontrados em estudos empíricos. Segundo Cunha e Coelho (2017) com a *dummy* se quer melhorar o percentual do coeficiente de determinação (R²) e sua contribuição é indicar a presença ou ausência de determinado atributo, assumindo apenas 0 ou 1.

3.3.7 Seleção de variáveis – método *stepwise*

Para seleção das variáveis, foi aplicado o método *stepwise*. Para Kim e Sohn (2010), o método *stepwise* é usado para selecionar variáveis significativas. O processo de seleção gradual termina quando nenhuma outra variável puder ser adicionada para o modelo em um determinado nível de significância.

Essa técnica é bastante utilizada nos modelos desenvolvidos com a com a técnica de RL. O método *stepwise* é utilizado em métodos de estimação, com seleção sequencial de variáveis, objetivando identificar a variável independente com o maior poder preditivo no modelo de RL (HAIR *et al.*, 2009).

3.3.8 Validação e ajuste dos modelos

Para as técnicas selecionadas, foram utilizadas uma base teste (BT) de 20% dos dados totais, escolhidos aleatoriamente, para validação dos modelos. Os 80% restantes da base, denominada de base principal (BP), foram utilizados para ajuste e elaboração dos modelos. Esse procedimento foi efetuado para todas as 4 classes de empresas e para os DG. Este percentual também é encontrado nos trabalhos de Selau e Ribeiro (2009) e Silva, Ribeiro e Matias (2016).

3.3.9 Pressupostos da técnica RL

Para desenvolvimento dos modelos de previsão de risco de crédito, foi selecionada a técnica RL. Na Tabela 3, são apresentados os pressupostos necessários para o desenvolvimento da técnica e os testes desenvolvidos neste trabalho, apresentados no capítulo 4.

Tabela 3 - Pressupostos para utilização das técnicas

Técnicas	Termos	Pressupostos, Dias Filho e Corrar (2017)	Testes
Regressão	Não	a. [...] Inexistência de correlação entre os erros e as variáveis independentes	a. Tolerância
Logística	Linear	b. Ausência de multicolinearidade perfeita entre as variáveis independentes	b. Inverso da Tolerância (VIF)
			c. Verossimilhança
			d. Hosmer e Lemeshow

Fonte: Elaborada pelo autor

3.3.10 Elaboração dos modelos

Com base na seleção das variáveis, com maior poder preditivo, utilizando o método *stepwise*, elaborou-se os modelos, com base na fórmula, da técnica RL, resultado em 5 modelos, sendo uma para cada classe de empresas: MEI, ME, PE, MédE e DG. Todas as classes apresentaram normalidade e ausência de multicolinearidade. Esta técnica foi selecionada para desenvolvimento dos modelos.

3.3.11 Teste da acurácia para a Técnicas da RL

Um dos meios utilizados para verificação da capacidade dos modelos pode ser a capacidade de predição do modelo, também entendido como acurácia. Caracteriza-se pela expressão das técnicas, em percentuais da capacidade de classificação (discriminação) dos adimplentes e inadimplentes que compõe a base de dados. Conforme Portal Action (2018), acurácia (ACC) é a proporção de predições corretas, sem considerar o que é positivo e o que negativo e sim o acerto total. Dentro do processo da RL também é gerado a classificação da acurácia geral. Posteriormente efetuou-se uma análise comparativa da acurácia (predição) geral da técnica da RL para as 4 classes de empresas *versus* DG.

4 RESULTADOS E DISCUSSÃO

Neste capítulo foram realizadas a análise das variáveis, seguido da aplicação da técnica estatística RL. A técnica foi aplicada sobre a base de dados de contratos da uma cooperativa de crédito com corte em 10 de outubro de 2017. A base de dados, inicialmente foi separada por faixa de faturamento, conforme apresentado na Tabela 2. Ainda, com as informações coletadas, foi utilizada a base de dados para formulação de um modelo geral denominado DG.

4.1 TRATAMENTO DAS VARIÁVEIS INDEPENDENTES

4.1.1 Análise das variáveis

Ao iniciar a análise da base de dados, foi observado o comportamento de cada variável independente em relação a variável dependente e conseqüentemente entre as próprias variáveis independentes. As variáveis V3 (Renda Mensal) e V15 (Classificação), foram retiradas do modelo devido a serem inter-relacionadas com outras variáveis, demonstrando correlação. A existência de correlação é determinada quando variáveis independentes, entre si, explicam o mesmo fato com informações similares, este fenômeno é conhecido como multicolinearidade. Para Hair *et al.* (2009), a multicolinearidade, medida de tolerância, denota que duas ou mais variáveis independentes estão altamente correlacionadas, quando uma variável pode ser prevista por outras variáveis com baixo poder explicativo para o conjunto.

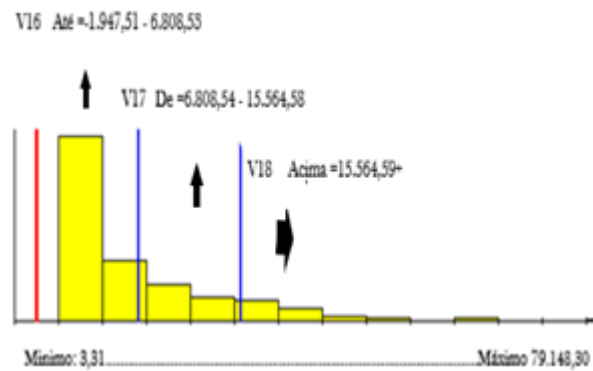
Por sua vez, a variável V5 (Risco) é calculada, pela Instituição Financeira, com base na evolução da variável V8 (Dias de Atraso). Desta forma a variável V5 e V8 foram descartadas por apresentar a mesma característica da variável dependente.

4.1.2 Discretização das variáveis

Para um melhor ajuste dos modelos, observando as classes das empresas, as variáveis, idade, renda anual, tempo de filiação, saldo devedor, número de parcelas e valor do contrato, foram discretizadas. Segundo Garcia *et al.* (2013), a discretização pode ser observada como um método de redução de dados, reduzindo em subconjuntos um grande volume de dados, uma vez que transforma atributos numéricos em atributos discretos ou nominais, com número finito de intervalos.

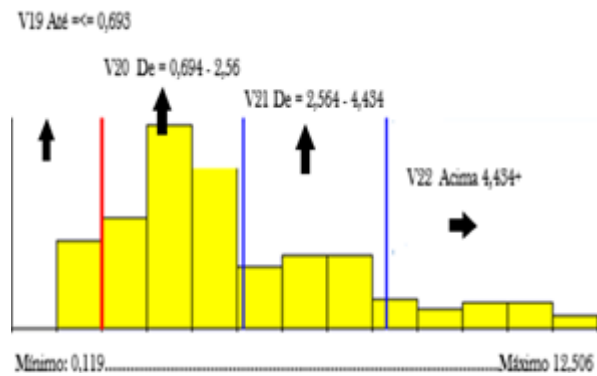
As Figuras 10 a 15 ilustram a discretização das variáveis saldo devedor, tempo de filiação, renda anual, idade, número de parcelas e valor do contrato, respectivamente, para as empresas classificadas como MEI.

Figura 10 - Discretização: Saldo Devedor



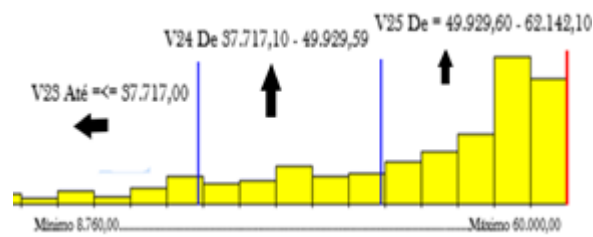
Fonte: Elaborada pelo autor

Figura 11 - Discretização: Tempo de Filiação



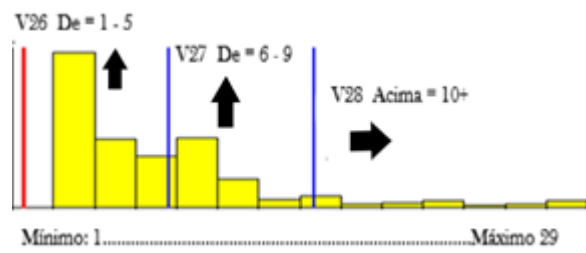
Fonte: Elaborada pelo autor

Figura 12 - Discretização: Renda anual



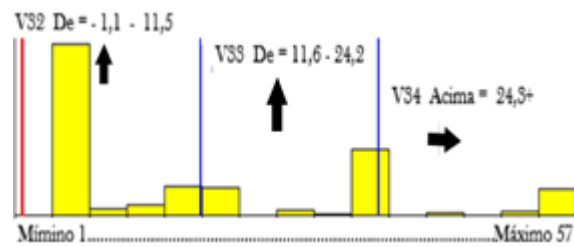
Fonte: Elaborada pelo autor

Figura 13 - Discretização: Idade



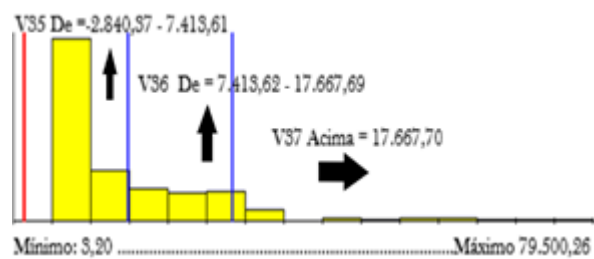
Fonte: Elaborada pelo autor

Figura 14 - Discretização: Número de Parcelas



Fonte: Elaborada pelo autor

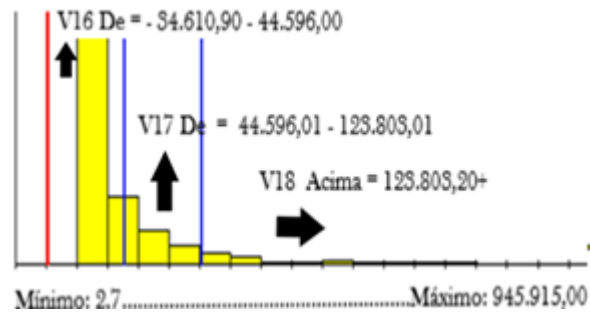
Figura 15 - Discretização: Valor do Contrato



Fonte: Elaborada pelo autor

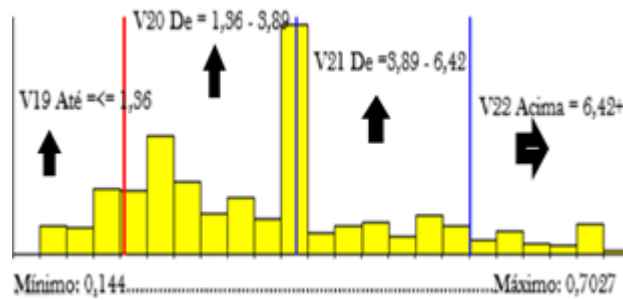
As Figuras 16 a 21 ilustram a discretização das variáveis saldo devedor, tempo de filiação, renda anual, idade, número de parcelas e valor do contrato, respectivamente, para as empresas classificadas como ME.

