

**UNIVERSIDADE FEDERAL DO PARANÁ  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

**FERNANDA MEDEIROS ASSEF**

**ALGORITMOS DE CLASSIFICAÇÃO EM APLICAÇÃO FINANCEIRA:  
AVALIAÇÃO DE RISCO DE CRÉDITO PARA PESSOA JURÍDICA**

**CURITIBA  
2018**

**FERNANDA MEDEIROS ASSEF**

**ALGORITMOS DE CLASSIFICAÇÃO EM APLICAÇÃO FINANCEIRA:  
AVALIAÇÃO DE RISCO DE CRÉDITO PARA PESSOA JURÍDICA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Paraná, como requisito parcial à obtenção do título de mestre em Engenharia de Produção.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Teresinha  
Arns Steiner

Co-orientação: Prof. Dr. Cassius Tadeu  
Scarpin

**CURITIBA  
2018**

---

A844a Assef, Fernanda Medeiros  
Algoritmos de classificação em aplicação financeira: avaliação de risco  
de crédito para pessoa jurídica / Fernanda Medeiros Assef. – Curitiba, 2018.  
147f. : il. [algumas color.]; 30 cm.

Dissertação (mestrado) - Universidade Federal do Paraná, Setor de  
Tecnologia, Programa de Pós-Graduação em Engenharia de Produção,  
2018.

Orientadora: Maria Teresinha Arns Steiner

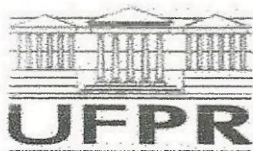
Coorientador: Cassius Tadeu Scarpin.

Bibliografia: p. 112-125.

1. Redes neurais - computação. 2. Funções de Base Radial. I.  
Universidade Federal do Paraná. II. Steiner, Maria Teresinha Arns. III.  
Scarpin, Cassius Tadeu. IV. Título.

CDD: 519.535

---



## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA DE PRODUÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **FERNANDA MEDEIROS ASSEF** intitulada: **ALGORITMOS DE CLASSIFICAÇÃO EM APLICAÇÃO FINANCEIRA: AVALIAÇÃO DE RISCO DE CRÉDITO PARA PESSOA JURÍDICA**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 20 de Fevereiro de 2018.

MARIA TERESINHA ARNS STEINER  
Presidente da Banca Examinadora (UFPR)

FABIANO OSCAR DROZDA  
Avaliador Interno (UFPR)

PEDRO JOSÉ STEINER NETO  
Avaliador Externo (UP)

MARIANA KLEINA  
Avaliador Interno (UFPR)

## AGRADECIMENTOS

Agradeço a Deus, por me conceder saúde e me guiar por caminhos seguros.

Agradeço aos meus pais, Maria Laura Medeiros Assef e Julio Cesar Assef, por me incentivarem nos momentos mais difíceis, pela compreensão e amor incondicional.

Agradeço ao meu irmão, Vinicius Medeiros Assef, pelo ombro amigo em tantas situações, pelos momentos de descontração em horas difíceis, me mostrando sempre uma luz.

Agradeço ao Anis Assad Neto, meu amor, sempre ao meu lado nos momentos de diversão e de trabalho. Sua presença faz meus dias serem melhores do que eu poderia imaginar. Obrigada pelo carinho e motivação.

Agradeço ao David Gabriel de Barros Franco pelo auxílio com o manuseio do Matlab e pela paciência com que me esclareceu as diversas dúvidas que tive.

Agradeço ao Edevar Tochetto pelo fornecimento dos dados utilizados neste trabalho e pelo auxílio na compreensão dos mesmos.

Agradeço ao Edson da Silva Lima Junior e a Rafaella Laureano Scaramella Figueira da Silva sempre prontos para colaborar com os meus pedidos tanto na graduação quanto no mestrado.

Agradeço ao Professor Dr. Cassius Tadeu Scarpin, co-orientador deste trabalho, por me incentivar à pesquisa científica.

Agradeço a todos os professores que tive durante minha vida, pelo conhecimento transmitido e pela formação humana que me dispensaram.

Agradeço, em especial, à Professora Dra. Maria Teresinha Arns Steiner, por verdadeiramente dar significado à palavra orientação. Sua competência e segurança foram fundamentais para o êxito de mais esta etapa em minha vida profissional. Obrigada pela paciência em cada ensinamento e por me mostrar que não se sabe tudo sempre.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo incentivo financeiro que foi fundamental para minha dedicação à este trabalho.

*“Não é na ciência que está a felicidade, mas na aquisição da ciência”*

Edgar Allan Poe

## RESUMO

A avaliação de risco de crédito propõe um problema clássico de tomada de decisão, onde o mérito de determinado indivíduo receber (ou não) o crédito é baseado em estimativas do potencial de devolução deste crédito acrescido pelas taxas bancárias. Desde que foi introduzida na década de 1950, a avaliação de risco tem sido amplamente utilizada. A avaliação do risco pode estar associada ao empréstimo por uma instituição financeira a uma determinada empresa (pessoa jurídica) ou a um indivíduo (pessoa física). A presente pesquisa visa analisar o primeiro caso: crédito bancário concedido a entidades jurídicas (empresas), cujas respostas, já conhecidas por meio de registros históricos, estão divididas em classes, sendo estas: clientes adimplentes (pagam seus empréstimos em dia), inadimplentes (aqueles que não pagam) e temporariamente inadimplentes (clientes que pagam, mas com atraso), com o objetivo de auxiliar a decisão a ser tomada pela analista de determinada instituição financeira quanto a conceder (ou não) o crédito a novas empresas solicitantes. Para melhor ilustrar a metodologia aqui utilizada fez-se uso de 5.432 dados (instâncias; 2.600 Adimplentes, 1.281 Inadimplentes e 1.551 Temporariamente Inadimplentes), cada um dos quais com 15 atributos, de uma grande instituição bancária brasileira. A metodologia fez uso dos seguintes métodos: Redes Neurais Artificiais (RNAs), mais especificamente, o modelo o Perceptron de Camada Múltipla (MLP) e as Funções de Base Radial (RBF) e, também, o modelo estatístico de Regressão Logística (RL). Para a implementação das RNAs, o *software* MATLAB foi utilizado e, para o modelo estatístico, foi utilizado o SPSS. O melhor desempenho foi apresentado o apresentado pelas MLP, cujas melhores acurácias foram de 74,70%; 91,4% e 74,6% para as classes “Adimplentes ou Outra”, “Inadimplentes ou Outra” e Temporariamente Inadimplentes ou Outra”, respectivamente. Desta forma, para a classificação de um novo cliente, teríamos que aplicar o modelo MLP para as três classes, verificando qual delas fornece o maior valor para a acurácia.

**Palavras-chave:** Redes Neurais Artificiais (RNA), Avaliação de Risco de Crédito, *Perceptron* de Múltiplas Camadas, Funções de Bases Radiais, Regressão Logística

## ABSTRACT

The credit risk assessment proposes a classic problem of decision making, where the merit of a given individual receives (or not) the credit is based on approximations of the potential for repayment of this credit plus bank fees. Since its introduction in the 1950s, risk assessment has been widely used. The risk assessment may be linked to the lending by a financial institution to a particular company (legal entity) or to an individual (individual). The present study aims at analyzing the first case: bank credit granted to legal entities (enterprises), whose answers, already known through historical records, are divided into classes, which are: clients default (pay their loans on time), defaulters (those who do not pay) and temporarily defaulters (customers who pay but are late), in order to assist the decision to be taken by the analyst of a particular financial institution to grant (or not) the credit to new applicant companies. In order to better illustrate the methodology used, 5,432 data were used (instances: 2,600 Default, 1,281 Defaulters and 1,551 Temporarily Defaulters), each one with 15 attributes, of a large Brazilian banking institution. The methodology used the following methods: Artificial Neural Networks (ANNs), more specifically, the Multiple Layer Perceptron (MLP) and the Radial Base Functions (RBF) model, as well as the Logistic Regression (RL). For the implementation of ANNs, MATLAB software was used and, for the statistical model, SPSS was used. The best performance was presented by the MLP, whose best accuracy was 74.70%; 91.4% and 74.6% for the "Default and Others", "Defaulters and Other" and "Temporarily Defaulters and Other" classes, respectively. Thus, for the classification of a new customer, we would have to apply the MLP model to the three classes, verifying which one provides the highest value for the accuracy.

**Keywords:** Artificial Neural Networks (ANNs), Credit Risk Assessment, Multilayer Perceptron, Radial Basis Functions, Logistic Regression.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Etapas do Procedimento da Pesquisa.....	19
Figura 2 - Esquematização de um Neurônio Biológico.....	30
Figura 3 – Modelo de um Neurônio MCP.....	31
Figura 4 – Modelo de um Neurônio Não-Linear.....	32
Figura 5 – Gráfico da Função de Limiar.....	33
Figura 6 – Gráfico da Função por Partes.....	34
Figura 7 – Gráfico da Função Sigmoide-Logística.....	34
Figura 8 - Rede Feedforward de Camada Única.....	36
Figura 9 - Rede Feedforward de Múltiplas Camadas.....	36
Figura 10 – Rede Neural Recorrente.....	37
Figura 11 – Rede Neural Multilayer Perceptron (MLP).....	40
Figura 12 - Classificação e Enquadramento da Pesquisa.....	80
Figura 13 - Acurácia e MSE das RBF para o modelo “Adimplentes ou Outra”.....	97
Figura 14 - Acurácia e MSE das RBF para o modelo “Inadimplentes ou Outra”.....	99
Figura 15 - Acurácia e MSE das RBF para o modelo “Temporariamente Inadimplentes ou Outra”.....	101

## LISTA DE TABELAS

Tabela 1 - Eixos de Divisão dos Trabalhos Correlatos .....	45
Tabela 2 - Binarização do atributo "Risco" .....	82
Tabela 3 - Binarização do atributo "Data de Constituição" .....	82
Tabela 4 - Binarização do atributo "Quantidade de Funcionários" .....	83
Tabela 5 - Resultados Obtidos pelas MLP para cada classe testada .....	89
Tabela 6 - Acurácia dos Testes com MLP para o modelo "Adimplentes ou Outra" .....	91
Tabela 7 - Acurácia dos Testes com MLP para o modelo "Inadimplentes ou Outra" ...	93
Tabela 8 - Acurácia dos Testes com MLP para o modelo "Temporariamente Inadimplentes ou Outra" .....	95
Tabela 9 - Resultados Obtidos pelas RBF para as três classes utilizadas.....	96
Tabela 10 - Resultados Obtidos pela RL considerando as três classes .....	102
Tabela 12 - Matriz de Confusão (Passo 0 - RL) .....	102
Tabela 13 - Variáveis consideradas na Equação (Passo 0).....	103
Tabela 14 - Teste Hosmer e Lemeshow .....	103
Tabela 15 - Matriz de Confusão de RL para o modelo "Adimplentes ou Outra" .....	103
Tabela 17 - Matriz de Confusão (Passo 0 - RL) .....	104
Tabela 18 - Teste Hosmer e Lemeshow .....	105
Tabela 19 - Matriz de Confusão de RL para o modelo "Inadimplentes ou Outra" .....	105
Tabela 21 - Matriz de Confusão (Passo 0 - RL) .....	106
Tabela 22 - Teste Hosmer e Lemeshow .....	106
Tabela 23 - Matriz de Confusão de RL para o modelo "Temporariamente Inadimplentes ou Outra" .....	106
Tabela 24 - Resumo das Acurácias (%) dos modelos aplicados às três classes .....	109
Tabela 25 - Variáveis Categóricas da RL no modelo: Adimplentes ou Outra .....	126
Tabela 26 - Variáveis Categóricas da RL no modelo: Inadimplentes ou Outra .....	129
Tabela 27 - Variáveis Categóricas da RL no modelo: Temporariamente Inadimplentes ou Outra .....	133
Tabela 28 - Variáveis da Regressão do modelo: Adimplentes ou Outra.....	138
Tabela 29 - Variáveis da Regressão do modelo: Inadimplentes ou Outra.....	142
Tabela 30 - Variáveis da Regressão do modelo: Temporariamente Inadimplentes ou Outra .....	145

## LISTA DE QUADROS

Quadro 1 - Amostra Parcial da Coleta de Dados realizada (20 do total de 5.432 instâncias)	24
Quadro 2 - Exemplos de Funções de Ativação	33
Quadro 3 - Funções de Bases Radiais	42
Quadro 4 – Alguns Trabalhos Correlatos com Abordagem em Falência	51
Quadro 5 – Alguns Trabalhos Correlatos com Abordagem em Mercado de Ações	60
Quadro 6 – Alguns Trabalhos Correlatos com Abordagem em Risco de Crédito	70
Quadro 7 – Alguns Trabalhos Correlatos com Aplicações Gerais de RNA	77
Quadro 8 - Funções de Treinamento do Matlab	84
Quadro 9 - Parâmetros da MLP utilizada	86
Quadro 10 - Parâmetros da RBF utilizada	87
Quadro 11 - Matriz de confusão	88

## LISTA DE ABREVIATURAS E SIGLAS

ACP	Análise de Componente Principal
ADL	Análise Discriminante Linear
ADQ	Análise Discriminante Quadrática
AG	Algoritmo Genético
ANC	Análise de Crédito
ANCP	Análise baseada em Núcleo de Componente Principal
ANOVA	Análise de Variância
ARFCP	Análise Robusta <i>Fuzzy</i> Componentes Principais
ARIMA	<i>AutoRegressive Integrated Moving Average</i>
AS	<i>Simulated Annealing</i>
BACEN	Banco Central do Brasil
BH	Busca Harmônica
B-OLS	<i>Beta Ordinary Least Squares</i>
BP	<i>Backpropagation</i>
CART	<i>Classification and Regression Trees</i>
ELM	<i>Extreme Learning Machines</i>
ENN	<i>Edited Nearest Neighbour</i>
FRAME	<i>Functional Response Additive Model Estimation</i>
GARCH	<i>Generalized Auto-Regressive Conditional Heteroskedasticity</i>
GRNN	<i>Generalized Regression Neural Networks</i>
IA	Inteligência Artificial
KELM	<i>Kernel Extreme Learning Machines</i>
KNN	k-Vizinhos mais Próximos
LMCP	Longas Memórias de Curto Prazo
LVQ	<i>Learning Vector Quantization</i>
MAPE	<i>Mean Absolut Percentage Error</i>
MARS	Regressão Adaptativa Multivariada
MCOC	Classificador de Otimização Multi-Critério
MLP	<i>Multilayer Perceptron</i>
MSE	<i>Mean Square Error</i>
NASDAQ	<i>National Association of Securities Automated Quotations</i>
NCN	<i>Nearest Centroid Neighbours Editing</i>
PSO	<i>Particle Swarm Optimization</i>
RBF	<i>Radial Basis Functions</i>
RL	Regressão Logística
RMSE	Raiz Quadrada do Erro Médio
RNA	Redes Neurais Artificiais
RNC	Redes Neurais Convolucionais
RNR	Redes Neurais Recorrentes
RP	Reconhecimento de Padrões
RVM	<i>Relevance Vector Machines</i>
SERNFIS	<i>Self Evolving Recurrent Neuro-Fuzzy Inference System</i>
SJR	<i>Scimago Journal &amp; Country Rank</i>
SMAPE	<i>Symetric Mean Absolut Percentage Error</i>

SNC	Sistema Nervoso Central
SOM	<i>Self-Organizing Maps</i>
SPSS	<i>Statistical Package for Social Sciences</i>
SVM	<i>Support Vector Machines</i>
SZSE	<i>Shenzhen Stock Exchange</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
1.1	PROBLEMATIZAÇÃO	15
1.2	OBJETIVOS	16
1.2.1	Objetivo Geral	16
1.2.2	Objetivos Específicos	17
1.3	JUSTIFICATIVA	17
1.4	LIMITAÇÕES DO TRABALHO	18
1.5	ETAPAS DA PESQUISA	19
1.6	ESTRUTURA DO TRABALHO	19
<b>2</b>	<b>DESCRIÇÃO DO PROBLEMA</b>	<b>21</b>
2.1	RISCO DE CRÉDITO	21
2.2	O PROBLEMA REAL ABORDADO	22
2.2.1	Coleta de Dados	23
<b>3</b>	<b>REVISÃO DA LITERATURA</b>	<b>25</b>
3.1	RISCO DE CRÉDITO BANCÁRIO	25
3.1.1	Legislação Específica	26
3.2	RECONHECIMENTO DE PADRÕES	27
3.3	REDES NEURAIS ARTIFICIAIS	28
3.3.1	Neurônio Artificial	31
3.3.2	Paradigmas de aprendizagem	34
3.3.3	Arquiteturas de Redes Neurais Artificiais	35
3.3.4	O aprendizado de uma Rede Neural Artificial	37
3.3.5	Modelos de Redes Neurais Artificiais	38
3.4	REGRESSÃO LOGÍSTICA	44
3.5	TRABALHOS CORRELATOS	44
3.5.1	Falência	46
3.5.2	Mercado de Ações	52
3.5.3	Risco de Crédito	62
3.5.4	Aplicações Gerais de RNA	72
<b>4</b>	<b>MÉTODO DE PESQUISA</b>	<b>79</b>
4.1	CLASSIFICAÇÃO DA PESQUISA	79
4.2	METODOLOGIA APLICADA	80
4.3	PRÉ-PROCESSAMENTO DOS DADOS COLETADOS	81
4.4	REDES NEURAIS ARTIFICIAIS NO MATLAB	83
4.5	REGRESSÃO LOGÍSTICA COM SPSS	85
4.6	PARÂMETROS DOS MÉTODOS APLICADOS	86
4.6.1	Método de Avaliação de Desempenho	87
<b>5</b>	<b>OBTENÇÃO E ANÁLISE DOS RESULTADOS</b>	<b>89</b>
5.1	RESULTADOS MLP	89
5.1.1	Situação: Adimplentes ou Outra	90
5.1.2	Situação: Inadimplentes ou Outra	92
5.1.3	Situação: Temporariamente Inadimplentes ou Outra	94
5.2	RESULTADOS RBF	96

5.2.1	Situação: Adimplentes ou Outra.....	96
5.2.2	Situação: Inadimplentes ou Outra .....	98
5.2.3	Situação: Temporariamente Inadimplentes ou Outra .....	100
5.3	RESULTADOS RL .....	102
5.3.1	Situação: Adimplentes ou Outra.....	102
5.3.2	Situação: Inadimplentes ou Outra .....	104
5.3.3	Situação: Temporariamente Inadimplentes ou Outra .....	105
<b>6</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>108</b>
	<b>REFERÊNCIAS .....</b>	<b>112</b>
	<b>APÊNDICE A – CÓDIGO DAS VARIÁVEIS CATEGÓRICAS .....</b>	<b>126</b>
	<b>APÊNDICE B – VARIÁVEIS PRESENTES NA REGRESSÃO .....</b>	<b>137</b>

## 1 INTRODUÇÃO

De acordo com o Serviço de Proteção ao Crédito (2017), em Janeiro deste ano, o número de empresas devedoras no Brasil cresceu 5,28% frente à mesma época no ano anterior. Apesar de essa variação ser levemente superior àquela registrada pelo indicador no último mês, Dezembro de 2016, o crescimento da quantidade de pessoas jurídicas no país mostrou perda de força ao longo de todo o ano de 2016. Com relação à Região Sul, o número de pessoas jurídicas devedoras cresceu 3,17% em Janeiro de 2017.

As determinações em torno do tema de crédito bancário apresentam ser um problema clássico de tomada de decisão sob incerteza: os tratamentos para indivíduos devem ser selecionados com base em estimativas de resultados futuros potenciais e outros fatores ligados à situação financeira do mesmo atualmente. Segundo Fahner (2012), as organizações desenvolvem ou aspiram desenvolver generalizações dos modelos chamados de avaliação de crédito, ou *credit scoring*, para fornecer tais estimativas de forma a prever, ou classificar, objetivos como receita, lucro, pré-pagamento, inadimplência ou perda.

Pesquisadores de diversas áreas visam facilitar as negociações do setor financeiro, disponibilizando crédito entre outros produtos financeiros, prevendo tendências, simulando o comportamento tanto financeiro quanto o de investidores, além de avaliação de objetivos, gerenciamento de carteira de ativos, preços de oferta pública inicial, determinação de estrutura de capital ótima, movimentação de preços de títulos, previsão de inadimplência e falência, entre outros tipos de problemas. Nesse sentido, são utilizados diferentes metodologias de solução para cada tipo de assunto. Entre os métodos utilizados é possível citar os estatísticos não paramétricos (sendo alguns exemplos os métodos de *k*-vizinhos mais próximos e árvores de decisão), os métodos estatísticos paramétricos (por exemplo, análise discriminante e regressão logística) e ainda abordagens denominadas de *soft-computação* (por exemplo, algoritmos inteligentes artificiais e conjuntos brutos). Os métodos de inteligência artificial (IA), especialmente Redes Neurais Artificiais (RNA), são a ferramenta mais popular usada nos mercados financeiros (BAHRAMMIRZAEI, 2010).

Passa este trabalho, será considerado então o Banco XYZ como um dos órgãos encarregados de dispor crédito, oferecendo uma gama de produtos bancários, tanto para setor público quanto para o privado, neste trabalho será apresentado um modelo para

determinação de adimplência ou inadimplência de pessoas jurídicas que fazem uso dos serviços do Banco XYZ.

Para o desenvolvimento desse trabalho será utilizada uma metodologia de IA, apresentada anteriormente, as RNA. A RNA em questão será trabalhada para realizar a classificação dos clientes em três classes: adimplentes, inadimplentes e os temporariamente inadimplentes.

## 1.1 PROBLEMATIZAÇÃO

A história da IA é uma história de fantasias, possibilidades, demonstrações e promessas. Desde que Homero, por volta do século VIII, escreveu em *Ilíada* sobre ajudantes mecanizados do deus Hefesto, os assistentes mecânicos imaginários fazem parte de nossa cultura. No entanto, apenas recentemente, considerando esse amplo histórico, foi possível construir máquinas experimentais que testassem hipóteses sobre os mecanismos de pensamento e do comportamento inteligente demonstrando, desta forma, mecanismos que anteriormente existiam apenas como possibilidades teóricas (BUCHANAN, 2006).

Como início da Neurocomputação, é muitas vezes considerado o artigo de pesquisa de McCulloch e Pitts (1943), que mostraram que mesmo tipos simples de RNA poderiam, a princípio, computar qualquer função aritmética ou lógica.

De acordo com Gurney (2004), as RNA são frequentemente usadas para análise estatística e modelagem de dados, em que seu papel é percebido como uma alternativa à regressão não linear padrão ou técnicas de análise de agrupamento. Assim, elas são tipicamente utilizadas em problemas que podem ser formulados em termos de classificação ou de previsão. Alguns exemplos incluem reconhecimento de imagem e fala (SILVA, VELLASCO e CATALDO, 2017), reconhecimento textual de caracteres (LU e SU, 2017) e domínios de especialização humana, como diagnóstico médico (ALEJO et al., 2003; PEREIRA et al., 2016; ELSALAMONY, 2017), pesquisa geológica para o petróleo (FATH, 2017) e previsão de indicadores de mercado financeiro (GÖÇKEN et al., 2016).

De acordo com Burrell e Folarin (1997), antes da utilização de RNA e outras técnicas de IA para realizar diversos tipos de avaliações voltadas a mercado financeiro se utilizavam técnicas de modelagem econométrica, como, por exemplo, a hipótese de mercado eficiente e a análise estatística.

As RNA podem ser usadas para substituir ou complementar essas técnicas tradicionais de modelagem em diferentes áreas de finanças, como previsão de desempenho de ações, previsões de séries temporais financeiras, previsão de falências, melhoria de rating de obrigações, análise de risco de empréstimos e gerenciamento de investimentos. Por causa de sua arquitetura paralela, as RNA podem superar a maioria das limitações computacionais, uma vez que são treinadas em dados de exemplo, elas são mais adaptáveis às mudanças (BAKIRCIOĞLU e KOÇAK, 2000).

No presente trabalho pretende-se trabalhar o conceito de RNA dentro do mercado financeiro, mais especificamente no problema de classificação de crédito. Diferentemente de muitos trabalhos que se preocupam apenas em verificar a “adimplência” ou “inadimplência” de clientes, será aqui abordada uma terceira classe que é a de clientes “temporariamente inadimplentes”. Esta classe representará a gama de clientes que ficam temporariamente na inadimplência, mas que não chegam a gerar prejuízos para a instituição, pois após um período de tempo conseguem pagar sua dívida.

De acordo com Lakatos e Marconi (2003), o problema de pesquisa indica exatamente qual é a dificuldade que se pretende analisar. Logo, formular o problema de pesquisa significa expor de forma clara e precisa qual a dificuldade existente. Nesse sentido, tem-se como o problema de pesquisa: “Como classificar adequadamente os clientes, incorporando uma nova classe para o problema de classificação de crédito bancário dos clientes jurídicos do Banco XYZ?”.

## 1.2 OBJETIVOS

São aqui apresentados o objetivo geral e, também, os objetivos específicos do trabalho.

### 1.2.1 Objetivo Geral

O objetivo geral deste trabalho é realizar uma comparação entre três técnicas de reconhecimento de padrões, as RNA *Multilayer Perceptron* (MLP), as RNA *Radial Basis Functions* (RBF) e por fim a Regressão Logística (RL), para a resolução do problema de classificação de risco de crédito de pessoas jurídicas, clientes de um Banco XYZ.

### 1.2.2 Objetivos Específicos

Para o cumprimento deste objetivo geral são propostos os seguintes objetivos específicos:

- Descrever as características de um problema de concessão de crédito bancário;
- Descrever as características as técnicas de Reconhecimento de Padrões utilizadas, eles sendo: RNA MLP, RNA RBF e a RL.
- Propor três situações de “um contra todos” (*one-against-all*) para os métodos a serem testados, sendo estas: adimplentes, inadimplente, e temporariamente inadimplentes;
- Gerar diferentes topologias para as RNA testadas, tendo como finalidade a obtenção do número mais adequado de neurônios na camada oculta para a resolução do problema em questão;
- Comparar as soluções obtidas em termos de acurácia e erro dos métodos avaliados em relação aos seus conjuntos de teste.

### 1.3 JUSTIFICATIVA

Os Mercados Financeiros são cruciais para a promoção da eficiência econômica pela canalização dos fundos improdutivo para atividades produtivas. Mercados financeiros eficientes são a chave para o crescimento da economia, e seu mau funcionamento é uma das causas para a pobreza em vários países do mundo (MISHKIN & EAKINS, 2012). Instituições Financeiras, como unidade de organização que compõe o mercado financeiro, são as caracterizadas como:

“Consideram-se Instituições Financeiras, para os efeitos da legislação em vigor, as pessoas jurídicas públicas e privadas, que tenham como atividade principal ou acessória a coleta, a intermediação ou a aplicação de recursos financeiros próprios ou de terceiros, em moeda nacional ou estrangeira, e a custódia de valor de propriedade de terceiros”. E complementa em seu parágrafo único: “Para os efeitos desta Lei e da legislação em vigor equiparam-se às instituições financeiras as pessoas físicas que exerçam qualquer das atividades referidas neste artigo, de forma permanente ou eventual.”. (BRASIL. Lei 4.595, 1964, art. 17).

O Banco XYZ como instituição financeira oferece diferentes produtos a seus clientes jurídicos. Alguns desses produtos visam, por exemplo, a obtenção de capital de giro de forma a ser possível realizar o pagamento de fornecedores, salários, impostos ou para equilibrar o fluxo de caixa, ou ainda para financiamento de investimentos na empresa em questão, de forma a tornar realizável a compra de bens e modernização, ou reforma, das instalações de seus clientes.

O não pagamento da dívida de um cliente é um dos principais fatores que consternam qualquer instituição financeira. Os detrimientos causados devido à inadimplência podem ser significativos, e a adoção de abordagens computacionais que visem à redução desse agente podem trazer vantagens às instituições que disponibilizam crédito.

Um parecer equivocado de um candidato durante a avaliação de crédito classificando-o como “inadimplente” tem como consequência a geração de ganhos não realizados, enquanto que, a classificação incorreta de um cliente classificado como “adimplente” tem como decorrência provável o prejuízo pela perda do valor envolvido na operação financeira.

Um aumento no volume de concessão de crédito justifica a necessidade de uma instituição financeira em optar por um sistema computacional que faça a avaliação de crédito do cliente, dessa forma, essa adoção pode levar a um ganho à empresa. Para esse fim, torna-se necessária a avaliação dos processos informáticos voltados à classificação de risco de crédito, onde devem ser analisadas as informações disponibilizadas pela empresa para o consentimento de crédito.

Com o objetivo de fundamentar o trabalho como competência de um Engenheiro de Produção, se verifica que de acordo com a Associação Brasileira de Engenharia de Produção (1998), cabe aos profissionais de Engenharia de Produção conteúdos profissionais relacionados à área de Gestão Econômica. Dessa forma, a pesquisa busca, ainda, oferecer incentivo a trabalhos ligados à Engenharia Econômica dentro de Programas de Pós-Graduação em Engenharia de Produção.

#### 1.4 LIMITAÇÕES DO TRABALHO

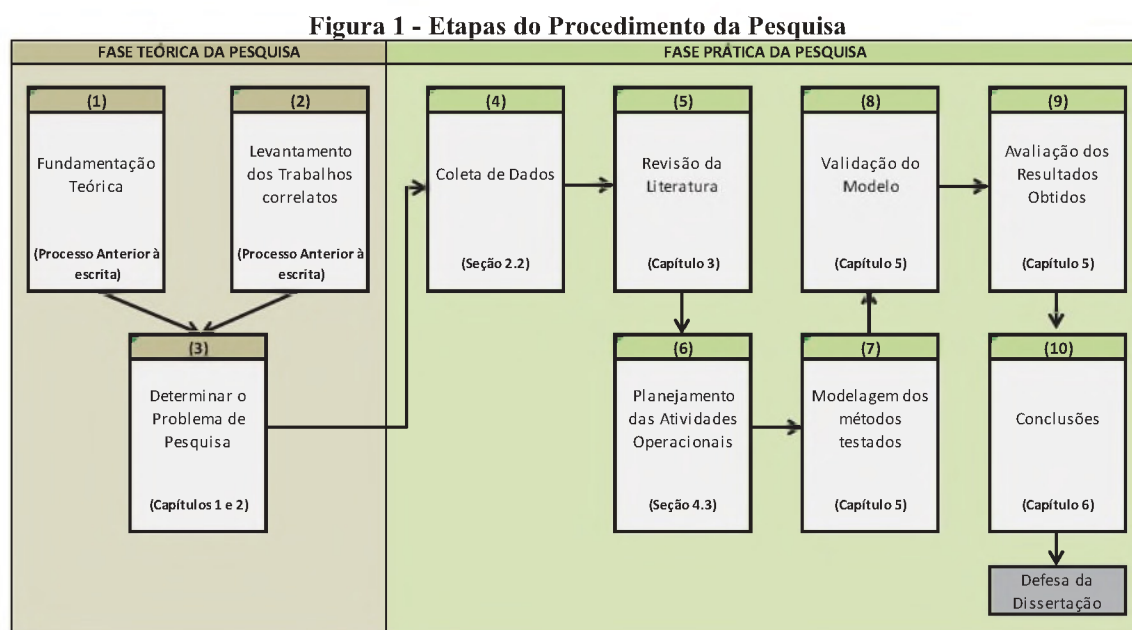
A presente pesquisa se restringe a encontrar uma solução realizável para o problema de classificação de clientes voltado aos consumidores da instituição financeira em forma de pessoas jurídicas com a aplicação de diferentes topologias de RNA do tipo

MLP, RBF e por fim uma solução através de aplicação da RL. As eventuais decisões gerenciais consequentes da categorização concluída não serão responsabilidade do trabalho aqui exposto, uma vez que a metodologia aplicada serve para o auxílio dessa tomada de decisão.

Conforme mencionado anteriormente, a classificação gerada pelos modelos propostos ocorrerá, em três diferentes classes (adimplente, inadimplente ou temporariamente inadimplente). As características dessas classes serão discutidas na seção correspondente à metodologia do trabalho.

## 1.5 ETAPAS DA PESQUISA

A Figura 1 - Etapas do Procedimento da Pesquisa a seguir mostra todo o procedimento que será adotado para a realização da pesquisa proposta. Esse procedimento é constituído, sobretudo, por 10 passos, auto-explicativos, a serem realizados para a condução da pesquisa.



FONTE: A autora, 2017.

## 1.6 ESTRUTURA DO TRABALHO

Com a finalidade de alcançar os objetivos previamente delineados na estruturação deste trabalho, o presente trabalho se apresenta segmentado em cinco capítulos.

Neste primeiro capítulo foram apresentados os elementos relacionados ao planejamento desta pesquisa incluindo, conseqüentemente, uma introdução sobre o tema, o problema abordado para pesquisa, seus objetivos, sua justificativa e as limitações do trabalho.

O capítulo 2 descreve o problema aqui abordado junto ao Banco XYZ, tratando sobre a quantidade de instâncias e de atributos que serão analisados.

O capítulo 3 versa sobre os principais temas relativos ao desenvolvimento do trabalho utilizando-se de um levantamento bibliográfico. Tal divisão compreende uma fundamentação teórica acerca do problema de Risco de Crédito, das técnicas de Reconhecimento de Padrões (RP), dos modelos utilizados de RNA, por fim, do funcionamento da RL, além de uma seção de trabalhos correlatos aos temas abordados.

No capítulo 4 são investigados os aspectos metodológicos pertencentes ao trabalho, tais como a natureza e o enquadramento da pesquisa. Ademais, é retratado qual o sistema utilizado para a coleta de dados e ainda a descrição da aplicação da metodologia proposta ao problema estudado.

No capítulo 5 são apresentados os resultados obtidos com a metodologia aplicada com os respectivos parâmetros utilizados. E, por fim, o capítulo 6 apresenta as conclusões obtidas ao longo da pesquisa realizada.

## 2 DESCRIÇÃO DO PROBLEMA

Neste capítulo discorre-se, inicialmente, sobre o risco de crédito, tema do problema real abordado neste trabalho e, em seguida, apresenta-se a descrição do problema, com suas instâncias, atributos e classes.

### 2.1 Risco de Crédito

A avaliação de risco de crédito é um problema de mineração de dados muito desafiador e importante no domínio da análise financeira. A avaliação de risco de crédito é um método utilizado para a previsão da possibilidade de concessão (ou não) de crédito. Desde que foi introduzido na década de 1950, tem sido amplamente utilizado, mais recentemente, para a concessão de empréstimo, especialmente para cartões de crédito e consentimento de crédito (LUO, KONG e NIE, 2016). Esse problema consiste basicamente na avaliação do risco associado com o empréstimo de uma organização financeira para determinada empresa ou indivíduo.

Para a construção de um modelo para este tipo de problema é preciso obter os dados históricos, que deverão refletir o desempenho das empresas nos períodos de crise econômica. A avaliação do risco de crédito para a tomada de decisão é de extrema relevância para as instituições financeiras, devido, principalmente ao elevado risco associado à decisão da concessão de crédito inadequada.

De acordo com Paleologo, Elisseeff e Antonini (2010), existem quatro diferentes tipos de classificação, resumidas a seguir:

- a. Aplicação: refere-se à avaliação da solvência para novos candidatos. Quantificam-se os riscos associados aos pedidos de crédito, avaliando os dados sociais, demográficos, financeiros e outros coletados no momento do pedido;
- b. Comportamental: envolve princípios que são semelhantes a de aplicação, com a diferença de que aborda aos clientes existentes. Os modelos comportamentais analisam os padrões comportamentais do consumidor para suportar processos dinâmicos de gerenciamento de portfólio;
- c. Coleta: é usado para dividir os clientes com diferentes níveis de insolvência em grupos, separando os que exigem ações mais decisivas daqueles que não

precisam ser atendidos imediatamente. Esses modelos diferenciam-se de acordo com o grau de inadimplência e permitem uma melhor gestão dos clientes inadimplentes.

d. Detecção de fraude: classificam os candidatos de acordo com a probabilidade relativa de que um pedido pode ser fraudulento.

Conforme o trabalho de Zhang, Gao e Shi (2014), existe uma vasta gama de opções de metodologias para a solução de problemas de classificação de risco de crédito. Esses métodos incluem principalmente regressão logística, regressão probit, análise de vizinho mais próximo, rede bayesiana, RNA, árvore de decisão, algoritmo genético, tomada de decisão de critérios múltiplos, máquina de suporte de vetores, dentre muitos outros. Os modelos de avaliação de crédito com RNA têm alta precisão, mas algumas habilidades de modelagem são necessárias - por exemplo, para projetar topologias de rede adequadas. Já os modelos baseados em Support Vector Machines (SVM) têm indicado resultados promissores na avaliação do risco de crédito, mas o classificador SVM precisa resolver um problema de programação quadrática convexa, que é muito caro computacionalmente nas aplicações do mundo real.

## 2.2 O Problema Real abordado

Os dados utilizados neste trabalho são reais e foram obtidos junto ao Banco XYZ, considerando os clientes da cidade de Curitiba – PR, Brasil, a respeito de pessoas jurídicas do estado, no que diz respeito ao crédito bancário. O Banco XYZ procura atender às necessidades desse segmento de mercado, colocando à disposição linhas de crédito tanto para capital de giro quanto para investimentos. A clientela da carteira de pessoa jurídica do Banco XYZ é constituída tanto de micro, pequenos, médios e grandes empresários.

