

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO
Gestão Organizacional e Regionalidade

**PREVISÃO DE INSOLVÊNCIA CORPORATIVA NO BRASIL CONSIDERANDO A
REGIONALIDADE**

DENIZE LEMOS DUARTE

Orientador: Prof. Dr. Flávio L. M. Barboza

UBERLÂNDIA-MG
2019

DENIZE LEMOS DUARTE

**PREVISÃO DE INSOLVÊNCIA CORPORATIVA NO BRASIL CONSIDERANDO A
REGIONALIDADE**

Dissertação apresentada ao Programa de Pós-Graduação em Administração da Faculdade de Gestão e Negócios da Universidade Federal de Uberlândia, como requisito parcial para obtenção do título de Mestre em Administração.

Área de Concentração: Gestão Organizacional e Regionalidade

Linha de Pesquisa: Gestão Financeira e Controladoria

Orientador: Prof. Dr. Flávio L. M. Barboza

**Uberlândia - MG
2019**

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU com dados informados pelo(a) próprio(a) autor(a).

D812 Duarte, Denize Lemos, 1975
2020 Previsão de insolvência corporativa no Brasil considerando a regionalidade [recurso eletrônico] / Denize Lemos Duarte. - 2020.

Orientador: Flávio Luiz de Moraes Barboza.
Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Administração.
Modo de acesso: Internet.
Disponível em: <http://doi.org/10.14393/ufu.di.2020.95> Inclui bibliografia.

1. Administração. I. Barboza, Flávio Luiz de Moraes, 1980-, (Orient.). II. Universidade Federal de Uberlândia. Pós-graduação em Administração. III. Título.

CDU: 658

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Administração

Av. João Naves de Ávila, nº 2121, Bloco 5M, Sala 109 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4525 - www.fagen.ufu.br - pgaadm@fagen.ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Administração			
Defesa de:	Dissertação de Mestrado Acadêmico PPGAdm - Número 228			
Data:	17 de fevereiro de 2020	Hora de início:	09:00	Hora de encerramento:
Matrícula do Discente:	11812ADM022			
Nome do Discente:	Denize Lemos Duarte			
Título do Trabalho:	PREVISÃO DE INSOLVÊNCIA CORPORATIVA NO BRASIL CONSIDERANDO A REGIONALIDADE			
Área de concentração:	Regionalidade e Gestão			
Linha de pesquisa:	Gestão Organizacional e Regionalidade			
Projeto de Pesquisa de vinculação:				

Reuniu-se no Bloco 1F, sala 223, Campus Santa Mônica, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Administração, assim composta: Professores Doutores: Rodrigo Fernandes Malaquias (FAGEN/UFU), Poueri do Carmo Mário (UFMG) e Flávio Luiz de Moraes Barboza orientador(a) do(a) candidato(a). Ressalta-se que o(a) Prof(a) Dr(a). Poueri do Carmo Mário participou da defesa por meio de webconferência e os demais membros da banca e o(a) aluno(a) participaram in loco.

Iniciando os trabalhos a presidente da mesa, a Prof(a). Dr(a). Flávio Luiz de Moraes Barboza, apresentou a Comissão Examinadora e o(a) candidato(a), agradeceu a presença do público, e concedeu ao(a) Discente a palavra para a exposição do seu trabalho. A duração da apresentação do(a) Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(as) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado(a)

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por Flávio Luiz de Moraes Barboza, Professor(a) do



Magistério Superior, em 17/02/2020, às 10:31, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rodrigo Fernandes Malaquias, Professor(a) do Magistério Superior**, em 17/02/2020, às 10:31, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **POUERI DO CARMO MÁRIO, Usuário Externo**, em 17/02/2020, às 11:06, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site
https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1864617** e o código CRC **1092C940**.

Reitor da Universidade Federal de Uberlândia

Valder Steffen Júnior

Diretora da Faculdade de Gestão e Negócios

Kárem Cristina de Sousa Ribeiro

Coordenadora do Programa de Pós-Graduação

Cíntia Rodrigues de Oliveira Medeiros

*Com carinho e amor aos meus pais, meu marido e meu filho.
Aos meus pais, eterna gratidão; ao meu marido, todo carinho e cumplicidade e ao meu
filho, amor imensurável.*

AGRADECIMENTOS:

No momento de conclusão deste trabalho, muitas colaborações se fazem presentes em cada fase, seja no apoio, seja com críticas construtivas, em formas de carinho e de amizade, das lembranças e vivências que nos fortalecem e nos encorajam a seguir muito mais adiante.

Meu agradecimento mais profundo ao meu marido Wanderson Santos e ao nosso filho Gabriel Duarte por estarem a todo momento ao meu lado, incondicionalmente. Nos momentos mais difíceis, que não foram raros neste último ano, sempre me fazendo acreditar que chegaria ao final desta difícil, porém gratificante etapa de minha vida. Sou muito grata por cada gesto de carinho, pela paciência e apoio, sem vocês, eu não teria conseguido, vocês são a minha vida.

Aos meus pais Sr. Clarindo e Sra. Dirce que me amparam em todos os momentos, são meus exemplos de determinação e dignidade, obrigada por tudo que já fizeram e ainda fazem por mim.

Agradeço ainda o carinho e preocupação dos meus irmãos Renan, Sérgio e Sandro. Também meu sobrinho Daniel que está sempre presente para alegrar nossos momentos.

Sou eternamente grata ao meu orientador Professor Doutor Flávio Luiz de Moraes Barboza pela compreensão, pela paciência, serenidade e todo conhecimento transmitido, para mim será sempre um grande exemplo. A ele devo a construção de um aprendizado que vai muito além de uma dissertação, não tenho palavras para expressar minha imensa gratidão.

A Professora Dra. Karém Cristina de Sousa Ribeiro agradeço pelas valiosas contribuições e discussões, não poderia deixar de agradecer todo seu apoio desde o período de graduação, foram muito importantes para minha vida. Também não poderia deixar de agradecer aos Professores Dr. Valdir Machado Valadão Júnior, ao Dr. Rodrigo Fernandes Malaquias, a Dra. Fernanda Maciel Peixoto e a Dra. Cíntia R. de Oliveira Medeiros, pela dedicação incansável no exercício de nos transmitir o conhecimento. São minhas referências de compromisso com a docência e pela sabedoria no ato de ensinar. Estendo meus agradecimentos ao Prof. Dr. Poueri do Carmo Mário pelas excelentes contribuições. Aos colaboradores da secretaria da Coordenação da PPGA, ao Vinícius e em especial a Juliana pela competência e boa vontade em auxiliar sempre que necessário.

Aos colegas companheiros da turma do mestrado de 2018, pela amizade e companheirismo e pela convivência extraordinária.

Sobretudo a Deus, em quem confio fielmente, por ser tão generoso em minha vida, me proporcionando discernimento, determinação e me amparando nos momentos de preocupação.

“A lei de ouro do comportamento é a tolerância mútua, já que nunca pensaremos todos da mesma maneira, já que nunca veremos senão uma parte da verdade e sob ângulos diversos.”

Mahatma Gandhi

RESUMO

A previsão do fenômeno de dificuldade financeira no contexto de análise da concessão de créditos financeiros tem um papel relevante para o mercado, pois está associada a perdas e a elevados custos envolvidos no desdobramento do processo de insolvência e de recuperação dos créditos. Esta pesquisa foi desenvolvida para comparar e avaliar dois modelos de previsão de insolvência, um baseado em aprendizado de máquina, *Random Forest* (RF), e outro em estatística tradicional, Regressão Logística (LR), com dados de empresas brasileiras no período de 2005 a 2018. Verificou-se ainda o desempenho dos modelos em empresas regionais (mesorregião do Triângulo Mineiro e Alto Paranaíba e do Sul Goiano). No intuito de aprofundar o conhecimento, realizou-se uma revisão sistemática de literatura sobre dificuldades financeiras e falência, na qual detectou-se, que a tecnologia de inteligência artificial está em constante aprimoramento para prever empresas em dificuldade financeira. Num segundo momento, focou-se nos modelos, os quais são analisados em termos de desempenho. Os principais resultados conforme as métricas de acurácia, *Brier Score*, erros de previsão e área sob a Curva ROC (AUC), demonstraram que o classificador RF supera o modelo de LR, considerando a AUC, a precisão preditiva no cenário nacional foi de 96,7%, e 93,4% e nas empresas da região foi de 96,3% e 90,9% respectivamente, porém, em relação a outros estudos, o modelo LR apresentou um resultado satisfatório.

Palavras-chave: Falência, Regionalidade, Risco de Crédito, Regressão Logística, Florestas Aleatórias.

ABSTRACT

The prediction of financial distress in the context of the credit analysis plays a crucial role for the market because the link with losses and high costs involved in the unfolding of the insolvency and credit recovery process. This research develops a comparison and evaluates two insolvency prediction models, one based on machine learning, called Random Forest (RF), and another on traditional statistics, Logistic Regression (LR), by using data from Brazilian companies between 2005 to 2018. We also verify the performance of the models considering a regionality property (trough the mesoregion of Triângulo Mineiro and Alto Paranaíba and Sul Goiano). In order to deepen the knowledge, we carried out a systematic literature review on financial distress and bankruptcy where we detected that artificial intelligence technology is constantly improving to predict companies in financial distress. After that, we focused on the analyzes of the model performances. The main results according to the accuracy, brier score, forecast errors and area under the ROC curve (AUC) metrics showed that the RF classifier surpasses the LR model, considering the AUC, the predictive capacity in the national scenario was 96, 7%, and 93.4% and in the region's companies was 96.3% and 90.9% respectively. However, compared to other studies, the LR model presented a satisfactory result.

Keywords: *Bankruptcy, Regionality, Credit Risk, Logistic Regression, Random Forest.*

LISTA DE ILUSTRAÇÕES

Figura 1 - Evolução de empresas em situação de Falência Decretada, RJ Deferida e Concedida no Brasil	24
Figura 2 - Resumo da metodologia ProKnow-C, similar a De Azevedo, Ensslin e Jungles (2014) e Ensslin et al. (2017)	32
Figura 3 - Mapa de autores mais ativos e suas ligações por <i>aging</i> de publicação	36
Figura 4 - Evolução anual de estudos publicados.....	38
Figura 5 - Mapa com periódicos em destaque sobre o tema desta pesquisa.....	40
Figura 6 - Mapa de publicações por país de procedência.....	41
Figura 7 - Mapa de <i>Keywords</i> por citação.....	42
Figura 8 - Classificação de artigos por contexto da pesquisa.....	44
Figura 9 – Tipo de Pesquisa do Portfólio Bibliométrico.....	45
Figura 10 - Distribuição da base de dados por <i>Focus</i>	47
Figura 11 - Mapeamento de países por concentração de frequência de utilização do conjunto da base de dados	48
Figura 12 – Horizonte de tempo (<i>database</i>) utilizado no conjunto das bases de dados	49
Figura 13 - Distribuição das variáveis de interesse	50
Figura 14 - Distribuição das variáveis independentes	51
Figura 15 - Gráficos de dispersão dos modelos nacionais RF e LR, no cenário nacional e regional, respectivamente com delimitação da área cinzenta	78
Figura 16 - Curva ROC dos modelos testados RF e LR nacional e regional respectivamente ...	79
Figura 17 - Contribuição relativa das variáveis de entrada no desempenho dos modelos RF e LR, nacional e regional, respectivamente	82
Quadro 1 - Variáveis explicativas do modelo	72

LISTA DE TABELAS

Tabela 1 - Autores mais ativos evidenciados sobre o tema desta pesquisa por número de publicações	35
Tabela 2 - Autores mais evidenciados sobre o tema da pesquisa por número de publicações..	37
Tabela 3 – Periódicos em destaque sobre o tema desta pesquisa	39
Tabela 4 – Periódicos com mais publicações e citações do PB de 1992 a 2019	40
Tabela 5 - Categorização para análise do PB	43
Tabela 6 - Métodos e técnicas mais utilizados no PB	46
Tabela 7 - Evolução das pesquisas que utilizaram as técnicas RF e LR	61
Tabela 8 – Matriz de Confusão.....	72
Tabela 9 - Estatísticas descritivas do cenário nacional e regional	73
Tabela 10 - Resultados e comparação dos métodos RF e LR nacional e regional	75
Tabela 11 – Pontos de corte para definição da área cinzenta	77
Tabela 12 - Avaliação do poder discricionário do modelo RF em relação as variáveis de entrada por relevância	83
Tabela 13 - Avaliação do poder discricionário do modelo LR em relação as variáveis de entrada por relevância	84

LISTA DE ABREVIATURAS E SIGLAS

Abreviatura	Significado
AB	AdaBoost
AC	AdaCost
AI	Artificial Intelligence
ANN	Artificial Neural Networks
BPNN	Back Propagation Neural Network
BA	Bagging
BRT	Bagging Random Tree
PB	Portfólio Bibliométrico
BC	Binary Classification
BOO	Boosting
CBFS	Case Based Forecasting System
CBR	Case Based Reasoning
CLN	Categorical Learning Network
CHAID	Chi-Square Automatic Interaction Detector
CW	Class Switching
CART	Classification and Regression Trees
CT	Combination of techniques
CL	Competitive Learning
CA	Consensus Approach
CVNN	Cross Validation Neural Network
CB	Cost Boost
CSL	Cost Sensitive Learning Method
COX	Cox's Model
CDT	Credal Decision Tree
DM	Data Mining
DSS	Decision Support System
DT	Decision Trees
DECORATE	Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples
EWS	Early Warming Systems

ELECTRE	Elimination And Choice Net Translating Reality
EM	Elysee Method
EABC	Evolutionary Artificial Bee Colony
ES	Expert Systems
ELM	Extreme Learning Machine
FNN	Feedforward Neural Network
FDP	Financial Distress
FLD	Fisher's Linear Discriminant
FFAO	Fruit Fly Algorithm Optimized
FS	Fuzzi Score
FSVM	Fuzzi Support Vector Machine
FIFLN	Fuzzy Integral Base Functional Link
FL	Fuzzy Logic
FRBC	Fuzzy Rule Based Classifier
GNG	Gabriel Neighbour hood Graph
GP	Gaussian Process
GRNN	General Regression Neural Network
GSOFNN	Generic Self Organizing Fuzzy Neural Network
GA	Genetic Algorithm
GPR	Genetic Programming
GBOO	Gradient Boosting
GRA	Grey Relational Analysis
HIS	Hibrid Intelligence Systems
HFP	Hybrid Failure Prediction
ISTRIS	Incorporating Sentiment And Textual Information Into a Randon Subspace
IFWTSVM	Incremental Filter Wrapper Twin Support Vector Machine
IMSW	Interative Model With Subject Weight
KM	K-Means
KNN	K-Nearest Neighbor
LVQ	Learning Vector Quantization
LNMC	Levenberg Marquardt Neural Nets
LDA	Linear Discriminant Analysis
LR	Logistic Regression

LGD	Loss Given Default
ML	Machine Learning
MV	Majority Voting
MCMAC	Modified Cerebellar Model Articulation Controller
MCOC	Multi Criteria Optimization Classifier
MTL	Multi Task Learning
MLP	MultiLayer Perceptron
MCS	Multiple Classifier Systems
MCDM	Multiple Criteria Decision Making
MCLP	Multiple Criteria Linear Programming
MDA	Multiple Discriminant Analysis
MARS	Multivariate Adaptive Regression Splines
NB	Naive Bayes
NMC	Nearest Mean Classifier
O	Others
OS	Over Sampling
PLS	Partial Least Squares
PLSDA	Partial Least Squares Discriminant Analysis
PSO	Particle Swarm Optimization
PNN	Probabilistic Neural Network
PM	Probit Method
PBM	Profile Based Model
QDA	Quadratic Discriminant
RBF	Radial Basis Function
RBFN	Radical Basis Function Network
RF	Random Forest
RS	Random Subspace
RB	Realibility Based
RPA	Recursive Partitioning Algorithm
RTF	Rotation Forest
RST	Rough Set Theory
SV	Selection of variables
SOM	Self Organizing Map

SMO	Sequential Minimal Optimization
SAT	Single Attribute Test
SLFN	Single Hidden Layer Feed Forward Neural Network
SLP	Single Layer Perceptron
DAS	Stepwise Discriminant Analysis
SVM	Support Vector Machine
TV	Threshold Varying
TBOO	Time Boost
TSCM	Two Stage Classification Method
US	Under Sampling
VPRS	Variable Precision Rough Set Model
VB	Voting Based
WVRN	Weighted Vote Relational Neighbour
ZS	Z Score

SUMÁRIO

CAPÍTULO I: INTRODUÇÃO.....	18
1.1 CONTEXTUALIZAÇÃO DO TEMA DE PESQUISA.....	18
1.2 OBJETIVOS E PROBLEMA DE PESQUISA	20
1.3 RELEVÂNCIA DO ESTUDO	22
CAPÍTULO II: PREVISÃO DE DIFICULDADES FINANCEIRAS, FALÊNCIA, RISCO DE CRÉDITO E APRENDIZADO DE MÁQUINA - UMA REVISÃO SISTEMÁTICA DA LITERATURA	26
2.1 LITERATURA E ABORDAGEM GERAL	28
2.2 APRENDIZADO DE MÁQUINA EM RISCO DE CRÉDITO.....	29
2.3 METODOLOGIA.....	30
2.4 A BASE DE DADOS	31
2.5 RESULTADO DAS PESQUISAS	32
2.6 ANÁLISE DESCRIPTIVA QUANTITATIVA	34
2.7 ARTIGOS COM RELEVÂNCIA CIENTÍFICA	37
2.8 ANÁLISE QUANTITATIVA DOS ARTIGOS DO PB	43
2.9 CONSIDERAÇÕES FINAIS	51
CAPÍTULO III: PREVISÃO DE INSOLVÊNCIA NO MERCADO CORPORATIVO BRASILEIRO CONSIDERANDO A REGIONALIDADE	55
3.1 REFERENCIAL TEÓRICO	59
3.2 AS TÉCNICAS DE REFERÊNCIA PARA O ESTUDO	65
3.3 A BASE DE DADOS	69
3.4 CRITÉRIOS DE AVALIAÇÃO	71
3.5 VARIÁVEIS ABORDADAS NO ESTUDO	67
3.6 RESULTADOS ENCONTRADOS	73
3.7 CONSIDERAÇÕES FINAIS	85
CAPÍTULO IV – CONCLUSÕES.....	88
REFERÊNCIAS	90
APÊNDICES	110

CAPÍTULO I: INTRODUÇÃO

1.1 Contextualização do tema de pesquisa

No início das atividades capitalistas, as transações comerciais eram realizadas somente com pagamentos imediatos, o que tornava uma limitação ao poder aquisitivo. A concessão de crédito foi uma alternativa para viabilizar essas transações, que antes eram restritas, de forma a expandir o processo econômico, além de torná-lo mais eficiente entre as partes. Neste contexto, um dos propósitos do mercado de capitais é a transferência de fundos entre credores (poupadores) e devedores (tomadores), que interagem entre si de forma eficiente por meio de transações financeiras (COPELAND; WESTON, 1988; GRILLI; TEDESCHI; GALLEGATI, 2015). Scherr (1989) estabelece então que crédito é o poder ou a capacidade de obtenção de bens ou serviços em troca do compromisso futuro de pagamento.

Brealey e Myers (1996) relatam que nas empresas um terço do ativo circulante é composto pela conta clientes, o que confirma a demanda e a relevância por constante aperfeiçoamento contínuo nas modelagens e políticas as previsões de perdas financeiras. As empresas concedem crédito com o objetivo de: aumentar vendas; aumentar o valor das operações a termo; aumentar, reter ou conquistar clientes. Dessa maneira, conceder crédito é uma forma de investir no cliente. Corroborando com esta afirmativa, Copeland e Weston (1988) relatam que os provedores de capital após esgotarem as suas oportunidades produtivas de maior retorno, estão dispostos a ceder seus excessos de recursos, pois a taxa de empréstimo é maior do que aquela que se pode captar, mas a liquidação do montante cedido no momento acordado, é fundamental para garantir a rentabilidade desejada.

Desta forma, múltiplas variáveis interferem nos padrões de comportamento dos pagamentos, inclusive as condições macroeconômicas em que a empresa está inserida, e que, de forma acelerada podem impactar nas previsões de recebimentos, comprometendo a potencialidade de caixa do negócio, assim, definem-se as perdas de crédito provenientes da dificuldade financeira (LOPEZ; SAIDENBERG, 2000). Steiner et al. (1999) apontam que qualquer erro na decisão para concessão de créditos pode representar muitas perdas em relação aos ganhos obtidos de concessões bem-sucedidas, pois além do valor do crédito concedido, existem os custos de oportunidade inclusos bem como os elevados custos para recuperação. Shleifer e Vishny (1997) discorrem que a proteção legal de credores, em determinados países, é mais eficaz do que a proteção de acionistas uma vez que a inadimplência é uma violação direta de um contrato de dívida formal, o que contribui para o recebimento das dívidas.

Em particular, Barboza, Kimura e Altman (2017) relatam que a estimativa da inadimplência se trata de um fenômeno decorrente de ação humana, e por isso, é algo que possui uma certa medida de incerteza. Se for levado em consideração que existe a possibilidade de se errar essa medida, de modo que, seu resultado leve a algum dano ou perda para outro agente (contraparte), tem-se então uma situação de migração do risco, ou melhor, uma situação de inadimplência e, consequentemente de perdas financeiras (LOPEZ; SAIDENBERG, 2000).

A dificuldade financeira é conceituada como um evento oneroso, pois cria uma tendência para que as empresas necessitem tomar decisões que sejam prejudiciais aos acionistas e as partes interessadas (OPLER; TITMAN, 1994). As dificuldades financeiras das empresas aumentam o risco, quando diversas empresas são incapazes de retornar o capital para a economia devido a sua fragilidade financeira, isso faz com que as dívidas se tornem incobráveis, trazendo enormes problemas para todo sistema financeiro (SUN et al., 2017). Esta migração de risco em grande escala, de forma desencadeada, interrompe e reduz o fluxo de capital, desacelerando a economia e ocasionando as crises financeiras.

O cenário frequente dessas crises e bolhas financeiras mundiais impactam diretamente o crescimento econômico (SUN et al., 2017). Nas últimas décadas, vem ocorrendo uma transferência significativa de recursos dos setores produtivos para o setor financeiro. Rochon e Rossi (2010) alertam sobre a desregulamentação, o crescimento dos instrumentos financeiros complicados, as atividades predatórias dos empréstimos bancários, as práticas fraudulentas das instituições, que trouxeram mudanças ao funcionamento dos sistemas econômicos. Assim, esse processo de realocação do capital na economia é um dos elementos responsáveis pela crescente instabilidade financeira (GRILLI; TEDESCHI; GALLEGATI, 2015; SUN et al., 2017).

Diante da fragilidade econômica, as organizações entram em dificuldade financeira comprometendo a estimação do capital econômico e aumentando o risco nos mercados (SUN et al., 2017). De acordo com o Banco Central do Brasil (BACEN), o saldo das operações de crédito, com recursos livres e direcionados, de Pessoas Jurídicas, no Sistema Financeiro Nacional (SFN), totalizaram R\$ 1,4 trilhões, sendo que, deste montante, 2,6% se torna inadimplência (com atraso acima de 90 dias, ainda não baixados para prejuízo contábil), ou seja, cerca de R\$ 37 bi (BACEN, 2019). Tratam de operações de crédito regulamentadas pelo Conselho Monetário Nacional (CMN) ou vinculadas a recursos orçamentários e referem-se a empréstimos com recursos livres e financiamentos com recursos direcionados.

Diversas pesquisas procuram desenvolver modelos que busquem melhorar a previsão e a precisão para o controle da insolvência, evitando perdas financeiras e custos envolvidos (HUANG et al., 2004; SUN; HUANG; HE, 2014; BARBOZA; KIMURA; ALTMAN, 2017; ALAKA et al., 2018). Horta, Alves e Carvalho (2013) expõem que os modelos de *financial distress prediction* (FDP), baseados em *data mining*, são ferramentas avançadas, capazes de discriminar informações relevantes, de forma imparcial, possibilitando uma obtenção mais confiável no processo de tomada de decisão na avaliação de concessão de créditos financeiros. Juntamente com o desenvolvimento da tecnologia de banco de dados e das técnicas de mineração, existe a busca por métodos inovadores baseados em inteligência artificial.

Bellovary, Giacomo e Akers (2007) concluem que apesar dos diferentes modelos de previsão, os testes empíricos da maioria dos modelos demonstram uma alta capacidade de precisão preditiva. Complementando, Pinheiro et al. (2009) discorrem sobre a relevância da atualização dinâmica desses modelos em decorrência das interferências nos coeficientes associados às variáveis endógenas e exógenas temporais que compõem os modelos.

O modelo proposto nesta dissertação considera variáveis financeiras das empresas do mercado de capitais brasileiro para comparação dos principais modelos de previsão *Random Forest* e *Logit*, em conformidade com os estudos de Barboza, Kimura e Altman (2017). Para a variável dependente usou-se a dificuldade financeira, a falência e a recuperação judicial deferida, adaptado do estudo de Manzaneque, Priego e Merino (2016), e para as variáveis independentes foram usadas aquelas extraídas da literatura, as quais se mostravam mais relevantes e tinham dados disponíveis para o devido emprego no estudo proposto. Em contrapartida, a discussão sobre regionalidade, nesse tipo de assunto, os estudos são inexistentes. A relevância de se tratar uma determinada temática, na sua centralidade, para o espaço regional, contribui para a municipalização e expansão local, trazendo resultados positivos de responsabilidade e recursos, abrindo espaço para a geração de mecanismos de crescimento, sócio-cultural, de riqueza e desenvolvimento.

1.2 Objetivos e problema de pesquisa

Considerando a relevância na gestão da incerteza e de riscos, na maximização da riqueza dos acionistas, na circulação corrente de recursos na economia e consequentemente no crescimento e maior estabilidade econômica de um país, o problema que estimulou a realização deste estudo foi: o modelo RF possui maior capacidade preditiva que o modelo *Logit* no

mercado de capitais brasileiro considerando os cenários nacional e regional para o período de 2005 a 2018?

O objetivo deste trabalho é avaliar e comparar dois modelos utilizando mecanismos de aprendizado de máquina e Regressão Logística para prever previamente o fenômeno da insolvência de empresas brasileiras e avaliar o comportamento dos modelos em empresas da região (mesorregião do Triângulo Mineiro e Alto Paranaíba e do Sul Goiano). Assim, o presente estudo optou por trabalhar com a revisão teórica sistemática, complementada pelo conhecimento e embasamento epistemológico de diagnóstico coletivo e das diferentes visões no intuito de contribuir para uma maior e melhor análise das variantes que compõe o risco de crédito, a fim de identificar a previsibilidade de insolvência a partir do *behaviour* das organizações de capital aberto listadas na bolsa brasileira, Brasil, Bolsa e Balcão (B3), no período de 2005 a 2018. Este período foi definido em decorrência da Lei de Recuperação Judicial e Extrajudicial e de Falência nº 11.101 de 2005.

Enfim, o objetivo geral deste trabalho consistiu em avaliar, comparar e desenvolver dois modelos, utilizando mecanismos de aprendizado de máquina, RF, e estatística tradicional, *Logit*, para prever previamente o fenômeno de insolvência em empresas de capital aberto no Brasil, avaliando o desempenho destes dois modelos em nível nacional e, depois, somente no âmbito da regionalidade, isto é, para empresas da região de interesse (mesorregião do Triângulo Mineiro e Alto Paranaíba e do Sul Goiano).

De forma específica, os objetivos deste trabalho são:

- a) avaliar as diversas pesquisas acadêmicas com as múltiplas visões sobre risco de crédito com enfoque em dificuldades financeiras, classificação de crédito e inteligência artificial, sua evolução ao longo dos anos, realizando uma revisão sistemática para identificar os diferentes construtos desse campo da literatura;
- b) identificar e descrever a utilização de técnicas de análise, que envolvem AI, em pesquisas sobre risco de crédito, dificuldades financeiras e falência;
- c) apresentar as tendências de publicações, conjunto de dados, periódicos, autores e co-autorias;
- d) analisar se a adoção de um modelo de aprendizado de máquina agrupa e melhora a previsão de dificuldades financeiras e falência em empresas do mercado de capitais brasileiro;

- e) examinar se o modelo *Random Forest*, baseado nas variáveis selecionadas para este estudo, oferece maior precisão e melhora a capacidade preditiva de dificuldades financeiras em relação ao método estatístico tradicional de Regressão Logística;
- f) conhecer os fatores e variáveis que interferem no risco diante do fenômeno de dificuldade financeira e na previsão de insolvência para o mercado brasileiro;
- g) verificar e comparar o comportamento dos modelos de previsão nas empresas da região de interesse (mesorregião do Triângulo Mineiro e Alto Paranaíba e do Sul Goiano).

A inadimplência pode ser tratada como um fenômeno decorrente do processo decisório, partido pela decisão de um gestor ou por conselhos de executivos, assim faz com que sua estimativa se torne uma medida de incerteza (BARBOZA; KIMURA; ALTMAN, 2017). Diante do exposto, de acordo com a literatura apresentada e baseando-se nos estudos empíricos sobre o assunto, chegou-se nas seguintes hipóteses:

H₁. O método baseado em aprendizado de máquina é mais preciso que o método baseado em estatística tradicional regressão logística.

H₂. Os modelos de previsão apresentam o desempenho compatível em relação ao poder preditivo e a precisão quando aplicado em empresas regionais.

1.3 Relevância do estudo

Os processos constantes de homogeneização nas práticas mercadológicas mundiais decorrentes da globalização, exigem organizações cada vez mais dispostas a responder às essas demandas para se manterem em níveis competitivos, prevendo as mudanças no intuito de maximizar a riqueza dos acionistas e perpetuar o negócio. Para tanto, diversas são as estratégias adotadas, uma delas, trata-se da gestão do risco inerente à concessão de créditos, que dispõe de variáveis, pressupostos e modelagens de previsão, admitidos na estimação do capital econômico suficiente para comportar os fluxos das atividades demandadas pelas empresas e instituições, bem como a regulação dos mercados e limites toleráveis de perdas.

Esta dissertação contribui ao verificar a teoria, fornecendo informações em relação aos métodos e técnicas mais utilizados e também realiza uma análise ampla, para aprofundar o conhecimento dos fatores e variáveis integrados ao risco de crédito diante do fenômeno de dificuldade financeira e na previsão de insolvência, determinando assim, quais são as tendências

das pesquisas neste campo de estudos. Apresenta conjuntamente, evidências para fornecer orientação teórica e suporte à decisão sobre estes aspectos em um mercado emergente e na regionalidade. O modelo também traz testes para auxiliar na orientação diante das afirmações da literatura em relação as variáveis selecionadas demonstrando a influência na precisão preditiva do fenômeno.

Além disso, o estudo traz contribuições no que se refere ao cunho social, tendo em vista que a melhoria no processo de concessão de créditos e mitigação de riscos, pode contribuir para o aumento na mobilidade do capital nos diversos setores, direta ou indiretamente, e consequentemente o crescimento econômico, gerando empregos, possibilitando maior poder de aquisição, maior acesso ao capital e melhora no padrão social do mercado brasileiro (ADDO; GUEGAN; HASSANI, 2018).

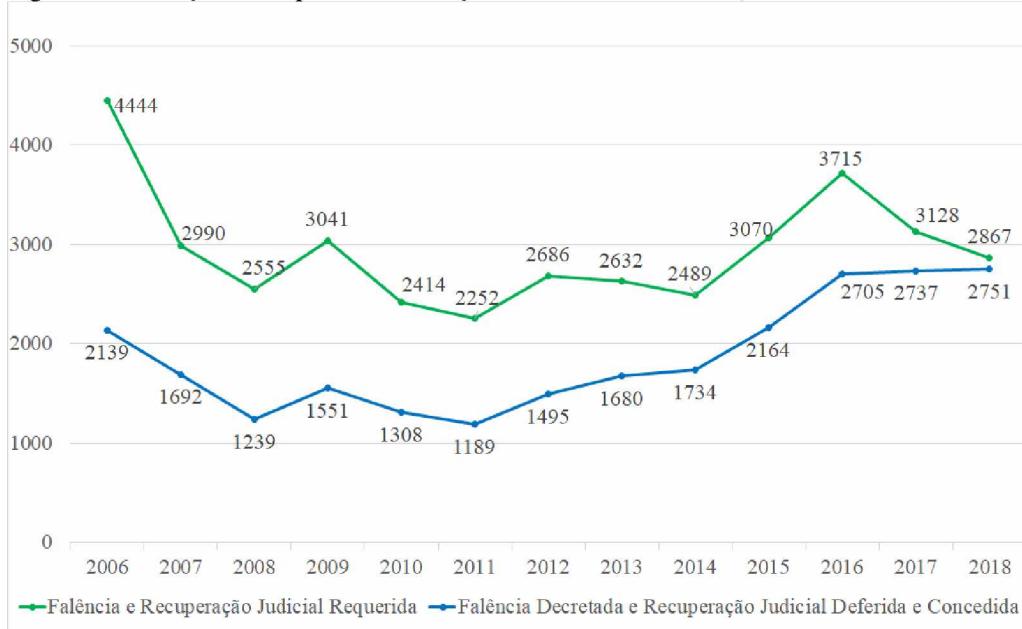
Em relação a regionalidade, as cidades das mesorregiões do Triângulo Mineiro, Alto Paranaíba e do Sul Goiano foram escolhidas pois, em termos econômicos, destacam-se na sua participação no PIB nacional, que, segundo o IBGE (2017), é de 2,34%, ficando entre as dez maiores no país, diante das 137 mesorregiões. A economia se ressalta devido a oferta de produtos agropecuários, na indústria alimentícia, no setor sucroalcooleiro, no setor terciário, com evidência para os segmentos de serviços de telecomunicações, comércio atacadista e turismo. E ainda é o principal polo produtor de fertilizantes fosfatados do Brasil (IBGE, 2017).

O Triângulo Mineiro e Alto Paranaíba, traz uma malha rodoviária que se evidencia devido a expansão das atividades agroindustriais, como a pecuária extensiva, sendo a que possui maior rebanho bovino do estado de Minas Gerais, e com a alta produção de grãos, sendo a segunda mais importante na cultura do café. As mesorregiões do Triângulo Mineiro e Alto Paranaíba e do Sul Goiano, são privilegiadas por sua localização, possuindo uma das mais relevantes e estratégicas malhas viárias do país, interligando diversas regiões (BERNARDES; FERREIRA, 2013; IBGE, 2017).

Em termos de utilidade para o mercado, os resultados do estudo são interessantes para organizações, Instituições Financeiras e partes interessadas, que podem investigar a comparação dos modelos apresentados, sua capacidade de precisão preditiva em relação a outros modelos já estudados, de forma a orientar as análises práticas na concessão de créditos e, consequentemente, tratar a mitigação dos riscos, possibilitando o alcance de maiores lucros, redução de perdas, que trata do valor principal concedido, dos custos de erro na classificação envolvidos no processo de insolvência e na recuperação de créditos e também os custos de

oportunidade do principal e dos juros não recebidos (AGARWAL; TAFFLER, 2008; ABDOU; POINTON, 2011).

Figura 1 – Evolução de empresas em situação de Falência Decretada, RJ Deferida e Concedida no Brasil



Fonte: Serasa Experian (2019)

No Brasil, conforme Figura 1, nota-se uma crescente evolução no volume de empresas em situação de Falência Decretada e Recuperação Judicial (RJ) Deferida e Concedida. Comparando a quantidade de empresas de capital aberto e fechado que entraram nesta situação em 2008 e 2018 houve um aumento de 122% no período de dez anos, partindo de 1.239 para 2.751 empresas.

Nesta dissertação, a amostra é formada por empresas listadas na bolsa, uma vez que todas as informações empregadas sobre suas particularidades são divulgadas ao mercado. Desenvolveu-se uma análise utilizando dados de empresas brasileiras, por se tratar de uma referência econômica com participação relevante na América Latina e no mundo. Como demonstrado na Figura 1, neste mercado, o volume de empresas em situação de dificuldade financeira vem crescendo ao longo dos anos, assim, torna-se relevante analisar os modelos de predição de perdas financeiras diante das características das respectivas empresas para que os interessados possam direcionar suas decisões e consequentemente melhorar sua lucratividade.

Os integrantes do mercado de capitais são as grandes instituições e organizações, que possuem operações significativas em níveis globais, quando as organizações, participantes do mercado, não conseguem liquidar suas obrigações por problemas de fluxo de caixa, a exposição

ao risco de crédito avança, desencadeando graves efeitos para todo o sistema financeiro, trazendo altos custos individuais e coletivos, sociais e econômicos (ZHANG; WANG; JI, 2013; SUN; FUJITA; CHEN; LI, 2017; LI; PEREZ-SAIZ, 2018).

Considerando as discussões teóricas e o aspecto histórico, as pesquisas sobre risco de crédito possuem não apenas relevância prática, para o mercado e o sistema financeiro como um todo, uma vez que está conectado à operações nas quais abrange toda sociedade, mas também tem sua importância teórica devido a sua relação, intrínseca ou não, com outros estudos que estão vinculados à teoria de finanças.

O presente estudo está estruturado em quatro capítulos. Sendo este, o Capítulo 1, que trata de uma contextualização que introduz de forma sintetizada os conteúdos que serão abordados nos capítulos subsequentes. O Capítulo 2 dispõe de uma revisão teórica sistemática da literatura sobre análise de risco de crédito, dificuldade financeira, falência e modelos e técnicas, baseados em inteligência artificial, em estudos publicados no Brasil e no mundo, identificando lacunas e recomendações de pesquisas futuras, enquanto o Capítulo 3 avalia e desenvolve um modelo computacional, utilizando mecanismos de aprendizado de máquina e outro modelo por meio da estatística tradicional para prever o fenômeno de insolvência em empresas de capital aberto no Brasil e na região de interesse. Por fim, o Capítulo 4 encerra este trabalho apresentando as conclusões que abrangem os achados dos Capítulos 2 e 3, expondo as limitações, os pressupostos deste estudo e as sugestões para futuras pesquisas.

CAPÍTULO II: PREVISÃO DE DIFICULDADES FINANCEIRAS, FALÊNCIA, RISCO DE CRÉDITO E APRENDIZADO DE MÁQUINA - UMA REVISÃO SISTEMÁTICA DA LITERATURA

Com o rápido desenvolvimento nos sistemas de informação e a acelerada disseminação da globalização econômica, a forma de se conduzir uma avaliação eficaz do risco de crédito é uma questão relevante no campo de finanças. Um elevado índice de insucessos empresariais pode ser comprometedor para investidores, acionistas, fornecedores, sociedade e de forma geral, para a economia de um país (GENG; BOSE; CHEN, 2015; ZELENKOV; FEDOROVA; CHEKRIZOV, 2017; ALAKA et al., 2018). Uma melhora na precisão da previsibilidade de dificuldade financeira, mesmo que pequena, pode proporcionar maiores retornos, minimizar os efeitos negativos na disfunção do sistema, nos segmentos de mercado e até mesmo no crescimento econômico (WANG et al., 2012).

Diante do enfraquecimento do crescimento econômico, seguido pelo efeito oneroso das dificuldades financeiras das organizações, nas últimas décadas, os mercados se mostram receosos com a concessão de crédito e com a própria fragilidade do sistema financeiro como um todo (ANTUNES; RIBEIRO; PEREIRA, 2017). Após a crise financeira global de 2007/2008, o mercado mundial procurou melhorar o risco de mercado e as determinações relacionadas ao apreçamento de instrumentos financeiros. Essas diretrizes, pretendem aperfeiçoar a capacidade das organizações e instituições da absorção de choques e enfrentamento para as eventuais crises financeiras (BIS, 2019; BACEN, 2019). Com isso surgiu o Basileia III, que abrange as medidas que fortalecem a regulação, supervisão e a gestão de risco.