Figura 16 - Discretização: Saldo Devedor



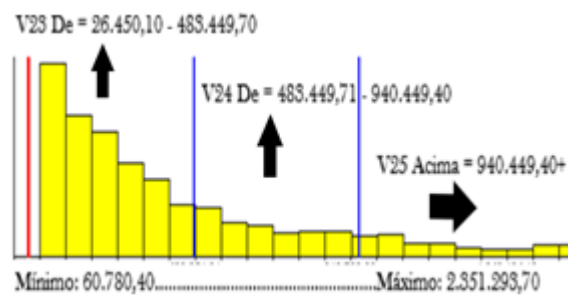
Fonte: Elaborada pelo autor

Figura 17 - Discretização: Tempo Filiação



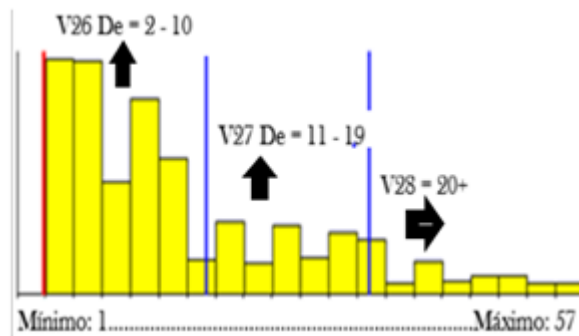
Fonte: Elaborada pelo autor

Figura 18 - Discretização: Renda Anual



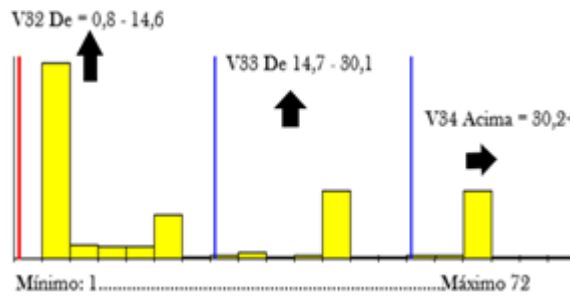
Fonte: Elaborada pelo autor

Figura 19 - Discretização: Idade



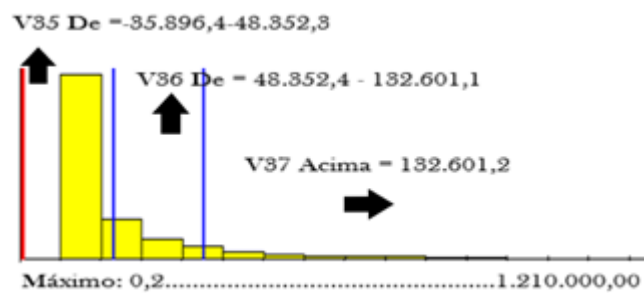
Fonte: Elaborada pelo autor

Figura 20 - Discretização: Número de Parcela



Fonte: Elaborada pelo autor

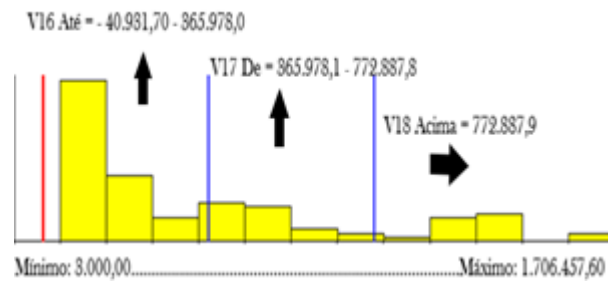
Figura 21 - Valor Contrato



Fonte: Elaborada pelo autor

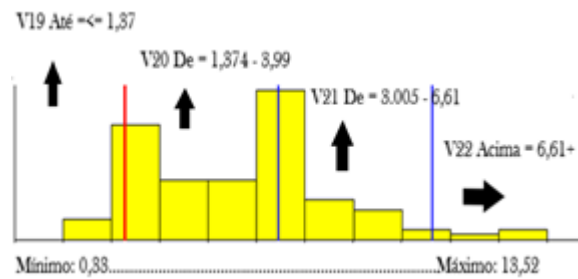
As Figuras 22 a 27 ilustram a discretização das variáveis saldo devedor, tempo de filiação, renda anual, idade, número de parcelas e valor do contrato, respectivamente, para as empresas classificadas como PE.

Figura 22 - Discretização: Saldo Devedor



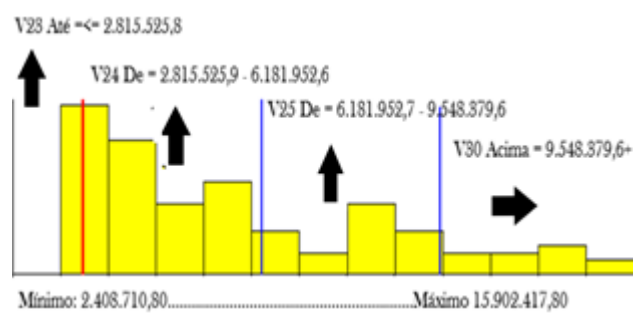
Fonte: Elaborada pelo autor

Figura 23 - Discretização: Tempo Filiação



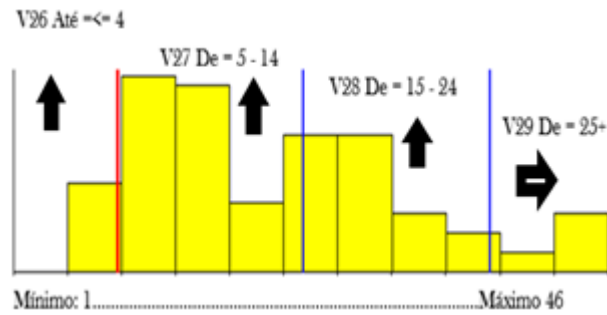
Fonte: Elaborada pelo autor

Figura 24 - Discretização: Renda anual



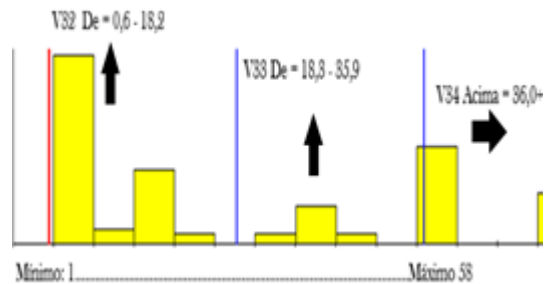
Fonte: Elaborada pelo autor

Figura 25 - Discretização: Idade



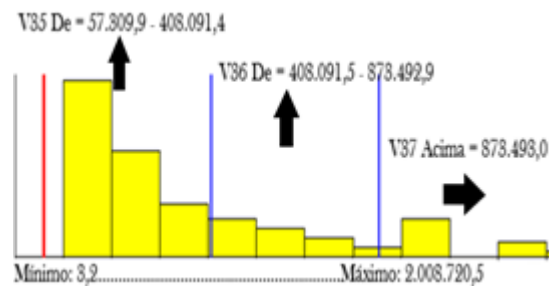
Fonte: Elaborada pelo autor

Figura 26 - Discretização: Número de Parcela



Fonte: Elaborada pelo autor

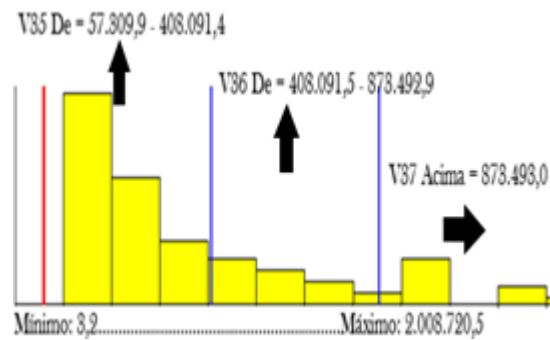
Figura 27 - Discretização: Valor do Contrato



Fonte: Elaborada pelo autor

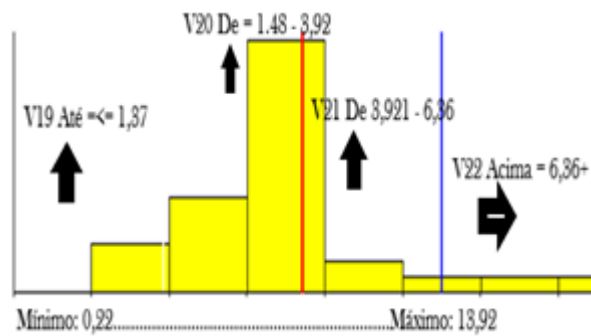
As Figuras 28 a 33 ilustram a discretização das variáveis saldo devedor, tempo de filiação, renda anual, idade, número de parcelas e valor do contrato, respectivamente, para as empresas classificadas como MéDE.

Figura 28 - Discretização: Saldo Devedor



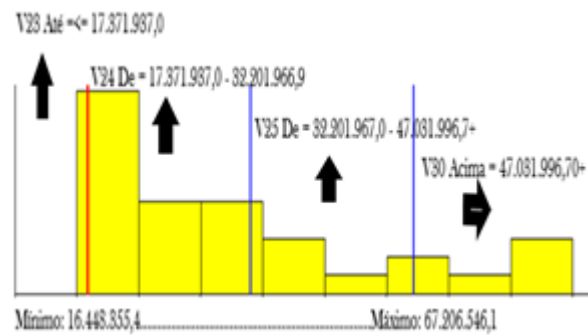
Fonte: Elaborada pelo autor

Figura 29 - Discretização: Tempo de Filiação



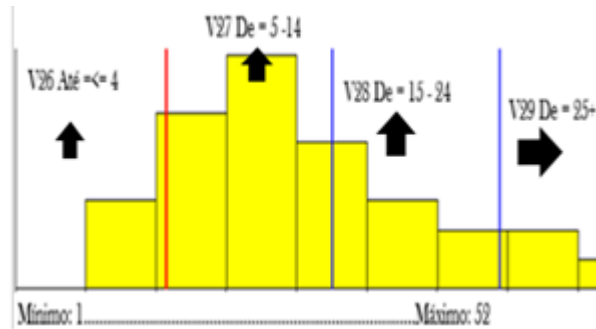
Fonte: Elaborada pelo autor

Figura 30 - Discretização: Renda Anual



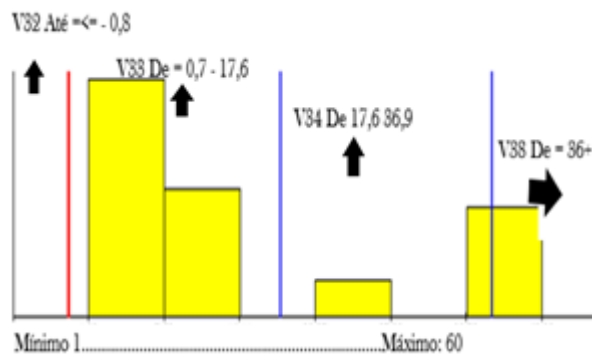
Fonte: Elaborada pelo autor

Figura 31 - Discretização: Idade



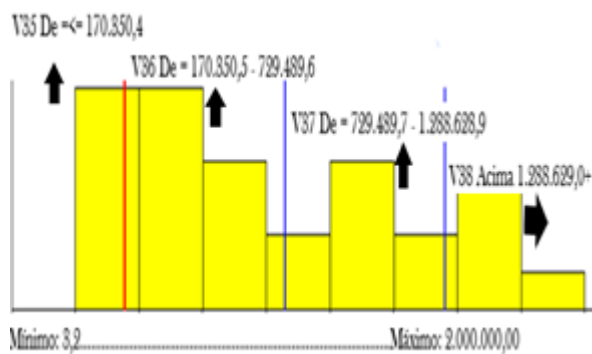
Fonte: Elaborada pelo autor

Figura 32 - Discretização: Número de Parcelas



Fonte: Elaborada pelo autor

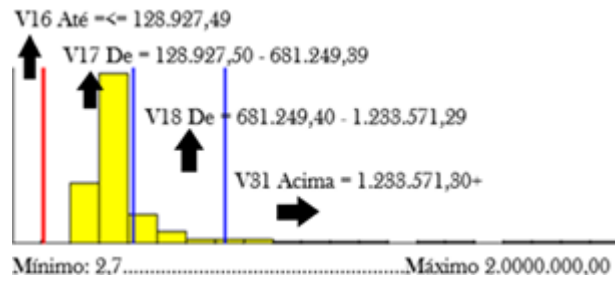
Figura 33 - Discretização: Valor do Contrato