Atualmente, o Banco XYZ utiliza, como ferramenta para realizar sua análise de crédito, um aplicativo interno chamado Análise de Crédito (ANC). O ANC utiliza de Regressão Linear Múltipla para fazer o levantamento do risco que o banco tem ao realizar o empréstimo para determinado cliente. É esse aplicativo que contém as informações cadastrais e contábeis das empresas, qual a gerência do Banco se vale para apoiar suas decisões de conceder (ou não) crédito bancário.

Neste trabalho, inicialmente contavam-se com 8.358 clientes, realizando uma limpeza nos dados, excluindo assim os que não tiveram movimentações de empréstimos, foram levantados os dados históricos de 5.432 clientes em forma de pessoa jurídica, dos quais 2.600 são comprovadamente adimplentes; 1.281 são inadimplentes; e, por fim,

1.551 clientes são considerados temporariamente inadimplentes de acordo com os especialistas do Banco XYZ.

De cada um destes clientes foram considerados 15 atributos, listados a seguir:

1. Natureza Jurídica (Sociedade Limitada; Ltda – Sociedade Cotas de Responsabilidade; Empresa Individual de Responsabilidade Ltda; Firma Individual; Empresário Individual; Sociedade Simples; Microempreendedor Individual);
2. Faturamento Bruto Anual (apresenta uma variação de 0 a R\$769.665.000,00);
3. Quantidade de Funcionários (apresenta uma variação de 0 a 5.113);
4. Atividade (apresenta uma infinidade de variações);
5. Data da Constituição;
6. Data de abertura;
7. Segmento (Micro; Pequena; Corporate; Empresa; Empresarial Middle; Empresarial Upper Middle; Estadual; PJ – não classificável / sem perfil de segmentação);
8. Risco (variação de A a E);
9. Escala de Estabelecimento de Limite de Crédito;
10. Limite Estabelecido (variação de 0 a R\$ 35.000.000,00);
11. Aplicação (variação de 0 a R\$ 76.147.526,00);
12. Débito Vencido (variação de 0 a R\$ 12.777.748,00);
13. Débito Total (variação de 0 a R\$ 32.123.211,00);
14. Endividamento Sistema Financeiro Nacional (SFN) Total (variação de 201 a R\$ 355.563.956,00);
15. Restrições (Restrições Fracas / Informações; Restrições Impeditivas; Sem Restrições).

### **2.2.1 Coleta de Dados**

No QUADRO 1 a seguir são apresentados alguns dados parciais referentes à coleta de dados realizada.

**Quadro 1 - Amostra Parcial da Coleta de Dados realizada (20 do total de 5.432 instâncias)**

#	Natureza Jurídica	Faturamento Bruto Anual	Quantidade de Funcionários	Atividade	Data de Constituição	Cliente Desde	Segmento	Risco	Escala Estabelecimento Limite de Crédito	Restrição	Situação
1	Natureza 1	2.512.000	11	Atividade 1	29/11/2004	17/05/2005	PEQUENA EMPRESA	C	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO IMPEDITIVA	ADIMPLLENTE
2	Natureza 1	3.893.314	6	Atividade 2	11/03/1992	23/01/2006	PEQUENA EMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	ADIMPLLENTE
3	Natureza 1	4.638.333	30	Atividade 3	09/09/1988	23/11/2004	PEQUENA EMPRESA	B	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO IMPEDITIVA	ADIMPLLENTE
4	Natureza 1	13.227.379	0	Atividade 4	17/12/2007	10/02/2013	EMPRESA	B	AGÊNCIAS/ESCRIT. PRIVATE-CRÉDITO	RESTRICÇÃO IMPEDITIVA	INADIMPLLENTE
5	Natureza 1	672.080	0	Atividade 5	23/05/2013	25/11/2013	MICROEMPRESA	B	AGÊNCIAS/ESCRIT. PRIVATE-CRÉDITO	SEM RESTRICÇÕES	TEMPORARIAMENTE
6	Natureza 1	1.415.162	6	Atividade 6	23/05/2013	02/01/2013	PEQUENA EMPRESA	E	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO IMPEDITIVA	INADIMPLLENTE
7	Natureza 1	416.889	5	Atividade 7	25/06/2012	02/01/2013	PEQUENA EMPRESA	E	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO FRACA/INFO	ADIMPLLENTE
8	Natureza 1	914.286	3	Atividade 8	06/02/1975	02/10/1976	MICROEMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO FRACA/INFO	ADIMPLLENTE
9	Natureza 2	87.834	1	Atividade 9	27/04/1973	24/03/1977	MICROEMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO FRACA/INFO	INADIMPLLENTE
10	Natureza 3	216.061	3	Atividade 10	04/01/1986	20/05/1986	MICROEMPRESA	B	AGÊNCIAS/ESCRIT. PRIVATE-CRÉDITO	RESTRICÇÃO FRACA/INFO	INADIMPLLENTE
11	Natureza 3	231.716	6	Atividade 11	07/08/1985	01/11/1989	MICROEMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO FRACA/INFO	ADIMPLLENTE
12	Natureza 1	2.111.938	12	Atividade 12	17/07/1989	20/07/1990	MICROEMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO IMPEDITIVA	ADIMPLLENTE
13	Natureza 1	251.120	14	Atividade 13	07/05/2001	05/09/2014	PEQUENA EMPRESA	C	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	ADIMPLLENTE
14	Natureza 1	155.800	1	Atividade 14	25/07/1991	27/07/1999	MICROEMPRESA	C	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO IMPEDITIVA	INADIMPLLENTE
15	Natureza 1	1.602.162	8	Atividade 15	05/11/2001	21/01/2002	MICROEMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	ADIMPLLENTE
16	Natureza 2	2.658.491	19	Atividade 16	28/04/2004	18/06/2004	PEQUENA EMPRESA	A	AGÊNCIAS/ESCRIT. PRIVATE-CRÉDITO	SEM RESTRICÇÕES	TEMPORARIAMENTE
17	Natureza 1	7.825.019	20	Atividade 17	29/05/1992	07/04/2005	PEQUENA EMPRESA	A	AGÊNCIAS/ESCRIT. PRIVATE-CRÉDITO	SEM RESTRICÇÕES	TEMPORARIAMENTE
18	Natureza 1	121.000	0	Atividade 18	03/11/2003	29/07/2005	EMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	TEMPORARIAMENTE
19	Natureza 4	3.020.230	17	Atividade 19	27/05/1988	18/10/2005	MICROEMPRESA	B	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	ADIMPLLENTE
20	Natureza 1	2.028.000	25	Atividade 20	14/09/2004	31/07/2006	PEQUENA EMPRESA	A	AGÊNCIAS/ESCRIT. PRIVATE-CRÉDITO	SEM RESTRICÇÕES	TEMPORARIAMENTE
21	Natureza 1	2.810.498	3	Atividade 21	23/08/1993	16/06/2006	PEQUENA EMPRESA	B	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	ADIMPLLENTE
22	Natureza 1	2.037.535	11	Atividade 22	15/05/2006	11/09/2011	PEQUENA EMPRESA	E	DICRE/9969-L.C. VAR/MP/ SUBCOM C	RESTRICÇÃO IMPEDITIVA	INADIMPLLENTE
23	Natureza 1	271.472	1	Atividade 23	05/09/2006	07/03/2006	PEQUENA EMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	ADIMPLLENTE
24	Natureza 1	615.746	4	Atividade 24	06/02/2006	08/02/2006	MICROEMPRESA	C	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO FRACA/INFO	TEMPORARIAMENTE
25	Natureza 1	199.350	0	Atividade 25	21/08/2001	14/09/2006	MICROEMPRESA	B	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO FRACA/INFO	ADIMPLLENTE
26	Natureza 1	775.100	3	Atividade 26	17/04/1990	11/01/2006	MICROEMPRESA	B	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	TEMPORARIAMENTE
27	Natureza 4	2.376.895	12	Atividade 27	21/12/1994	21/05/2009	MICROEMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	ADIMPLLENTE
28	Natureza 1	13.296.652	45	Atividade 28	11/01/1986	11/10/2006	PEQUENA EMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	INADIMPLLENTE
29	Natureza 1	918.700	1	Atividade 29	20/09/2006	27/11/2006	EMPRESA	A	DICRE/9969-L.C. VAR/MP/ SUBCOM C	RESTRICÇÃO FRACA/INFO	ADIMPLLENTE
30	Natureza 1		13	Atividade 30	05/03/2006	01/09/2007	MICROEMPRESA	B	AGÊNCIA-SISTEMA-CRÉDITO	SEM RESTRICÇÕES	TEMPORARIAMENTE
					26/02/2002	14/02/2007	PEQUENA EMPRESA	A	AGÊNCIA-SISTEMA-CRÉDITO	RESTRICÇÃO FRACA/INFO	TEMPORARIAMENTE

FONTE: A autora, 2017.

### 3 REVISÃO DA LITERATURA

Fazendo uso de uma breve revisão bibliográfica que, de acordo com Köche (2011), representa uma pesquisa que deverá explicar determinado problema, com utilização do conhecimento acessível baseado em teorias publicadas, neste capítulo será introduzido o referencial teórico que serve de embasamento ao presente trabalho. Com a fundamentação teórica obtida aliada a simulações computacionais das técnicas aqui tratadas, este estudo objetiva alcançar uma técnica eficaz na classificação de clientes em forma de pessoa jurídica entre as três classes apresentadas anteriormente.

A seguir, será apresentada uma fundamentação teórica a respeito tanto do problema de risco de crédito bancário, quanto do estudo de Reconhecimento de Padrões, bem como Redes Neurais Artificiais e seu funcionamento, além de um levantamento de trabalhos correlatos a respeito de aplicações relacionadas ao tema em questão. A pergunta a ser respondida pela revisão de literatura é relativa às lacunas na literatura a respeito de RNA em Aplicações Financeiras.

#### 3.1 RISCO DE CRÉDITO BANCÁRIO

Até a década de 70, instituições financeiras focavam em metodologias semi-quantitativas para a análise de medida e administração do crédito, que eram, sobretudo, avaliações subjetivas do crédito de determinado cliente por meio, basicamente, de análise de demonstração financeira (ALTMAN e SAUNDERS, 1998).

A avaliação de risco de crédito propõe um problema clássico de tomada de decisão sob incerteza, onde o mérito de determinado indivíduo receber, ou não, o crédito é baseado em estimativas do potencial de renda futura (FAHNER, 2012). O risco de crédito propriamente dito é o risco ligado ao não pagamento de um empréstimo, ou outra linha de crédito, por parte do devedor (CHEN, WANG e WU, 2010). De acordo com o autor, uma vez que novidades financeiras e suas derivativas aumentam rapidamente na competitiva indústria financeira, a medida de risco de crédito e sua gestão.

O crédito tem uma atribuição importante tanto na vida de boa parte das pessoas, bem como para as organizações que envolvem investimento monetário de alguma forma (CHIRINKO, GUILL e HEBERT, 1991; KOCH e MACDONALD, 2010; SELWYN,

1999). Segundo Feng *et al.* (2014), a obtenção de crédito é essencial para a suavização das operações.

A avaliação de crédito tem sua complexidade devido à variedade de riscos envolvidos (BABIC e PLAZIBAT, 1998; SELWYN, 1999). Além de ter a possibilidade de recusar o crédito a um cliente que seria merecedor, o detentor do crédito frequentemente opta por garantir a segurança do pagamento recusando então os empréstimos mais arriscados (ALTMAN e SAUNDERS, 1998). De acordo com Ignatius *et al.* (2016) a determinação do merecimento de crédito de um cliente não é realizada de forma tão direta em consequência da possibilidade do cliente possuir múltiplas (talvez até não declaradas) fontes de rendimento, além do fato da exposição de risco ser dependente da ocupação do cliente e do próprio estilo de vida.

Segundo Zhang, Tadikamalla, e Shang (2016), a avaliação de risco de crédito é essencial para o controle dos próprios riscos da organização detentora do crédito. Para estes autores, é necessário, portanto, um sistema rigoroso e conciso que seja capaz de avaliar a confiabilidade e a competência para pagamento do tomador de empréstimo. A análise de risco de crédito não apenas pode detectar o risco do crédito com antecedência e realizar assim as melhores decisões, mas também aperfeiçoar o portfólio do investimento das instituições que concedem crédito.

Para Zhang, Tadikamalla, e Shang (2016), os métodos de avaliação de crédito podem ser classificados em cinco categorias, sendo elas: Métodos Estatísticos (BOLTON, 2009; DIAKOULAKI, MAVROTAS e PAPAYANNAKIS, 1992); Teoria de Decisão (MESSIER e HANSEN, 1985; BRYANT, 2001; ZHANG *et al.*, 2010); Redes Neurais Artificiais (DESAI, CROOK e OVERSTREET Jr., 1996; WEST, 2000; PAVLENKO e CHERNYAK, 2010); *Support Vector Machines* (GESTEL *et al.*, 2003; BELLOTI e CROOK, 2009; ZHANG, GAO e SHI, 2014); e por fim, Análise por Envoltória de Dados (SHANMUGAM e JOHNSON, 2007; IAZZOLINO, BRUNI e BERARDI, 2013).

### **3.1.1 Legislação Específica**

No Brasil, a concessão de crédito bancário é subordinada à legislação do Banco Central do Brasil (BACEN). Esta instituição financeira criada em dezembro de 1.964, disciplinada pela Lei nº 4.595 de 31 de dezembro de 1.964, tem autonomia federal vinculada ao Ministério da Fazenda, com sede e foro em Brasília e atua em todo o território brasileiro. Das atribuições desta instituição financeira temos:

“O Banco Central tem por finalidade a formulação, execução, o acompanhamento e o controle das políticas monetária, cambial, de crédito e de relações financeiras com o exterior; a organização, disciplina e fiscalização do Sistema Financeiro Nacional (SFN) e do Sistema de Consórcio; a gestão do Sistema de Pagamentos Brasileiro (SPB) e de serviços do meio circulante.” (BANCO CENTRAL DO BRASIL, 2017).

Na resolução nº 3.721 do Banco Central do Brasil, em seu artigo 1º, a instituição resolve que:

“As instituições financeiras e demais instituições autorizadas a funcionar pelo Banco Central do Brasil devem implementar estrutura de gerenciamento do risco de crédito compatível com a natureza das suas operações e a complexidade dos produtos e serviços oferecidos e proporcional à dimensão da exposição ao risco de crédito da instituição”.

Na mesma resolução, o BACEN define risco de crédito como a probabilidade de perdas associadas ao não cumprimento pelo tomador ou contraparte das obrigações financeiras dentro dos termos contratados, à desvalorização de contrato de crédito decorrente de perecimento categorização de risco do tomador, à redução de ganhos ou remunerações, às vantagens concedidas na renegociação e aos custos de recuperação.

### 3.2 RECONHECIMENTO DE PADRÕES

O RP é a disciplina científica cujo objetivo é a classificação de objetos em um número de categorias ou classes, sendo que esses objetos são referidos usando padrões (instâncias) como termo genérico. Como visto em Theodoridis e Koutroumbas (2009), o RP tem uma longa história, mas antes da década de 1960 foi principalmente parte da produção de pesquisa na área de estatística. Como tudo mais, o advento dos computadores aumentou a demanda por aplicações práticas de RP, o que, por sua vez, estabeleceu novas demandas para novos desenvolvimentos teóricos.

O RP está aplicado em diversas áreas como: reconhecimento de caracteres e imagens (KARCZMAREK, KIERSZTYN e PEDRYCZ, 2017; SHIVAKUMARA *et al.*, 2017.), diagnósticos com auxílio computacional (YE e FU, 2016; JHA e TOPOL, 2016; RAITH *et al.*, 2017), reconhecimento de voz (SIMMONS, CAIRD e STEEL, 2017), mineração de dados (WITTEN *et al.*, 2016; ARMENGOL, BOIXADER e GRIMALDO, 2017) e descoberta de conhecimento em bases de dados (CHAUHAN e JAISWAL, 2016; ABDUL RHAMAN *et al.*, 2016).

Particularmente, a mineração de dados, que visa à obtenção de padrões em bases de dados, é de intenso interesse em uma ampla gama de aplicações como medicina,

biologia, análise de mercado financeiro, gestão de negócios, exploração de ciência, recuperação de imagens e ainda na música. Sua popularidade deriva do fato de que na era da informação há uma demanda cada vez maior para recuperar informações e transformá-las em conhecimento. Essas informações atualmente existem em grandes quantidades de dados e nas mais variadas formas, incluindo texto, imagens, áudio e vídeo, armazenados em diferentes lugares distribuídos em todo o mundo (THEODORIDIS e KOUTROUMBAS, 2009).

### 3.3 REDES NEURAIS ARTIFICIAIS

Uma RNA consiste em um modelo computacional que simula o comportamento do neurônio biológico, aprendendo a partir da alteração de seu estado interno a partir de *inputs* provenientes do ambiente externo. Segundo Haykin (2001):

“Uma RNA é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torna-lo disponível para o uso.”

RNA são modelos computacionais baseados na estrutura e no funcionamento do cérebro biológico que, apesar de não rodar programas, controla o comportamento do indivíduo, sendo responsável por seu pensamento, percepção, cognição e emoção (BRAGA, CARVALHO e LUDEMIR, 2000).

O sucesso na utilização das RNAs para resolução de diversos problemas se deve a algumas características de sua metodologia usada na resolução desses problemas, tais como esses apresentados em Bittencourt (2006):

- a. Ter capacidade de aprender através de exemplos, generalizando o aprendizado para reconhecer novas instâncias similares;
- b. Apresentar bom desempenho em tarefas onde o conhecimento do especialista não está disponível;
- c. Ser adequada a problemas onde não existe conhecimento a respeito dos modelos matemáticos dos domínios de aplicação;
- d. Possuir elevada imunidade ao ruído, sendo capaz de manter o desempenho em presença de algumas informações falsas ou ausentes;
- e. Apresentar a possibilidade de simulação de raciocínio “a priori” e impreciso.

O trabalho de McCulloch e Pitts publicado, em 1943, foi o princípio da neurocomputação, dessa forma sendo o início da concretização de RNA. Em 1949, Hebb escreveu um livro intitulado “The Organization of Behavior” que buscava a ideia de que o condicionamento psicológico clássico é onipresente em animais porque é uma propriedade de neurônios individuais. Essa ideia não era nova, mas Hebb levou mais longe do que qualquer outro antes dele. Ele propôs uma lei de aprendizagem específica para as sinapses de neurônios (YADAV, YADAV e KUMAR, 2015).

O primeiro neuro-computador bem sucedido (Mark I Perceptron) foi desenvolvido durante 1958 por Frank Rosenblat. Como é conhecido hoje, Rosenblatt é o fundador da Neurocomputing. Seu principal interesse era o RP (ROSEMBLAT, 1958) que, por sua vez, é o principal objetivo das RNA.

Selfridge (1958), em seu trabalho intitulado “Pandemonium: a paradigm for learning”, dissertou a respeito de um programa de manipulação de símbolos para RP. O referido autor discute sobre a aprendizagem e uma abordagem multiagente para a resolução de problemas. Seus trabalhos introduziram conceitos em RNA que foram importantes no princípio da pesquisa não apenas para a RNA, mas também, para o RP e para a aprendizagem de máquinas.

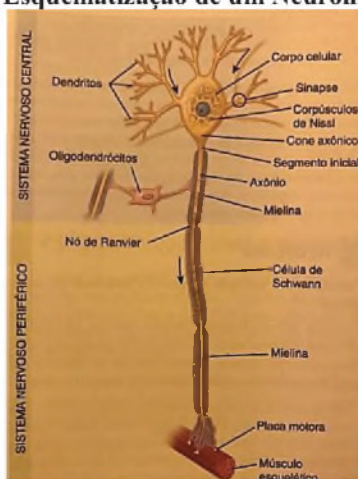
O termo "redes neurais" é muito evocativo. Sugere máquinas que agem como cérebros e é potencialmente carregado com as conotações da ficção científica do mito de Frankenstein. Uma RNA é uma composição interconectada de simples elementos de processamento, chamados de unidades ou nós, cuja funcionalidade é vagamente baseada no neurônio animal. A capacidade de processamento da rede é armazenada nas forças de ligação interunidades, ou pesos, obtidos por um processo de adaptação, ou por meio da aprendizagem a partir de um conjunto de padrões de formação. (GURNEY, 2004).

O neurônio animal, neurônio biológico, é a unidade estrutural e funcional do sistema nervoso. De acordo com Pawlina (2016), embora essas unidades exibam a maior variação de tamanho e formato dentre qualquer grupo de células do corpo, elas podem ser agrupadas em três categorias gerais.

- a. Os neurônios sensitivos são responsáveis por transmitir impulsos dos receptores para o Sistema Nervoso Central (SNC);
- b. Os neurônios motores transmitem impulsos do SNC ou dos gânglios – acúmulos de corpos celulares de neurônios situados fora do SNC – para as células efetoras, células que exercem funções imunológicas;
- c. Os interneurônios formam uma rede de comunicação e de integração entre os sensitivos e os motores;

Segundo Pawlina (2016), são encontrados diferentes tipos de neurônios biológicos, sendo eles neurônios unipolares, bipolares, pseudounipolares e multipolares. A título de ilustração, utilizaremos para a explicação dos componentes do neurônio, o classificado como multipolar uma vez que constitui a maioria dos neurônios no tecido nervoso. Esses neurônios possuem um corpo celular, onde se encontra o núcleo da célula, vários dendritos, que têm a responsabilidade de receber os sinais, e um axônio, responsável pela condução dos impulsos elétricos que partem do corpo celular. A estrutura do neurônio apresentado encontra-se na Figura 2 - Esquemática de um Neurônio Biológico, a seguir.

**Figura 2 - Esquemática de um Neurônio Biológico**



FONTE: PAWLINA, 2016

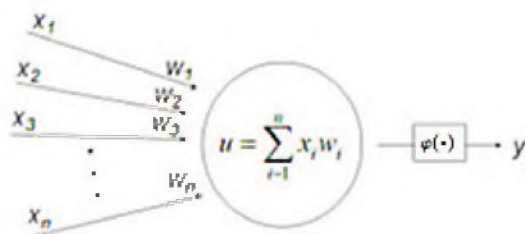
As Redes Neurais Artificiais tradicionais codificam então informações utilizando a taxa de disparo (nível de atividade) dos neurônios biológicos, e os *outputs* dos neurônios geralmente são expressos como variáveis análogas em dado intervalo. (LIN, WANG e HAO, 2017).

### 3.3.1 Neurônio Artificial

O primeiro neurônio artificial foi um modelo matemático, proposto pelos pesquisadores McCulloch e Pitts como uma simplificação do neurônio biológico, para simular o comportamento dos neurônios (BRAGA, CARVALHO e LUDEMIR, 2000). Esse modelo ficou conhecido como neurônio MCP, por conta dos nomes dos autores.

A Figura 3 abaixo apresenta o neurônio MCP. Ele possui diversas entradas (dendritos) ponderadas por determinados pesos – que seriam o comportamento das sinapses – e um único canal (axônio), que é o condutor da saída binária até as extremidades que distribuem essa mesma saída aos neurônios qual estejam ligadas.

Figura 3 – Modelo de um Neurônio MCP



FONTE: Adaptado de REZENDE, 2003.

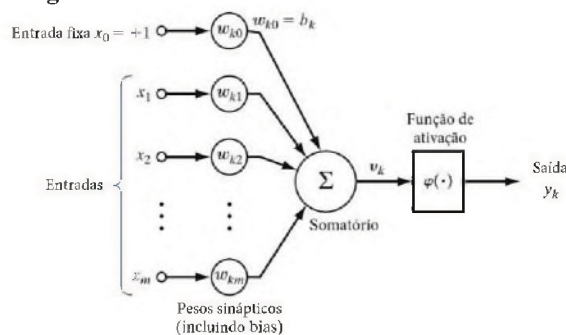
Onde  $x_i$  (com  $i = 1, \dots, n$ ) é a entrada,  $w_1, w_2, \dots, w_n$  são os respectivos pesos,  $u$  é a saída do combinador linear,  $\varphi(\cdot)$  é a função de ativação e, por fim,  $y$  é a saída do neurônio.

No neurônio MCP, mostrado na Figura 3, a saída linear ( $u$ ) é o resultado do combinador linear que calcula o produto escalar (ou produto interno) das entradas ( $x_i$ ) e dos pesos ( $w_i$ ). Esses pesos não são ajustáveis, ou seja, são fixos, não podendo ser modificados, temos também a função de ativação ( $\varphi(\cdot)$ ) que deve ser linear. (REZENDE, 2003)

A partir do neurônio MCP foram propostos outros modelos de neurônio que permitem a saída de outros valores além de zero ou um. O neurônio não-linear é um desses modelos.

O neurônio não-linear mais utilizado atualmente pela comunidade científica tem sua representação gráfica mostrada na Figura 4. Em comparação com o neurônio MCP, as principais mudanças feitas no neurônio não-linear são os pesos ( $w_i$ ) que são ajustáveis, a função de ativação ( $\varphi(\cdot)$ ) que pode ser não linear e o termo *bias* que é acrescido ao produto escalar das entradas ( $x_i$ ) pelos pesos ( $w_i$ ) (MODOLO, 2016).

**Figura 4 – Modelo de um Neurônio Não-Linear**



FONTE: Adaptado de HAYKIN, 2001.

Boukadida et al. (2011) apresentam que, da mesma forma que sua contraparte biológica, o neurônio artificial é a unidade básica que compõe a RNA. Os elementos básicos que compõem o neurônio artificial são:

- i. Um conjunto de nós de entrada que recebe o sinal de entrada correspondente ou um vetor de padrões;
- ii. Um conjunto de conexões sinápticas cujas forças são representadas por um conjunto de pesos;
- iii. Uma função de ativação  $\varphi(\cdot)$  que relaciona a entrada sináptica total com a saída (ativação) do neurônio.

Além desses elementos, há também o *bias* que é um valor positivo ou negativo que é aplicado externamente, ou seja, não faz parte das entradas recebidas pela rede ou de outros neurônios. Seu valor aumenta ou diminui a entrada líquida da função de ativação, de acordo com seu valor positivo ou negativo, respectivamente. O termo *bias* permite deslocar o hiperplano da origem no eixo das abscissas e, com sua utilização, o limiar não precisa ser alterado durante o treinamento, apenas os pesos (HAYKIN, 2001). Este autor ainda descreve matematicamente o neurônio artificial  $k$  da Figura 4 pelas equações (1) e (2) a seguir.

$$i_k = \sum_{j=1}^m w_{kj}x_j + \theta_k \quad (1)$$

$$u_k = \varphi(i_k) \quad (2)$$

Onde  $x_1, x_2, \dots, x_m$  são as entradas (atributos);  $w_{k1}, w_{k2}, \dots, w_{kt}$  são os pesos sinápticos;  $i_k$  é o campo local induzido;  $\theta_k$  é o *bias*;  $\varphi(\cdot)$  é a função de ativação definida posteriormente; e  $a_k$  é o sinal de saída, todos relacionados ao neurônio  $k$ .

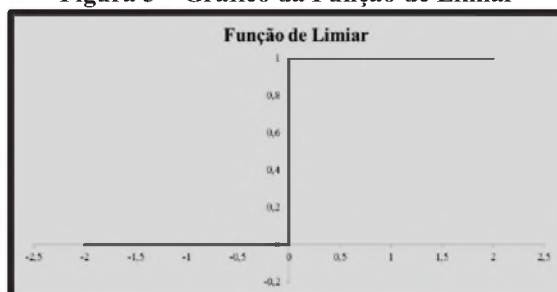
As características do neurônio não-linear permitem usar diferentes tipos de funções de ativação, facilitando a adequação da saída do neurônio ao tipo do problema. A Quadro 2 mostra quatro tipos de funções de ativação que podem ser utilizadas no neurônio não-linear: limiar, linear por partes, sigmoide logística e tangente hiperbólica (QUILES, 2004).

<b>Quadro 2 - Exemplos de Funções de Ativação</b>	
TIPO DE FUNÇÃO	FUNÇÃO
<b>Limiar</b>	$\varphi(\cdot) = \begin{cases} 1, & \text{se } u \geq 0 \\ 0, & \text{se } u < 0 \end{cases}$
<b>Linear por Partes</b>	$\varphi(\cdot) = \begin{cases} 1, & \text{se } u \geq \frac{1}{2} \\ u, & \text{se } -\frac{1}{2} < u < \frac{1}{2} \\ 0, & \text{se } u \leq -\frac{1}{2} \end{cases}$
<b>Sigmoide Logística</b>	$\varphi(\cdot) = \frac{1}{1 + e^{-au}}$
<b>Tangente Hiperbólica</b>	$\varphi(\cdot) = \tanh u$

FONTE: Adaptado de QUILES, 2004.

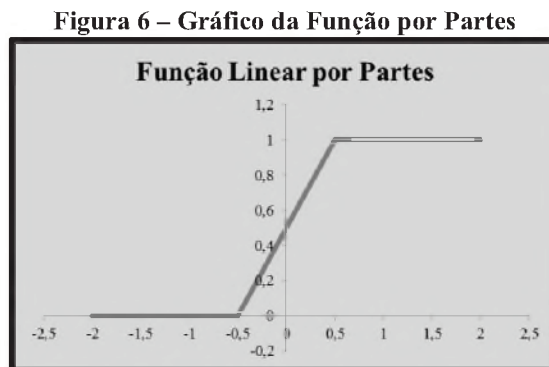
A Função de Limiar define a saída do neurônio para “0” se o argumento de função for menor que “0”, ou “1” se seu argumento for maior ou igual a “0”. Esta função de ativação está representada pela Figura 5 a seguir.

**Figura 5 – Gráfico da Função de Limiar**



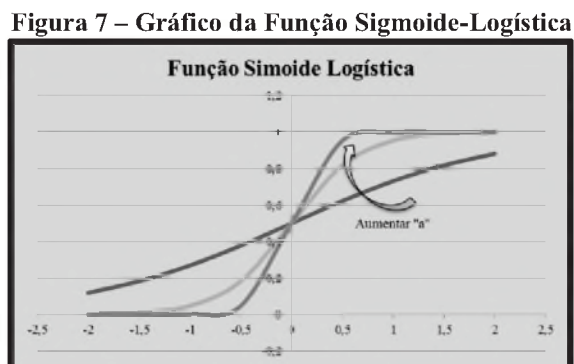
FONTE: A autora, 2017.

A função linear por partes é uma variação da função limiar que cresce linearmente e permite, além da saída dos valores zero e um, a saída de valores reais no intervalo de 0” a “1”. O gráfico desta função é representado pela Figura 6 a seguir.



FONTE: A autora, 2017.

A função sigmoide logística, apresentada pela Figura 7, também tem como saída valores reais no intervalo entre zero e um. O parâmetro  $\alpha$  permite que se obtenham diferentes inclinações. Ela é a função mais utilizada atualmente na construção de RNAs porque exibe um balanceamento adequado entre o comportamento linear e o comportamento não-linear (MÓDOLO, 2016).



FONTE: A autora, 2017.

### 3.3.2 Paradigmas de aprendizagem

De acordo com Braga, Carvalho e Ludemir (2000), existem dois fundamentos que se distinguem conforme as características do conjunto de exemplos utilizados no treinamento. Quando cada exemplo possui uma saída esperada, a aprendizagem é

apresentada como supervisionada e quando não existe uma saída esperada a aprendizagem é considerada não-supervisionada.

Na aprendizagem supervisionada existe um supervisor externo que fornece a saída desejada para cada entrada da rede. O objetivo é ajustar os parâmetros da rede de maneira que, para cada entrada, a saída da rede seja aquela fornecida. Cada saída da rede tem seu erro calculado que é usado pelo algoritmo de treinamento para ajuste dos pesos sinápticos visando minimizar esse erro (MÓDOLO, 2016).

Ainda segundo este autor, esse paradigma de aprendizagem é capaz de realizar tarefas como classificação de padrões e aproximação de funções. A regra de aprendizagem por correção de erro é um exemplo desse paradigma. O modelo de RNAs *Perceptron* e, também, o *Perceptron* de Múltiplas Camadas se enquadram nesse paradigma.

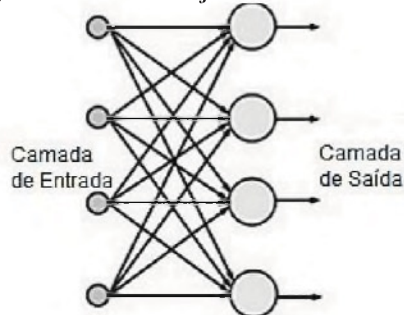
Já a aprendizagem não-supervisionada funciona sem um supervisor que forneça a saída esperada ou acompanhe o processo de aprendizado. O conjunto de exemplos usados no treinamento contém apenas os padrões de entrada. O algoritmo de treinamento busca criar novos grupos automaticamente estabelecendo uma harmonia com as regularidades estatísticas das entradas e formando novas representações internas. O algoritmo somente consegue encontrar os padrões ou características no conjunto de exemplos de entrada se houver uma redundância nos dados de entrada (FREEMAN e SKAPURA, 1991).

### **3.3.3 Arquiteturas de Redes Neurais Artificiais**

A arquitetura que será usada na RNA depende do tipo de problema que será tratado. Tanto neurônios biológicos quanto artificiais, pelo menos isoladamente, possuem baixa capacidade computacional, mas, quando conectados formando uma rede, têm a possibilidade de resolver problemas de alta complexidade (BRAGA, CARVALHO e LUDEMIR, 2011).

Na forma mais simples de uma rede em camadas, pode existir uma camada de entrada de nós de origem que se projeta diretamente para uma camada de saída de neurônios, mas não vice-versa (HAYKIN, 2009). Essa arquitetura é denominada de rede *feedforward* de camada única, representada pela Figura 8 a seguir.

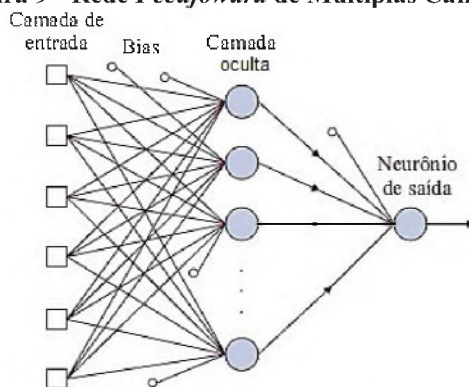
**Figura 8 - Rede *Feedforward* de Camada Única**



FONTE: Adaptado de Haykin, 2001.

A segunda classe de uma rede neuronal *feedforward* distingue-se pela presença de uma ou mais camadas ocultas, operando em paralelo, cujos nós de computação são correspondentemente chamados de neurônios ocultos. O termo "oculto" refere-se ao fato de que esta parte da RNA não é vista diretamente da entrada ou da saída da rede. Esta topologia é representada pela Figura 9 a seguir.

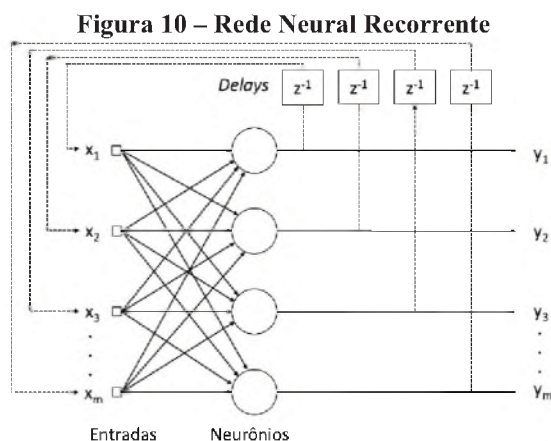
**Figura 9 - Rede *Feedforward* de Múltiplas Camadas**



FONTE: Adaptado de Haykin, 2009.

Como podemos ver de acordo com a Figura 9 anterior, os nós (neurônios) na camada de entrada da rede fornecem informações a respeito das instâncias, que constituem os sinais de entrada aplicados aos neurônios da segunda camada (camada oculta). Para a saída, os sinais da segunda camada são utilizados como entradas para a terceira camada, e assim por diante para o resto da rede (HAYKIN, 2009).

Redes Neurais Recorrentes (RNR), em sua essência, são redes neurais que empregam repetição, onde, basicamente, usando informações de uma iteração anterior utiliza-a de entrada para a RNA. Todas as Redes Neurais Recorrentes podem ser descritas como um relacionamento de recorrência (KETKAR, 2017). A Figura 10 apresenta uma imagem da RNR.



FONTE: Adaptado de Haykin, 2009.

### 3.3.4 O aprendizado de uma Rede Neural Artificial

Como aprender é um conceito usado em várias áreas, não existe uma definição de aprendizagem que seja consenso. Mas, no contexto das RNAs pode-se definir aprendizagem como:

“Aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre.” (HAYKIN, 2001).