A criação do Acordo de Basileia, deu ênfase ao setor bancário, mas teve o intuito de alertar os mercados em todo mundo de que existe uma preocupação com o futuro do sistema financeiro global e suas perspectivas em relação a sua contribuição para o desenvolvimento social e econômico (BARBOZA; KIMURA; ALTMAN, 2017). De acordo com a Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais - ANBIMA (2019), o Acordo de Basileia III iniciou no Brasil em outubro de 2013, sendo o começo da fase de transição para os novos padrões prudenciais e deve ser concluída integralmente em 2022.

Neste contexto, apresentado pela vulnerabilidade do sistema financeiro, existe a necessidade de se quantificar e interpretar as variáveis e informações referentes ao fenômeno de dificuldade financeira que envolve o risco de crédito e suas operações. Existem diversos

estudos de métodos estatísticos, baseados em regressões, análises discriminantes multivariadas e baseados em *Artificial Intelligence* (IA) disponíveis, mas não existe uma concordância daquele que traz a melhor estratégia, maior assertividade e precisão diante do fenômeno da insolvência (ADNAN AZIZ; DAR, 2006; ABDOU; POINTON, 2011).

Ohlson (1980) utilizou a regressão logística para a estimação do padrão de probabilidade para dificuldades financeiras. Beaver (1966), Altman (1968), Taffler (1982) e Altman, Marco e Varetto (1994) usaram a *Multiple Discriminate Analysis* (MDA) através de indicadores contábeis e financeiros. Altman, Baidya e Dias (1979) utilizaram a análise discriminante para classificar e prever problemas financeiros em empresas brasileiras. Destacaram-se também outros estudos de análise discriminante e regressão logística como: Martin (1977); Griffin e Lemmon (2002); Hillegeist, Keating, Cram e Lundstedt (2004); De Andres, Landajo e Lorca (2005). Begley, Ming e Watts (1996) criticaram os modelos apresentados por Altman (1968) e por Ohlson (1980), sendo que este último apresentou, nesta análise, melhor desempenho.

Alguns estudos utilizaram algoritmos e inteligência artificial para classificar a idoneidade creditícia e a previsibilidade de dificuldades financeiras ou falência: Tam e Kiang (1992); Altman, Marco e Varetto (1994); Shin, Lee e Kim (2005); Pan (2012); Olson, Delen e Meng (2012); Sun, Li, Huang e He (2014); Barboza, Kimura e Altman (2017); Sun, Lang, Fujita e Li (2018); e, García, Marqués e Sánchez (2019). Addo, Guegan e Hassani (2018) discorrem que os algoritmos baseados em redes neurais artificiais não fornecem necessariamente o melhor desempenho e que é importante analisar e verificar a qualidade dos dados para evitar vieses na classificação, mas esta demanda é relevante em qualquer modelo, seja ela baseado em redes neurais ou outros baseados em estatística tradicional.

Desse modo, este estudo adotou uma perspectiva para avaliar as diversas pesquisas acadêmicas com as múltiplas visões sobre risco de crédito com enfoque em dificuldade financeira, classificação de crédito, falência, aprendizado de máquina e inteligência artificial (IA) e, sua evolução ao longo dos anos. Esse objetivo foi alcançado realizando uma revisão sistemática para identificar os diferentes construtos desse campo da literatura, explorando e descrevendo a utilização dos métodos e técnicas de análise que envolvem AI, e ainda apresentando as tendências das publicações, dos conjuntos de dados, periódicos, autores e co-autorias deste portfólio de pesquisas.

Para desenvolvimento da primeira parte deste trabalho empregou-se a metodologia *Knowledge Development Process-Constructivist* (ProKnow-C), o que resultou, como suporte

empírico uma seleção de 165 publicações científicas nacionais e internacionais mais relevantes em relação as suas citações e diante das bases de dados selecionadas para esta pesquisa. Diante desta averiguação, pode-se afirmar que a aplicação da tecnologia computacional no âmbito de análise da gestão do risco de crédito, para instituições e organizações, vem angariando atenção de forma singular. Constatou-se que a demanda para se identificar e introduzir novas variáveis, classificadores e métodos mais assertivos é constante.

O esforço para se aprimorar a interpretação dos dados e modelos, é intenso e procuram tratar as lacunas neste campo de estudos. Desta forma, este estudo contribui para a categorização das linhas de pesquisas e traz uma síntese da literatura, em que os resultados servem como referência, além de sugerir pesquisas futuras sobre o tema. Esta revisão de literatura constitui um arcabouço de 165 publicações com contribuições que auxiliam na análise de lacunas referente ao risco de crédito diante do fenômeno de dificuldade financeira e na previsão de insolvência. Os achados desta pesquisa podem contribuir ainda para os interessados nesta temática, que podem examinar sobre as tendências neste campo de estudos no Brasil e no mundo.

2.1 Literatura e Abordagem Geral

Beaver (1966) define a dificuldades financeiras das firmas como sendo a incapacidade de uma determinada empresa em pagar suas obrigações financeiras com seus credores, sendo eles fornecedores, instituições financeiras e acionistas. O autor ressalta ainda que a maior preocupação não é em tratar diretamente o problema da inadimplência, mas sim os preditores financeiros dos eventos que são considerados para a probabilidade do fenômeno para se prever as dificuldades financeiras.

O risco de crédito tem sido foco de estudos há décadas e se trata de um importante tema na área de gestão financeira (ALAKA et al., 2018). Beaver (1966) estudou os índices financeiros como preditores de dificuldades financeiras usando uma perspectiva *Bayesian*, que trata a avaliação de hipóteses probabilísticas por razões de verossimilhança, onde os eventos possíveis são vistos como sendo dicotômicos, em que a empresa sofre dificuldades financeiras ou não. O autor concluiu que a relação entre a razão de verossimilhança e o fluxo de caixa para a dívida total foi o melhor preditor.

Altman (1968) usou Análise Discriminante Multivariada para classificar as empresas falidas e não falidas para prever o fenômeno de falência e concluiu que a falência pode ser

prevista com precisão até dois anos antes da falha real através da combinação de índices contábeis e financeiros. Ohlson (1980) aplicou modelos de Regressão Logística para prever dificuldades financeiras. Zmijewski (1984) usou um modelo *probit* para previsão de falência.

Diversas são as metodologias utilizadas para auxiliar e melhorar a precisão na previsão do risco de crédito. Desde métodos estatísticos tradicionais aqueles modelos baseados em AI, que podem ser univariados ou multivariados, que se concentram em indícios que levam a falha da organização e os dados são extraídos principalmente de relatórios financeiros e contábeis. Os modelos que são teóricos se concentram nas causas qualitativas da falha, e extraem informações que podem argumentar a falha utilizando a teoria multivariada e empregando técnicas estatísticas para fornecer o suporte quantitativo para o argumento teórico (ADNAN; AZIZ; DAR, 2006).

2.2 Aprendizado de Máquina em Risco de Crédito

O aprendizado de máquina envolve o estudo da teoria do reconhecimento de padrões, a construção de algoritmos e aprendizagem computacional em AI, que pode treinar, aprender e fazer previsões sobre determinados dados (ZHU et al., 2016). Odom e Sharda (1990), Tam e Kiang (1990) foram um dos precursores a empregar o método de redes neurais para prever falências e tratar o risco de crédito. Neste artigo, os autores trouxeram uma abordagem de rede neural para previsão de falência em instituições financeiras e empresas e fizeram uma comparação com outros modelos mais utilizados, como análise discriminante. Os resultados apresentados pelos autores foram que as redes neurais se mostraram mais robustas e tiveram melhor desempenho do que os outros modelos.

Neste período, as redes neurais já se mostraram como um instrumento competitivo entre os métodos estatísticos tradicionais para se avaliar a condição financeira de bancos e empresas, em especial na redução da taxa de erro de classificação do tipo I. Em 1992, Tam e Kiang (1992) publicaram outro trabalho que teve ainda mais destaque, onde usaram também uma abordagem de rede neural comparando outros modelos. Eles selecionaram 59 bancos em crise e 59 bancos em situação normal no mercado americano entre 1985 e 1987, com base em 19 índices financeiros e compararam quatro modelos diferentes, como *Artificial Neural Networks* (ANN), MDA, Logit e *Decision Trees* (DT). De acordo com seus experimentos, a ANN e a DT tiveram melhor desempenho preditivo. Os resultados empíricos destes estudos foram promissores para

a avaliação das condições financeiras das empresas em termos de precisão preditiva usando a ANN.

Tam e Kiang (1992) discorrem que uma ANN representa uma função discriminante não-linear como um padrão de conexões entre suas unidades de processamento. Huang, Tzeng e Ong (2006) comentam que as ANN foram desenvolvidas para simular a neurofisiologia do cérebro humano como um tipo de modelo de regressão não linear flexível, discriminante e de agrupamento. Os autores afirmam que a arquitetura da ANN pode ser representada como um sistema de três camadas, sendo uma de entrada, que processa os recursos, a outra camada de ocultação com pesos adequados usando uma função de ativação como tangente hiperbólica, softmax ou função logística e a última camada trata da saída.

Brockett et al. (1994) ponderam que um modelo de rede neural pode se adaptar aos dados, sistemas ou problemas e vão se modificando, esta característica não pode ser encontrada em outros modelos estáticos. Assim, muitos trabalhos baseados em ANN, *Support Vector Machine* (SVM) e outros algoritmos geralmente se encaixam bem nos dados para tratar a previsão de dificuldade financeira e falência, mas devido à sua complexidade, eles são considerados como tecnologias de “caixa preta” (OLSON; DELEN; MENG, 2012).

No intuito de se melhorar a precisão na predição do risco de crédito, é possível encontrar algumas pesquisas que utilizam combinações de técnicas e classificadores que integram múltiplos métodos, inclusive fazendo comparações entre eles. Essa combinação de classificadores, técnicas e métodos tem trazido melhor desempenho tanto de predição, quanto na redução dos erros Tipo I e II, mas além dessas medições, é necessário se avaliar o desempenho, o custo e o benefício dessas combinações (WANG; MA, 2012; MARQUÉS; GARCÍA; SÁNCHEZ, 2012; HAJEK; OLEJ; MYSKOVA, 2014; LESSMANN; BAESENS; SEOW; THOMAS, 2015; ABELLÁN; CASTELLANO, 2017; DU JARDIN, 2017; DU JARDIN, 2018).

2.3 Metodologia

A fim de abordar o objetivo estabelecido para esta parte da dissertação, que é avaliar as diversas pesquisas acadêmicas com as múltiplas visões sobre risco de crédito com enfoque em dificuldades financeiras, classificação de crédito, falência e inteligência artificial (IA) e, sua evolução ao longo dos anos, realizou-se uma revisão sistemática para identificar os diferentes construtos desse campo da literatura em artigos publicados no Brasil e no mundo.

A revisão bibliométrica é definida como uma aplicação matemática e estatística de métodos para tratar temas diversos oriundos de vários meios de comunicação (PRITCHARD, 1969). Cole e Eales (1917) discorrem que através de meios estatísticos, pode-se traçar os ramos de um determinado tema que está atraindo maior evidência, e que a influência é exercida pelos trabalhos mais relevantes publicados. Assim, esta revisão bibliométrica possibilita a constatação de lacunas na literatura que podem ser exploradas futuramente.

Para realização deste trabalho, seguiu-se a orientação do processo para formação do conhecimento e triagem dos artigos pelo método *Knowledge Development Process-Constructivist* (ProKnow-C), proposto por De Azevedo, Ensslin e Jungles (2014) e Ensslin et al. (2017), que fornece uma análise de processos de forma estruturada para construir um arcabouço teórico consistente. Este método se baseia na delimitação do campo do conhecimento, na seleção das bases de dados, das palavras-chave, de filtros temporais e critérios para inclusão e exclusão que estão desalinhadas ao que concerne o tema central desta dissertação, além de análise de citações, co-citações, autorias, co-autorias e de periódicos.

2.4 A base de dados

A fase inicial da pesquisa científica começa com a delimitação do problema de pesquisa, que irá estimular o pesquisador a buscar detalhadamente informações e dados sobre um determinado tema em bases de dados bibliográficas (ENSSLIN et al., 2017). Para avaliação, medição e interpretação dos resultados foram realizadas análises quantitativas sobre o Portfólio Bibliométrico (PB) (WANG; VEUGELERS; STEPHAN, 2017).

Com o eixo da pesquisa definido, foram estabelecidos os critérios e filtros para delimitação da base de dados, sendo selecionados os tipos de publicações (artigos e revisões), e identificou-se também os bancos de dados que foram acessados para consulta com o intuito de cobrir o maior número possível de estudos: *Web of Science* e *Scopus (Elsevier)*. Foram excluídos do portfólio bibliométrico: livros, capas de livros, trabalhos duplicados e fora do tema desta pesquisa (DE AZEVEDO; ENSSLIN; JUNGLES, 2014; ENSSLIN et al., 2017). Não foi definido filtro temporal na consulta nos bancos de dados.

Para iniciar a pesquisa, foram então atribuídas as palavras-chave que conduziram a busca pelo referencial para formar o PB: ("bankruptcy" or "business failure" or "credit risk" or "credit scoring" or "expose at default" or "financial distress" or "loss given default" or "probability of default" or "rating" or "risk of default") and ("machine learning" or "artificial

intelligence” or “support vector machine” or “genetic algorithm” or “boosting” or “bagging” or “data mining” or “ensemble”).

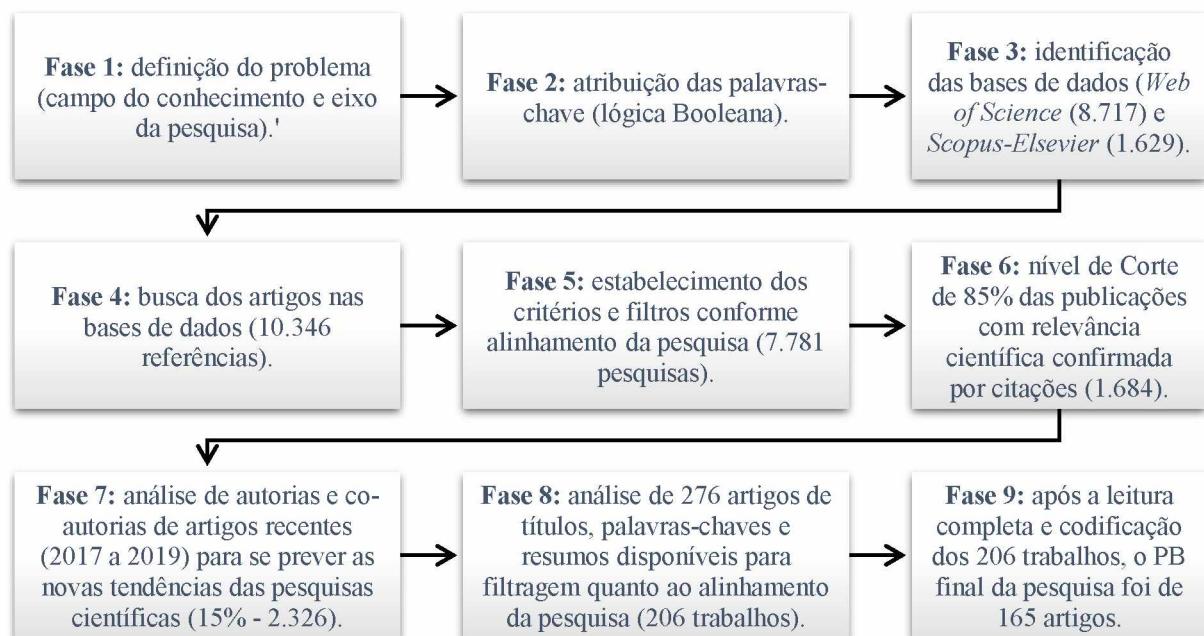
Após a observação e leitura de títulos, palavras-chaves e resumos dos primeiros artigos encontrados nas buscas iniciais, evidenciou-se o uso dessas palavras-chave para delimitar o objetivo deste trabalho. Mas, para explorar melhor o tema deste estudo e aumentar o volume de dados da amostra inicial, fez-se a união destes dois blocos de palavras-chave que deram início a composição do PB.

2.5 Resultado das Pesquisas

Com o emprego dos dois blocos de palavras-chave citados anteriormente, sem utilizar limitação temporal, a pesquisa selecionou na primeira amostra 10.346 estudos, entre artigos e revisões, sendo 8.717 do banco de dados *Web of Science* e 1.629 do banco de dados *Scopus* (Fases 1 a 4 - Figura 2).

A Figura 2 apresenta o resumo do processo de seleção dos artigos para a construção do referencial teórico deste estudo (Portfólio Bibliométrico).

Figura 2 - Resumo da metodologia ProKnow-C, similar a De Azevedo, Ensslin e Jungles (2014) e Ensslin et al. (2017)



Fonte: elaborado pelos autores.

Dessa base total de 10.346 estudos, foram excluídas 780 pesquisas duplicadas, o que resultou em uma amostra de 9.566 referências. Após uma codificação e ordenação dos títulos e

palavras-chave de cada artigo, os registros foram lidos para observar o quanto estavam alinhados com o tema deste estudo e, após esta revisão, 1.785 trabalhos foram excluídos por não terem aderido ao tema principal desta pesquisa, restando 7.781 artigos para posterior análise (Fase 5 - Figura 2).

Em decorrência do grande volume ainda existente na base de dados, optou-se por fazer a verificação da relevância científica dos trabalhos e as datas de publicação. Usou-se o número de citações e co-autorias, que medem o impacto das contribuições dos autores e identifica suas colaborações científicas, cujos caminhos podem ser seguidos por meio de redes, e de acordo com a evolução e as melhorias desenvolvidas, novas oportunidades se tornam disponíveis para investigação (BORGMAN, 1989). Para identificar os artigos mais relevantes, ou seja, aqueles mais citados, foi realizado um corte que corresponde 85% da soma total de citações dos 7.781 estudos. Este limite foi baseado no postulado de Pareto, que afirma que a maioria dos efeitos se originam de um pequeno número de causas (DE AZEVEDO; ENSSLIN; JUNGLES, 2014).

Assim, os estudos foram classificados em ordem decrescente do número de citações individuais, sendo considerado o maior número de citações por documento, os artigos foram selecionados até que o percentual de citações acumuladas fosse igual ao valor de corte (85%). Após esse processo, o conjunto bibliográfico concluiu com 1.684 trabalhos (Fase 6 - Figura 2).

Para o restante da base de 15% menos citados, usou-se as autorias e co-autorias da base final dos 85% mais relevantes para se prever as novas tendências em campos científicos publicados de 2017 a 2019 (GLANZEL, 2002). A base inicial foi de 2.326 estudos. Utilizando este critério e também a análise de títulos e palavras-chaves quanto ao alinhamento do tema desta pesquisa, foram adicionados 30 estudos resultantes desta investigação e assim foram integrados ao PB (Fase 7 - Figura 2).

Nos 1.684 estudos, foram identificados pela leitura de títulos e palavras-chave 1.438 artigos que não estavam alinhados com o objetivo da pesquisa, o que resultou numa base de 246 artigos alinhados ao tema desta pesquisa, que somado aos 30 trabalhos, totalizaram 276 pesquisas (Fase 8 - Figura 2).

No conjunto total (276 estudos), foi avaliado se o resumo estava alinhado com o objetivo central desta pesquisa. Esta análise resultou na exclusão de 70 artigos. A base do PB então resultou em 206 trabalhos alinhados ao tema desta pesquisa (Fase 8 - Figura 2).

Realizou-se então a busca dos trabalhos e os 206 estavam disponíveis e foram armazenados, inclusive atualizado o volume de citações por artigo com busca no site Google

Acadêmico. Procedeu-se a leitura completa e a codificação destes trabalhos que resultaram uma base final do PB 165 artigos. Foram excluídos 41 artigos com base de dados de pessoa física, que não estão ligados ao tema central deste trabalho. Um quadro constando algumas informações referente aos artigos selecionados pelo método ProKnow-C é apresentado no Apêndice A ao final desta pesquisa (Fase 9 - Figura 2).

2.6 Análise Descritiva Quantitativa

No levantamento de artigos que conduziu este estudo segundo a metodologia ProKnow-C, constatou 165 trabalhos relacionados à aplicação dos métodos de AI, ML e Redes Neurais Artificiais (ANN) para tratar dificuldade financeira, risco de crédito, falência e inadimplência. Esse acervo englobou um total de 302 pesquisadores distintos. Entre eles, 55 autores publicaram mais de um trabalho no portfólio bibliométrico relacionados a gestão de risco de crédito, previsão de falência e dificuldades financeiras, utilizando modelos de aprendizado de máquina e redes neurais.

A tabela 1 apresenta os 15 autores que publicaram mais de 4 artigos. Nesta tabela as citações independem da autoria ou co-autoria, ou seja, são adicionais por autor e artigo. O autor Chih-Fong Tsai é o mais ativo, com 13 trabalhos publicados (citações 1.259), os mais citados são os de 2008 e 2009, e suas publicações como primeiro autor, foram até 2014, nos artigos mais recentes ele está como co-autor. Geralmente seus trabalhos utilizam as mesmas fontes de dados (Alemanha, Japão e Austrália), e não informa o horizonte de tempo da base, nos seus artigos são realizadas comparações entre técnicas individuais e de conjuntos, híbrida ou não, e a maioria utiliza as técnicas MLP e SVM. Hui Li e Jie Sun ficam em segundo lugar com 10 trabalhos cada (747 citações cada um), eles publicam juntos alternando nos trabalhos como primeiros autores. Geralmente utilizam o mesmo conjunto de dados chineses, e mesmo horizonte de tempo da base, procuram utilizar modelos híbridos e trabalham com diversas técnicas. Em Terceiro lugar, com 7 estudos publicados sobre o tema desta pesquisa ficam Vicente García, Ana I. Marqués, J. Salvador Sánchez (319 citações cada um), geralmente os autores García Marqués e Vicente García se alternam como primeiros autores e Sánchez está na co-autoria em todos os trabalhos e apesar de serem de universidades distintas, sempre publicam trabalhos conjuntamente. Quando se eleva o número de publicações para 6, o total de pesquisas cai para 9 autores mais ativos.

Tabela 1 – Autores mais ativos evidenciados sobre o tema desta pesquisa por número de publicações

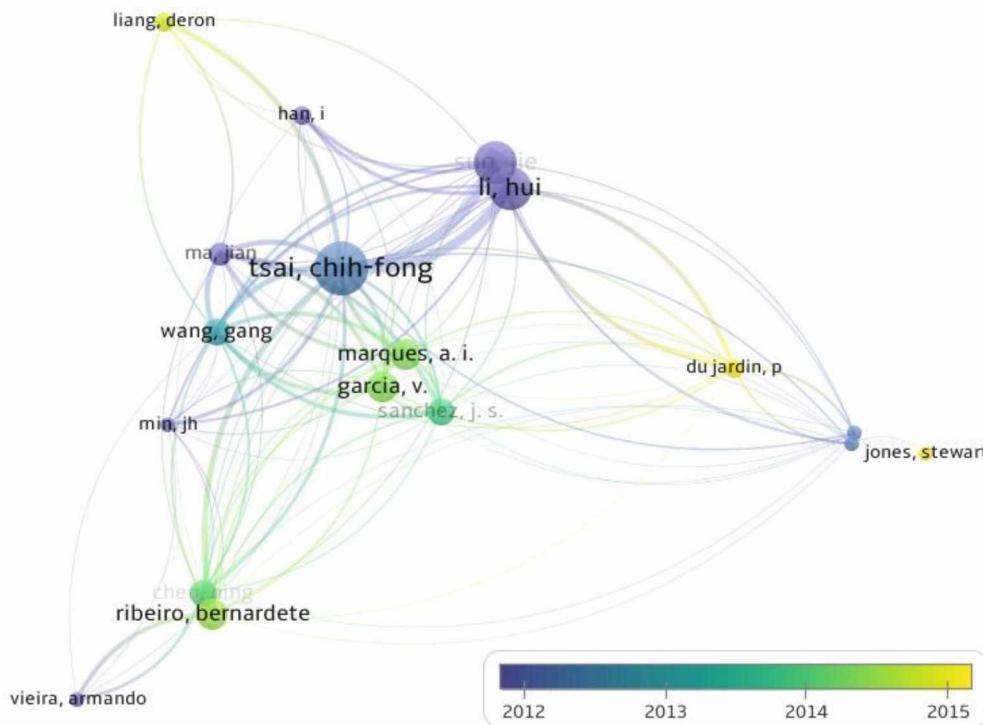
Autor	Publicações	Citações
Tsai, CF	13	1.259
Li, H	10	747
Sun, J	10	747
García, V	7	319
Marqués, AI	7	319
Sánchez, JS	7	319
Ribeiro, B	7	229
Wang, G	6	570
Chen, N	6	215
Han, I	4	1.031
Kim, MJ	4	429
Ma, J	4	319
Liang, D	4	198
Du Jardin, P	4	101
Jones, S	4	101

Fonte: resultados da pesquisa.

A Figura 3 demonstra os 15 autores mais dominantes ativamente e suas ligações. Para se avaliar a relevância destes estudos foi considerado o número de citações por estudo. Du Jardin, Jones, Liang e Bernardete, são os autores que mais publicaram nos últimos 5 anos. Os autores Hui Li e Jie Sun, neste PB, todos os seus trabalhos foram publicados conjuntamente. Os autores Bernardete Ribeiro e Armando S. Vieira estão sempre publicando como autores e co-autores.

O autor Chih Fong Tsai de Taiwan é o mais citado entre os autores mais ativos, pois ele foi precursor ao estudar o tema. Seus artigos tratam de modelos híbridos, que utilizam combinações de técnicas de conjuntos classificadores para previsão de FD, falência e para prever o risco de crédito. Hui Li, da China, é especialista em ML, ANN e *data mining*, desenvolve modelos híbridos com *case based reasoning* e que também utilizam combinações de técnicas de conjuntos classificadores para previsão de dificuldade financeira e risco de crédito. Em um dos artigos selecionados no PB, o autor propõe uma abordagem *Elimination And Choice Net Translating Reality-ELECTRE* (LI; SUN, 2009).

Figura 3 – Mapa de autores mais ativos e suas ligações por *aging* de publicação



Fonte: VOSviewer – resultados da pesquisa.

A tabela 2 apresenta os autores e co-autores que mais foram citados em relação ao tema desta pesquisa por ordem de citações e publicações respectivamente, que são: Varetto, F (1617), foi o mais citado, participou como co-autor em um artigo publicado no *Journal of Banking & Finance*, no qual Edward Altman (um dos autores mais relevantes no que tange risco de crédito) foi o primeiro autor. Kiang, MY (1444) e Tam, KY (1444) foram os precursores a utilizar modelos de aprendizado de máquina para prever falência. Altman, E (1393) é o autor seminal que trata de predição com o modelo de MDA e referência no tema. Marco, G (1336) também é mais citado onde foi co-autor em um artigo publicado no *Journal of Banking & Finance*, onde Altman foi o primeiro autor.

Tabela 2 – Autores mais evidenciados sobre o tema desta pesquisa por número de citações

Autor	Publicações	Citações
Vareto, F	2	1.617
Kiang, MY	1	1.444
Tam, KY	1	1.444
Altman, E	2	1.393
Marco, G	1	1.336
Tsai, CF	13	1.259
Shin, KS	2	1.110
Lee, YC	3	1.052
Han, I	4	1.031
Min, JH	3	958
Chen, HC	1	949
Chen, WH	1	949
Hsu, CJ	1	949
Huang, Z	1	949
Wu, SS	1	949
Sharda, R	1	821
Wilson, RL	1	821
Pan, WT	1	815

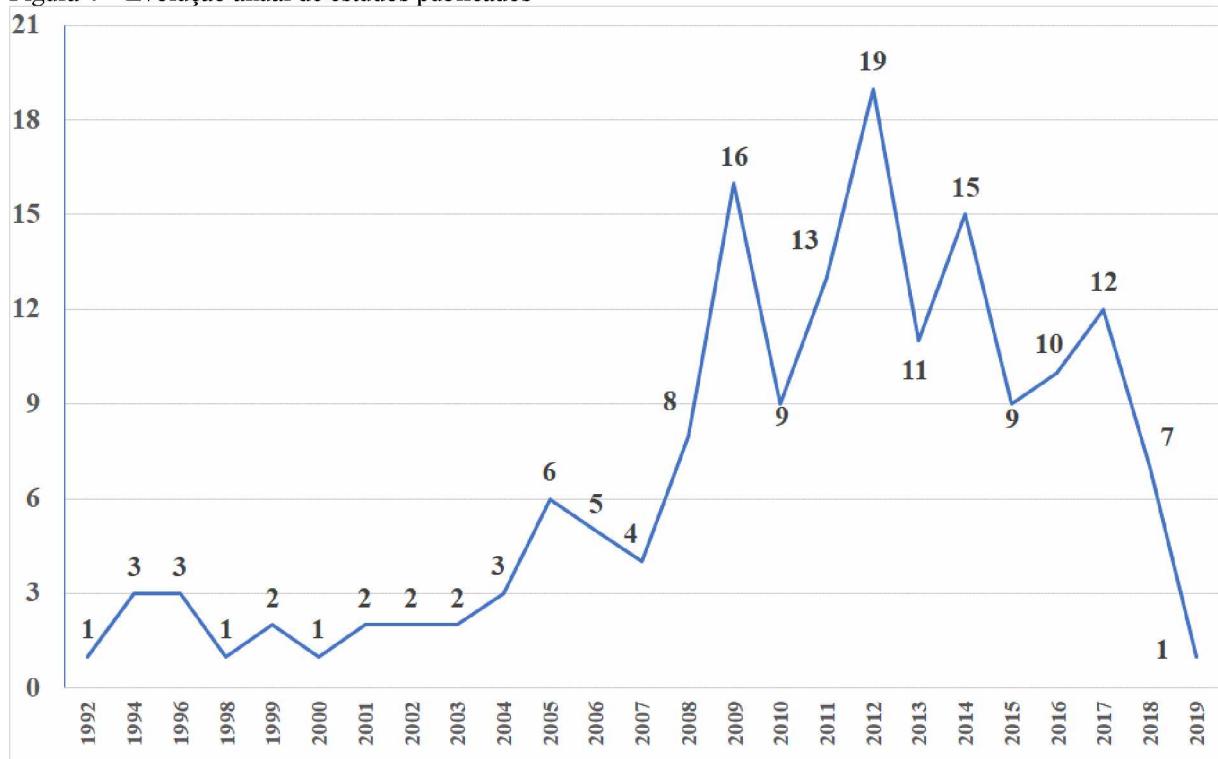
Fonte: resultados da pesquisa.

Esta é uma breve revisão que discorre sobre principais autores que abordaram em relação ao tema central desta pesquisa e pode-se destacar a China, Taiwam, USA, Espanha e Coreia do Sul como as escolhas que lideram os estudos ligados a gestão de risco de crédito, previsão de falência e FD, utilizando modelos de aprendizado de máquina, ANN e AI. Considerando a análise até o momento, os autores que desenvolvem modelos nessa área procuram melhorar as técnicas, fazendo cruzamentos e agrupamentos de bases de dados e modelos distintos, para melhorar a precisão da predição dos modelos.

2.7 Artigos com relevância científica

De acordo com a análise dos artigos e das revisões, foram expostas as características fundamentais de cada estudo. Esses atributos compreendem a periodicidade das publicações, os periódicos em que os trabalhos foram publicados, a autoria e co-autoria dos documentos, as características das bases de dados utilizadas, os métodos, técnicas e classificadores e os países que serviram de base para esta pesquisa. Os documentos examinados no PB foram publicados entre o período de 1992 e 2019.

Figura 4 – Evolução anual de estudos publicados



Fonte: resultados da pesquisa.

A Figura 4 demonstra a evolução quantitativa de artigos publicados anualmente. Nota-se que houve um aumento considerável de artigos publicados após o ano de 2008, o que pode ser explicado pela crise financeira ocorrida em meados de 2007/2008, também pela melhoria na infraestrutura e evolução dos recursos computacionais, que são capazes de armazenar e transacionar uma vasta quantidade de dados, bem como o aperfeiçoamento dos métodos de AI (MORO; CORTEZ; RITA, 2015). Cole e Eales (1917) descrevem sobre a questão de muitas pesquisas serem concentradas em um período específico e em uma determinada região, por mais indiferente que seja, esta é uma indicação segura do atendimento aos interesses e atividades contemporâneos e que podem-se traçar os ramos daquele tema que estava atraindo maior atenção. Isso reforça o interesse pelo tema, que explora os mecanismos de aprendizado de máquina para prever dificuldades financeiras nas organizações.

Tabela 3 – Periódicos em destaque sobre o tema desta pesquisa

Periódico	JCR	SJR
<i>Information Fusion</i>	10.716	2,238
<i>Ieee Transactions on Fuzzy Systems</i>	8.759	2,794
<i>Ieee Transactions on Systems Man Cybernetics-Systems</i>	7.351	2,147
<i>Tourism Management</i>	6.012	2,924
<i>Neural Networks</i>	5.785	1,970
<i>Information Sciences</i>	5.524	1,620
<i>Omega-International Journal of Management Science</i>	5.341	0,426
<i>Knowledge-Based Systems</i>	5.101	1,460
<i>Artificial Intelligence Review</i>	5.095	1,055
<i>Applied Soft Computing</i>	4.873	1,216
<i>Neural Computing & Applications</i>	4.664	0,637
<i>Technological and Economic Development of Economy</i>	4.344	0,774
<i>Expert Systems With Applications</i>	4.292	1,190
<i>Management Science</i>	4.219	6,080
<i>Information & Management</i>	4.120	0,683

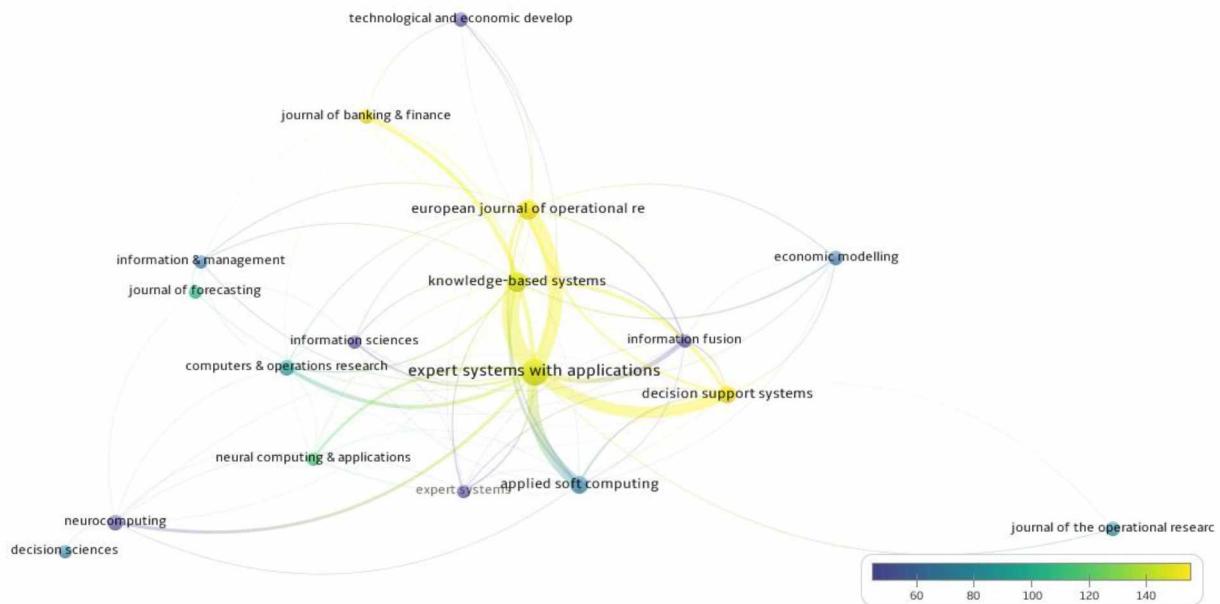
Fonte: resultados da pesquisa.

Para se medir a relevância dos periódicos do PB, foram considerados os indicadores de classificação *Journal Citation Reports* (JCR), para comparação direta com os periódicos da área de conhecimento assim priorizou a classificação por este índice, (JCR, 2019) e o *Scientific Journal Rankings* (SJR), para se comparar com outros periódicos referentes a outras áreas do conhecimento (SCIMAGO, 2018).

A Tabela 3 traz a posição dos 15 periódicos mais relevantes de acordo com seu fator de impacto dentro da análise deste PB. O critério utilizado para o mapeamento dos periódicos em destaque sobre o tema desta pesquisa foi o número de citações e de publicações. Além disso, a Figura 5 traz este mapeamento de periódicos mais citados e a relação entre eles nos estudos que compõem este PB.

Pode-se observar que, os trabalhos do PB foram publicados em 44 periódicos distintos, sendo que, o *Expert Systems With Applications*, o *Knowledge-Based Systems*, o *European Journal of Operational Research* e o *Decision Support Systems*, foram os que tiveram maior destaque em relação a citações e também publicações. Assim pode-se inferir então, que a maioria das publicações que compõe o PB desta pesquisa, estão voltadas para estudos empíricos de construção e comparabilidade de modelos para previsão e gestão do risco de crédito.

Figura 5 – Mapa com periódicos em destaque sobre o tema desta pesquisa



Fonte: VOSviewer – resultados da pesquisa.

A tabela 4 traz numericamente a distribuição do PB entre os periódicos. Em relação a média de citações por publicação o *Management Science* tem 159% mais citações que o segundo nesta posição que é o *Journal of Banking & Finance* que também está 54% acima do seguinte.

Tabela 4 – Periódicos com mais publicações e citações do PB de 1992 a 2019.