Fonte: Elaborada pelo autor

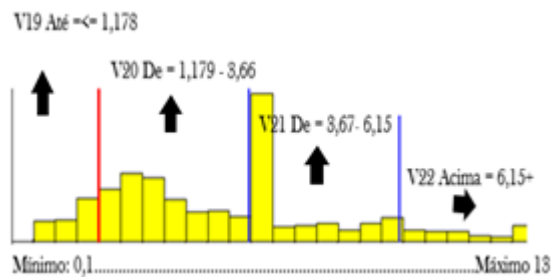
As Figuras 34 a 39 ilustram a discretização das variáveis saldo devedor, tempo de filiação, renda anual, idade, número de parcelas e valor do contrato, respectivamente, para as empresas classificadas como DG.

Figura 34 - Discretização: Saldo Devedor



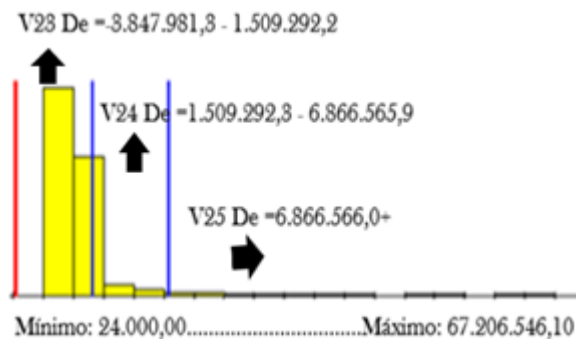
Fonte: Elaborada pelo autor

Figura 35 - Discretização: Tempo Filiação



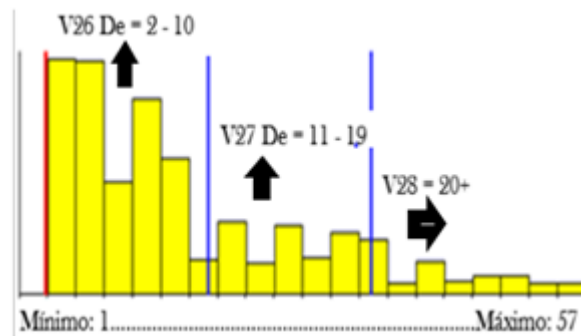
Fonte: Elaborada pelo autor

Figura 36 - Discretização: Renda Anua



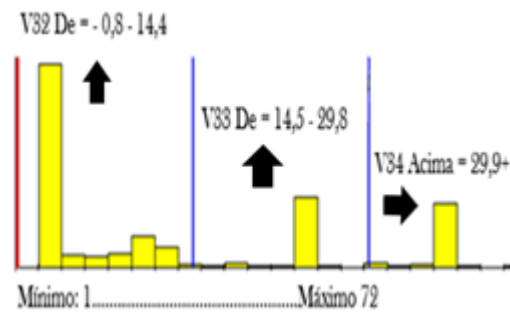
Fonte: O autor (2018)

Figura 37 - Discretização: Idade



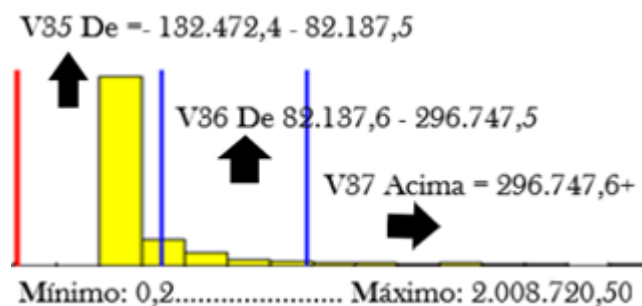
Fonte: Elaborada pelo autor

Figura 38 - Discretização: Número de Parcelas



Fonte: Elaborada pelo autor

Figura 39 - Discretização: Valor do Contrato



Fonte: Elaborada pelo autor

4.1.3 Transformação das variáveis em *dummies*

Quanto à variável *dummy*, assume-se que a relação existente entre a variável dependente e a variável explicativa, para as observações, é estável, sendo que a variável

categórica é desdobrada em duas ou mais variáveis *dummies* (MISSIO *et al*, 2007). Nas Tabelas 4 a 8 são ilustradas as codificações de cada nova variável *dummy* e a faixa de classificação, em substituição às variáveis primitivas, V6, V7, V11, V12, V13 e V14, para cada classe de empresas, separadas por categoria de faturamento e para os DG.

Tabela 4 - Transformação das variáveis dummy para MEI

Cód.	Saldo Devedor	SD1	SD2	SD3	SD4
V16	Até 1.947,51	1	0	0	0
V31	De 1.947,52 a 6.808,53	0	1	0	0
V17	De 6.808,54 a 15.564,58	0	0	1	0
V18	Acima de 15.564,59+	0	0	0	1
Cód.	Tempo Filiação	TF1	TF2	TF3	TF4
V19	Até 0,693	1	0	0	0
V20	De 0,694 a 2,56	0	1	0	0
V21	De 2,564 a 4,434	0	0	1	0
V22	Acima de 4,434+	0	0	0	1
Cód.	Renda Anual	RA1	RA2	RA3	RA4
V23	Até 37.717,00	1	0	0	-
V24	De 37.717,10 a 49.929,50	0	1	0	-
V25	De 49.929,60 a 62.142,10	0	0	1	-
Cód.	Idade	ID1	ID2	ID3	ID4
V26	De 1 a 5	1	0	0	-
V27	De 6 a 9	0	1	0	-
V28	Acima de 10+	0	0	1	-
Cód.	Número Parcelas	NP1	NP2	NP3	NP4
V32	De 1,1 a 11,5	1	0	0	-
V33	De 11,6 a 24,2	0	1	0	-
V34	Acima 24,3+	0	0	1	-
Cód.	Valor Contrato	VC1	VC2	VC3	VC4
V35	De 2.840,37 a 7.413,61	1	0	0	-
V36	De 7.413,62 a 17.667,59	0	1	0	-
V37	Acima 17.667,59+	0	0	1	-

Fonte: Elaborada pelo autor

Tabela 5 - Transformação das variáveis dummy para ME

Cód.	Saldo Devedor	SD1	SD2	SD3	SD4
V16	De 34.610,90 a 44.596,00	1	0	0	-
V17	De 44.596,10 a 123.803,1	0	1	0	-
V18	Acima 123.803,20+	0	0	1	-
Cód.	Tempo Filiação	TF1	TF2	TF3	TF4
V19	Até 1,36	1	0	0	0
V20	De 1,364 a 3,89	0	1	0	0
V21	De 3,895 a 6,426	0	0	1	0
V22	Acima de 6,426+	0	0	0	1
Cód.	Renda Anual	RA1	RA2	RA3	RA4
V23	De 26.450,10 a 483.449,70	1	0	0	-
V24	De 483.449,80 a 940.449,40	0	1	0	-
V25	Acima de 940.449,40+	0	0	1	-
Cód.	Idade	ID1	ID2	ID3	ID4
V26	Até 2	1	0	0	0
V27	De 3 – 11	0	1	0	0
V28	De 12 – 20	0	0	1	0
V29	Acima de 21+	0	0	0	1
Cód.	Número Parcelas	NP1	NP2	NP3	NP4
V32	De 1 a 14,6	1	0	0	-
V33	De 14,7 a 30,1	0	1	0	-
V34	Acima de 30,2+	0	0	1	-
Cód.	Valor Contrato	VC1	VC2	VC3	VC4
V35	De 35.896,4 a 48.352,3	1	0	0	-
V36	De 48.352,4 a 132.601,1	0	1	0	-
V37	Acima de 132601,2+	0	0	1	-

Fonte: Elaborada pelo autor

Tabela 6 - Transformação das variáveis dummy para PE

Cód.	Saldo Devedor	SD1	SD2	SD3	SD4
V16	De 40.931,70 a 365.978,0	1	0	0	-
V17	De 365.978,10 a 772.887,80	0	1	0	-
V18	Acima de 772.887,90+	0	0	1	-
Cód.	Tempo Filiação	TF_1	TF_2	TF_3	TF_4
V19	Até 1,37	1	0	0	0
V20	De 1,374a 3,99	0	1	0	0
V21	De 3,995 a 6,61	0	0	1	0
V22	Acima de 6,61+	0	0	0	1
Cód.	Renda Anual	RA1	RA2	RA3	RA4
V23	Até 2.815.525,80	1	0	0	0
V24	De 2.815.525,90 a 6.181.952,60	0	1	0	0
V25	De 6.181.952,70 a 9.548.379,50	0	0	1	0
V30	Acima de 9.548.379,50+	0	0	0	1
Cód.	Idade	ID1	ID2	ID3	ID4
V26	Até 4	1	0	0	0
V27	De 5 a 14	0	1	0	0
V28	De 15 a 24	0	0	1	0
V29	De 25+	0	0	0	1
Cód.	Número Parcelas	NP1	NP2	NP3	NP4
V32	De 1 a 18,2	1	0	0	-
V33	De 18,3 a 35,9	0	1	0	-
V34	Acima de 36,0+	0	0	1	-
Cód.	Valor Contrato	V_C_1	V_C_2	V_C_3	V_C_4
V35	De 57.309,9 a 408.091,4	1	0	0	0
V36	De 408.091,5 a 873.492,9	0	0	1	0
V37	Acima de 873.493,0+	0	0	0	1