A equação (3) a seguir define a atualização de pesos da RNA na iteração  $n$ .

$$w_{kj}(n + 1) = w_{kj}(n) + \Delta w_{kj}(n) \quad (3)$$

Onde  $w_{kj}$  representa os pesos do neurônio  $k$  e  $\Delta w_{kj}$  os ajustes dos pesos, todos com  $0 < j < m$ , com  $j = 0$  representando o *bias*. A regra de aprendizagem, normalmente referida como regra delta, usada para calcular a variação do peso ( $\Delta w$ ) considera, além dos valores da entrada ( $x$ ) e do erro ( $e$ ), o valor da taxa de aprendizagem ( $\eta$ ), que é uma constante positiva do conjunto dos números reais no intervalo entre 0 e 1, usada para ajustar a proporção do erro ( $e$ ) que será usada para cálculo do peso ( $w$ ). O cálculo da variação do peso ( $\Delta w$ ) é dado pela equação (4).

$$\Delta w_{kj}(n) = \eta \cdot e_k(n) \cdot x_j(n) \quad (4)$$

Segundo Haykin (2000), “os ajustes corretivos dos pesos são projetados para aproximar passo a passo o sinal de saída  $y_k$  da resposta desejada  $d_j$ ”. Para tanto, usa-se

minimizar a função do erro quadrático médio (5) (BRAGA, CARVALHO e LUDEMIR, 2011; HAYKIN, 2001).

$$\varepsilon(n) = \frac{1}{N} \sum_{k=1}^N e_k^2(n) \quad (5)$$

A taxa de aprendizagem tem uma influência muito grande para que seja alcançada a estabilidade e convergência do processo de aprendizagem interativo. O tipo de aprendizagem aqui descrito é a aprendizagem por correção de erro é utilizada nas RNAs *Perceptron* e também nas RNAs *Perceptron* de Múltiplas Camadas.

### 3.3.5 Modelos de Redes Neurais Artificiais

Vários modelos de RNAs foram desenvolvidos com variações no processo de aprendizagem e na arquitetura da rede. Cada um dos modelos de aprendizagem tem seu algoritmo de treinamento, de acordo com as regras e paradigmas de aprendizagem adotados, além de assumir uma determinada arquitetura de rede, de acordo com o número de camadas, o tipo de conexão entre os neurônios e a topologia da rede (MÓDOLO, 2016).

Os modelos clássicos, *Perceptron* de uma única camada e *Perceptron* de Múltiplas Camadas (MLP), adotam a aprendizagem supervisionada por correção de erro. Eles se diferenciam um do outro pelo número de camadas. Ambos são adequados a problemas de classificação, mas as RNAs *Perceptron* resolvem somente problemas linearmente separáveis.

Segundo Haykin (2001), o *Perceptron* de camada única é a representação mais simples de uma RNA utilizada para a classificação de padrões. Outro ponto a respeito desta RNA de camada única é destacado por Braga, Carvalho e Ludemir (2011), onde os autores comentam que, independentemente do valor  $\eta$ , deverá existir uma convergência em um tempo finito, no caso das classes serem linearmente separáveis.

A seguir, é apresentado o algoritmo do *Perceptron* de camada única, com  $k$  neurônios (HAYKIN, 2001; BRAGA, CARVALHO e LUDEMIR, 2011).

#### 1. Inicialização

Gerar valores aleatórios entre  $[-1,1]$  para os vetores sinápticos  $w_k = 0$  e o bias  $\theta_k = 0$ .

## 2. Ativação

Realizar o somatório apresentado em (6), onde  $m$  representa a quantidade de pesos sinápticos para o neurônio  $k$  e  $p$  é o padrão que passa pelo neurônio na iteração  $n$ .

$$u_k(n) = \sum_{j=1}^m w_{kj}(n) \cdot x_j(n) + b_k \quad (6)$$

Agora, é executada a função de transferência, dada pela equação (7), onde  $\varphi(\cdot)$  é uma função de limiar de forma a calcular o valor obtido para o padrão.

$$y_k = \varphi(u_k(n)) \quad (7)$$

Assim, calcula-se o sinal do erro dado por (8), onde  $d_j$  representa o valor desejado para o padrão  $p$ .

$$e_k = d_j(n) - y_k(n) \quad (8)$$

## 3. Atualização dos Pesos

Calcular os novos pesos sinápticos do neurônio  $k$  dados pela equação (9).

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n) \quad (9)$$

onde  $\Delta w_{kj}(n) = \eta \cdot e_k(n) \cdot x_j(n)$ , com  $0 < \eta < 1$ .

## 4. Análise da Situação Atual

Após todos os exemplos (instâncias) de treinamento terem passado pela rede uma única vez teremos encerrada a 1ª iteração; então deve-se calcular o erro quadrático global dado por (10), para todos os  $N$  exemplos de treinamento.

$$\frac{1}{2} \sum_{k=1}^N e_k^2(n) \quad (9)$$

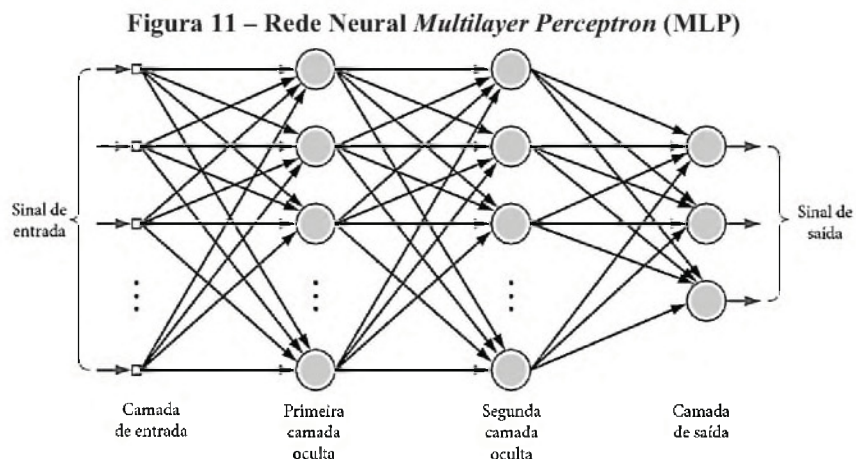
## 5. Continuação

Incremente a iteração  $n$  em uma unidade e volte ao passo 2.

O procedimento então deverá continuar acontecendo até que um erro  $e$  suficientemente pequeno seja encontrado.

Um MLP, nada mais é que uma generalização do *Perceptron* de camada única (HAYKIN, 2009). As redes neurais MLP também utilizam a regra de aprendizagem por correção de erro do paradigma supervisionado. Sua arquitetura tem pelo menos duas camadas que são acíclicas e totalmente conectadas (HAYKIN, 2001).

É possível descrever o MLP como uma rede constituída por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída (BRAGA, CARVALHO e LUDEMIR, 2011). O treinamento do MLP ocorre em três etapas: a alimentação para frente (*feedforward*) da rede com os padrões de entrada, o cálculo e retropropagação (*backpropagation; BP*) do erro, e o ajuste dos pesos (FAUSETT, 1994). A seguir, na Figura 11, temos a representação de um *Perceptron* de múltiplas camadas.



FONTE: HAYKIN, 2009.

O algoritmo do *backpropagation* tem sua atuação nos pesos sinápticos, de forma a minimizar a função erro, por meio da técnica do gradiente descendente (COSTA Jr. e CRUZ, 1998). O erro mínimo local achado pode ser satisfatório, mas se não é, novos pesos iniciais deverão ser utilizados ou, então, uma rede com mais neurônios poderá fazer um trabalho melhor.

O algoritmo de treinamento *back-propagation*, composto pelas etapas propagação *forward* e propagação *backward*, para a estrutura com duas camadas (entrada; oculta e de saída), tem seus passos descritos a seguir.

### 1. Propagação Forward

#### 1. Inicialização

Definir valores aleatórios tanto para os pesos sinápticos da camada oculta ( $w_{kj}$  e  $b_k$ ) quanto para os da camada de saída ( $w_{lj}$  e  $b_l$ ) entre  $[-1,1]$ .

#### 2. Ativação – Camada Oculta

Realizar o somatório dado em (10) para cada neurônio  $k$  da camada oculta, onde  $x^p_j$  são os dados de entrada da rede.

$$u^p_k(n) = \sum_{j=1}^m w_{kj}(n) \cdot x^p_j(n) + b_k(n) \quad (10)$$

Em seguida, calcule o valor da função de transferência (11) para cada um dos neurônios da camada oculta, onde  $\varphi(\cdot)$  geralmente é uma função sigmoide.

$$y^p_k(n) = \varphi_k(u^p_k(n)) \quad (11)$$

### 3. Ativação – Camada de Saída

Execute o somatório dado em (12) para cada neurônio  $l$  da camada de saída, caso exista mais de um.

$$u^p_l(n) = \sum_{j=1}^m w_{lj}(n) \cdot y^p_k(n) + b_l(n) \quad (10)$$

Então, calcule o valor da função de transferência (12) para cada neurônio presente na camada de saída, onde, geralmente,  $\varphi(\cdot)$  É uma função linear.

$$y^p_l(n) = \varphi_l(u^p_l(n)) \quad (11)$$

## 2. Propagação Backward:

### 1. Ajuste dos pesos – Camada de Saída para Camada Oculta

Inicialmente, calcula-se o erro da camada de saída utilizando a equação em (12), onde  $d_j$  é o valor esperado de saída e  $\varphi'(\cdot)$  é a derivada da função de transferência do neurônio de saída.

$$\delta^p_l(n) = (d^p_j - y^p_l) \cdot \varphi'_l(u^p_l(n)) \quad (12)$$

Em seguida (13), calculamos  $\Delta^p w_{lj} = \alpha (w_{lj}(n-1)) + \eta \delta_l(n) y_l(n)$ , que representa o ajuste qual o peso deverá sofrer e realizamos a atualização dos pesos sinápticos na camada de saída de acordo com a regra delta.

$$w_{lj}(n+1) = w_{lj}(n) + \Delta^p w_{lj} \quad (13)$$

onde,  $\eta$  é a taxa de aprendizagem e  $\alpha$  a taxa de *momentum*.

### 2. Ajuste dos pesos – Camada Oculta para Camada de Entrada

Primeiramente calculamos o erro da camada oculta utilizando a equação dada em (14).

$$\delta^p_k(n) = \varphi'_k(y^p_k(n)) \sum_l \delta^p_l(n) w_{lj}(n) \quad (14)$$

E, de forma análoga ao passo 1, é calculado  $\Delta^p w_{kj} = \alpha (w_{kj}(n-1)) + \eta \delta_k(n) y_k(n)$  para ser possível realizar a atualização dos pesos utilizando a equação (15).

$$w_{kj}(n+1) = w_{kj}(n) + \Delta^p w_{kj} \quad (15)$$

### 3. Cálculo do erro global

Após todos os exemplos de treinamento terem passado pela rede uma única vez teremos findado a 1ª iteração, então calcule o erro global dado por (16).

$$\varepsilon(n) = \frac{1}{N} \sum_{l=1}^N e_l^2(n) \quad (16)$$

onde  $N$  representa o número de elementos de treinamento.

Apesar do sucesso do BP, existem alguns aspectos que fazem com que esse algoritmo não seja útil universalmente. Para Kröse e Smagt (1996), esse método possui duas principais deficiências. A primeira, a “paralisia da rede”, que diz respeito a, no momento do treinamento, os pesos podem acabar sendo ajustados para valores altos fazendo com que a função sigmoide tenha valores próximos a zero, dessa forma o treinamento poderia cair em uma paralisação virtual. A outra deficiência diz respeito aos mínimos locais, pois como a superfície de erro de uma rede é feita de picos e vales, por causa do gradiente descendente, a rede pode acabar presa em um mínimo local quando outro valor muito menor pode estar próximo.

Outro tipo de RNA muito utilizado são as redes de Funções de Bases Radiais, RBF (*Radial Basis Function*). São redes multicamadas onde a ativação de um neurônio pode ser em função da distância entre seus vetores de entrada e de peso. O nome, RBF, se deve ao fato da utilização, pelos neurônios da camada oculta, de funções de bases radiais, onde as mais comuns são apresentadas a seguir pelo Quadro 3.

**Quadro 3 - Funções de Bases Radiais**

Tipo de Função	Função
Função gaussiana	$f(x) = e^{-\frac{(r)^2}{c^2}}$
Função multiquadrática	$f(x) = (r^2 + c^2)^k, k > 0, k \notin \mathbb{N}$
Função thin-plate-spline	$f(x) = r^2 \ln(r)$

FONTE: Adaptado de CHEN, FU e CHEN, 2014.

Onde  $r = x - \mu$ , sendo  $x$  o valor de entrada,  $\mu$  o centro da função, e ainda  $c$  uma constante positiva conhecida como parâmetro de forma (CHEN, FU e CHEN, 2014).

As redes RBF têm apresentado excelentes desempenhos em diversos problemas de interesses práticos (KHAN *et al.*, 2017). Em Tatar *et al.* (2015) reservatórios de salmora tiveram suas propriedades físico-químicas analisadas utilizando um modelo híbrido de RBF com algoritmo genético. Em Li e Verme (2016), outra aplicação é apresentada, onde o núcleo da RBF é utilizado para prever o gradiente de pressão no contexto de física nuclear. Temos ainda o trabalho desenvolvido por Liu, Bi e Fan (2017) que, utilizando técnicas de *machine learning* como as RBF, procurou detectar sentimentos expressos em textos, como por exemplo em redes sociais. Para Lu *et al.* (2016), as RBF foram utilizadas para representação dos formatos de órgãos para reconstrução dos mesmos em 3D.

As redes RBF possuem determinadas vantagens quando comparadas com outras redes. Em Leśniak (2016), é comentado que as RBF podem modelar qualquer função não-linear por meio de uma única camada, não sendo necessário decidir o número de camadas da rede. A segunda vantagem é que uma simples função linear executada na camada de saída pode ser otimizada com técnicas de modelagem linear que são rápidas e não têm problemas como mínimos locais.

Matematicamente, o processo de treinamento de uma RBF é definido por (LÓPEZ *et al.*, 2016):

- Em um conjunto de  $m$  elementos pertencentes a um conjunto de treinamento  $\vec{E}_{m \times n} = \{e_1, e_2, \dots, e_m\}$ , e onde  $\vec{W} = \{w_1, w_2, \dots, w_n\}$  são os pesos dos neurônios que compõe a rede, a expressão qual a RBF é baseada está descrita em (17):

$$rbf(\vec{x}) = \sum_{i=1}^m w_i \cdot \varphi(\|\vec{e}_i - \vec{x}\|) \quad (17)$$

- Onde  $\varphi_i(\cdot)$  é uma função de base radial. Segundo Haykin (1994), se colocar  $\varphi_i(\|\vec{e}_i - \vec{x}\|)$  na forma matricial tem-se (18):

$$\begin{pmatrix} \varphi_1(\vec{e}_1) & \cdots & \varphi_1(\vec{e}_m) \\ \vdots & \ddots & \vdots \\ \varphi_m(\vec{e}_1) & \cdots & \varphi_m(\vec{e}_m) \end{pmatrix} \times \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \quad (18)$$

- Dessa forma, para obter-se o  $\vec{W}$  desconhecido (que seria a função do estágio de treinamento), tem-se que:

$$\begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \times \begin{pmatrix} \varphi_1(\vec{e}_1) & \cdots & \varphi_1(\vec{e}_m) \\ \vdots & \ddots & \vdots \\ \varphi_m(\vec{e}_1) & \cdots & \varphi_m(\vec{e}_m) \end{pmatrix}^{-1} \quad (19)$$

### 3.4 REGRESSÃO LOGÍSTICA

Métodos de regressão são componentes de qualquer análise de dados que se refira à descrição do relacionamento entre uma variável de resposta e uma ou mais variáveis exploratórias. O modelo de regressão logística é o modelo utilizado mais frequentemente para análise desse tipo de dados (HOSMER, LEMESHOW e STURDIVANT, 2013).

O que distingue a Regressão Logística (RL) de modelos de Regressão Lineares é que a saída do primeiro pode ser uma saída binária, ou seja, dicotômica. Esta variável binária que está sendo modelada é a variável resposta, ou variável dependente. De acordo com Hilbe (2016), para que um modelo se adapte bem aos dados deve-se assumir que: os preditores não são correlacionados entre eles; são significativamente relacionados com a resposta; e as observações ou elementos dos dados do modelo também não possuem correlação.

De acordo com Harrell Jr. (2015), o modelo estatístico da Regressão Logística é em termos da probabilidade de  $Y = 1$  dado um valor  $X$ , o valor dos preditores. Esta probabilidade é dada pela equação (20) abaixo:

$$Prob\{Y = 1|X\} = [1 + e^{-X\beta}]^{-1} \quad (20)$$

onde a função  $P = [1 + e^{-X\beta}]^{-1}$  é denominada de função logística. De acordo com Harrell Jr (2015), uma vez que a distribuição de uma variável binária aleatória  $Y$  é completamente definida pela verdadeira probabilidade de  $Y = 1$ , e sendo que o cálculo não faz premissas sobre a distribuição dos preditores, o modelo logístico não faz suposições ligadas à distribuição.

### 3.5 TRABALHOS CORRELATOS

Foram pesquisados os principais problemas ligados à aplicação financeira: Falência; Risco de Crédito; Pontuação de Crédito e Mercado de Ações. Quanto aos

métodos foram filtrados por: Redes Neurais Artificiais, Estatística, *Support Vector Machines*, Heurísticos e ainda Programação Linear. Por fim, as aplicações levantadas foram tanto para previsão quanto para classificação.

Em seguida, uma série de exclusões foi realizada com o objetivo de filtrar a quantidade de artigos a serem analisados. As exclusões são apresentadas a seguir:

1. Ano: foram selecionados apenas trabalhos desenvolvidos nos últimos cinco anos (início, portanto, em 2012).
2. Área de aplicação: As áreas verificadas foram: Ciência da Computação; Engenharia; Matemática; Ciência da Decisão; Negócios, Administração e Contabilidade; Economia, Econometria e Finanças; Ciências Sociais; e Multidisciplinar.
3. Idioma: Foram selecionados os artigos publicados em Inglês.
4. Tipo da Fonte: Selecionou-se os trabalhos de conferências e periódicos.
5. Possibilidade de encontrar: excluiu-se os artigos que não possuíam ou autores ou o DOI do documento, impossibilitando que fosse possível o encontrar.
6. Título: A partir da leitura dos títulos dos artigos foram selecionados os que realmente teriam relevância com os temas pesquisados.

Feitas esses filtros restaram 333 documentos para a análise dos eixos: Falência, Mercado de Ações e Risco de Crédito. Quando separados nos eixos (anexo no formato Excel) foram verificados a classificação Qualis e o H index dos periódicos de forma a excluir os documentos publicados em periódicos de classificação Qualis inferior a B3. Caso o periódico não tivesse Qualis excluiu-se os quais o periódico tivesse H index inferior a 60 ou *Scimago Journal & Country Rank* (SJR) inferior a 1. Terminada esta exclusão, 108 artigos foram analisados. Abaixo, temos uma Tabela 1 resumo de como estão divididos os trabalhos selecionados.

**Tabela 1 - Eixos de Divisão dos Trabalhos Correlatos**

		PROBLEMA			APLICAÇÃO (ÕES)		METODOS				
		Falência	Mercado de Ações	Risco de Crédito	Previsão	Classificação	Redes Neurais Artificiais	Técnicas Estatísticas	<i>Support Vector Machines</i>	Heurísticas	Programação Linear
PROBLEMA	Falência	24			24	3	14	12	10	9	1
	Mercado de Ações		40		40	0	32	11	14	22	0
	Risco de Crédito			44	0	44	29	27	28	25	2
APLICAÇÃO (ÕES)	Previsão	24	40	0	64		46	23	24	31	1
	Classificação	3	0	44		47	31	29	30	27	2

FONTE: A autora, 2017.

A seguir, para cada um dos eixos de aplicações em sistemas financeiros serão apresentados os trabalhos levantados. Iniciando por falência, em seguida a apresentação das aplicações com mercados de ações e por fim classificação de risco de crédito.

### 3.5.1 Falência

A falência de empresas, de acordo com Ding *et al.* (2012), é considerada uma das maiores ameaças a determinado negócio. Isso não apenas aumenta a perda financeira dos credores de determinada companhia, mas também tem um impacto negativo na sociedade e na economia em si.

Dentro do levantamento de trabalhos relacionados a esse tema, foi percebido que boa parte dos mesmos foca em aplicações voltadas à previsão de falência. Existem diversas ferramentas que podem ser aplicadas com esse foco, como, por exemplo, as de IA como as RNA, técnicas estatísticas, SVM, heurísticas, e até Programação Linear.

Considerando as aplicações que utilizam as RNA, Jeong, Min e Kim (2012) sugeriram uma nova abordagem para ajuste dos parâmetros das RNA com o objetivo de aprimorar a acurácia da rede. O método sugerido apresentou melhores resultados quanto aos demais testados pelos autores, como as árvores de decisão, o método aditivo generalizado, o modelo linear generalizado e uma análise discriminante multivariada.

Tsai e Cheng (2012) apresentaram um comparativo entre diferentes técnicas, sendo elas: RNA, Árvores de Decisão, RL e SVM, com o objetivo de prever falência de empresas em quatro diferentes bases de dados (nomeadas de australiana, alemã, japonesa e de uma determinada competição). Os resultados mostraram que para a base de dados alemã foi mais difícil obter altas acurácias de previsão. Com base nos dados dos autores, a “melhor” técnica utilizada foi a de SVM por possuir maior taxa de acurácia de previsão nas demais bases de dados testadas.

O trabalho desenvolvido por Sánchez-Lasheras *et al.* (2012) propôs um método híbrido entre o *Self-Organizing Maps* (SOM) e Regressão Adaptativa Multivariada (MARS). Os autores puderam demonstrar que este híbrido teve uma acurácia maior que outras técnicas consideradas *benchmark*, utilizando apenas o algoritmo *backpropagation* e a segunda utilizando apenas o MARS, em empresas espanholas do setor de construção.

De acordo com Du Jardin e Séverin (2012), a maioria dos modelos perde sua habilidade em generalizar informações quando os dados utilizados para a previsão e

estimação são coletados em diferentes períodos de tempo. De forma a superar esta falha dos métodos, os autores propuseram um modo de utilizar o mapa de Kohonen como um modelo de previsão, comparando-o com Análise Discriminante, RL, Método de Cox e ainda as RNA. Os pesquisadores então concluíram que a generalização do erro foi mais estável utilizando o mapa de Kohonen do que os demais métodos, mas em termos de acurácia, todos os métodos tiveram resultados semelhantes.

Outro trabalho que utilizou RNA foi o desenvolvido por Korol (2013), buscando a previsão de falência em empresas da América Latina e Europa Central especificamente. Através de uma comparação entre uma metodologia heurística (Árvores de Decisão), modelo estatístico (Análise Discriminante) e as RNA, o autor conseguiu acurácia acima de 80% para os três métodos com antecedência de um ano prévio à falência das empresas. Já para dois anos precedentes, a acurácia cai para aproximadamente 75% na aplicação com Análise Discriminante, mas os demais métodos mantêm acurácia semelhante ao aviso de falência com um ano de antecedência.

Lee e Choi (2013), ainda envolvendo uma aplicação com RNA, propõem um modelo de previsão de falência para empresas coreanas. Utilizando dados de empresas de construção, de varejo e de manufatura, os autores compararam o desempenho do algoritmo *backpropagation* (BP) com a análise discriminante multivariada, onde as RNA BP apresentaram melhores resultados.

No artigo desenvolvido por Federova, Gilenko e Dovzhenko (2013), para previsão de falência em empresas russas, foram testadas diferentes combinações de algoritmos (Análise Discriminante Multivariada, RL, Árvores de Regressão e RNA). Outro objetivo do trabalho era também de analisar se os indicadores financeiros estipulados pela legislação russa eram efetivos para esse tipo de previsão. As melhores combinações realizadas utilizavam as RNA e que apenas um dos indicadores era realmente influente na previsão realizada.

Utilizando uma base de dados de 1980 a 2011, Tinoco e Wilson (2013) buscaram investigar empiricamente se seria útil combinar contabilidade e uma base de dados macroeconômica baseada no mercado. O modelo construído foi comparado com outras aplicações utilizando RNA, onde se equiparou em termos de acurácia.

O modelo de RNA desenvolvido por López e Sanz (2015) leva em consideração ainda alguns fatores específicos da crise financeira que havia ocorrido um tempo antes. Foram combinados MLP e SOM de forma a desenvolver uma ferramenta que mostrasse a probabilidade de falência até três anos antes. O modelo gerado conseguiu detectar com

acurácia acima de 95% as falências no período estabelecido, superando assim modelos tradicionais deste problema.

No trabalho desenvolvido por Slavici, Maris e Pirtea (2016), foi proposto um algoritmo efetivo para o problema em questão baseado no algoritmo *Kernel Extreme Learning Machine* (KELM) e comparado com outras técnicas de aprendizado de máquina (SVM e RNA) e ainda modelos heurísticos. No trabalho dos autores em questão, o algoritmo desenvolvido possuiu melhores resultados aos demais comparados se mostrando uma técnica competitiva e possível de ser utilizada.

Mais recentemente, utilizando as RNAs, Kim, Jo e Shin (2016) desenvolveram uma otimização baseada em *clusters* em combinação com as RNA. Foi visto que o modelo teve sucesso de aplicação para o caso de pequenas e médias empresas na indústria de manufatura.

Em Wagle, Yang e Benslimane (2017), discutiu-se a aplicação e os benefícios de técnicas de *Data Mining* em modelos de previsão. Foram comparadas técnicas como RNA, SVM, regressão logística e redes Bayesianas, onde as RNA apresentaram melhor acurácia dentro dos parâmetros testados para a base de dados levantada pelos autores.

Um artigo mais recente de aplicação de RNA para o problema de falência foi desenvolvido por Barboza, Kimura e Altman (2017). Neste trabalho, os autores buscaram testar diferentes técnicas de aprendizado de máquinas (RNA, SVM, e utilizando também técnicas de *ensemble*, ou seja, comitê de máquinas) e algumas metodologias estatísticas, como a RL e a Análise Discriminante, com o objetivo de prever a falência de empresas norte-americanas. Ao contrário que outros autores verificaram, no trabalho desenvolvido observou-se que a melhor metodologia utilizava a técnica de Florestas Randômicas, apresentando acurácia de 87% enquanto as metodologias estatísticas testadas apresentaram apenas entre 50 – 69%.

Não necessariamente todas as propostas de solução para o problema de falência de empresas necessitam utilizar RNA no método proposto. Como é o caso do trabalho desenvolvido por Ribeiro *et al.* (2012), onde os autores propuseram uma solução utilizando apenas SVM e para todos os testes desenvolvidos atingiram uma acurácia superior a 85%.

Outro trabalho que, de forma análoga, utilizou apenas um tipo de técnica para a previsão de falência foi o de Ding *et al.* (2012). Neste trabalho, os autores procuraram fazer esta previsão utilizando a Regressão Logística como metodologia. Utilizando um

levantamento de empresas falidas do ano de 1981 a 2006, os autores conseguiram 80% de acurácia com o método utilizado.

Xiong et al. (2013) semelhantemente a Ribeiro *et al.* (2012) também utilizaram SVM, juntamente com uma metodologia de pontuação de crédito, para realizar a previsão de falência utilizando uma base de dados de determinado banco canadense. Os resultados chegaram a uma acurácia superior a 90% na previsão desejada.

Em Virag e Nyitrai (2013) pode-se encontrar uma combinação de metodologias estatísticas com o SVM. Utilizando uma base de dados da Hungria, os autores combinaram uma análise de variância (ANOVA) com o SVM para realizar a previsão desejada. Dentre os resultados obtidos, os autores conseguiram melhorar a acurácia obtida anteriormente utilizando RNA, chegando a uma taxa de 83% aproximadamente.

Em artigo publicado numa conferência, Lu *et al.* (2014) utilizando uma combinação de SVM com métodos heurísticos usaram dados de uma competição para prever falência. Os resultados dos testes apresentaram acurácia variando de 89% aproximadamente até superior a 95%.

O trabalho desenvolvido por Zhao et al. (2016) foram combinados os algoritmos KELM e SVM para a previsão de falência utilizando uma base de dados de empresas polonesas dos anos de 1997 a 2001. A acurácia média dos autores em questão chegou a 82%.

Mais recentemente, Wang et al. (2017) desenvolveram uma metodologia utilizando apenas métodos heurísticos para a resolução do problema. O modelo desenvolvido utilizou uma base de dados japonesa pública e, através do MATLAB, os autores conseguiram atingir uma taxa de acurácia média de 86% para os piores resultados, e superiores a 98% quando nos melhores resultados entre as metodologias testadas.

Kim, Mun e Bae (2017), com o uso da linguagem R de programação, implementaram o SVM, comparando-o com outras técnicas (RL, Análise Discriminante Multivariada e RNA) para uma base de dados coreana de falência de empresas de manufatura se estendendo do ano de 2000 a 2013. O melhor resultado obtido foi o SVM implementado chegando a ter acurácia acima de 90% enquanto as RNA ficaram superiores a 80% apenas.

Jones (2017) trouxe técnicas de *ensemble* combinadas com métodos estatísticos buscando atingir uma acurácia satisfatória para uma base de dados estadunidense com 1115 arquivos de falência com 91 variáveis de previsão. Prevendo com três nos

precedentes à falência, o autor conseguiu atingir uma acurácia de 91%, tendo resultados melhores para um ano de antecedência (93%) e no ano da falência (95%).

Por fim, o último trabalho encontrado a respeito deste problema, Li, Crook e Andreeva (2017), foi o único artigo que apresentou Programação Linear numa tentativa de prever falência. Os autores chegaram a atingir uma acurácia de 88% no tempo computacional previsto para o programa.

A seguir é apresentado no Quadro 4 um resumo dos artigos voltados à falência com as respectivas metodologias utilizadas.

**Quadro 4 – Alguns Trabalhos Correlatos com Abordagem em Falência**

AUTORES	ANO	PUBLICAÇÃO	CLASSIFICAÇÃO PERIÓDICO			# CITAÇÕES	ABORDAGEM	MÉTODO						
			Qualis	H Index				Redes Neurais Artificiais	Estatística	Support Vector Machines	Heurísticos	Programação Linear	SOFTWARE	
BARBOZA F.; KIMURA H.; ALTMAN E.;	2017	Periódico	A1	131		2	Previsão	X	X					R
LI Z.; CROOK J.; ANDREEVA G.;	2017	Periódico	A1	131		2	Previsão					X		
JONES S.;	2017	Periódico	-----	55		1	Previsão		X					TreeNet
KIM S.; MUN B.M.; BAE S.J.;	2017	Periódico	B1	46		0	Previsão			X				R
WANG M.; CHEN H.; LI H.; CAI Z.; ZHAO X.;	2017	Periódico	A2	76		0	Previsão				X			MATLAB
TONG C.; LI J.; XU X.;	2017	Periódico	A2	76		0	Previsão					X		SPSS/WEKA
WAGLE M.; YANG Z.; BENSUIMANE Y.;	2017	Conferência	-----	-----		0	Previsão	X	X					MATLAB/WEKA
ZHAO D.; HUANG C.; WEI Y.; YU F.; WANG M.;	2016	Periódico	-----	32		6	Previsão			X				
CHEN H.;	2016	Periódico	A1	131		6	Previsão	X	X					
KIM H.-J.; JO N.-O.; SHIN K.-S.;	2016	Periódico	A1	131		6	Previsão	X	X					
SLAVICIT.; MARIS S.; PIRTEA M.;	2016	Periódico	B1	39		1	Previsão	X						MATLAB
LÓPEZ ITURRIAGA F.J.; SANZ I.P.;	2015	Periódico	A1	131		49	Previsão	X						SPSS
LU Y.; ZHU J.; ZHANG N.; SHAO Q.;	2014	Conferência	-----	-----		1	Previsão			X				MATLAB
TSAI C.-F.; HSU Y.-F.; YEN D.-C.;	2014	Periódico	A2	89		37	Ambos	X	X					WEKA
TINOCO M. H.; WILSON N.;	2013	Periódico	B1	35		106	Previsão	X	X					SAS
FEDOROVA E.; GILENKO E.; DOVZHENKO S.;	2013	Periódico	A1	131		32	Ambos	X	X					
VIRAG M.; NYITRAI T.;	2013	Periódico	-----	73		1	Ambos		X	X				WinsVM
LEE S.; CHOI W.S.;	2013	Periódico	A1	131		38	Previsão	X	X					Neuroshell
KOROL T.;	2013	Periódico	B1	45		44	Previsão	X	X					
XIONG T.; WANG S.; MAYERS A.; MONGA E.;	2013	Periódico	A1	131		22	Ambos			X				
DING A.-A.; TIAN S.; YU Y.; GUO H.;	2012	Periódico	-----	147		11	Previsão		X					COMPUSTAT
RIBEIRO B.; SILVA C.; CHEN N.; VIEIRA A.;	2012	Periódico	A1	131		29	Previsão			X				
CARVALHO DAS NEVES J.;	2012	Periódico	A1	200		22	Previsão	X	X					
DU JARDIN P.; SEVERIN E.;	2012	Periódico	A1	131		25	Previsão	X	X					
SÁNCHEZ-LASHERAS F.; DE ANDRÉS J.; LORCA P.;	2012	Periódico	A1	74		28	Previsão	X	X					
DE COS JUEZ F.J.;	2012	Periódico	A1	131		25	Previsão	X	X					
TSAI C.-F.; CHENG K.-C.;	2012	Periódico	A1	131		25	Previsão	X	X					
JEONG C.; MIN J. H.; KIM M.S.;	2012	Periódico	A1	131		25	Previsão	X	X					
TOTAL	24	22				20		14	12	10	9	1		

FONTE: A autora, 2017.

### 3.5.2 Mercado de Ações

Segundo Devi, Bhaskaran e Kumar (2015), a previsão do mercado de ações é uma tarefa muito difícil e altamente complicada vez que é afetada por diferentes fatores tais como, por exemplo, condições econômicas, interesses dos investidores e eventos políticos.

De acordo com Chong *et al.* (2017), com o aumento na disponibilidade de dados das ações comerciais e o desempenho não tão satisfatório de modelos existentes na literatura, estudos abrangentes que examinem de forma objetiva a adequação de métodos para a previsão e a análise de mercado de ações são oportunos.

Durante a análise dos artigos levantados relacionados a esta aplicação, percebeu-se que eram utilizadas RNA, Estatística, SVM e Modelos Heurísticos para a resolução deste problema. Diferentemente dos trabalhos de falência, esta seção não possui aplicação de Programação Linear nos trabalhos pesquisados. Outra característica desse tipo de aplicação é que os trabalhos com este problema possuem abordagem de previsão apenas, não tendo sido observados trabalhos de classificação.

Dos trabalhos encontrados, observando apenas os que tiveram RNA nas aplicações, o desenvolvido por Dai, Wu e Lu (2012) utilizou a Análise de Componente Não Linear Independente e as RNA combinadas com o objetivo de obter uma acurácia satisfatória na previsão desejada. Após analisar o desempenho individual das técnicas escolhidas e a combinação entre elas, foi percebido que a associação das mesmas obteve melhor desempenho chegando a atingir, no mínimo, 80% de acurácia nos testes desenvolvidos.

Em Asadi *et al.* (2012) é utilizada apenas a RNA para a previsão de preços de ações em seis diferentes conjuntos de dados. O teste realizado utilizando o algoritmo proposto por Levenberg-Marquardt para o treinamento das RNA foi o de melhor resultado em relação à acurácia da previsão, apresentando melhoria em relação aos algoritmos anteriormente utilizados nas mesmas bases de dados.

No trabalho desenvolvido por Ticknor (2013) foi apresentado, além das RNA as redes com regularização Bayesiana, que designam uma natureza probabilística aos pesos da rede, permitindo, assim, que o próprio método penalize o modelo de forma automática e ótima. Os resultados obtidos pelo autor mostraram que o desempenho da metodologia proposta é tão bom quanto os métodos mais avançados, mas sem a necessidade do pré-processamento de dados, teste de sazonalidade, e, ainda, análise de ciclo.

Kazem *et al.* (2013) em sua metodologia, compararam diferentes híbridos de SVM com uma aplicação de RNA para uma base de dados das ações da Microsoft, da Intel e das ações de determinado banco. Em termos de erro mínimo quadrado, dois, dos quatro métodos propostos tiveram valores menores que as RNA nos testes realizados.

Com base em dados de ações de empresas do Taiwan, Wei (2013) desenvolveu um híbrido entre RNA e Algoritmo Genético (AG) para previsão dos preços de ações e o comparou com trabalhos similares de outros autores. O autor pôde mostrar que o método desenvolvido conseguiu superar as demais metodologias em termos de raiz quadrada do erro médio (RMSE).

Em relação ao Brasil, Oliveira *et al.* (2013) a partir de uma base de ações da Petrobrás utilizaram as RNA para a previsão dos preços das ações. Os autores conseguiram atingir uma acurácia de aproximadamente 93% no conjunto de teste.

Xi *et al.* (2014) utilizaram as RNA do tipo de funções de bases radiais (RBF) em dados de uma empresa de produtos a base de ferro e aço, onde os resultados apresentados pelo método foi dado como satisfatório pelos autores do trabalho.

Na pesquisa desenvolvida por Bisoi e Dash (2014), utilizando um híbrido de algoritmo evolucionário e RNA, aplicado a uma base de dados de ações de Bombay. Utilizando quatro estratégias de treinamento para o algoritmo desenvolvido, os autores conseguiram de forma que o modelo proposto obteve o menor erro dentre as comparações realizadas.