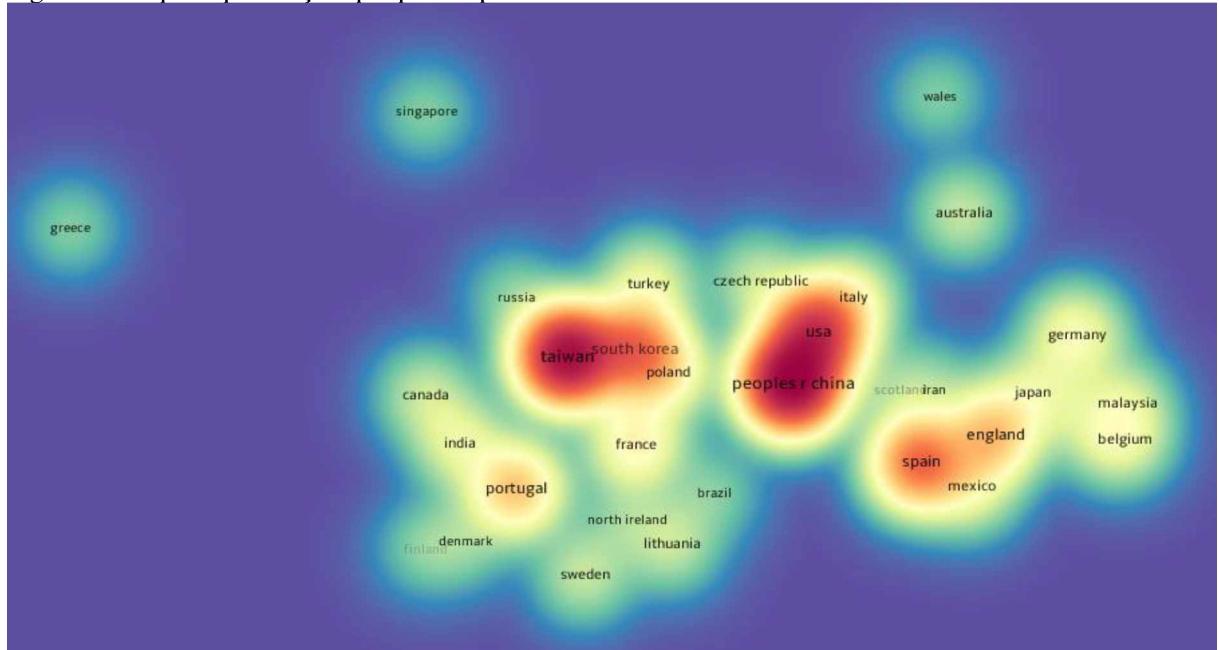
Periódico	Publicações	Citações	Média Citações/ Publ.	% partic. Cit.	JCR	SJR
<i>Expert Systems With Applications</i>	57	8.193	144	35,1%	4.292	1,190
<i>Decision Support Systems</i>	9	2.863	318	12,3%	3.847	1,536
<i>Knowledge-Based Systems</i>	15	2.155	144	9,2%	5.101	1,460
<i>European Journal of Operational Research</i>	14	2.110	151	9,0%	3.806	2,205
<i>Journal of Banking & Finance</i>	3	1.671	557	7,2%	2.205	1,599
<i>Management Science</i>	1	1.444	1.444	6,2%	4.219	6,080
<i>Applied Soft Computing</i>	8	576	72	2,5%	4.873	1,216
<i>Ieee Transactions on Fuzzy Systems</i>	1	361	361	1,5%	8.759	2,794
<i>Computers & Operations Research</i>	4	360	90	1,5%	3.002	1,859
<i>Journal of Management Information Systems</i>	1	324	324	1,4%	3.013	2,388
Outros	52	3.300	63	14,1%		
Total	165	23357	142	100,0%		

Fonte: resultados da pesquisa.

Analisou-se também os países de origem das pesquisas, que contemplou 34 países distintos, referente as instituições em que os autores são associados, determinando a procedência desses trabalhos. A figura 6, traz o mapeamento por país de origem das pesquisas

deste PB, em destaque de grandeza de citações tem-se os autores dos USA (5.880), Taiwan (4.421), China (2.730), Coreia do Sul (2.388) e Itália (1.801), o que representa 71% do total das pesquisas deste portfolio bibliométrico.

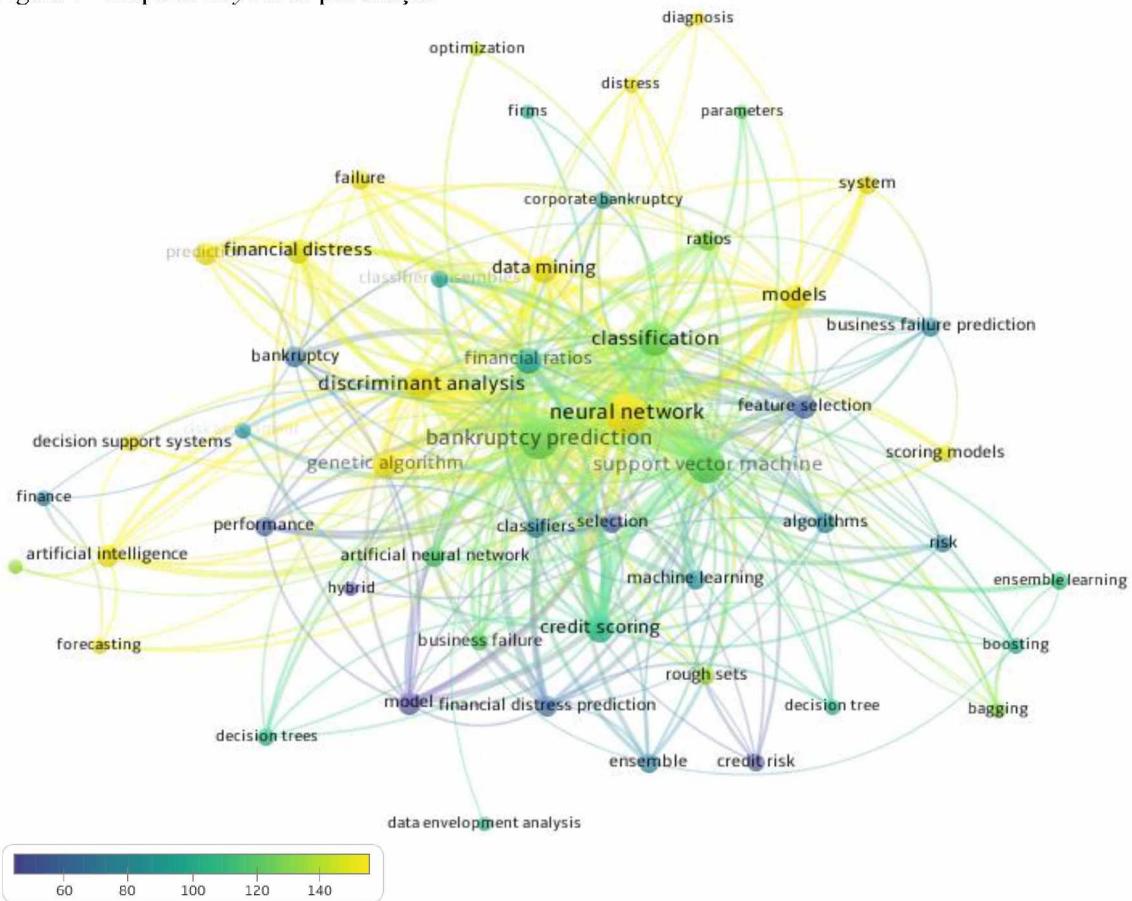
Figura 6 – Mapa de publicações por país de procedência



Fonte: VOSviewer – resultados da pesquisa.

A figura 6 traz o mapa com as regiões de maior densidade, que demonstra onde se concentra o maior volume pesquisas, que são China (33), Taiwan (30), USA (20), Espanha (14) e Coreia do Sul (13), o que representa 67% do PB. Segundo Tollefson (2018) o governo chinês vem ampliando gradualmente os investimentos em ciência, desenvolvimento e inovação tecnológica, o que pode ter contribuído para o grande volume de pesquisas científicas. Ao contrário, as regiões com menor densidade no mapa, são as que apresentam menor número de pesquisas em relação ao tema central deste PB, que são: Países de Gales, Suíça, África do Sul, Cingapura, Escócia, Rússia, Irlanda do Norte, Irã, Grécia, Finlândia, Dinamarca e Brasil, onde cada uma produziu apenas 1 artigo com relevância científica. Estudos que contemplam pesquisadores dessas regiões e de outras, onde não houve nenhum artigo relacionado ao tema central deste trabalho, como países da América Latina, Israel, Hungria entre outros, que poderiam trazer enriquecimento ao acervo de pesquisas com resultados distintos e múltiplas visões.

Figura 7 - Mapa de *Keywords* por citação



Fonte: VOSviewer – resultados da pesquisa.

A Figura 7 classifica as palavras-chave (*keywords*) dos artigos que fazem parte deste PB. É possível constatar que as palavras que mais se destacaram foram: *neural network* (97), *bankruptcy prediction* (91), *support vector machine* (70), *classification* (60), *discriminant analysis* (45), *credit scoring* (33), *data mining* (31), *genetic algorithm* (29), *financial ratios* (27), *models* (26), *financial distress* (23), *feature selection* (22), *artificial intelligence* (20), *model* (20), *prediction* (20), confirmado a exatidão quanto a delimitação do PB em relação ao tema central desta pesquisa e a aderência quanto a ferramenta utilizada (ProKnow-C). Identificou-se também que nos níveis extremos do mapeamento tem-se as técnicas *ensemble*, *boosting* e *bagging* que estão sendo mais utilizadas ultimamente, pois segundo alguns estudos, estes métodos melhoram o desempenho diante dos modelos individuais (WANG; HAO; MA; JIANG, 2011; LIANG; TSAI; DAI; EBERLE, 2017; entre outros).

2.8 Análise quantitativa dos artigos do PB

Encerrada a primeira etapa de análise dos principais trabalhos relacionados ao tema central desta pesquisa, o PB foi classificado e codificado conforme as categorias descritas na tabela 5, cada uma dessas categorias de classificação foi numerada de 1 a 8, sendo: 1 - Contexto, 2 – Tipo de pesquisa, 3 – Métodos e técnicas, 4 - Foco, 5 – País de origem da base, 6 – Variável dependente, 7 – Variáveis independentes, 8 – Horizonte de tempo (base). Utilizou-se as letras do alfabeto para codificação das subcategorias para aprofundar ainda mais a análise. As subcategorias da categoria 3 – Métodos e técnicas, onde se agrupam os métodos, técnicas e classificadores, foram codificadas pelas abreviações para facilitar a logística dos dados.

Tabela 5 - Categorização para análise do PB

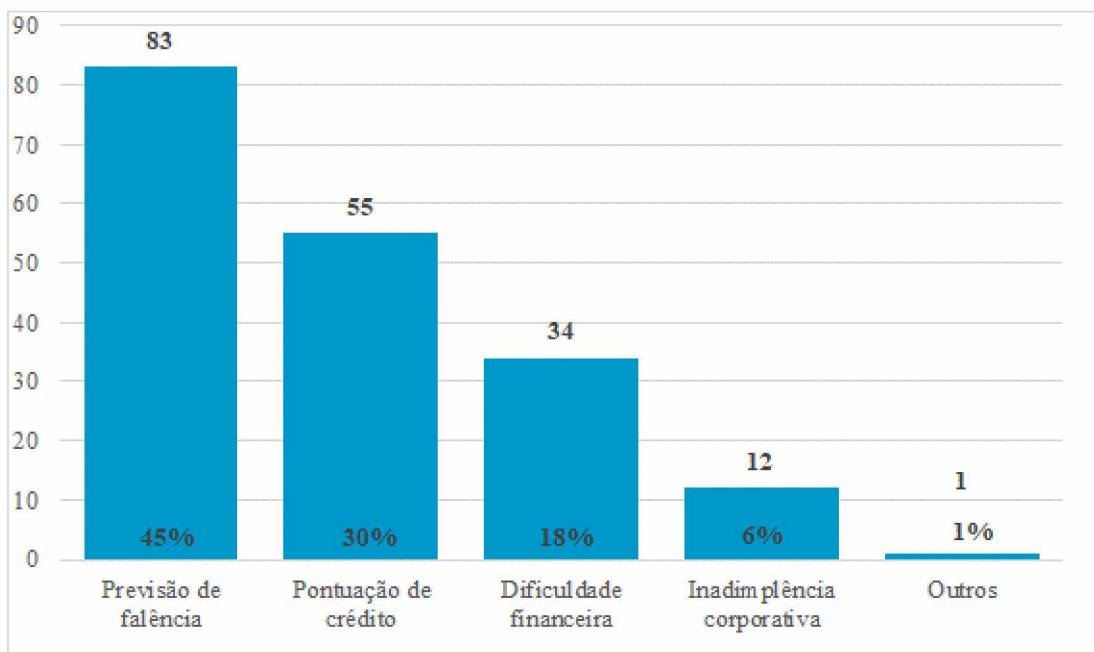
Categorização	Significado	Código das alternativas	Significado
1	Contexto	A B C D E	Inadimplência corporativa Dificuldade financeira Previsão de falência Pontuação de crédito Outros
2	Tipo de pesquisa	A B	Empírico Revisão
3	Métodos e técnicas	SVM LR ANN BA LDA MLP MDA DT BOO AB NB CART KNN RF BPNN GA RS RBF CBR RST SOM RTF ELM ML SMO O	<i>Support Vector Machine</i> <i>Logistic Regression</i> <i>Artificial Neural Networks</i> <i>Bagging</i> <i>Linear Discriminant Analysis</i> <i>MultiLayer Perceptron</i> <i>Multiple Discriminant Analysis</i> <i>Decision Tree</i> <i>Boosting</i> <i>AdaBoost</i> <i>Naive Bayes</i> <i>Classification and Regression Trees</i> <i>K-Nearest Neighbor</i> <i>Random Forest</i> <i>Back Propagation Neural Network</i> <i>Genetic Algorithm</i> <i>Random Subspace</i> <i>Radial Basis Function</i> <i>Case Based Reasoning</i> <i>Rough Set Theory</i> <i>Self Organizing Map</i> <i>Rotation Forest</i> <i>Extreme Learning Machine</i> <i>Machine Learning</i> <i>Sequential Minimal Optimization</i> Outros
4	Foco	A B C	Instituições Financeiras Não Financeiras Outros

5	País de origem da base	A	Estados Unidos da América
		B	Europa
		C	Ásia
		D	América Latina
		E	Austrália
		F	Outros / Não mencionado
6	Variável dependente	A	Falência
		B	Dificuldade Financeira
		C	Inadimplência
		D	<i>Loss Given Default (LGD)</i>
		E	Outros
7	Variáveis independentes	A	Indicadores financeiros
		B	Índices não financeiros
		C	<i>LGD</i>
		D	Outros
8	Horizonte de tempo (base)		Início final

Fonte: elaborado pelos autores.

As variáveis de interesse que tratam do contexto dos trabalhos, Figura 8, foram classificadas e agrupadas, e percebeu-se que a maior parte dos estudos, 83 (45%), procuram identificar o comportamento das empresas que entraram em falência em relação as empresas que não faliram. Já os modelos que criam score e classificam o risco foram 55 estudos (28%) e 34 trabalhos (18%) procuraram prever a dificuldade financeira das organizações.

Figura 8 – Classificação dos artigos por contexto da pesquisa.



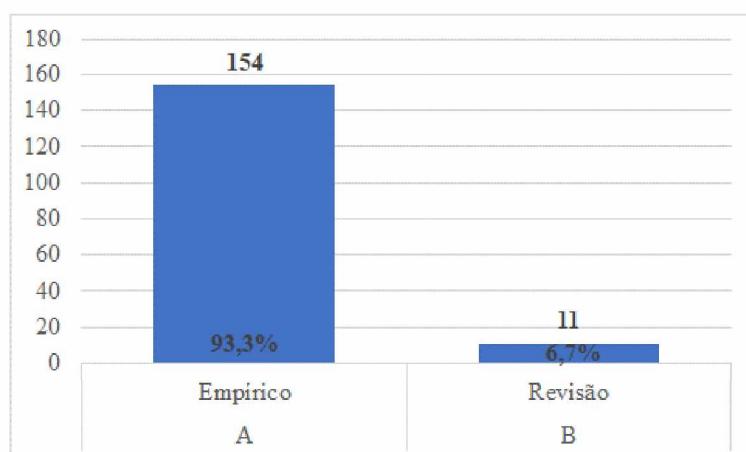
Fonte: resultados da pesquisa.

Dentro do próprio ciclo de vida da organização ela passa por fases de dificuldades financeiras, quando esta não está totalmente estabilizada para comportar tais eventos, o desfecho pode resultar em falência ou recuperação judicial, por isso, a precisão na previsão dessas ocorrências para os tomadores de decisões e demais envolvidos é relevante e contribui para que a situação não se torne ainda mais grave, aumentando o número de falhas e se tornando um processo contagioso e devastador para a economia de uma país (ALTMAN; BAIDYA; DIAS, 1979; GENG; BOSE; CHEN, 2015; ZELENKOV; FEDOROVA; CHEKRIZOV, 2017; ALAKA et al., 2018).

No que tange ao tipo de pesquisa conforme a figura 9, a maior parte dos trabalhos são estudos empíricos (152), cerca de 93,3% e o restante, 11 artigos (6,7%), tratam de revisões de literatura. Os trabalhos do PB desenvolvem modelos que utilizam a tecnologia moderna para simplificar processos e aumentar a eficiência na precisão e predição, procurando gerenciar de maneira proativa o risco de crédito, as dificuldades financeiras e as falências corporativas (GARCÍA; MARQUÉS; SANCHÉZ, 2019).

Uma situação importante para se destacar entre os modelos computacionais seria a interface com a análise financeira qualitativa para se explicar melhor e facilitar a implementação e teste dos modelos construídos (HAJEK; OLEJ; MYSKOVA, 214). A análise financeira qualitativa poderia atender as perspectivas regionais e condições regulatórias, os estudos de Ribeiro, Chen e Kovacec (2017) trazem uma abordagem baseada em algoritmo para construção de gráficos, utilizando relações binárias de dados qualitativos de risco de crédito, e os resultados demonstram que o desempenho da precisão de previsibilidade pode ser melhorado.

Figura 9 – Tipo de Pesquisa do Portfólio Bibliométrico



Fonte: resultados da pesquisa.

Para se prever com precisão a dificuldade financeira e a falência das organizações, vários métodos foram desenvolvidos e testados. Eles podem ser divididos nas seguintes categorias: métodos estatísticos tradicionais e métodos baseados em AI. Os métodos estatísticos tradicionais funcionam bem e são mais simples e compreensíveis. Os modelos baseados em redes neurais artificiais, máquinas de vetores de suporte e outros algoritmos, geralmente apresentam desempenho superior e se encaixam de forma satisfatória na base de dados para a previsibilidade, mas devido à complexidade dos métodos e à falta de compreensão de analistas e gestores, os modelos são considerados tecnologias de caixa preta (OLSON; DELEN; MENG, 2012).

Tabela 6 – Métodos e técnicas mais utilizados no PB

<i>Abrev.</i>	<i>Method</i>	<i>Frequency</i>	<i>% Prop.</i>
LR	Regressão Logística	76	10,1%
SVM	<i>Support Vector Machine</i>	75	10,0%
ANN	<i>Artificial Neural Networks</i>	73	9,7%
DT	<i>Decision Tree</i>	50	6,7%
BA	<i>Bagging</i>	29	3,9%
LDA	<i>Linear Discriminant Analysis</i>	28	3,7%
MLP	<i>MultiLayer Perceptron</i>	27	3,6%
MDA	<i>Multiple Discriminant Analysis</i>	27	3,6%
BOO	<i>Boosting</i>	25	3,3%
AB	<i>AdaBoost</i>	20	2,7%
NB	<i>Naive Bayes</i>	19	2,5%
CART	<i>Classification and Regression Trees</i>	18	2,4%
KNN	<i>K-Nearest Neighbor</i>	18	2,4%
RF	<i>Random Forest</i>	17	2,3%
BPNN	<i>Back Propagation Neural Network</i>	15	2,0%
GA	<i>Genetic Algorithm</i>	15	2,0%
RS	<i>Random Subspace</i>	14	1,9%
RBF	<i>Radial Basis Function</i>	12	1,6%
CBR	<i>Case Based Reasoning</i>	10	1,3%
RST	<i>Rough Set Theory</i>	9	1,2%
SOM	<i>Self Organizing Map</i>	9	1,2%
RTF	<i>Rotation Forest</i>	7	0,9%
ELM	<i>Extreme Learning Machine</i>	6	0,8%
ML	<i>Machine Learning</i>	5	0,7%
O	Outros	145	19,4%

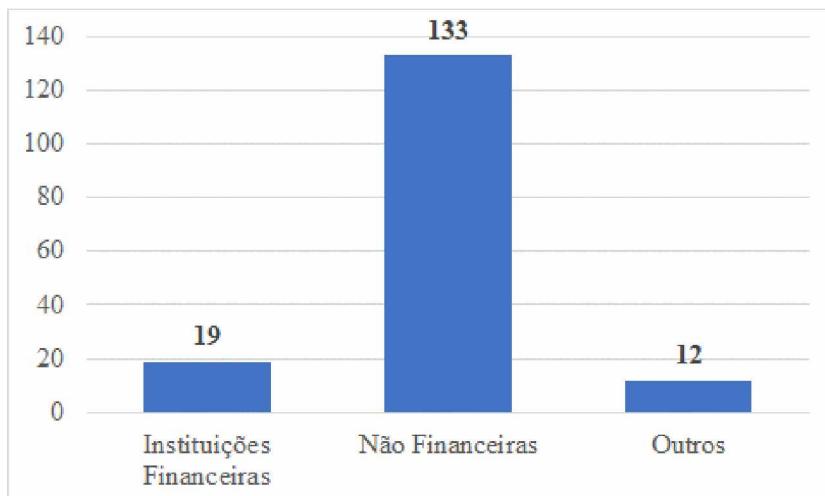
Fonte: resultados da pesquisa.

A tabela 6 apresenta os métodos e técnicas mais utilizados no PB, onde a Regressão Logística (10,1%) foi a mais utilizada, seguido do *Support Vector Machine* (10,0%) e *Artificial Neural Network* (9,7%). Embora existam múltiplos métodos e técnicas, conjuntos, individuais e híbridos, para previsão de dificuldade financeira e falência, eles se comportam de maneira

diferente em várias situações, conforme a base de dados e as variáveis selecionadas, neste PB, foram utilizadas 104 métodos e técnicas. Assim, não é possível especificar um único método apropriado que possa se adaptar e funcionar melhor que qualquer outro (WANG; CHEN; CHU, 2018).

Em relação ao conjunto de dados utilizados pelos estudos, quando se trata de previsão de dificuldade financeira e falências, ele é caracterizado pelo desequilíbrio das classes, onde a distribuição de instâncias das classes de dados é distorcida, pois o número de empresas em condições ruins é bem menor do que as empresas em condições normais (WANG; CHEN; CHU, 2018). Um algoritmo utilizado para equacionar este problema é o SMOTE (*Synthetic Minority Over-Sampling Techinique*), que ajuda a manter o equilíbrio de distribuição entre as classes, contribuindo para os resultados preditivos diante de conjuntos de dados desequilibrados (MARQUÉS; GARCÍA; SANCHÉZ, 2013; ZHOU, 2013; HAJEK; OLEJ; MYSKOVA, 2014; SUN; LANG; FUJITA; LI, 2018; WANG; CHEN; CHU, 2018).

Figura 10 – Distribuição da base de dados por *Focus*



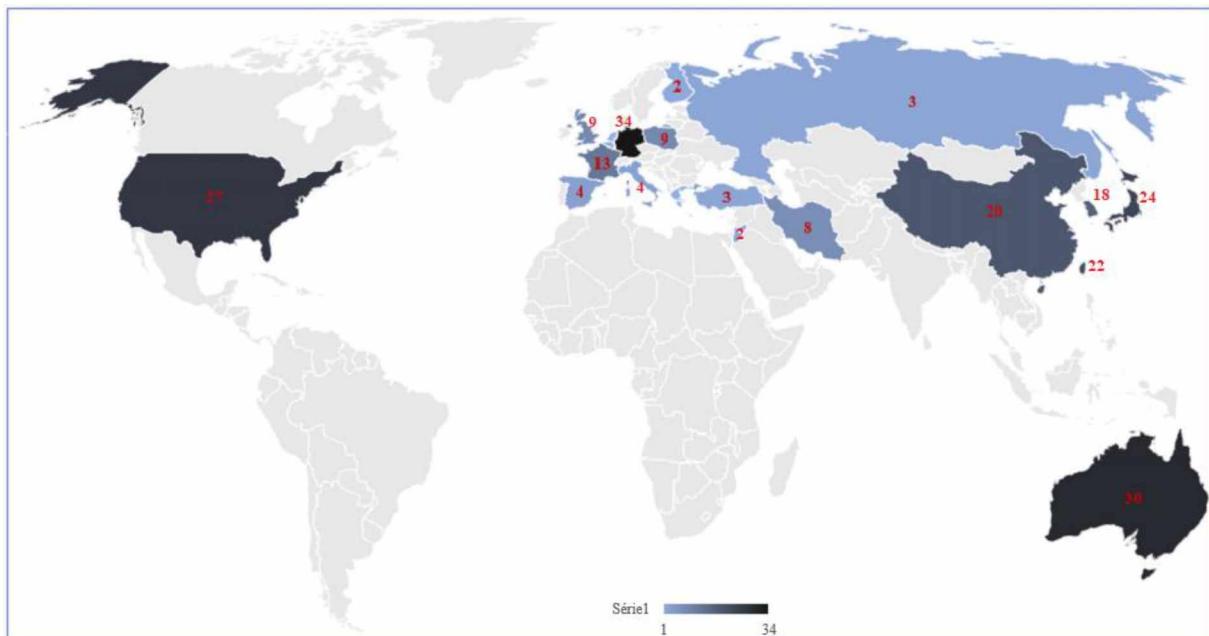
Fonte: resultados da pesquisa.

A figura 10 traz os artigos deste PB, que focaram em trabalhar, em sua maioria (81%), usando conjunto de dados de empresas não financeiras, 13% utilizaram dados de financeiras ou seguradoras e o restante, 8%, não foi possível identificar. Cada segmento apresenta um sistema regulatório específico, são segmentos de mercado distintos, portanto faz-se necessária essa especificação, e para filtrar ainda mais este PB, optou-se, nesta pesquisa, por excluir os estudos com conjuntos de dados de pessoas físicas (exemplo de artigo desconsiderado: GUO; ZHOU;

LUO; LIU; XIONG, 2016). Dentre os artigos que não possuíam esta informação estão Hu (2009) e Lahsasna, Ainan e Teh (2010).

A partir desta revisão de literatura, observou-se que em relação à origem dos bancos de dados, ou seja, dos países utilizados para os experimentos, que a maioria dos trabalhos empregaram conjuntos de dados da Alemanha (12,5%), Austrália (11,1%), USA (10%), Japão (8,9%), Taiwan (8,1%), China (7,4%), Coreia (6,6%) e França (4,8%). A utilização destas bases de dados para se tratar a gestão do risco de crédito, permite que outros pesquisadores possam comparar novos modelos e também seus resultados. Mas quando esses bancos de dados são utilizados demasiadamente, podem não apresentar as condições específicas de outros países, mercados, ou segmentos, devido a questões regulatórias e as condições socioeconômicas específicas, assim os experimentos podem levar a conclusões distorcidas e talvez não atendam as demandas dos diversos mercados, pois os fatores que afetam a falência variam entre os diferentes países (BALCAEN; OOGHE, 2006; GARCÍA; MARQUÉS; SANCHÉZ, 2014; LIANG; TSAI; DAI; EBERLE, 2017; TIAN; YU, 2017). Assim, faz-se necessário acrescentar pesquisas de outros países, como Brasil, países da América Latina, Israel, que, nesta base de dados do PB, não apresentou nenhuma pesquisa. A figura 11 traz os países mais utilizados como conjuntos de dados para experimento e análise.

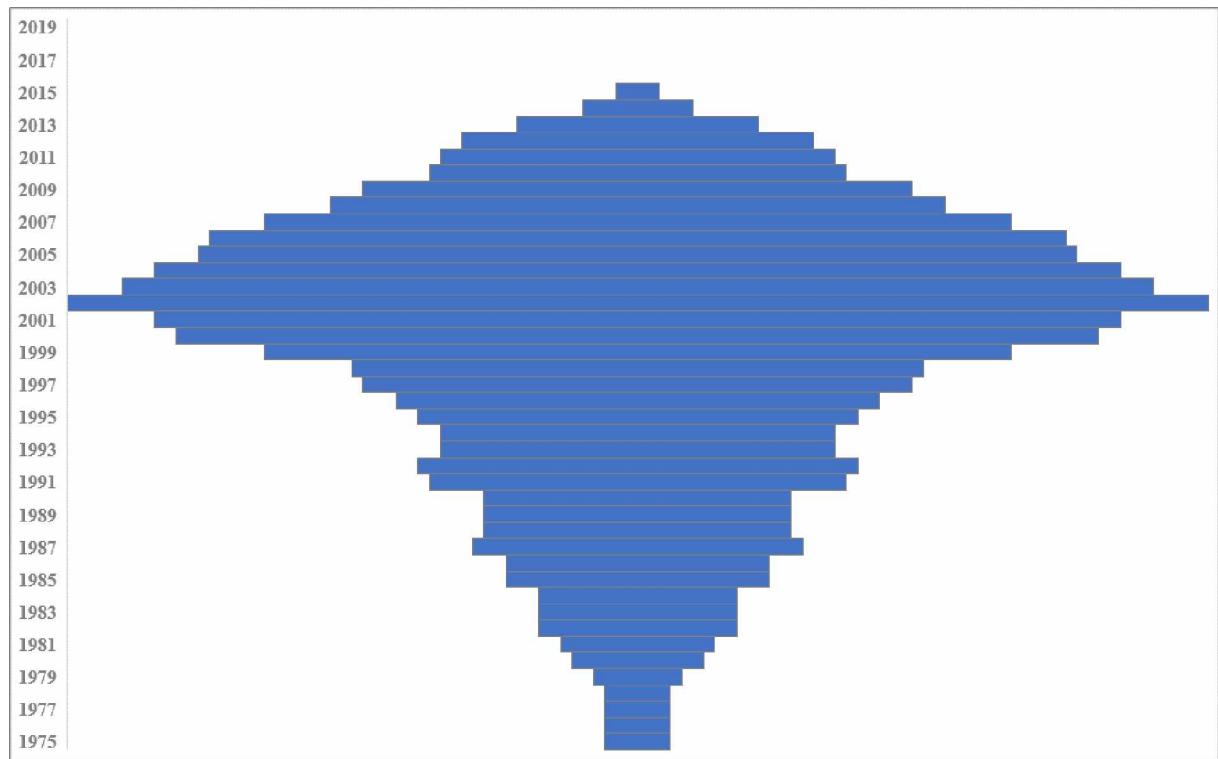
Figura 11 – Mapeamento de países por concentração de frequência de utilização do conjunto da base de dados



Fonte: resultados da pesquisa.

Na avaliação do horizonte de tempo de utilização dos conjuntos de dados que compõe o PB, figura 12, verifica-se que houve uma concentração de dados do período de 1999 a 2007, as bases mais recentes utilizadas foram apenas até o ano de 2015 e somente duas pesquisas chegaram até este período. A demanda para se atualizar os modelos, com conjuntos de dados maiores e mais atualizados, por um período de tempo maior poderia trazer implicações mais amplas em termos de resultados. Altman (2002) já demonstrava esta preocupação onde para o autor os principais fatores motivacionais técnicos incluíam o refinamento das técnicas tradicionais, inovações nas soluções analíticas, bases de dados maiores e melhoradas, para se traduzir a gestão do risco de crédito.

Figura 12 – Horizonte de tempo (*database*) utilizado no conjunto das bases de dados

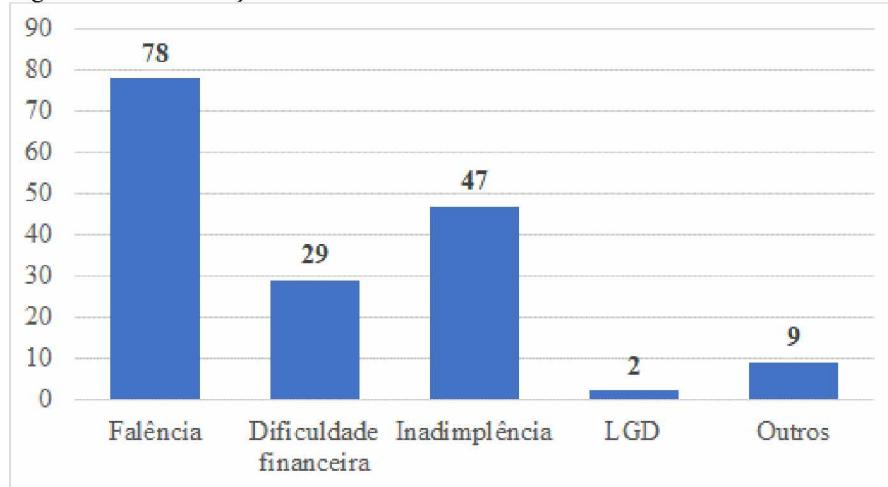


Fonte: resultados da pesquisa.

Para avaliar o desempenho dos modelos testados, a maioria dos pesquisadores utilizaram a precisão (ACC), a área sob a Curva ROC (AUC), os erros do tipo I e do tipo II e o *Brier Score*, erro quadrático médio da previsão (ALA'RAJ; ABBOD, 2016). Karan, Ulucan e Kaya (2013) utilizam as seguintes métricas de avaliação de desempenho: ACC, Error Type I, Error Type II, AUC, Gini, Kolmogorov-Smirnov (KS) e *Hand's H*. Estas medidas dispõem do desempenho preditivo do experimento proposto e serve de comparação entre os diversos modelos (ALA'RAJ; ABBOD, 2016; WANG et al., 2018).

Os estudos utilizam múltiplas variáveis quantitativas e qualitativas, para desenvolver os inúmeros modelos, no intuito de monitorar as ocorrências em andamento das organizações e detectar mudanças bruscas de comportamento em fatores que irão compor o risco. Baseado nos conjuntos de dados do PB desta pesquisa, as variáveis de interesse foram classificadas e agrupadas, e percebeu-se que a maior parte dos estudos (47%), procuram identificar o comportamento das empresas que entraram em falência em relação às empresas que não faliram. Já os modelos que criam score e classificam o risco de crédito foram 28% e 18% procuraram o identificar a dificuldade financeira das organizações antecipadamente.

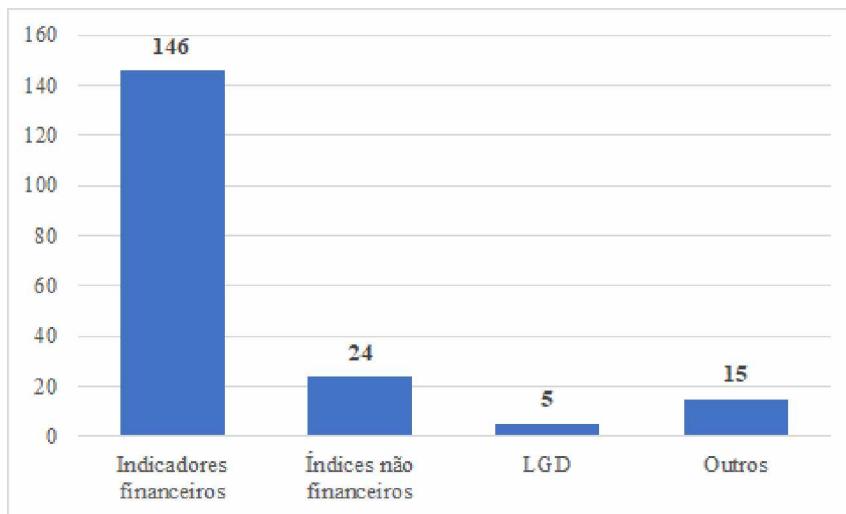
Figura 13 – Distribuição das variáveis de interesse



Fonte: resultado das pesquisas.

Ao se classificar as variáveis explicativas, a maioria dos estudos (77%) utilizam indicadores financeiros e contábeis, pela facilidade de apurar e trabalhar com os dados. As variáveis qualitativas além de terem certa subjetividade, ainda existe a dificuldade para coleta dos dados, devido ao intenso trabalho manual, principalmente dados de Governança Corporativa, em alguns países. Assim, essas variáveis foram identificadas em 13% dos estudos. Liang, Lu, Tsai e Shih (2016) e Jones (2017) utilizaram dados de Governança Corporativa em seus modelos como: Estrutura do Conselho, Estrutura de Propriedade, Reputação da Auditoria, Remuneração do CEO, entre outras.

Figura 14 – Distribuição das variáveis independentes



Fonte: resultado das pesquisas.

Algumas pesquisas não detalham a descrição completa das variáveis utilizadas nos experimentos ou não forneceram o tamanho da base de dados utilizada ou o período do conjunto de dados, o que dificulta ainda mais a compreensão dos modelos e métodos desenvolvidos (YU et al., 2011; ABELLÁN; MANTAS, 2014).

Neste PB, os métodos baseados em aprendizado de máquina e redes neurais, como SVM, DT, ANN e RF, conjuntos, individuais ou híbridos, tiveram melhor desempenho em relação a acurácia que os métodos estatísticos tradicionais. O método de LR se destacou entre os estatísticos tradicionais, assim na terceira etapa deste trabalho, comparou o desempenho dos métodos RF e LR, individuais, no mercado de capitais brasileiro e nas empresas da região de interesse.

2.9 Considerações Finais

Com o desenvolvimento nos sistemas de informação e a disseminação da globalização econômica, a forma de se conduzir o risco de crédito é uma questão de relevância no campo de contabilidade e finanças. O efeito causado pelo índice de insucessos empresariais pode ser comprometedor para investidores, acionistas, fornecedores, sociedade e de forma geral, para a economia de um país. Uma melhora na precisão da previsibilidade de dificuldades financeiras pode proporcionar maiores retornos e ainda minimizar seus efeitos negativos para todos os envolvidos.

Esta pesquisa foi desenvolvida a fim de avaliar as diversas pesquisas acadêmicas com as múltiplas visões sobre risco de crédito com enfoque em dificuldades financeiras, classificação de crédito, falência, para as quais usaram metodologias recentes e com perspectivas diferentes da estatística tradicional, tais como aprendizado de máquina e inteligência artificial e, sua evolução ao longo dos anos. Esse objetivo foi alcançado realizando uma revisão sistemática para identificar os diferentes construtos desse campo da literatura, explorando e descrevendo a utilização dos métodos e técnicas de análise que envolvem AI, e ainda apresentando as tendências das publicações, dos conjuntos de dados, periódicos, autores e co-autorias deste portfólio de pesquisas. Empregou-se a metodologia ProKnow-C, que resultou, como suporte empírico, na seleção de 165 referências científicas. Nesta seleção não houve limitação temporal.

Diante da averiguação final, pode-se observar que a aplicação da tecnologia computacional no âmbito da gestão do risco de crédito, vem angariando atenção de forma singular, principalmente após a crise financeira ocorrida em meados de 2007/2008. Nota-se que houve um aumento considerável de artigos publicados após esse período, o que pode ser explicado pela crise ocorrida, outra questão que também pode ser levantada é a melhoria na infraestrutura e evolução dos recursos computacionais, que são capazes de armazenar e transacionar uma vasta quantidade de dados, bem como o aperfeiçoamento dos métodos baseados em AI (MORO; CORTEZ; RITA, 2015).

Entretanto, apesar de todos os benefícios trazidos pela evolução dos modelos e técnicas baseados em AI e ANN, em relação a performance na precisão de predição, alguns autores discorrem sobre suas desvantagens e limitações. São citadas questões a serem aperfeiçoadas como a arquitetura utilizada onde, as decisões são baseadas em tentativas e erro, o que torna o processo demasiadamente lento em decorrência dos critérios e volume de dados. As ANN *feedforward* necessitam de reestruturação quando ocorrem ínfimas variações nos dados, outra questão a se discutir é a existência da grande dificuldade por parte dos usuários em interpretar a saída dos dados nos modelos de predição. Outra situação seria em relação as bases de dados, onde alguns autores alegam que os modelos não são capazes de tratar o desbalanceamento das bases de dados, podendo até comprometer sua performance. Estas limitações dificultam a aplicação desses modelos na prática tornando-os assim uma tecnologia de caixa preta (ZMIJEWSKI, 1984; WANG et al., 2011; OLSON; DELEN; MENG, 2012; ZHANG; WANG; JI, 2013; CHUANG, 2013). Seria interessante avaliar o nível de dependência entre as variáveis,

explicando o grau de influência de cada uma, dentro dos diversos modelos e técnicas, ou seja, implementar a perspectiva da análise fundamentalista aos modelos computacionais.