Fonte: Elaborada pelo autor

Tabela 7 - Transformação das variáveis dummy para MédE

Cód.	Saldo Devedor	SD1	SD2	SD3	SD4
V16	Até 128.927,49	1	0	0	0
V17	De 128.927,50 a 681.249,39	0	1	0	0
V18	De 681.249,40 a 1.233.571,29	0	0	1	0
V31	Acima de 1.233.571,30+	0	0	0	1
Cód.	Tempo Filiação	TF1	TF2	TF3	TF4
V19	Até 1,473	1	0	0	0
V20	De 1,474 a 3,92	0	1	0	0
V21	De 3,921 a 6,368	0	0	1	0
V22	Acima de 6,368+	0	0	0	1
Cód.	Renda Anual	RA1	RA2	RA3	RA4
V23	Até 17.371.937,00	1	0	0	0
V24	De 17.371.937,10 a 32.201.966,90	0	1	0	0
V25	De 32.201.967,00 a 47.031.996,70	0	0	1	0
V30	Acima de 47.031.996,70+	0	0	0	1
Cód.	Idade	ID1	ID2	ID3	ID4
V26	Até 4	1	0	0	0
V27	De 5 a 14	0	1	0	0
V28	De 15 a 24	0	0	1	0
V29	Acima de 25+	0	0	0	1
Cód.	Número Parcelas	NP1	NP2	NP3	NP4
V32	De 1 a 17,6	1	0	0	-
V33	De 17,7 a 35,9	0	1	0	-
V34	Acima de 36+	0	0	1	-
Cód.	Valor Contrato	VC1	VC2	VC3	VC4
V35	De < 170.350,4	1	0	0	0
V36	De 170.350,5 a 729.489,6	0	1	0	0
V37	De 729.489,7 a 1.288.628,9	0	0	1	0
V38	Acima de 1.288.629,0+4.1	0	0	0	1

Fonte: Elaborada pelo autor

Tabela 8 - Transformação das variáveis dummy para DG

Cód.	Saldo Devedor	SD1	SD2	SD3	SD4
V16	De -121.272,7 a 75.167,3	1	0	0	0
V17	De 75.167,4 a 271.607,4	0	1	0	0
V18	Acima de 271.607,5+	0	0	1	0
Cód.	Tempo Filiação	TF1	TF2	TF3	TF4
V19	Até 1,17826	1	0	0	0
V20	De 1,17826 a 3,6645138	0	1	0	0
V21	De 3,66451388 a 6,150764	0	0	1	0
V22	Acima de 6,150764+	0	0	0	1
Cód.	Renda Anual	RA1	RA2	RA3	RA4
V23	De -3.847.981,3 a 1.509.292,2	1	0	0	0
V24	De 1.509.292,3 a 6.866.565,9	0	1	0	0
V25	Acima de 6.866.566,0+	0	0	1	0
Cód.	Idade	ID1	ID2	ID3	ID4
V26	De 2 a 10	1	0	0	0
V27	De 11 a 19	0	1	0	0
V28	Acima de 20+	0	0	1	0
Cód.	Número Parcelas	NP1	NP2	NP3	NP4
V32	De 1 a 14,4	1	0	0	0
V33	De 14,5 a 29,8	0	1	0	0
V34	Acima de 29,9+	0	0	1	0
Cód.	Valor Contrato	VC1	VC2	VC3	VC4
V35	De -132.472,4 a 82.137,5	1	0	0	0
V36	De 82.137,6 a 296.747,5	0	1	0	0
V37	Acima de 296.747,6+	0	0	1	0

Fonte: Elaborada pelo autor

4.2 REGRESSÃO LOGÍSTICA

Para estruturação dos modelos de pontuação de crédito, utilizou-se de 5 variáveis primitivas sendo: V1, V2, V4, V9, V10 e 23 variáveis *dummies*, (ou artificiais), em substituição as variáveis primitivas V6, V7, V11, V12, V13 e V14, totalizando 28 variáveis, conforme descrito na Figura 44. A utilização de variáveis artificiais no modelo, aumentou o número de variáveis na composição, bem como melhorou o percentagem da acurácia.

Para interpretação da Figura 40, é destacado a seguintes abreviaturas: P = Primitivas; CO = Correlacionadas com as variáveis independentes ou com comportamentos semelhantes

a variável dependente; PS = Primitivas Substituídas; PU = Primitivas Utilizadas e Ar = Artificiais.

Figura 40 - Cálculo das variáveis utilizadas nos testes de RL

Variáveis															Cálculo		
V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	=	15	P
		V3		V5			V8							V15	-	4	CO
V1	V2		V4		V6	V7		V9	V10	V11	V12	V13	V14		=	11	P
					V6	V7				V11	V12	V13	V14		-	6	PS
V1	V2		V4					V9	V10						=	5	PU
V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27				+	23	Ar
V28	V29	V30	V31	V32	V33	V34	V35	V36	V37	V38							
Total de Variáveis															=	28	P e Ar

Fonte: Elaborada pelo autor

Para seleção das variáveis independentes, as quais possuem o maior poder preditivo, a fim de incorporar o modelo de previsão para cada classe de empresas, utilizou-se o método *stepwise*. Esse método está disponível no *Software SPSS®*. Para o nível de corte adotou-se o valor de 0,50. Esse valor é utilizado para seleção e agrupamento de contratos “cliente” como inadimplente ou adimplente, os quais assumem a codificação de valor “0” e “1”, respectivamente, conforme ilustra Tabela 9.

Tabela 9 - Valores assumidos pela variável dependente RL

Variável	Valor
Inadimplente	0
Adimplente	1

Fonte: Elaborada pelo autor

Na Tabela 10 são apresentadas as variáveis inclusas nos modelos, escolhidas pelo método *stepwise*. Este método efetua uma seleção sequencial, objetivando identificar a variável que apresenta um maior poder preditivo para a regressão (HAIR *et al.*, 2009). Para cada classe de empresas o método selecionou variáveis significantes dentro da base de dados correspondente.

No volume de 28 variáveis independentes utilizadas como entrada nos modelos de pontuação de crédito, após utilização do método *stepwise*, obteve-se uma saída de 9 variáveis com poder de discriminação, para a técnica de RL, conforme apresentado na tabela 10.

Tabela 10 - Variáveis incluídas no modelo de RL

VARIÁVEIS	MEI	ME	PE	MédE	DG
V2	-	-	-	-	X
V9	X	X	X	X	X
V19	X	X	-	X	X
V21	-	X	-	-	-
V26	-	-	-	-	X
V27	-	X	-	-	-
V34	-	-	-	X	-
V35	-	X	-	-	-
V36	-	-	-	-	X

Fonte: Elaborada pelo autor

- Obs: Os espaços representados por um traço “-“, referem-se a ausência da variável que compõe o modelo dentro de cada classe.

A Tabela 11 ilustra o volume de casos que foram utilizados para a BT e BP. Observa-se que, dentro do banco de dados a ME é que contém o maior volume de casos e conseqüentemente o maior volume de negociações da instituição. Outro fator de destaque está na quantidade de variáveis independentes selecionadas pelo método *stepwise*.

Tabela 11 - Número para amostra teste e principal de casos válidos

Contratos	MEI	ME	PE	MédE	DG
(20%)	48	228	28	8	313
(80%)	212	852	87	28	1.178
(100%)	260	1080	115	36	1491
Porcentagem em relação ao total de contratos	17,44%	72,43%	7,71%	2,41%	100,00%

Fonte: Elaborada pelo autor

4.2.1 Teste de multicolinearidade

Este teste consiste no exame de correlação entre as variáveis independentes, ocorrendo quando duas ou mais variáveis explicativas tentam explicar o mesmo fato (CUNHA; COELHO, 2017). A colinearidade pode ser mediada pela tolerância e sua inversa, chamada de Fator de Inflação de Variância (VIF), sendo medidas bastante comuns para colinearidade (HAIR *et al.* 2009). A tolerância (T) é calculada como $1 - R^2$ e o VIF é calculado por meio do inverso da tolerância, quando o VIF for 1 e a tolerância for 1, implica dizer que não há

multicolinearidade (HAIR *et al.*, 2009). A tabela 12 apresenta a estatística de colinearidade para as 4 classes de empresas e DG.

Tabela 12 - Teste dos coeficientes de Colinearidade

	MEI	MEI	ME	ME	PE	PE	MédE	MédE	D G	D G
	T*	VIF**	T	VIF	T	VIF	T	VIF	T	VIF
V2	-	-	-	-	-	-	-	-	0,978	1,022
V9	0,970	1,030	0,854	1,171	1,000	1,000	0,896	1,116	0,943	1,061
V19	0,970	1,030	0,914	1,095	-	-	0,900	1,111	0,981	1,019
V21	-	-	0,867	1,153	-	-	-	-	-	-
V26	-	-	-	-	-	-	-	-	0,937	1,068
V27	-	-	0,972	1,029	-	-	-	-	-	-
V34	-	-	-	-	-	-	0,960	1,041	-	-
V35	-	-	0,881	1,135	-	-	-	-	-	-
V36	-	-	-	-	-	-	-	-	0,903	1,107

Fonte: Elaborada pelo autor

- *Tolerância
- **Obs: Os valores representados por “-“, referem-se a variáveis independentes, não utilizadas na classe de empresas correspondente no modelo respectivo.

Analisando os resultados, os valores de tolerância ficaram muito próximos de 1, e o VIF também bastante próximo de 1 e distante de 10. Segundo Hair *et al.* (2009), uma referência de corte muito comum é um valor de tolerância de 0,10, o que corresponde a um valor VIF de 10.

4.2.2 Teste da Verossimilhança

O teste de Log Likelihood Value, objetiva estimar a probabilidade de um evento ocorrer, aferindo a capacidade do modelo (DIAS FILHO; CORRAR 2017). O teste, também se mostra importante para verificar se o modelo apresenta melhora com a inclusão ou retirada de variáveis independentes, conforme ilustrado na Figura 41. Na RL, é estimado um modelo base, o qual tem a função de servir como padrão para comparações, utilizando a soma dos quadrados das médias para estabelecer o valor do logaritmo da verossimilhança $\{-2LL\}$ (HAIR *et al.*, 2009).

Figura 41 - Teste de verossimilhança, Cox & Snell e Nagelkerke

MEI				ME			
Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke	Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	169,284	0,168	0,268	1	640,259	0,278	0,422
2	149,379	0,242	0,388	5	585,234	0,323	0,491
PE				MédE			
Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke	Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	57,707	0,226	0,376	1	177,214	0,173	0,279
				3	150,313	0,266	0,427

Fonte: Elaborada pelo autor

Valores menores de medida $-2LL$, melhoram o ajuste do modelo, sendo esta técnica utilizada pelo método *stepwise* para melhora do passo (etapa) anterior (HAIR *et al.*, 2009). Observa-se que em todas as fases houve redução da verossimilhança para todas as etapas das classes de empresas. Dias Filho e Corrar (2009), salientam que o $-2LL$ sozinho apresenta pouca significância intrínseca, desta forma, para contornar esta questão, se estabelece uma base de comparação, assim verificando se o indicador aumenta ou diminui.

Outro instrumento utilizado para medição de modelos concorrentes é Cox & Snell. Esta medida, não pode atingir o valor máximo de 1, ficando limitada (HAIR *et al.*, 2009). Para tanto Nagelkerke, propôs um domínio de 0 a 1, modificando a medida de Cox & Snell. As duas medidas refletem a quantia de variação explicada pelo modelo logístico, com o valor de 1,0 indicando o ajuste perfeito (HAIR *et al.*, 2009).