O modelo Dash, Dash e Bisoi (2015), onde foi proposto um GARCH (*Generalized Auto-Regressive Conditional Heteroskedasticity*) exponencial (EGARCH), os autores utilizaram RNA e ainda um sistema de lógica *fuzzy* para duas bases de dados: BSE Sensex e CNX Nifty do mercado de ações da Índia. Os resultados obtidos indicaram que o modelo proposto oferece melhoria na previsão de volatilidade em comparação com outros métodos aplicados às mesmas bases de dados.

Em Patel *et al.* (2015), os autores buscaram a previsão dos preços em duas bases de dados a CNX Nifty e a S&P Bombay Stock Exchange. Utilizando diferentes híbridos, os autores propuseram uma comparação entre algoritmos de forma a observar o que possuísse melhor desempenho. Os híbridos propostos foram: SVR + RNA; SVR + Florestas Randômicas, estes híbridos foram comparados com os algoritmos mencionados separadamente e ainda com outros propostos em demais trabalhos. Os resultados variaram em relação ao conjunto de dados utilizado, mas para os autores foi de forma satisfatória uma vez que o pior desempenho atingiu 73,3% de acurácia (Redes Bayesianas).

Em outro trabalho no mesmo ano, Patel *et al.* (2015) apresentaram outra comparação entre SVR, RNA e FR e os comportamentos dos híbridos entre esses modelos, utilizando a mesma base de dados do trabalho apresentado anteriormente. A partir dos resultados observados os autores concluíram que o método de melhor desempenho foi o híbrido de SVR-RNA.

Shan *et al.* (2015) propuseram um híbrido baseado na transformação de *wavelet* utilizando métodos lineares como o ARIMA (*AutoRegressive Integrated Moving Average*) e não lineares com as RNA (tanto utilizando BP quanto as RBF). O modelo proposto possuiu os menores erros comparados com os demais testados, chegando a um SMAPE (*Symmetric Mean Absolut Percentage Error*) 0,0064 para a base de dados das ações de Shanghai.

Utilizando informações da Bovespa, Laboissiere, Fernandes e Lage (2015) propuseram uma metodologia que previsse o máximo e o mínimo dos preços de ações de três empresas de distribuição de energia utilizando RNA. De acordo com os resultados obtidos, os autores concluíram que o modelo proposto tinha desempenho satisfatório, uma vez que o maior MAPE (*Mean Absolut Percentage Error*) apresentado foi inferior a 0,9% para a previsão dos preços no próprio dia.

A metodologia proposta por Devi, Bhaskaran e Kumar (2015) consistia numa comparação entre modelos muito divulgados nesta aplicação, sendo eles RNA e SVM, outro método apresentado pelo autor foi a busca Cuckoo, algoritmo heurístico baseado em inteligência de Enxame de Partículas. O híbrido proposto entre SVM e a busca Cuckoo foi o que apresentou melhores resultados em termos de taxa de erro.

Em Sands *et al.* (2015) novamente pode-se encontrar um comparativo entre RNA, SVM e modelos heurísticos. Utilizando, portanto, SVM, RNA e PSO (*Particle Swarm Optimization*), os autores observaram que o SVM possuiu melhor desempenho chegando a 95% de acurácia.

Sem realizar comparativo entre outras técnicas, observando apenas o comportamento dos resultados a partir de alterações nos parâmetros das RNA, Chen, Zhou e Dai (2015) mostraram que as acurácias dos testes se alteravam conforme mudanças ocorriam na arquitetura da rede e no que eles chamaram de métodos de otimização, onde foram testados 5 tipos. O melhor tipo de rede encontrado pelos autores foi o que considerava a normalização das informações de entrada, utilizando apenas uma única empresa como base de informação e os preços de baixa, alta e de fechamento das ações.

De acordo com Göçken *et al.* (2016), o preço do mercado de ações é um dos principais indicadores do crescimento econômico de um país. Em seu trabalho, os autores buscaram avaliar a eficiência da utilização de indicadores técnicos (como o movimento médio de preço de fechamento e seu *momentum*) no mercado de ações turco.

Para poder analisar o relacionamento entre esses indicadores e o mercado de ações em si, Göçken *et al.* (2016) propuseram um híbrido de RNA com dois modelos heurísticos o AG e a Busca Harmônica (BH). Na comparação realizada, os autores perceberam que o híbrido RNA + BH teve melhor desempenho na previsão dos preços das ações.

Qiu, Song e Akagi (2016) também propuseram híbridos de RNA com modelos heurísticos, neste caso os métodos para realizar a hibridização foram: técnicas de busca globais, AG e *Simulated Annealing* (SA). Os híbridos foram aplicados o índice Japanese Nikkei 225 para a previsão do retorno dessas ações. Analisando os resultados obtidos, os autores perceberam que o híbrido RNA + AG possuiu melhor desempenho que os demais.

Semelhantemente, Dash e Dash (2016) apresentaram um modelo que chamaram de *Self Evolving Recurrent Neuro-Fuzzy Inference System* (SERNFIS). Esse método é uma RNA Recorrente que utiliza regra *fuzzy* “*if-then*” com dois *loops* de *feedback* de informação; para estimação dos parâmetros da RNA utilizaram ainda o método heurístico de média harmônica diferencial modificada. Os resultados obtidos apresentaram resultados satisfatórios aos autores.

Novamente hibridizando RNA com normas *fuzzy*, Oztekin *et al.* (2016) buscaram prever o movimento de preços das ações no dia da Istanbul BIST 100 Index integrando três modelos: RNA, sistemas de inferência adaptativas *neuro-fuzzy* e SVM. Os melhores resultados foram os que previam movimentos de queda, em comparação com os movimentos de alta de preço.

Pehlivanlı, Aşikgil e Gülay (2016) apresentaram um modelo de previsão dos preços de ações para o dia seguinte à aplicação do algoritmo. Para tanto, os autores fizeram uso, principalmente, de RNA, regressão logística, Florestas Randômicas e SVM. As acurácias variaram de acordo com o tamanho das empresas, os autores puderam perceber que o melhor método para empresas de pequeno porte foi o Florestas Randômicas (75,6%), para de médio RNA MLP (70,1%) e SVM (79,6%) para empresas de grande porte.

Zhang, Li e Pan (2016) utilizaram o AdaBoost como método de *ensemble* juntamente com AG para as RNA e SVM. As bases de dados utilizada pelos autores foi a *National Association of Securities Automated Quotations* (NASDAQ) e Shenzhen Stock

Exchange (SZSE) com o objetivo de prever a tendência dessas bolsas de valores. Os autores, a partir dos resultados obtidos, perceberam então que o híbrido AdaBoost-AG-SVM possuiu melhor acurácia dentro das condições por eles testadas.

O método utilizado por Khatri e Srivastava (2016) foi o de RNA com variação de número de neurônios na camada oculta. Foram utilizadas cinco bases de dados distintas, sendo elas: Apple, Google, Microsoft, Oracle e Facebook. Realizados os testes, os autores puderam concluir que o menor MSE testado é para a base da Apple com erro de 0,14 para 10 neurônios na camada oculta.

Qiu, Li e Song (2016) verificaram a possibilidade de previsão de preços das ações utilizando um híbrido de AG-RNA utilizando a base da Nikkei 225 Index para diferentes funções de ativação para a RNA. No primeiro teste realizado, a melhor acurácia (77,11%) foi para a RNA que utilizava função tangente-sigmóide, e em seguida logística-sigmóide, no segundo a melhor acurácia foi de 75,44% utilizando nos dois momentos a função logística-sigmóide.

No trabalho desenvolvido por Zhong e Enke (2017), utilizando como base de dados as ações da S&P 500 Index ETF (SPY) com cerca de 60 atributos financeiros e econômicos, foi proposto um modelo de RNA para cada Análise de Componentes utilizada, sendo elas: Análise de Componente Principal (ACP), Análise Robusta *Fuzzy* Principal de Componentes (ARFCP) e ainda a Análise baseada em Núcleo de Componente Principal (ANCP). Os autores, com base nos resultados obtidos, perceberam que a melhor opção seria a utilização das RNA após a ACP.

Xiong e Lu (2017) utilizaram quatro modelos para a previsão sendo eles: ARIMA, RNA *backpropagation*, modelo de Khashei e Bijari, e ainda uma proposta de híbrido ARIMA-RNA. Os autores puderam observar que a metodologia ARIMA-RNA possuiu melhor acurácia para os quatro testes realizados, variando de 59,09% a 78,79%.

Li *et al.* (2017) utilizaram em seu trabalho as RNA para a previsão do preço de fechamento da S&P 500. Foram testadas três modelos de RNA do tipo *deep belief*. A metodologia aplicada que possuiu o menor MSE obtido foi a denominada de LDBN.

Gunduz, Yaslan e Cataltepe (2017) propuseram uma RNA Convolutiva aplicada a dados da Borsa Istanbul 100 e a compararam com a Regressão Logística. Dentre os resultados obtidos os autores puderam observar que o modelo proposto possuiu o melhor desempenho nos testes realizados.

No artigo de congresso de Yu, Wang e Zhang (2017) foi proposta uma RNA BP para a previsão de preços de ações da Vanke de junho de 2008 a junho de 2012. Dois

testes foram realizados, sendo o primeiro baseado em opinião da internet e o segundo não, os autores puderam visualizar que o que não possuía opiniões oriundas da internet possuiu maior acurácia – 93,34% – que o segundo.

Outro trabalho desenvolvido com a utilização de RNA foi o de Li, Bu e Wu (2017). Neste, os autores propuseram um modelo de RNA para a base de dados do índice CSI300 de Shanghai e compararam ainda com o SVM. Realizados os testes, os autores concluíram que o melhor desempenho foi para as RNA com 87,86% de acurácia no teste.

Em Weng, Ahmed e Megahed (2017) foram testados e comparados 7 cenários as RNA, SVM e Árvores de Decisão. O cenário 7 foi o que possuiu melhor desempenho, os autores levantaram informações relativas: ao mercado, indicadores técnicos, *Wikipedia Traffic*, *Google News*, e atributos gerados. Este cenário possuiu 85,8% acurácia de acerto nos testes realizados.

O último trabalho dos levantados, que apresentou RNA em sua aplicação, foi o desenvolvido por Chong, Han e Park (2017). No trabalho desenvolvido, os autores utilizaram RNA do tipo *deep learning* para análise e previsão do mercado de ações. Os resultados obtidos sugeriram que o modelo proposto pode extrair mais informações e ainda melhorar o desempenho da previsão.

Tendo analisado os trabalhos que utilizaram RNA em sua metodologia, são apresentados a seguir os que utilizaram diferentes métodos na pesquisa dos últimos 5 anos.

Fan *et al.* (2014) propuseram um modelo estatístico denominado *Functional Response Additive Model Estimation* (FRAME), que utiliza modelo de regressão linear envolvendo os preditores funcionais  $x_j(t)$ , os escalares  $Z_k$  e as respostas funcionais  $Y(s)$ . O modelo foi utilizado para prever a taxa de decaimento de demanda por filmes de Hollywood usando o mercado de ações virtual. O modelo foi comparado com o ACP e algumas variantes, mostrando ter desempenho melhor que as demais.

Wu, Zheng e Olsen (2014) apresentaram uma combinação de modelo GARCH e SVM. O método visava verificar se existe uma correlação sólida entre a volatilidade da tendência do preço de ações e sentimento de mercado. Os autores conseguiram mostrar que existe essa correlação e ainda apresentaram 81,82% de acurácia ao testá-la.

A proposta de método realizada por Zhang *et al.* (2015) representou um modelo evolucionário de reversão de tendência (*eTrendRev*, como os autores o chamaram). Essa heurística desenvolvida foi aplicada numa série da NASDAQ de janeiro de 2001 a dezembro de 2013, possuindo então mais de 3.000 observações. A partir dos resultados

obtidos os autores puderam constatar que o *eTrendRev* possuiu maior acurácia entre os modelos testados (árvores de decisão e redes Bayesianas).

O trabalho desenvolvido por Nayak, Mishra e Rath (2015) utilizou da hibridização para propor um modelo que fosse composto de SVM e k-vizinhos mais próximos (KNN). A metodologia proposta foi comparada com outros métodos conhecidos da área e os autores perceberam que o algoritmo apresentado possuía melhores resultados em termos de erro que os demais.

Chang, Wu e Lin (2016) apresentaram um modelo baseado em regra *fuzzy* (Takagi-Sugeno – modelo TS) para identificar indicadores das ações e servir como input para o SVM. Essa abordagem foi comparada com modelos de regressão linear e RNA indicando que o TS não apenas rende melhores resultados que os outros tratamentos, mas também permite uma identificação dinâmica da complexidade do sistema de previsão das ações.

O artigo desenvolvido por Chen e Chen (2016) apresentou três propostas de modelos heurísticos a serem comparados entre si e ainda com um trabalho benchmark de outro autor. As bases de dados utilizadas foram as da NASDAQ e da TAIEX em quatro e dois períodos de teste, respectivamente. Os autores, a partir dos resultados obtidos, perceberam que o modelo proposto tinha comportamento diferente para cada uma das bases testadas, mas ainda sim essa metodologia foi superior ao trabalho utilizado como comparativo.

No trabalho desenvolvido por Hong, Linton e Zhang (2017) foi proposta um modelo estatístico multivariado para verificar bolhas de mercado de ações. Os autores perceberam evidências de redução da previsibilidade linear em período recente, mas ainda sim rejeitaram a hipótese nula neste período. Os principais achados deste trabalho, de acordo com os autores, não são substancialmente afetados ao permitir um padrão de fator de risco variável.

O último trabalho relativo ao problema de Mercado de ações a ser apresentado nesta seção foi o desenvolvido por Chen e Hao (2017). Este artigo apresenta uma estrutura de hibridização para SVM com KNN para prever os índices de preços de ações de forma eficiente. A base de dados utilizada constituía de informações da Shanghai Exchange Composite Index (SSE Composite Index) e da Shenzhen Stock Exchange Composite Index (SZSE COMP SUB IND) para análise de mercado de ações chinês. Dentre os resultados obtidos, os autores perceberam que a metodologia proposta era melhor, em termos de erro, para todos os três horizontes de tempo testados.

A seguir é apresentado um quadro resumo (Quadro 5) dos artigos voltados à análise de mercado de ações com as respectivas metodologias utilizadas por cada autor mencionado.

**Quadro 5 – Alguns Trabalhos Correlatos com Abordagem em Mercado de Ações**

AUTORES	ANO	PUBLICAÇÃO	CLASSIFICAÇÃO PERIÓDICO			MÉTODO					SOFTWARE	
			Qualis	H Index	# CITAÇÕES	ABORDAGEM	Redes Neurais Artificiais	Estatística	Support Vector Machines	Heurísticos		Programação Linear
CHONG E.; HAN C.; PARK F. C.;	2017	Periódico	A1	131,00	2	Previsão	X	X				PYTHON
CHEN Y.; HAO Y.;	2017	Periódico	A1	131,00	2	Previsão			X	X		MATLAB
WENG B.; AHMED M.A.; MEGAHED F.M.;	2017	Periódico	A1	131,00	3	Previsão	X		X	X		R
LI J.; BU H.; WU J.;	2017	Conferência	-----	-----	0	Previsão	X	X	X			-----
YU Y.; SHANSHAN W.; LIJUN Z.;	2017	Conferência	-----	-----	0	Previsão	X					-----
GUNDUZ H.; YASLAN Y.; CATALIPE Z.;	2017	Periódico	A1	74,00	0	Previsão	X	X				-----
LI X.; YANG L.; XUE F.; ZHOU H.;	2017	Conferência	-----	-----	0	Previsão	X					-----
XIONG L.; LU Y.;	2017	Conferência	-----	-----	0	Previsão	X	X				SPSS/MATLAB
ZHONG X.; ENKE D.;	2017	Periódico	A1	131,00	13	Previsão	X	X		X		MATLAB
HONG S.Y.; LINTON O.; ZHANG H.J.;	2017	Periódico	-----	31,00	2	Previsão		X				MATLAB
QIU, M.; LI C.; SONG Y.;	2016	Conferência	-----	-----	2	Previsão	X			X		-----
KHATRI S.K.; SRIVASTAVA A.;	2016	Conferência	-----	-----	0	Previsão	X					MATLAB
ZHANG X.-D.; LI A.; PAN R.;	2016	Periódico	A2	89,00	2	Previsão	X		X			MATLAB
PEHIVANLI A.Ç.; AŞIKGİL B.; GÜLAY G.;	2016	Periódico	A2	89,00	2	Previsão	X	X		X		-----
OZTEKIN A.; KIZILASLAN R.; FREUND S.; ISERI A.;	2016	Periódico	A1	200,00	11	Previsão	X		X	X		-----
DASH R.; DASH P.;	2016	Periódico	A1	131,00	5	Previsão	X			X		-----
CHEN T.-L.; CHEN F.-Y.;	2016	Periódico	A1	131,00	10	Previsão				X		-----
QIU M.; SONG Y.; AKAGI F.;	2016	Periódico	-----	117,00	19	Previsão	X			X		MATLAB



### 3.5.3 Risco de Crédito

Dos três problemas levantados a respeito de aplicações financeiras, o último a ser apresentado é o problema de Risco de Crédito. A avaliação de risco de crédito é um problema de mineração de dados muito desafiador e importante no domínio da análise financeira.

A avaliação de crédito bancário é um método utilizado para a previsão da possibilidade de concessão (ou não) de crédito. Desde que foi introduzido na década de 1950, tem sido amplamente utilizado, mais recentemente, para a concessão de empréstimo, especialmente para cartões de crédito e consentimento de crédito (LUO, KONG e NIE, 2016). Esse problema consiste basicamente na avaliação do risco associado com o empréstimo de uma organização financeira para determinada empresa ou indivíduo.

De forma análoga às seções anteriores (falência e mercado de ações), os trabalhos levantados para esta seção serão apresentados inicialmente pelos que fizeram utilização de RNA em suas metodologias em seguida dos que não fizeram. Os artigos selecionados estão apresentados em ordem cronológica de acordo com os modelos utilizados.

O primeiro trabalho levantado que fez uso de RNA no período pesquisado foi o de Loterman *et al.* (2012). O estudo em larga escala realizado pelos autores comparou 24 técnicas para classificação de crédito bancário, incluindo heurísticas, RNA, SVM e modelos estatísticos. Essas técnicas foram avaliadas em seis bases de dados reais e para cada base foi obtido resultados diferentes quanto aos desempenhos das técnicas utilizadas. Duas técnicas tiveram desempenhos melhores em mais de uma base de dados em termos de MAPE, sendo elas: as que utilizaram RNA (desempenho superior em três das bases de dados) e a que utilizou B-OLS (*Beta Ordinary Least Squares*).

Brown e Mues (2012) também fizeram uma extensa comparação de técnicas, desta vez para cinco bases de dados reais. As técnicas utilizadas foram, resumidamente, RL, RNA, Árvores de Decisão, SVM e Florestas Randômicas. Os melhores desempenhos de algoritmo encontrados pelos autores foram para os métodos de Florestas Randômicas e para a técnica de gradiente *boosting*.

Em Wang *et al.* (2012) foram comparados os desempenhos de RNA do tipo RBF, árvores de decisão e RL em duas bases de dados de crédito da UCI. Os autores constataram que para as bases de dados utilizadas os melhores resultados ficaram todos para a técnica de RL com acurácia superior a 88% em todos os casos apresentados.

Hens e Tiwari (2012) desenvolveram um híbrido de SVM com AG e compararam esta metodologia com RNA. A acurácia do modelo proposto, apesar de superior, ficou bem próxima à de RNA, as taxas obtidas foram de 86,76% e 86,71% respectivamente.

No trabalho desenvolvido por Marqués, Garcia e Sánchez (2012) foram apresentadas sete técnicas de classificação para este tipo de problema, variando na utilização de técnicas de *ensemble* ou não. Os melhores resultados foram para os testes que utilizavam técnicas de ensemble e o melhor desempenho obtido foi para as árvores de decisão, seguido por resultados próximos obtidos pelas RNA. Os piores resultados obtidos foram dos classificadores *naive* de Bayes.

Akkoç (2012) propõe um híbrido de três estágios para a classificação de crédito bancário numa base de dados turca. Inicialmente três técnicas foram analisadas separadamente, sendo elas: análise discriminante linear, RL, RNA, seguidamente essas técnicas foram hibridizadas para a comparação de resultados. O híbrido apresentou melhor desempenho que as demais técnicas, possuindo uma acurácia 60% em média com relação aos testes realizados.

Oreski, Oreski e Oreski (2012) propuseram um híbrido de AG com RNA para uma base de dados de um banco da Croácia. A melhor acurácia obtida chegou a 91,07% e os autores puderam concluir que o sistema híbrido proposto é competitivo e pode ser utilizado como seleção de atributos de forma a descobrir os mais significantes na determinação do risco de crédito.

No trabalho desenvolvido por Kao, Chiu e Chiu (2012) foi proposto um modelo de classificação de clientes com base em variáveis Bayesianas latentes. O método proposto foi comparado com técnicas como: Análise discriminante, RNA, SVM, CART (*Classification and Regression Trees*) e MARS. O modelo apresentado apresentou 92,9% de acerto em sua classificação, seguido do CART com aproximadamente 90% de taxa de acurácia.

Em base de dados de banco peruano, Blanco *et al.* (2013) utilizaram as técnicas de RNA, Análise Discriminante Linear (ADL), análise discriminante quadrática (ADQ) e regressão logística para a previsão de adimplência ou não de 5500 clientes. Os resultados revelaram que as RNA tiveram desempenho superior às demais técnicas utilizadas possuindo menor custo de erro de classificação.

Utilizando os métodos de classificação: ADL, ADQ, RL, RNA do tipo MLP, SVM, árvores de classificação, e métodos de ensemble como o bagging e o algoritmo de boosting, Cubiles-de-la-Veja *et al.* (2013) fez uma análise de risco de crédito em uma

instituição financeira peruana. Os resultados obtidos pelos autores apresentaram que o modelo de MLP reduz em 13,7% os custos de erro de classificação em relação à aplicação dos demais métodos testados.

Oreski e Oreski (2014) apresentaram um híbrido de AG com RNA em uma base de dados da Croácia. Os resultados obtidos mostraram que o algoritmo proposto tinha uma habilidade de classificação promissora para seleção de atributos e para a própria análise de risco de crédito de clientes em relação às demais técnicas semelhantes.

Tsai e Hung (2014) comparam ensembles de RNA e um híbrido de RNA sobre três séries de dados benchmarking, uma australiana, uma alemã e a terceira japonesa. Os resultados experimentais obtidos mostraram aos autores que tanto o híbrido quanto os ensembles testados superaram o desempenho da utilização pura de RNA.

Em Bravo, Thomas e Weber (2015) foi proposta uma metodologia que dividia os clientes em “bons pagantes”, “não podem pagar” e “não vão pagar”, semelhantemente às classificações sugeridas por este trabalho. Os autores em questão utilizaram algoritmo de clusterização, RL e RL Multinomial, juntamente com RNA para poder fazer a análise de crédito bancário proposta. Dentre os resultados obtidos, os autores puderam perceber que este tipo de divisão pode melhorar significativamente as informações utilizadas pela metodologia clássica deste tipo de problema, que ocorre na divisão dos clientes em pagantes ou não pagantes.

Para Koutanaei, Sajedi e Khanbabaie (2015) as técnicas de data mining possuem inúmeras aplicações na classificação de crédito de clientes bancários. Desta forma eles desenvolveram um modelo híbrido que, num primeiro estágio, faz o pré-processamento das informações utilizando análise de componentes principais e algoritmo genético. Em seguida, os parâmetros obtidos são utilizados como inputs para o SVM e para as RNA de forma a ser realizada uma comparação entre esses métodos. Os resultados obtidos mostraram que a RNA, com a técnica de ensemble AdaBoost (*adaptive boosting*) tem acurácia superior às demais testadas.

No trabalho desenvolvido por Harris (2015) foram investigadas técnicas estatísticas (regressão linear e RL), RNA, heurísticas (árvores de decisão) e o SVM para a classificação de clientes para uma base de dados alemã e de outra originária de um banco em Barbados. De acordo com os resultados obtidos os autores concluíram que para os dados utilizados a melhor técnica apresentada foi a de SVM.

Danenas e Garvas (2015) apresentam técnicas de RNA, PSO e SVM para a classificação de clientes de uma base de dados de 1999 – 2007 do EDGAR. A pesquisa

experimental realizada propôs que os SVM possuíram melhor desempenho que as demais técnicas testadas com acurácia superior a 90% em sua classificação.

O trabalho desenvolvido por Zhao *et al.* (2015) propõe uma análise de RNA do tipo MLP buscando melhorar a performance utilizando três estratégias: a primeira, otimizando a distribuição das informações das bases de dados utilizando um método denominado de *Average Random Choosing*; a segunda, comparando os efeitos dos números de instâncias de treinamento-validação-teste; a terceira, encontrar o melhor número de neurônios para a camada oculta. Utilizando uma base de clientes de um banco alemão, os autores chegaram a uma acurácia de 87% para 31 neurônios na camada oculta.

Silva *et al.* (2015), em artigo de conferência, também fizeram uso MLP para a análise de crédito bancário. O trabalho explorou a ideia de geração de casos em regiões de padrões que são mais difíceis de produzir respostas consistentes. A proporção de adimplentes para inadimplentes no banco de dados utilizado pelos autores era de 80% a 20%, tendo maior tendência das informações para os adimplentes, baixando a acurácia dos métodos testados.

Abedini, Ahmadzadeh e Noorossana (2015) propuseram uma abordagem com técnica de ensemble e apresentaram uma estrutura de classificação utilizando RNA MLP, SVM, RL, e ainda a RL com *bagging* e as MLP utilizando Florestas Randômicas. A comparação dos resultados mostrou que o modelo proposto tem desempenho superior à aplicação das técnicas por si só. A melhor acurácia apresentada foi de 76,67% para a aplicação de MLP com a técnica de *ensemble* das Florestas Randômicas.

Zakirov, Bondarev e Momtselidye (2016) apresentaram uma estrutura de comparação de diferentes métodos de classificação para análise de crédito de dados real. Com base nos resultados obtidos os autores concluíram que a melhor metodologia testada foi a de Florestas Randômicas.

Aplicando os métodos de CART, MLP, SVM, redes de Bayes, redes de Bayes *naïve*, Ozturk, Namli e Erdal (2016) exploraram a previsão de diferentes técnicas de inteligência artificial para a análise de crédito bancário. Os resultados obtidos mostraram que as técnicas tiveram acurácia de previsão acima de 90% quando o desempenho de modelos estatísticos convencionais possuem acurácia em torno de 70%.

Abdou *et al.* (2016) aplicaram três métodos de classificação, sendo eles: RL, CART e RNA de correlação para uma base de dados de bancos em Camarões. Dentre os resultados apresentados, os autores perceberam que o melhor desempenho ocorria para

as RNA testadas onde a identificação dos “bons pagadores” acontecia com 97,43% de acurácia.

Punniyamoorthy e Sridevi (2016) utilizaram RNA e compararam com um modelo de SVM para a classificação de clientes bancários do Reino Unido. Dentre os resultados obtidos, os autores puderam constatar que o método utilizado de SVM possuiu melhores resultados que as RNA chegando a uma acurácia de 80% enquanto as RNA ficaram em torno de 77%.

Ala'raj e Abbod (2016) apresentaram uma abordagem com os seis classificadores mais conhecidos para o problema de risco de crédito, entre eles: regressão logística, redes neurais artificiais, SVM, Florestas Randômicas, árvores de decisão e redes Bayesianas *naïve*. Foram usadas 5 bases de dados, uma alemã, uma australiana, uma japonesa, uma iraniana e uma polonesa, para serem realizados os testes dos modelos mencionados anteriormente. Dos resultados obtidos a que possuiu melhor desempenho geral foi a Florestas Randômicas, melhor em 4 das cinco bases, seguido das RNA, melhor acurácia na base alemã.

Andric e Kalpic (2016) apresentaram uma comparação entre três algoritmos de classificação, regressão logística, redes neurais artificiais e *gradient boosting*. Para a primeira base de dados, a alemã, houve um empate de acurácia entre a RNA e o *gradient boosting*, ambos ficaram com 75% de acerto, já para a segunda, a japonesa, o *gradient boosting* possuiu melhor resultado dentre as técnicas testadas apresentando 91% de acurácia.

Em outro trabalho desenvolvido por Ala'raj e Abbod (2016) foram realizadas três comparações entre cinco métodos comuns na literatura de classificação de clientes bancários (RNA, SVM, Florestas Randômicas, árvores de decisão, redes Bayesianas *naïve*), as análises foram: comparação individual desses métodos para escolha do melhor resultado e assim aplicação de modelos de ensemble, em seguida comparação do modelo proposto com os classificadores básicos e métodos tradicionais e por fim, comparação do método proposto com o já embasado em literatura. Na primeira etapa os autores concluíram que as Florestas Randômicas possuíam melhor acurácia que os demais modelos testados, situação que se manteve na segunda comparação. Tendo os resultados, os autores desenvolveram o modelo denominado de ConsA e perceberam que seus resultados foram superiores a todos os demais testes realizados, sendo 79% a menor acurácia apresentada pelo método (para a base de dados alemã).

Guo e Dong (2017) propuseram um modelo de PSO multiobjetivo para a classificação de clientes e o compararam com as redes Bayesianas *naïve*, RL, otimização mínima sequencial, árvores de decisão e RNA. A acurácia do modelo proposto foi superior aos demais testados (74,64%) para a primeira base de dados testada, mas para a segunda as RNA tiveram desempenho superior às demais, com 61,15% de acurácia, para a última base de dados testada o melhor desempenho foi para a regressão logística com acurácia de 76,70%.

No trabalho desenvolvido por Luo, Wu e Wu (2017) foi investigado o desempenho de modelos de pontuação de crédito aplicado em base de dados CDS. Foram utilizadas RNA do tipo *deep learning* e MLP, RL e SVM para as comparações dos testes. A melhor acurácia encontrada foi para as RNA de *deep learning* (100%) seguida das MLP (87,85%).

Por fim, o último trabalho que apresentou RNA em sua metodologia foi desenvolvido por Bequé e Lessmann (2017), onde os autores compararam RNA e outros métodos de classificação com a técnica de *ensemble* conhecida como *extreme learning machines* (ELM). As ELM são uma metodologia proposta como alternativa às RNA *feed-forward* padrões. Analisando os resultados obtidos, os autores constataram que as ELM provaram possuir um poder discriminante tanto para a comparação isolada de técnicas quando para a comparação utilizando ensembles.

Após analisados os trabalhos que utilizaram RNA em sua metodologia, são apresentados a seguir os que utilizaram diferentes métodos na pesquisa dos últimos 5 anos ligados à Risco de Crédito.

Tong, Mues e Thomas (2012) utilizando programação em R, fizeram uma análise de crédito bancário utilizando técnicas estatísticas como regressão logística, modelo de cura mista e regressão de Cox. Os desempenhos dos modelos testados foram extremamente semelhantes em termos de discriminação dos clientes. Mas o modelo de cura mista permitiu a distinção de duas subpopulações.

García, Marqués e Sánchez (2012) fizeram uma extensa pesquisa analisando 20 algoritmos, utilizando diferentes tipos de SVM, técnicas estatísticas e heurísticas, para aplicação em 8 bases de dados distintas. Dentre os resultados levantados, os melhores desempenhos ficaram com três modelos heurísticos, sendo eles *Edited Nearest Neighbour* (ENN), *Nearest Centroid Neighbours Editing* (NCN) e por fim *CHC evolutionary algorithm*.

Konno e Saito (2013) desenvolveram o trabalho acerca de classificação de crédito de empresas japonesas utilizando um modelo heurístico para a divisão em grupos de superfície. Dentre os resultados obtidos, os autores puderam melhorar o desempenho natural do agrupamento por elipses incorporando uma ideia de margem máxima de hiperplanos. Não mencionando quais os tipos de divisão realizados, os autores atingiram 96% de acurácia no modelo de classificação proposto.

Outro trabalho desenvolvido em avaliação de risco de crédito foi proposto por Li, Wei e Hao (2013). Nesta pesquisa, foi proposta uma metodologia de ensemble juntamente com a técnica de SVM e ACP em duas bases de dados consideradas benchmark de crédito bancário. Com base nos experimentos computacionais realizados, os autores puderam apresentar que o método proposto não apenas teve melhor acurácia (89%) dentre os modelos testados, mas também mais estabilidade dentre as instâncias testadas.

Nikolic *et al.*(2013) utilizando uma abordagem de RL propuseram um modelo de avaliação de risco de crédito de pessoas jurídicas de diferentes bases de dados ao longo de 5 anos (de 2007 a 2011). As acurácias deste trabalho variaram de 50,99% a 53,91% dentre as instâncias testadas de acordo com a quantidade de variáveis analisadas.

Marqués, García e Sánchez (2013) fizeram uma aplicação de avaliação de crédito com base de dados reais onde foi testada a habilidade de métodos de reamostragem de dados. Foram comparadas as técnicas de RL e SVM, onde, analisando os resultados computacionais, os autores perceberam que os modelos de reamostragem podem melhorar o desempenho das técnicas de classificação.

No artigo desenvolvido por Zhu *et al.* (2013), foram testados oito métodos, dentre eles: ADL, ADQ, Árvores de Decisão, RL e SVM aplicados em duas bases de dados da UCI. Dentre os resultados obtidos, os autores puderam perceber que o algoritmo que possuiu melhor desempenho foi o de Árvores de Decisão.

Niklis, Duompos e Zopounidis (2014) fizeram a comparação de diferentes tipos de SVM com a técnica de RL para uma avaliação de risco de crédito de dados reais dos quais a base iria do ano de 2005 ao ano de 2010. O melhor resultado obtido foi para uma das técnicas de SVM testadas atingindo 79,6% de acurácia.

Em Zhang, Gao e Shi (2014) propuseram uma comparação entre modelos heurísticos do tipo *fuzzy*, Classificador de Otimização Multi-Critério (MCOC) e SVM. Os resultados avaliados pelos autores mostraram que as análises de risco de crédito realizadas pelo MCOC possuíram melhor desempenho que os demais métodos apresentados.

No trabalho desenvolvido por Tong e Li (2015), novamente utilizando SVM e possíveis combinações desta técnica, os autores buscaram realizar a avaliação de risco de crédito de empresas chinesas. O melhor resultado obtido, de erro de 9%, ficou com a técnica denominada de *Relevance Vector Machines* (RVM), uma variante do SVM.

Outra análise de diferentes técnicas de SVM foi realizada por Yi e Zhu (2015), estas foram aplicadas em uma base de dados australiana e outra alemã, para testar seus potenciais quanto à classificação de crédito bancário. O melhor resultado para a base de dados alemã foi através de utilização de funções de bases radiais juntamente com o SVM, atingindo uma acurácia de 76%. Já para a base australiana, a maior acurácia obtida foi de 84,83% utilizando funções polinomiais no SVM testado.

Li (2016) desenvolveu para uma conferência uma análise de gestão de risco de crédito utilizando SVM baseado em lógica *fuzzy*. Dentre os resultados obtidos no trabalho desenvolvido, os autores encontraram uma taxa de acurácia de 77,7% para as bases de dados testadas.

Lanzarini (2017) apresentou um modelo alternativo que pudesse gerar regras que funcionassem não apenas com atributos numéricos como também nominais. A técnica utilizada baseava-se em LVQ - PSO (*Learning Vector Quantization* combinada com PSO). A melhor acurácia obtida pelo algoritmo proposto foi de 81,05% dentre os testes realizados.

Em Li (2017) foi proposto um modelo de SVM semi-supervisionado para avaliação de crédito de consumidores. Com exceção de uma instância, o modelo proposto foi superior em todos os demais testes realizados, variando sua taxa de acerto de 82,7% até 97,6%.

Por fim, o último trabalho desta seção foi desenvolvido por Maldonado, Pérez e Bravo (2017). Nesta pesquisa, foram desenvolvidas duas propostas baseadas em SVM para que fosse possível realizar a seleção de atributos e ainda a classificação de créditos bancários. Dentre os SVM testados, o que teve melhor acurácia foi o Fisher-SVM (taxa de acerto de 70,6%). Este algoritmo apresentado pelos autores utilizava um algoritmo de pontuação de Fisher para a seleção de atributos de acordo com seus custos e em seguida foi aplicado o SVM nos selecionados.

A seguir é apresentado novamente um quadro (Quadro 6) resumo dos artigos apresentados acima.