Os artigos mais citados são o de Tam e Kiang (1992), que foram os precursores na implementação dos métodos baseados em AI e ANN para prever falência e o de Altman, Marco e Varetto (1994), que também utilizou métodos baseados em ANN, mas neste caso, Altman foi o primeiro a aplicar a metodologia de análise discriminante para prever dificuldade financeira das firmas. Os dois artigos representam 11,9% do total de citações entre os 165 artigos do PB.

Os modelos apresentados por Altman (1968) e Ohlson (1980), LR, LDA e MDA, ainda são considerados relevantes pela capacidade preditiva, simplicidade e consistência, o que poucos modelos baseados em AI conseguiram apresentar (BARBOZA; KIMURA; ALTMAN, 2017). Dentre os 104 métodos e técnicas apresentados no PB analisado, estes três métodos ainda estão entre os 8 mais aplicados e comparados nos estudos, sugerindo que mesmo diante dos modelos baseados em AI, os métodos estatísticos tradicionais ainda estão presentes pelo entendimento e facilidade de aplicação.

Em relação ao horizonte de tempo de utilização dos conjuntos de dados, verificou-se que houve uma concentração de utilização do período de 1999 a 2007, as bases mais recentes utilizadas foram até o ano de 2015 e somente duas pesquisas chegaram até este período. A demanda para se atualizar os modelos, com conjuntos de dados maiores, mais atualizados e por um período de tempo maior, pode trazer implicações mais amplas em termos de desempenho, isso traz um reflexo positivo que pode ser aplicado nas organizações e nas instituições, melhorando a operação e aumentando a lucratividade.

Observou-se que em relação a origem dos conjuntos de dados, ou seja, os países que foram utilizados para os experimentos, a maioria dos trabalhos empregaram conjuntos de dados da Alemanha, Austrália, US, Japão, Taiwan, China, Coreia e França. A utilização destas bases para se tratar a gestão do risco de crédito, permite que outros pesquisadores possam comparar novos modelos e seus resultados. Entretanto, quando esses dados são utilizados exageradamente, podem não apresentar as condições específicas de outras regiões, mercados, ou segmentos, devido a questões regulatórias e as condições socio econômicas específicas, assim os experimentos podem levar a conclusões distorcidas e talvez não atenderem as demandas dos diversos mercados, pois os fatores que afetam a falência variam entre os diferentes países (BALCAEN; OOGHE, 2006; GARCÍA; MARQUÉS; SANCHÉZ, 2014; LIANG; TSAI; DAI; EBERLE, 2017; TIAN; YU, 2017). Assim, faz-se necessário acrescentar

pesquisas de outros países, como Brasil, países da América Latina, Israel, que, nesta base de dados do PB, não tiveram nenhuma pesquisa neste portfólio.

A presente pesquisa não pretende generalizar os resultados em relação ao tema tratado, nem tão pouco entender de forma pormenorizada toda produção envolvida nos estudos. Assim, este trabalho apresentou por meio de uma revisão sistemática da literatura diversas análises através de indicadores, mapas e redes, demonstrando o acompanhamento da evolução de uma determinada área do conhecimento, sua arquitetura, características, distribuição e relevância. Algumas limitações foram identificadas como a restrição das bases de dados para consulta e a formulação de outras análises, questões e objetivos, como a análise de desempenho por método, permitindo assim a identificação de novas oportunidades para pesquisas futuras.

CAPÍTULO III: PREVISÃO DE INSOLVÊNCIA NO MERCADO CORPORATIVO BRASILEIRO CONSIDERANDO A REGIONALIDADE.

O *Bank for International Settlements* (BIS) fornece as definições de risco de crédito, risco de liquidez, risco de mercado, risco operacional e outros tipos de risco. Este estudo abordou a gestão do risco de crédito que é definido como o risco de uma contraparte não conseguir cumprir plenamente suas obrigações financeiras quando vencidas ou a qualquer momento no futuro, sendo este considerado inadimplente potencial (BIS, 2012).

Os participantes do mercado de capitais são as grandes instituições e organizações, que possuem atividades significativas e operações a níveis globais. O estresse financeiro desses participantes pode ser considerado como um evento exógeno, desencadeado por condições não necessariamente relacionadas às suas operações, assim se constitui o risco sistemático (LI; PEREZ-SAIZ, 2018). Quando as organizações não conseguem liquidar suas obrigações por problemas de fluxo de caixa, a exposição ao risco de crédito aumenta, o que desencadeia graves efeitos para todo o sistema financeiro, trazendo altos custos individuais, sociais e econômicos (ZIEBA; TOMCZAK; TOMCZAK, 2013; ZHANG; WANG; JI, 2013; SUN; FUJITA; CHEN; LI, 2016).

Assim, a gestão do risco de crédito tem sido um importante tópico de pesquisas. Desde 1930, já existiam estudos focados no uso de análise de razão simples para se prever falências futuras (BELLOVARY; GIACOMINO; AKERS, 2007). Nas últimas décadas os estudos evoluíram de forma significativa, em finanças, na contabilidade e na sociedade em geral (ZIEBA; TOMCZAK; TOMCZAK, 2013; ZHANG; WANG; JI, 2013; LIN; LU; TSAI, 2018; GARCÍA; MARQUÉS; SANCHÉZ, 2019). As crises financeiras desencadeadas no mundo todo, fazem com que se torne necessária a inovação dos modelos estatísticos, com bancos de dados maiores, mais atualizados e com períodos de tempo, de fontes de análises, mais extensas (ALTMAN, 2002; LIN; LU; TSAI, 2018).

Tan (2012) analisou 277 empresas de oito economias do Leste Asiático, avaliando a relação entre dificuldades financeiras e desempenho durante a crise financeira asiática de 1997-1998. Os resultados encontrados neste estudo, foi que empresas com baixa alavancagem financeira tendem a ter um desempenho melhor do que as empresas com alta alavancagem. Além disso, a crise amplia a relação negativa entre dificuldades financeiras e desempenho e empresas com alta alavancagem apresentam desempenho pior durante uma crise.

Diversos estudos sobre análise e classificação do risco de crédito, previsão de dificuldades financeiras e falências foram realizados ao longo dos anos. Beaver (1966) aplicou um teste de classificação dicotômica, logo após, Altman (1968) trouxe um modelo de Análise Discriminante Multivariada (MDA) e empregou índices financeiros e contábeis como variáveis preditoras. Deakin (1972) revisou esses estudos, de Altman (1968) e Beaver (1966), utilizando uma função quadrática para construir um modelo de classificação mais preciso e, assim, aumentar a precisão para se prever falências. Em seguida, vieram outros modelos que foram desenvolvidos em diversos trabalhos como Regressão Logística (Martin, 1977; Ohlson, 1980; Mensah, 1984; Platt; Platt, 1990; Zhang; Hu; Eddy Patuwo; Indro, 1999; Li; Sun, 2010; Geng; Bose; Chen, 2015; Lin; Lu; Tsai, 2018) e *Probit* (ZMIJEWSKI, 1984; DIMITRAS; ZANAKIS; ZOPOUNIDIS 1996; JONES; JOHNSTONE; WILSON, 2015).

Com a evolução dos sistemas de informação e o processo de globalização, as relações entre os diversos negócios no mundo tornaram-se mais complexos e se expandiram, o que mudou o comportamento de relacionamento entre as organizações. As empresas cotadas nas bolsas de valores, geralmente operam em diversos países, independentemente da localização de sua matriz ou de fatores específicos de seu país de origem. Assim, surge a necessidade de se revisar os modelos, para que sejam mais precisos e rigorosos para tratar a gestão do risco de crédito nos diversos mercados. As pesquisas trouxeram alternativas como o *Z-score* (Altman, 2000) e os modelos baseados em Redes Neurais Artificiais (ANN) que apresentaram melhor desempenho e superaram os métodos estatísticos tradicionais (TAM; KIANG, 1992; ALTMAN; MARCO; VARETTO, 1994; WILSON; SHARDA, 1994; BROCKETT; COOPER; GOLDEN; PITAKTONG, 1994; JO; HAN, 1996; LEE; HAN; KWON, 1996; SERRANO-CINCA, 1996; VARETTO, 1998; SUNG; CHANG; LEE, 1999; ZHANG et al., 1999; ETHERIDGE; SRIRAM; HSU, 2000; entre outros).

Assim, ao longo dos anos, diversos modelos e técnicas baseados em AI, aprendizado de máquina e estatística foram desenvolvidos, aperfeiçoados e comparados entre si, para contribuir para a melhoria de precisão na previsibilidade de dificuldades financeiras, falência e na classificação do risco de crédito.

Entre eles, podem se destacar como os mais usados:

- a) *Support Vector Machine* (Huang; Chen; Hsu; Chen; Wu, 2004; Shin; Lee; Kim, 2005; Hung; Chen, 2009; Wang; Hao; Ma; Jiang, 2011; Hajek; Michalak, 2013;

Geng; Bose; Chen, 2015; Barboza; Kimura; Altman, 2017; Wang; Chen; Chu, 2018);

- b) Regressão Logística (Tam; Kiang, 1992; Zhang; Hu; Eddy Patuwo; Indro, 1999; Min; Lee; Han, 2006; Brown; Mues, 2012);
- c) *MultiLayer Perceptron* (Serrano-Cinca, 1996; Tsai; Wu, 2008; Wang; Ma; Huang; Xu, 2012; Liang; Tsai; Dai; Eberle, 2017);
- d) *Decision Tree* (Tam; Kiang, 1992; Sung; Chang; Lee, 1999; Hung; Chen, 2009; Kim; Kang, 2010; Du Jardin, 2017; Sun; Lang; Fujita; Li, 2018) e,
- e) *Random Forest* (CHANDRA; RAVI; BOSE, 2009; WANG; MA; HUANG; XU, 2012; ALA'RAJ; ABBOD, 2016; BARBOZA; KIMURA; ALTMAN, 2017; GARCÍA; MARQUÉS; SÁNCHEZ, 2019).

Contudo, apesar dos benefícios trazidos por estes modelos baseados em AI, em relação a performance na precisão de predição, alguns autores discorrem sobre suas desvantagens e limitações, dentre elas pode-se destacar: a arquitetura utilizada onde, as decisões são baseadas em tentativas e erro tornando processo demasiadamente lento; as ANN *feedforward* necessitam de reestruturação em decorrência de pequenas variações nos dados; existe uma grande dificuldade por parte do usuário em interpretar a saída dos dados nos modelos de predição. Estas limitações dificultam a aplicação dos modelos, tornando-os assim uma tecnologia de caixa preta (OLSON; DELEN; MENG, 2012; ZHANG; WANG; JI, 2013; CHUANG, 2013; DU JARDIM, 2016). Os modelos apresentados por Altman (1968) e Ohlson (1980) ainda são considerados relevantes pela precisão de predição, simplicidade e interpretabilidade, o que poucos modelos conseguiram apresentar (JONES; JOHNSTONE; WILSON, 2015; BARBOZA; KIMURA; ALTMAN, 2017).

Lessmann, Baesens, Seow e Thomas (2015) discorrem que os avanços computacionais permitem e facilitam a comparação entre os diversos modelos que são desenvolvidos, inclusive para implementação e aceitação pelas organizações. Os autores alegam ainda que os classificadores avançados não requerem intervenção manual e por isso são mais precisos.

Este capítulo da dissertação tem como objetivo avaliar e comparar dois modelos, utilizando mecanismos de aprendizado de máquina e estatística tradicional para prever o fenômeno de insolvência em empresas de capital aberto no Brasil e em empresas regionais (mesorregião do Triângulo Mineiro e Alto Paranaíba e do Sul Goiano) entre o período de 2005 a 2018.

Complementarmente, é avaliado se os modelos de aprendizado de máquina baseado na técnica *Random Forest* (RF) e de Regressão Logística agregam e melhoram a gestão do risco corporativo em empresas do mercado de capitais brasileiro; ainda, é feita também uma comparação entre os dois modelos baseados nas variáveis selecionadas para este estudo, qual oferece maior precisão e melhora a capacidade preditiva de dificuldades financeiras e falência; são ressaltados os fatores e variáveis que integram o risco diante da incerteza do fenômeno de dificuldades financeiras e na previsão de insolvência para o mercado brasileiro e também são apresentadas evidências para utilização de conceitos teóricos sobre os aspectos de risco de crédito e ademais, assim como analisa-se o comportamento dos modelos em empresas da região de interesse (mesorregião do Triângulo Mineiro e Alto Paranaíba e do Sul Goiano).

Neste capítulo da dissertação, o desempenho e a precisão de previsão de falências e dificuldades financeiras no mercado de capitais brasileiro entre os métodos Regressão Logística e *Random Forest*, obtiveram de precisão no cenário nacional 93,4% e 96,7%, e no cenário regional 90,9% e 96,3%, respectivamente. Esses achados contribuem para a literatura sobre a predição do risco de crédito e reforça a significante melhora dos resultados obtidos por meio dos mecanismos e técnicas de aprendizado de máquina em relação ao modelo estatístico tradicional *Logit*.

Os resultados corroboram com os estudos de Brockett et al. (1994), Zhang et al. (1999), Cheng; Chen; Fu (2006); Chandra, Ravi e Bose (2009), Hsieh, Hsiao, Yeh (2012), Jones, Johnstone e Wilson (2016) e Barboza, Kimura e Altman (2017). Apesar destas pesquisas utilizarem outras técnicas e modelos, mas fizeram também uma comparação entre estatísticas tradicionais e inteligência artificial, utilizaram conjuntos de dados de outros países, e não fizeram comparação com modelos regionais e nem análise das variáveis mais relevantes, como foi feito neste trabalho.

Apesar de existirem diversos estudos nacionais e internacionais sobre risco de crédito, dificuldade financeira e falência, existe a necessidade constante da busca por posições que possam tratar melhor a mitigação do risco, de forma integrada e sistematizada a partir das pesquisas existentes sobre este tema.

Os resultados encontrados são relevantes para o Brasil e outras economias emergentes, e até no aspecto regional, principalmente para aqueles com o mesmo ou maior nível de perdas financeiras. Nos últimos anos, principalmente após eventos de crises financeiras, ocorreu um aumento no volume de pesquisas relacionadas dificuldades financeiras e de modelos baseados

em AI, conforme apresentado no Capítulo II. Assim, este trabalho ainda pode ser expandido e também pode auxiliar outros estudos sobre este tema em mercados de economias emergentes. Espera-se que os resultados obtidos nesta pesquisa, bem como as técnicas e métodos empregados, possam enriquecer a ciência e ainda contribuir para novas alternativas de pesquisas futuras.

3.1 Referencial Teórico

Diante das crises financeiras mundiais, o tema sobre risco de crédito se desenvolveu rapidamente nos últimos anos e se tornou um ponto importante nos sistemas de gerenciamento das instituições financeiras e das organizações. A previsão de risco de crédito, dificuldades financeiras e falências corporativas vem sendo amplamente estudado nas áreas de contabilidade, finanças e sistemas da informação (ZHANG; WANG; JI, 2013; LIN; LU; TSAI, 2018; GARCÍA; MARQUÉS; SANCHÉZ, 2019).

ALTMAN, DANOVY E FALINI (2013) discorrem que a firma atravessa por fases de sucesso e fracasso em algum momento em seu ciclo de vida, e isso trata-se de um processo natural. O problema ocorre quando o período negativo se torna irreversível e o desfecho pode resultar em falência ou recuperação judicial. Esse grau de incerteza em relação ao momento exato em que ocorre o fracasso, fez com que a busca por métodos e técnicas, com dados quantitativos e qualitativos, tentassem prever o instante em que se ocorre a falência. Os autores aplicaram o método Z-Score, para dimensionar zona de incerteza ou área cinzenta, onde o horizonte futuro de continuidade da firma é duvidoso. Utilizando dados de empresas italianas, os autores identificaram que a tratativa dessas empresas, contidas na área cinzenta, é extremamente informativa no processo decisório.

Beaver (1966), foi um dos precursores a pesquisar a previsão de falências, investigou a capacidade preditiva utilizando índices financeiros. Em seguida Altman (1968) trouxe um modelo de MDA, depois Ohlson (1980) com o modelo de LR. Gombola e Ketz (1983) ressaltam sobre os padrões de classificação dos índices financeiros, inclusive sobre a relevância na consideração do fluxo de caixa operacional como uma dimensão de desempenho da firma. Casey e Bartczak (1985) também utilizaram a LR, MDA e *Probit* para previsão de falência e criticaram Gombola e Ketz (1983) sobre a utilização do fluxo de caixa operacional na previsão de falência.

Tam e Kiang (1992) iniciaram com uma abordagem de ANN e compararam com outros quatro modelos distintos, como ANN, MDA, *Logit* e DT. Os resultados empíricos destes trabalhos foram o marco para a busca constante por melhorias nos modelos e técnicas de avaliação das condições financeiras das organizações em termos de precisão preditiva (ZIEBA; TOMCZAK; TOMCZAK, 2016). Nesta mesma época, Salchenberger, Cinar e Lash (1992), compararam os modelos ANN e LR para prever dificuldades financeiras com índices financeiros e dados de instituições financeiras, no período de 1986 a 1987 e identificaram que a ANN exige menos premissas, atinge um grau mais alto de precisão de previsão e são mais robustas.

Os modelos baseados em ML e AI não assumem conhecimento preliminar, eles extraem de forma automática as informações de eventos passados, que são representados por um conjunto de variáveis explicativas, criando um comportamento baseado em indicadores financeiros ou macroeconômicos, ou indicadores qualitativos baseados em Governança Corporativa, ou condições socio geográficas, que são constituídos como variáveis contínuas ou ponderadas como ocorre em informações qualitativas (JONES; JOHNSTONE; WILSON, 2015; LIANG et al., 2016; JONE, 2017; GARCÍA et al., 2019).

Zieba, Tomczak e Tomczak (2016) trazem uma abordagem para a previsão de falências que utiliza *Extreme Gradient Boosting* para treinamento de *Decision Trees* e desenvolveram um recurso sintético que combina as medidas econométricas utilizando operações aritméticas. Os autores criticam os modelos lineares, como *Logit* e *Probit*, diante dos volumes crescentes das bases de dados, alegando que são incapazes de refletir os relacionamentos entre as métricas econômicas. Jones, Johnstone e Wilson (2016) comentam que os classificadores mais simples como *Logit* e LDA funcionam bem para a previsão de falências, mas os modelos mais avançados baseados em AI são mais recomendados pois predizem melhor, são fáceis de estimar e implementar, contam com pouca intervenção manual e são razoavelmente interpretáveis, que pode ser realizado por meio de métricas como o nível de relevância das variáveis.

Barboza, Kimura e Altman (2017) usaram modelos de aprendizado de máquina (SVM, BA, BOO e RF) para prever falência e compararam os resultados com as técnicas de LDA, LR e ANN. Os autores utilizaram dados de 1985 a 2013 de empresas norte-americanas, integrando informações do banco de dados do *Salomon Center* e *Compustat*. Eles identificaram que as técnicas de ML trazem uma melhora significativa na precisão da previsão de falências. No estudo a técnica de RF foi a que obteve o melhor desempenho com precisão de 87%, enquanto

a LR e a LDA tiveram 69% e 50%, respectivamente. Os autores criticaram o método SVM, que não foi superior aos outros mecanismos de aprendizado de máquina analisados.

Bekhet e Eletter (2014) propõe dois modelos para scores de crédito usando técnicas de mineração de dados (*data mining*) para apoiar as decisões de empréstimo para os bancos da Jordânia no período de 2006 a 2011. No estudo, os autores demonstraram que o modelo de Regressão Logística obteve melhor taxa de classificação do que o modelo *Radial Basis Function* (RBF), mas em relação a identificação dos clientes que podem entrar em *default* o RBF teve melhor resultado.

Brown e Mues (2012) usaram LR, LDA, ANN, DT, SVM e RF para previsão de inadimplência com base em cinco conjunto de dados de instituições financeiras da região de Benelux, Alemanha e Austrália. Os resultados do estudo indicaram que RF e o *Gradient Boosting* tiveram bom desempenho no contexto de pontuação de crédito e foram capazes de lidar comparativamente bem com conjunto de dados desbalanceados. Os autores ressaltaram que os métodos tradicionais como LDA e LR, deram resultados razoavelmente competitivos com as técnicas mais complexas, mesmo quando as amostras se tornaram muito mais desequilibradas, sugerindo que essas abordagens são bastante robustas também em classes de dados desequilibradas.

No mercado brasileiro, Scalzer et al. (2019) utilizaram a LR com dados de 60 distribuidores de energia no período de 2009 a 2015, para investigar quais indicadores financeiros poderiam prever dificuldades financeiras em relação às metas impostas pelo órgão regulador e identificaram que o retorno sobre ativos, a liquidez corrente, e a liquidez imediata se destacaram em seu poder preditivo. Em relação a definição das variáveis, Becker et al. (1998) discorrem sobre a relevância de se considerar aquela mais assertiva para o modelo, e que uma única variável pode substituir inúmeras outras que possam estar omitidas.

A tabela 7 apresenta as pesquisas que aplicaram as técnicas de *Random Forest* e Regressão Logística, nas últimas décadas, de forma individual ou conjunta, ainda comparando os com outros modelos e seus respectivos conjuntos de bases para pesquisa.

Tabela 7 – Evolução das pesquisas que utilizaram as técnicas RF e LR

Autores	Cit.	Contexto	Técnicas / Métodos	Horizonte de tempo (base)	País de origem do conjunto de dados
Tam; Kiang (1992)	1444	Previsão de falência	LDA, LR, KNN, ANN, DTID3	1985-1987	USA

Brockett et al. (1994)	196	Dificuldade financeira	LR, ANN, MDA, O	1987-1990	USA
Zhang et al. (1999)	691	Previsão de falência	LR, ANN, CVNN		Outros
Beynon; Peel (2001)	305	Inadimplência empresarial	RST, VPRS, RPA, LR, EM, MDA		Reino Unido
Lin; McClean (2001)	176	Inadimplência empresarial	LDA, LR, ANN, DTC5.0	1980-1999	Reino Unido
Huang et al. (2004)	949	Pontuação de crédito	LR, SVM, ANN, BPNN	1998-2002	USA, Taiwan
Liu; Schumann (2005)	90	Pontuação de crédito	KNN, LR, NN, MLP, DTM5		Alemanha
Yongqiao Wang; Lai (2005)	361	Pontuação de crédito	LR, ANN, FSVM, RBF		Reino Unido, Japão
Min; Lee (2005)	759	Previsão de falência	ANN, SVM, BPNN, LR, MDA	2000-2002	Coreia
Ko; Lin (2006)	49	Dificuldade financeira	GA, PSO, ANN, LR, LDA	1993-2002	Taiwan
Min; Lee; Han (2006)	381	Previsão de falência	SVM, ANN, LR, GA, CBR	1999-2000	Coreia
Hu; Ansell (2007)	78	Pontuação de crédito	LR, ANN, SMO, NB	1994-2002	USA
Cheng; Chen; Fu (2006)	80	Dificuldade financeira	LR, RBFN, BPNN	1996-2004	Taiwan
Florez-Lopez (2007)	39	Pontuação de crédito	CART, DTC4.5, MDA, LR	1999-2000	Reino Unido, Alemanha, França, Itália, Espanha, Países Baixos, Outros
Hu; Ansell (2007)	78	Pontuação de Crédito	LR, ANN, SMO, NB	1994-2002	USA
Wu et al. (2007)	373	Previsão de falência	SVM, GA, MDA, LR, PM, ANN	2001-2003	Taiwan
Yu; Wang; Lai (2008)	253	Pontuação de crédito	LR, ANN, SVM, FSVM, VB, RB		Reino Unido
Liou (2008)	82	Inadimplência empresarial	LR, ANN, DT	2003-2004	Taiwan
Chandra; Ravi; Bose (2009)	56	Inadimplência empresarial	MLT, CART, RF, LR, SVM, DT, BOO		Outros
Hu (2009)	41	Previsão de falência	LDA, LR, SLP, MLP, SVM, PM, FIFLN, ELECTRE	1975-1982	Outros
Zhao; Sinha; Ge (2009)	101	Previsão de falência	LR, DTC4.5, ANN, KNN	1991-1992	Outros
Härdle et al. (2009)	72	Pontuação de crédito	LR, LDA, SVM	1997-2002	Alemanha
Cho; Kim; Bae (2009)	77	Previsão de falência	MDA, LR, ML, ANN, DT, SVM, RULE, IMSW	1999-2002	Coreia
Li; Sun; Sun (2009)	85	Dificuldade financeira e previsão de falência	MDA, LR, ANN, SVM, CBR	2000-2005	China
Min; Jeong (2009)	94	Previsão de falência	MDA, LR, DT, ANN, BC	2001-2004	Outros

Twala (2010)	134	Pontuação de crédito	ANN, DT, KNN, LR, BA, BOO, NB	1985-1987	Alemanha, Japão, Austrália
Li; Sun (2010)	54	Dificuldade financeira e previsão de falência	CBR, LR, MDA, KNN	2000-2005	China
Tsai; Chen (2010)	99	Pontuação de crédito	DTC4.5, NB, LR, ANN, KM	2004-2006	Taiwan
Chen (2011)	130	Dificuldade financeira	LR, DT, CART		Taiwan
Chaudhuri; De (2011)	133	Previsão de falência	LR, LDA, PNN, FSVM	2000-2001	USA
Li; Sun (2011)	48	Dificuldade financeira e previsão de falência	CBR, MDA, LR	2000-2005	China
Lin; Liang; Chen (2011)	72	Dificuldade financeira	SVM, LR, MDA, RBFN	2000-2008	Taiwan
Finlay (2011)	121	Pontuação de crédito	LR, LDA, CART, ANN, KNN, AB, BA, BOO	2002	Reino Unido
Wang et al. (2011)	249	Pontuação de crédito	BA, BOO, LR, DT, ANN, SVM	2006-2007	Alemanha, Austrália, China
Wang; Ma (2011)	37	Pontuação de crédito	LR, DTC4.5, ANN, BA, BOO, RS	2006-2007	China
Tsai; Cheng (2012)	41	Previsão de falência	MLP, DTC4.5, LR, SVM		Alemanha, Japão, Austrália, Outros
Koyuncugil; Ozgulbas (2012)	130	Dificuldade financeira	CHAID-EWS-LR-MDA-SOM-CART	2007	Turquia
Brown; Mues (2012)	303	Pontuação de crédito	LR, ANN, DT, SVM, RF, LDA, GBOO		BENELUX
Marqués; García; Sánchez (2012)	67	Pontuação de crédito	AB, BA, RS, RF, ANN, LR, MLP, SVM, DTC4.5		USA, Alemanha, Polônia, Japão, Irã, Austrália
Wang et al. (2012)	127	Pontuação de crédito	RS, BA, DT, LR, LDA, MLP, RBFN, RF, RTF		Alemanha, Austrália
Kim; Ahn (2012)	88	Pontuação de crédito	SVM, MDA, LR, CBR, ANN		Coreia
Marqués; García; Sánchez (2012)	68	Pontuação de crédito	BA, RS, DECORATE, AB, RTF, NN, NB, LR, MLP, RBF, ANN, SVM, DTC4.5		Alemanha, Japão, Austrália, Irã
Du Jardin; Séverin (2012)	39	Previsão de falência	MDA, LR, ANN, COX	1991-2009	França
Li; Sun (2012)	54	Dificuldade financeira e previsão de falência	SVM, ANN, MDA, LR, KNN	1998-2010	China
Hsieh; Hsiao; Yeh (2012)	40	Pontuação de crédito	LR, ANN, SVM, EABC, BPNN, GA	1999-2007	Taiwan
Öğüt et al. (2012)	56	Pontuação de crédito	LR, MDA, SVM, PNN	2003-2009	Turquia
Olson; Delen; Meng (2012)	190	Previsão de falência	LR, RBF, DTC5.0, CART, SVM, DTJ8	2005-2009	USA
Wang; Ma (2012)	86	Pontuação de crédito	RSB, SVM, ANN, BA, RS, BOO, LR, DTC4.5, MLP, AB, SMO	2006-2007	China

Marqués; García; Sánchez (2013)	69	Pontuação de crédito	SVM, LR	USA, Alemanha, Polônia, Japão, Irã, Austrália	
Hajek; Michalak (2013)	62	Pontuação de crédito	MLP, RBF, SVM, NB, RF, LDA, NMC, ANN	USA, Reino Unido, Alemanha, França, Outros	
Serrano-Cinca; Gutiérrez-Nieto (2013)	69	Dificuldade financeira	LDA, LR, MLP, KNN, NB, SVM, BOO, DTC4.5, BRT, PLSDA	USA	
Zhou (2013)	87	Previsão de falência	LDA, LR, DT, ANN, SVM	1981-2009	USA, Japão
Chuang (2013)	47	Inadimplência empresarial	CBR, RST, CART, GRA, LR	1999-2006	Taiwan
Fedorova; Gilenko; Dovzhenko (2013)	65	Previsão de falência	MDA, LR, CART, ANN, AB	2007-2011	Rússia
Lin; Chang; Hsu (2013)	48	Dificuldade financeira	RF, BPNN, LDA, ML, O	2009-2011	Taiwan, Polônia, Alemanha, Austrália
Tsai (2014)	80	Previsão de falência	MLP, LR, CART, SOM, KM	Alemanha, Japão, Austrália, Outros	
Wang; Ma; Yang (2014)	69	Previsão de falência	LR, NB, DTC4.5, ANN, SVM, BA, BOO, MLP, SMO	Outros	
Wu; Olson; Luo (2014)	26	Pontuação de crédito	LR, DT, ANN	Outros	
Kou; Peng; Lu (2014)	25	Inadimplência empresarial	NB, LR, KNN, DTC4.5, CART, MCDM	2002-2004	China
Bekhet; Eletter (2014)	63	Pontuação de crédito	LR, RBF	2006-2011	Jordânia
Hajek; Olej; Myskova (2014)	29	Dificuldade financeira	SVM, ANN, DT, NB, LR, RBF, MLP, SMO	2008	USA
Lessmann et al. (2015)	237	Pontuação de crédito	QDA, SVM, NB, LR, KNN, ELM, CART, ANN, RF, BOO, BA, KM, O	Alemanha, Austrália, Reino Unido	
Koutanaei; Sajedi; Khanbabaei (2015)	59	Pontuação de crédito	ANN, SVM, CART, NB, AB, BA, RF	Irã	
Jones; Johnstone; Wilson (2015)	54	Pontuação de crédito	AB, BOO, RF, ANN, SVM, LR, LDA, QDA, PROBIT	1983-2013	USA
Danenas; Garsva (2015)	64	Previsão de falência	SVM, LR, ANN, RBF	1999-2007	USA
Geng; Bose; Chen (2015)	122	Dificuldade financeira	ANN, DTC5.0, SVM, LR, LDA, MV, DM	2001-2008	China
Jones; Johnstone; Wilson (2016)	31	Previsão de falência	AB, BOO, RF, ANN, SVM, LR, LDA, QDA	USA	
Ala'raj; Abbad (2016)	62	Pontuação de crédito	MCS, LR, ANN, SVM, RF, DT, NB	Polônia, Alemanha, Japão, Irã, Austrália	
Ala'raj; Abbad (2016)	39	Pontuação de crédito	GNG, MARS, CA, ANN, SVM, RF, DT, NB	USA, Polônia, Alemanha, Jordânia, Japão, Irã, Austrália	

Du Jardin (2016)	30	Previsão de falência	MDA, LR, ANN, CART, BA, BOO, RS, PBM, SOM	2003-2012	França
Zięba; Tomczak; Tomczak (2016)	74	Previsão de falência	AB, AC, LDA, MLP, LR, DTC4.5, RF, GB	2007-2013	Polônia
Chang; Hsu (2016)	1	Dificuldade financeira	SVM, DEA, IFWTSVM, ANN, RF	2012-2014	Taiwan
Zelenkov; Fedorova; Chekhrizov (2017)	13	Previsão de falência	KNN, LR, NB, DT, SVM, ADB, RF, BA, QDA, ZS, MLP, GBOO, TSCM		Rússia
Abellán; Castellano (2017)	45	Previsão de falência e Pontuação de crédito	AB, BA, RS, DECORATE, RTF, LR, DTC4.5, MLP, SVM, CDT		USA, Polônia, Alemanha, Japão, Irã, Austrália
Bequé; Lessmann (2017)	17	Pontuação de crédito	ELM, ANN, LR		Alemanha, Austrália
Barboza; Kimura; Altman (2017)	57	Previsão de falência	MDA, LR, ML, SVM, BA, RF, BOO, ANN	1985-2013	USA
Du Jardin (2017)	29	Previsão de falência	MDA, LR, SVM, FNN, DTC4.5, ELM, RS, RTF, DECORATE, BA, BOO, AB, COX, SOM	1996-2001	França
Antunes; Ribeiro; Pereira (2017)	14	Previsão de falência	GP, SVM, LR	2006-2007	França
Lin; Lu; Tsai (2018)	2	Previsão de falência	LR, NB, BPNN, DTC4.5, SVM, KNN, BA, BOO, GA		Taiwan, Alemanha, Austrália
Cheng; Jones; Moser (2018)	0	Previsão de falência	LR, GBOO	1992-2012	USA
Du Jardin (2018)	3	Previsão de falência	MDA, LR, SVM, FNN, CART, ELM, RS, RTF, BA, BOO, AB, COX, SOM	2006-2014	França
García et al. (2019)	2	Previsão de falência	FLD, LDA, SVM, LR		Coreia
García; Marqués; Sánchez (2019)	1	Previsão de falência e Pontuação de crédito	RF, RTF, DECORATE, AB, BOO, BA, RS		Finlândia, Polônia, Japão, Alemanha, China, Austrália, Irã

Fonte: resultados da pesquisa.

3.2 As técnicas de referência para o estudo

O objetivo deste trabalho é avaliar e comparar dois modelos utilizando mecanismos de aprendizado de máquina, *Random Forest*, e estatística tradicional, Regressão Logística para prever o fenômeno de insolvência em empresas de capital aberto no Brasil. A fim de investigar a questão da regionalidade nesse contexto, de forma inovadora, é também avaliado tal modelagem aplicada a uma base de dados da região de interesse, isto é, se a adoção destes modelos agrupa e melhora a gestão do risco de crédito, se oferecem maior precisão e melhoram a capacidade preditiva de falências, e também são comparados os desempenhos dos dois

métodos. Estes métodos foram selecionados pois, conforme analisado nas pesquisas apresentadas no capítulo II desta dissertação, são os métodos mais empregados individualmente e também se utilizou como artigo base os estudos de Barboza, Kimura e Altman (2017).

A Regressão Logística (LR) é uma das técnicas estatísticas tradicionais mais utilizadas para previsão de dificuldades financeiras das organizações. Suas classes são definidas de forma binária onde a empresa é categorizada como em dificuldades financeiras, falida ou com recuperação judicial deferida ou não se encontra nesta situação (TAM; KIANG, 1992; ZHANG et al., 1999; MIN; LEE, 2005; DANENAS; GARSVA, 2015; BARBOZA; KIMURA; ALTMAN, 2017; DU JARDIN, 2018). A LR é um modelo de regressão em que as variáveis dependentes são distribuídas de forma binomial, assumindo dois valores possíveis, neste estudo, assume 1 se a empresa estiver falida ou em recuperação judicial ou em dificuldades financeiras ou 0 se ela não se encontra nesta situação.

Neste caso, x é um vetor de coluna das variáveis explicativas e $\pi(x) = P(y = 1 | x)$ é a probabilidade de resposta a ser modelada. O modelo de LR é apresentado pela seguinte equação:

$$\text{logit}(\pi) = \ln\left(\frac{\pi}{1 - \pi}\right) = \alpha + \beta^T x \quad (1)$$

Onde, α é o parâmetro de intercepção e β^T contém os coeficientes da variável (OHLSON, 1980).

A LR adota uma estimativa de probabilidade após transformar a variável dependente em uma variável *logit*, ou seja, calcula-se o logaritmo natural das probabilidades para se prever a ocorrência de um determinado evento. Huang, Tzeng e Ong (2006) discorrem que o modelo de Regressão Logística (LR) é uma das ferramentas estatísticas mais utilizadas para tratar problemas de classificação. Os autores sustentam que este modelo ajusta vários tipos de funções de distribuição como *Gamble*, *Poisson*, distribuições normais, é o mais adequado para tratar problemas de inadimplência e ainda, aumenta a precisão.

Huang et al. (2004) discorrem que as técnicas de aprendizado de máquina extraem matematicamente conhecimento de um conjunto de dados e constroem estruturas de modelos diferentes a partir do comportamento das variáveis na base de dados. Os autores alegam que a principal diferença entre os modelos estatísticos tradicionais e métodos de aprendizado de máquina é que os métodos estatísticos geralmente necessitam da intervenção humana para impor estruturas a diferentes modelos, como a linearidade na análise de regressão múltipla, e para construir o modelo estimando, os pressupostos devem ser testados para adequação dos dados e das observações.

O modelo *Randon Forest* (RF) é considerado uma técnica avançada de *Decision Tree* (DT), como proposto por Breiman (2001), que consiste em conjuntos de árvores de classificação ou regressão que são treinadas com amostras de dados. Este método se baseia numa seleção de variáveis de forma aleatória, enquanto o processo de geração da floresta acontece. Depois que uma grande variedade de árvores é gerada, as quais fornecem um voto (cada) para a decisão em qual classe a observação é identificada. O total de árvores apresenta a votação na classe mais popular, ou seja, no final, uma votação majoritária é feita para previsão (LIAW; WIENER, 2002). Estes procedimentos de votação em floresta são definidos coletivamente como *Randon Forest*.

Esse algoritmo é uma ferramenta de classificação que constitui numa seleção de classificadores estruturados em florestas $\{h(x, \theta_k), k = 1, \dots\}$ onde $\{\theta_k\}$ são vetores aleatórios distribuídos independentes e cada árvore lança um voto unitário para as demais classes de variáveis x . Esse algoritmo agrupa uma aleatoriedade extra ao modelo, pois, ao invés de examinar pela melhor característica dos dados ao fazer a partição dos nós, ele procura a melhor característica em um subconjunto aleatório (BREIMAN, 2001).

No modelo RF não foram definidos os parâmetros de quantidade de árvores e de atributos usados para criar cada floresta, assim se manteve o padrão do algoritmo que são 500 árvores. Em relação as variáveis o valor configurado é de um terço do número total de variáveis preditoras (LIAW; WIENER, 2002).

3.3 Variáveis abordadas no estudo

A análise contemplou empresas em situação de falência e com pedido de recuperação judicial deferido conforme a legislação vigente considerando as premissas do estudo de Barboza, Kimura e Altman (2017). Para as empresas em dificuldades financeiras, o conceito foi adaptado do estudo de Manzaneque, Priego e Merino (2016), neste sentido a variável dependente leva valor 1 quando a empresa se encontra em dificuldades financeiras ou falência ou recuperação judicial deferida e 0, caso contrário. As empresas foram consideradas em dificuldades financeiras ou falência ou recuperação judicial nas seguintes condições:

- (a) quando a companhia se encontra em situação de falência decretada ou teve o pedido de recuperação judicial deferido; e / ou,
- (b) quando a geração operacional de caixa bruta medida pelo *Earnings Before Interest, Taxes, Depreciation and Amortization* (EBITDA), ou seja, a geração de recursos

considerando apenas as atividades operacionais, desconsiderando os efeitos financeiros e de impostos, é menor do que as despesas financeiras por dois anos consecutivos; e / ou,

- (c) quando empresa se encontre com a situação patrimonial extrema, ou seja, o Patrimônio Líquido da entidade seja negativo por dois períodos consecutivos.