A Tabela 13, ilustra os resultados para o teste de *Hosmer e Lemeshow*. Segundo Dias Filho e Corrar (2017), este teste tem finalidade de verificar se existem diferenças significativas entre as classificações realizadas pelo modelo e as realidades observadas. Sua análise é com base na significância do modelo, o qual é favorável quando o nível de significância é igual ou superior a 0,05. Conforme Portal Action (2018), o teste de *Hosmer-Lemeshow* é bastante utilizado em regressão logística, e tem por finalidade testar a bondade do ajuste, pode-se entender que, o teste verifica se o modelo proposto pode explicar bem o que se observa, de forma diferente, pode-se dizer que o teste avalia o modelo ajustado por meio das distâncias entre as probabilidades ajustadas e as probabilidades observadas.

Das classes apresentadas na Tabela 13, a única classe que não apresentou um nível significativo foi a PE, assim rejeitando a hipótese nula de não haver diferenças significativas. Entretanto, o *stepwise* apresentou apenas uma etapa, não havendo outro nível para efetuar comparação.

Tabela 13 - Teste de Hosmer e Lemeshow

MEI	MEI	MEI	MEI	ME	ME	ME	ME
Etapa	Qui-quadrado	df	Sig.	Etapa	Qui-quadrado	df	Sig.
1	3,735	5	0,588	1	14,107	6	0,028
2	3,410	5	0,637	5	15,001	8	0,059
PE	PE	PE	PE	MédE	MédE	MédE	MédE
Etapa	Qui-quadrado	df	Sig.	Etapa	Qui-quadrado	df	Sig.
1	18,694	6	0,005	1	9,569	5	0,088
				3	10,731	7	0,151

Fonte: Elaborada pelo autor

4.2.3 Modelos de pontuação de crédito com a técnica estatística de RL

Conforme Dias Filho e Corrar (2017), a Equação Logística, descreve a probabilidade relativa à ocorrência de determinado evento, podendo assumir o seguinte formato:

Figura 42 - Equação Logística

$$P(\text{evento}) = \frac{1}{1 + e^{-(b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki})}}$$

Fonte: Adaptado de Dias Filho e Corrar (2017)

As Figuras de 43 a 47, ilustram os pesos atribuídos a cada variável independente incorporada ao modelo correspondente a classe de empresas que foi selecionado pelo método *stepwise*, bem como pode ser observado a Equação Logística.

Figura 43 - Equação Logística para MEI

Variáveis	Peso	Equação Logística
V9	0,952	$\frac{1}{1 + e^{-(-1,274 + 0,952(V9) + 20,861(V19))}}$
V19	20,861	
(Constante)	-1,274	

Fonte: Elaborada pelo autor

O modelo, ilustrado na Figura 43, que compõe a Equação Logística para as empresas classificadas como MEI, destaca a inclusão de 2 variáveis, sendo uma primitiva (V9) e outra artificial (V19).

Figura 44 - Equação Logística para ME

Variáveis	Peso	Equação Logística
V9	0,942	$\frac{1}{1 + e^{-(-1,734 + 0,942(V9) + 2,222(V19) + 0,628(V21) - 1,020(V27) + 0,744(V35))}}$
V19	2,222	
V21	0,628	
V27	-1,020	
V35	0,744	
(Constante)	-1,734	

Fonte: Elaborada pelo autor

O modelo, ilustrado na Figura 44, que compõe a Equação Logística para as empresas classificadas como ME, destaca a inclusão de 5 variáveis, sendo uma primitiva (V9) e 4 variáveis artificiais (V19), (V21), (V27) e (V35)

Figura 45 - Equação Logística para PE

Variáveis	Peso	Equação Logística
V9	0,753	$\frac{1}{1 + e^{-(-0,608 + 0,753(V9))}}$
(Constante)	-0,608	

Fonte: Elaborada pelo autor

O modelo, ilustrado na Figura 45, que compõe a Equação Logística para as empresas classificadas como PE, destaca a inclusão de apenas uma variável, sendo ela primitiva (V9). Esta questão de haver apenas uma variável para compor o modelo está de acordo com a questão acadêmica, devido a respeitar as premissas da técnica, porém pode dificultar sua aceitação a nível de gestão. O aprimoramento do modelo com a inclusão de variáveis de gestão ou de restrição de risco podem melhorar a confiabilidade no modelo.

Figura 46 - Equação Logística para MédE

Variáveis	Peso	Equação Logística
V9	0,753	$\frac{1}{1 + e^{-(-1,505 + 0,992(V9) + 21,013(V19) + 2,284(V34))}}$
V19	21,013	
V34	2,284	
(Constante)	-1,505	

Fonte: Elaborada pelo autor

O modelo, ilustrado na Figura 46, que compõe a Equação Logística para as empresas classificadas como MédE, destaca a inclusão de 3 variáveis, sendo uma primitiva (V9) e 2 variáveis artificiais (V19) e (V34).

Figura 47 - Equação Logística para DG

Variáveis	Peso	Equação Logística
V2	-0,229	$\frac{1}{1 + e^{-(-0,31 - 0,229(V2) + 0,909(V9) + 2,057(V19) - 1,106(V26) - 1,099(V36))}}$
V9	0,909	
V19	2,057	
V26	-1,106	
V36	-1,099	
(Constante)	-0,031	

Fonte: Elaborada pelo autor

O modelo, ilustrado na Figura 47, que compõe a Equação Logística para as empresas classificadas como DG, destaca a inclusão de 5 variáveis, sendo duas primitivas (V2), (V9) e 3 variáveis artificiais (V19), (V26) e (V36).

Pode-se observar que o modelo de DG, por meio do método *stepwise*, incluiu o máximo de 5 variáveis, apresentando o mesmo número de variáveis do modelo da ME, sendo esta, a classe que apresenta maior volume de dados.

4.2.4 Teste da acurácia para Regressão Logística

Nas Figuras de 48 a 52, são apresentados os resultados dos testes de acurácia da BT, e da BP para cada uma das classes de empresas e DG. Na Tabela 14 é apresentado um resumo das acurácias dos testes, já apresentados nas Figuras 48 a 52, juntamente com o teste de acurácia sobre os DG, envolvendo 1.491 casos.

Ao observar a Figura 48 os resultados do teste de acurácia para MEI, apresentaram uma superioridade de 3,8% pontos percentuais, da BP sobre a BT, mas ambos os indicadores acima de 65%. A acurácia geral atingiu percentual de 83,0%, no qual é interessante observar que a predição para o adimplente atingiu percentagem de 93,6%, apresentando maior poder discriminatório para identificação do adimplente. Devido à baixa quantidade de dados que compôs a amostra teste, não foram selecionados observações para inadimplentes na BT. O valor de corte atribuído foi de 0,50.

Figura 48 - Resultados dos testes de acurácia para classe de empresas classificadas como MEI

	MEI Base Principal			MEI Base Teste		
	INAD	ADIM	%	INAD	ADIM	%
INADIMLENTE	16	25	39,0	0	10	0
ADIMPLENTE	11	160	93,6	0	38	100
Percentagem Global			83,0			79,2

Fonte: Elaborada pelo autor

Na Figura 49, a acurácia para a classe de ME, a BT e BP apresentam uma pequena diferença de 2,8% pontos percentuais, esta proximidade demonstra uma boa adaptabilidade do modelo, com acurácia geral de 84,9%. A Microempresa, na BP, apresentou um poder de previsão, para o inadimplente de 64,1%, e uma acurácia bastante significativa para o adimplente de 91,1%.

Figura 49 - Resultados dos testes de acurácia para as empresas classificadas como ME

	ME Base Principal			ME Base Teste		
	INAD	ADIM	%	INAD	ADIM	%
INADIMPLENTE	125	70	64,1	34	14	70,8
ADIMPLENTE	59	603	91,1	14	166	92,2
Percentagem Global			84,9			87,7

Fonte: Elaborada pelo autor

A Figura 50 Ilustra os resultados da acurácia para as empresas classificadas na classe de PE. O nível de acurácia entre a BT e BP apresentaram uma diferença de 11,5% pontos percentuais, demonstrando a capacidade de predição do modelo, sobre a BP de 88,5%, composta de 212 contratos. A percentagem de acerto, sobre o inadimplente, deve ser analisada com cuidado, devido a apresentar percentual abaixo de 50%, este fato pode estar relacionado com o volume de contratos incluídos no teste.

Figura 50 - Resultados dos testes de acurácia para classe de empresas classificadas como PE

	PE Base Principal			PE Base Teste		
	INAD	ADIM	%	INAD	ADIM	%
INADIMPLENTE	6	9	40	4	6	40,0
ADIMPLENTE	1	71	98,6	2	36	94,7
Percentagem Global			88,5			83,3

Fonte: Elaborada pelo autor

Na Figura 51 é Ilustrada a acurácia obtida para a classe das MédE. O volume de contratos utilizados na BP foi de 227 contratos, apresentando uma acurácia de 83,3%. A BT apresentou uma acurácia de 100% tanto para adimplentes como para inadimplentes, no entanto, o volume de contratos que compôs a BT, foi de apenas 12, o que pode distorcer os valores observados.

Figura 51 - Resultados dos testes de acurácia para as empresas classificadas como MédE

	MédE Base Principal			MédE Base Teste		
	INAD	ADIM	%	INAD	ADIM	%
INADIMPLENTE	15	28	35	2	0	100
ADIMPLENTE	10	174	94,6	0	10	100
Percentagem Global			83,3			100

Fonte: Elaborada pelo autor

Posteriormente a aplicação da técnica da RL, nas classes: MEI, ME, PE e MédE, sobre o volume de contratos/clientes, utilizou-se a base geral, aqui denominada de DG, com 1.491 observações válidas, nos quais foram aplicados os mesmos procedimentos das classes de empresas.

Os resultados estão apresentados na Figura 52, sendo que a BT e a BP apresentaram acurácias praticamente iguais, ou seja 85,4% e 85,0%, respectivamente, apresentando uma boa capacidade de previsão do modelo, mostrando, principalmente que, a acurácia destinada a prever o adimplente, teve bom desempenho de predição. Ao analisar o inadimplente, o desempenho foi apenas regular, ficando na casa de 59% e 52,5%.

Conforme ilustrado na Figura 52, a acurácia do modelo de DG de 85,0%, utilizada sobre a totalidade da base de dados.

Figura 52 - Resultados dos testes de acurácia para a base de DG

	DG Base Principal			DG Base Teste		
	INAD	ADIM	%	INAD	ADIM	%
INADIMPLENTE	134	121	52,5	49	34	59,0
ADIMPLENTE	55	861	94,0	12	220	94,8
Percentagem Global	-	-	85,0	-	-	85,4

Fonte: Elaborada pelo autor

Na observação da Tabela 14, os DG ficaram com percentuais próximos ao das classes de empresas, principalmente MEI e ME, porém, mesmo pequenas variações percentuais podem representar valores significantes. Uma observação importante, é que os modelos com acurácia similares, são aqueles que possuem as maiores bases de dados, entretanto as classes da PE e MédE, são as classes que possuem menores distâncias dentro da base de dados, conforme pode-se observar nas discretizações.