**Quadro 6 – Alguns Trabalhos Correlatos com Abordagem em Risco de Crédito**

AUTORES	ANO	PUBLICAÇÃO	CLASSIFICAÇÃO PERIÓDICO			# CITAÇÕES	ABORDAGEM	MÉTODO						SOFTWARE
			Qualis	H Index				Redes Neurais Artificiais	Estatística	Support Vector Machines	Heurísticos	Programação Linear		
BEQUÉ A.; LESSMANN S.;	2017	Periódico	A1	131		1	Classificação	X		X				WEKA
LUO C.; WU D.; WU D.;	2017	Periódico	A2	76		2	Classificação	X	X	X				R
MALDONADO S.; PÉREZ J.; BRAVO C.;	2017	Periódico	A1	200		2	Classificação			X			X	-----
GUO Y.; DONG C.;	2017	Conferência	-----	-----		0	Classificação	X	X			X		WEKA
LI Y.; TIAN Y.; LI K.; YHOU F.; YANG W.;	2017	Periódico	A1	131		2	Classificação			X				-----
LANZARINI L.C.; VILLA MONTE A.; BARRIVIERA A.F.; JIMBO SANTANA P.;	2017	Periódico	B1	28		1	Classificação					X		-----
ALA'RAJ.M.; ABBOD M.F.;	2016	Periódico	A1	131		4	Classificação	X		X		X		-----
LI Y.-X.;	2016	Conferência	-----	-----		2	Classificação			X				-----
ANDRIĆ K.; KALPIĆ D.;	2016	Conferência	-----	-----		3	Classificação	X	X			X		-----
ALA'RAJ.M.; ABBOD M.F.;	2016	Periódico	A1	74		10	Classificação	X	X	X		X		MATLAB
PUNNIYAMOORTHY M.; SRIDEVI P.;	2016	Periódico	B3	44		2	Classificação	X		X				MATLAB
ABDOU H.A.; TSAFAK M.D.D.; NTIM C.G.; BAKER R.D.;	2016	Periódico	A1	74		4	Classificação	X	X			X		SPSS
OYTURK H.; NAMLI E.; ERDAL H.I.;	2016	Periódico	B1	45		5	Classificação	X		X		X		-----
YAKIROV D.; BONDAREV A.; MOMTSELIDYE N.;	2016	Conferência	-----	-----		4	Classificação	X		X		X		R
ABEDINI M.; AHMADYADEH F.; NOOROSSANA R.;	2016	Periódico	B1	28		1	Classificação	X	X	X		X		WEKA
YI B.; YHU J.;	2015	Conferência	-----	-----		2	Classificação			X				-----
SILVA C.; VASCONCELOS G.; BARROS H.; FRANÇA G.;	2015	Conferência	-----	-----		1	Classificação	X						WEKA
YHAO Y.; XU S.; KANG B.H.; KABIR M.M.J.; LIU Y.; WASINGER R.;	2015	Periódico	A1	131		33	Classificação	X						MATLAB
DANENAS P.; GARSVA G.;	2015	Periódico	A1	131		26	Classificação	X	X	X		X		WEKA
HARRIS T.;	2015	Periódico	A1	131		60	Classificação	X	X	X		X		-----
KOUTANAEI F.N.; SAJEDI H.; KHANBABAIEI M.;	2015	Periódico	B3	51		19	Classificação	X	X	X		X		RAPIDMINER / WEKA
TONG G.; LI S.;	2015	Periódico	B1	39		1	Classificação			X				MATLAB
BRAVO C.; THOMAS L.C.; WEBER R.;	2015	Periódico	B1	83		11	Classificação	X	X					-----



### 3.5.4 Aplicações Gerais de RNA

Nesta seção, foi realizado o mesmo procedimento que o levantamento dos trabalhos correlatos apresentados anteriormente (falência, mercado de ações e risco de crédito) com exceção que não foi limitado o tipo de aplicação da RNA. De modo a resumir o que foi obtido de aplicações práticas utilizando RNA, foi escolhido apresentar o que foi pesquisado nesta área no último ano, chegando a um total de 25 artigos.

O trabalho desenvolvido por Sánchez *et al.* (2017), por exemplo, propôs um modelo matemático que utiliza das RNA para verificar o potencial energético de ondas marítimas. Analisando as informações de equipamentos subaquáticos que fazem a medida do comprimento de ondas das costas leste e oeste dos Estados Unidos da América (EUA) e da costa da Bahia no Brasil, os autores verificaram que o potencial energético da costa oeste dos EUA teria maior potencial energético que a costa brasileira analisada.

Na pesquisa de Hussain e Alalili (2017) foi apresentado um modelo híbrido para verificar o potencial de energia solar em determinadas regiões de interesse. Utilizando-se de uma comparação com quatro modelos diferentes de RNA, os autores puderam visualizar que, para os dados testados, a melhor opção em termos de acurácia era com as redes GRNN (*Generalized Regression Neural Networks*), sendo superior a 97% em ambas as abordagens testadas.

Em Garcia *et al.* (2017) verifica-se uma aplicação de previsão de queda de pressão durante a evaporação de determinado composto químico. A RNA proposta foi treinada utilizando informações de 127 experimentos encontrados na literatura possuiu acurácia de 90% considerando uma significância de 10% e de 99% para uma significância de 30%, mostrando que as RNA testadas, comparadas com correlações na literatura, demonstraram alto poder preditivo de acordo com os autores.

Moore *et al.* (2017) desenvolveram seu trabalho com RNA, em combinação com a ACP, numa aplicação que visava estimar a idade de moscas varejeiras, informação importante no campo da ciência forense. Com acurácia de 85,4% em sua metodologia, os autores observaram grande potencial na ferramenta em análise forense de cadáveres.

Santhosh *et al.* (2017) utilizaram as RNA para previsão de confiabilidade de cabos de controle e instrumentação denominados de XLPE I&C. Diferentes arquiteturas de RNA foram testadas e, de acordo com os resultados computacionais obtidos, os autores verificaram que as RNA forneceram resultados semelhantes a experimentos práticos realizados com o mesmo objeto.

Denisov (2017) trabalhou com detecção e classificação de poluentes através de utilização de enzimas biosensoras. A arquitetura da RNA testada ficou constituída de 61 neurônios de entrada, 3 camadas ocultas de neurônios, e, por fim, 3 neurônios de saída. Com isso, o autor conseguiu demonstrar que as RNA são uma ferramenta adequada para este tipo de problema.

Em Božnar *et al.* (2017) verificou-se uma aplicação utilizando RNA em previsão de radiação solar com informações obtidas durante a primavera de 2013 na Eslovênia com coleta de dados a cada meia hora. Utilizando RNA do tipo MLP, os autores conseguiram atingir resultados satisfatórios em relação à aplicação realizada, verificando que os erros obtidos chegavam a 26,4% que, considerando as variáveis do problema, foi considerado baixo neste.

Em um estudo ligado à epigenética forense, Vidaki *et al.* (2017) utilizaram RNA para prever idade de organismos uni e multicelulares presentes em seres em decomposição. O banco de dados utilizado era composto por 1156 amostras de sangue. Os autores mostraram que a utilização de RNA no campo da pesquisa realizada tem crescido apresentando grande potencial em relação à acurácia, um exemplo seria que o erro no trabalho desenvolvido ficou em torno de 6%.

Buscando descrever um modelo de comportamento para pilares de fundação no ramo de construções civis, Nejad e Jaksa (2017) utilizaram RNA em um banco de dados de 500 instâncias obtidas em literatura similar. O artigo em questão comparava os resultados obtidos pelas RNA com os que foram obtidos por métodos tradicionais para este tipo de problema. Os resultados obtidos pelas RNA foram considerados aceitáveis apenas para pilares com base superior a 138 mm, chegando a ter um erro de 3,21mm (cerca de 2%).

Cabrera *et al.* (2017) propuseram um modelo de RNA que pudesse prever o potencial de geração de água potável de um protótipo de desalinação de água marítima utilizando osmose invertida. Os modelos de RNA foram desenvolvidos para gerar valores de fluxo de alimentação e pressão de operação tendo em conta não apenas a energia elétrica disponível, mas também a temperatura e a condutividade da água de alimentação. Os autores concluíram que os modelos de RNA foram capazes de gerenciar com sucesso a energia elétrica aleatória e ainda o teste de hipóteses realizado não mostrou diferenças estatísticas significativas entre os erros obtidos.

Sitton, Zeinali e Story (2017) utilizaram RNA para uma pesquisa que visava a classificação de tipos de solos para construção de blocos de terra compactados. A

arquitetura de rede testada ficou constituída de uma camada de entrada com cinco neurônios, duas camadas ocultas ambas com 10 neurônios e uma camada de saída com apenas um neurônio. Os resultados foram variados para cada um dos 12 tipos de solos testados, sendo que a maior acurácia (99,40%) ficou com o solo do tipo 6 e a menor (14,5%) ficou com o solo do tipo 8.

Bouhoune *et al.* (2017) apresentaram um modelo híbrido utilizando RNA e lógica *fuzzy* para controle de uma máquina de indução. Os desempenhos do híbrido proposto aplicado ao motor de indução trifásico fornecido foram investigados por simulações e as provas de desempenho, por resultados comparativos a outros controles anteriormente testados sem o modelo proposto. Os resultados mostram a viabilidade e o bom desempenho obtidos pelo controle proposto melhorando o tempo de resposta e a robustez do sistema.

Cascardi, Micelli e Aiello (2017) propuseram um modelo de RNA para a previsão de força compressiva em colunas circulares de concreto. Os resultados evidenciaram que o modelo proposto é adaptado para o projeto de concreto confinado e garante uma precisão melhorada em relação aos concorrentes disponíveis, atingindo 8% no máximo de erro para as instâncias testadas.

Yang *et al.* (2017) desenvolveram um modelo híbrido utilizando otimização evolucionária com ACP para o treinamento de um modelo de RNA. O modelo proposto tratou de uma RNA de três camadas aplicada em 17 bases de dados benchmark. Os resultados obtidos mostraram que a RNA apresentada obteve melhores resultados que outras RNA baseadas em algoritmos evolucionários apresentadas na literatura.

Mohammed *et al.* (2017) desenvolveram uma RNA de classificação para identificação de carcinoma nasofaríngeo buscando encontrar as melhores características que levem a esse tipo de carcinoma. Tanto o treinamento quanto o teste obtiveram resultados satisfatórios para os autores, apresentando acurácia de 88,03%.

Tahir e Mandal (2017) trabalharam numa base de dados relacionados à carga de curvatura de cascas cilíndricas finas sob compressão axial, utilizando RNA para prever esse encurvamento. Os autores puderam observar que os resultados obtidos pelas RNA são próximos a resultados de experimentos realizados com os materiais.

Ascione *et al.* (2017) desenvolveram uma estrutura para adaptação de energia a um custo ótimo para diferentes tipos de construções através de um algoritmo multiobjetivo baseado em RNA. Dentre os resultados obtidos, os autores puderam

observar um erro médio relativo de 2,55% mostrando que os resultados obtidos pela estrutura desenvolvida tiveram desempenho satisfatório para o problema aplicado.

Em outra aplicação à construção civil, Kotsovou, Cotsovos e Lagaros (2017) utilizaram as previsões de um modelo RNA sobre o modo de falha e capacidade de carga das articulações para comparar com suas contrapartes armazenadas no banco de dados. Tal comparação confirmou as deficiências já relatadas dos métodos atuais, demonstrando que o modelo RNA pode fornecer alternativas confiáveis.

Cervone et al. (2017) utilizaram uma técnica de ensemble em RNA para prever a energia produzida por células fotovoltaicas. Os resultados mostram que uma solução ensemble combinada com RNA produz melhores resultados e que a solução proposta é adequada para computação em escala maciça.

Luján *et al.* (2017) desenvolveram um modelo que previsse a eficiência volumétrica de motores a combustão interna propondo um modelo de RNA MLP. Os resultados apresentados mostraram aos autores que o método proposto funciona com maior velocidade de aprendizagem, recursos computacionais reduzidos e menor complexidade em relação a outros métodos testados na literatura.

Elnesr e Alazba (2017) propuseram um modelo de RNA para a simulação de distribuição de água através de observação de uma goteira. O trabalho desenvolvido utilizou um conjunto de 51 variáveis de entrada em um modelo de goteira projetado em ambiente 2D/3D. A análise dos resultados obtidos mostrou para os autores que a variável que mais afeta o comportamento estudado é o tempo de infiltração da água.

Mirabdolazimi e Shafabakhsh (2017) utilizaram as RNA para realizar a previsão do comportamento da mistura de asfalto com uma determinada fibra. O modelo proposto apresentou resultados semelhantes a experimentos realizados pelos autores com esse tipo de asfalto, aonde chegaram à conclusão que a ferramenta proposta tem potencial para este tipo de problema.

Anderlini *et al.* (2017) propuseram uma RNA para o controle reativo de um conversor de energia das ondas marítimas. O objetivo da otimização é maximizar a absorção de energia, limitando o deslocamento para evitar falhas. Uma vez que o treinamento ocorreu, o algoritmo apresentou uma previsão de absorção de energia semelhante ao que ocorre no momento.

Kocadagli e Langari (2017) propuseram um modelo de RNA que, através de exames de eletroencefalogramas, buscou descrever o comportamento no momento de ataques epiléticos. De acordo com os resultados da análise, esta abordagem não só permite

analisar profundamente os sinais dos eletroencefalograma para a detecção de epilepsia, mas também fornece as melhores configurações de modelo para RNAs em termos de confiabilidade e complexidade.

Lins *et al.* (2017) utilizaram as RNA para a previsão de demência em pacientes idosos. A análise demonstrou que o uso do modelo proposto produz valores médios para uma acurácia superior a 90%.

Em uma aplicação teórica, Manngård, Kronqvist e Böling (2018) buscaram encontrar a estrutura e os parâmetros de uma RNA propondo um modelo denominado de otimização esparsa. Experimentos foram realizados com banco de dados considerados benchmark, tanto para o problema de regressão quanto para o de classificação. Os autores concluíram então que o modelo proposto tinha desempenho eficiente em uma aplicação com dados ligados à distribuição de energia.

A seguir, apresenta-se no Quadro 7 o resumo das aplicações gerais de RNA cujos artigos foram apresentados anteriormente.

**Quadro 7 – Alguns Trabalhos Correlatos com Aplicações Gerais de RNA**

AUTORES	ANO	CLASSIFICAÇÃO PERIÓDICO			APLICAÇÃO							
		Qualis	H Index	# CITAÇÕES	ABORDAGEM	Medicina	Biologia	Construção	Energia	Engenharia	Teórica	SOFTWARE
M. MANNING; J. KRONQVIST; J. M. BÖLING,	2018	A2	94	0	Regressão/ Classificação				X	X	X	-----
A.J.C.C. LINS, M.T.C. MUNIZ, A.N.M. GARCIA, A.V. GOMES, R.M. CABRAL, C.J.A. BASTOS-FILHO,	2017	B1	71	1	Regressão/ Classificação	X						C++
O. KOCADAGU, R. LANGARI	2017	A1	131	0	Classificação	X						MATLAB
E. ANDERLINI, D.I.M. FOREHAND, E. BANNON, M. ABUSARA	2017	B3	10	0	Previsão				X			MATLAB
S. M. MIRABDOLAZIMI; Gh. SHAFABAKHSH	2017	----	99	1	Previsão			X				-----
M.N. EINESR, A.A. ALAZBA	2017	A2	81	0	Previsão					X		NeuroCharter
J. M. LUIÁN, H. CLIMENT, I. M. GARCÍA-CUEVAS, A. MORATAL	2017	A1	112	1	Previsão				X	X		-----
G. CERVONE; I. CLEMENTE-HARDING; S. ALESSANDRINI; L. D. MONACHE	2017	A2	134	4	Previsão				X			JAVA/ Python
G. M. KOTSOVOU, D. M. COTISOVOS, N. D. LAGAROS	2017	A1	99	0	Previsão			X				-----
F. ASCIONE, N. BIANCO, C. DE STASIO, G. M. MAURO, G. P. VANOLI	2017	A1	123	7	Previsão			X	X			MATLAB
Z. R. TAHIR, P. MANDAL,	2017	A1	99	0	Previsão			X				MATLAB
M. A. MOHAMMED, M. K. A. GHANI, R. I. HAMED, D. A. IBRAHIM, M. K. ABDULLAH	2017	B1	21	7	Classificação	X						-----
T. YANG, A. A. ASANJAN, M. FARIDZAD, N. HAYATBINI, X. GAO, S. SOROOSHIAN	2017	A1	131	0	Previsão						X	-----

FONTE: A autora, 2017.



## 4 MÉTODO DE PESQUISA

Este capítulo é dedicado à descrição da metodologia empregada nos modelos do capítulo 5. Inicialmente é descrita a classificação da pesquisa realizada, seguida do pré-processamento dos dados coletados, logo após são descritos os métodos empregados em relação ao seu funcionamento nos *softwares* utilizados, por fim são apresentados os algoritmos de classificação utilizados com seus respectivos parâmetros, bem como os índices de avaliação do desempenho dos modelos.

### 4.1 CLASSIFICAÇÃO DA PESQUISA

De acordo com Gil (2008), o método indutivo parte do específico e coloca a generalização como um produto posterior do trabalho de coleta de dados particulares. Seguindo esse tipo de raciocínio, a generalização de uma determinada ideia não deve ser buscada inicialmente, mas constatada a partir da observação de casos concretos que, suficientemente, poderão confirmar essa realidade.

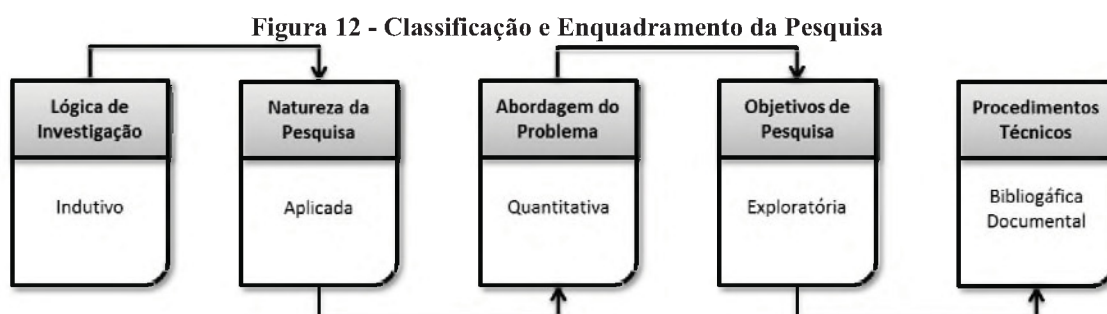
Observando as naturezas de pesquisa propostas por Silva e Menezes (2005), o trabalho desenvolvido é uma pesquisa aplicada que, por definição, deve gerar conhecimentos voltados à aplicação prática e dirigidos à solução de problemas específicos.

Quanto à abordagem do problema, de acordo com Silva e Menezes (2005), uma pesquisa pode ser quantitativa ou qualitativa. No caso da dissertação desenvolvida, trata-se de uma pesquisa predominantemente quantitativa, uma vez que considera valores quantificáveis, ou seja, traduz em números determinadas opiniões e informações para aí analisa-las e classifica-las.

Segundo Gil (2002), com base nos objetivos da pesquisa, existem três grandes grupos de classificações: as pesquisas exploratórias, as descritivas e as explicativas. O trabalho aqui descrito é considerado como pesquisa exploratória, considerando assim que esta dissertação tem por objetivo proporcionar maior familiaridade com determinada problemática, com o propósito de tornar o problema estudado mais explícito ou então a construção de hipóteses ao seu redor. Ou seja, o principal objetivo deste tipo de pesquisa é o aprimoramento de ideias ou, ainda, a descoberta de intuições.

Existem diferentes procedimentos técnicos que podem ser adotados para a análise de um problema de pesquisa. Para o trabalho apresentado, segundo a classificação de Gil (2002), são utilizados procedimentos tais como: pesquisa bibliográfica, pesquisa documental e estudo de caso. Inicialmente temos a pesquisa bibliográfica desenvolvida com base em material elaborado previamente, dessa forma, sendo constituída especialmente de livros e artigos científicos; utiliza-se também a pesquisa documental, pois dependerá dos contratos de concessão e informações dos clientes jurídicos da instituição financeira de estudo.

A seguir, na Figura 12, é apresentado um resumo das informações anteriormente descritas.



Fonte: A autora, 2017.

## 4.2 METODOLOGIA APLICADA

O problema de classificação de clientes aqui abordado tem como objetivo a aplicação de modelos empíricos para apoiar a tomada de decisão no negócio de crédito de varejo. Esse tipo de problema é uma estimativa baseada em modelos da probabilidade de que um mutuário mostre algum comportamento indesejável no futuro. Na aplicação de classificação de risco de crédito, os credores empregam modelos preditivos, chamados *scorecards*, para estimar a probabilidade de um candidato ser padrão (ou não) (LESSMANN *et al.*, 2015).

Para o estabelecimento da metodologia do trabalho foi necessário realizar um levantamento dos trabalhos recentes com aplicações de RNA, verificando quais são as principais aplicações de RNA no setor financeiro e a lacuna existente dentro desse ramo de pesquisa. Dentre os trabalhos pesquisados, observou-se que não é com frequência que se encontra um comparativo entre topologias de redes neurais MLP, mais raros os que fazem comparações não apenas com as arquiteturas de MLP como também com outros

métodos. A pesquisa em questão visa preencher esta lacuna propondo uma comparação entre algoritmo de redes neurais MLP, RBF e a técnica estatística de RL, testando também variadas topologias para as RNA.

Dessa forma, pretende-se testar três modelos RNA (discriminar clientes adimplentes, inadimplentes e temporariamente inadimplentes) com diferentes topologias para cada um dos métodos de validação apresentados anteriormente. A implementação das RNA (MLP e RBF) ocorrerá com a utilização do *software* Matlab 2017a da MATHWORKS INC e da RL com a utilização do *software* SPSS.

O treinamento do MLP foi inicializado em uma rede, onde na camada oculta começará com 1 neurônio, variando até chegar a 20 neurônios nesta camada; para a validação da rede foi utilizado o modelo do *hold-out* estratificado, dividindo as instâncias em 2/3 para treinamento e 1/3 para o teste. As RNA foram testadas para 31 diferentes conjuntos de pesos iniciais, de forma a analisar seu desempenho para essas variadas situações. Com relação às RBF, o número de neurônios na camada oculta deverá ser no máximo igual ao número de instâncias utilizadas para o treinamento e, de forma análoga às MLP, os dados foram divididos em 2/3 para treinamento e 1/3 para a validação da rede. Por fim, o último método, a construção do modelo de RL ocorreu com o auxílio do *software* SPSS da desenvolvedora IBM. Os demais parâmetros dos modelos testados são discutidos na seção 4.6 desde capítulo. Ao final, deseja-se encontrar, dentro das metodologias testadas, a que possui maior acurácia para os conjuntos de testes.

### 4.3 PRÉ-PROCESSAMENTO DOS DADOS COLETADOS

Feita a coleta de dados descrita na seção 2.2.1, foi necessário realizar a binarização das informações levantadas, isto é, cada um dos 15 atributos descritos foi transformado em “0” ou “1”, estando os mesmos originalmente em formato nominal (qualitativo) ou numérico (quantitativo).

Atributos como a natureza jurídica, a atividade da empresa, o segmento de atuação no mercado, o risco atribuído ao empréstimo, a escala de estabelecimento de limite de crédito e a restrição junto ao banco, tem sua binarização de acordo com a sub-divisão de cada um deles.. Apenas para o caso da atividade da empresa, com a variedade existente (629 opções) foi necessário analisar as seções do código CNAE (Classificação Nacional de Atividades Econômicas) de cada cliente, reduzindo-os para 19.

Para exemplificar o processo, existem cinco valores respectivos ao “risco de empréstimo”, dessa forma, para esse atributo, cada cliente terá um vetor binário onde o 1 estará na posição representada pelo risco que ele possui para o banco, e as demais posições serão zeradas, como é possível verificar na Tabela 2 a seguir, que contém um exemplo gerado aleatoriamente não representando clientes específicos.

**Tabela 2 - Binarização do atributo "Risco"**

CLIENTE	<i>RISCO</i>				
	A	B	C	D	E
1	1	0	0	0	0
2	0	1	0	0	0
3	0	1	0	0	0
4	0	0	1	0	0
5	0	0	0	0	1

Fonte: A autora, 2017

Com relação aos atributos relacionados às datas, como a data de constituição da empresa e a data desde a qual a empresa é cliente do banco, foi necessário inicialmente transformar as datas em valores numéricos, ordena-los do menor para o maior, e dividir os quartis para que fosse possível agrupar o máximo de informações num espaço reduzido. Assim o atributo poderia ser binarizado da mesma forma que o exemplo acima, a seguir, na Tabela 3, temos o atributo “Data de Constituição” binarizado de forma a exemplificar o procedimento apresentado, o exemplo a seguir, de forma similar ao da Tabela 2, foi gerado de forma aleatória.

**Tabela 3 - Binarização do atributo "Data de Constituição"**

CLIENTE	<i>DATA DE CONSTITUIÇÃO</i>				
	15/02/1939	03/06/1997	20/02/2004	26/02/2009	
	01/06/1997	19/02/2004	19/02/2009	02/12/2016	
1	0	0	1	0	
2	1	0	0	0	
3	1	0	0	0	
4	0	1	0	0	
5	0	0	0	1	

Fonte: A autora, 2017.

Por fim, o último procedimento de binarização a ser apresentado foi o realizado em valores numéricos, como é o caso do Faturamento Bruto Anual e a Quantidade de Funcionários. Em processo análogo aos valores em formato de datas, também foram divididos em faixas que representassem seus quartis, a seguir na Tabela 4, a seguir tem-se o exemplo aleatório da binarização da “Quantidade de Funcionários”.

**Tabela 4 - Binarização do atributo "Quantidade de Funcionários"**

CLIENTE	QUANTIDADE DE FUNCIONÁRIOS			
	0	2	5	12
	1	4	11	5113
1	0	0	1	0
2	0	0	1	0
3	1	0	0	0
4	0	1	0	0
5	0	0	0	1

Fonte: A autora, 2017.

#### 4.4 REDES NEURAIAS ARTIFICIAIS NO MATLAB

As ferramentas das RNAs para MATLAB contém algoritmos, modelos pré-treinados, aplicativos para serem criados, visualizados e simulados para quaisquer RNA (BEALE, HAGAN e DEMUTH, 2017). Segundo os autores, dentre os principais atributos oferecidos no MATLAB, alguns exemplos dos que são considerados chave são: as Redes Neurais Convolucionais (RNC), redes neurais de longas memórias de curto prazo (LMCP) – para classificação de séries temporais, codificadores automáticos – para aprendizado de atributos, algoritmos de aprendizado não-supervisionado e algoritmos de treinamento supervisionado, MLP, RBF, redes neurais recorrentes.

A forma mais simples de criação de uma RNA do tipo MLP no MATLAB é utilizando uma função de criação de rede, aqui, no caso, a função “*feedforwardnet*”. Os sub-objetos chave da rede são as entradas (*inputs*), o número de camadas (*layers*), as saídas (*outputs*), os *biases* e ainda os pesos (*weights*). Após as definições feitas é escolhida a função de ativação dos neurônios, podendo divergir entre camadas (BEALE, HAGAN e DEMUTH, 2017). O MATLAB oferece ainda diferentes algoritmos de treinamentos para as redes neurais artificiais, esses treinamentos são apresentados na Quadro 8 abaixo.

Quadro 8 - Funções de Treinamento do Matlab

<i>Função no MATLAB</i>	<i>Algoritmo Utilizado</i>
<i>trainlm</i>	Levenberg-Marquardt
<i>traingd</i>	Gradiente Descendente
<i>traingdx</i>	Taxa de Aprendizado Variável Gradiente Descendente
<i>traingdm</i>	Gradiente Descendente com <i>Momentum</i>
<i>trainbr</i>	Regularização Bayesiana
<i>trainbfg</i>	BFGS Quasi-Newton
<i>trainrp</i>	<i>Backpropagation</i> resiliente
<i>trainscg</i>	Gradiente Escalonado Conjugado
<i>traincgb</i>	Gradiente Conjugado com reinícios de Powell/Beale
<i>traincgf</i>	Gradiente Conjugado de Fletcher-Powell
<i>traincgp</i>	Gradiente Conjugado de Polak-Ribière
<i>trainoss</i>	Secante de um passo

FONTE: Adaptado de BEALE, HAGAN e DEMUTH, 2017

Dessas, a função que possui o treinamento mais rápido é a “*trainlm*”, a função *default* para a *feedforwardnet*. O método Quasi-Newton também é considerado rápido, mas ambas as funções são menos eficientes para grandes redes (com centenas de pesos), uma vez que necessitam mais memória e tempos computacionais para esses casos (BEALE, HAGAN e DEMUTH, 2017).

Para as RBF, a função “*newrb*” iterativamente criará uma rede com função de base radial um neurônio de cada vez. Nesta função, neurônios são adicionados à rede até que o MSE chegue a um valor inferior ao objetivo (parâmetro da função) ou então que o número de iterações seja atingido. A “chamada” para esta função é dada, portanto, por  $net = newrb(P, T, GOAL, SPREAD)$ , onde “*net*” representa a rede que está sendo criada, e os parâmetros da função são: “P”, o vetor de entrada, ou seja, as informações que o algoritmo deverá analisar, “T” o vetor de respostas para os valores de entrada, “*goal*” deverá apresentar o erro mínimo admitido pelo usuário, e “*spread*” que representa a propagação da camada de bases radiais. Quando desejado realizar a simulação para a previsão das respostas do algoritmo treinado, utiliza-se a função  $simulation = sim(net, test\_input)$ , que deverá realizar a simulação utilizando a rede treinada “*net*”

com “*test\_input*”, que são os valores de entrada das variáveis separadas para teste (THE MATHWORKS INC, 2017b).

Para utilizar uma RNA, seja do tipo MLP ou RBF, no MATLAB é necessário inicialmente carregar a base de dados no formato “*mat*” em seguida é possível iniciar a criação da rede chamando a função *feedforwardnet* com o nome da rede a ser elaborada.

#### 4.5 REGRESSÃO LOGÍSTICA COM SPSS

A RL permite a suposição e o teste de modelos que prevejam resultados de categorização, sejam com duas (binomial) ou mais categorias (polinomial). O *software* SPSS (*Statistical Package for Social Sciences*) foi escolhido para ser utilizado na aplicação da técnica de RL, este *software* é utilizado para solucionar uma variedade de problemas tanto em negócios quanto em pesquisas. Possui um amplo agrupamento de recursos, incluindo diferentes análises, testes de hipóteses e relatórios, com o objetivo de facilitar o gerenciamento de dados, a seleção e execução de análises e, enfim, o compartilhamento de seus resultados (IBM SPSS *Statistics*, 2016).

Segundo Field (2009), o SPSS utiliza principalmente duas janelas, *data editor* – o editor de dados, onde os dados entram e são executadas funções estatísticas – e o *syntax editor* – editor de sintaxe, permite que os comandos sejam colocados manualmente.

A RL é um modelo que prevê a probabilidade de um dado evento ocorrer em certa situação, baseado em observações semelhantes já conhecidas. Para o SPSS, existem dois diferentes métodos de RL, sendo o primeiro o de entrada forçada (onde todas as covariáveis são alocadas em um único bloco e as estimativas dos parâmetros são calculadas para cada bloco separadamente); e o método *stepwise* (passo a passo). Particularmente, sobre o *stepwise*, quando este é empregado, o computador deverá iniciar o modelo incluindo apenas uma constante e, em seguida, irá adicionar os previsores um a um com base em um critério denominado *score*. Esse critério faz com que a variável mais significativa, seja adicionada ao modelo, e este processo segue até que nenhum predictor que restar tenha um *score* significativo (FIELD, 2009).

De forma a executar uma RL no SPSS, de acordo com Field (2009), a entrada dos dados deve, no editor de dados, ser arranjada em uma coluna para cada variável, incluindo as respostas de cada instância adicionada. Feita a entrada dos dados, o processo se resume basicamente às seleções: *Analyze* (menu de análises do SPSS), *Regression* (opções de

regressões realizadas pelo *software*) e, por fim, *Binary Logistic* (que se traduz em Logística Binária).

#### 4.6 PARÂMETROS DOS MÉTODOS APLICADOS

Tendo definidos os respectivos *softwares* para cada metodologia a ser testada por esta pesquisa, agora são apresentados, de forma mais especificada, os parâmetros para o funcionamento de cada modelo. Ao final da seção demonstram-se os métodos de avaliação de desempenho aplicados aos modelos testados.

Como mencionado anteriormente, para as redes MLP, foram testadas diferentes quantidades de neurônios na camada oculta, quantidades essas variadas de 1 a 20. Cada arquitetura será testada para 30 conjuntos de pesos randômicos diferentes e para o conjunto *default* do *software* utilizado. Em relação à função de ativação, escolheu-se a tangente-sigmoidal, tanto da camada de entrada para a camada oculta quanto da camada oculta para a camada de saída. Os dados, conforme já foi mencionado, foram divididos em um conjunto com 2/3 das instâncias para treinamento e 1/3 para a validação da rede. A seguir, no Quadro 9, é possível observar um resumo das informações apresentadas.

<b>Quadro 9 - Parâmetros da MLP utilizada</b>	
<i>PARÂMETROS</i>	<i>MLP</i>
Total de Instâncias	5432
Instâncias utilizadas no Treinamento	3639
Instâncias utilizadas no Teste	1793
Treinamento	Supervisionado
Quantidade de Conjuntos de Pesos Sinápticos testados	31 conjuntos testados
Número de Camadas	3 camadas
Variação dos Neurônios na Camada Oculta	1 - 20 neurônios
Algoritmo de Treinamento	Levenberg-Marquardt
Função de Ativação camada de entrada-camada oculta	Tangente Sigmoidal
Função de Ativação camada oculta-camada de saída	Tangente Sigmoidal

Fonte: A autora, 2017.

Com relação à RBF, as instâncias também foram divididas em 2/3 para treinamento da rede e 1/3 para a sua avaliação e, de forma análoga às MLP, foram utilizadas 3 camadas (a camada de entrada da rede, a camada oculta e a camada de saída). É possível observar no Quadro 10 um resumo dos parâmetros utilizados para sua execução nesta pesquisa.

**Quadro 10 - Parâmetros da RBF utilizada**

<i>INFORMAÇÕES</i>	<i>RBF</i>
Total de Instâncias	5432
Instâncias utilizadas no Treinamento	3639
Instâncias utilizadas no Teste	1793
Número de Camadas	3 camadas
Numero Máximo de Neurônios na Camada Oculta	3639
<i>Spread</i> utilizado	1

Fonte: A autora, 2017.

Finalmente, ao utilizar o SPSS, foi utilizado o teste Hosmer-Lemeshow *goodness-of-fit* para verificar se o modelo é considerado um bom ajuste para os dados (caso o valor  $p$  seja superior à significância de 0,05). Outro indicador escolhido foi o intervalo de confiança para  $Exp(B)$ , que quantifica alterações na probabilidade da classificação caso ocorram mudanças na variável explicativa associada. Caso, por exemplo  $Exp(B) = 1$ , isso corresponde a nenhuma alteração nas chances após uma alteração de uma unidade em um preditor, um valor de “2” significa que as chances de alteração no resultado são duas vezes maiores e “0,2” significa que as chances podem diminuir em 80% (RESTORE, 2011).

#### 4.6.1 Método de Avaliação de Desempenho

Foram considerados dois métodos de avaliação de desempenho para os modelos testados (acurácia): 1) através da matriz de confusão gerada pelos softwares utilizados; 2) através do MSE (*Mean Square Error*).

Uma vez que se trata de uma aplicação onde as situações testadas são “um contra todos”, por exemplo a classificação do método testado para a situação dos Adimplentes

será: Adimplente ou Outra classe (onde esta poderá então ser Inadimplente ou Temporariamente Inadimplente), as métricas podem ser analisadas a partir de uma matriz de confusão 2x2, dada no Quadro 11. Para esta matriz, cada entrada  $(i, j)$  contém o número de casos corretos/incorretos, para exemplificar, é considerado positivo o cliente adimplente e negativo o que for categorizado na classificação “outros”.

**Quadro 11 - Matriz de confusão**

	<i>Positivo Previsto</i>	<i>Negativo Previsto</i>
Positivo Real	Positivo Verdadeiro (PV)	Negativo Falso (NF)
Negativo Real	Positivo Falso (PF)	Negativo Verdadeiro (NR)

Fonte: A autora, 2017.

Tratando-se de um problema binário o critério para o cálculo da taxa de acurácia é dado pela equação (21) a seguir.

$$Acurácia = \frac{\textit{sensitividade} + \textit{especificidade}}{2} \quad (21)$$

Onde a sensibilidade é dada por  $\frac{PV}{(PV+NF)}$  e mede a porcentagem de exemplos positivos que foram classificados corretamente, enquanto isso a especificidade é dada por  $\frac{NR}{(NR+PF)}$  que corresponde à porcentagem de exemplos negativos que foram corretamente previstos (MARQUÉS, GARCIA e SÁNCHEZ, 2013).

Especificamente para as RNA testadas, após executar a simulação com os parâmetros apresentados o Matlab 2017a apresenta os resultados de forma gráfica e numérica calculada pelo MSE, dado pela equação 22 abaixo.

$$MSE = \frac{1}{N} * \sum_{i=1}^N (d_i - y_i)^2 \quad (21)$$

Onde,  $N$  é o valor de instâncias (exemplos) testados,  $d_i$  representa o valor esperado, ou seja, a resposta original da instância  $i$  testada e  $y_i$  é o valor encontrado pela rede.

## 5 OBTENÇÃO E ANÁLISE DOS RESULTADOS

Retomando os objetivos específicos apresentados no Capítulo 1 deste trabalho, inicialmente foram descritas as características de um problema de concessão de crédito bancário ao decorrer do Capítulo 3, ao longo da seção 3.1 e dos trabalhos correlatos com esta aplicação. O segundo objetivo tratava da proposição de quais modelos deveriam ser testados pelos métodos testados, sendo adimplentes ou outra classificação, inadimplentes ou outra e temporariamente inadimplentes e outra, este foi cumprido no Capítulo 4, na seção de metodologia aplicada.