Os apontamentos foram considerados um ano antes do fenômeno em si, ou seja, se a data de anúncio da falência ocorrer em até seis meses após o fechamento do ano fiscal, os dados coletados são apurados um ano antes do fato. Os dados financeiros relativos as empresas que compõe este estudo, foram extraídos da base de dados Economática, no período de 2005 a 2018, este período foi definido em decorrência da Lei de Recuperação Judicial e Extrajudicial e de Falência nº 11.101 de 9 de fevereiro de 2005.

A maior parte dos estudos relacionados ao tema central desta pesquisa, conforme revisão de literatura realizada no capítulo II desta dissertação, utilizaram bases de dados com períodos médios de 7 anos, a base de dados aplicada para este estudo será de 14 anos. As empresas do setor financeiro foram excluídas da amostra devido às características contábeis e setoriais específicas do próprio segmento. As empresas que não possuíam informações foram extraídas da base de dados (395 empresas).

Dessa forma, no total, foram utilizadas 15 variáveis independentes, sendo todas elas disponíveis na base de dados Economática. As variáveis foram escolhidas a partir do portfólio final de estudos selecionados pela revisão sistemática de literatura realizada no Capítulo II desta dissertação. Em relação as contas que tiveram sua uniformização em decorrência das Normas Internacionais de Contabilidade (IFRS), emitidas pelo *International Accounting Standards Board* (IASB), foram tratadas conforme base de dados Economática e o Ativo Imobilizado foi agregado ao Ativo não Circulante.

No quadro 1 foram relacionadas as variáveis explicativas do modelo conforme a literatura sobre risco de crédito. Os indicadores utilizados no estudo compreendem índices de liquidez, de rentabilidade, de atividade, de estrutura, geração de caixa e valor.

Quadro 1 - Variáveis explicativas do modelo

Cod.	Variáveis	Fórmula	Autor
X ₁	CALO	(Ativo Circulante - Estoques - Recebíveis) / (Passivo Circulante-Fornecedores)	Salchenberger; Cinar; Lash (1992), adaptado de Zieba; Tomczak; Tomczak (2016)
X ₂	L	Alavancagem: (Passivo Circulante + Passivo não Circulante) / Ativo Total	Cheng; Chen; Fu (2006); Tan (2012); Hajek; Michalak (2013); Korol (2013); Wang; Chen; Chu (2018)
X ₃	TTA	(Ativo Total _t – Ativo Total _{t-1}) / Ativo Total _{t-1}	Barboza; Kimura; Altman (2017); Ribeiro; Chen; Kovacecc (2017)
X ₄	CPB	CPB Price-to-Book _t – Price-to-Book _{t-1} tal que P/B = Valor de Mercado da ação / Valor contábil da ação	Barboza; Kimura; Altman (2017)
X ₅	AT	Giro de Ativos: Receita Líquida / Ativo Total	Zieba; Tomczak; Tomczak (2016); Chang; Hsu (2016); Barboza; Kimura; Altman (2017)
X ₆	CL	Liquidez Corrente = Ativo Circulante / Passivo Circulante (PC)	Cheng; Chen; Fu (2006); Wang; Chen; Chu (2018)
X ₇	EIE	Despesas Financeiras (DF) / EBIT	Adaptado de Koyuncugil; Ozgulbas (2012); Chang; Hsu (2016)
X ₈	ROA	EBIT / Ativo Total	Cheng; Chen; Fu (2006); Tan (2012); Zieba; Tomczak; Tomczak (2016); Barboza; Kimura; Altman (2017); Wang; Chen; Chu (2018)
X ₉	GMNR	Margem Bruta / Receita Líquida	Sun; Jia; Li (2011); Ribeiro; Chen; Kovacecc (2017); Wang; Chen; Chu (2018)
X ₁₀	OLEB	(Passivo – Fornecedores) / EBIT	Hajek; Michalak (2013)
X ₁₁	LNA	Tamanho = ln (Ativo Total)	Altman (1968); Becker et al. (1998); Hajek; Michalak (2013); Korol (2013); Zieba; Tomczak; Tomczak (2016)
X ₁₂	MVS	Valor de Mercado das ações * número de ações <i>outstanding</i> / (Passivo Circulante + Passivo não Circulante)	Barboza; Kimura; Altman (2017)
X ₁₃	OLCF	(Passivo Circulante + Passivo não Circulante)/EBITDA	Adaptado Gombola; Ketz (1983); adaptado de Ribeiro; Chen; Kovacecc (2017)
X ₁₄	FEV	(DF _t – DF _{t-1}) / DF _{t-1}	Adaptado de Koyuncugil; Ozgulbas (2012); adaptado de Yu; Miche; Severin; Lendasse (2014)
X ₁₅	CPV	(PC + PnC) _t - (PC + PnC) _{t-1} / (PC + PnC) _{t-1}	Adaptado de Zelenkov; Fedorova; Chekrizov (2017)

Fonte: elaborado pelos autores.

3.4 A base de dados

A amostra compreendeu empresas do mercado de capitais brasileiro, onde foram analisados os relatórios de Informações Periódicas Eventuais (disponíveis no site da Comissão de Valores Mobiliários - CVM) ou no Relatório Anual da Empresa, para identificar as empresas que entraram em situação de falência ou teve o pedido de recuperação judicial deferido e também o momento em que se ocorreu o fenômeno, conforme definido pela Lei de Recuperação Judicial e Extrajudicial e de Falência nº 11.101 de 9 de fevereiro de 2005.

Os dados foram winsorizados ao nível de 2%, para redução dos *outliers*. Foi realizado o teste *Variance Inflation Factor* (VIF) para detectar multicolinearidade, que apresentou uma média de 2,10, indicando que os dados apresentam indícios que não possuem

multicolinearidade, atendendo aos pressupostos do teste (VIF médio < 10) (HAIR et al., 2014). Procedeu-se o teste F e verificou que o modelo que está sendo estimado de fato existe, onde a Prob > F = 0,000. Os Testes de Wald e Wooldridge, detectaram heterocedasticidade e autocorrelação. As análises devem ser consideradas com cautela pois os dados não são normais e não foram tratados. Utilizou-se a técnica *stepwise* na Regressão Logística, conforme Apêndice C, para encontrar a melhor combinação de variáveis preditoras para detectar o modelo mais adequado e funcional ao nível de confiança de 95% (TSAI, 2009; LIANG et al., 2016; JONES; JOHNSTONE; WILSON, 2016). Os testes estatísticos e a técnica *stepwise* foram executados utilizando as ferramentas Stata/MP 14.0. Após esta etapa os dados foram transferidos para a ferramenta RStudio.

Zmijewski (1984) examinou dois vieses de estimativa que podem resultar quando modelos de previsão para dificuldades financeiras são estimados em amostras não aleatórias. O primeiro viés resulta de empresas com problemas de "sobreamostragem" e se enquadra no tópico de vieses de amostras baseadas em escolhas. O segundo resulta na utilização de critérios de seleção de amostras de dados completos e se enquadra no tópico de vieses de seleção de amostras.

As duas questões surgem devido a restrições de seleção de amostra e coleta de dados de empresas em dificuldades financeiras. A primeira restrição é a taxa de frequência extremamente baixa dessas empresas e a segunda é que os dados dessas empresas muitas vezes não estão disponíveis. Assim, em decorrência do viés de desbalanceamento na base de dados e dados faltantes utilizou-se o algoritmo *Synthetic Minority Oversampling Technique* (SMOTE). Esse algoritmo faz um ajuste na classe minoritária, para que todas as classes sejam representadas igualmente no conjunto de dados de treinamento (CHAWLA et al., 2002; MARQUÉS; GARCÍA; SANCHÉZ, 2013; ZHOU, 2013; HAJEK; OLEJ; MYSKOVA, 2014; SUN et al., 2018; WANG; CHEN; CHU, 2018).

A amostra final de 336 empresas analisadas, no período de 2005 a 2018, resultou em 2.085 observações, sendo que dessas observações, 77,3% são de situações e períodos onde as empresas não foram classificadas em dificuldade financeira ou falência ou recuperação judicial deferida e 22,7% das observações foram identificadas com tal situação. Desse modo, a amostra final foi balanceada utilizando a técnica SMOTE proposto por Chawla et al. (2002), no qual novas observações sintéticas são criadas baseadas nas observações minoritárias existentes. O resultado foi de uma base total balanceada com 50% de observações classificadas como sem

dificuldades financeiras e 50% em dificuldades financeiras ou falência ou recuperação judicial deferida.

Empregou-se uma técnica robusta de validação cruzada, comumente utilizada nos modelos de aprendizado de máquina, com 10 subconjuntos, para a base de treinamento, que foi dividida aleatoriamente em 2/3 do total da amostra, para se desenvolver um modelo que melhor generaliza os resultados. O restante da base foi utilizado para teste (ZHANG et al., 1999; LI; SUN; SUN, 2009; ZHANG; GAO; SHI, 2014; CHENG; JONES; MOSE, 2018). A seleção das empresas foi realizada de modo aleatório.

Em relação as empresas da região de interesse, foram apuradas as cidades que pertencem as mesorregiões do Triângulo Mineiro e do Alto Paranaíba (66 cidades) e do Sul Goiano (82 cidades), conforme IBGE (2019). Verificou-se no site de cada empresa se ela possui sede ou filiais ou plantas ou lojas ou subsídios ou representantes, pertencente a alguma cidade que compreende a região de interesse, para agrupamento da amostra no cenário regional. Assim, a composição desses dados totalizou em 78 empresas, com 551 observações, sendo que, 86,4% são de empresas que não foram classificadas em dificuldade financeira ou falência ou recuperação judicial deferida e 13,6% das observações foram identificadas com tal situação. As empresas que compõem a região de interesse foram relacionadas no Apêndice B.

Os modelos de previsão e as técnicas SMOTE e de validação cruzada foram desenvolvidos no R, usando a plataforma RStudio 1.2.1335, que é um software estatístico de fonte aberta com módulos que admitem ajustes aos modelos de forma a conciliar o volume de dados e suas parametrizações.

3.5 Critérios de avaliação

Os critérios para avaliação dos modelos, geralmente utilizados na gestão de risco de crédito, também foram utilizados nesta pesquisa. Empregaram-se os seguintes para avaliação do desempenho e comparação dos modelos desta pesquisa: a acurácia (ACC), o Tipo de Erro I (TE I), o Tipo de Erro II (TE II), a Área sobre a curva (AUC) que é a área sob a curva *Receiver Operating Characteristic* (ROC) e o *Brier score* (BS) (DE CASTRO VIEIRA et al., 2019). Essas medidas foram selecionadas pois são as mais utilizadas para validação de modelos de previsão e abrangem os aspectos relacionados ao desempenho dos modelos (ALA'RAJ; ABBOD, 2016).

A tabela 8 exibe a matriz de confusão que traz a definição destas medidas.

Tabela 8 – Matriz de Confusão

Resultado do teste		Situação atual	
		Positivo (Risco)	Negativo (Não Risco)
	Positivo (Risco)	Verdadeiro Positivo (TP)	Falso Positivo (FP)
	Negativo (Não Risco)	Falso Negativo (FN)	Verdadeiro Negativo (TN)

Fonte: Wang, Chen e Chu (2018)

As medidas são calculadas conforme as seguintes equações (CHANG; HSU, 2016; ZELENKOV; FEDOROVA; CHEKRIZOV, 2017; WANG; CHEN; CHU, 2018):

$$ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Erro\ Tipo\ I = \frac{FP}{FP + TN} \quad (3)$$

$$Erro\ Tipo\ II = \frac{FN}{FN + TP} \quad (4)$$

O TE I trata das empresas que não estão em dificuldades financeiras, mas o modelo prediz que sim estão, e o TE II trata das empresas que estão em dificuldades financeiras, mas o modelo prediz que não (WHALEN, 1991; CHEN; DU, 2009; CHANG; HSU, 2016; GARCÍA et al., 2019). Whalen (1991) dispõe que ambos os erros de classificação são importantes na avaliação da capacidade de predição do modelo e que um modelo adequado deve exibir baixas taxas do TE II, pois essa falha causa demora na resolução da perda, altos custos, ou ambos. Ainda, segundo o autor, para um modelo ser considerado satisfatório, deve apresentar também baixas taxas de TE I, identificando talvez, antecipadamente, uma possível falha futura que possa vir a ocorrer no horizonte temporal, assim as empresas que estão nestas condições, devem ter um acompanhamento ao longo do tempo.

Uma das principais métricas é a AUC (*Area Under the Curve*), que fornece um critério de precisão para o conjunto de validação. A curva ROC mostra valores da taxa de verdadeiros positivos (TP) contra valores da taxa de falsos positivos (FP) em várias configurações de limite. A curva é determinada pelas taxas de acertos conhecidas como sensibilidade e especificidade calculadas na *confusion matrix*. Em outras palavras, o score AUC mede a eficiência do modelo em relação à discriminação entre as duas classes (WANG; CHEN; CHU, 2018; DE CASTRO VIEIRRA et al., 2019).

O *Brier Score* (BS) é uma medida de estimativa, proposta por Brier (1950), que trata do erro quadrático médio da probabilidade, ou seja, ele mostra se está errando muito na estimativa

ou não, assim, quanto menor a pontuação do BS melhor é o desempenho do classificador, ou seja, melhor é a previsão (ALA'RAJ; ABBOD, 2016, DE CASTRO VIEIRA et al., 2019).

$$BS = \frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2 \quad (5)$$

Em que n é o número de previsões emitidas, P_i é a probabilidade prevista para ocorrência do evento; e O_i é a variável binária (1/0) do evento real observado ou não (TASCHE, 2008; ALESSANDRINI; SPERATI; PINSON, 2013).

3.6 Resultados encontrados

A Tabela 9 traz as estatísticas descritivas da base de dados de treinamento e teste, dos cenários nacional e regional, evidenciando as informações no que tange ao número de observações, média, desvio-padrão, valores mínimos, máximos e coeficiente de variação (CV).

Tabela 9 – Estatísticas descritivas dos cenários nacional e regional

Variável	Obs.	Média	Desvio Padrão	Mín.	Máx.	CV	Obs.	Média	Desvio Padrão	Mín.	Máx.	CV
Cenário Nacional							Cenário Regional					
CALO	2.085	1,18	1,15	-0,12	7,21	0,98	551	1,21	0,85	0,03	7,21	0,70
L	2.085	0,73	0,71	0,09	5,64	0,97	551	0,62	0,52	0,15	5,64	0,83
TTA	2.085	0,15	0,40	-0,90	2,43	2,68	551	0,21	0,44	-0,69	2,43	2,14
CPB	2.085	0,69	9,79	-15,19	228,07	14,25	551	0,21	2,69	-15,19	26,21	12,99
AT	2.085	0,72	0,54	-0,02	3,07	0,75	551	0,83	0,57	0,00	3,07	0,68
CL	2.085	1,72	1,15	0,00	6,59	0,67	551	1,79	0,89	0,04	6,59	0,50
EIE	2.085	0,42	4,15	-24,90	17,72	9,86	551	0,50	3,35	-24,90	17,72	6,68
ROA	2.085	0,05	0,12	-0,50	0,40	2,15	551	0,08	0,09	-0,37	0,40	1,09
GMNR	2.085	0,31	0,21	-0,35	0,91	0,67	551	0,34	0,18	-0,35	0,91	0,53
OLEB	2.085	4,27	32,00	-164,69	143,18	7,48	551	5,36	25,53	-164,69	143,18	4,76
LNA	2.085	14,63	1,80	7,91	20,44	0,12	551	15,20	1,06	8,90	18,55	0,07
MVS	2.085	1,41	1,82	0,00	9,95	1,30	551	1,88	2,13	0,00	9,95	1,13
OLCF	2.085	6,00	63,85	-170,53	2.206,67	10,64	551	8,51	50,70	-170,53	975,42	5,96
FEV	2.085	0,32	1,30	-2,36	8,81	4,01	551	0,34	1,19	-2,36	8,81	3,51
CPV	2.085	0,17	0,45	-0,92	2,80	2,58	551	0,22	0,49	-0,87	2,80	2,27

Fonte: resultados da pesquisa.

Em que: CV: Coeficiente de Variação; CALO: (Ativo Circulante - Estoques - Recebíveis) / (Passivo Circulante-Fornecedores); L: (Passivo Circulante + Passivo não Circulante) / Ativo Total; TTA: ((Ativo Total_t - Ativo Total_{t-1})/Ativo Total_{t-1}); CPB: (CPB Price-to-Book_t - Price-to-Book_{t-1}); AT: Giro de Ativos: (Receita Líquida / Ativo Total); CL: (Ativo Circulante / Passivo Circulante (PC)); EIE: (Despesas Financeiras (DF) / EBIT); ROA: (EBIT / Ativo Total); GMNR: (Margem Bruta / Receita Líquida); OLEB: ((Passivo - Fornecedores) / EBIT); LNA: Tamanho = ln (Ativo Total); MVS: (Valor de Mercado / (Passivo Circulante + Passivo não Circulante)); OLCF: ((Passivo Circulante + Passivo não Circulante)/EBITDA); FEV: (DF_t - DF_{t-1}/DF_{t-1}); CPV: (PC + PnC)_t - (PC + PnC)_{t-1} / (PC + PnC)_{t-1}.

No cenário nacional, a variável menos dispersa, ou seja, que possui menor coeficiente de variação (0,123) é a LNA, que trata do logaritmo natural do Ativo Total, o que sugere que as empresas da amostra não se distanciam muito em relação ao seu porte médio baseado no

Ativo Total. Observa-se que as empresas da amostra apresentam, em média, bom CL, com baixo CV, o que sugere que possuem boa liquidez a curto prazo, ou seja, possuem capital disponível para uma possível liquidação das obrigações de curto prazo. Em relação ao CPB, que é a variável com maior dispersão em relação a média, sugere que as empresas possuem valores de mercado, que segundo Fama (1970), é determinado pela expectativa dos investidores, bem mais altos ou mais baixos que seu valor patrimonial. O OLFC também teve alta dispersão, são empresas que possuem dívidas muito altas ou muito baixas em relação a sua capacidade de geração operacional de caixa.

Para validação dos modelos, avaliação do desempenho e da capacidade na precisão preditiva, empregou-se a acurácia (ACC), o Erro Tipo I (TE I), o Erro Tipo II (TE II), a *Area Under the Curve* (AUC) e o *Brier score* (BS). A principal métrica de avaliação é a AUC, que se trata da medida de capacidade de discriminação baseada na curva ROC, ela fornece um critério de precisão para o conjunto de validação. A AUC deve ser superior a 0,5 para que o modelo seja aceitável, e quanto mais próximo de 1 for, melhor será o poder de previsão do modelo. A ACC representa a proporção das classes de empresas classificadas corretamente e mede o poder preditivo do modelo.

A tabela 10 apresenta os resultados de validação dos modelos testados nesta pesquisa, pode-se observar que a capacidade preditora e de discriminação entre as classes das empresas do modelo RF é mais satisfatório, com a ACC de 90,8%, o que demonstra um bom poder preditivo, e a AUC de 96,7%, apresentando uma boa precisão na predição em relação ao modelo de LR. O modelo de LR também teve bons resultados em relação a capacidade de predição e a precisão, sendo a ACC de 87,2% e a AUC de 93,4%. Na região, o modelo RF também foi mais satisfatório, com a ACC de 92,6%, apresentando uma boa precisão na predição em relação ao modelo de LR que apresentou uma ACC de 89,9%.

Assim, o modelo RF foi superior em relação a capacidade de predição e precisão em relação ao modelo LR, tanto no cenário nacional quanto no regional. Desta forma o método de aprendizado de máquina, baseado em AI superou o modelo estatístico tradicional, corroborando com os resultados de Brockett et al. (1994), Zhang et al. (1999), Chandra, Ravi e Bose (2009), Brown e Mues (2012), Hsieh, Hsiao, Yeh (2012), Jones, Johnstone e Wilson (2015), Lessmann et al. (2015), Jones, Johnstone e Wilson (2016), Barboza, Kimura e Altman (2017), Bequé e Lessmann (2017) e contrariando os resultados de Liou (2008), Bekhet e Eletter (2014), Wu,

Olson e Luo (2014). Esta pesquisa se diferencia pois trata de uma economia emergente, comparando os resultados do modelo com empresas regionais.

Tabela 10 – Resultados e comparação dos métodos RF e LR nacional e regional

Modelos / Cenários	TP	TN	FP	FN	TE I	TE II	ACC	AUC	BS
RF - Nacional	151	480	55	9	0,103	0,056	0,908	0,967	0,071
RF - Regional	24	150	10	4	0,063	0,143	0,926	0,963	0,056
LR - Nacional	140	466	69	20	0,129	0,125	0,872	0,934	0,101
LR - Regional	19	150	10	9	0,063	0,321	0,899	0,909	0,084

Fonte: resultados da pesquisa.

O BS trata-se do erro quadrático médio da probabilidade, ele é uma medida de estimativa e mensura a precisão das previsões de probabilidades dos conjuntos de dados e quanto menor for o BS, melhores são as previsões (ALA'RAJ; ABBOD, 2016). Seguindo este conceito, o modelo de RF apresentou o melhor BS, com 0,071 e o de LR teve 0,101, no cenário nacional, já no cenário regional o RF obteve um BS de 0,056 e o de LR 0,084. Considerando esta métrica, que segundo De Castro Vieira et al. (2019), tem a função de medir a precisão das previsões probabilísticas e quanto menor, melhor é a previsão. O modelo RF foi mais preciso que o modelo de LR em ambos os cenários, mas no cenário regional os dois modelos melhoraram a precisão, isso sugere que dentro das probabilidades de ambos os modelos, a assertividade aumentou.

O TE I trata daquelas empresas que não se encontram em situação de dificuldade financeira ou falência ou em recuperação judicial deferida, mas que o modelo erroneamente as considera nesta condição, e o TE II trata da previsão incorreta das empresas que estão em situação de dificuldade financeira ou falência ou em recuperação judicial deferida e o modelo prediz que não estão nesta situação. Na tabela 10, pode-se verificar que o modelo RF foi superior nas duas modalidades de erro, onde o TE I foi de 10,3% e o TE II foi de 5,6% e no modelo de LR o TE I foi de 12,9% e o TE II foi de 12,5%, corroborando com os resultados de Zelenkov, Fedorova e Chekrizov (2017) em que os erros de classificação do modelo baseado em algoritmos, foram menores que os erros do modelo *Logit*.

Nas empresas da região, ambos os modelos tiveram piora no TE II, (RF: 14,3%; LR: 32,1%), no TE I, ao contrário teve melhora nos dois modelos (RF: 6,2%; LR: 6,2%), desta forma, regionalmente, como o TE I é aquele que menos onera, sendo ele menor, proporciona à empresa maior margem para alavancar o negócio (JONES; JOHNSTONE; WILSON, 2016).

Numa avaliação superficial dos erros de classificação entre os cenários nacional e regional, o modelo RF, no TE I, o modelo errou mais duas empresas, que estavam na região de

incerteza, com 0,45 e 0,48 de probabilidade, que passaram para 0,63 e 0,56 respectivamente, mas ainda assim, o percentual de erro reduziu, isso ocorreu pois a proporção da amostra aumentou, caindo assim o TE I de 10,3% para 6,2%. No TE II, o modelo se manteve, não alterou as classificações, mas ainda assim o TE II aumentou, em decorrência da proporção de empresas que caiu, e assim, fez com que o erro piorasse, isso pode ser visto na tabela 10. O Apêndice D, traz o detalhamento dessas empresas e as variáveis que podem ter contribuído para que o TE I piorasse.

No caso da LR, fazendo a mesma comparação superficial entre o cenário nacional e regional, no TE I, o modelo errou mais três empresas e acertou uma, mas ainda assim, se igualou ao modelo RF, pois a proporção de empresas da amostra aumentou, reduzindo o erro. No TE II, o modelo errou mais cinco empresas, assim houve um aumento significativo no erro, pois a proporção da razão de empresas caiu, isso pode ser visto na tabela 10. O Apêndice E, traz o detalhamento dessas empresas e as variáveis que podem ter influenciado para que o TE II piorasse.

Os riscos associados o falso negativo (TE II) podem ser mais onerosos em relação aos custos envolvidos do que aqueles que compreendem o falso positivo (TE I), onde os custos são menores (WHALEN, 1991). Um ponto de corte do erro deve ser avaliado, levando em consideração o planejamento estratégico da organização, os prazos de recebimentos de clientes, os custos diretos e indiretos envolvidos na recuperação dos créditos financeiros, bem como as perdas financeiras irrecuperáveis e o custo de oportunidade que, pode não ser remunerado na recuperação dos créditos. Uma alternativa é a avaliação do padrão das empresas em que os modelos as classificam na faixa da área cinzenta, onde nesse intervalo, de acordo com Altman, Danovi e Falini (2013), o futuro da empresa avaliada é incerto, tanto quanto ao fracasso como na sua continuidade.

A área cinzenta precisa ser tratada conforme a particularidade de cada segmento e negócio, inclusive diante das condições do próprio mercado, ponderando os custos individuais. E somado a esta questão, os cortes de crédito e as perdas financeiras são metas delimitadas conforme orientação estratégica da organização, assim seus gestores são direcionados diante das limitações que tratam os erros de classificação e dos custos que envolvem as perdas.

Elucidando sobre esta questão de erros de classificação, Jones, Johnstone e Wilson (2016) discorrem que, deixar de atender um “bom” cliente por falha de classificação, deixa-se de se ganhar os lucros da operação, pois diante das políticas de crédito definidas, optou-se por

não conceder o crédito, no intuito de manter a exposição ao risco mais baixa possível. No entanto, uma outra instituição pode se interessar em expandir sua carteira e assim, assumir mais riscos, de forma mais agressiva e desta forma, tenha uma tolerância mais alta de inadimplência em sua estratégia de negócios. Nesse caso, a organização definirá estrategicamente um ponto de corte tolerável para os TE I (corte, lucro) e TE II (perdas financeiras, custos de recuperação e de oportunidade, etc.).

Altman (1968) sugere a seleção desse ponto de corte, para que permita analisar a empresa que foi classificada incorretamente pelos modelos na amostra inicial, identificando a área cinzenta, onde não se tem a certeza sobre a situação da empresa, devido à suscetibilidade na classificação. Portanto, é desejável estabelecer uma orientação para os pontos de corte da "área cinzenta", a Tabela 11 expõe alguns pontos de corte para os modelos RF e LR, nos cenários nacional e regional.

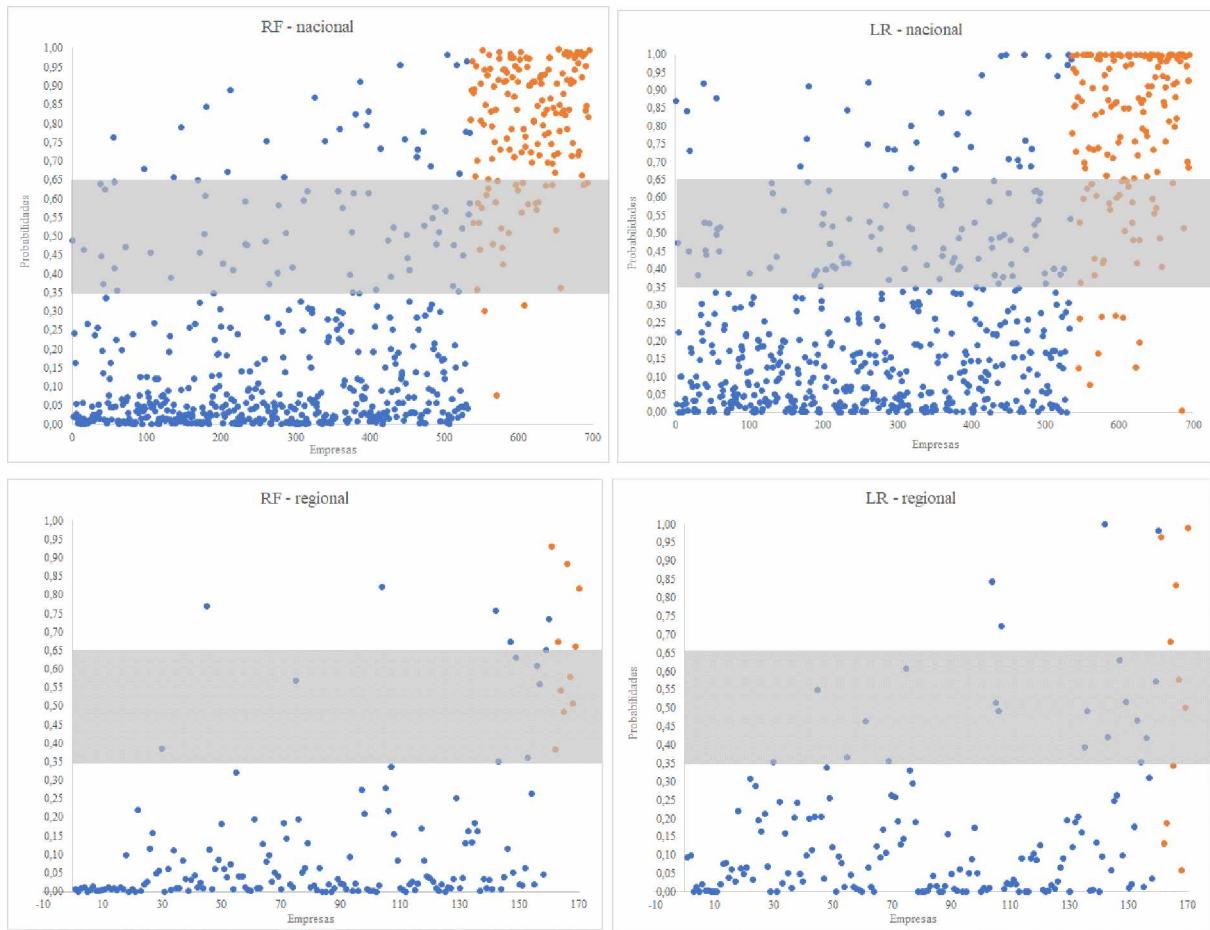
Tabela 11 – Pontos de corte para definição da área cinzenta

Área cinzenta	0,10		0,20		0,30		0,40		0,50	
	0,45 a 0,55	0,40 a 0,60	0,35 a 0,65	0,30 a 0,70	0,25 a 0,75	0,20 a 0,50	0,15 a 0,40	0,10 a 0,30	0,05 a 0,20	0,02 a 0,10
Modelos / Cenários	TE I	TE II								
RF - Nacional	9	3	17	4	26	6	32	8	35	8
RF - Regionalidade	0	1	2	1	4	3	6	3	7	3
LR - Nacional	11	4	20	10	26	13	33	14	41	16
LR - Regionalidade	3	1	4	1	6	1	6	4	7	5
RF - Nacional	0,087	0,038	0,073	0,032	0,057	0,019	0,046	0,007	0,040	0,007
RF - Regionalidade	0,062	0,111	0,051	0,111	0,038	0,040	0,026	0,040	0,020	0,040
LR - Nacional	0,111	0,103	0,095	0,067	0,084	0,048	0,072	0,041	0,057	0,028
LR - Regionalidade	0,045	0,296	0,038	0,296	0,026	0,296	0,026	0,208	0,020	0,174

Fonte: resultado das pesquisas.

A figura 15 apresenta a dispersão das empresas conforme as probabilidades de classificação dos modelos RF e de LR, nos cenários nacional e regional, e demonstra quais as empresas que se encontram na área cinzenta, ou seja, que a classificação foi próxima da incerteza, considerando um intervalo do ponto de corte na probabilidade de 35% a 65% (Tabela 11). Para o modelo nacional de RF, o TE I seria 5,7% e o TE II seria 1,9%, essas empresas que compõem a área cinzenta devem ser tratadas com cautela devido ao grau de incerteza do próprio método, assim essas avaliações poderiam ser realizadas por analistas manualmente, levando em consideração outras informações e particularidades do negócio.

Figura 15 – Gráficos de dispersão dos modelos nacionais RF e LR, no cenário nacional e regional, respectivamente com delimitação da área cinzenta



Fonte: resultado das pesquisas.

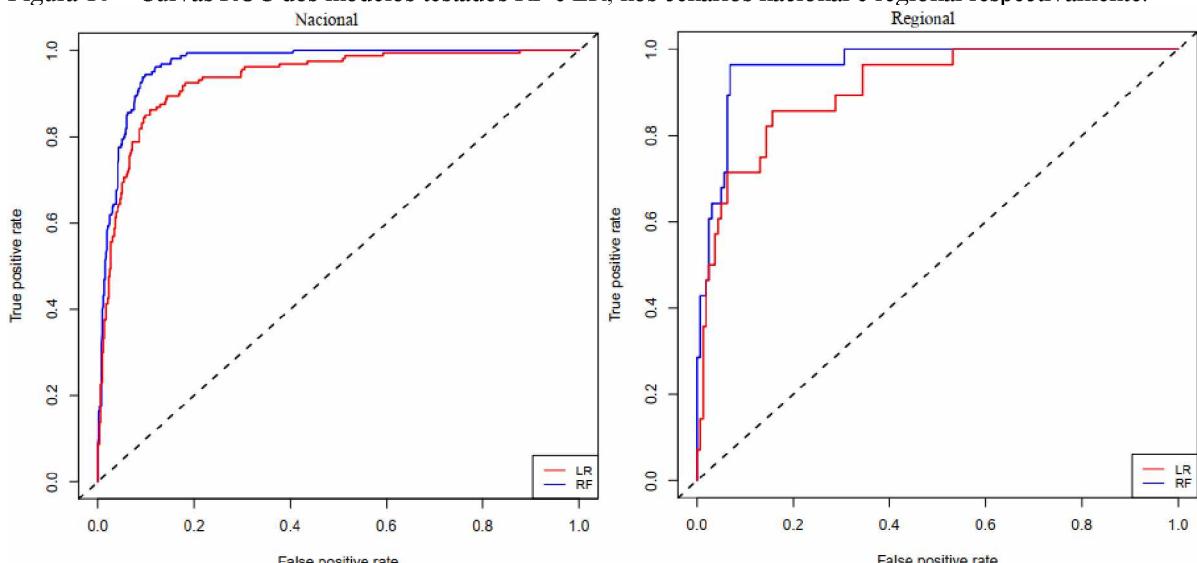
No modelo nacional de LR, considerando o mesmo ponto de corte (Tabela 11), o TE I seria 8,4% e o TE II seria 4,8%, da mesma forma seria viável uma avaliação criteriosa das empresas nesta classificação. Os pedidos manuais geram custos elevados de mão de obra especializada, assim cabe a empresa analisar a viabilidade de implementação de esteira manual, diante do seu volume de perdas geradas e custos envolvidos na avaliação manual. Beynon e Peel (2001) discorrem que a intervenção de um especialista humano pode não ser viável, uma vez que o uso do conhecimento pode ser impraticável, relativamente caro e ainda introduzir um nível inaceitável de viés subjetivo na análise.

Considerando o mesmo parâmetro (tabela 11) para as empresas da região, o modelo RF teria de TE I 3,8% e de TE II 4% e o modelo de LR o TE I seria 2,6% e o TE II de 29,6%. Em relação ao TE I, para as empresas da região o modelo LR foi menor considerando esse intervalo para a área cinzenta, o RF não teve nenhuma empresa classificada no TE II desconsiderando a área cinzenta. Em ambos os modelos e cenários, seria interessante analisar detalhadamente as características das empresas que foram classificadas na área cinzenta, para se entender qual o

padrão que influenciou os modelos. Um exemplo, no caso do modelo LR, no cenário regional, o erro aumentou em empresas do setor “Indústria de equipamentos e materiais para uso médico”, desta forma, em pesquisas futuras, pode-se entender melhor essas influências.

A curva ROC demonstra os valores das taxas de verdadeiros positivos (TP) em relação aos valores da taxa de falsos positivos (FP) em várias configurações de limite. A sua área fornece a medida em relação a precisão total independente. Os valores da área abaixo da diagonal (0,5) não são válidos, pois são considerados aleatórios, porém, os valores que se aproximam de 1,0, demonstram a capacidade do modelo em fazer previsões assertivas, apontando o poder de discriminação entre as classes, determinando os erros tipo I e tipo II, ou seja, quanto mais próximo do 1,0 a curva estiver, menores serão os erros (JONES; JOHNSTONE; WILSON, 2016).

Figura 16 – Curvas ROC dos modelos testados RF e LR, nos cenários nacional e regional respectivamente.



Fonte: resultados da pesquisa.

Na figura 16, os modelos RF e LR são comparados. Observou-se que o modelo RF teve uma performance satisfatória, porém, quando comparado com o modelo de LR, a variação não é tão significativa, mas, ainda superior. A nível nacional o modelo RF apresentou AUC de 0,967 e o LR de 0,934, já nas empresas da região o modelo RF apresentou AUC de 0,963 e o LR de 0,909, sugerindo que os dois classificadores são muito fortes, pois exibem um excelente equilíbrio entre a sensibilidade e a especificidade diante dos diferentes limiares de probabilidade (JONES; JOHNSTONE; WILSON, 2016).

As variáveis de entrada, que determinaram as probabilidades, com maior relevância para os dois modelos, são expostas como forma de rastrear e monitorar as características

determinantes dos fatores que influenciaram o poder de predição e precisão de cada modelo em específico. A Figura 17 apresenta a contribuição relativa de cada variável de entrada por ordem de relevância no desempenho de cada modelo, RF e LR, nacional e regional, respectivamente. Na tentativa de entender o comportamento dessas variáveis de maior relevância dos modelos, analisou-se conjuntamente as tabelas 11 e 12.

No modelo RF, observou-se que as variáveis EIE, ROA, OLEB, OLCF e L são as mais representativas, que tratam de rentabilidade, grau de alavancagem, custos da dívida e liquidez. No modelo LR, observou-se que as variáveis ROA, LNA, L e MVS foram as mais relevantes.

A partir desses resultados, apresentam-se a lucratividade, a rentabilidade, a liquidez e a alavancagem como características relevantes para se prever dificuldades financeiras, falência e recuperação judicial no mercado brasileiro, o que reforça os estudos de Altman (1968).

O grau de endividamento, representado por L, que trata da proporção dos ativos financiados com recursos de terceiros de curto e longo prazos e, consequentemente mede a dimensão dos recursos totais resultantes de credores, foi relevante para os dois modelos. Desta forma, quanto menor o grau de endividamento da empresa, maior será a sua segurança em termos de pagamentos com credores, assim, as empresas menos endividadas tem menor propensão de entrar em dificuldades financeiras e maior será a sua facilidade para obtenção de empréstimos com custos menores, o que corrobora com os achados de Beynon e Peel (2001), que apesar de utilizarem métodos diferentes e conjuntos de dados de outro país, também encontraram relevância nesta variável.

Outra questão é que empresas mais alavancadas com capital de origem de terceiros, recorrem mais a empréstimos e financiamentos no intuito de liquidar suas dívidas e tentarem manter sua situação de caixa líquida. Ohlson (1980) sugere que empresas mais alavancadas estão mais propensas a falência e que esta variável é relevante para determinação do modelo de previsão. Tan (2012) apresentou, em seu modelo, a existência de relação negativa entre alavancagem e desempenho, ou seja, empresas mais alavancadas apresentam menores retornos. Cheng, Chen e Fu (2006) demonstraram que empresas mais alavancadas são mais propensas a entrar em dificuldade financeira.