Tabela 14 - Comparativo da acurácia das classes de empresas com a base de DG

CLASSES	MEI	ME	PE	MédE	DG
Acurácia	83%	84,9%	88,5%	83,0%	85,0%

Fonte: Elaborada pelo autor

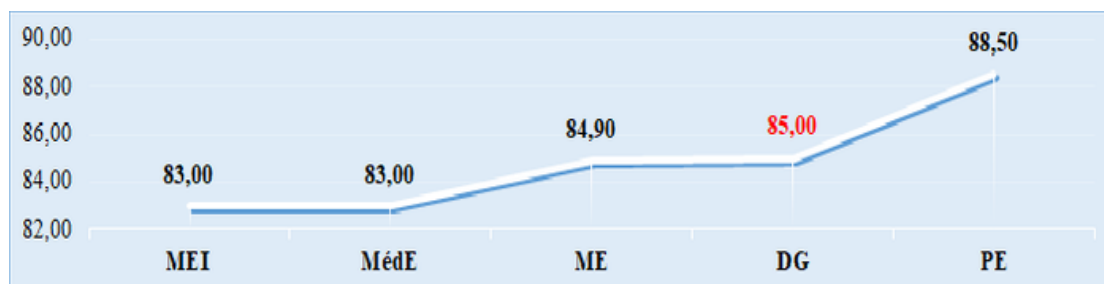
Antes de ser efetuada uma análise da predição das técnicas utilizadas neste trabalho é importante relatar a questão de atendimento aos pressupostos da técnica, conforme descritos na Tabela 3. Pode-se observar que a RL apresenta uma boa aderência aos pressupostos. Para Dias Filho e Corrar (2017), o modelo logístico acolhe com mais facilidade variáveis categóricas, sendo uma das razões que se torna boa alternativa à AD.

Ao analisar a Tabela 14, verifica-se que a classe de empresas PE, apresenta a melhor percentagem preditiva, porém ao observar a composição das variáveis do modelo para PE, conforme Figura 45, apenas 1 variável foi selecionada pelo método *stepwise*. Para uso prático do modelo para PE, recomenda-se que outras variáveis, não contempladas neste estudo, como dados levantados diretamente pela gestão ou de restrição de crédito, e verificadas novamente pelo método *stepwise*, sejam testadas para inclusão no modelo, objetivando dar maior robustez ao modelo.

Na Tabela 11, também é possível observar que a classe de empresas ME, é a que mais possui vínculo de negócios na carteira de crédito com a Instituição Financeira, representando 72,43% do total dos contratos. Este percentual demonstra que a grande maioria dos clientes da instituição, apresentam-se com faturamento dentro das características da ME. Devido a este fator a acurácia desta técnica torna-se importante para análise da carteira de clientes da instituição, a qual performou uma percentagem de 84,9%.

Em resposta ao objetivo específico "c", observou-se que as percentagens da acurácia das 4 classes de empresas: MEI, ME, PE e MédE, são próximos, porém, não iguais à dos DG. A Figura 53 ilustra a distribuição da acurácia em ordem crescente, onde pode-se observar que as classes MEI, MédE e ME encontram-se até 2 pontos percentuais abaixo dos DG e a classe PE encontra-se 3,5 pontos percentuais acima dos DG.

Figura 53 - Distribuição da Acurácia, em ordem crescente



Fonte: Elaborada pelo autor

Essas diferenças percentuais, principalmente a da ME que representa 72,43% dos negócios gerados pela Instituição, devem ser analisadas dentro de cada classe em comparativo ao risco que a classe apresenta junto a Instituição Financeira.

4.3 DISCUSSÃO DO PODER DISCRIMINATÓRIO DA TÉCNICA RL

Para obtenção dos pesos das variáveis, apresentados no item 4.3.3 para RL, em resposta ao objetivo específico “a”, observou-se que a variável independente artificial com maior poder discriminatório foi a V19 (tempo de filiação). Esta variável está presente nos modelos das classes da MEI, ME, MédE e também nos DG. Outra variável independente, com significância para o modelo de RL, é a V9 (Número de Produtos), que corresponde ao bens e serviços que o associado ou cliente possui junto a Instituição. A variável V9 figurou em todas as classes de empresas e DG. Esta variável é utilizada nos modelos de forma primitiva, não havendo modificações ou transformações em suas características.

As classes de empresas, que apresentam maior volume de dados (contratos), apresentam também uma maior proximidade de características como variáveis selecionadas pelo método *stepwase* e percentagem de acurácia. Essa constatação, destaca a importância da análise das premissas quanto ao número de casos que devem compor a amostra ou banco de dados para validação dos resultados.

No tocante ao atendimento dos pressupostos, necessários para a correta aplicação das técnicas, observou-se que a RL permite boa aderência das variáveis na composição dos modelos.

Conforme análise das classes de empresas e DG, observou-se que os resultados da acurácia geral ficaram com valores preditivos adequados com a literatura. Os percentuais acompanharam outros resultados em trabalhos como de Mselmi, Lahiani e Hamza (2017); Li *et al.* (2016), Smaranda (2014); Wang e Zhou (2011); Ciampi e Gordini (2008).

5 CONSIDERAÇÕES FINAIS

O mercado financeiro de fomento às empresas, principalmente para as PMEs, tem na análise de crédito seu ponto decisório entre fornecer ou não a solicitação de crédito. Esta decisão cabe aos especialistas, responsáveis pelo setor de análise de crédito, no entanto, este fato de decidir está atrelado ao risco. Conforme discorre Silva (2016), ao conceder um empréstimo ou financiamento a empresa assume o risco de não receber, ou seja, o cliente pode não cumprir a promessa de pagamento da parcela.

Em observação a crescente busca por técnicas de auxílio, que possam gerar indicativos ou mensurações de risco, este trabalho aplicou a técnica de RL sobre o banco de dados de uma Instituição Financeira. Devido as peculiaridades dos setores, tamanhos e “*know-how*”, entre outros fatores, que podem ter relação com o risco, buscou-se desenvolver modelos de pontuação de crédito para cada classe das empresas, bem como apresentou uma pontuação de crédito para os DG. As empresas de porte menor são mais sensíveis, sendo que, em épocas de restrições financeiras de mercado, são as primeiras a enfrentar dificuldades financeiras e consequentemente as últimas a sair destas dificuldades (SILVA, 2016).

Diante do exposto, trabalhou-se com 4 classes de empresas, objetivando elaborar um modelo para cada classe e mais 1 modelo para DG. Foi possível observar que, apesar de algumas variáveis serem comuns a todas as classes e DG, outras variáveis independentes incorporaram-se unicamente em um ou outro modelo. Esta constatação torna-se importante, pois permite verificar que para determinada classe de empresas, tal variável deve ser analisada de forma específica, justificando a elaboração do modelo por classe de empresas.

Quanto aos percentuais de acurácia da RL, conforme revisão de literatura, apresentado no item 2 deste trabalho, obteve-se bons resultados para as classes de empresas e DG, demonstrando que, apesar de existirem outras técnicas que surgiram posteriormente, como as computacionais ou algoritmos genéticos, a RL ainda é uma técnica eficiente e de ampla aplicabilidade.

Os valores alcançados das acurácias, de forma geral, para a técnica testada, foram superiores a 65%, ficando, os resultados obtidos, alinhados aos observados nas pesquisas recentes de Prado *et al.* (2016); Louzada, Ara e Fernandes (2016); Selau e Ribeiro (2011); Mselmi, Lahiani e Hamza (2017). Destaca-se que as variáveis utilizadas nos modelos, não são oriundas dos indicadores Contábeis, Financeiros ou Econômicos, e sim oriundos de fontes de cadastro e histórico de crédito e clientes.

Os modelos apresentados, com o uso da técnica da RL, podem servir de base para as instituições que trabalham com fomento financeiro e que tenham em seu portfólio empresas classificadas como MEI, ME, PE e MédE, principalmente na identificação do adimplente, pois as informações geradas ou fornecidas por este tipo ou classes de empresas são praticamente as mesmas, exemplo: faturamento, tempo de filiação, número de produtos,

idade, entre outras.

Vale ressaltar, que a incorporação de outras variáveis aos modelos é interessante, podendo aumentar o poder de acurácia, porém, neste caso, deve-se efetuar novamente o tratamento das premissas e do método *stepwise* para o conjunto integral das variáveis. Mais pontualmente, quanto ao modelo para PE, a inclusão de variáveis de gestão e variáveis de restrição podem, desde que apresentem poder preditivo, melhorar o modelo, proporcionando uma maior segurança para seu uso em termos de gestão.

Outro fator que contribuiu para a melhoria dos modelos apresentados foram a discretização de variáveis e a criação de variáveis artificiais “*dummies*”, processos que ajudam a amenizar as diferenças dentro das classes de dados.

Devido as menores exigências quanto às premissas, a RL mostrou-se adequada na aplicação dos testes junto a base de dados específica da Instituição Financeira, apresentando maior facilidade de manipulação. Com base nos resultados, observa-se que a técnica de RL teve boa adaptabilidade às estruturas e base de dados da Instituição Financeira, sendo uma boa opção, para compor o ferramental decisório, no tratamento dos dados, com a finalidade de predição dos adimplentes ou inadimplentes. Desta forma, em conformidade com o trabalho de revisão da literatura de Fuhr, Lima e Schenatto (2017), melhor descrito no item 3.2, justifica-se a escolha da técnica preditiva RL.

Ainda, pode-se destacar as peculiaridades dos modelos para Pessoa Jurídica ou Equiparada, que são: divisão da base em classes de empresas, respeitando suas diferenças e a utilização de variáveis não contábeis ou auditáveis na sua totalidade.

Por fim, deixa-se como sugestão de estudos, a utilização de outros métodos, como os computacionais, algoritmos genéticos ou sistemas híbridos, na busca de indicadores com maior poder preditivo para as 4 classes de empresas, bem como, o estudo para inclusão de outras variáveis na composição dos modelos.