Os últimos dois objetivos são verificados neste Capítulo, sendo eles: Gerar diferentes topologias para as RNA testadas, tendo como finalidade a obtenção do número mais adequado de neurônios na camada oculta para a resolução do problema em questão; E comparar as soluções obtidas em termos de acurácia e erro dos métodos avaliados em relação aos seus conjuntos de teste. Suas resoluções serão apresentadas nas seções seguintes.

### 5.1 RESULTADOS MLP

Na Tabela 5, a seguir, são observados os resultados obtidos pelas MLP testadas. Para a Tabela 5, a coluna “conjunto de pesos” identifica qual foi o conjunto dos testados que apresentou o resultado em questão, a “neurônios” indica o número de neurônios na camada oculta da simulação com melhor resultado. A coluna MSE refere-se ao erro quadrático médio apresentado por tal arquitetura e a última coluna, acurácia, apresenta a taxa de acerto do teste da RNA.

**Tabela 5 - Resultados Obtidos pelas MLP para cada classe testada**

<i>CLASSIFICAÇÃO</i>	<i>CONJUNTO DE PESOS</i>	<i>NEURÔNIOS</i>	<i>MSE</i>	<i>ACURÁCIA</i>
ADIMPLENTES OU OUTRA	18	20	0,19260	74,70%
INADIMPLENTES OU OUTRA	13	13	0,07884	91,40%
	26	14	0,078782	91,40%
TEMPORARIAMENTE INADIMPLENTES OU OUTRA	28	13	0,087617	91,40%
	15	19	0,18825	74,60%

Fonte: A autora, 2017.

As seções a seguir apresentam de forma mais detalhada todos os testes realizados com os parâmetros mencionados no Capítulo 4.

### **5.1.1 Situação: Adimplentes ou Outra**

Já a Tabela 6 apresenta os resultados para a MLP utilizando o modelo “Adimplente ou Outra”. Cada coluna da tabela se refere ao número de neurônios presente na camada oculta, variando de 1 a 20 neurônios na camada oculta, e cada linha representa o conjunto de peso utilizado para sua execução, variando, desta forma, de 1 a 30, mais o conjunto *default* do *software*.

Na tabela em questão, encontramos três principais situações: valores em tons de verde que representam acurácias próximas à melhor encontrada, valores em tons amarelados, que são taxas entre os superiores e inferiores e, por último, valores em tons avermelhados, que são as menores acurácias encontradas para esta aplicação.

Observa-se que, para a situação apresentada na Tabela 6, existe uma maior frequência de valores em tons de verde a partir dos testes realizados com mais de 10 neurônios na camada oculta, mas, também, há uma frequência superior de valores baixos de acurácia na área em questão. A maior taxa de acurácia encontrada ocorre para o conjunto de pesos 18 com 20 neurônios na camada oculta, conforme mencionado na Tabela 5, sendo seu valor de 74,70%. Já a menor taxa encontrada ocorre para o conjunto de pesos 13 com 19 neurônios na camada oculta, sendo a acurácia de 67,90%.

Tabela 6 - Acurácia dos Testes com MLP para o modelo “Adimplentes ou Outra”

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
30	71,50%	71,50%	71,20%	72,10%	71,10%	71,00%	71,30%	71,70%	72,10%	71,00%	72,70%	71,10%	72,80%	71,60%	72,10%	72,20%	72,60%	72,00%	72,80%	72,50%
29	71,40%	70,70%	70,60%	71,20%	71,40%	73,00%	70,90%	72,70%	72,10%	71,70%	71,90%	71,80%	73,50%	72,20%	71,10%	71,90%	70,80%	71,10%	71,30%	71,70%
28	71,00%	71,00%	70,60%	71,20%	71,70%	69,60%	72,70%	71,00%	71,90%	71,60%	71,20%	72,60%	71,90%	71,30%	71,60%	71,50%	73,10%	72,50%	72,30%	73,60%
27	71,40%	71,00%	71,90%	71,60%	72,10%	71,00%	71,60%	71,90%	70,90%	71,50%	70,90%	71,50%	72,40%	71,70%	72,60%	70,80%	73,40%	72,00%	71,70%	73,00%
26	71,60%	71,60%	71,90%	70,90%	71,50%	71,40%	72,10%	72,10%	71,90%	71,50%	71,70%	72,50%	72,60%	71,70%	72,40%	71,40%	72,10%	72,10%	70,40%	70,90%
25	71,20%	71,10%	70,90%	71,90%	70,20%	71,80%	70,00%	72,10%	70,00%	71,90%	73,40%	71,90%	69,60%	69,40%	72,60%	70,00%	72,30%	70,40%	71,30%	71,70%
24	71,20%	71,60%	72,00%	71,80%	71,00%	72,10%	72,20%	69,90%	72,40%	70,50%	71,10%	73,10%	70,50%	73,30%	72,10%	71,00%	72,60%	72,00%	70,00%	71,30%
23	71,60%	71,40%	71,00%	71,40%	71,60%	72,60%	71,00%	72,50%	71,80%	71,40%	71,20%	73,00%	72,00%	72,40%	71,90%	72,30%	71,10%	72,90%	72,80%	70,80%
22	71,00%	70,80%	71,90%	71,40%	70,90%	71,30%	70,60%	71,60%	71,40%	72,10%	70,80%	73,60%	73,30%	71,60%	71,00%	71,60%	70,90%	73,70%	72,30%	73,70%
21	71,40%	70,80%	71,00%	71,60%	71,30%	71,40%	72,30%	71,90%	71,70%	71,70%	72,80%	71,10%	71,10%	71,10%	72,10%	71,50%	72,10%	68,20%	73,40%	72,10%
20	71,20%	71,20%	71,50%	71,20%	71,60%	71,40%	71,70%	71,30%	71,70%	72,50%	72,70%	71,60%	70,50%	72,60%	71,50%	71,40%	72,50%	72,90%	71,40%	71,80%
19	71,30%	71,60%	70,80%	71,20%	71,80%	71,00%	71,20%	71,60%	72,10%	71,00%	72,00%	71,80%	72,00%	69,70%	72,20%	72,70%	71,90%	71,30%	73,80%	71,70%
18	71,20%	71,40%	70,30%	71,90%	71,40%	72,10%	71,20%	71,00%	71,00%	71,30%	71,40%	71,40%	71,50%	71,80%	72,30%	71,90%	71,50%	73,10%	71,20%	74,70%
17	71,30%	71,50%	71,80%	70,70%	72,00%	72,20%	71,40%	71,60%	71,80%	71,30%	71,30%	71,50%	72,10%	72,50%	73,70%	72,30%	70,70%	71,80%	71,40%	74,00%
16	71,20%	71,10%	71,00%	71,00%	71,90%	72,20%	71,70%	71,00%	71,70%	71,40%	71,80%	71,60%	70,50%	71,10%	72,40%	72,60%	71,90%	72,10%	70,70%	71,60%
15	71,20%	71,30%	71,70%	72,00%	71,40%	71,50%	73,20%	72,80%	71,40%	72,30%	72,10%	72,40%	70,50%	69,50%	72,50%	71,10%	72,30%	73,10%	73,30%	72,20%
14	71,30%	71,50%	69,70%	71,50%	71,70%	71,10%	71,20%	71,90%	71,70%	72,10%	72,10%	72,50%	73,00%	70,80%	72,10%	72,10%	71,10%	69,10%	70,90%	72,10%
13	71,00%	71,60%	72,00%	71,60%	71,00%	71,40%	71,20%	71,30%	72,30%	71,30%	71,20%	73,60%	71,10%	71,10%	72,80%	71,60%	71,00%	71,90%	67,90%	69,10%
12	71,40%	71,40%	71,20%	70,10%	71,60%	71,70%	72,00%	71,40%	71,90%	71,50%	70,30%	69,90%	71,80%	69,80%	71,30%	69,90%	71,10%	72,20%	70,20%	69,60%
11	71,00%	70,10%	71,30%	72,50%	71,90%	72,20%	70,80%	71,70%	71,90%	71,90%	71,40%	72,20%	69,70%	71,10%	72,80%	71,20%	71,00%	71,50%	71,10%	71,70%
10	71,20%	71,40%	71,70%	71,90%	71,00%	71,00%	71,60%	71,30%	71,90%	70,90%	72,00%	71,10%	71,60%	73,30%	71,50%	71,30%	72,60%	72,60%	70,40%	70,70%
9	71,20%	71,30%	71,40%	71,00%	71,00%	71,40%	71,30%	71,00%	71,00%	72,30%	71,60%	70,90%	71,90%	71,90%	71,40%	70,50%	73,30%	73,00%	69,60%	73,50%
8	71,20%	71,60%	71,70%	71,10%	71,20%	71,40%	73,20%	70,90%	72,50%	71,30%	70,60%	71,50%	71,10%	72,70%	71,70%	74,00%	71,60%	72,00%	73,50%	72,10%
7	71,50%	70,80%	71,50%	71,40%	72,00%	70,80%	72,20%	70,70%	70,80%	72,20%	71,60%	72,60%	71,50%	71,80%	71,40%	72,40%	70,50%	71,50%	71,50%	71,10%
6	71,20%	72,70%	70,50%	71,00%	71,90%	72,50%	71,70%	72,80%	71,00%	71,90%	71,80%	71,80%	71,30%	68,70%	71,80%	71,30%	71,80%	70,70%	71,00%	69,30%
5	71,20%	71,00%	71,80%	72,50%	71,80%	72,00%	72,70%	72,20%	72,10%	72,00%	72,80%	72,10%	73,00%	73,10%	71,20%	70,30%	72,30%	72,10%	73,00%	72,50%
4	71,30%	71,90%	71,30%	71,60%	72,00%	72,20%	71,90%	72,60%	72,00%	72,40%	71,60%	71,80%	70,10%	70,40%	71,40%	73,00%	72,80%	71,90%	71,40%	69,00%
3	71,30%	71,90%	71,30%	71,60%	72,00%	72,20%	71,90%	72,60%	72,00%	72,40%	71,60%	71,80%	70,10%	70,40%	71,40%	73,00%	72,80%	71,90%	71,40%	69,00%
2	71,10%	71,30%	71,40%	70,80%	71,50%	72,40%	72,40%	70,50%	71,40%	71,80%	71,50%	71,10%	71,00%	71,00%	70,70%	71,80%	71,90%	73,40%	71,10%	73,00%
1	71,10%	70,70%	72,20%	70,20%	71,40%	72,30%	71,50%	72,80%	73,10%	71,10%	72,60%	71,70%	72,10%	72,70%	70,70%	71,30%	71,90%	71,90%	71,60%	72,20%
Default	71,10%	70,10%	71,60%	72,10%	71,50%	71,80%	71,80%	72,10%	71,50%	69,80%	71,70%	71,40%	70,60%	73,80%	71,40%	71,20%	73,10%	71,40%	73,90%	72,20%

Neurônios na camada oculta

Fonte: A autora, 2017.

### 5.1.2 Situação: Inadimplentes ou Outra

A Tabela 7 a seguir apresenta os resultados para a MLP utilizando o modelo “Inadimplente ou Outra”. Cada coluna da tabela se refere ao número de neurônios presente na camada oculta, variando de 1 a 20 neurônios na camada oculta, e cada linha representa o conjunto de peso utilizado para sua execução, variando, desta forma, de 1 a 30, mais o conjunto *default* do *software*.

De forma análoga à seção anterior, encontramos três principais situações: valores em tons de verde que representam acurácias próximas à melhor encontrada, valores em tons amarelados, que são taxas consideradas médias, por último, valores em tons avermelhados, que são as menores acurácias.

Diferentemente dos resultados encontrados na Tabela 6, nota-se uma frequência superior de valores amarelados e esverdeados, e poucos em tons de vermelho, ou seja, poucos testes tiveram acurácias consideradas baixas para a situação apresentada, e, apesar do amarelo ser predominante, os demais testes se dividiram nos tons de verde e amarelo. Da mesma forma que a situação apresentada anteriormente, percebe-se uma frequência maior dos valores esverdeados, ou seja, acurácias próximas à maior encontrada, a partir de 10 neurônios na camada oculta. Nota-se também que os valores inferiores tiveram maior presença com poucos neurônios na camada oculta.

Conforme apresentado na Tabela 5, o melhor resultado para o modelo dos “Inadimplentes ou Outra”, com o valor de 91,4%, ocorre para os conjuntos de pesos 13, 26 e 28, onde os neurônios poderiam ser 13 (para os conjuntos 13 e 28) ou 14 (para o conjunto 26). Para a pior taxa de acurácia encontrada, observa-se que seu valor de 76,40% está presente para os conjuntos: *default*, 5, 10, 18, 19 e 23, cujos neurônios na camada oculta são, 5 (conjunto *default*), 2 (conjuntos 5, 10, 19 e 23), e 3 (conjuntos 18 e 23).

Tabela 7 - Acurácia dos Testes com MLP para o modelo “Inadimplentes ou Outra”

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
30	88,80%	88,60%	88,70%	88,20%	88,80%	88,20%	89,30%	89,20%	89,80%	89,00%	90,40%	89,80%	89,80%	89,30%	88,90%	90,10%	88,50%	89,80%	90,10%	90,10%
29	88,90%	87,50%	89,00%	89,60%	88,50%	90,60%	89,40%	90,30%	89,50%	90,10%	88,50%	89,50%	89,50%	87,40%	90,40%	90,20%	89,10%	89,30%	87,20%	89,30%
28	88,10%	88,80%	88,60%	88,40%	88,90%	88,70%	89,90%	88,90%	89,10%	89,20%	89,80%	88,80%	88,80%	89,30%	90,10%	91,30%	90,70%	89,70%	89,60%	91,20%
27	88,70%	89,60%	88,80%	89,90%	89,40%	89,30%	89,50%	89,10%	89,10%	90,90%	90,30%	89,20%	89,20%	89,70%	88,50%	89,70%	90,20%	88,70%	88,90%	91,10%
26	88,30%	88,80%	88,40%	88,70%	88,70%	89,10%	89,20%	89,40%	89,20%	89,40%	90,70%	89,00%	89,00%	91,40%	90,00%	90,00%	90,00%	88,60%	89,20%	90,40%
25	88,70%	89,00%	88,90%	88,60%	88,40%	88,90%	88,60%	89,20%	89,20%	88,20%	90,00%	89,40%	89,40%	90,10%	90,50%	90,50%	89,70%	90,50%	90,50%	89,00%
24	88,50%	88,20%	88,40%	89,10%	88,90%	88,20%	89,10%	90,00%	88,70%	89,50%	89,20%	89,40%	90,00%	88,20%	87,90%	89,30%	89,90%	89,20%	89,60%	89,40%
23	88,60%	76,40%	76,40%	89,20%	88,80%	89,10%	88,90%	88,80%	89,40%	89,70%	89,80%	89,10%	89,50%	88,80%	90,00%	89,50%	88,60%	89,80%	90,20%	87,70%
22	88,70%	88,80%	88,30%	89,10%	88,80%	89,90%	88,70%	90,80%	89,00%	88,40%	89,00%	89,00%	89,60%	88,50%	89,10%	89,20%	91,10%	89,30%	89,40%	89,80%
21	88,70%	88,80%	88,60%	89,30%	88,80%	89,10%	89,20%	88,90%	89,90%	89,40%	89,10%	89,20%	89,60%	90,40%	90,10%	88,50%	89,70%	89,80%	89,50%	89,30%
20	88,90%	88,10%	88,60%	90,10%	88,70%	88,90%	90,30%	89,20%	89,00%	89,20%	89,70%	89,30%	88,50%	89,10%	89,30%	89,40%	89,80%	89,80%	87,90%	90,30%
19	88,60%	76,40%	89,30%	88,80%	89,70%	88,80%	89,10%	89,90%	90,40%	88,90%	89,10%	89,80%	88,90%	89,10%	89,00%	90,40%	87,60%	89,30%	90,20%	88,70%
18	88,40%	89,90%	76,40%	88,60%	89,30%	90,80%	89,30%	89,00%	89,90%	89,10%	89,50%	89,00%	89,70%	90,20%	90,50%	90,00%	89,20%	89,00%	89,40%	89,60%
17	88,80%	88,00%	88,90%	88,40%	88,60%	89,60%	89,10%	89,20%	89,70%	89,30%	89,10%	87,50%	88,30%	87,50%	90,60%	89,40%	89,70%	89,90%	89,90%	88,40%
16	88,20%	88,50%	89,10%	88,20%	88,70%	89,00%	89,20%	88,30%	89,00%	88,70%	88,20%	89,30%	89,50%	89,60%	89,20%	88,70%	88,80%	89,60%	89,50%	89,90%
15	88,50%	88,90%	89,40%	89,10%	88,50%	88,90%	88,70%	89,00%	89,80%	89,20%	89,60%	90,00%	88,80%	90,50%	88,70%	89,90%	90,50%	90,60%	90,20%	89,60%
14	88,90%	88,30%	89,10%	89,10%	89,40%	88,20%	87,90%	89,50%	88,90%	89,20%	89,60%	89,00%	89,50%	90,20%	89,40%	89,70%	89,50%	90,10%	88,90%	89,50%
13	88,50%	89,10%	88,90%	89,10%	88,50%	86,80%	88,80%	90,10%	90,20%	89,90%	88,40%	89,00%	89,00%	88,80%	89,20%	89,50%	89,40%	90,90%	89,90%	90,20%
12	88,60%	88,60%	89,60%	88,30%	88,60%	88,90%	89,90%	89,50%	89,50%	88,80%	89,30%	89,10%	89,00%	89,00%	90,00%	89,90%	90,40%	90,50%	88,90%	89,30%
11	88,50%	88,30%	89,60%	88,80%	88,90%	89,40%	88,50%	89,00%	90,30%	88,60%	89,50%	90,10%	90,10%	88,30%	91,30%	88,20%	89,50%	88,50%	89,80%	89,30%
10	88,50%	76,40%	88,50%	88,70%	88,50%	90,00%	88,10%	88,40%	89,10%	88,80%	90,40%	88,80%	89,90%	90,00%	89,20%	90,50%	89,80%	90,40%	89,30%	89,80%
9	88,60%	88,90%	89,50%	88,30%	88,60%	89,30%	88,80%	88,80%	90,00%	90,30%	89,00%	88,50%	88,80%	89,40%	90,10%	89,10%	88,10%	89,90%	89,20%	89,20%
8	88,50%	88,10%	88,80%	89,50%	89,00%	89,20%	88,80%	88,90%	88,60%	90,30%	88,70%	90,20%	89,50%	89,00%	90,90%	90,00%	89,40%	89,50%	89,90%	89,70%
7	88,50%	88,70%	89,20%	88,70%	89,30%	89,50%	88,80%	89,90%	89,50%	89,30%	89,60%	88,90%	89,10%	90,90%	89,40%	90,20%	91,30%	90,30%	89,60%	88,90%
6	88,40%	89,00%	88,40%	88,50%	89,50%	89,20%	89,20%	88,90%	90,10%	88,60%	88,50%	88,80%	89,40%	90,40%	88,80%	89,70%	89,70%	90,80%	89,70%	90,30%
5	88,50%	76,40%	88,70%	89,40%	89,20%	88,60%	89,30%	89,00%	88,90%	89,50%	89,10%	90,10%	90,10%	89,30%	88,20%	90,40%	90,80%	90,90%	88,70%	89,30%
4	88,60%	88,60%	88,60%	88,50%	89,90%	88,80%	90,70%	89,50%	89,80%	89,80%	90,10%	89,60%	89,60%	89,80%	88,40%	89,40%	89,70%	89,60%	88,70%	89,70%
3	88,60%	88,60%	88,60%	88,50%	89,80%	88,80%	90,70%	89,50%	89,80%	89,80%	90,10%	89,60%	87,60%	89,80%	88,40%	89,40%	89,70%	89,60%	88,70%	89,70%
2	88,50%	89,20%	88,70%	88,70%	89,20%	90,00%	90,10%	88,50%	88,80%	89,30%	88,80%	89,50%	90,10%	89,90%	90,60%	89,20%	89,70%	88,90%	88,70%	90,20%
1	88,80%	89,30%	89,00%	88,70%	89,10%	88,70%	88,60%	89,70%	89,90%	88,90%	88,80%	88,70%	88,70%	89,40%	88,70%	89,20%	89,10%	87,90%	88,70%	88,90%
Default	88,80%	88,50%	89,40%	88,90%	76,40%	89,40%	89,70%	88,90%	89,70%	89,20%	89,50%	89,20%	89,20%	89,60%	89,70%	91,00%	89,20%	89,80%	90,30%	89,60%

Conjuntos de Pesos

Neurônios na camada oculta

Fonte: A autora, 2017.

### 5.1.3 Situação: Temporariamente Inadimplentes ou Outra

A Tabela 8 a seguir apresenta os resultados para a MLP utilizando o modelo “Inadimplente ou Outra”. Cada coluna da tabela se refere ao número de neurônios presente na camada oculta, variando de 1 a 20 neurônios na camada oculta, e cada linha representa o conjunto de peso utilizado para sua execução, variando, desta forma, de 1 a 30, mais o conjunto *default* do *software*.

De forma análoga à seção anterior, encontramos três principais situações: valores em tons de verde que representam acurácias próximas à melhor encontrada, valores em tons amarelados, que são taxas consideradas médias, por último, valores em tons avermelhados, que são as menores acurácias.

Diferentemente dos modelos anteriormente apresentados (“Adimplentes ou Outra” e “Inadimplentes ou Outra”), para a Tabela 8, resultados dos “Temporariamente Inadimplentes ou Outra”, não é possível distinguir facilmente um padrão de resultados de acordo com os parâmetros testados, mas, na situação geral, há ocorrência de mais valores avermelhados que os demais modelos testados, que tinham predominância da cor amarela, indicando que os resultados da tabela abaixo ficaram mais próximos, em geral, da menor acurácia encontrada.

É possível verificar que o melhor resultado encontrado é de acurácia igual a 74,60% para o conjunto de pesos 15, com 19 neurônios na camada oculta. A menor acurácia encontrada foi de 70,00% para o conjunto de pesos “1” com 18 neurônios na camada oculta.

Tabela 8 - Acurácia dos Testes com MLP para o modelo "Temporariamente Inadimplentes ou Outra"

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
30	71,70%	71,50%	72,30%	73,60%	72,10%	72,20%	71,90%	71,80%	71,90%	72,60%	71,80%	71,40%	71,70%	72,00%	72,20%	72,10%	73,40%	72,10%	72,10%	72,90%
29	71,40%	71,40%	71,80%	71,90%	71,90%	72,60%	71,90%	72,20%	72,00%	71,90%	72,10%	71,60%	72,10%	72,80%	71,90%	71,90%	71,60%	72,70%	72,40%	73,10%
28	71,40%	71,90%	71,80%	72,10%	71,30%	71,40%	71,60%	72,50%	71,20%	71,90%	72,10%	72,00%	72,80%	71,80%	71,80%	71,90%	72,10%	74,50%	71,50%	71,80%
27	71,60%	71,60%	71,80%	71,80%	71,50%	71,60%	72,60%	71,90%	72,70%	71,50%	72,40%	72,40%	71,90%	71,90%	72,00%	72,20%	71,70%	73,50%	73,60%	71,70%
26	71,50%	71,00%	72,10%	72,30%	71,70%	72,80%	71,90%	72,20%	72,00%	71,90%	71,10%	71,90%	72,50%	72,20%	72,10%	72,50%	71,70%	72,30%	72,30%	72,80%
25	71,40%	71,40%	71,60%	71,80%	72,00%	72,30%	71,70%	72,70%	71,90%	72,10%	73,10%	73,10%	72,60%	72,10%	72,20%	71,50%	72,60%	72,70%	72,60%	71,70%
24	72,10%	72,20%	71,40%	71,40%	71,80%	71,70%	71,50%	71,30%	71,70%	72,20%	72,10%	72,00%	71,90%	72,80%	72,20%	71,50%	72,20%	71,90%	73,00%	72,80%
23	71,40%	71,40%	71,40%	71,70%	71,00%	72,70%	71,80%	73,00%	71,40%	71,60%	71,50%	71,90%	71,60%	72,10%	72,10%	73,30%	71,80%	72,20%	71,80%	71,30%
22	71,40%	71,60%	71,80%	71,30%	72,10%	71,90%	71,70%	71,40%	72,10%	71,70%	72,20%	71,90%	72,10%	72,10%	71,80%	71,60%	72,80%	71,60%	72,60%	71,30%
21	72,00%	71,40%	71,50%	72,10%	72,50%	72,10%	71,50%	71,40%	72,10%	72,00%	72,30%	71,80%	71,90%	71,40%	73,10%	72,60%	72,40%	72,00%	73,30%	72,60%
20	71,30%	71,40%	72,40%	72,80%	72,30%	71,70%	72,40%	71,40%	71,60%	71,80%	72,20%	72,10%	72,20%	72,80%	73,20%	72,80%	72,60%	72,60%	71,50%	72,50%
19	71,90%	71,40%	72,20%	71,40%	71,80%	71,40%	71,60%	71,70%	71,90%	72,00%	72,30%	72,60%	72,60%	73,00%	72,50%	72,40%	73,20%	72,30%	72,50%	72,30%
18	71,40%	72,00%	71,50%	71,40%	71,80%	72,10%	71,40%	71,80%	72,00%	71,90%	72,60%	72,30%	71,90%	74,20%	72,10%	72,30%	72,10%	72,00%	72,80%	71,30%
17	71,40%	72,10%	71,50%	71,60%	71,90%	72,00%	72,60%	72,60%	72,30%	71,70%	72,30%	72,10%	72,60%	73,20%	71,70%	74,00%	72,60%	72,10%	71,30%	71,60%
16	71,40%	72,90%	72,40%	72,30%	72,00%	71,50%	72,10%	71,40%	73,20%	71,30%	72,40%	71,30%	72,20%	71,60%	71,40%	71,50%	71,50%	72,20%	71,90%	71,40%
15	71,40%	72,60%	71,60%	71,40%	72,00%	72,00%	72,20%	72,10%	71,50%	72,40%	73,10%	73,20%	72,30%	71,50%	71,60%	71,20%	73,00%	71,90%	74,60%	72,40%
14	71,50%	71,40%	71,70%	71,40%	71,50%	71,70%	72,00%	71,50%	71,80%	71,40%	72,80%	71,80%	71,70%	72,60%	72,00%	72,20%	71,80%	71,80%	73,30%	72,60%
13	71,60%	71,50%	72,20%	72,20%	73,10%	72,40%	72,40%	71,80%	71,50%	72,20%	72,70%	72,10%	71,70%	72,20%	72,80%	71,90%	72,10%	73,10%	72,60%	72,10%
12	71,70%	71,60%	71,70%	71,70%	71,80%	72,40%	72,30%	72,90%	72,20%	72,20%	72,80%	71,70%	71,30%	72,50%	70,70%	71,50%	71,80%	73,60%	72,40%	71,90%
11	71,50%	71,80%	71,50%	72,00%	72,00%	71,80%	72,00%	72,70%	71,60%	71,80%	72,20%	71,70%	72,50%	71,50%	72,70%	73,30%	71,90%	72,10%	72,70%	72,70%
10	71,40%	71,30%	71,80%	70,90%	72,10%	72,00%	72,10%	71,40%	72,90%	71,40%	71,40%	71,40%	72,20%	72,20%	73,50%	72,10%	74,00%	70,90%	72,30%	71,50%
9	71,40%	72,10%	71,60%	71,70%	72,00%	71,90%	72,70%	72,10%	71,70%	72,30%	71,70%	71,60%	74,00%	72,30%	71,70%	71,30%	73,10%	71,60%	71,90%	73,00%
8	71,40%	71,40%	71,90%	71,50%	71,90%	71,40%	72,50%	71,90%	71,90%	71,90%	71,80%	71,40%	72,00%	72,40%	72,10%	73,90%	73,10%	72,20%	71,40%	73,70%
7	71,90%	71,40%	72,40%	71,10%	71,60%	71,90%	71,60%	71,90%	72,10%	73,00%	72,20%	71,70%	72,10%	72,60%	71,60%	72,30%	72,50%	71,50%	71,50%	72,30%
6	71,90%	71,20%	71,30%	71,80%	71,50%	73,40%	71,70%	72,10%	72,40%	71,40%	71,60%	72,20%	72,10%	72,60%	71,60%	72,50%	70,80%	73,00%	72,30%	71,80%
5	71,40%	71,40%	72,00%	71,20%	71,20%	72,50%	72,40%	72,10%	72,00%	72,00%	72,00%	72,10%	71,50%	72,10%	71,90%	71,40%	72,30%	72,70%	72,30%	72,00%
4	72,10%	71,90%	71,70%	71,90%	72,30%	72,20%	72,20%	71,90%	71,80%	71,60%	72,10%	71,70%	72,60%	72,30%	73,50%	71,20%	72,00%	72,10%	72,30%	71,90%
3	72,10%	71,90%	71,70%	71,90%	72,30%	72,30%	72,20%	71,90%	71,80%	71,60%	72,10%	71,70%	72,60%	72,30%	73,50%	71,20%	72,00%	72,10%	72,70%	71,90%
2	71,40%	71,60%	71,90%	71,20%	71,40%	73,00%	71,80%	71,80%	72,10%	72,30%	72,00%	73,20%	71,60%	72,40%	71,80%	72,20%	71,80%	72,60%	73,20%	72,20%
1	71,40%	71,50%	72,00%	71,60%	71,60%	72,30%	71,80%	71,60%	72,50%	71,80%	72,20%	72,20%	72,50%	72,10%	72,10%	72,50%	73,10%	70,00%	72,00%	72,40%
Default	71,60%	71,40%	71,70%	72,10%	72,30%	72,30%	71,50%	71,60%	72,10%	72,20%	72,30%	72,70%	72,00%	72,30%	72,40%	72,10%	72,90%	71,60%	71,90%	72,00%

Neurônios na camada oculta

Fonte: A autora, 2017.

## 5.2 RESULTADOS RBF

O resultado apresentado a seguir, Tabela 9, referem-se às RBF testadas, onde foram feitas duas análises, de média das acurácias e de desvio-padrão dos valores de neurônios testados.

**Tabela 9 - Resultados Obtidos pelas RBF para as três classes utilizadas**

CLASSIFICAÇÃO	ACURÁCIA		MSE	
	MÉDIA	DESVIO-PADRÃO	MÉDIA	DESVIO-PADRÃO
ADIMPLENTES OU OUTRA	58,84%	3,09%	0,32453	0,398830
INADIMPLENTES OU OUTRA	80,72%	2,93%	0,21608	0,393438
TEMPORARIAMENTE INADIMPLENTES OU OUTRA	70,60%	0,74%	0,28952	0,386438

Fonte: A autora, 2017.

As seções a seguir apresentam de forma mais detalhada todos os testes realizados com os parâmetros mencionados no Capítulo 4.

### 5.2.1 Situação: Adimplentes ou Outra

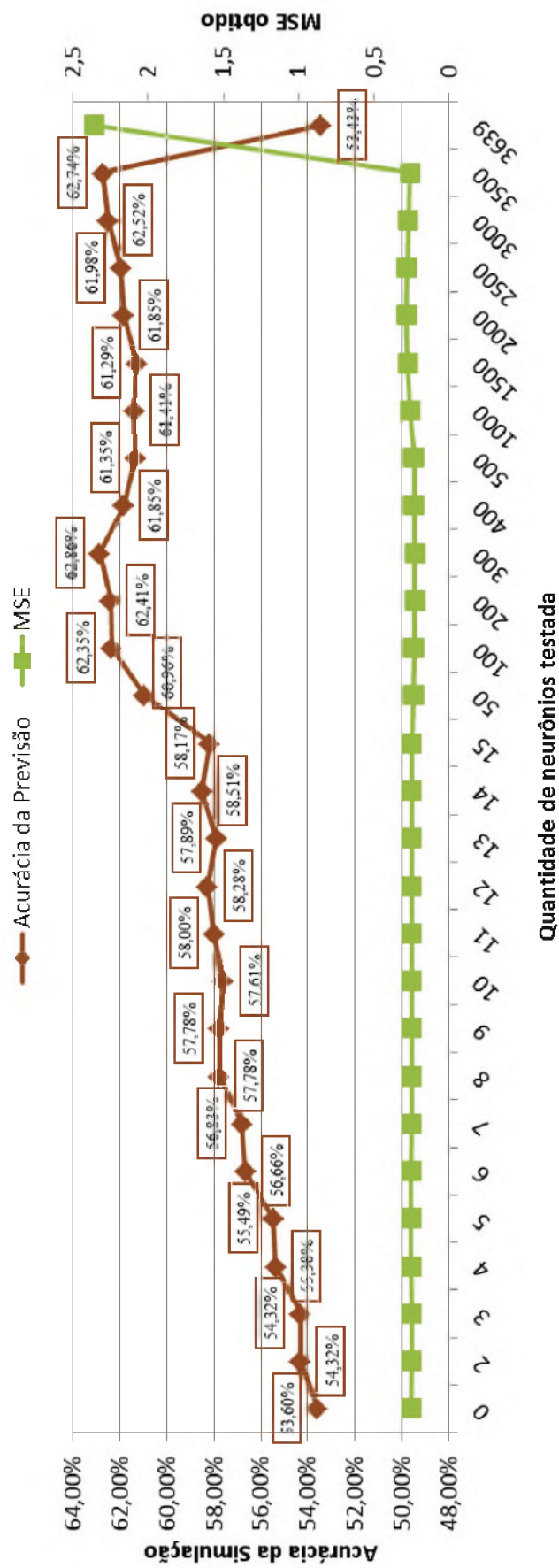
Tendo em vista que a Tabela 9 apresenta apenas as médias dos resultados em questão, esta seção demonstrará, então, o comportamento tanto da acurácia das simulações realizadas, bem como do MSE obtido para cada configuração. Vale ressaltar que, sempre que o conjunto de dados for o mesmo, para um certo número de neurônios, o MSE será o mesmo.

Os dados resultantes dos testes realizados com o método de RNA com RBF, para o modelo dos “Adimplentes ou Outra” são apresentados pela Figura 13 a seguir. Nela, é possível observar um comportamento praticamente constante do MSE do conjunto de teste até a última simulação realizada (3.639 neurônios).

Enquanto isso, a acurácia aparenta crescer até chegar ao máximo de 62,86% (300 neurônios na camada oculta). Mas no momento que a acurácia cai para a última simulação de 3.639 neurônios, chegando a 53,46%, o MSE aumenta, sugerindo a possibilidade de *overfitting* (sobreajuste) da rede.

Figura 13 - Acurácia e MSE das RBF para o modelo “Adimplentes ou Outra”

### Resultados Adimplentes



Fonte: A autora, 2017.

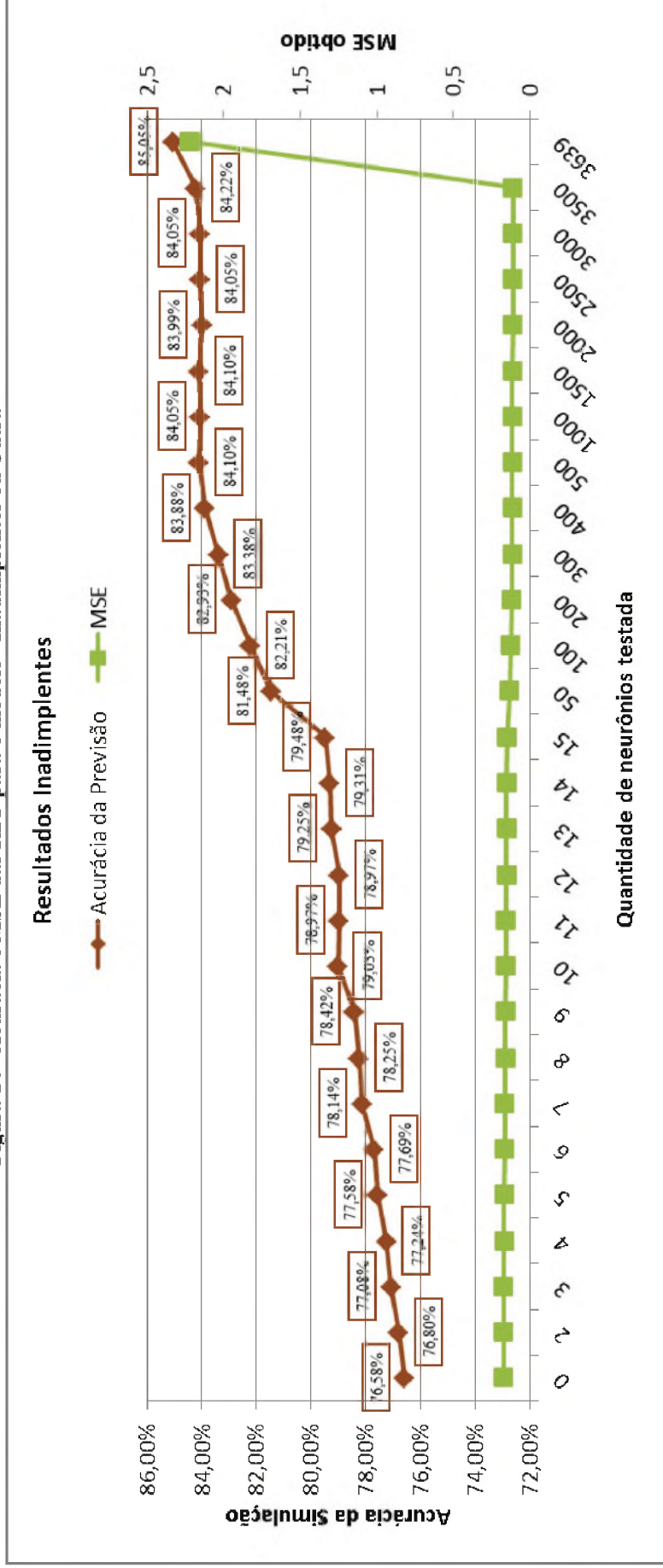
### 5.2.2 Situação: Inadimplentes ou Outra

Os dados resultantes dos testes realizados com o método de RNA com RBF, para o modelo dos “Inadimplentes ou Outra” são apresentados pela Figura 14 a seguir.