Quando se avalia o ROA, mede-se a eficiência da empresa em administrar seus ativos para gerar rentabilidade. Altman (1968) discorre que esta é uma medida da verdadeira produtividade da empresa. Assim, os resultados sugerem que as firmas que possuem maior rentabilidade em relação aos seus investimentos em ativos, tem menos propensão para entrar

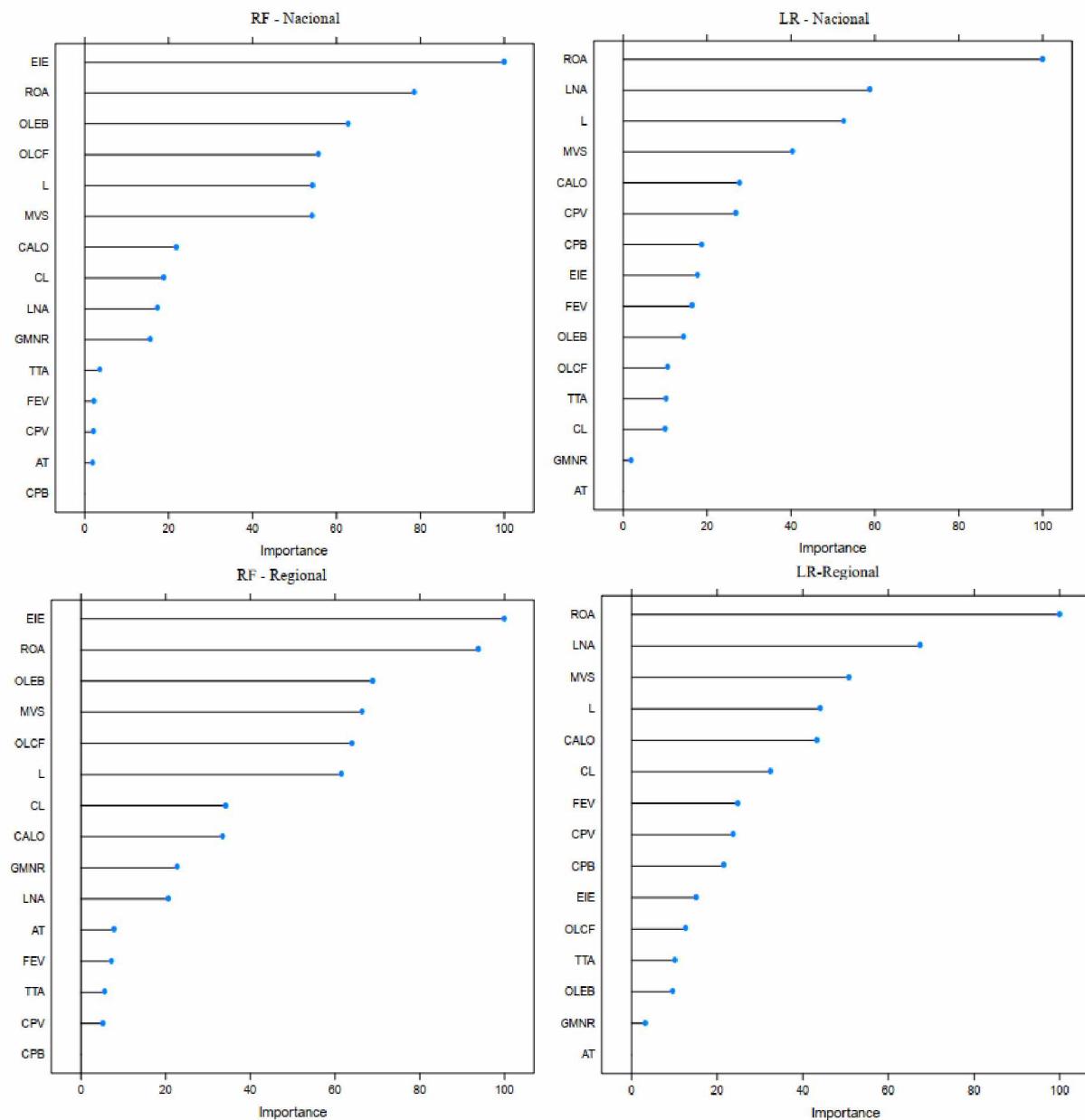
em situação de estresse financeiro, já aquelas que não possuem capacidade de gerar lucro a partir de seus ativos, estão mais propensas a entrarem em situação de estresse financeiro. Cheng, Chen e Fu (2006) e Scalzer et al. (2019), em seus modelos, também trataram o nível de rentabilidade como relevante e associado negativamente com a dificuldade financeira. Os achados de Tan (2012) também demonstram que as empresas com dificuldades financeiras apresentam menor desempenho (ROA).

A variável OLCF, representa o quanto a firma é capaz de liquidar suas dívidas brutas utilizando sua capacidade de geração operacional de caixa, demonstrando que a própria operação consegue sanar as dívidas com credores e governos, sem recorrer às suas disponibilidades e aplicações, esta variável foi mais relevante para o modelo RF. Estes resultados reforçam os achados de Gombola e Ketz (1983) sobre a utilização do fluxo de caixa operacional como variável relevante no desempenho das firmas e contraria os resultados de Casey e Bartczak (1985) que também utilizaram a LR para previsão de falência, mas os resultados com o índice de fluxo de caixa operacional não forneceram incremento no poder preditivo. Outra variável de representatividade neste modelo foram os gastos provenientes de despesas financeiras em relação a rentabilidade (EIE), sugerindo que o reflexo de custos financeiros maiores, decorrentes de capital de terceiros, podem onerar o retorno demonstrando a ineficiência da empresa (CHANG; HSU, 2016).

As empresas de maior porte representadas pelo logaritmo natural do ativo total (LNA), ou seja, empresas maiores apresentam menor probabilidade de entrar em dificuldade financeira, isso pode ocorrer em decorrência do volume de investimentos em ativos, operações, filiais, produtos, entre outros. Esses resultados contribuem com Becker et al. (1998), que afirmam que o tamanho da firma é uma variável relevante e que pode substituir diversas outras variáveis que possam estar omitidas nos modelos. Ohlson (1980) discorre que o tamanho da empresa é um importante preditor de falência.

Em ambos os modelos, empresas mais valorizadas no mercado de capitais em relação aos seus passivos (MVS), são menos propensas a entrar em dificuldades financeiras. Altman (1968) dispõe que esta medida demonstra quanto dos ativos da empresa podem diminuir em valor (medido pelo valor de mercado mais a dívida) antes que os passivos excedam os ativos e a empresa se torne inadimplente. Corroborando com os estudos de Altman (1968), essa variável foi um relevante preditor para os modelos.

Figura 17 – Contribuição relativa das variáveis de entrada no desempenho dos modelos RF e LR, nacional e regional, respectivamente.



Fonte: resultados da pesquisa.

Em que: CALO: (Ativo Circulante - Estoques - Recebíveis) / (Passivo Circulante- Fornecedores); L: (Passivo Circulante + Passivo não Circulante) / Ativo Total; TTA: (Ativo Total_t – Ativo Total_{t-1}) / Ativo Total_{t-1}; CPB: CPB Price-to-Book_t – Price-to-Book_{t-1}; AT: Giro de Ativos: Receita Líquida / Ativo Total; CL: Ativo Circulante / Passivo Circulante (PC); EIE: Despesas Financeiras (DF) / EBIT; ROA: EBIT / Ativo Total; GMNR: Margem Bruta / Receita Líquida; OLEB: (Passivo – Fornecedores) / EBIT; LNA: Tamanho = ln (Ativo Total); MVS: Valor de Mercado / (Passivo Circulante + Passivo não Circulante); OLCF: (Passivo Circulante + Passivo não Circulante)/EBITDA; FEV: DF_t – DF_{t-1}/DF_{t-1}; CPV: (PC + PnC)_t - (PC + PnC)_{t-1} / (PC + PnC)_{t-1}

Outra situação apontada é que empresas com maiores investimentos em aplicações financeiras em relação aos seus passivos onerosos (CALO), desconsiderando assim os ativos e passivos de funcionamento (clientes, fornecedores e estoques), também são menos propensas a

entrar em situação de dificuldades financeiras, possivelmente essas empresas procuram manter liquidez no curíssimo prazo e conseguem assegurar sua operação com o caixa gerado das próprias atividades. Scalzer et al. (2019) também encontraram relevância nessa variável de liquidez imediata. Os autores discorrem que as empresas que possuem este índice maior, estão mais focados em tratar a liquidez instantânea para garantir maior solvência financeira.

A variável CL também foi relevante nos modelos, com ênfase no cenário regional, e ela representa um excedente de caixa para a empresa saldar suas obrigações no momento determinado. Empresas que possuem baixa liquidez corrente, necessitam de geração de caixa operacional no curto prazo ou da venda de algum ativo fixo para sanar suas dívidas de curto prazo. Cheng, Chen e Fu (2006) e Scalzer et al. (2019) também encontraram relevância nessa variável em seus modelos, os autores apresentaram que este índice pode estar associado a maiores necessidades de capital de giro no curto prazo, sendo negativamente relacionado à dificuldade financeira.

Diversos autores consideram a tecnologia de AI e aprendizado de máquina uma tecnologia de caixa preta (ZMIJEWSKI, 1984; WANG et al., 2011; OLSON; DELEN; MENG, 2012; ZHANG; WANG; JI, 2013; CHUANG, 2013). Assim, realizou-se uma análise das variáveis para melhor entender as decisões tomadas nos nós do modelo RF, construído a partir da base de treinamento, seguindo o comportamento e as particularidades do mercado brasileiro, que foi condizente com o esperado e é apresentado na tabela 12.

Tabela 12 – Avaliação do poder discricionário do modelo RF em relação as variáveis de entrada por relevância

Grau Imp.	Sit./ Var.	0 Mín.	0 Máx.	0 Média	0 Q1	0 Med.	0 Q3	1 Mín.	1 Máx.	1 Média	1 Q1	1 Med.	1 Q3	RF Melhor
1	EIE	-15,1	17,7	0,6	0,2	0,4	0,8	-24,9	17,7	0,7	-0,4	0,4	1,3	menor
2	ROA	0,0	0,4	0,1	0,0	0,1	0,1	-0,5	0,4	-0,1	0,0	0,0	0,1	maior
3	OLEB	-111,9	143,2	7,2	2,8	5,1	9,1	-165	143	2,4	-4,1	5,1	12,1	menor
4	OLCF	-35,4	35,2	5,2	2,6	4,2	6,9	-170,5	322,5	-4,8	-3,0	5,1	9,7	menor
5	L	0,1	1,9	0,5	0,4	0,6	0,7	0,1	5,6	1,2	0,5	0,6	0,8	menor
6	MVS	0,0	10,0	1,8	0,5	1,0	2,4	0,0	10,0	0,6	0,1	0,4	1,0	maior
7	CALO	0,0	7,2	1,3	0,6	1,0	1,6	-0,1	7,2	0,9	0,3	0,8	1,4	maior
8	CL	0,3	6,6	1,9	1,1	1,5	2,1	0,0	6,6	1,3	0,9	1,5	2,2	maior
9	LNA	10,6	20,2	15,1	13,9	15,1	16,3	7,9	19,0	13,6	13,3	14,1	15,5	maior
10	GMNR	0,0	0,9	0,4	0,2	0,3	0,5	-0,4	0,9	0,2	0,1	0,3	0,4	maior
11	TTA	-0,7	2,4	0,2	0,0	0,1	0,2	-0,9	2,4	0,1	0,0	0,0	0,2	maior
12	FEV	-2,4	8,8	0,2	-0,2	0,1	0,4	-2,4	8,8	0,4	-0,1	0,1	0,5	menor
13	CPV	-0,9	2,8	0,2	0,0	0,1	0,2	-0,8	1,3	0,1	0,0	0,1	0,2	menor
14	AT	0,0	3,1	0,7	0,4	0,6	0,9	0,0	3,1	0,7	0,3	0,5	0,9	maior
15	CPB	-15,2	90,9	0,2	-0,6	0,0	0,8	-15,2	21,1	0,0	-0,7	0,0	0,4	maior

Fonte: resultados da pesquisa.

Em que: Grau Imp.: Grau de Importância; Sit./Var.: Situação/Variável; CALO: (Ativo Circulante - Estoques - Recebíveis) / (Passivo Circulante- Fornecedores); L: (Passivo Circulante + Passivo não Circulante) / Ativo Total; TTA: ((Ativo Total_t – Ativo Total_{t-1})/Ativo Total_{t-1}); CPB: (CPB Price-to-Book_t – Price-to-Book_{t-1}); AT: Giro de Ativos: (Receita Líquida / Ativo Total); CL: (Ativo Circulante / Passivo Circulante (PC)); EIE: (Despesas Financeiras (DF) / EBIT); ROA: (EBIT / Ativo Total); GMNR: (Margem Bruta / Receita Líquida); OLEB:

((Passivo – Fornecedores) / EBIT); LNA: Tamanho = ln (Ativo Total); MVS: (Valor de Mercado / (Passivo Circulante + Passivo não Circulante)); OLCF: ((Passivo Circulante + Passivo não Circulante)/EBITDA); FEV: ($DF_t - DF_{t-1}/DF_{t-1}$); CPV: $(PC + PnC)_t - (PC + PnC)_{t-1} / (PC + PnC)_{t-1}$

Foi realizada uma análise descritiva a partir das empresas em dificuldades financeiras ou falidas ou em recuperação judicial deferida (1) e caso contrário (0) para verificar como o modelo se comportou, na condição de que se for maior ou menor o indicador, melhor foi para o modelo, levando em consideração que o agrupamento das empresas que não estão em estresse financeiro, representadas pelo zero, foi a referência para o melhor, e utilizou-se ainda os valores mínimo e máximo, do primeiro (Q1) e do terceiro quartil (Q3), da média e da mediana nesta avaliação. Ressaltando que esta avaliação deve ser considerada com cautela, para o modelo RF, pois ele não trata as variáveis de forma linear, ou seja, pode ser que em determinado momento de decisão do nó, o modelo possa ter invertido o voto de decisão.

No modelo LR nacional apresentado na tabela 13, as saídas também foram condizentes com o esperado, apresentando os mesmos resultados do modelo RF.

Tabela 13 – Avaliação do poder discricionário do modelo LR em relação as variáveis de entrada por relevância

Grau Imp.	Sit./ Var.	0 Mín.	0 Máx.	0 Média	0 Q1	0 Med.	0 Q3	1 Mín.	1 Máx.	1 Média	1 Q1	1 Med.	1 Q3	LR Melhor
1	ROA	-0,2	0,4	0,1	0,0	0,1	0,1	-0,5	0,4	-0,1	0,0	0,0	0,1	maior
2	LNA	10,6	20,2	15,1	13,8	15,1	16,3	7,9	18,8	13,6	13,3	14,1	15,6	maior
3	L	0,1	1,1	0,5	0,4	0,6	0,7	0,2	5,6	1,2	0,5	0,6	0,8	menor
4	MVS	0,0	10,0	1,8	0,5	1,0	2,4	0,0	2,5	0,6	0,1	0,4	1,1	maior
5	CALO	0,0	7,2	1,3	0,6	1,0	1,6	-0,1	7,2	0,9	0,3	0,8	1,4	maior
6	CPV	-0,9	2,8	0,2	0,0	0,1	0,2	-0,8	1,1	0,1	0,0	0,1	0,2	menor
7	CPB	-15,2	90,9	0,2	-0,6	0,0	0,8	-15,2	21,1	0,0	-0,8	0,0	0,4	maior
8	EIE	-24,9	17,7	0,6	0,2	0,4	0,8	-24,9	17,7	0,7	-0,2	0,5	1,3	menor
9	FEV	-2,4	8,8	0,2	-0,2	0,1	0,4	-2,4	8,8	0,4	-0,2	0,1	0,5	menor
10	OLEB	-111,9	143	7,2	2,8	5,1	9,1	-164,7	143	2,4	-4,1	5,2	12,1	menor
11	OLCF	-57,6	64,0	5,2	2,6	4,2	7,0	-170,5	322,5	-4,8	-2,9	5,0	9,7	menor
12	TTA	-0,7	2,4	0,2	0,0	0,1	0,2	-0,9	2,4	0,1	0,0	0,0	0,2	maior
13	CL	0,3	6,6	1,9	1,1	1,5	2,1	0,0	6,6	1,3	0,9	1,5	2,1	maior
14	GMNR	-0,1	0,9	0,4	0,2	0,3	0,5	-0,4	0,9	0,2	0,1	0,3	0,4	maior
15	AT	0,0	3,1	0,7	0,4	0,6	0,9	0,0	3,1	0,7	0,3	0,5	0,9	maior

Fonte: resultados da pesquisa.

Em que: Grau Imp.: Grau de Importância; Sit./Var.: Situação/Variável; CALO: (Ativo Circulante - Estoques - Recebíveis) / (Passivo Circulante- Fornecedores); L: (Passivo Circulante + Passivo não Circulante) / Ativo Total; TTA: ((Ativo Total_t – Ativo Total_{t-1})/Ativo Total_{t-1}); CPB: (CPB Price-to-Book_t – Price-to-Book_{t-1}); AT: Giro de Ativos: (Receita Líquida / Ativo Total); CL: (Ativo Circulante / Passivo Circulante (PC)); EIE: (Despesas Financeiras (DF) / EBIT); ROA: (EBIT / Ativo Total); GMNR: (Margem Bruta / Receita Líquida); OLEB: ((Passivo – Fornecedores) / EBIT); LNA: Tamanho = ln (Ativo Total); MVS: (Valor de Mercado / (Passivo Circulante + Passivo não Circulante)); OLCF: ((Passivo Circulante + Passivo não Circulante)/EBITDA); FEV: ($DF_t - DF_{t-1}/DF_{t-1}$); CPV: $(PC + PnC)_t - (PC + PnC)_{t-1} / (PC + PnC)_{t-1}$

Diante das avaliações realizadas nas tabelas 11 e 12, observou-se que a tendência das variáveis em relação ao grau de importância e de quanto maior ou menor melhor o indicador, é

o mesmo para os dois modelos. Os pesos de relevância, entre os indicadores, são distintos entre os modelos e os coeficientes da LR por variável, que é apresentado no Apêndice F.

Pode-se inferir também que, sendo a relevância das variáveis o que determina os modelos, esta situação interfere diretamente na saída de dados, estabelecendo os resultados finais de precisão, predição e erros, apesar do modelo RF, não tratar suas decisões de forma linear. Assim, sugere-se para novas pesquisas, modificações em relação a estrutura dos modelos no que tange as variáveis direcionadas para cada modelo em específico e assim a construção partirá de variáveis distintas para cada um, atingindo, cada modelo, seu melhor desempenho conforme sua condição individualmente.

Os Apêndices F e G trazem os modelos de LR bem como o algoritmo utilizado no modelo RF, respectivamente.

3.7 Considerações Finais

A temática sobre risco de crédito recebe atenção crescente tanto para melhora da precisão da classificação, da predição, quanto para redução nos tipos de erro I e II (AGHAEIRAD; CHEN; RIBEIRO, 2016; BARBOZA; KIMURA; ALTMAN, 2017). Diversos modelos de aprendizado de máquina têm surgido para suprir esta demanda. Os modelos são bem sucedidos quanto aos resultados. Altman (2002) já demonstrava a preocupação em relação aos principais fatores motivacionais técnicos que incluem o refinamento das técnicas tradicionais, as novas soluções analíticas, a questão em relação aos conjuntos de dados maiores e melhorados, para se tratar a gestão do risco de crédito.

O objetivo deste capítulo foi de avaliar e comparar dois modelos, um computacional, utilizando mecanismos de aprendizado de máquina, RF e outro baseado em estatística tradicional, *Logit*, para prever previamente o fenômeno de insolvência em empresas de capital aberto no Brasil e na região (mesorregião do Triângulo Mineiro e Alto Paranaíba e do Sul Goiano) entre o período de 2005 a 2018. A amostra compreendeu uma base para treinamento e outra de teste, ambas de empresas listadas na B3.

As pontuações de ACC, AUC, BS, TE I e TE II indicam que o classificador de aprendizado de máquina, *Random Forest*, superou o modelo de Regressão Logística em relação aos dados amostrados de empresas brasileiras e também regionalmente. Considerando o índice de precisão AUC, os resultados mostraram que o modelo estatístico tradicional tem capacidade preditiva inferior quando comparado ao modelo de aprendizagem de máquina, a nível nacional

os resultados foram de 93,4% e 96,7%, e nas empresas da região foram de 90,9% e 96,3%, respectivamente.

Os resultados deste trabalho corroboram com os estudos de Brockett et al. (1994), Zhang et al. (1999), Chandra, Ravi e Bose (2009), Hsieh, Hsiao, Yeh (2012), Jones, Johnstone e Wilson (2016) e Barboza, Kimura e Altman (2017), no sentido de comparação entre estatísticas tradicionais e inteligência artificial, mas eles utilizaram conjuntos de dados de outros países. Esta pesquisa se diferencia, além de tratar de um mercado emergente e também apresentando os resultados do modelo regionalmente, esta pesquisa ainda dispõe das relevâncias das variáveis que constituem os modelos.

Novos estudos podem adaptar os modelos conforme a relevância das variáveis para cada técnica em específico. O classificador RF mostrou-se preciso e com alta capacidade de predição, podendo ser promissor para futuras pesquisas e para implementação prática na área de finanças no mercado brasileiro e na regionalidade. Os classificadores avançados não são passíveis de subjetividade pois não requerem intervenção humana, o que os torna mais precisos (Lessmann et al., 2015). A previsão de dificuldade financeira das firmas é um campo que necessita constantemente de investigação e melhorias, apenas um por cento de aumento na precisão e predição, pode trazer benefícios em relação as perdas financeiras e o retorno das organizações (WANG et al., 2012; BARBOZA; KIMURA; ALTMAN, 2017).

Diversos trabalhos têm encontrado resultados significativamente melhores utilizando combinações de diferentes classificadores, híbridos ou uma combinação mais extensa de estratégias de conjuntos ou mesmo modelos globais de previsão para obter melhor desempenho, que podem ser aplicados no mercado brasileiro (ALFARO et al., 2008; FINLAY, 2011; CHOU; HSIEH; QIU, 2017). Outra questão interessante para aplicação em pesquisas futuras é a inclusão de indicadores qualitativos e de governança corporativa (Kim; Han, 2003; Chen; Shih, 2006; Yu; Wang; Lai, 2008; Hsieh; Hsiao; Yeh, 2012; Liang et al., 2016; Jones, 2017; Tobback et al., 2017) ou mesmo de informações que tratam os sentimentos dos analistas (Hajek; Olej; Myskova, 2014), que são questões encorajadoras e podem ser aplicadas no mercado de capitais brasileiro ou até mesmo na regionalidade. Sugere-se ainda que sejam analisados os padrões e as características das empresas que se encontram na área cinzenta, para que os modelos fiquem ainda mais assertivos, verificando se existem características comuns que possam ser atribuídas ao modelo para melhora das probabilidades.

Os resultados desta dissertação são interessantes para os profissionais da área de finanças, instituições financeiras e gestores das organizações, que podem analisar tanto o modelo, quanto a relevância das variáveis que contribuem no poder de predição e precisão, bem como a redução dos custos referentes aos erros de classificação. Os achados deste estudo são reforçados por resultados anteriores, aplicados em outros mercados e que podem aprimorar a avaliação do risco de crédito, na redução de perdas financeiras, na perda de vendas em decorrência do não atendimento por falha de classificação, na redução de custos relacionados a recuperação dos créditos financeiros, custos operacionais relacionados ao processo de classificação manual, e portanto, contribuindo para a maximização dos resultados do negócio.

CAPÍTULO IV – CONCLUSÕES

Os mecanismos de previsão de dificuldades financeiras e falências passam por constantes transformações em decorrência das crises financeiras, da globalização dos mercados e da evolução dos sistemas de informação. Os modelos de aprendizado de máquina apresentam bons resultados de desempenho e diversas pesquisas procuram verificar aquele modelo ou técnica que possui melhor precisão e predição, com menores erros de classificação, para isso, fazem testes com modelos individuais, conjuntos ou híbridos para tratar o risco de crédito.

Esta dissertação incorporou uma avaliação das pesquisas acadêmicas com as múltiplas visões sobre risco de crédito com enfoque em dificuldades financeiras e falência, classificação de crédito e inteligência artificial, demonstrando que houve uma evolução de estudos ao longo dos anos, principalmente após a crise financeira ocorrida em meados de 2007/2008. A revisão de literatura realizada, permitiu o desenvolvimento e a comparação de dois modelos, utilizando mecanismos de aprendizado de máquina, RF, e estatística tradicional, LR, para prever previamente o fenômeno de insolvência em empresas de capital aberto brasileiras, avaliando o desempenho destes dois modelos a nível nacional e também nas empresas da região de interesse (mesorregião do Triângulo Mineiro e Alto Paranaíba e do Sul Goiano).

Em relação ao poder de predição, ACC, os resultados apresentaram que o modelo estatístico tradicional de LR possui capacidade preditiva menor quando comparado com o modelo RF, tanto no cenário nacional (87,2% e 90,8%) quanto no regional (89,9% e 92,6%). Em relação ao TE II, o modelo de LR, nacional e regional, (12,5% e 32,1%) melhora a predição, mas piora nas empresas da região, o mesmo acontece com o modelo RF (5,6% e 14,3%) também apresenta avanço na predição, mas também piora o TE II. No TE I o modelo RF, nacional e regional, (10,3% e 6,2%) melhora na predição nas empresas da região e o modelo LR (12,9% e 6,2%) também melhora na predição das empresas da região e se iguala ao modelo de aprendizado de máquina. Estes resultados são robustos a outras variáveis apresentadas pela literatura como fatores relevantes para compreender o aprimoramento na previsão de dificuldades financeiras e falência utilizando modelos de aprendizado de máquina e estatística tradicional. Como principais resultados, as hipóteses centrais desta dissertação foram confirmadas onde o modelo RF, baseado em ML, superou o modelo estatístico tradicional LR, mantendo este comportamento tanto no cenário nacional quanto no cenário regional.

Novas pesquisas podem implementar outras técnicas e modelos, de conjuntos ou híbridos, replicando outros trabalhos com dados de empresas brasileiras. Apesar de que, existe uma dificuldade de se conseguir informações de empresas brasileiras, salvo do mercado de

capitais. Assim, abrindo um leque de oportunidade na área de risco de crédito dessa economia emergente que é referência na América Latina e no mundo.

REFERÊNCIAS

- ABDOU, H. A.; POINTON, J. Credit scoring, statistical techniques and evaluation criteria: a review of the literature. **Intelligent Systems in Accounting, Finance and Management**, v. 182, n. 3, p. 59–88, 2011. doi.org/10.1002/isaf.325
- ABELLÁN, J.; CASTELLANO, J. G. A comparative study on base classifiers in ensemble methods for credit scoring. **Expert Systems with Applications**, v. 73, p. 1–10, 2017. doi:10.1016/j.eswa.2016.12.020
- ABELLÁN, J.; MANTAS, C. J. Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. **Expert Systems with Applications**, v. 41, n. 8, p. 3825–3830, 2014. doi:10.1016/j.eswa.2013.12.003
- ADDO, P. M.; GUEGAN, D.; HASSANI, B. Credit Risk Analysis Using Machine and Deep Learning Models. **Risks**, v. 6, n. 2, p. 38, 2018. doi.org/10.3390/risks6020038
- ADNAN AZIZ, M., DAR, H. A. Predicting corporate bankruptcy: where we stand? **Corporate Governance: The International Journal of Business in Society**, v. 6, n. 1, p. 18–33, 2006. doi:10.1108/14720700610649436
- AGARWAL, V.; TAFFLER, R. Comparing the performance of market-based and accounting-based bankruptcy prediction models. **Journal of Banking & Finance**, v. 32, n. 8, p. 1541–1551, 2008. doi.org/10.1016/j.jbankfin.2007.07.014
- AGHAEIRAD, A.; CHEN, N.; RIBEIRO, B. Improve credit scoring using transfer of learned knowledge from self-organizing map. **Neural Computing and Applications**, v. 28, n. 6, p. 1329–1342, 2016. doi:10.1007/s00521-016-2567-2
- AHN, H.; KIM, K. Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach. **Applied Soft Computing**, v. 9, n. 2, p. 599–607, 2009. doi:10.1016/j.asoc.2008.08.002
- ALAKA, H. A.; OYEDELE, L. O.; OWOLABI, H. A.; KUMAR, V.; AJAYI, S. O.; AKINADE, O. O.; BILAL, M. Systematic review of bankruptcy prediction models: Towards a framework for tool selection. **Expert Systems with Applications**, v. 94, p. 164–184, 2018. doi:10.1016/j.eswa.2017.10.040
- ALA’RAJ, M.; ABBOD, M. F. Classifiers consensus system approach for credit scoring. **Knowledge-Based Systems**, v. 104, p. 89–105, 2016. doi:10.1016/j.knosys.2016.04.013
- ALA’RAJ, M.; ABBOD, M. F. A new hybrid ensemble credit scoring model based on classifiers consensus system approach. **Expert Systems with Applications**, v. 64, p. 36–55, 2016. doi:10.1016/j.eswa.2016.07.017
- ALESSANDRINI, S.; SPERATI, S.; PINSON, P. A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data. **Applied energy**, v. 107, p. 271-280, 2013. doi.org/10.1016/j.apenergy.2013.02.041

ALFARO, E.; GARCÍA, N.; GÁMEZ, M.; ELIZONDO, D. Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. **Decision Support Systems**, v. 45, n. 1, p. 110–122, 2008. doi:10.1016/j.dss.2007.12.002

ALTMAN, E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. **The Journal of Finance**, v. 23, n. 4, p. 589-609, 1968.
<https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>

ALTMAN, E. I. Managing Credit Risk: A Challenge for the New Millennium. **Economic Notes**, v. 31, n. 2, p. 201–214, 2002. doi:10.1111/1468-0300.00084

ALTMAN, E. I.; BAIDYA, T. K. N.; DIAS, L. M. R. Assessing Potential Financial Problems for Firms in Brazil. **Journal of International Business Studies**, v. 10, n. 2, p. 9-24, 1979. doi.org/10.1057/palgrave.jibs.8490787

ALTMAN, E. I.; DANOVİ, A.; FALINI, A. Z-Score Models' application to Italian companies subject to extraordinary administration. **Journal of Applied Finance (Formerly Financial Practice and Education)**, v. 23, n. 1, 2013.

ALTMAN, E. I.; MARCO, G.; VARETTO, F. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks the Italian experience. **Journal of Banking Finance**, v. 18, n. 3, p. 505–529, 1994. doi:10.1016/0378-4266(94)90007-8

ANTUNES, F.; RIBEIRO, B.; PEREIRA, F. Probabilistic modeling and visualization for bankruptcy prediction. **Applied Soft Computing**, v. 60, p. 831–843, 2017.
doi:10.1016/j.asoc.2017.06.043

ASSOCIAÇÃO BRASILEIRA DAS ENTIDADES DOS MERCADOS FINANCEIRO E DE CAPITAIS – ANBIMA. **Basileia III no Brasil**. disponível em:
<http://www.anbima.com.br/pt_br/informar/regulacao/informe-de-legislacao/basileia-iii-no-brasil.htm>. Acesso em: mar. 2019.

Brasil Bolsa e Balcão – B3. **Índice Bovespa**. Disponível em:
<<http://www.bmfbovespa.com.br>>. Acesso em: mar. 2019

BAHRAMMIRZAEE, A. A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. **Neural Computing and Applications**, v. 19, n. 8, p. 1165–1195, 2010. doi:10.1007/s00521-010-0362-z

BALCAEN, S.; OOGHE, H. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. **The British Accounting Review**, v. 38, n. 1, p. 63–93. 2006. doi.org/10.1016/j.bar.2005.09.001

BANCO CENTRAL DO BRASIL – BACEN. Disponível em:
<https://www.bcb.gov.br/fis/supervisao/basileia.asp>. Acesso em: mar. 2019

BARBOZA, F.; KIMURA, H.; ALTMAN, E. Machine learning models and bankruptcy prediction. **Expert Systems with Applications**, v. 83, p. 405–417, 2017.
doi:10.1016/j.eswa.2017.04.006

- BEAVER, W. H. Financial ratios as predictors of failure. **Journal of Accounting Research**, v. 4, Empirical Research in Accounting: Selected, p. 71-111, 1966. doi: 10.2307/2490171
- BECKER, C. L.; DEFOND, M. L.; JIAMBALVO, J.; SUBRAMANYAM, K. R; The effect of audit quality on earnings management. **Contemporary accounting research**, v. 15, n. 1, p. 1-24, 1998. doi.org/10.1111/j.1911-3846.1998.tb00547.x
- BEGLEY, J.; MING, J.; WATTS, S. Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models. **Review of accounting Studies**, v. 1, n. 4, p. 267-284, 1996. doi.org/10.1007/BF00570833
- BEKHET, H. A.; ELETTER, S. F. K. Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. **Review of Development Finance**, v. 4, n. 1, p. 20–28, 2014. doi:10.1016/j.rdf.2014.03.002
- BELLOVARY, J. L.; GIACOMINO, D. E.; AKERS, M. D. A review of bankruptcy prediction studies: 1930 to present. **Journal of Financial Education**, v. 33, p. 1-42, 2007. doi: tw/10.2307/41948574
- BEQUÉ, A.; LESSMANN, S. Extreme learning machines for credit scoring: An empirical evaluation. **Expert Systems with Applications**, v. 86, p. 42–53, 2017. doi:10.1016/j.eswa.2017.05.050
- BERNARDES, F. F.; FERREIRA, W. R. A logística em transporte no Triângulo Mineiro e Alto Paranaíba: operacionalizando os sistemas agrícolas. **Observatorium: Revista Eletrônica de Geografia**, v. 5, n. 13, 2013.
- BEYNON, M. J.; PEEL, M. J. Variable precision rough set theory and data discretisation: an application to corporate failure prediction. **Omega**, v. 29, n. 6, p. 561–576, 2001. doi:10.1016/s0305-0483(01)00045-7
- BORGMAN, C. L. Bibliometrics and scholarly communication: Editor's introduction. **Communication Research**, v. 16, n. 5, p. 583-599, 1989. doi: 10.1177/009365089016005002
- BOSE, I. Deciding the financial health of dot-coms using rough sets. **Information & Management**, v. 43, n. 7, p. 835–846, 2006. doi:10.1016/j.im.2006.08.001
- BOYACIOGLU, M. A.; KARA, Y.; BAYKAN, Ö. K. Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. **Expert Systems with Applications**, v. 36, n. 2, p. 3355–3366, 2009. doi:10.1016/j.eswa.2008.01.003
- BREALEY, R. A; MYERS, S. C. **Principles of financial management**. New York: McGraw-Hill, 1996.
- BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001. doi.org/10.1023/A:1010933404324

- BRIER, G.W. Verification of forecasts expressed in terms of probability. **Monthly weather review**, v. 78, n. 1, p.1-3, 1950. doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2
- BROCKETT, P. L.; COOPER, W. W.; GOLDEN, L. L.; PITAKTONG, U. A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency. **The Journal of Risk and Insurance**, v. 61, n. 3, p. 402-402, 1994. doi:10.2307/253568
- BROWN, I.; MUES, C. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. **Expert Systems with Applications**, v. 39, n. 3, p. 3446–3453, 2012. doi:10.1016/j.eswa.2011.09.033
- CASEY, C.; BARTCZAK, N. Using operating cash flow data to predict financial distress: Some extensions. **Journal of Accounting Research**, p. 384-401, 1985. doi: 10.2307/2490926
- CHANDRA, D. K.; RAVI, V.; BOSE, I. Failure prediction of dotcom companies using hybrid intelligent techniques. **Expert Systems with Applications**, v. 36, n. 3, p. 4830–4837, 2009. doi:10.1016/j.eswa.2008.05.047
- CHANG, T.-M.; HSU, M.-F. Integration of incremental filter-wrapper selection strategy with artificial intelligence for enterprise risk management. **International Journal of Machine Learning and Cybernetics**, v. 9, n. 3, p. 477–489, 2016. doi:10.1007/s13042-016-0545-8
- CHAUDHURI, A.; DE, K. Fuzzy Support Vector Machine for bankruptcy prediction. **Applied Soft Computing**, v. 11, n. 2, p. 2472–2486, 2011. doi:10.1016/j.asoc.2010.10.003
- CHAWLA, N. V; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, v. 16, p. 321-357, 2002. doi.org/10.1613/jair.953
- CHEN, F.-L.; LI, F.-C. Combination of feature selection approaches with SVM in credit scoring. **Expert Systems with Applications**, v. 37, n. 7, p. 4902–4909, 2010. doi:10.1016/j.eswa.2009.12.025
- CHEN, M.-Y. Predicting corporate financial distress based on integration of decision tree classification and logistic regression. **Expert Systems with Applications**, v. 38, n. 9, p. 11261–11272, 2011. doi:10.1016/j.eswa.2011.02.173
- CHEN, N.; RIBEIRO, B.; CHEN, A. Financial credit risk assessment: a recent review. **Artificial Intelligence Review**, v. 45, n. 1, p. 1–23, 2015. doi:10.1007/s10462-015-9434-x
- CHEN, N.; RIBEIRO, B.; VIEIRA, A.; CHEN, A. Clustering and visualization of bankruptcy trajectory using self-organizing map. **Expert Systems with Applications**, v. 40, n.1, p. 385–393, 2013. doi:10.1016/j.eswa.2012.07.047
- CHEN, N.; RIBEIRO, B.; VIEIRA, A. S.; DUARTE, J.; NEVES, J. C. A genetic algorithm-based approach to cost-sensitive bankruptcy prediction. **Expert Systems with Applications**, v. 38, n. 10, p. 12939–12945, 2011. doi:10.1016/j.eswa.2011.04.090

CHEN, W.-H.; SHIH, J.-Y. A study of Taiwan's issuer credit rating systems using support vector machines. **Expert Systems with Applications**, v. 30, n. 3, p. 427–435, 2006. doi:10.1016/j.eswa.2005.10.003

CHEN, W.-S.; DU, Y.-K. Using neural networks and data mining techniques for the financial distress prediction model. **Expert Systems with Applications**, v. 36, n. 2, p. 4075–4086, 2009. doi:10.1016/j.eswa.2008.03.020

CHEN, Y.-S.; CHENG, C.-H. Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry. **Knowledge-Based Systems**, v. 39, p. 224–239, 2013. doi:10.1016/j.knosys.2012.11.004

CHENG, C.-B.; CHEN, C.-L.; FU, C.-J. Financial distress prediction by a radial basis function network with logit analysis learning. **Computers & Mathematics with Applications**, v. 51, n. 3-4, p. 579-588, 2006. doi.org/10.1016/j.camwa.2005.07.016

CHENG, C.; JONES, S.; MOSER, W. J. Abnormal trading behavior of specific types of shareholders before US firm bankruptcy and its implications for firm bankruptcy prediction. **Journal of Business Finance & Accounting**, v. 45, n. 9-10, p. 1100-1138, 2018. doi:10.1111/jbfa.12338

CHO, S.; KIM, J.; BAE, J. K. An integrative model with subject weight based on neural network learning for bankruptcy prediction. **Expert Systems with Applications**, v. 36, n. 1, p. 403–410, 2009. doi:10.1016/j.eswa.2007.09.060

CHOU, C.-H.; HSIEH, S.-C.; QIU, C.-J. Hybrid genetic algorithm and fuzzy clustering for bankruptcy prediction. **Applied Soft Computing**, v. 56, p. 298–316, 2017. doi:10.1016/j.asoc.2017.03.014

CHUANG, C.-L. Application of hybrid case-based reasoning for enhanced performance in bankruptcy prediction. **Information Sciences**, v. 236, p. 174–185, 2013. doi:10.1016/j.ins.2013.02.015

CIELEN, A.; PEETERS, L.; VANHOOF, K. Bankruptcy prediction using a data envelopment analysis. **European Journal of Operational Research**, v. 154, n. 2, p. 526–532, 2004. doi:10.1016/s0377-2217(03)00186-3

COLE, F. J.; EALES, N. B. The history of comparative anatomy. Part I—A statistical analysis of the literature. **Science Progress** (1916-1919), v. 11, n. 44, p. 578–96, 1917. doi:10.2307/43426882.