REFERÊNCIAS

- ABDOU, H. A.; POINTON, J. Credit Scoring, Statistical Techniques And Evaluation Criteria: A Review Of The Literature. **Journal of Small Business Management.** , v. 88, p. 59-88, 2011.
- ALTMAN, E I. The Prediction of Corporate Bankruptcy: A Discriminant Analysis. **The Journal of Finance**, v. 23, n. 1, p. 193-194, 1968.
- ALTMAN, E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. **The Journal of Finance** , v. XXIII, n. 4, p. 589-609, 1968.
- ALTMAN, E. I; CAOUILLE, J. B.; NARAYANAN, P. Credit-risk measurement and management: The ironic challenge in the next decade. **Financial Analysts Journal**, v. 54, n. 1, p. 7-11, 1998.
- ALVIM, P. C. R. D. C.. O papel da informação no processo de capacitação tecnológica das micro e pequenas empresas. **Ciência da Informação**, v. 27, p. 28-35, 1998.
- ARAÚJO, E. A; CARMONA, C. U. M. Construção de Modelos de Credit Scoring com Análise Discriminante e Regressão Logística para a Gestão do Risco de Inadimplência de uma Instituição de Microcrédito. **REAd**, v. 15, n. 1, 2009.
- BANCO CENTRAL DO BRASIL - BACEN. **Composição e segmentos do Sistema Financeiro Nacional**. Disponível em:
<http://www.bcb.gov.br/pec/depep/spread/rebc_2014.pdf>. Acesso em: 18 fev. 2017.
- _____. **Composição e segmentos do Sistema Financeiro Nacional**. Disponível em:
<<https://www.bcb.gov.br/pre/composicao/composicao.asp>>. Acesso em: 11 nov. 2017.
- _____. **Dados sobre a Evolução do Sistema Financeiro**. Disponível em:
<<https://www.bcb.gov.br/htms/novaPaginaSPB/PapelDosIntermediariosFinanceiros>>. Acesso em: 11 nov. 2017.
- _____. **O que é e o que faz o Banco Central**. Disponível em:
<<https://www.bcb.gov.br/pre/portalCidadao/bcb/bcFaz.asp?idpai=SFNCOMP>>. Acesso em: 17 ago. 2018.
- BANCO CENTRAL DO BRASIL - BACEN. Resolução. Diário Oficial 30 de abril de 2009. Disponível em: <<http://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo>>. Acesso em: 11 nov. 2017.
- BEAVER, W. H. Financial Ratios as Predictors of Failure. **Wiley on behalf of Accounting Research Center, Booth School of Business, University of Chicago** , v. 4, p. 71-111, 1966.
- BERGER, N. A.; COWAN, A. M.; FRAME, W. S. The surprising use of credit scoring in small business lending by community banks and the attendant effects on credit availability, risk, and profitability. **Journal of Financial Services Research**, v. 39, n. 1, p. 1-17, 2011.
- BERTI, A. **Consultoria e Diagnóstico Empresarial**. 2. ed. Curitiba: Juruá, 2012.

BEUREN, I. M. **Como elaborar trabalhos monográficos em contabilidade**: : teoria e prática. 3. ed. São Paulo: Atlas, 2014. (9. Reimpr.).

BNDES – BANCO NACIONAL DE DESENVOLVIMENTO. **Classificação de porte de empresa**. Disponível em:

<https://www.bndes.gov.br/wps/portal/site/home/imprensa/noticias/conteudo/20100622_modificacao_porte_empresa>. Acesso em: 11 nov. 2017.

BOLETIM Responsabilidade Social e Ambiental do Sistema Financeiro. **BANCO CENTRAL DO BRASIL -BACEN**, v. 5, n. 50, 2010.

BRASIL. Comitê de Pronunciamentos Contábeis. Pronunciamento n. 00 02 de dezembro de 2011. Diário Oficial 15 de dezembro de 2011. Disponível em:

<<http://www.cpc.org.br/CPC/Documentos-Emitidos/Pronunciamentos/Pronunciamento?Id=80>>. Acesso em: 5 jul. 2018.

_____. Presidência da República - Casa Civil. LEI COMPLEMENTAR n. 123 14 de dezembro de 2006. Diário Oficial. Disponível em:

<http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp123.htm>. Acesso em: 7 jul. 2018.

BRITO, G. A. S.; ASSAF NETO, A. Modelo de Classificação de Risco de Crédito de Empresas: A Model for the Classification of Companies Credit Risk. **Revista Contabilidade e Finanças**. São Paulo, v. 19, n. 46, p. 18-29, 2008.

BRUNI, A. L. **SPSS: Guia prático para pesquisadores**. 1. ed. São Paulo: Atlas, 2012.

CAMARGOS, M. A.; CAMARGOS, M. C. S.; ARAÚJO, E. A. Inadimplência em um programa de crédito de uma instituição financeira pública de minas gerais: uma análise utilizando regressão logística. **REGE**. São Paulo - SP, v. 19, n. 3, p. 467-486, 2012.

CHEN, S.; GOO, Y. J.; SHEN, Z. A Hybrid Approach of Stepwise Regression, Logistic Regression, Support Vector Machine, and Decision Tree for Forecasting Fraudulent Financial Statements. **The Scientific World Journal**, n. 1-9, p. 2014.

CIAMPI, F.; GORDINI, N. Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises. **Journal Of Small Business Management**, v. 51, n. 1, p. 23-45, 2013.

_____. Using Economic-Financial Ratios for Small Enterprise Default Prediction Modeling: An Empirical Analysis. **Oxford Business & Economics Conference Program**, 2008.

CUNHA, J. V. A.; COELHO, A. C. Regressão Linear Múltipla. In: CORRAR, L. J. (Coord.); PAULO, E. (Coord.); DIAS FILHO, J. M. (Coord.). **Análise Multivariada**: para os Cursos de Administração, Ciências Contábeis e Economia. São Paulo: Atlas, 2017. cap. 3.

DEAKIN, E. B. A Discriminant Analysis of Predictors of Business Failure: Research Reports. **Wiley-Blackwell on behalf of Accounting Research Center, Booth School of Business, University of Chicago**, v. 10, n. 1, p. 167-179, 1972.

DIAS FILHO, J. M.; CORRAR, L. J. Regressão Logística. In: CORRAR, L. J. (Coord.); PAULO, E. (Coord.); DIAS FILHO, J. M. (Coord.). **Análise Multivariada**: Para os Cursos de Administração, Ciências Contábeis e Economia. São Paulo: Atlas, 2017. cap. 5.

DIEESE. DEPARTAMENTO INTERSINDICAL DE ESTATÍSTICA E ESTUDOS SOCIOECONÔMICOS. **Nota Técnica**: Evolução do crédito na economia brasileira, 2008-2013. 2014. Disponível em: <<https://www.dieese.org.br/notatecnica/2014/notaTec135Credito.pdf>>. Acesso em: 27 ago. 2018.

DUARTE, S. V.; FURTADO, M. S. V. **Trabalho de conclusão de curso em Ciências Sociais Aplicadas**. 1. ed. Saraiva, 2014.

DURAND, D.. Risk elements in consumer instalment financing: National Bureau of Economics. New York, 1941.

FONSECA, R. A. et al. Importância Das Cooperativas De Crédito Como Agentes De Desenvolvimento Regional. **Revista Opara: Ciências Contemporâneas Aplicadas**, v. 3, n. 1, 2013.

FUHR, F.; LIMA, J. D.; SCHENATTO, F. J. Uma Revisão Sistemática da Literatura Sobre Credit Scoring. In: CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO, VII. 2017.

GALINARI, R. D.; COSTA, S. C. C.; TEIXEIRA JR, L. S. O apoio do BNDES a micro, pequenas e médias empresas por meio de redes empresariais: experiências recentes e perspectivas. **BNDES Setorial 43, publicação semestral**, p. 81-118, 2016.

GARCÍA, S. et al. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. **IEEE Transactions on Knowledge and Data Engineering**, v. 25, n. 4, p. 734-750, 2013.

GARCÍA, V.; MARQUÉS, A. I. ; SÁNCHEZ, J. S. An insight into the experimental design for credit risk and corporate bankruptcy prediction systems. **Journal of Intelligent Information Systems**, 2014.

GEPP, A.; KUMAR, K; BHATTACHARYA, S. Business Failure Prediction using Decision Trees. **Journal Of Forecasting**, v. 29, n. 6, p. 536-555, 2010.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2008.

_____. **Métodos e Técnicas de Pesquisa Social**. 6. ed. São Paulo: Atlas, 2008.

GONÇALVES, E. B. ; GOUVÊA, M.A. ; MANTOVANI, D. M. N. Análise de risco de crédito com o uso de regressão logística. **Revista Contemporânea de Contabilidade**, v. 10, n. 20, p. 139-160, 2013.

GONÇALVES, E. B. **Análise de Risco de Crédito com o uso de Modelos de Regressão Logística, Redes Neurais e Algoritmos Genéticos**: Dissertação de Mestrado apresentada para o Departamento de Administração da Faculdade de economia, Administração e Contabilidade da Universidade de São Paulo. 2005. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/12/12139/tde-09042008-144032/pt-br.php>>. Acesso em: 26 mai. 2005.

GORDINI, N. A genetic algorithm approach for {SMEs} bankruptcy prediction: Empirical evidence from Italy. **Expert Systems With Applications**, v. 41, n. 14, p. 6433-6445, 2014.

HAIR JR, J. F. et al. **Análise Multivariada De Dados**. Tradução Tradução Adonai Schlup Sant'Anna . 6. ed. Porto Alegre: Bookman, 2009.

HILLEGEIST, S. A.; CRAM, E. K.; LUNDSTEDT, D. P. Assessing the Probability of Bankruptcy. **Review of Accounting Studies**, v. 9, p. 5-34, 2004.

HOJI, M. **Administração Financeira e Orçamentária**: Matemática financeira aplicada, Estratégias financeiras, Orçamento empresarial. 9. ed. São Paulo: Atlas, 2010.

_____. **Administração financeira**: uma abordagem prática: matemática financeira aplicada, estratégias financeiras, análise, planejamento e controle financeiro. 4. ed. São Paulo: Atlas, 2003.

HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**: p. cm. (Wiley series in probability and mathematical statistics. Applied probability and statistics section). 2. ed. A Wiley-Interscience publication, 1989.

KANITZ, S.C. Como Prever Falências de Empresas. **Negócios em Exame**, p. 95-102. 1974. Disponível em: <<https://pt.scribd.com/document/61123852/EXAME-ComoPreverFalenciaEmpresa-Kanitz>>. Acesso em: 7 jul. 2018.

KAO, L. J. ; CHIU, C. C.; CHIU, F. Y. A Bayesian latent variable model with classification and regression tree approach for behavior and credit scoring. **Knowledge-Based Systems**, v. 36, p. 245-252, 2012.

KIM, H. S.; SOHN, S. Y. Support vector machines for default prediction of SMEs based on technology credit. **European Journal of Operational Research**, p. 838-846, 2010.

KLEINBAUM, D. G.; KLEIN, M. **Logistic Regression: A Self-Learning Text**. Third . ed. Atlanta, GA: Springer, 2010.

LAKATOS, E. M. ; MARCONI, M. D. A. **Fundamentos de metodologia científica**. 7. ed. São Paulo: Atlas, 2010.

LEMOS, E. P.; STEINER, M. T. A.; NIEVOLA, J. C. Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining. **R. Adm.** São Paulo, v. 40, n. 3, p. 225-234, 2005.

LI, H.; SUN, J.; WU, J. Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. **Expert Systems with Applications**, v. 37, n. 8, p. 5895-5904, 2010.

LI, K. et al. Financial Innovation Credit Default Model For SME Lending. **Expert Systems With Applications**, v. 61, p. 343-355, 2016.

LIMA, J. D.; MARQUES, J. M. Encontro Nacional de Engenharia de Produção. In: ANÁLISE ECONÔMICO-FINANCEIRA DE EMPRESAS SOB A ÓTICA DA ESTATÍSTICA MULTIVARIADA, XXII. 2002, Curitiba.