Similarmente ao teste anterior, percebe-se um comportamento constante quanto à curva do MSE até, novamente, a última simulação realizada onde o erro cresce rapidamente, sugerindo que para esta quantidade de neurônios na camada oculta a rede pode apresentar *overfitting*.

Agora, diferentemente à situação apresentada na seção 5.2.1, a acurácia para este modelo tende apenas a crescer, chegando ao máximo obtido para os 3.639 neurônios na camada oculta, com 85,05% de acurácia na simulação.

Figura 14 - Acurácia e MSE das RBF para o modelo “Inadimplentes ou Outra”



Fonte: A autora, 2017.

### 5.2.3 Situação: Temporariamente Inadimplentes ou Outra

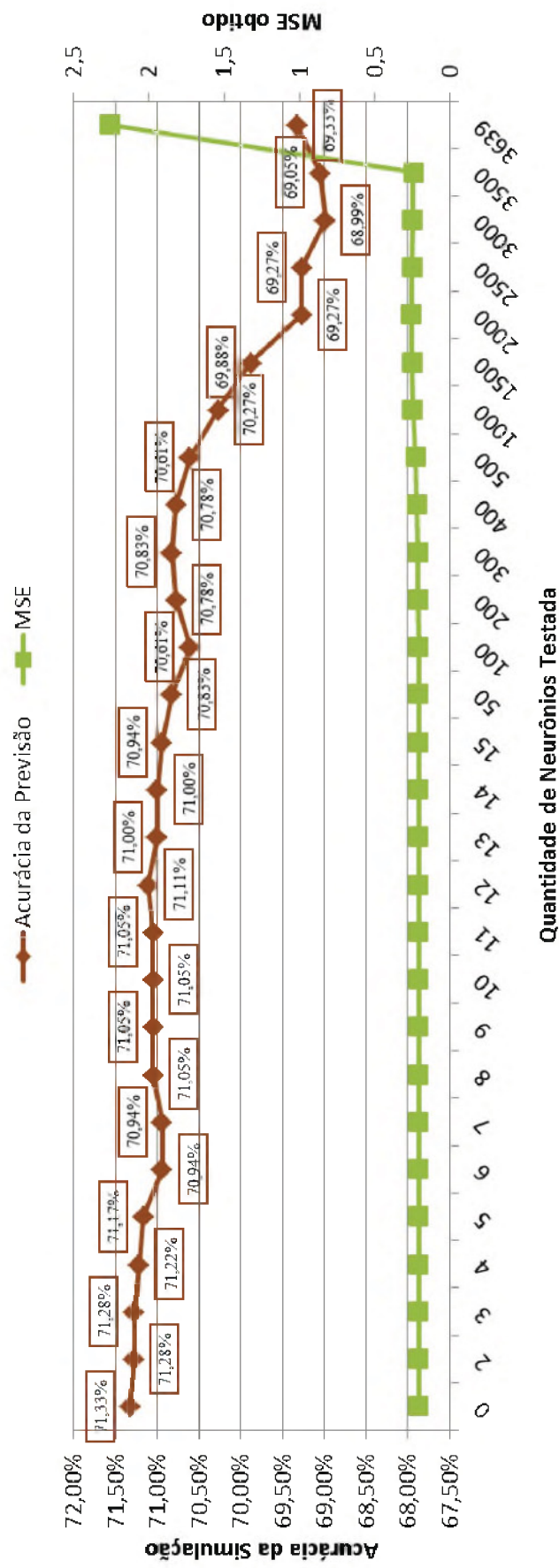
Os dados resultantes dos testes realizados com o método de RNA com RBF, para o modelo dos “Temporariamente Inadimplentes ou Outra” são apresentados pela Figura 15 a seguir.

O comportamento deste modelo difere quanto à acurácia, para este caso, ela tende a continuamente cair, desta forma seu máximo acaba ocorrendo para 0 neurônios na camada oculta, representando 71,33% de acurácia.

O MSE neste caso tem o mesmo comportamento para as últimas duas situações apresentadas (Adimplentes ou Outra; e Inadimplentes ou Outra), permanecendo, portanto, constante até a última simulação realizada. Novamente indicando *overfitting* da rede para a configuração apresentada, 3.639 neurônios na camada oculta e superiores.

Figura 15 - Acurácia e MSE das RBF para o modelo “Temporariamente Inadimplentes ou Outra”

### Resultados Temporariamente Inadimplentes



Fonte: A autora, 2017.

### 5.3 RESULTADOS RL

A seguir são encontrados os resultados da RL, sendo que os experimentos foram realizados considerando inicialmente as três classes em conjunto (Tabela 10) e, em seguida, considerando as classificações de “Adimplentes ou Outra”; “Inadimplentes ou Outra”; e “Temporariamente Inadimplentes ou Outra”.

**Tabela 10 - Resultados Obtidos pela RL considerando as três classes**

<i>CLASSE</i>	<i>ACURÁCIA</i>
<b>ADIMPLENTES OU OUTRA</b>	71,50%
<b>INADIMPLENTES OU OUTRA</b>	88,50%
<b>TEMPORARIAMENTE INADIMPLENTES OU OUTRA</b>	71,60%

Fonte: A autora, 2017.

De forma análoga às seções 5.1 e 5.2, mais detalhes sobre esses métodos serão apresentados a seguir.

#### 5.3.1 Situação: Adimplentes ou Outra

Inicialmente, o *software* SPSS apresenta a quantidade de casos utilizados para a RL juntamente com se foram encontrados casos que não possuíam valores. Onde foi visto que todos os casos seriam selecionados.

Após feita a análise anterior, as quantidades de clientes para cada atributo são listadas para cada um dos atributos (esta listagem pode ser encontrada no APÊNDICE A deste trabalho), e, em seguida, é gerado o próximo conjunto de *outputs*, exibidos pelas Tabela 11 e Tabela 12 a seguir.

**Tabela 11 - Matriz de Confusão (Passo 0 - RL)**

<i>Observado</i>	<i>Previsto</i>		<i>Porcentagem Correta</i>
	0	1	
0	2832	0	100%
1	2600	0	0%
<b>Porcentagem Geral</b>			<b>52,1%</b>

Fonte: A autora, 2017.

**Tabela 12 - Variáveis consideradas na Equação (Passo 0)**

	<i>B</i>	<i>S.E.</i>	<i>Wald</i>	<i>Graus de Liberdade</i>	<i>Sig.</i>	<i>Exp(B)</i>
Passo 0 Constante	-0,085	0,027	9,903	1	0,002	,918

Fonte: A autora, 2017.

As tabelas em questão formam a base do modelo de RL, lembrando que neste momento o modelo não possui as variáveis explicativas. As previsões demonstradas na Tabela 11 são feitas na categoria que possuir maior frequência nos atributos utilizados dentro do conjunto de dados, ou seja, neste modelo, mais clientes são previstos como “0” (classe “Outros”) porque esta tem maior representatividade nos dados.

Na Tabela 12 são apresentadas as variáveis contidas na equação, onde é apresentado o valor de  $B_0$  (-0,085), apenas a constante, as demais variáveis, como mencionado anteriormente, não estão contidas na equação até o momento.

Seguindo para o Passo 1 da RL, onde é verificado se o modelo é considerado bom ou não. Verificamos na Tabela 13 que o valor  $p$  (Sig) é superior à 0,05, ou seja, o modelo gerado é considerado como bem suportado.

**Tabela 13 - Teste Hosmer e Lemeshow**

<i>Passo</i>	<i>Chi-quadrado</i>	<i>Graus de liberdade</i>	<i>Sig.</i>
1	3,118	8	,927

Fonte: A autora, 2017.

Agora, para verificar a acurácia do modelo proposto, na Tabela 14 é apresentada a matriz de confusão final do algoritmo utilizado.

**Tabela 14 - Matriz de Confusão de RL para o modelo “Adimplentes ou Outra”**

<i>Observado</i>	<i>Previsto</i>		<i>Porcentagem Correta</i>
	0	1	
0	1891	941	66,8%
1	609	1991	76,6%
<b>Porcentagem Geral</b>			<b>71,5%</b>

Fonte: A autora, 2017.

Verifica-se, assim, que para a maioria dos casos a RL acertou a previsão (71,5% de acurácia global). É observado também que a RL tem maior Erro tipo 2, apresentando mais falsos positivos (941 clientes foram considerados adimplentes, quando não são desta classe). Enquanto isso, para o Erro tipo 1 (falsos negativos), 11% dos casos (609 clientes) não foram considerados adimplentes quando, na verdade, eram.

As variáveis consideradas na equação de regressão podem ser encontradas no APÊNDICE B deste trabalho.

### 5.3.2 Situação: Inadimplentes ou Outra

De forma análoga ao modelo apresentado anteriormente, a primeira informação que o *software* fornece é a quantidade de casos utilizados para a RL juntamente com se foram encontrados casos que não possuíam valores, onde percebeu-se que todos os casos poderiam ser utilizados pelo método.

Feita esta análise, as quantidades de clientes adimplentes para são listadas para cada um dos atributos (esta listagem pode ser encontrada no APÊNDICE A deste trabalho), e, em seguida, é gerado o próximo conjunto de *outputs*, exibido pela Tabela 15 a seguir. Não serão apresentadas as variáveis que fazem parte do modelo neste ponto pois, da mesma forma que para os “Adimplentes ou Outra”, neste momento apenas uma constante está presente no modelo proposto

**Tabela 15 - Matriz de Confusão (Passo 0 - RL)**

<i>Observado</i>	<i>Previsto</i>		<i>Porcentagem Correta</i>
	0	1	
0	4151	0	100%
1	1281	0	0%
<b>Porcentagem Geral</b>			<b>76,4%</b>

Fonte: A autora, 2017.

É possível perceber que a acurácia para o Passo 0 neste momento já é superior ao mesmo passo para o modelo dos “Adimplentes ou Outra”, isso se deve ao fato que existem menos clientes Inadimplentes do que os demais tipos.

Seguindo para o Passo 1 da RL, onde é verificado se o modelo é considerado bom ou não. Verificamos na Tabela 16 que o valor *p* (Sig) desta vez é inferior a 0,05, ou seja,

o modelo gerado não é bem suportado. Isso pode ocorrer devido ao fato que a quantidade de clientes considerados inadimplentes é inferior às demais classes.

**Tabela 16 - Teste Hosmer e Lemeshow**

<i>Passo</i>	<i>Chi-quadrado</i>	<i>Graus de liberdade</i>	<i>Sig.</i>
1	18,616	8	,017

Fonte: A autora, 2017.

Agora, para verificar a acurácia do modelo proposto, na Tabela 17 é apresentada a matriz de confusão final do algoritmo utilizado.

**Tabela 17 - Matriz de Confusão de RL para o modelo “Inadimplentes ou Outra”**

<i>Observado</i>	<i>Previsto</i>		<i>Porcentagem Correta</i>
	0	1	
0	3976	175	95,8%
1	451	830	64,8%
<b>Porcentagem Geral</b>			<b>88,5%</b>

Fonte: A autora, 2017.

É observado pela Tabela 17 que para a maioria dos casos a RL acertou a previsão (88,5% de acurácia global). Verifica-se também que a rede tem maior Erro tipo 1, apresentando mais falsos negativos (451 clientes foram considerados como outros, quando são inadimplentes). Enquanto isso, para o Erro tipo 2 (falsos positivos), 3% dos casos (175 clientes) foram considerados inadimplentes quando, na verdade, não eram.

As variáveis consideradas na equação de regressão podem ser encontradas no APÊNDICE B deste trabalho.

### 5.3.3 Situação: Temporariamente Inadimplentes ou Outra

De mesma maneira aos modelos apresentados anteriormente (Adimplentes ou Outra; e Inadimplentes ou Outra), a primeira informação que o *software* fornece é a

quantidade de casos utilizados para a RL juntamente com se foram encontrados casos que não possuíam valores, novamente, todos os casos foram utilizados na regressão.

As quantidades de clientes adimplentes para são listadas para cada um dos atributos (esta listagem pode ser encontrada no APÊNDICE A deste trabalho). Em seguida, é gerado o próximo conjunto de *outputs*, exibidos pela TABELA a seguir. Não serão apresentadas as variáveis que fazem parte do modelo neste ponto, pois, da mesma forma que para os “Adimplentes ou Outra” e para os “Inadimplentes ou Outra”, neste momento apenas uma constante está presente no modelo proposto.

A Tabela 18 a seguir apresenta a matriz de confusão do Passo 0 do algoritmo de RL utilizado, nela é possível ver uma acurácia de 71,4% uma vez que a quantidade de clientes considerados “Temporariamente Inadimplentes” pelo banco é inferior às somas das quantidades presentes nas demais classes.

**Tabela 18 - Matriz de Confusão (Passo 0 - RL)**

<i>Observado</i>	<i>Previsto</i>		<i>Porcentagem Correta</i>
	0	1	
0	3881	0	100%
1	1551	0	0%
<b>Porcentagem Geral</b>			<b>71,4%</b>

Fonte: A autora, 2017.

Seguindo para o Passo 1 da RL, onde é verificado se o modelo é considerado bom ou não. Verificamos na Tabela 19 que o valor  $p$  (Sig) desta vez é superior a 0,05, ou seja, o modelo gerado é considerado bem suportado.

**Tabela 19 - Teste Hosmer e Lemeshow**

<i>Passo</i>	<i>Chi-quadrado</i>	<i>Graus de liberdade</i>	<i>Sig.</i>
1	13,311	8	,102

Fonte: A autora, 2017.

Agora, para verificar a acurácia do modelo proposto, na Tabela 20 é apresentada a matriz de confusão final do algoritmo utilizado.

**Tabela 20 - Matriz de Confusão de RL para o modelo “Temporariamente Inadimplentes ou Outra”**

	<i>Previsto</i>	<i>Porcentagem Correta</i>
--	-----------------	----------------------------

<i>Observado</i>	0	1	
0	3771	110	97,2%
1	1435	116	7,5%
<b>Porcentagem Geral</b>			<b>71,6%</b>

Fonte: A autora, 2017.

Verifica-se, então, pela Tabela 20 que para a maioria dos casos a RL acertou a previsão (71,6% de acurácia global), mas a maior parte do acerto deste modelo foi para os clientes classificados como “Outros” do que para os “Temporariamente Inadimplentes” em si. É observado também que a RL tem maior Erro tipo 1, apresentando mais falsos negativos (1434 clientes foram considerados como outros, quando são temporariamente inadimplentes). Enquanto isso, para o Erro tipo 2 (falsos positivos), 2% dos casos (110 clientes) foram considerados temporariamente inadimplentes quando, na verdade, não eram.

As variáveis consideradas na equação de regressão podem ser encontradas no APÊNDICE B deste trabalho.

## 6 CONSIDERAÇÕES FINAIS

Neste trabalho foram propostos três métodos de reconhecimento de padrões para a classificação de clientes “adimplentes” ou “inadimplentes” ou “temporariamente inadimplentes” de pessoas jurídicas que fazem uso de serviços de crédito bancário visando a classificação de novas solicitações. Foi considerada uma base de dados que incluía 5.432 clientes (onde 2.600 são comprovadamente adimplentes; 1.281 são inadimplentes; e, por fim, 1.551 clientes são considerados temporariamente inadimplentes) disponibilizados por uma instituição financeira aqui denominada de Banco XYZ, sendo que para cada instância foram considerados 15 atributos (listados no Capítulo 2 deste trabalho).

Os métodos de reconhecimento de padrões escolhidos foram as RNA MLP, as RNA RBF e a RL. Os três métodos foram testados em modelo de “um contra todos” para que a saída da rede fosse binária para cada uma das três classes de interesse deste trabalho.

As características da RNA do tipo MLP foram a utilização do algoritmo de aprendizagem Levenberg-Marquardt, com uma topologia três camadas (Camada de Entrada, Camada Oculta e Camada de Saída). Para a camada oculta, o número de neurônios presentes nela variou de 1 a 20.

Os mesmos testes foram realizados para as RNA do tipo RBF, sendo que o maior número de neurônios testados foi representado pelo número de instâncias a serem treinadas pela rede, ou seja, após o *hold-out* estratificado, foi obtido um conjunto de treinamento com 3.639 instâncias, logo o número de neurônios na camada oculta para as RNA RBF foram 3.639.

E o último método foi a RL que teve como objetivo comparar o método estatístico com os de Inteligência Artificial mencionados anteriormente. Para este método foi utilizado o método de entrada de dados *stepwise* e para a avaliação do ajuste da regressão utilizou-se o Teste de Homer-Lemeshow para um intervalo de confiança de 95%.

Para cada método então foram realizados testes de três situações, uma representando cada uma das classes mencionadas anteriormente. Ou seja, para o método utilizado foram testadas as classificações Adimplentes, Inadimplentes e Temporariamente Inadimplentes, no modelo de “um contra todos”, o método então classificaria, para a primeira situação, se o cliente pertence ou não à classe dos Adimplentes (0 ou 1, respectivamente), caso a resposta fosse nula, o cliente poderia então

ser Inadimplente ou Temporariamente Inadimplente, sendo necessário o uso da rede para as outras duas situações para a tomada de decisão. A seguir, na Tabela 21 encontram-se as acurácias de cada um dos nove testes realizados.

**Tabela 21 - Resumo das Acurácias (%) dos modelos aplicados às três classes**

<i>Métodos</i>	<i>Classes</i>		
	<i>Adimplentes ou Outra</i>	<i>Inadimplentes ou Outra</i>	<i>Temporariamente Inadimplentes ou Outra</i>
MLP	74,7	91,4	74,6
RBF	58,84	80,72	70,60
RL	71,5	88,5	71,6

Fonte: A autora, 2017.

Pode ser observado que os resultados das redes MLP foram superiores em termos de acurácia em relação aos outros dois métodos para cada uma das situações testadas, tendo seu melhor resultado para a classificação dos Inadimplentes (91,40% de acurácia) e acurácias semelhantes para a classificação dos Adimplentes e dos Temporariamente Inadimplentes (74,70% e 74,60% respectivamente).

O segundo melhor método testado foi a RL sendo seus resultados superiores a 70,00% para as três classificações realizadas, lembrando que para a classe dos Inadimplentes o modelo não foi considerado como tendo um bom ajuste para os dados, mas atingindo acurácia de 88,50%. E as redes RBF tiveram resultados inferiores para as três situações quando comparadas com os outros métodos testados, principalmente para a classificação dos Adimplentes, onde a rede tem acurácia média de 58,84%.

Adicionalmente, pode-se observar que a taxa de acurácia do modelo dos Inadimplentes é significativamente superior às demais classes (Adimplentes e dos Temporariamente Inadimplentes), sendo que estas apresentaram resultados semelhantes em todas as três técnicas, mostrando que a classe dos Inadimplentes está melhor discriminada do que as demais. Este fato pode ser considerado valioso, já que exatamente os Inadimplentes, ou seja, aquelas empresas que se concedido o crédito, darão prejuízo à instituição bancária.

Assim, dada uma nova instância, poderemos classificá-la utilizando os pesos obtidos para cada uma das classes (Adimplentes, conjunto de pesos 18, 20 neurônios na camada oculta, Tabela 6; Inadimplentes, conjunto de pesos 13, 13 neurônios na camada oculta, Tabela 7 e Temporariamente Inadimplentes, conjunto de pesos 15, 19 neurônios na camada oculta, Tabela 8) e observar quais as classificações obtidas em cada uma dessas

três redes e tomar a decisão considerando as acurácias apresentadas para cada uma destas configurações. Vale ressaltar que caso esta nova instância seja enquadrada na classe “Temporariamente Inadimplente”, evidenciando uma empresa que provavelmente pagará, mas com atraso de até 180 dias, os gerentes da instituição terão que decidir se concretizarão o crédito ou não.

A contribuição deste trabalho pode ser verificada tanto em termos de aplicação prática quanto acadêmica. Quanto aplicação prática do problema de concessão de crédito, a dissertação desenvolvida apresenta uma nova classe, a de Temporariamente Inadimplentes, sendo que a adição desta nova situação pode trazer benefícios tanto à Instituição Financeira em conhecer melhor seus futuros clientes quanto aos clientes que antes poderiam ter seu pedido de crédito negado devido à determinado risco que ele poderia apresentar ao banco.

Em relação à contribuição acadêmica, o trabalho aqui desenvolvido, pela sua característica exploratória, gerou maior familiaridade com o problema de concessão de crédito bancário, definindo suas características e ainda sua relação com os métodos de reconhecimento de padrões utilizados (RNA MLP, RNA RBF e RL) considerando uma nova classe.

Como trabalhos futuros sugere-se a aplicação de técnicas de aperfeiçoamento de desempenho da classificação pelas RNA, como técnicas de *ensemble*, utilização de híbridos conhecidos, como RNA + AG ou outro método heurístico. Ainda é possível testar novas configurações para as RNA utilizadas neste trabalho, alterando a quantidade de neurônios na camada oculta, os conjuntos de pesos utilizados ou ainda não fixando a semente de aleatório para cada conjunto de peso. Quanto às RBF, é possível testar outras funções de ativação, não apenas a *default* do *software* utilizado, ou, ainda, testar novos valores de propagação do neurônio na camada oculta para verificar o desempenho desta rede em diferentes situações.

Fica ainda como sugestão, a utilização de outros métodos de reconhecimento de padrões, como clusterização ou árvores de decisão, comparando-os com os resultados obtidos neste trabalho. Assim seria possível ampliar a gama de técnicas utilizadas, fazendo assim uma “varredura” de alguns dos métodos disponíveis na literatura.

Outra sugestão seria realizar uma análise do impacto financeiro que o banco sofreria com a implementação do método de classificação gerado, verificando tanto as perdas quanto os anhos que a instituição teria, analisando então a lucratividade ou mesmo

o prejuízo que a implementação da presente metodologia apresentaria à instituição bancária em análise.

## REFERÊNCIAS

- ABDOU H. A.; TSAFACK M. D. D.; NTIM C. G.; BAKER R. D.; Predicting creditworthiness in retail banking with limited scoring data. *Knowledge-Based Systems*. v. 103, p. 89 – 103, 2016.
- ABEDINI M.; AHMADYADEH F.; NOOROSSANA R. Customer credit scoring using a hybrid data mining approach. *Kybernetes*. v. 45, n. 10, 2016.
- ANDERLINI, E.; FOREHAND, D.I.M.; BANNON, E.; ABUSARA, M. Reactive control of a wave energy converter using artificial neural networks. *International Journal of Marine Energy*. v. 19, p. 207-220, 2017.
- AKKOÇ S.; An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research*. v. 222, p. 168 – 178, 2012.
- ALA'RAJ M.; ABBOD M. F.; Classifiers consensus system approach for credit scoring. *Knowledge-Based Systems*. v. 104, p. 89 – 105, 2016.
- 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.
- ALEJO, R. P.; RUIZ-CABELLO, J.; CORTIJO, M.; RODRIGUEZ, I.; ECHAVE, I.; REGADERA, J.; ARRAZOLA, J.; AVILÉS, P.; BARREIRO, P.; GARGALO, D.; GRAÑA, M. Computer-assisted enhanced volumetric segmentation magnetic resonance imaging data using a mixture of artificial neural networks. *Magnetic Resonance Imaging*, v. 21, p. 901 – 912, 2003.
- ALTMAN, E. I.; SAUNDERS, A.. Credit Risk Measurement: developments over the 20 years. *Journal of Banking and Finance*. vol. 21, p. 1721 – 1744, 1998.
- ANDRIC K.; KALPIC D.; The effect of class distribution on classification algorithms in credit risk assessment. 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2016 - Proceedings. 2016.
- ARMENGOL, E.; BOIXADER, D.; GRIMALDO, F.. Special Issue on Pattern Recognition Techniques in Data Mining. 2017.
- ASADIS.; HADAVANDI E.; MEHMANPAZIR F.; NAKHOSTIN M.M.; Hybridization of evolutionary Levenberg-Marquardt neural networks and data pre-processing for stock market prediction. *Knowledge-Based Systems*. v. 35, p. 245 – 258, 2012.
- ASCIONE, F.; BIANCO, N.; DE STASIO, C.; MAURO, G. M.; VANOLI, G. P. CASA, cost-optimal analysis by multi-objective optimisation and artificial neural networks: A new framework for the robust assessment of cost-optimal energy retrofit, feasible for any building. *Energy and Buildings*. v. 146, p. 200-219, 2017.
- BABIC, Z.; PLAZIBAT, N.. Ranking of Enterprises based on multi-criterial analysis. *International Journal of Production Economics*. vol. 56/57, p. 29 – 35, 1998.
- BAHRAMMIRZAEI, A. A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing & Applications*, v. 19, p. 1165 – 1195, 2010.
- BAKIRCIOĞLU, H.; KOÇAK, T. Survey of random neural networks applications. *European Journal of Operational Research*, v. 1260, p. 319 – 330, 2000.
- BANCO CENTRAL DO BRASIL (Brasil). Regimento Interno do Banco Central do Brasil. Brasília: Banco Central do Brasil. p. 1, 2017.

- BANCO CENTRAL DO BRASIL (Brasil). Dispõe sobre a implementação de estrutura de gerenciamento de risco de crédito. Brasília: Banco Central do Brasil. p. 1, 2009.
- BARBOZA F.; KIMURA H.; ALTMAN E.; Machine learning models and bankruptcy prediction. *Expert Systems with Applications*. v. 83, p. 405 – 417, 2017.
- BEALE, M. H.; HAGAN, M. T.; DEMUTH, H. B.. *Neural Network Toolbox™: Getting Started Guide*. MathWorks. 134 p., 2017.
- BEALE, M. H.; HAGAN, M. T.; DEMUTH, H. B.. *Neural Network Toolbox™: Users Guide*. MathWorks. 512 p., 2017.
- BELLOTI, T.; CROOK, J.. Support Vector Machines for credit scoring and discovery of significant features. *Expert Systems with Applications*. vol. 36, n. 2, p. 3302 – 3308, 2009.
- BEQUÉ A.; LESSMANN S.; Extreme learning machines for credit scoring: An empirical evaluation. *Expert Systems with Applications*. v. 86, p. 42 – 53, 2017.
- BISOI R.; DASH P.K.; A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter. *Applied Soft Computing Journal*. v. 19, p. 41 – 56, 2014.
- BITTENCOURT, G. *Inteligência Artificial: Ferramentas e Teorias*. 3. ed. Florianópolis: Editora da UFSC, 2006. ISBN 85-328-0138-2.
- BLANCO A.; PINO-MEJÍAS R.; LARA J.; RAYO S.; Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with Applications*. v. 40, p. 356 – 364, 2013.
- BOLTON, C.. *Logistic Regression and its Application in Credit Scoring*, Universidade de Pretoria. 238 p., 2009.
- BOUHOUNE, K.; AZID, K.; BOUCHERIT, M.S.; CHÉRITI, A. Hybrid control of the three phase induction machine using artificial neural networks and fuzzy logic. *Applied Soft Computing*. v. 55, p. 289-301, 2017.
- BOUKADIDA, H.; HASSEN, N.; GAFSI, Z.; BESBES, K. A Highly Time-Efficient Digital Multiplier based on the A2 Binary Representation. *International Journal of Engineering Science and Technology*, v. 3, n. 5, p. 4498 – 4509, 2011.
- BOŽNAR, M. Z.; GRAŠIČ, B.; OLIVEIRA, A. P.; SOARES, J.; MLAKAR, P. Spatially transferable regional model for half-hourly values of diffuse solar radiation for general sky conditions based on perceptron artificial neural networks. *Renewable Energy*. v. 103, p. 794-810, 2017.
- BRAGA, A. D. P.; CARVALHO, A. C. P. D. L. F.; LUDEMIR, T. B. *Redes Neurais Artificiais: Teoria e Aplicações*. 2. ed. Rio de Janeiro: LTC, 2000.
- BRASIL. Lei 4.595, de 31 de Dezembro de 1964. Dispõe sobre a Política e as Instituições Monetárias, Bancárias e Creditícias, Cria o Conselho Monetário Nacional e dá outras providências. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/L4595.htm](http://www.planalto.gov.br/ccivil_03/leis/L4595.htm)>. Acesso em: 20 de setembro de 2017.
- BRAVO C.; THOMAS L.C.; WEBER R.; Improving credit scoring by differentiating defaulter behaviour. *Journal of the Operational Research Society*. v. 66, p.771 – 781, 2015.
- BROWN I.; MUES C.; An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*. v. 39, p. 2446 – 3453, 2012.
- BRYANT, K. ALEES: an agricultural loan evaluation expert system. *Expert Systems with Applications*. vol. 21, p. 75 – 85, 2001.
- BUCHANAN, B. G. A (Very) Brief History of Artificial Intelligence. *AI Magazine*, v. 26, p. 53 – 60, 2006.

- BURRELL, P. R.; FOLARIN, B. O. The Impact of Neural Networks in Finance. *Neural Computing & Applications*, v. 6, p. 193 – 200, 1997.
- CABRERA, P.; CARTA, J. A.; GONZÁLEZ, J.; MELIÁN G. Artificial neural networks applied to manage the variable operation of a simple seawater reverse osmosis plant. *Desalination*. v. 416, p. 140-156, 2017.
- CASCARDI, A.; MICELLI, F.; AIELLO, M. A. An Artificial Neural Networks model for the prediction of the compressive strength of FRP-confined concrete circular columns. *Engineering Structures*. v. 140, p. 199-208, 2017.
- CERVONE, G.; CLEMENTE-HARDING, L.; ALESSANDRINI, S.; MONACHE, L. D. Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. *Renewable Energy*. v. 108, p. 274-286, 2017.
- CHANG P.-C.; WU J.-L.; LIN J.-J.; A Takagi-Sugeno fuzzy model combined with a support vector regression for stock trading forecasting. *Applied Soft Computing Journal*. v. 38, p. 831 – 842, 2016.
- CHAUHAN, D.; JAISWAL, V.. An efficient data mining classification approach for detecting lung cancer disease. In: *Communication and Electronics Systems (ICCES)*, International Conference on. IEEE, p. 1 – 8, 2016.
- CHEN, X.; WANG, X.; WU, D. D.. Credit risk measurement and early warning of SMEs: An empirical study of listed SMEs in China. *Decision Support Systems*. vol. 49, p. 301 – 310, 2010.
- CHEN, W.; FU, Z. J.; CHEN, C. S.. *Recent Advances in Radial Basis Function Collocation Methods*. Springer. 90 p. 2014.
- CHEN K.; ZHOU Y.; DAI F.; A LSTM-based method for stock returns prediction: A case study of China stock market. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*. 2015.
- CHEN T.-L.; CHEN F.-Y.; An intelligent pattern recognition model for supporting investment decisions in stock market. *Information Sciences*. v. 346 – 347, p. 261 – 274, 2016.
- CHEN Y.; HAO Y.; A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications*. v. 80, p. 240 – 355, 2017.
- CHIRINKO, R. S.; GUILL, G. D.; HEBERT, P.. Developing a systematic approach to credit risk management. *Journal of Retail Banking Services*. vol. 3, p. 29 – 37, 1991.
- CHONG E.; HAN C.; PARK F.C.. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*. v. 83, p. 187 – 205, 2017.
- COSTA Jr., R. A. L. F. P. P.; CRUZ, F. R. B.. Minimização do erro no algoritmo back-propagation aplicado ao problema de manutenção de motores. *Pesquisa Operacional*. vol. 18, n. 1, p. 21 – 36, 1998.
- CUBILES-DE-LA-VEGA M.-D.; BLANCO-OLIVER A.; PINO-MEJÍAS R.; LARA-RUBIO J.; Improving the management of microfinance institutions by using credit scoring models based on Statistical Learning techniques. *Expert Systems with Applications*. v. 40, p. 6910 – 6917, 2013.
- DAI W.; WU J.-Y.; LU C.-J.; Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. *Expert Systems with Applications*. v. 39, p. 4444 – 4452, 2012.
- DANENAS P.; GARSVA G.; Selection of Support Vector Machines based classifiers for credit risk domain. *Expert Systems with Applications*. v. 42, p. 3194 – 3204, 2015.

- DASH R.; DASH P.; Efficient stock price prediction using a Self Evolving Recurrent Neuro-Fuzzy Inference System optimized through a Modified technique. *Expert Systems with Applications*. v. 52, p. 75 – 90, 2016.
- DASH R.; DASH P.K.; BISOI R.; A differential harmony search based hybrid interval type2 fuzzy EGARCH model for stock market volatility prediction. *International Journal of Approximate Reasoning*. v. 59, p. 81 – 104, 2015.
- DENISOV, I. A. Luciferase-based bioassay for rapid pollutants detection and classification by means of multilayer artificial neural networks. *Sensors and Actuators B: Chemical*. v. 242, p. 653-657, 2017.
- DESAI, V.; CROOK, J. OVERSTREET Jr., G.. A comparison of neural networks and linear scoring models in credit union environment. *European Journal of Operational Research*. vol. 95, p. 24 – 37, 1996.
- DEVI K.N.; BHASKARAN V.M.; KUMAR G.P.; Cuckoo optimized SVM for stock market prediction. *ICIIECS 2015 - 2015 IEEE International Conference on Innovations in Information, Embedded and Communication Systems*. 2015.
- DIAKOULAKI, D.; MAVROTAS, G.; PAPAYANNAKIS, L.. A multicriteria approach for evaluating the performance of industrial firms. *Omega*. vol. 20, n. 4, p. 467 – 474, 1992.
- DING A. A.; TIAN S.; YU Y.; GUO H.; A class of discrete transformation survival models with application to default probability prediction. *Journal of the American Statistical Association*. v. 107, n. 499, p. 990 – 1003, 2012.
- 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, p. 378 – 396, 2012.
- ELNESR, M. N.; ALAZBA, A. A. Simulation of water distribution under surface dripper using artificial neural networks. *Computers and Electronics in Agriculture*. v. 143, p. 90-99, 2017.
- ELSALAMONY, H. A. Anaemia cells detection based on shape signature using neural networks. *Measurement*, v. 104, p. 50 – 59, 2017.
- FAHNER, G.. Estimating casual effects of credit decisions. *International Journal of Forecasting*. vol. 28, n. 1, p. 248 – 260, 2012.
- FAN Y.; FOUTZ N.; JAMES G.M.; JANK W.; Functional response additive model estimation with online virtual stock markets. *Annals of Applied Statistics*. v. 8, n. 4, p. 2435 – 2460, 2014.
- FATH, A. H. Application of radial basis function neural networks in bubble point oil formation volume factor prediction for petroleum system. *Fluid Phase Equilibria*, v. 437, p. 14 – 22, 2017.
- FAUSETT, Laurene V. *Fundamentals of neural networks: architectures, algorithms, and applications*. 1. ed. USA: Prentice Hall, 1994. 461p.
- FEDOROVA E.; GILENKO E.; DOVZHENKO S.; Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications*. v. 40, p. 7285 – 7293, 2013.
- FENG, Y.; MU, Y.; HU, B.; KUMAR, A.. Commodity options purchasing and credit financing under capital constraint. *International Journal of Production Economics*. vol. 153, n. 7, p. 230 – 237, 2014.
- FIELD, A. *Descobrimdo a Estatística usando o SPSS*. 2ª ed. Porto Alegre: Artmed, 684p. 2009
- FREEMAN, J. A.; SKAPURA, D. M.. *Neural Networks: Algorithms, Applications and Programming Techniques*. Addison-Wesley Publishing Company. p.386, 1991.