COPELAND, T. E.; WESTON, J. F. **Financial Theory and Corporate Policy**. Mass. Addison-Wesley Publishing Company, p. 464-471 1988.

COMISSÃO DE VALORES MOBILIÁRIOS - CVM. **Empresas Listadas**. Disponível em: <<http://sistemas.cvm.gov.br/asp/cvmwww/ciasabertas/formciaaberta.asp>>. Acesso em: abr. 2019

DANENAS, P.; GARSVA, G. Selection of Support Vector Machines based classifiers for credit risk domain. **Expert Systems with Applications**, v. 42, n. 6, p. 3194–3204, 2015. doi:10.1016/j.eswa.2014.12.001

DEAKIN, E. B. A discriminant analysis of predictors of business failure. **Journal of accounting research**, p. 167-179, 1972. doi:10.2307/2490225

DE ANDRES, J.; LANDAJO, M.; LORCA, P. Forecasting business profitability by using classification techniques: A comparative analysis based on a Spanish case. **European Journal of Operational Research**, v. 167, n. 2, p. 518–542, 2005. doi.org/10.1016/j.ejor.2004.02.018

DE ANDRÉS, J.; LORCA, P.; DE COS JUEZ, F. J.; SÁNCHEZ-LASHERAS, F. Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). **Expert Systems with Applications**, v. 38, n. 3, p. 1866–1875, 2011. doi:10.1016/j.eswa.2010.07.117

DE AZEVEDO, R. C.; ENSSLIN, L.; JUNGLES, A. E. A review of risk management in construction: opportunities for improvement. **Modern Economy**, v. 5, n. 4, p. 367, 2014. doi:10.4236/me.2014.54036

DE CASTRO VIEIRA, J. R.; BARBOZA, F.; SOBREIRO, V. A.; KIMURA, H. Machine learning models for credit analysis improvements: Predicting low-income families' default. **Applied Soft Computing**, v. 83, p. 105640, 2019. doi.org/10.1016/j.asoc.2019.105640

DIMITRAS, A. I.; ZANAKIS, S. H.; ZOPOUNIDIS, C. A survey of business failures with an emphasis on prediction methods and industrial applications. **European Journal of Operational Research**, v. 90, n. 3, p. 487–513, 1996. doi:10.1016/0377-2217(95)00070-4

DOUMPOS, M.; ZOPOUNIDIS, C. A Multicriteria Outranking Modeling Approach for Credit Rating. **Decision Sciences**, v. 42, n. 3, p. 721–742, 2011. doi:10.1111/j.1540-5915.2011.00328.x

DU JARDIN, P. A two-stage classification technique for bankruptcy prediction. **European Journal of Operational Research**, v. 254, n. 1, p. 236–252, 2016. doi:10.1016/j.ejor.2016.03.008

DU JARDIN, P. Dynamics of firm financial evolution and bankruptcy prediction. **Expert Systems with Applications**, v. 75, p. 25–43, 2017. doi:10.1016/j.eswa.2017.01.016

DU JARDIN, P. Failure pattern-based ensembles applied to bankruptcy forecasting. **Decision Support Systems**, v. 107, p. 64–77, 2018. doi:10.1016/j.dss.2018.01.003

DU JARDIN, P.; SÉVERIN; E. Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. **European Journal of Operational Research**, v. 221, n. 2, p. 378–396, 2012. doi:10.1016/j.ejor.2012.04.006

ENSSLIN, L.; ENSSLIN, S. R.; DUTRA, A.; NUNES, N. A.; REIS, C. BPM governance: a literature analysis of performance evaluation. **Business Process Management Journal**, v. 23, n. 1, p. 71–86, 2017. doi.org/10.1108/BPMJ-11-2015-0159

- ETHERIDGE, H. L.; SRIRAM, R. S.; HSU, H. Y. K. A Comparison of Selected Artificial Neural Networks that Help Auditors Evaluate Client Financial Viability. **Decision Sciences**, v. 31, n. 2, p. 531–550, 2000. doi:10.1111/j.1540-5915.2000.tb01633.x
- FAMA, E. F. Efficient Capital Markets: a review of theory and empirical work, **Journal of Finance**, v. 25, n. 2; p. 383-417, 1970. doi.org/10.1111/j.1540-6261.1970.tb00518.x
- FAMA, E. F. Term premiums and defaults premiums in money markets. **Journal of Financial Economics**, v. 17, n. 1, p. 175-196, 1986. doi.org/10.1016/0304-405X(86)90010-3
- FEDOROVA, E.; GILENKO, E.; DOVZHENKO, S. Bankruptcy prediction for Russian companies: Application of combined classifiers. **Expert Systems with Applications**, v. 40, n. 18, p. 7285–7293, 2013. doi:10.1016/j.eswa.2013.07.032
- FENG, X.; XIAO, Z.; ZHONG, B.; DONG, Y.; QIU, J. Dynamic weighted ensemble classification for credit scoring using Markov Chain. **Applied Intelligence**, v. 49, n. 2, p. 555-568, 2019. doi:10.1007/s10489-018-1253-8
- FENG, X.; XIAO, Z.; ZHONG, B.; QIU, J.; DONG, Y. Dynamic ensemble classification for credit scoring using soft probability. **Applied Soft Computing**, v. 65, p. 139–151, 2018. doi:10.1016/j.asoc.2018.01.021
- FERNÁNDEZ, M. T. T.; GUTIÉRREZ, F. J. C. Variables y Modelos Para La Identificación y Predicción Del Fracaso Empresarial: Revisión de La Investigación Empírica Reciente. **Revista de Contabilidad**, v. 15, n. 1, p. 7–58, 2012. doi:10.1016/s1138-4891(12)70037-7
- FINLAY, S. Multiple classifier architectures and their application to credit risk assessment. **European Journal of Operational Research**, v. 210, n. 2, p. 368–378, 2011. doi:10.1016/j.ejor.2010.09.029
- FLOREZ-LOPEZ, R. Modelling of insurers' rating determinants. An application of machine learning techniques and statistical models. **European Journal of Operational Research**, v. 183, n. 3, p. 1488–1512, 2007. doi:10.1016/j.ejor.2006.09.103
- 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**, v. 44, n. 1, p. 159–189, 2014. doi:10.1007/s10844-014-0333-4
- GARCÍA, V.; MARQUÉS, A. I.; SÁNCHEZ, J. Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction. **Information Fusion**, v. 47, p. 88-101, 2019. doi.org/10.1016/j.inffus.2018.07.004
- GARCÍA, V.; MARQUÉS, A. I.; SÁNCHEZ, J. S.; OCHOA-DOMÍNGUEZ, H. J. Dissimilarity-Based Linear Models for Corporate Bankruptcy Prediction. **Computational Economics**, v. 53, n. 3, p. 1019-1031, 2019. doi:10.1007/s10614-017-9783-4
- GENG, R.; BOSE, I.; CHEN, X. Prediction of financial distress: An empirical study of listed Chinese companies using data mining. **European Journal of Operational Research**, v. 241, n. 1, p. 236–247, 2015. doi.org/10.1016/j.ejor.2014.08.016

GLANZEL, W. Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies. **Library Trends**, v. 50, n. 3, p. 461-473, 2002.

GOMBOLA, M. J.; KETZ, J. E. A note on cash flow and classification patterns of financial ratios. **Accounting Review**, p. 105-114, 1983.

GORZAŁCZANY, M. B.; RUDZIŃSKI, F. A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability. **Applied Soft Computing**, v. 40, p. 206–220, 2016. doi:10.1016/j.asoc.2015.11.037

GRiffin, J. M.; LEMMON, M. L. Book-to-market equity, distress risk, and stock returns. **The Journal of Finance**, v. 57, n. 5, p. 2317-2336, 2002. doi.org/10.1111/1540-6261.00497

GRILLI, R.; TEDESCHI, G.; GALLEGATI, M. Markets connectivity and financial contagion. **Journal of Economic Interaction and Coordination**, v. 10, n. 2, p. 287-304, 2015. doi.org/10.1007/s11403-014-0129-1

GUO, Y.; ZHOU, W.; LUO, C.; LIU, C.; XIONG, H. Instance-based credit risk assessment for investment decisions in P2P lending. **European Journal of Operational Research**, v. 249, n. 2, p. 417-426, 2016. doi.org/10.1016/j.ejor.2015.05.050

HAIR, J.F.; BLACK, W.C.; BABIN, B.J; ANDERSON, R.E. **Multivariate Data Analysis**. 7 ed., v. 1, Londres: Prentice-Hall, Pearson, 2014.

HAJEK, P.; MICHALAK, K. Feature selection in corporate credit rating prediction. **Knowledge-Based Systems**, v. 51, p. 72–84, 2013. doi:10.1016/j.knosys.2013.07.008

HAJEK, P.; OLEJ, V.; MYSKOVA, R. Forecasting Corporate Financial Performance using sentiment in annual reports for stakeholders' decision-making. **Technological and Economic Development of Economy**, v. 20, n. 4, p. 721–738, 2014. doi:10.3846/20294913.2014.979456

HÄRDLE, W.; LEE, Y.-J.; SCHÄFER, D.; YEH, Y.-R. Variable selection and oversampling in the use of smooth support vector machines for predicting the default risk of companies. **Journal of Forecasting**, v. 28, n. 6, p. 512–534, 2009. doi:10.1002/for.1109

HEO, J.; YANG, J. Y. AdaBoost based bankruptcy forecasting of Korean construction companies. **Applied Soft Computing**, v. 24, p. 494–499, 2014. doi:10.1016/j.asoc.2014.08.009

HILLEGEIST, S. A.; KEATING, E. K.; CRAM, D. P.; LUNDSTEDT, K. G. Assessing the probability of bankruptcy. **Review of accounting studies**, v. 9, n. 1, p. 5-34, 2004. doi.org/10.1023/B:RAST.0000013627.90884.b7

HORTA, R. A. M.; ALVES, F. J. S.; CARVALHO, F. A. A. Seleção de atributos na previsão de insolvência: aplicação e avaliação usando dados brasileiros recentes. **Revista de Administração Mackenzie**, v. 15, n. 1, p. 125-151, 2013.

- HSIEH, T.-J.; HSIAO, H.-F.; YEH, W.-C. Mining financial distress trend data using penalty guided support vector machines based on hybrid of particle swarm optimization and artificial bee colony algorithm. **Neurocomputing**, v. 82, p. 196–206, 2012. doi:10.1016/j.neucom.2011.11.020
- HU, Y.-C.; ANSELL, J. Measuring retail company performance using credit scoring techniques. **European Journal of Operational Research**, v. 183, n. 3, 1595–1606, 2007. doi:10.1016/j.ejor.2006.09.101
- HU, Y.-C. Bankruptcy prediction using ELECTRE-based single-layer perceptron. **Neurocomputing**, v. 72, n.13-15, p. 3150–3157, 2009. doi:10.1016/j.neucom.2009.03.002
- HUANG, S.-M.; TSAI, C.-F.; YEN, D. C.; CHENG, Y.-L. A hybrid financial analysis model for business failure prediction. **Expert Systems with Applications**, v. 35, n. 3, p. 1034–1040, 2008. doi:10.1016/j.eswa.2007.08.040
- HUANG, Z.; CHEN, H.; HSU, C.-J.; CHEN, W.-H.; WU, S. Credit rating analysis with support vector machines and neural networks: a market comparative study. **Decision Support Systems**, v. 37, n. 4, p. 543–558, 2004. doi:10.1016/s0167-9236(03)00086-1
- HUANG, J.-J.; TZENG, G.-H.; ONG, C.-S. Two-stage genetic programming 2SGP for the credit scoring model. **Applied Mathematics and Computation**, v. 174, n. 2, p. 1039–1053, 2006. doi.org/10.1016/j.amc.2005.05.027
- HUNG, C.; CHEN; J.-H. A selective ensemble based on expected probabilities for bankruptcy prediction. **Expert Systems with Applications**, v. 36, n. 3, p. 5297–5303, 2009. doi:10.1016/j.eswa.2008.06.068
- INSTITUTO BRASILEIRO DE GEROGRAFIA E ESTATÍSTICA – IBGE. **Divisão Territorial Brasileira – DTB**. Disponível em: <<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/divisao-regional/23701-divisao-territorial-brasileira.html?=&t=downloads>>. Acesso em: dez. 2019.
- INSTITUTO BRASILEIRO DE GEROGRAFIA E ESTATÍSTICA – IBGE. SIDRA. Produto Interno Bruto. Disponível em: <<https://cidades.ibge.gov.br/brasil/mg/joao-pinheiro/pesquisa/38/46996>>. Acesso em: fev. 2019.
- JO, H.; HAN, I. Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction. **Expert Systems with Applications**, v. 11, n. 4, p. 415–422, 1996. doi:10.1016/s0957-4174(96)00056-5
- JONES, S.; JOHNSTONE, D.; WILSON, R. An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes. **Journal of Banking & Finance**, v. 56, p. 72–85, 2015. doi:10.1016/j.jbankfin.2015.02.006
- JONES, S.; JOHNSTONE, D.; WILSON, R. Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks. **Journal of Business Finance & Accounting**, v. 44, n. 1-2, p. 3–34, 2016. doi:10.1111/jbfa.12218

JONES, S. Corporate bankruptcy prediction: a high dimensional analysis. **Review of Accounting Studies**, v. 22, n. 3, p. 1366–1422, 2017. doi:10.1007/s11142-017-9407-1

JOURNAL CITATION REPORTS-JCR. InCites Journal Citation Reports. Disponível em <https://jcr-clarivate.ez34.periodicos.capes.gov.br/JCRJournalHomeAction.action?>. Acesso em jun. 2019.

KARAN, M. B.; ULUCAN, A.; KAYA, M. Credit risk estimation using payment history data: a comparative study of Turkish retail stores. **Central European Journal of Operations Research**, v. 21, n. 2, p. 479–494, 2012. doi:10.1007/s10100-012-0242-y

KIM, K.; AHN, H. A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach. **Computers & Operations Research**, v. 39, n. 8, p. 1800–1811, 2012. doi:10.1016/j.cor.2011.06.023

KIM, M.-J.; HAN, I. The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms. **Expert Systems with Applications**, v. 25, n. 4, p. 637–646, 2003. doi:10.1016/s0957-4174(03)00102-7

KIM, M.-J.; KANG, D.-K.; KIM, H. B. Geometric mean based boosting algorithm with oversampling to resolve data imbalance problem for bankruptcy prediction. **Expert Systems with Applications**, v. 42, n. 3, p. 1074–1082, 2015. doi:10.1016/j.eswa.2014.08.025

KIM, M.-J.; KANG, D.-K. Ensemble with neural networks for bankruptcy prediction. **Expert Systems with Applications**, v. 37, n. 4, p. 3373–3379, 2010. doi:10.1016/j.eswa.2009.10.012

KIM, M.-J.; KANG, D.-K. Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction. **Expert Systems with Applications**, v. 39, n. 10, p. 9308–9314, 2012. doi:10.1016/j.eswa.2012.02.072

KIM, S. Y.; UPNEJA, A. Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. **Economic Modelling**, v. 36, p. 354–362, 2014. doi:10.1016/j.econmod.2013.10.005

KO, P.-C.; LIN, P.-C. An evolution-based approach with modularized evaluations to forecast financial distress. **Knowledge-Based Systems**, v. 19, n. 1, p. 84–91, 2006. doi:10.1016/j.knosys.2005.11.006

KOROL, T. Early warning models against bankruptcy risk for Central European and Latin American enterprises. **Economic Modelling**, v. 31, p. 22–30, 2013. doi:10.1016/j.econmod.2012.11.017

KOU, G.; PENG, Y.; LU, C. MCDM approach to evaluating bank loan default models. **Technological and Economic Development of Economy**, v. 20, n. 2, p. 292–311, 2014. doi:10.3846/20294913.2014.913275

KOUTANAEI, F. N.; SAJEDI, H.; KHANBABAEI, M. A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. **Journal of Retailing and Consumer Services**, v. 27, p. 11–23, 2015. doi:10.1016/j.jretconser.2015.07.003

KOYUNCUGIL, A. S.; OZGULBAS, N. Financial early warning system model and data mining application for risk detection. **Expert Systems with Applications**, v. 39, n. 6, p. 6238–6253, 2012. doi:10.1016/j.eswa.2011.12.021

KWAK, W.; SHI, Y.; ELDRIDGE, S. W.; KOU, G. Bankruptcy prediction for Japanese firms: using Multiple Criteria Linear Programming data mining approach. **International Journal of Business Intelligence and Data Mining**, v. 1, n. 4, p. 401-416, 2006. doi:10.1504/ijbidm.2006.010782

LAHSASNA, A.; AINON, R. N.; TEH, Y. W. Credit Scoring Models Using Soft Computing Methods: A Survey. **Int. Arab J. Inf. Technol.**, v. 7, n. 2, p. 115-123, 2010. doi:10.1109/ICCCE.2008.4580635

LEE, K. C.; HAN, I.; KWON, Y. Hybrid neural network models for bankruptcy predictions. **Decision Support Systems**, v. 18, n. 1, p. 63–72, 1996. doi:10.1016/0167-9236(96)00018-8

LEE, Y.-C. Application of support vector machines to corporate credit rating prediction. **Expert Systems with Applications**, v. 33, n. 1, p. 67–74, 2007. doi:10.1016/j.eswa.2006.04.018

LESSMANN, S.; BAESENS, B.; SEOW, H.-V.; THOMAS, L. C. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. **European Journal of Operational Research**, v. 247, n. 1, p. 124–136, 2015. doi:10.1016/j.ejor.2015.05.030

LI, H.; SUN, J.; SUN, B.L. Financial distress prediction based on OR-CBR in the principle of k-nearest neighbors. **Expert Systems with Applications**, v. 36, n. 1, p. 643–659, 2009. doi:10.1016/j.eswa.2007.09.038

LI, H.; SUN, J. Hybridizing principles of the Electre method with case-based reasoning for data mining: Electre-CBR-I and Electre-CBR-II. **European Journal of Operational Research**, v. 197, n. 1, p. 214–224, 2009. doi:10.1016/j.ejor.2008.05.024

LI, H.; SUN, J. Business failure prediction using hybrid2 case-based reasoning (H2CBR). **Computers & Operations Research**, v. 37, n. 1, p. 137–151, 2010. doi:10.1016/j.cor.2009.04.003

LI, H.; SUN, J. Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples – Evidence from the Chinese hotel industry. **Tourism Management**, v. 33, n. 3, p. 622–634, 2012. doi:10.1016/j.tourman.2011.07.004

LI, H.; SUN, J. Principal component case-based reasoning ensemble for business failure prediction. **Information & Management**, v. 48, n. 6, p. 220–227, 2011. doi:10.1016/j.im.2011.05.001

LI, F.; PEREZ-SAIZ, H. Measuring Systemic Risk Across Financial Markets Infrastructures. **Journal of Financial Stability**, v. 34, p. 1-11, 2018. doi.org/10.1016/j.jfs.2017.08.003

LIANG, D.; LU, C.-C.; TSAI, C.-F.; SHIH, G.-A. Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. **European Journal of Operational Research**, v. 252, n. 2, p. 561–572, 2016. doi:10.1016/j.ejor.2016.01.012

LIANG, D.; TSAI, C.-F.; DAI, A.-J.; EBERLE, W. A novel classifier ensemble approach for financial distress prediction. **Knowledge and Information Systems**, v. 54, n. 2, p. 437–462, 2017. doi:10.1007/s10115-017-1061-1

LIANG, D.; TSAI, C.-F.; WU, H.-T. The effect of feature selection on financial distress prediction. **Knowledge-Based Systems**, v. 73, p. 289–297, 2015. doi:10.1016/j.knosys.2014.10.010

LIAW, A.; WIENER, M. Classification and regression by Random Forest. **R news**, v. 2, n. 3, p. 18-22, 2002.

LIN, F.; LIANG, D.; CHEN, E. Financial ratio selection for business crisis prediction. **Expert Systems with Applications**, v. 38, n. 12, p. 15094–15102, 2011. doi:10.1016/j.eswa.2011.05.035

LIN, F. Y.; MCCLEAN, S. A data mining approach to the prediction of corporate failure. **Knowledge-Based Systems**, v. 14, n 3-4, p. 189–195, 2001. doi:10.1016/s0950-7051(01)00096-x

LIN, R.-H.; WANG, Y.-T.; WU, C.-H.; CHUANG, C.-L. Developing a business failure prediction model via RST, GRA and CBR. **Expert Systems with Applications**, v. 36, n. 2, p. 1593–1600, 2009. doi:10.1016/j.eswa.2007.11.068

LIN, S.-J.; CHANG, C.; HSU, M.-F. Multiple extreme learning machines for a two-class imbalance corporate life cycle prediction. **Knowledge-Based Systems**, v. 39, p. 214–223, 2013. doi:10.1016/j.knosys.2012.11.003

LIN, S.-W.; SHIUE, Y.-R.; CHEN, S.-C.; CHENG, H.-M. Applying enhanced data mining approaches in predicting bank performance: A case of Taiwanese commercial banks. **Expert Systems with Applications**, v. 36, n. 9, p.11543–11551, 2009. doi:10.1016/j.eswa.2009.03.029

LIN, W.-C.; LU, Y.-H.; TSAI, C.-F. Feature selection in single and ensemble learning-based bankruptcy prediction models. **Expert Systems**, v. 36, n. 1, p. e12335, 2018. doi:10.1111/exsy.12335

LIN, W. Y.; HU, Y. H.; TSAI, C. F. Machine Learning in Financial Crisis Prediction: A Survey. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 42, n. 4, p. 421–436, 2012. doi:10.1109/tsmcc.2011.2170420

LIOU, F. Fraudulent financial reporting detection and business failure prediction models: a comparison. **Managerial Auditing Journal**, v. 23, n. 7, p. 650–662, 2008. doi:10.1108/02686900810890625

- LIU, Y.; SCHUMANN, M. Data mining feature selection for credit scoring models. **Journal of the Operational Research Society**, v. 56, n. 9, p. 1099–1108, 2005.
doi:10.1057/palgrave.jors.2601976
- LOPEZ, J. A.; SAIDENBERG, M. R. Evaluating credit risk models. **Journal of Banking & Finance**, v. 24, n. 1-2, p. 151–165, 2000. doi:10.1016/s0378-4266(99)00055-2
- MANZANEQUE, M.; PRIEGO, A. M.; MERINO, E. Corporate governance effect on financial distress likelihood: Evidence from Spain. **Revista de Contabilidad**, v. 19, n. 1, p. 111–121, 2016. doi:10.1016/j.rcsar.2015.04.001
- MARQUÉS, A. I.; GARCÍA, V.; SÁNCHEZ, J. S. On the suitability of resampling techniques for the class imbalance problem in credit scoring. **Journal of the Operational Research Society**, v. 64, n. 7, p. 1060–1070, 2013. doi:10.1057/jors.2012.120
- MARQUÉS, A. I.; GARCÍA, V.; SÁNCHEZ, J. S. Exploring the behaviour of base classifiers in credit scoring ensembles. **Expert Systems with Applications**, v. 39, n. 11, p. 10244–10250, 2012. doi:10.1016/j.eswa.2012.02.092
- MARQUÉS, A. I.; GARCÍA, V.; SÁNCHEZ, J. S. Two-level classifier ensembles for credit risk assessment. **Expert Systems with Applications**, v. 39, n. 12, p. 10916–10922, 2012. doi:10.1016/j.eswa.2012.03.033
- MARQUÉS, A. I.; GARCÍA, V.; SÁNCHEZ, J. S. A literature review on the application of evolutionary computing to credit scoring. **Journal of the Operational Research Society**, v. 64, n. 9, p. 1384–1399, 2013. doi:10.1057/jors.2012.145
- MARTIN, D. Early warning of bank failure. **Journal of Banking Finance**, v. 1, n. 3, p. 249–276, 1977. doi:10.1016/0378-4266(77)90022-x
- MCKEE, T. E.; LENSBERG, T. Genetic programming and rough sets: A hybrid approach to bankruptcy classification. **European Journal of Operational Research**, v. 138, n. 2, p. 436–451, 2002. doi:10.1016/s0377-2217(01)00130-8
- MCKEE, T. E. Rough sets bankruptcy prediction models versus auditor signalling rates. **Journal of Forecasting**, v. 22, n. 8, p. 569–586, 2003. doi:10.1002/for.875
- MENSAH, Y. M. An Examination of the Stationarity of Multivariate Bankruptcy Prediction Models: A Methodological Study. **Journal of Accounting Research**, v. 22, n. 1, p. 380, 1984. doi:10.2307/2490719
- MIN, J. H.; JEONG, C. A binary classification method for bankruptcy prediction. **Expert Systems with Applications**, v. 36, n. 3, p. 5256–5263, 2009. doi:10.1016/j.eswa.2008.06.073
- MIN, J.; LEE, Y. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. **Expert Systems with Applications**, v. 28, n. 4, p. 603–614, 2005. doi:10.1016/j.eswa.2004.12.008
- MIN, J. H.; LEE, Y.-C. A practical approach to credit scoring. **Expert Systems with Applications**, v. 35, n. 4, p. 1762–1770, 2008. doi:10.1016/j.eswa.2007.08.070

MIN, S.-H.; LEE, J.; HAN, I. Hybrid genetic algorithms and support vector machines for bankruptcy prediction. **Expert Systems with Applications**, v. 31, n. 3, p. 652–660, 2006. doi:10.1016/j.eswa.2005.09.070

MORO, S.; CORTEZ, P.; RITA, P. Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. **Expert Systems with Applications**, v. 42, n. 3, p. 1314–1324, 2015. doi:10.1016/j.eswa.2014.09.024

NANNI, L.; LUMINI; A. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. **Expert Systems with Applications**, v. 36, n. 2, p. 3028–3033, 2009. doi:10.1016/j.eswa.2008.01.018

ODOM, M. D.; SHARDA, R. A neural network model for bankruptcy prediction. In: **IJCNN International Joint Conference on Neural Networks, 1990**. IEEE, San Diego, CA, p. 163–168, 1990. doi:10.1109/ijcnn.1990.137710

ÖĞÜT, H.; DOĞANAY, M. M.; CEYLAN, N. B.; AKTAŞ; R. Prediction of bank financial strength ratings: The case of Turkey. **Economic Modelling**, v. 29, n. 3, p. 632–640, 2012. doi:10.1016/j.econmod.2012.01.010

OHLSON, J. A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. **Journal of Accounting Research**, v. 18, n. 1, p. 109, 1980. doi:10.2307/2490395

OLSON, D. L.; DELEN, D.; MENG, Y. Comparative analysis of data mining methods for bankruptcy prediction. **Decision Support Systems**, v. 52, n. 2, p. 464–473, 2012. doi:10.1016/j.dss.2011.10.007

OPLER, T. C.; TITMAN, S. Financial distress and corporate performance. **The Journal of finance**, v. 49, n. 3, p. 1015-1040, 1994. doi.org/10.1111/j.1540-6261.1994.tb00086.x

PALEOLOGO, G.; ELISSEEFF, A.; ANTONINI, G. Subagging for credit scoring models. **European Journal of Operational Research**, v. 201, n. 2, p. 490–499, 2010. doi:10.1016/j.ejor.2009.03.008

PAN, W.-T. A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example. **Knowledge-Based Systems**, v. 26, p. 69–74, 2012. doi:10.1016/j.knosys.2011.07.001

PENDHARKAR, P. C. A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem. **Computers & Operations Research**, v. 32, n. 10, p. 2561–2582, 2005. doi:10.1016/j.cor.2004.06.023

PINHEIRO, L. E. T.; SANTOS, C. P.; COLAUTO, R. D.; PINHEIRO, J. L. Validação de modelos brasileiros de previsão de insolvência. **Contabilidade Vista & Revista**, v. 18, n. 4, 83-103, 2009.

PLATT, H. D.; PLATT, M. B. Development of a class of stable predictive variables: the case of bankruptcy prediction. **Journal of Business Finance & Accounting**, v. 17, n. 1, p. 31-51, 1990. doi.org/10.1111/j.1468-5957.1990.tb00548.x

- PRITCHARD, A. Statistical bibliography or bibliometrics. **Journal of documentation**, v. 25, n. 4, p. 348-349, 1969.
- RIBEIRO, B.; CHEN, N.; KOVACEC, A. Shaping graph pattern mining for financial risk. **Neurocomputing**, 2017. doi:10.1016/j.neucom.2017.01.119
- RIBEIRO, B.; SILVA, C.; CHEN, N.; VIEIRA, A.; CARVALHO DAS NEVES, J. Enhanced default risk models with SVM+. **Expert Systems with Applications**, v. 39, n. 11, p. 10140–10152, 2012. doi:10.1016/j.eswa.2012.02.142
- ROCHON, L.-P.; ROSSI, S. Has "It" Happened Again?. **International Journal of Political Economy**, v. 39, n. 2, p. 5-9, 2010. doi.org/10.2753/IJP0891-1916390201
- SALCEDO-SANZ, S.; FERNÁNDEZ-VILLACAÑAS, J.-L.; SEGOVIA-VARGAS, M. J.; BOUSÓN-CALZÓN, C. Genetic programming for the prediction of insolvency in non-life insurance companies. **Computers & Operations Research**, v. 32, n. 4, p. 749–765, 2005. doi:10.1016/j.cor.2003.08.015
- SALCHENBERGER, L. M.; CINAR, E. M.; LASH, N. A. Neural Networks: A New Tool for Predicting Thrift Failures. **Decision Sciences**, v. 23, n. 4, p. 899–916, 1992. doi:10.1111/j.1540-5915.1992.tb00425.x
- SCALZER, R. S.; RODRIGUES, A.; MACEDO, M. Á. D. S.; WANKE, P. Financial distress in electricity distributors from the perspective of Brazilian regulation. **Energy policy**, v. 125, p. 250-259, 2019. doi.org/10.1016/j.enpol.2018.10.018
- SCHERR, F. C. **Modern Working Capital Management**. São Paulo: Prentice-Hall, 1989.
- SCIMAGO-SJR. **Scimago Journal & Country Rank**. Disponível em: <https://www.scimagojr.com/journalrank.php>. Acesso em: jun. 2019.
- SERASA EXPERIAN. **Indicadores econômicos**. Disponível em: <https://www.serasaexperian.com.br>. Acesso em: abr. 2019.
- SERRANO-CINCA, C.; GUTIÉRREZ-NIETO; B. Partial Least Square Discriminant Analysis for bankruptcy prediction. **Decision Support Systems**, v. 54, n. 3, p. 1245–1255, 2013. doi:10.1016/j.dss.2012.11.015
- SERRANO-CINCA, C. Self organizing neural networks for financial diagnosis. **Decision Support Systems**, v. 17, n. 3, p. 227–238, 1996. doi:10.1016/0167-9236(95)00033-x
- SHIN, K.-S.; LEE, Y.-J. A genetic algorithm application in bankruptcy prediction modeling. **Expert Systems with Applications**, v. 23, n. 3, p. 321–328, 2002. doi:10.1016/s0957-4174(02)00051-9
- SHIN, K.-S.; LEE, T. S.; KIM, H. An application of support vector machines in bankruptcy prediction model. **Expert Systems with Applications**, v. 28, n. 1, p. 127–135, 2005. doi:10.1016/j.eswa.2004.08.009

SHLEIFER, A.; VISHNY, R. W. A Survey of Corporate Governance. **The Journal of Finance**, v. 52, n. 2, p. 737-783, 1997. doi.org/10.1111/j.1540-6261.1997.tb04820.x

STEINER, M. T. A.; CARNIERI, C.; KOPITTKE, B. H.; STEINER NETO, P. J. Sistemas especialistas probabilísticos e redes neurais na análise do crédito bancário. **Revista de Administração**, São Paulo, v. 34, n. 3, 1999.

SUN, J.; FUJITA, H.; CHEN, P.; LI, H. Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble. **Knowledge-Based Systems**, v. 120, p. 4–14, 2017. doi:10.1016/j.knosys.2016.12.019

SUN, J.; JIA, M.; LI, H. AdaBoost ensemble for financial distress prediction: An empirical comparison with data from Chinese listed companies. **Expert Systems with Applications**, v. 38, n. 8, p. 9305–9312, 2011. doi:10.1016/j.eswa.2011.01.042

SUN, J.; LANG, J.; FUJITA, H.; LI, H. Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates. **Information Sciences**, v. 425, p. 76–91, 2018. doi:10.1016/j.ins.2017.10.017

SUN, J.; LI, H.; HUANG, Q.-H.; HE, K.-Y. Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. **Knowledge-Based Systems**, v. 57, p. 41–56, 2014. doi:10.1016/j.knosys.2013.12.006

SUN, J.; LI, H. Data mining method for listed companies' financial distress prediction. **Knowledge-Based Systems**, v. 21, n. 1, p. 1–5, 2008. doi:10.1016/j.knosys.2006.11.003

SUNG, T. K.; CHANG, N.; LEE, G. Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction. **Journal of Management Information Systems**, v. 16, n. 1, p. 63–85, 1999. doi:10.1080/07421222.1999.11518234

TAFFLER, R. J. Forecasting Company Failure in the UK Using Discriminant Analysis and Financial Ratio Data. **Journal of the Royal Statistical Society. Series A General**, v. 145, n. 3, p. 342-358, 1982. doi.org/10.2307/2981867

TAM, K. Y.; KIANG, M. Predicting Bank Failures: A Neural Network Approach. **Applied Artificial Intelligence**, v. 4, n. 4, p. 265–282, 1990. doi:10.1080/08839519008927951

TAM, K. Y.; KIANG, M. Y. Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. **Management Science**, v. 38, n. 7, p. 926–947, 1992. doi:10.1287/mnsc.38.7.926

TAN, T. K. Financial distress and firm performance: Evidence from the Asian financial crisis. **Journal of Finance and Accountancy**, v. 11, p. 1, 2012.