LOUZADA, F.; ARA, A.; FERNANDES, G. B. Classification methods applied to credit scoring: Systematic review and overall comparison. **Surveys in Operations Research and Management Science**, v. 21, n. 2, p. 117-134, 2016.

LUGOVSKAYA, L.. Predicting default of Russian SMEs on the basis of financial and non-financial variables. **Journal of Financial Services Marketing**, v. 4, n. 14, p. 301-313, 2010.

LUNET, N.; SEVERO, M.; BARROS, H. **Desvio Padrão ou Erro Padrão**: Arquivos de Medicina. 2006. Disponível em: <<http://www.scielo.mec.pt/pdf/am/v20n1-2/v20n1-2a08.pdf>>. Acesso em: 7 jul. 2006.

MAGRO, C. B. D.; MONDINI, V. E. D.; HEIN, N. Gestão dos Riscos de Inadimplência dos Tomadores de Crédito: Um Estudo em uma Cooperativa de Crédito. **Inadimplência dos Tomadores de Crédito: Um Estudo em uma Cooperativa de Crédito. Pensar Contábil**, v. XVII, p. 55-63, 2015.

MANUAL DE PROCEDIMENTOS CONTÁBEIS PARA MICRO E PEQUENAS EMPRESAS: CFC / SEBRAE. Pedro Coelho Neto (coord.) Daniel Salgueiro da Silva, José Antonio de Godoy, José Xavier Cunha. 5. ed. 2002. Disponível em: <<http://www.portalcfc.org.br/wordpress/wp-content/uploads/2013/01/ManuMicro.pdf>>. Acesso em: 19 ago. 2018.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003.

_____. **Fundamentos de metodologia científica**. 8. ed. São Paulo: Atlas, 2017.

MARIO, P. C. Análise Discriminante. In: CORRAR, L. J. (Coord.); PAULO, E. (Coord.); DIAS FILHO, J. M. (Coord.). **Análise Multivariada**: para os Cursos de Administração, Ciências Contábeis e Economia. São Paulo: Atlas, 2017. cap. 4.

MARTENS, D. et al. Credit rating prediction using Ant Colony Optimization. **Journal of the Operational Research Society**, v. 61, n. 4, p. 561-573, 2010.

MENDES, A. S. R. **Transição de Basileia II para o Basileia III** : “Qual o enfoque que é dado aos Riscos nos Acordos de Basileia?”. 2013 Dissertação (Instituto Superior de Economia e Gestão) - Universidade Técnica de Lisboa. Disponível em: <[filhttps://www.iseg.ulisboa.pt/aquila/getFile.do?fileId=441875&method=getFile](https://www.iseg.ulisboa.pt/aquila/getFile.do?fileId=441875&method=getFile)>. Acesso em: 1 set. 2017.

MENEZES, L. L.; RICCIO, E. L. Relacionamento entre a controladoria e a gestão da informação econômico-financeira na nova economia. **Revista Brasileira de Gestão de Negócios**, v. 7, n. 18, p. 34-50, 2005.

MILERIS, R. Macroeconomic Determinants of Loan Portfolio Credit Risk in Banks. . **INZINERINE EKONOMIKA-ENGINEERING ECONOMICS**, v. 23, n. 5, p. 496-504, 2012.

MISSIO, F.; JACOBI, L. Variáveis dummy: especificações de modelos com parâmetros variáveis. **Ciência e Natura**, p. 111-135, 2007.

MSELMY, N.; LAHIANI, A.; HAMZA, T. Financial distress prediction: The case of French small and medium-sized firms. **International Review of Financial Analysis**, 2017.

OHLSON, J. A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. **Journal of Accounting Research**, v. 18, n. 1, p. 109-131, 1980.

OLIVEIRA, A. B. S. **Métodos da Pesquisa Contábil**. São Paulo: Atlas, 2011.

PIGNATA, F. A.; CARVALHO, D. O. Efeitos da crise econômica no Brasil. **Revista Eletrônica “Diálogos Acadêmicos”**, v. 09, n. 2, p. 04-18, 2015.

PIROLLA, F. R. **Redução de dimensionalidade usando agrupamento e discretização ponderada para a recuperação de imagens por conteúdo**. São Carlos, 2012 Dissertação (Programa de Pós-Graduação em Ciência da Computação) - Universidade Federal de Santa Catarina - UFSC. Disponível em: <<https://repositorio.ufscar.br/bitstream/handle/ufscar/511/4756.pdf?sequence=1>>. Acesso em: 20 mai. 2017.

PORTAL ACTION. **4.5 Predição**. Disponível em: <<http://www.portalaction.com.br/analise-de-regressao/45-predicao>>. Acesso em: 19 ago. 2018.

PRADO, J. W. et al. Multivariate analysis of credit risk and bankruptcy research data: a bibliometric study involving different knowledge fields (1968–2014). . **UFLA. Scientometrics**, v. 106, n. 3, p. 1007-1029, 2016.

PRATES, D. M.; FARHI, M. A. A crise financeira internacional, o grau de investimento e a taxa de câmbio do real. **IE/UNICAMP**, n. 164, 2009.

PRICE WATER HOUSE COOPERS (PWC). **Basileia III**: Principais características e potenciais impactos. 2013. Disponível em: <[http://www.abbc.org.br/images/content/PwC_Basileia III](http://www.abbc.org.br/images/content/PwC_Basileia%20III)>. Acesso em: 20 mai. 2017.

RIBEIRO, A.; FREIRE, E. J. ; BARELLA, L. A . A informação contábil como instrumento de apoio às micro e pequenas empresas: percepção dos gestores de micro e pequenas empresas de Paranaíta–MT, quanto à utilização de informações da contabilidade no processo de tomada de decisão. **REFAF, Revista Eletrônica**, v. 2, n. 1, 2013.

RODRIGUES, A.; PAULO, E. Introdução à Análise Multivariada. In: CORRAR, L. J. (Coord.); PAULO, E. (Coord.); DIAS FILHO, J. M. (Coord.). **Análise Multivariada**: para Curso de Administração, Ciências Contábeis e Economia. São Paulo: Atlas, 2017. cap. 1.

RODRIGUES, C.; BLATTMANN, U. Information management and the importance of use of sources for generation of knowledge: Gestão da informação e a importância do uso de fontes de informação para geração de conhecimento. **Perspectivas em Ciência da Informação**, v. 19, n. 3, p. 4-29, 2014.

ROGERS, D.; SILVA, W. M.; ROGERS, P. Credit Rating Change and Capital Structure in Latin America. **Brazilian Administration Review - BAR**, v. 13, n. 2, p. 1-22, 2016.

SALAZAR, G. T. **Administração dos Fluxos de Caixa**. São Paulo: Atlas, 2012.

SANTOS, A. S.; SANTOS, R. A. S.; SANTOS, S.. A importância das cooperativas de crédito no atual cenário econômico brasileiro, v. 1, p. 1-15, 2016. Disponível em: <<http://revistas.cesmac.edu.br/index.php/administracao/article/view/510>>. Acesso em: 19 jan. 2017.

SANTOS, V.; DOROW, D. R.; BEUREN, I. M. Práticas de micro e pequenas empresas. **Revista Ambiente Contábil**, v. 8, n. 1, p. 153-186, 2016.

SEBRAE - SERVIÇO BRASILEIRO DE APOIO ÀS MICRO E PEQUENAS EMPRESAS. **Anuário do trabalho na micro e pequena empresa: 2015**: Departamento Intersindical de Estatística e Estudos Socioeconômicos [responsável pela elaboração da pesquisa, dos textos, tabelas, gráficos e mapas]. 2015. Disponível em: <<http://www.sebrae.com.br/Sebrae/PortalSebrae/Anexos/Anuario-do-trabalho-na-micro-e-pequena-empresa-2014.pdf>>. Acesso em: 6 fev. 2018.

_____. Análise do CAGED. Julho/2018: Empreendedorismo que transforma. **Sebrae**. 2018. 3 p. Disponível em: <<http://www.sebrae.com.br/Sebrae/PortalSebrae/Anexos/Relatorio-do-CAGED-07-2018.pdf>>. Acesso em: 28 set. 2018.

_____. **MEI - ME – EPP**. Disponível em: <<http://www.sebrae-sc.com.br/leis/default.asp?vcdtexto=4154>>. Acesso em: 28 abr. 2018.

SELAU, L. P. R.; RIBEIRO, J. L. D. A systematic approach to construct credit risk forecast models. **Pesquisa Operacional**, v. 31, n. 1, p. 41-56, 2011.

SILVA, J. P. D. **Análise Financeira das Empresas**. São Paulo: Atlas, 2004.

_____. **Gestão e Análise de Risco de Crédito**. São Paulo: Atlas, 2008.

_____. **Gestão e Análise de Risco de Crédito**: Edição revisada e atualizada. 9. ed. São Paulo: Cengage Learning, 2016.

SILVA, R. A.; RIBEIRO, E. M. S.; MATIAS, A. B. Aprendizagem Estatística Aplicada à Previsão de Default de Crédito. **Revista de Finanças Aplicadas**, v. 7, n. 2, p. 1-19, 2016.

SMARANDA, C. Scoring Functions and Bankruptcy Prediction Models : Case Study for Romanian Companies. **Procedia Economics and Finance**, p. 217-226, 2014.

SOARES, M. M.; BALLIANA, G. M. O cooperativismo de crédito no Brasil. In: VENTURA, E. C. F. (Coord.); FONTES FILHO, J. R. (Coord.); SOARES, M. M. (Coord.). **GOVERNANÇA COOPERATIVA**: Diretrizes e mecanismos para fortalecimento da governança em cooperativas de crédito. Brasília: BCB, 2009. cap. 1.

STEINER, M. T. A. et al. Extração de regras de classificação a partir de redes neurais para auxílio à tomada de decisão na concessão de crédito bancário. **Pesquisa Operacional**, v. 27, p. 407-426, 2007.

STEINER, M. T. A. et al. Sistemas especialistas probabilísticos e redes neurais na análise de crédito bancário. **Revista Administração**, v. 34, n. 3, p. 56-67, 1999.

STELL, W. CHANCE: The Sixth C of Credit “The scope and speed of change argue for its inclusion among the Cs of credit”. **The RMA Journal**, 2009.

VAHID, P. R.; AHMADI, A. Modelling Corporate Customers Credit Risk Considering the Ensemble Approaches in Multiclass Classification: Evidence from Iranian Corporate Credits. **Journal of Credit Risk**, v. 12, n. 3, p. 71-95, 2016.

WANG, W. T.; ZHOU, X. Could traditional financial indicators predict the default of small and medium-sized enterprises? . **International Conference on Economics and Finance Research**, v. 4, p. 72-76, 2011.

ZHANG, D. et al. A hybrid credit scoring model based on genetic programming and support vector machines: In Proceedings . **4th International Conference on Natural Computation, ICNC**, v. 7, p. 8-12, 2008.

ZHU, Y. et al. Predicting China's SME credit risk in supply chain financing by logistic regression, artificial neural network and hybrid models. **Sustainability (Switzerland)**, v. 8, n. 5, p. 1-17, 2016.