TASCHE, D. Validation of internal rating systems and PD estimates. In: **The analytics of risk model validation**. Academic Press, p. 169-196, 2008. doi.org/10.1016/B978-075068158-2.50014-7

- TIAN, Y.; SHI, Y.; LIU, X. recent advances on support vector machines research. **Technological and Economic Development of Economy**, v. 18, n. 1, p. 5–33, 2012. doi:10.3846/20294913.2012.661205
- TIAN, S.; YU, Y. Financial ratios and bankruptcy predictions: An international evidence. **International Review of Economics Finance**, v. 51, p. 510–526, 2017. <https://doi.org/10.1016/j.iref.2017.07.025>
- TOBACK, E.; BELLOTTI, T.; MOEYERSOMS, J.; STANKOVA, M.; MARTENS, D. Bankruptcy prediction for SMEs using relational data. **Decision Support Systems**, v. 102, p. 69–81, 2017. doi:10.1016/j.dss.2017.07.004
- TOLLEFSON, J. China declared world's largest producer of scientific articles. **Nature**, v. 553, n. 7689, p. 390–390, 2018. doi:10.1038/d41586-018-00927-4
- TSAI, C.-F.; CHEN, M.-L. Credit rating by hybrid machine learning techniques. **Applied Soft Computing**, v. 10, n. 2, p. 374–380, 2010. doi:10.1016/j.asoc.2009.08.003
- TSAI, C.-F.; CHENG, K.-C. Simple instance selection for bankruptcy prediction. **Knowledge-Based Systems**, v. 27, p. 333–342, 2012. doi:10.1016/j.knosys.2011.09.017
- TSAI, C.-F.; HSU, Y.-F.; YEN, D. C. A comparative study of classifier ensembles for bankruptcy prediction. **Applied Soft Computing**, v. 24, p. 977–984, 2014. doi:10.1016/j.asoc.2014.08.047
- TSAI, C.; WU, J. Using neural network ensembles for bankruptcy prediction and credit scoring. **Expert Systems with Applications**, v. 34, n. 4, p. 2639–2649, 2008. doi:10.1016/j.eswa.2007.05.019
- TSAI, C.-F. Feature selection in bankruptcy prediction. **Knowledge-Based Systems**, v. 22, n. 2, p. 120–127, 2009. doi:10.1016/j.knosys.2008.08.002
- TSAI, C.-F. Combining cluster analysis with classifier ensembles to predict financial distress. **Information Fusion**, v. 16, p. 46–58, 2014. doi:10.1016/j.inffus.2011.12.001
- TSAI, C.-F. Financial decision support using neural networks and support vector machines. **Expert Systems**, v. 25, n. 4, p. 380–393, 2008. doi:10.1111/j.1468-0394.2008.00449.x
- TUNG, W. L.; QUEK, C.; CHENG, P. GenSo-EWS: a novel neural-fuzzy based early warning system for predicting bank failures. **Neural Networks**, v. 17, n. 4, p. 567–587, 2004. doi:10.1016/j.neunet.2003.11.006
- TWALA, B. Multiple classifier application to credit risk assessment. **Expert Systems with Applications**, v. 37, n. 4, p. 3326–3336, 2010. doi:10.1016/j.eswa.2009.10.018
- VARETTO, F. Genetic algorithms applications in the analysis of insolvency risk. **Journal of Banking & Finance**, v. 22, n. 10-11, p. 1421–1439, 1998. doi:10.1016/s0378-4266(98)00059-4

- VERIKAS, A.; KALSYTE, Z.; BACAUSKIENE, M.; GELZINIS, A. Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. **Soft Computing**, v. 14, n. 9, p. 995–1010, 2009. doi:10.1007/s00500-009-0490-5
- WANG, G.; CHEN, G.; CHU, Y. A new random subspace method incorporating sentiment and textual information for financial distress prediction. **Electronic Commerce Research and Applications**, v. 29, p. 30–49, 2018. doi:10.1016/j.elerap.2018.03.004
- WANG, G.; HAO, J.; MA, J.; JIANG, H. A comparative assessment of ensemble learning for credit scoring. **Expert Systems with Applications**, v. 38, n. 1, p. 223–230, 2011. doi:10.1016/j.eswa.2010.06.048
- WANG, G.; MA, J. Study of corporate credit risk prediction based on integrating boosting and random subspace. **Expert Systems with Applications**, v. 38, n. 11, p. 13871–13878, 2011. doi:10.1016/j.eswa.2011.04.191
- WANG, G.; MA, J. A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine. **Expert Systems with Applications**, v. 39, n. 5, p. 5325–5331, 2012. doi:10.1016/j.eswa.2011.11.003
- WANG, G.; MA, J.; YANG, S. An improved boosting based on feature selection for corporate bankruptcy prediction. **Expert Systems with Applications**, v. 41, n. 5, p. 2353–2361, 2014. doi:10.1016/j.eswa.2013.09.033
- WANG, G.; MA, J.; HUANG, L.; XU, K. Two credit scoring models based on dual strategy ensemble trees. **Knowledge-Based Systems**, v. 26, p. 61–68, 2012. doi:10.1016/j.knosys.2011.06.020
- WANG, J.; VEUGELERS, R.; STEPHAN, P. Bias against novelty in science: A cautionary tale for users of bibliometric indicators. **Research Policy**, v. 46, n. 8, p. 1416–1436, 2017. doi:10.1016/j.respol.2017.06.006
- WHALEN, G. A proportional hazards model of bank failure: an examination of its usefulness as an early warning tool. **Economic Review**, v. 27, n. 1, p. 21–31, 1991.
- WILSON, R. L.; SHARDA, R. Bankruptcy prediction using neural networks. **Decision Support Systems**, v. 11, n. 5, p. 545–557, 1994. doi:10.1016/0167-9236(94)90024-8
- WU, C.-H.; TZENG, G.-H.; GOO, Y.-J.; FANG, W.-C. A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. **Expert Systems with Applications**, v. 32, n. 2, p. 397–408, 2007. doi:10.1016/j.eswa.2005.12.008
- WU, D. D.; OLSON, D. L.; LUO, C. A Decision Support Approach for Accounts Receivable Risk Management. **IEEE Transactions on Systems, Man, and Cybernetics: Systems**, v. 44, n. 12, p. 1624–1632, 2014. doi:10.1109/tsmc.2014.2318020
- YANG, Z.; YOU, W.; JI, G. Using partial least squares and support vector machines for bankruptcy prediction. **Expert Systems with Applications**, v. 38, n. 7, p. 8336–8342, 2011. doi:10.1016/j.eswa.2011.01.021

YONGQIAO WANG; SHOUYANG WANG; LAI; K. K. A new fuzzy support vector machine to evaluate credit risk. **IEEE Transactions on Fuzzy Systems**, v. 13, n. 6, p. 820–831, 2005. doi:10.1109/tfuzz.2005.859320

YU, L.; WANG, S.; LAI, K. Credit risk assessment with a multistage neural network ensemble learning approach. **Expert Systems with Applications**, v. 34, n. 2, p. 1434–1444, 2008. doi:10.1016/j.eswa.2007.01.009

YU, L.; YAO, X.; WANG, S.; LAI, K. K. Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection. **Expert Systems with Applications**, v. 38, n. 12, p. 15392–15399, 2011. doi:10.1016/j.eswa.2011.06.023

YU; Q.; MICHE; Y.; SÉVERIN; E.; LENDASSE; A. Bankruptcy prediction using Extreme Learning Machine and financial expertise. **Neurocomputing**, v. 128, p. 296–302, 2014. doi:10.1016/j.neucom.2013.01.063

ZELENKOV, Y.; FEDOROVA, E.; CHEKRIZOV, D. Two-step classification method based on genetic algorithm for bankruptcy forecasting. **Expert Systems with Applications**, v. 88, p. 393–401, 2017. doi:10.1016/j.eswa.2017.07.025

ZHANG, G.; HU, M. Y.; PATUWO, B. E.; INDRO, D. C. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. **European Journal of Operational Research**, v. 116, n. 1, p. 16–32, 1999. doi:10.1016/s0377-2217(98)00051-4

ZHANG, Y.; WANG, S.; JI, G. A Rule-Based Model for Bankruptcy Prediction Based on an Improved Genetic Ant Colony Algorithm. **Mathematical Problems in Engineering**, p. 1–10, 2013. doi:10.1155/2013/753251

ZHANG, Z.; GAO, G.; SHI, Y. Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. **European Journal of Operational Research**, v. 237, n. 1, p. 335–348, 2014. doi:10.1016/j.ejor.2014.01.044

ZHAO, D.; HUANG, C.; WEI, Y.; YU, F.; WANG, M.; CHEN, H. An Effective Computational Model for Bankruptcy Prediction Using Kernel Extreme Learning Machine Approach. **Computational Economics**, v. 49, n. 2, p. 325–341, 2016. doi.org/10.1007/s10614-016-9562-7

ZHAO, H.; SINHA, A. P.; GE, W. Effects of feature construction on classification performance: An empirical study in bank failure prediction. **Expert Systems with Applications**, v. 36, n. 2, p. 2633–2644, 2009. doi:10.1016/j.eswa.2008.01.053

ZHONG, H.; MIAO, C.; SHEN, Z.; FENG, Y. Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings. **Neurocomputing**, v. 128, p. 285–295, 2014. doi:10.1016/j.neucom.2013.02.054

ZHOU, L. Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. **Knowledge-Based Systems**, v. 41, p. 16–25, 2013. <https://doi.org/10.1016/j.knosys.2012.12.007>

ZHU, Y.; XIE, C.; WANG, G.-J.; YAN, X.-G. Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance. **Neural Computing and Applications**, v. 28, n 1, p. 41–50, 2016.
<https://doi.org/10.1007/s00521-016-2304-x>

ZIEBA, M.; TOMCZAK, S. K.; TOMCZAK, J. M. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. **Expert Systems with Applications**, v. 58, p. 93–101, 2016. doi:10.1016/j.eswa.2016.04.001

ZMIJEWSKI, M. E. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. **Journal of Accounting Research**, v. 22, p. 59-82, 1984.
doi:10.2307/2490859

APÊNDICE A – Portfólio Bibliométrico final com os 165 artigos selecionados pelo método ProKnow-C

Publ.	DOI	1	2	3	4	5	6	7	8	
1992	10.1287/mnsc.38.7.926	C	A	LDA-LR-KNN-ANN-DTID3	A	A	A	A	1985-1987	
1994	10.1016/0378-4266(94)90007-8	C	B	A	LDA-ANN	B	B	A	A	1982-1992
1994	10.2307/253568	B	A	LR-ANN-MDA-O	B	A	B	A B	1987-1990	
1994	10.1016/0167-9236(94)90024-8	C	A	MDA-ANN	B	F	A	A	1975-1982	
1996	10.1016/S0957-4174(96)00056-5	C	A	LDA-ANN-CBFS	B	C	A	A	1991-1993	
1996	10.1016/0167-9236(96)00018-8	C	A	MDA-ANN-DTID3-SOM	B	C	A	A	1979-1992	
1996	10.1016/0167-9236(95)00033-X	B	A	SOM-LDA-MLP-DSS	B	B	B	A	1975-1985	
1998	10.1016/S0378-4266(98)00059-4	B	A	LDA-GA-ANN	B	B	B	A	1982-1995	
1999	10.1080/07421222.1999.11518234	C	A	GA-DT-MDA-ANN-BOO	B	C	A	A	1991-1998	
1999	10.1016/S0377-2217(98)00051-4 10.1111/j.1540-	C	A	LR-ANN-CVNN	C	F	A	D	-	
2000	5915.2000.tb01633.x	B	A	BPNN-PNN-CLN	A	A	C	C	1986-1988	
2001	10.1016/S0305-0483(01)00045-7	A	A	RST-VPRS-RPA-LR-EM-MDA	B	B	C	A	-	
2001	10.1016/S0950-7051(01)00096-X	A	A	LDA-LR-ANN-DT	B	B	C	A	1980-1999	
2002	10.1016/S0377-2217(01)00130-8	C	A	GA-RST	B	A	C	A	1991-1997	
2002	10.1016/S0957-4174(02)00051-9	C	A	GA	B	C	A	A	1995-1997	
2003	10.1016/S0957-4174(03)00102-7	C	A	ANN-GA	B	C	A	A B	2001-2002	
2003	10.1002/for.875	C	A	RST	B	A	A	A	1991-1997	
2004	10.1016/S0377-2217(03)00186-3	C	A	DEA-DT-O	B	B	A	A	1994-1996	
2004	10.1016/S0167-9236(03)00086-1	D	A	LR-SVM-ANN-BPNN	A	A C	C	A	1998-2002	
2004	10.1016/j.neunet.2003.11.006	C	A	GSOFNN-MCMAC-COX	A	A	A	A	1980-2000	
2005	10.1057/palgrave.jors.2601976	D	A	KNN-LR-NN-MLP-DTM5	C	B	C	D	-	
2005	10.1016/j.eswa.2004.12.008	C	A	ANN-SVM-BPNN-LR-MDA	B	C	A	A	2000-2002	
2005	10.1016/j.cor.2004.06.023	C	A	GA-ANN-DT-LDA-BPNN-TV	B	F	A	A	1987-1995	
2005	10.1016/j.cor.2003.08.015	B	A	GP-SVM-RST-GPR	A	B	B	A	1983-1993	
2005	10.1016/j.eswa.2004.08.009	C	A	SVM-BPNN	B	C	A	A	1996-1999	
2005	10.1109/TFUZZ.2005.859320	D	A	LR-ANN-FSVM-RBF	B	B C	C	A	-	

2006	10.1016/j.im.2006.08.001	B	A	ANN-RST	B	F	C	A	2002
2006	10.1016/j.eswa.2005.10.003	A	A	SVM-BPNN	A	C	C	A B	-
2006	10.1016/j.knosys.2005.11.006	B	A	GA-PSO-ANN-LR-LDA	B	C	B	A	1993-2002
2006	10.1504/IJBIDM.2006.010782	C	A	MCLP	B	C	A	A	1989-1999
2006	10.1016/j.eswa.2005.09.070	C	A	SVM-ANN-LR-GA-CBR	B	C	A	A	1999-2000
2007	10.1016/j.ejor.2006.09.103	D	A	CART-DT-MDA-LR	A	B	D	A	1999-2000
2007	10.1016/j.ejor.2006.09.101	D	A	LR-ANN-SMO-NB	B	A	A	A	1994-2002
2007	10.1016/j.eswa.2006.04.018	D	A	BPNN-SVM-MDA-CBR	B	C	C	A	1997-2002
2007	10.1016/j.eswa.2005.12.008	C	A	SVM-GA-MDA-LR-PM-ANN	B	C	A	A	2001-2003
2008	10.1016/j.dss.2007.12.002	C	A	AB-ANN	B	B	A	A B	-
2008	10.1016/j.eswa.2007.08.040	C	A	LDA-DT-BPNN-HIS	B	C	A	A	2001-2004
2008	10.1108/02686900810890625	A	A	LR-ANN-DT	B	C	D	A	2003-2004
2008	10.1016/j.eswa.2007.08.070	C	A	DEA-MDA	B	C	A	A	-
2008	10.1016/j.knosys.2006.11.003	B	A	DM	B	C	B	A	2000-2005
2008	10.1016/j.eswa.2007.05.019	C D	A	ANN-BPNN-MLP	B	B C E	A	A	-
2008	10.1111/j.1468-0394.2008.00449.x	C D	A	SVM-MLP-RBF	B	B C E F	A	A	-
2008	10.1016/j.eswa.2007.01.009	D	A	LR-ANN-SVM-FSVM-VB-RB	B	B	C	A B	-
2009	10.1016/j.asoc.2008.08.002	C	A	MDA-CBR-GA	A	C	A	A	1996-2000
2009	10.1016/j.eswa.2008.01.003	C	A	MLP-CL-LVQ-SOM-ANN-SVM-MDA-KM	A	C	A	A	1997-2003
2009	10.1016/j.eswa.2008.05.047	A	A	MLT-CART-RF-LR-SVM-DT-BOO	B	F	C	A	-
2009	10.1016/j.eswa.2008.03.020	C	A	ANN-DM	B	C	A	A B	1999-2006
2009	10.1016/j.eswa.2007.09.060	C	A	MDA-LR-ML-ANN-DT-SVM-RULE-IMSW	B	C	A	A	1999-2002
2009	10.1002/for.1109	D	A	LR-LDA-SVM	B	B	B	A	1997-2002
2009	10.1016/j.neucom.2009.03.002	C	A	LDA-LR-SLP-MLP-SVM-PM-FIFLN-ELECTRE	C	F	A	A	1975-1982
2009	10.1016/j.eswa.2008.06.068	C	A	DT-ANN-BPNN-SVM-SMO	B	F	A	A	1997-2001
2009	10.1016/j.eswa.2007.09.038	B C	A	MDA-LR-ANN-SVM-CBR	B	C	B	A	2000-2005
2009	10.1016/j.ejor.2008.05.024	B C	A	ELECTRE-CBR	B	C	B	A	2000-2005
2009	10.1016/j.eswa.2007.11.068	A	A	RST-CBR-GRA-HFP	B	C	C	A B	1999-2006
2009	10.1016/j.eswa.2009.03.029	B	A	PSO-SVM-DT	A	C	B	A B	2000-2005

2009	10.1016/j.eswa.2008.06.073	C	A	MDA-LR-DT-ANN-BC	B	F	A	A	2001-2004
2009	10.1016/j.eswa.2008.01.018	C D	A	RS-BA-RTF-SVM-MLP-CW-RBF	B	B C E	A	A	-
2009	10.1016/j.knosys.2008.08.002	C D	A	MLP	B	B C E F	A	A	-
2009	10.1016/j.eswa.2008.01.053	C	A	LR-DT-ANN-KNN	B	F	A	A	1991-1992
2010	10.1007/s00521-010-0362-z	B D	B	ANN-ES-HIS	B	F	A	C	-
2010	10.1016/j.eswa.2009.12.025	D	A	SVM-DT-LDA-RST-FS	B	B E	C	A	-
2010	10.1016/j.eswa.2009.10.012	C	A	ANN-BOO-BA-DT-ML	B	C	A	A	2002-2005
2010	10.1109/ICCCE.2008.4580635	D	B	O	C	F	E	D	-
2010	10.1016/j.cor.2009.04.003	B C	A	CBR-LR-MDA-KNN	B	C	B	A	2000-2005
2010	10.1016/j.ejor.2009.03.008	D	A	SVM-RBF-KNN-DT-AB	B	B	C	A B	-
2010	10.1016/j.asoc.2009.08.003	D	A	DT-NB-LR-ANN-KM	B	C	C	A	2004-2006
2010	10.1016/j.eswa.2009.10.018	D	A	ANN-DT-KNN-LR-BA-BOO-NB	A	A B C E	A	A	1985-1987
2010	10.1007/s00500-009-0490-5	C	B	O	C	E	E	D	-
2011	10.1016/j.asoc.2010.10.003	C	A	LR-LDA-PNN-FSVM	B	A	A	A	2000-2001
2011	10.1016/j.eswa.2011.02.173	B	A	LR-DT-CART	B	C	B	A B	-
2011	10.1016/j.eswa.2011.04.090	C	A	GA-LVQ-CSL	B	B	A	A B	2006
2011	10.1016/j.eswa.2010.07.117	C	A	FCN-MARS-LDA-ANN	B	C	A	A	-
2011	10.1111/j.1540-5915.2011.00328.x	D	A	O	B	B	C	A	1998-2003
2011	10.1016/j.ejor.2010.09.029	D	A	LR-LDA-CART-ANN-KNN-AB-BA-BOO	B	B	C	A	2002
2011	10.1016/j.im.2011.05.001	B C	A	CBR-MDA-LR	B	C	B	A	2000-2005
2011	10.1016/j.eswa.2011.05.035	B	A	SVM-LR-MDA-RBFN	B	C	B	A B	2000-2008
2011	10.1016/j.eswa.2011.01.042	B	A	AB-DT-SVM-SAT	B	C	B	A	2000-2008
2011	10.1016/j.eswa.2010.06.048	D	A	BA-BOO-LR-DT-ANN-SVM	B	C E B	C	A	2006-2007
2011	10.1016/j.eswa.2011.04.191	D	A	LR-DT-ANN-BA-BOO-RS	B	C	C	A	2006-2007
2011	10.1016/j.eswa.2011.01.021	C	A	RBF-MLP-SVM-ANN-PLS	B	F	A	A	-
2011	10.1016/j.eswa.2011.06.023	D	A	SVM	B	B E	C	D	-
2012	10.1016/j.eswa.2011.09.033	D	A	LR-ANN-DT-SVM-RF-LDA-GBOO	B	B	B	A	-
2012	10.1016/j.ejor.2012.04.006	C	A	MDA-LR-ANN-COX	B	B	A	A	1991-2009
2012	10.1016/j.neucom.2011.11.020	D	A	LR-ANN-SVM-EABC-BPNN-GA	B	C	A	A B	1999-2007

2012	10.1016/j.cor.2011.06.023	D	A	SVM-MDA-LR-CBR-ANN	B	C	C	A	-
2012	10.1016/j.eswa.2012.02.072	C	A	BA-BOO-DT-ANN-SVM	B	C	A	A	2002-2005
2012	10.1016/j.eswa.2011.12.021	B	A	CHAID-EWS-LR-MDA-SOM-CART	B	C	B	A	2007
2012	10.1016/j.tourman.2011.07.004	B C	A	SVM-ANN-MDA-LR-KNN	B	C	B	A	1998-2010
2012	10.1109/TSMCC.2011.2170420	D	B	O BA-RS-DECORATE-AB-RTF-NN-NB-LR-MLP-RBF-ANN-SVM-DT	C	F	E	D	1995-2010
2012	10.1016/j.eswa.2012.02.092	D	A	AB-BA-RS-RF-ANN-LR-MLP-SVM-DT	B	A B C E	C	A	-
2012	10.1016/j.eswa.2012.03.033	D	A	LR-MDA-SVM-PNN	B	A B C E	C	A	-
2012	10.1016/j.econmod.2012.01.010	D	A	LR-RBF-DT-CART-SVM-DTJ8	A	C	C	A	2003-2009
2012	10.1016/j.dss.2011.10.007	C	A	FFAO-ANN-GRNN	B	A	A	A	2005-2009
2012	10.1016/j.knosys.2011.07.001	B	A	SVM-MTL	B	B	C	A B	2003-2004
2012	10.1016/j.eswa.2012.02.142	A	A	O	C	F	E	D	2002-2006
2012	10.1016/S11138-4891(12)70037-7	A	B	SVM	C	F	E	D	-
2012	10.3846/20294913.2012.661205	C	B	MLP-DT-LR-SVM	B	B C E F	A	A	-
2012	10.1016/j.knosys.2011.09.017	C	A	RS-BA-DT-LR-LDA-MLP-RBFN-RF-RTF	B	B E	C	A	-
2012	10.1016/j.knosys.2011.06.020	D	A	RSB-SVM-ANN-BA-RS-BOO-LR-DT-MLP-AB-SMO	B	C	C	A	2006-2007
2013	10.1016/j.eswa.2012.07.047	C	A	SOM	B	B	A	A B	2003-2006
2013	10.1016/j.knosys.2012.11.004	D	A	RST-ML-HIS	A	A B C F	C	A	1998-2007
2013	10.1016/j.ins.2013.02.015	A	A	CBR-RST-CART-GRA-LR	B	C	A	A	1999-2006
2013	10.1016/j.eswa.2013.07.032	C	A	MDA-LR-CART-ANN-AB	B	B	A	A	2007-2011
2013	10.1016/j.knosys.2013.07.008	D	A	MLP-RBF-SVM-NB-RF-LDA-NMC-ANN	B	A B E	C	A	-
2013	10.1016/j.econmod.2012.11.017	B	A	LDA-CART-ANN	B	B D	B	A	1996-2009
2013	10.1016/j.knosys.2012.11.003	B	A	RF-BPNN-LDA-ML-O	B	B C E	B	A B	2009-2011
2013	10.1057/jors.2012.120	D	A	SVM-LR-SMOTE	B	A B C E	C	A	-
2013	10.1057/jors.2012.145	D	B	O	C	F	C	D	-
2013	10.1016/j.dss.2012.11.015	B	A	LDA-LR-MLP-KNN-NB-SVM-BOO-DT-BRT-PLSDA	A	A	B	A	-
2013	10.1016/j.knosys.2012.12.007	C	A	LDA-LR-DT-ANN-SVM-SMOTE-OS-US	B	A C	A	A	1981-2009
2014	10.1016/j.eswa.2013.12.003	C D	A	RS-LNMC-CDT-DT-RS-BA	B	B C E	A	-	-

2014	10.1016/j.rdf.2014.03.002	D	A	LR-RBF	A	C	C	C	2006-2011
2014	10.1007/s10844-014-0333-4	C D	B	O	C	F	E	C	2002-2013
2014	10.3846/20294913.2014.979456	B	A	SVM-ANN-DT-NB-LR-RBF-MLP-SMO-OS-SMOTE	B	A	B	A B	2008
2014	10.1016/j.asoc.2014.08.009	C	A	AB-ANN-SVM-DT-ZS	B	C	A	A	2008-2012
2014	10.1016/j.economod.2013.10.005	B	A	TD-AB	B	A	B	A	1988-2010
2014	10.3846/20294913.2014.913275	A	A	NB-LR-KNN-DT-CART-MCDM	B	C	C	A	2002-2004
2014	10.1016/j.knosys.2013.12.006	B C	B	O	E	F	E	D	-
2014	10.1016/j.inffus.2011.12.001	C	A	MLP-LR-CART-SOM-KM	B	B C E F	A	A	-
2014	10.1016/j.asoc.2014.08.047	C	A	MLP-SVM-DT-BOO-BA	B	B C E	A	A	-
2014	10.1016/j.eswa.2013.09.033	C	A	LR-NB-DT-ANN-SVM-BA-BOO-MLP-SMO	B	F	A	A	-
2014	10.1109/TSMC.2014.2318020	D	A	LR-DT-ANN	B	F	C	A	-
2014	10.1016/j.neucom.2013.01.063	C	A	ELM-LDA-SVM	B	B	A	A	2002-2003
2014	10.1016/j.ejor.2014.01.044	D	A	SVM-FSVM-MCOC	C	A B E	C	D	-
2014	10.1016/j.neucom.2013.02.054	D	A	ELM-SVM-BPNN-ANN-EKM-SLFN	B	A	C	A	-
2015	10.1007/s10462-015-9434-x	A C	B	O-CSL	C	F	E	D	-
2015	10.1016/j.eswa.2014.12.001	C	A	SVM-LR-ANN-RBF	B	A	A	A	1999-2007
2015	10.1016/j.ejor.2014.08.016	B	A	ANN-DT-SVM-LR-LDA-MV-DM	B	C	B	A	2001-2008
2015	10.1016/j.jbankfin.2015.02.006	D	A	AB-BOO-RF-ANN-SVM-LR-LDA-QDA-PROBIT	B	A E	C	A B D	1983-2013
2015	10.1016/j.eswa.2014.08.025	C	A	AB-GMC-CB	B	C	A	A	2001-2004
2015	10.1016/j.jretconser.2015.07.003	D	A	ANN-SVM-CART-NB-AB-BA-RF	A	C	C	A B	-
2015	10.1016/j.ejor.2015.05.030	D	A	QDA-SVM-NB-LR-KNN-ELM-CART-ANN-RF-BOO-BA-KM-O	B	B E	C	A	-
2015	10.1016/j.knosys.2014.10.010	B	A	KNN-CART-MLP-SVM	B	B C E	B	A	-
2015	10.1016/j.eswa.2014.09.024	E	B	O	A	F	E	D	2002-2013
2016	10.1007/s00521-016-2567-2	C D	A	SOM-FNN	B	B C E	A	A	2002-2006
2016	10.1016/j.knosys.2016.04.013	D	A	MCS-LR-ANN-SVM-RF-DT-NB-CA	B	B C E	A	A	-
2016	10.1016/j.eswa.2016.07.017	D	A	GNG-MARS-CA-ANN-SVM-RF-DT-NB	B	A B C E	A	A	-
2016	10.1007/s13042-016-0545-8	B	A	SVM-DEA-IFWTSVM-ANN-RF	B	C	B	A	2012-2014
2016	10.1016/j.ejor.2016.03.008	C	A	MDA-LR-ANN-CART-BA-BOO-RS-PBM-SOM	B	B	A	A	2003-2012
2016	10.1016/j.asoc.2015.11.037	D	A	FRBC	B	B C E	C	A	-

2016	10.1111/jbfa.12218	C	A	AB-BOO-RF-ANN-SVM-LR-LDA-QDA	B	A	A	A	-
2016	10.1016/j.ejor.2016.01.012	C	A	SVM-KNN-NB-CART-MLP-SDA	B	C	A	A B	1999-2009
2016	10.1016/j.knosys.2016.12.019	B	A	SVM-AB-TBOO	B	C	B	A	2000-2012
2016	10.1016/j.eswa.2016.04.001	C	A	AB-AC-LDA-MLP-LR-DT-RF-GB	B	B	A	A	2007-2013
2017	10.1016/j.eswa.2016.12.020	C D	A	AB-BA-RS-DECORATE-RTF-LR-DT-MLP-SVM-CDT	B	A B C E	A	A	-
2017	10.1016/j.asoc.2017.06.043	C	A	GP-SVM-LR	B	B	A	A	2006-2007
2017	10.1016/j.eswa.2017.04.006	C	A	MDA-LR-ML-SVM-BA-RF-BOO-ANN	B	A	A	A	1985-2013
2017	10.1016/j.eswa.2017.05.050	D	A	ELM-ANN-LR	A	B E	C	A	-
2017	10.1016/j.asoc.2017.03.014	C	A	FL-GA-BPNN	B	C	A	A	-
2017				MDA-LR-SVM-FNN-DT-ELM-RS-RTF-DECORATE-BA-BOO-					
2017	10.1016/j.eswa.2017.01.016	C	A	AB-COX-SOM	B	B	A	A	1996-2001
2017	10.1007/s10614-017-9783-4	C	A	FLD-LDA-SVM-LR	A	C	A	C	-
2017	10.1007/s11142-017-9407-1	C	A	GBOO	B	A	A	A B D	1987-2013
2017	10.1007/s10115-017-1061-1	B	A	SVM-KNN-CART-MLP-NB-BA-BOO-MV	B	B C E	B	A	2005-2015
2017	10.1016/j.neucom.2017.01.119	A	A	MLP-SVM-RBF-CART-BOO-KNN	B	B	C	A B	2002-2006
2017	10.1016/j.dss.2017.07.004	C	A	DM-WVRN	B	B	A	A B	2011-2014
2017	10.1016/j.eswa.2017.07.025	C	A	KNN-LR-NB-DT-SVM-ADB-RF-BA-QDA-ZS-MLP-GBOO-TSCM	B	B	A	A	-
2018	10.1111/jbfa.12338	C	A	LR-GBOO	B	A	A	A B	1992-2012
2018				MDA-LR-SVM-FNN-CART-ELM-RS-RTF-BA-BOO-AB-COX-					
2018	10.1016/j.dss.2018.01.003	C	A	SOM	B	B	A	A	2006-2014
2018	10.1016/j.asoc.2018.01.021	D	A	DT-ANN-SVM-NB-KNN-BA	B	B C E	C	A	-
2018	10.1007/s10489-018-1253-8	D	A	DT-ANN-SVM-NB-GA-BA-BOO	B	B C E	C	A	-
2018	10.1111/exsy.12335	C	A	LR-NB-BPNN-DT-SVM-KNN-BA-BOO-GA	B	B C E	A	A	-
2018	10.1016/j.ins.2017.10.017	D	A	DT-SMOTE-BA-OS-US	B	C	C	A	2007-2012
2018	10.1016/j.elerap.2018.03.004	B	A	SVM-SMOTE-BA-BOO-RS-OS-US-CSL-ISTRs	B	C	B	A D	2011-2015
2019	10.1016/j.inffus.2018.07.004	C D	A	RF-RTF-DECORATE-AB-BOO-BA-RS	B	B C E	A	A	-

Apêndice B – Relação de empresas que compreendem a região de interesse (mesorregião do Triângulo Mineiro e Alto Paranaíba e do Sul Goiano).

Empresa	Sector	Empresa	Sector
Aliansce	Locadora de imóveis	Locamerica	Locadora de automóveis
Alpargatas	Indústria de calçados	Lojas Americ	Loja de departamentos
Ambev S/A	Indústria de bebidas	Lojas Hering	Loja de roupas
Anima	Escola de ensino superior	Lojas Marisa	Loja de roupas
Bardella	Industria de máquinas agrícolas, de construção e mineração	Lojas Renner	Loja de roupas
Bematech	Indústria de computadores e produtos eletrônicos	Magaz Luiza	Loja de departamentos
BR Malls Par	Locadora de imóveis	Magnesita	Extração de minerais não metálicos
BR Pharma	Loja de artigos para saúde e cuidados pessoais	Metalfrio	Indústria de equipamentos de refrigeração
Brasilagro	Agricultura	Metisa	Outras indústrias de produtos de metal
BRF SA	Abatedouros	Mills	Serviços de engenharia e arquitetura
Brookfield	Construção de edifícios residenciais	MMX Miner	Mineração de metais
Carrefour BR	Loja de departamentos	Movida	Locadora de automóveis
Cemig	Geração, transmissão e distribuição de energia elétrica	MRV	Construção de edifícios residenciais
Cia Hering	Indústria de roupas de malha	Nadir Figuei	Indústria de vidro e produtos de vidro
Copasa	Água, esgoto e outros sistemas	Natura	Comércio atacadista de bens não duráveis variados
Cosan	Comércio atacadista de petróleo e produtos de petróleo	Net	TV a cabo
Cosan Ltd	Indústria de açúcar e produtos de confeitaria	Oi	Telecomunicações
Cristal	Indústria química	Omega Ger	Administração de empresas e empreendimentos
Cvc Brasil	Transporte turístico	P.Acucar-Cbd	Loja de departamentos
Cyrela Realt	Construção de edifícios residenciais	Portobello	Indústria de produtos de cerâmica e refratários
Dasa	Laboratório de exames médicos	Raiadrogasil	Loja de artigos para saúde e cuidados pessoais
Datasul	Serviços de processamento de dados, hospedagem e outros serviços	Sadia S/A	Abatedouros
Direcional	Construção de edifícios residenciais	Sao Martinho	Indústria de açúcar e produtos de confeitaria
Duratex	Indústria de móveis e afins	Saraiva Livr	Editoras de jornais, revistas, livros e similares
Ecorodovias	Atividades auxiliares ao transporte rodoviário	Souza Cruz	Indústria de fumo
Embratel Part	Telecomunicações	Springer	Indústria de eletrodomésticos
Energisa	Geração, transmissão e distribuição de energia elétrica	Sul America	Seguradora
Eternit	Indústria de outros produtos de minerais não metálicos	Tam S/A	Transporte aéreo regular
Eucatex	Indústria de produtos de madeira compensada e afins	Teka	Indústria de roupas de tecido
Fer Heringer	Indústria de fertilizantes e pesticidas	Telef Brasil	Telecomunicações
Fras-Le	Indústria de autopeças	Totvs	Editoras de software
Gerdau	Transformação de aço em produtos de aço	Ultrapar	Comércio atacadista de petróleo e produtos de petróleo
Gol	Transporte aéreo regular	Usiminas	Transformação de aço em produtos de aço
Haga S/A	Indústria de ferragens	Valid	Impressão e atividades auxiliares

Ipiranga Dis	Posto de gasolina	Viavarejo	Loja de departamentos
Itautec	Indústria de computadores e periféricos	Vigor Food	Indústria de laticínios
JBS	Abatedouros	Vivo	Telecomunicações
Josapar	Moinho de grãos	Weg	Indústria de motores, turbinas e transmissores de energia
Karsten	Indústria de roupas de tecido	Yara Brasil	Indústria de fertilizantes e pesticidas

APÊNDICE C - Comando *stepwise* Stata para modelo *Logit* mais adequado e funcional

. *stepwise, pr(0.05): logit B Setor CALO OL L CATL FATA TTA WCTA CPB CFGD AT CL EIE NINR ROA OM GMNR CLTL OLEB CROA CROE CROI LNA MVS GS OLCF OLIE CFC EI FEV CPV PIB SELIC IBV IPCA* ** ao nível de confiança de 95%.

Em que: CALO: (Ativo Circulante - Estoques - Recebíveis) / (Passivo Circulante- Fornecedores); OL: Alavancagem Operacional: [(EBITDA t - EBITDA t-1) / (Receita Líquida t - Receita Líquida t-1)]; L: Alavancagem: (Passivo Circulante + Passivo não Circulante) / Ativo Total; CATL: Ativo Circulante / Passivo Total; FATA: Ativo não Circulante / Ativo Total; TTA: (Ativo Totalt – Ativo Totalt-1) / Ativo Totalt-1; WCTA: Capital de Giro Líquido: (Ativo Circulante - Passivo Circulante) / Ativo Total; CPB: CPB Price-to-Bookt – Price-to-Bookt-1 tal que P/B = Valor de Mercado da ação / Valor contábil da ação; CFGD: EBITDA / Dívida Bruta; AT: Giro de Ativos: Receita Líquida / Ativo Total; CL: Liquidez Corrente = Ativo Circulante / Passivo Circulante (PC); EIE: Despesas Financeiras (DF) / Lucro antes de juros e impostos; NINR: Lucro Líquido / Receita Líquida; ROA: EBIT / Ativo Total; OM: EBIT / Receita Líquida; GMNR: Margem Bruta / Receita Líquida; CLTA: Passivo Circulante / Ativo Total; CLTL: Passivo Circulante / [Passivo Circulante + Passivo não Circulante (PnC)]; OLEB: (Passivo – Fornecedores) / EBIT; CROA: ROAt – ROAt-1 em que ROA = EBIT / Ativo Total; CROE: ROEt – ROEt-1 em que ROE = Lucro Líquido / Patrimônio Líquido; CROI: ROIt – ROIt-1 em que ROI (Lucro Líquido / Receita Líquida) * (Receita Líquida / Ativo total); LNA: Tamanho = ln (Ativo Total); MVS: Valor de Mercado das ações * número de ações outstanding / (Passivo Circulante + Passivo não Circulante); GS: (Receita Líquidat – Receita Líquidat-1) / Receita Líquidat-1; OLCF: (Passivo Circulante + Passivo não Circulante)/ EBITDA; OLIE: Despesas Financeiras / (Lucro Líquido – Depreciação e Amortização); CFC: Ciclo Financeirot – Ciclo Financeirot-1; EI: Investimentos / Patrimônio Líquido; FEV: DFt – DFt-1/ DFt-1; CPV: (PC + PnC)t - (PC + PnC)t-1 / (PC + PnC)t-1; ETA: Patrimônio Líquido / Ativo Total; PIB: Crescimento do Produto Interno Bruto; SELIC: Taxa básica de juros anual; IBV: Crescimento Índice Bovespa; IPCA: Índice de Preços ao Consumidor.

APÊNDICE D - Análise do modelo RF – nacional e regional

Empresa	Ano	B	EIE	ROA	OLEB	OLCF	L	MVS	CALO	CL	LNA	GMNR	TTA	FEV	CPV	AT	CPB	RF- Nac.	RF- Reg.	RF Reg.	RF Nac.
																			Prob.	Prob.	
Portobello	2016	0	1,54	0,06	11,82	8,66	0,81	0,32	0,44	1,31	14,03	0,36	-0,08	0,08	-0,11	0,82	-0,06	0	1	0,63	0,48
Viavarejo	2010	0	1,48	0,03	16,31	19,82	0,74	0,86	1,15	1,46	16,11	0,22	2,43	3,64	2,80	0,87	-0,49	0	1	0,56	0,45
Melhor			menor	maior	menor	menor	menor	maior	maior	maior	maior	maior	menor	menor	menor	maior	maior				
Média (0)			0,62	0,10	7,23	5,23	0,54	1,80	1,32	1,90	15,06	0,35	0,19	0,24	0,18	0,73	0,25				
Q3 (0)			0,85	0,12	9,11	6,91	0,67	2,36	1,56	2,13	16,30	0,45	0,18	0,36	0,22	0,93	0,76				

Em que: B: situação real da empresa; Prob: Probabilidade do modelo; Nac.: Nacional; Reg.: Regional.

APÊNDICE E - Análise do modelo LR – nacional e regional

Empresa	Ano	B	ROA	LNA	L	MVS	CALO	CPV	CPB	EIE	FEV	OLEB	OLCF	TTA	CL	GMNR	AT	LR-Nac.	LR-Reg.	LR-Prob.	LR-Reg.	LR-Nac.
Bardella	2008	0	0,02	13,43	0,68	0,29	0,32	0,85	-0,12	1,14	0,74	26,48	15,39	0,46	1,33	0,15	0,75	0,51	0,44	1	0	
Bardella	2014	0	0,02	13,58	0,57	0,12	1,14	0,13	-0,14	1,29	-0,12	24,50	13,25	0,04	1,20	0,17	0,61	0,72	0,66	1	0	
Portobello	2016	0	0,06	14,03	0,81	0,32	0,44	-0,11	-0,06	1,54	0,08	11,82	8,66	-0,08	1,31	0,36	0,82	0,52	0,54	1	0	
Brasilagro	2008	1	-0,02	13,46	0,18	3,13	3,43	-0,24	-0,31	-0,17	-0,74	-11,08	-11,98	-0,06	3,83	0,12	0,01	0,19	0,26	0	1	
Cyrela Realt	2017	1	0,00	16,16	0,39	1,25	0,65	-0,21	0,20	9,96	-0,26	143,18	48,42	-0,12	3,01	0,27	0,26	0,06	0,08	0	1	
Fer Heringer	2011	1	0,12	14,77	0,82	0,25	0,63	0,31	0,04	1,56	0,70	4,21	6,00	0,28	0,98	0,14	1,81	0,30	0,27	0	1	
JBS	2008	1	0,04	16,59	0,62	0,69	1,22	0,85	-0,99	2,46	2,41	11,32	8,89	0,91	1,68	0,10	1,88	0,31	0,27	0	1	
Metalfrio	2007	1	-0,03	13,34	0,43	3,93	2,87	0,39	3,00	-1,03	0,79	-10,64	-27,29	1,45	3,43	0,14	0,92	0,33	0,42	0	1	
Viavarejo	2007	0	0,03	14,85	0,74	0,62	0,76	0,27	2,94	1,61	-0,04	18,16	15,74	0,21	1,25	0,24	1,22	0,42	0,40	0	1	
Melhor			maior	maior	menor	maior	maior	menor	maior	menor	menor	menor	menor	maior	maior	maior	maior	maior	maior			
Média (0)			0,10	15,06	0,54	1,80	1,32	0,18	0,25	0,62	0,24	7,23	5,23	0,19	1,90	0,35	0,73					
Q3 (0)			0,12	16,29	0,67	2,35	1,57	0,22	0,76	0,85	0,37	9,13	6,99	0,18	2,13	0,45	0,93					

Em que: B: situação real da empresa; Prob: Probabilidade do modelo; Nac.: Nacional; Reg.: Regional.

APÊNDICE F – Modelo de previsão LR

Variável	Coeficientes	Erro Padrão	Valor Z	Pr(> z)
CALO	0,531	0,076	6,987	0,000 ***
L	3,043	0,290	10,475	0,000 ***
TTA	1,188	0,263	4,515	0,000 ***
CPB	-0,037	0,007	-5,708	0,000 ***
AT	-0,379	0,123	-3,077	0,002 **
CL	-0,389	0,087	-4,484	0,000 ***
EIE	0,102	0,018	5,576	0,000 ***
ROA	-17,102	0,997	-17,146	0,000 ***
GMNR	-1,091	0,326	-3,343	0,001 ***
OLEB	-0,013	0,003	-5,099	0,000 ***
LNA	-0,373	0,033	-11,345	0,000 ***
MVS	-0,745	0,085	-8,754	0,000 ***
OLCF	0,005	0,001	4,566	0,000 ***
FEV	0,249	0,046	5,388	0,000 ***
CPV	-1,793	0,262	-6,846	0,000 ***

Notas: (1) **. *** indicam nível de significância a 5% e 1% respectivamente.

APÊNDICE G – Modelo de previsão RF nacional

```
library(e1071)
library(randomForest)

train<-read.xlsx("dataset.xlsx")

trctrl <- trainControl(method = "cv", number = 10, allowParallel=TRUE)

model.rf<- train(B~, data = train,
method = "rf",
trControl=trctrl)

pred.rf<- predict(model.rf, newdata = test)
cm.rf=confusionMatrix(pred.rf, test$B )
cm.rf

CrossTable(test$B, pred.rf,
prop.chisq = FALSE, prop.c = FALSE, prop.r = TRUE,
dnn = c('actual default', 'predicted default'))
```