- GARCÍA V.; MARQUÉS A.I.; SÁNCHEZ J.S.; On the use of data filtering techniques for credit risk prediction with instance-based models. *Expert Systems with Applications*. v. 39, p. 13267 – 13276, 2012.
- GARCIA, J. J.; GARCIA, F.; BERMÚDEZ, J.; MACHADO L. Prediction of pressure drop during evaporation of R407C in horizontal tubes using artificial neural networks. *International Journal of Refrigeration*. 2017.
- GESTEL, T. V.; BAESSENS, B.; GARCIA, J.; DIJCKE, P. V.. A support vector machine approach to credit scoring. *Journal of Banking & Finance*. vol. 2, p. 73 – 82, 2003.
- GIL, A. C. Como elaborar projetos de pesquisa. 4a. ed. São Paulo: Atlas, 2002.
- \_\_\_\_\_. Métodos e técnicas de pesquisa social. 6a. ed. São Paulo: Atlas, 2008.
- GÖÇKEN M.; ÖZÇALICI M.; BORU A.; DOSDOĞRU A.T. Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction. *Expert Systems with Applications*. v. 44, p. 320 – 331, 2016.
- GUNDUZ H.; YASLAN Y.; CATALTEPE Z. Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations. *Knowledge-Based Systems*. v. 137, p. 138 – 148, 2017.
- GUO Y.; DONG C. A novel intelligent credit scoring method using MOPSO. *Proceedings of the 29th Chinese Control and Decision Conference, CCDC 2017*. 2017.
- GURNEY, K. An Introduction to neural networks, England: Taylor & Francis e-library, p. 17, 2004.
- HARRIS T. Credit scoring using the clustered support vector machine. *Expert Systems with Applications*. v. 42, p. 741 – 750, 2015.
- HARRELL JR., F. E.. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer. 598 p, 2015.
- HAYKIN, S.. *Neural Networks, A comprehensive Foundation*. IEEE Press. 1994.
- \_\_\_\_\_. *Redes neurais: princípios e prática*. 2a. ed. Porto Alegre: Bookman. 900p, 2001.
- \_\_\_\_\_. *Neural Networks and Learning Machines*. 3. ed. USA: Prentice Hall, 2009. 906p.
- HENS A.B.; TIWARI M.K.; Computational time reduction for credit scoring: An integrated approach based on support vector machine and stratified sampling method. *Expert Systems with Applications*. v. 39, n. 8, p. 6774 – 6781, 2012.
- HILBE, J. M.. *Practical Guide to Logistic Regression*. USA: Taylor & Francis. 162p, 2016.
- HONG S.Y.; LINTON O.; ZHANG H.J.; An investigation into multivariate variance ratio statistics and their application to stock market predictability. *Journal of Financial Econometrics*. v. 15, n. 2, p. 173 – 222, 2017.
- HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X.. *Applied Logistic Regression*. 3a ed. New Jersey: John Wiley & Sons Inc. 508p, 2013.
- HUSSAIN, S.; ALALILI, A. A hybrid solar radiation modeling approach using wavelet multiresolution analysis and artificial neural networks. *Applied Energy*. v. 208, p. 540-550, 2017.
- IAZZOLINO, G.; BRUNI, M. E.; BERALDI, P.. Using DEA and financial ratings for credit evaluation: an empirical analysis. *Applied Economic Letters*. vol. 20, n. 14, p. 1310 – 1317, 2013.
- IBM SPSS STATISTICS. *Guia do Usuário do Sistema Principal do IBM SPSS Statistics* 24, 2016.

- IGNATIUS, J.; HATAMI-MARBINI, A.; RAHMAN, A.; DHAMOTHARAN, L.; KHOSHNEVIS, P.. A fuzzy decision support system for credit scoring. *Neural Computing and Applications*, p. 1 – 17, 2016.
- JEONG C.; MIN J. H.; KIM M. S.; A tuning method for the architecture of neural network models incorporating GAM and GA as applied to bankruptcy prediction. *Expert Systems with Applications*. v. 39, p. 3650 – 3658, 2012.
- JHA, S.; TOPOL, E. J.. Adapting to artificial intelligence: radiologists and pathologists as information specialists. *JAMA*, v. 316, n. 22, p. 2353-2354, 2016.
- JONES S.. Corporate bankruptcy prediction: a high dimensional analysis. *Review of Accounting Studies*. v. 22, n. 3, p. 1366 – 1422, 2017.
- KAO L.-J.; CHIU C.-C.; CHIU F.-Y.; A Bayesian latent variable model with classification and regression tree approach for behavior and credit scoring. *Knowledge-Based Systems*. v. 36, p. 245 – 252, 2012.
- KARCZMAREK, P.; KIERSZTYN, A.; PEDRYCZ, W..An application of chain code-based local descriptor and its extension to face recognition. *Pattern Recognition*, v. 65, p. 26-34, 2017.
- KAZEM A.; SHARIFI E.; HUSSAIN F.K.; SABERI M.; HUSSAIN O.K.; Support vector regression with chaos-based firefly algorithm for stock market price forecasting. *Applied Soft Computing Journal*. v. 13, n. 2, p.947 – 958, 2013.
- KETHKAR, N.. *Deep Learning with Python: A Hands-on Introduction*. Apress: Índia. 160 p. 2017.
- KHAN, S.; NASSEM, I.; TOGNER, R.; BENNAMOUN, M.. A novel adaptive kernel for the RBF Neural Networks. *Circuits, Systems and Signal Process*. vol. 34, n. 4, p.1639 – 1653, 2017.
- KHATRI S.K.; SRIVASTAVA A.; Using sentimental analysis in prediction of stock market investment. 2016 5th International Conference on Reliability, Infocom Technologies and Optimization, ICRITO 2016: Trends and Future Directions. 2016.
- KIM H. J.; JO N. O.; SHIN K. S.; Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. *Expert Systems with Applications*. v. 59, p. 226 – 234, 2016.
- KIM S.; MUN B.M.; BAE S.J.; Data depth based support vector machines for predicting corporate bankruptcy. *Applied Intelligence*. p. 1 – 14, 2017.
- KOCADAGLI, O.; LANGARI, R. Classification of EEG signals for epileptic seizures using hybrid artificial neural networks based wavelet transforms and fuzzy relations. *Expert Systems with Applications*. v. 88, p. 419-434, 2017.
- KOCH, T. W.; MACDONALD, S. S.. *Bank Management*. 7ª edição, Mason, Ohio South-Western Cengage Learning, 872 p., 2010.
- KÖCHE, J. C.. *Fundamentos de metodologia científica*. 29. ed. RJ: Vozes, 182p., 2011.
- KONNO H.; SAITO M.; Classification of companies using maximal margin ellipsoidal surfaces. *Computational Optimization and Applications*. v. 55, p. 469 – 480, 2013.
- KOROL T.; Early warning models against bankruptcy risk for Central European and Latin American enterprises. *Economic Modelling*. v. 31, p. 22 – 30, 2013.
- KOTSOVOU, G. M.; COTSOVOS, D. M.; LAGAROS, N. D. Assessment of RC exterior beam-column Joints based on artificial neural networks and other methods. *Engineering Structures*. v. 144, p. 1-18, 2017.
- 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.

- KRÖSE, B. J. A.; SMAG, P. P. V. D.. An introduction to Neural Networks. University of Amsterdam. 8ª Edição, 134 p., 1996.
- LABOISSIERE L.A.; FERNANDES R.A.S.; LAGE G.G.; Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. *Applied Soft Computing Journal*. v. 35, p. 66 – 74, 2015.
- LAKATOS, E. M.; MARCONI, M. de A. Fundamentos de Metodologia Científica, São Paulo: Atlas, 5ª Edição, p. 126 – 127, 2003.
- LANZARINI L.C.; VILLA MONTE A.; BARIVIERA A.F.; JIMBO SANTANA P.; Simplifying credit scoring rules using LVQ + PSO. *Kybernetes*. v. 46, n. 1, p. 8 – 16, 2017.
- LEE S.; CHOI W.S.; A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*. v. 40, p. 2941 – 2946, 2013.
- LEŚNIAK, A.. Supporting Contractors' bidding decision: RBF neural network application. *International Conference of Numerical Analysis and Applied Mathematics*. vol. 1738, artigo n. 200002, 2016.
- LESSMANN, S.; BAESENS, B.; SEOW, H. V.; THOMAS, L. C. Benchmarking state-of-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, v. 247, n. 1, p.124 – 136, 2015.
- LI J.; WEI H.; HAO W.; Weight-selected attribute bagging for credit scoring. *Mathematical Problems in Engineering*. v. 2013, p. 1 – 14, 2013.
- LI, M. M.; VERMA, B.. Nonlinear curve fitting to stopping power data using RBF neural networks. *Expert Systems with Applications*. vol. 45, p. 161 – 171, 2016.
- LI Y.-X.; A new method of credit risk assessment of commercial banks. *Proceedings - 2016 International Conference on Robots and Intelligent System, ICRIS 2016*. p. 34 – 37, 2016.
- LI J.; BU H.; WU J.; Sentiment-aware stock market prediction: A deep learning method. *14th International Conference on Services Systems and Services Management, ICSSSM 2017 - Proceedings*. 2017.
- LI X.; YANG L.; XUE F.; ZHOU H.; Time series prediction of stock price using deep belief networks with intrinsic plasticity. *Proceedings of the 29th Chinese Control and Decision Conference, CCDC 2017*. 2017.
- LI Y.; TIAN Y.; LI K.; YHOU F.; YANG W.; Reject inference in credit scoring using Semi-supervised Support Vector Machines. *Expert Systems with Applications*. v. 74, p. 105 – 114, 2017.
- LI Z.; CROOK J.; ANDREEVA G.; Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems with Applications*. v. 80, p. 94 – 106, 2017.
- LIN, X.; WANG, X.; HAO, Z.. Supervised learning in multilayer spiking eural networks with inner products of spike trains. *Neurocomputing*. vol. 237, p. 59 – 70, 2017.
- LINS, A. J. C. C.; MUNIZ, M. T. C.; GARCIA, A. N. M.; GOMES, A. V.; CABRAL, R. M.; BASTOS-FILHO, C. J. A. Using artificial neural networks to select the parameters for the prognostic of mild cognitive impairment and dementia in elderly individuals. *Computer Methods and Programs in Biomedicine*. v. 152, p. 93-104, 2017.
- LIU, Y.; BI, J. W.; FAN, Z. P.. Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms. *Expert Systems with Applications*. vol. 80, p. 323 – 339, 2017.

- LÓPEZ F. J. I.; SANZ I.P.; Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems with Applications*. v. 42, n. 6, p. 2857 – 2869, 2015.
- LOTERMAN G.; BROWN I.; MARTENS D.; MUES C.; BAESENS B.; Benchmarking regression algorithms for loss given default modeling . *International Journal of Forecasting*. v. 28,p. 161 – 170, 2012.
- LU Y.; ZHU J.; ZHANG N.; SHAO Q.; A hybrid switching PSO algorithm and support vector machines for bankruptcy prediction. *Proceedings - 2014 International Conference on Mechatronics and Control, ICMC 2014. Julho/ 2014*, p.1329 – 1333, 2014.
- LU, G.; REN, L.; KOLAGUNDA, A.; KAMBHAMETTU, C.. Neural Network shape: Organ shape representation with radial basis function neural networks. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing*. p. 932 – 936, 2016.
- LU, B.; SU, S. Accurate recognition of words in scenes without character segmentation using recurrent neural network. *Pattern Recognition*, v. 63, p. 397 – 405, 2017.
- LUJÁN, J. M.; CLIMENT, H.; GARCÍA-CUEVAS, L. M.; MORATAL, A. Volumetric efficiency modelling of internal combustion engines based on a novel adaptive learning algorithm of artificial neural networks. *Applied Thermal Engineering*. v. 123, p. 625-634, 2017.
- LUO, S.; KONG, X.; NIE T. Spline Based Survival Model for Credit Risk Modelling. *European Journal of Operational Research*, v. 253, n. 3, p. 869 – 879, 2016.
- LUO, C.; WU D.; WU D.; A deep learning approach for credit scoring using credit default swaps. *Engineering Applications of Artificial Intelligence*. v. 65, p. 465 – 470, 2017.
- MALDONADO S.; PÉREZ J.; BRAVO C.; Cost-based feature selection for Support Vector Machines: An application in credit scoring. *European Journal of Operational Research*. v. 261, n. 2, p. 656 – 665, 2017.
- MANNGÅRD, M.; KRONQVIST, J.; BÖLING, J. M. Structural learning in artificial neural networks using sparse optimization. *Neurocomputing*. v. 272, p. 660-667, 2018.
- MARTÍNEZ LÓPEZ, F. J.; ARRIAZA, J. A. T.; PUERTAS, S. M.; LÓPEZ, M. M. P.. Multilevel neuronal architecture to resolve classification problems with large training sets: Parallelization of the training process. *Journal of Computational Science*. vol. 16, p. 59 – 64, 2016.
- MARQUÉS A. I.; GARCÍA V.; SÁNCHEZ J. S.; Exploring the behaviour of base classifiers in credit scoring ensembles. *Expert Systems with Applications*. v. 39, p. 10244 – 10250, 2012.
- \_\_\_\_\_.; On the suitability of resampling techniques for the class imbalance problem in credit scoring. *Journal of the Operational Research Society*. v. 64, p. 1060 – 1070, 2013.
- MCCULLOCH, W. S.; PITTS, W. H. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 115 – 133, 1943.
- MESSIER, W. F.; HANSEN, J. V.. Inducing Rules for expert system development an example using default and bankruptcy data. *Management Science*. vol. 9, p. 253 – 266, 1985.
- MIRABDOLAZIMI, S. M.; SHAFABAKHSH, G. Rutting depth prediction of hot mix asphalts modified with forta fiber using artificial neural networks and genetic programming technique. *Construction and Building Materials*. v. 148, p. 666-674, 2017.

- MISHKIN, F. S.; EAKINS, S. G. *Financial Markets & Institutions*, 7th Ed. USA: Prentice Hall, 2012. 710p.
- MÓDOLO, M.. *Classificação automática de supernovas usando redes neurais artificiais*. 2016. Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais. São José dos Campos. 2016.
- MOHAMMED, M. A.; GHANI, M. K. A.; HAMED, R. I.; IBRAHIM, D. A.; ABDULLAH, M. K. Artificial neural networks for automatic segmentation and identification of nasopharyngeal carcinoma. *Journal of Computational Science*. v. 21, p. 263-274, 2017.
- MOORE, H. E.; BUTCHER, J. B.; DAY, C. R.; DRIJFHOUT, F. P. Adult fly age estimations using cuticular hydrocarbons and Artificial Neural Networks in forensically important Calliphoridae species. *Forensic Science International*. v. 280, p. 233 - 244, 2017.
- NAYAK R.K.; MISHRA D.; RATH A.K.; A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices. *Applied Soft Computing Journal*. v. 35, p. 670 – 680, 2015.
- NEJAD, F. P.; JAKSA, M. B. Load-settlement behavior modeling of single piles using artificial neural networks and CPT data. *Computers and Geotechnics*. v. 89, p. 9-21, 2017.
- NIKLIS D.; DOUMPOS M.; YOPOUNIDIS C.; Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines. *Applied Mathematics and Computation*. v. 234, p. 69 – 81, 2014.
- NIKOLIC N.; YARKIC-JOKSIMOVIC N.; STOJANOVSKI D.; JOKSIMOVIC I.; The application of brute force logistic regression to corporate credit scoring models: Evidence from Serbian financial statements. *Expert Systems with Applications*. v. 40, p. 5932 – 5944, 2013.
- ORESKI S.; ORESKI G.; Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications*. v. 41, n. 4, p. 2052 – 2064, 2014.
- ORESKI S.; ORESKI D.; ORESKI G.; Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Systems with Applications*. v. 39, p. 12605 – 12617, 2012.
- OZTEKIN A.; KIZILASLAN R.; FREUND S.; ISERI A.; A data analytic approach to forecasting daily stock returns in an emerging market. *European Journal of Operational Research*. v. 253, n. 3, p. 697 – 710, 2016.
- OZTURK H.; NAMLI E.; ERDAL H.I.; Modelling sovereign credit ratings: The accuracy of models in a heterogeneous sample. *Economic Modelling*. v. 54, p. 469 – 478, 2016.
- PATEL J.; SHAH S.; THAKKAR P.; KOTECHA K.; Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*. v. 42, n. 4, p. 2162 – 2172, 2015.
- PATEL J.; SHAH S.; THAKKAR P.; KOTECHA K.; Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications*. v. 42, n. 1, p. 259 – 268, 2015.
- PAVLENKO, T.; CHERNYAK, O.. Credit risk modeling using Bayesian networks. *International Journal of Intelligent Systems*. vol. 25, n. 4, p. 326 – 344, 2010.
- PAWLINA, W. *Ross Histologia Texto e Atlas: Correções com Biologia Celular e Molecular*, 7ª Edição, Rio de Janeiro: Guanabara Koogan, p. 361 – 362, 2016.

- PEHLIVANLI A.Ç.; AŞIKGIL B.; GÜLAY G.; Indicator selection with committee decision of filter methods for stock market price trend in ISE. *Applied Soft Computing Journal*. v.49, p.792 – 800, 2016.
- PEREIRA, S.; PINTO, A.; ALVES, V.; SILVA, C. A. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Transactions on Medical Imaging* v. 35, n. 5, p. 1240 – 1251, 2016.
- PUNNIYAMOORTHY M.; SRIDEVI P.; Identification of a standard AI based technique for credit risk analysis. *Benchmarking*. v. 23, n. 5, p. 1381 – 1390, 2016.
- QUILES, M. G. Sistema de Visão Baseado em Redes Neurais para o Controle de Robôs Móveis. 2004. Dissertação (Mestrado em Ciência da Computação e Matemática Computacional) – Universidade de São Paulo. São Paulo. 2004.
- QIU, M.; LI C.; SONG Y.; Application of the Artificial Neural Network in Predicting the Direction of Stock Market Index. *Proceedings - 2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS 2016*. 2016.
- QIU M.; SONG Y.; AKAGI F.; Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. *Chaos, Solitons and Fractals*. v. 85, p. 1 – 7, 2016.
- RAITH, S.; VOGEL, E. P.; ANEES, N.; KEUL, C.; GÜTH, J. F.; EDELHOFF, D.; FISCHER, H.. Artificial Neural Networks as a powerful numerical tool to classify specific features of a tooth based on 3D scan data. *Computers in biology and medicine*. vol. 80, p. 65 – 76, 2017.
- RAHMAN, F. A.; SHAMSUDDIN, S. M.; HASSAN, S.; HARIS, N. A.. A Review of KDD-Data Mining Framework and Its Application in Logistics and Transportation. *International Journal of Supply Chain Management*, v. 5, n. 2, p. 77-84, 2016.
- RESTORE. Using Statistical Regression Methods in Education Research. Disponível em: <<http://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/modules/mod4/11/index.html>>, 2011. Acesso em: 12 de dezembro de 2017.
- REZENDE, S. O. Sistemas Inteligentes: Fundamentos e Aplicações. 1. ed. Barueri: Manole, 2003. ISBN 85-204-1683-7.
- RIBEIRO B.; SILVA C.; CHEN N.; VIEIRA A.; CARVALHO DAS NEVES J.; Enhanced default risk models with SVM+. *Expert Systems with Applications*. v. 39, p. 10140 – 10152, 2012.
- RISTER, B.; RUBIN, D. L. Piecewise convexity of artificial neural networks. *Neural Networks*. v. 94, p. 34-45, 2017.
- ROSEMBLAT, F. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, p. 386 – 408, 1958.
- SÁNCHEZ-LASHERAS F.; DE ANDRÉS J.; LORCA P.; DE COS JUEZ F.J.; A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. *Expert Systems with Applications*. v. 39, p. 7512 – 7523, 2012.
- SÁNCHEZ, A. S.; RODRIGUES, D. A.; FONTES, R. M.; MARTINS M. F.; KALID R. A.; TORRES E. A. Wave resource characterization through in-situ measurement followed by artificial neural networks' modeling. *Renewable Energy*. v.115, p. 1055-1066, 2017.
- SANDS T.M.; TAYAL D.; MORRIS M.E.; MONTEIRO S.T.; Robust stock value prediction using support vector machines with particle swarm optimization. 2015 IEEE Congress on Evolutionary Computation, CEC 2015 - Proceedings. 2015.

- SANTHOSH, T.V.; GOPIKA, V.; GHOSH, A.K.; FERNANDES, B.G. An approach for reliability prediction of instrumentation & control cables by artificial neural networks and weibull theory for probabilistic safety assessment of NPPs. *Reliability Engineering & System Safety*. v. 170, p. 31-44, 2017.
- SELFRIDGE, O. G. Pandemonium: a paradigm for learning. *Mechanization of Thought Processes: Proceedings of a Symposium Held at the National Physical Laboratory, Londres*, p. 513 – 526, 1958.
- SELWYN, P.. Financial credit-risk evaluation with neural and neuralfuzzy systems. *European Journal of Operational Research*. vol. 112, p. 310 – 321, 1999.
- SERVIÇO DE PROTEÇÃO AO CRÉDITO. Indicador de Inadimplência de Pessoas Jurídicas – SPC Brasil e CNDL. Disponível em: <<https://www.spcbrasil.org.br/pesquisas/indice/2581>>. Acesso em: 09 de março de 2017.
- SHANMUGAM, R.; JOHNSON, C. At a crossroad of data envelopment and principal component analysis. *Omega*. vol. 35, n. 4, p. 351 – 364, 2007.
- SHAN R.; DAI H.; ZHAO J.; LIU W. Forecasting study of Shanghai's and Shenzhen's stock markets using a hybrid forecast method. *Communications in Statistics: Simulation and Computation*. v. 44, n. 4, p. 1066 – 1077, 2015.
- SHIVAKUMARA, P.; RAGHAVENDRA, R.; QIN, L.; RAJA, K. B.; LU, T.; PAL, U. A new multi-modal approach to bib number/text detection and recognition in Marathon images. *Pattern Recognition*. vol. 61, p. 479 – 491, 2017.
- SILVA, C.; VASCONCELOS G.; BARROS H.; FRANÇA G. Case-based reasoning combined with neural networks for credit risk analysis. *Proceedings of the International Joint Conference on Neural Networks*. 2015.
- SILVA, E. L.; MENEZES, E. M. *Metodologia da Pesquisa e Elaboração de Dissertação*. 4ª ed., UFSC, Florianópolis, SC, 138p, 2005.
- SILVA, M.; VELLASCO, M. M. B. R., CATALDO, E. Evolving Spiking Neural Networks for Recognition of Aged Voices. *Journal of Voice*, v. 31, n. 1, p.24 – 33, 2017.
- SIMMONS, S. M.; CAIRD, J. K.; STEEL, P. A meta-analysis of in-vehicle and nomadic voice-recognition system interaction and driving performance. *Accident Analysis & Prevention*. vol. 106, p. 31 – 43, 2017.
- SITTON, J. D.; ZEINALI, Y.; STORY B. A. Rapid soil classification using artificial neural networks for use in constructing compressed earth blocks. *Construction and Building Materials*. v. 138, p. 214-221, 2017.
- SLAVICI T.; MARIS S.; PIRTEA M. Usage of artificial neural networks for optimal bankruptcy forecasting. Case study: Eastern European small manufacturing enterprises. *Quality and Quantity*. v. 50, n. 1, p. 385 – 398, 2016.
- TAHIR, Z. R.; MANDAL, P. Artificial neural network prediction of buckling load of thin cylindrical shells under axial compression. *Engineering Structures*. v. 152, p. 843-855, 2017.
- TATAR, A.; NASERI, S.; SIRACH, N.; LEE, M.; BAHADORI, A.. Prediction of reservoir brine properties using radial basis functions (RBF) neural network. *Petroleum*. vol. 1, n. 4, p. 349 – 357, 2015.
- THE MATHWORKS INC. Radial Basis Neural Networks. USA, 2017b. Disponível em: <<https://www.mathworks.com/help/nnet/ug/radial-basis-neural-networks.html#bss38h1-6>> Acesso em: 11 de dezembro de 2017.
- THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*, Oxford: Elsevier, 4ª Edição, p. 1, 2009.

- TICKNOR J.L.; A Bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*. v. 40, n. 14, p. 5501 – 5506, 2013.
- TINOCO M. H.; WILSON N.; Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*. v. 30, p. 394 – 419, 2013.
- TONG E. N. C.; MUES C.; THOMAS L. C.; Mixture cure models in credit scoring: If and when borrowers default. *European Journal of Operational Research*. v. 218, p. 132 – 139, 2012.
- TONG G.; LI S.; Construction and Application Research of Isomap-RVM Credit Assessment Model. *Mathematical Problems in Engineering*. v. 2015, p. 1 – 8, 2015.
- TSAI C.-F.; CHENG K. C.; Simple instance selection for bankruptcy prediction. *Knowledge-Based Systems*. v. 27, p. 333 – 342, 2012.
- TSAI C.-F.; HSU Y.-F.; YEN D.C.; A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing Journal*. v. 24, p. 977 – 984, 2014.
- TSAI C.-F.; HUNG C.; Modeling credit scoring using neural network ensembles. *Kybernetes*. v. 43, n. 7, p. 1114 – 1123, 2014.
- VIDAKI, A.; BALLARD, D.; ALIFERI, A.; MILLER, T. H.; BARRON, L. P.; COURT, D. S. DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. *Forensic Science International: Genetics*. v. 28, p. 225-236, 2017.
- VIRAG M.; NYITRAI T.; Application of support vector machines on the basis of the first Hungarian bankruptcy model. *Society and Economy*. v. 35, n. 2, p. 227 – 248, 2013.
- WAGLE M.; YANG Z.; BENSLIMANE Y.; Bankruptcy prediction using data mining techniques. 2017 8th International Conference on Information and Communication Technology for Embedded Systems, IC-ICTES 2017 - Proceedings. 2017.
- WANG J.; HEDAR A.-R.; WANG S.; MA J.; Rough set and scatter search metaheuristic based feature selection for credit scoring. *Expert Systems with Applications*. v. 39, p. 6123 – 6128, 2012.
- WANG M.; CHEN H.; LI H.; CAI Z.; ZHAO X.; TONG C.; LI J.; XU X.; Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction. *Engineering Applications of Artificial Intelligence*. v. 63, p. 54 – 68, 2017.
- WEI L.-Y.; A GA-weighted ANFIS model based on multiple stock market volatility causality for TAIEX forecasting. *Applied Soft Computing Journal*. v. 13, p. 911 – 920, 2013.
- WENG B.; AHMED M. A.; MEGAHEDE F. M.; Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems with Applications*. v. 79, p. 153 – 163, 2017.
- WEST, D.. Neural Network Credit Scoring Models. *Computers & Operations Research*. vol. 27, n. 11, p. 1131 – 1152, 2000.
- WITTEN, I. H.; FRANK, E.; HALL, M. A.; PAL, C. J.. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 621 p., 2016.
- WU D. D.; ZHENG L.; OLSON D. L.; A decision support approach for online stock forum sentiment analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. v. 44, n. 8, p. 1077 – 1087, 2014.

- XI L.; MUZHOU H.; LEE M.H.; LI J.; WEI D.; HAI H.; WU Y.; A new constructive neural network method for noise processing and its application on stock market prediction. *Applied Soft Computing Journal*. v. 15, p. 57 – 66, 2014.
- XIONG T.; WANG S.; MAYERS A.; MONGA E.; Personal bankruptcy prediction by mining credit card data. *Expert Systems with Applications*. v. 40, p. 665 – 676, 2013.
- XIONG L.; LU Y.; Hybrid ARIMA-BPNN model for time series prediction of the Chinese stock market. 2017 3rd International Conference on Information Management, ICIM 2017. 2017.
- YADAV, N., YADAV, A., KUMAR, M. An introduction to Neural Networks Methods for Differential Equations. Springer, p. 21, 2015.
- YANG, T.; ASANJAN, A. A.; FARIDZAD, M.; HAYATBINI, N.; GAO, X.; SOROOSHIAN, S. An enhanced artificial neural network with a shuffled complex evolutionary global optimization with principal component analysis. *Information Sciences*. v.s 418–419, p. 302-316, 2017.
- YE, J.; FU, J.. Multi-period medical diagnosis method using a single valued neutrosophic similarity measure based on tangent function. *Computer methods and programs in biomedicine*. vol. 123, p.142 – 149, 2016.
- YI B.; YHU J.; Credit scoring with an improved fuzzy support vector machine based on grey incidence analysis. *Proceedings of IEEE International Conference on Grey Systems and Intelligent Services, GSIS*. p. 173 – 178, 2015.
- YU Y.; SHANSHAN W.; LIJUN Z.; Stock price forecasting based on BP neural network model of network public opinion. 2017 2nd International Conference on Image, Vision and Computing, ICIVC 2017. 2017.
- ZAKIROV D.; BONDAREV A.; MOMTSELIDYE N.; A comparison of data mining techniques in evaluating retail credit scoring using R programming. *Proceedings of the 2015 12th International Conference on Electronics Computer and Computation, ICECCO 2015*. 2016.
- ZHANG, D.; ZHOU, X.; LEUNG, S. C.; ZHENG, J.. Vertical bagging decision trees model for credit scoring. *Expert Systems with Applications*. vol. 37, n. 12, p. 7838 – 7843, 2010.
- ZHANG Y.; 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.
- ZHANG X.; HU Y.; XIE K.; ZHANG W.; SU L.; LIU M.; An evolutionary trend reversion model for stock trading rule discovery. *Knowledge-Based Systems*. v. 79, p. 27 – 35, 2015.
- ZHANG X.-D.; LI A.; PAN R.; Stock trend prediction based on a new status box method and AdaBoost probabilistic support vector machine. *Applied Soft Computing Journal*. v. 49, p. 385 – 398, 2016.
- ZHANG, F.; TADIKAMALLA, P. R.; SHANG, J. Corporate credit-risk evaluation system: Integrating explicit and implicit financial performances. *International Journal of Production Economics*. vol. 177, p. 77 – 100, 2016.
- ZHANG, Z.; GAO, G.; SHI, Y.. Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzyfication and penalty factors. *European Journal of Operational Research*. vol. 237, n. 1, p. 335 – 348, 2014.
- ZHAO Y.; XU S.; KANG B. H.; KABIR M. M. J.; LIU Y.; WASINGER R. Investigation and improvement of multi-layer perception neural networks for credit scoring. *Expert Systems with Applications*. v. 42, n. 7, p. 3508 – 3516, 2015.

- 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.
- ZHONG X.; ENKE D.; Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications*. v. 67, p. 126 – 139, 2017.
- ZHU X.; LI J.; WU D.; WANG H.; LIANG C.; Balancing accuracy, complexity and interpretability in consumer credit decision making: A C-TOPSIS classification approach. *Knowledge-Based Systems*. v. 53, p. 259 – 267, 2013.

## APÊNDICE A – CÓDIGO DAS VARIÁVEIS CATEGÓRICAS

**Tabela 22 - Variáveis Categóricas da RL no modelo: Adimplentes ou Outra**

<i>Variáveis Categóricas</i>	<i>Código na</i>		<i>Código do</i>
	<i>Base de Dados</i>	<i>Frequência</i>	<i>Parâmetro</i>
SOCIEDADE ANONIMA - CAP FECHADO	0	5418	,000
	1	14	1,000
SOCIEDADE CIVIL COM FINS LUCRATIVOS	0	5416	,000
	1	16	1,000
SOCIEDADE EM NOME COLETIVO	0	5431	,000
	1	1	1,000
SOCIEDADE LIMITADA	0	1320	,000
	1	4112	1,000
SOCIEDADE SIMPLES	0	5269	,000
	1	163	1,000
ASSOCIACAO CIVIL	0	5429	,000
	1	3	1,000
AUTARQUIA ESTADUAL	0	5431	,000
	1	1	1,000
EMPRESA INDIVIDUAL DE RESPONSABIL.LTDA	0	5113	,000
	1	319	1,000
EMPRESARIO INDIVIDUAL	0	4976	,000
	1	456	1,000
FIRMA INDIVIDUAL	0	5362	,000
	1	70	1,000
LTDA - SOC COTAS DE RESPONSABILIDADE LI	0	5214	,000
	1	218	1,000
MICROEMPREENDEDOR INDIVIDUAL - EI	0	5373	,000
	1	59	1,000
FAT1	0	4279	,000
	1	1153	1,000
FAT2	0	4054	,000
	1	1378	1,000
FAT3	0	3925	,000
	1	1507	1,000
FAT4	0	4038	,000
	1	1394	1,000
FUNC1	0	3784	,000
	1	1648	1,000
FUNC2	0	4268	,000
	1	1164	1,000
FUNC3	0	4137	,000

	1	1295	1,000
FUNC4	0	4107	,000
	1	1325	1,000
A	0	5422	,000
	1	10	1,000
B	0	5427	,000
	1	5	1,000
C	0	4575	,000
	1	857	1,000
D	0	5431	,000
	1	1	1,000
E	0	5409	,000
	1	23	1,000
F	0	5149	,000
	1	283	1,000
G	0	3145	,000
	1	2287	1,000
H	0	5103	,000
	1	329	1,000
I	0	5015	,000
	1	417	1,000
J	0	5273	,000
	1	159	1,000
K	0	5396	,000
	1	36	1,000
L	0	5390	,000
	1	42	1,000
M	0	5138	,000
	1	294	1,000
N	0	5102	,000
	1	330	1,000
O	0	5431	,000
	1	1	1,000
P	0	5326	,000
	1	106	1,000
Q	0	5314	,000
	1	118	1,000
R	0	5386	,000
	1	46	1,000
S	0	5344	,000
	1	88	1,000
CONSTITUT1	0	4074	,000
	1	1358	1,000
CONSTITUT2	0	4073	,000

	1	1359	1,000
CONSTITUT3	0	4074	,000
	1	1358	1,000
CONSTITUT4	0	4075	,000
	1	1357	1,000
DESDE1	0	4076	,000
	1	1356	1,000
DESDE2	0	4073	,000
	1	1359	1,000
DESDE3	0	4072	,000
	1	1360	1,000
DESDE4	0	4075	,000
	1	1357	1,000
PEQUENAEMPRESA	0	2681	,000
	1	2751	1,000
EMPRESA	0	4762	,000
	1	670	1,000
MICROEMPRESA	0	3522	,000
	1	1910	1,000
EMPRESARIALMIDDLE	0	5334	,000
	1	98	1,000
PJNÃOCLASSIFICÁVEL	0	5431	,000
	1	1	1,000
CORPORATE	0	5431	,000
	1	1	1,000
ESTADUAL	0	5431	,000
	1	1	1,000
RISCO_A	0	4369	,000
	1	1063	1,000
RISCO_B	0	3051	,000
	1	2381	1,000
RISCO_C	0	4438	,000
	1	994	1,000
RISCO_D	0	5336	,000
	1	96	1,000
RISCO_E	0	4534	,000
	1	898	1,000
AG.VAREJO-REDE CRÉDITO	0	5430	,000
Limite	1	2	1,000
AGÊNCIAS/ESCRIT.PRIVATE-CRÉDITO	0	3562	,000
Limite	1	1870	1,000
AGÊNCIA-SISTEMA-CRÉDITO	0	2643	,000
Limite	1	2789	1,000
	0	5431	,000

DICRE/8554-COMITÊ LIM.CRÉDITO LARGE CORPORATE Limite	1	1	1,000
DICRE/8918-LC.ESPEC./SUBCOM "B" Limite	0	5429	,000
	1	3	1,000
DICRE/8922-LIM.CRÉD.UPPER MID./MIDDLE-SUBCOMITÊ "A" Limite	0	5428	,000
	1	4	1,000
DICRE/8922-UPPER MIDDLE/MIDDLE-SUBCOMITÊ "B" Limite	0	5428	,000
	1	4	1,000
DICRE/8922-UPPER MIDDLE/MIDDLE-SUBCOMITÊ "C" Limite	0	5267	,000
	1	165	1,000
DICRE/9570-LIM.CRÉD.SUBCOMITÊ "A" Limite	0	5430	,000
	1	2	1,000
DICRE/9969-LC.VAR/MPE/SUBCOM A Limite	0	5420	,000
	1	12	1,000
DICRE/9969-LC.VAR/MPE/SUBCOM B Limite	0	5402	,000
	1	30	1,000
DICRE/9969-LC.VAR/MPE/SUBCOM C Limite	0	4882	,000
	1	550	1,000
RESTRIÇÃO FRACA/INFO	0	4342	,000
	1	1090	1,000
RESTRIÇÃO IMPEDITIVA	0	3427	,000
	1	2005	1,000
SEM RESTRIÇÕES	0	3095	,000
	1	2337	1,000

Fonte: A autora, 2017.

**Tabela 23 - Variáveis Categóricas da RL no modelo: Inadimplentes ou Outra**

<i>Variáveis Categóricas</i>	<i>Código na Base de Dados</i>	<i>Frequência</i>	<i>Código do Parâmetro</i>
SOCIEDADE ANONIMA - CAP FECHADO	0	5418	,000
	1	14	1,000
SOCIEDADE CIVIL COM FINS LUCRATIVOS	0	5416	,000
	1	16	1,000
SOCIEDADE EM NOME COLETIVO	0	5431	,000
	1	1	1,000
SOCIEDADE LIMITADA	0	1320	,000
	1	4112	1,000
SOCIEDADE SIMPLES	0	5269	,000
	1	163	1,000
ASSOCIACAO CIVIL	0	5429	,000
	1	3	1,000
AUTARQUIA ESTADUAL	0	5431	,000

	1	1	1,000
EMPRESA INDIVIDUAL DE RESPONSABIL.LTDA	0	5113	,000
	1	319	1,000
EMPRESARIO INDIVIDUAL	0	4976	,000
	1	456	1,000
FIRMA INDIVIDUAL	0	5362	,000
	1	70	1,000
LTDA - SOC COTAS DE RESPONSABILIDADE LI	0	5214	,000
	1	218	1,000
MICROEMPREENDEDOR INDIVIDUAL - EI	0	5373	,000
	1	59	1,000
FAT1	0	4279	,000
	1	1153	1,000
FAT2	0	4054	,000
	1	1378	1,000
FAT3	0	3925	,000
	1	1507	1,000
FAT4	0	4038	,000
	1	1394	1,000
FUNC1	0	3784	,000
	1	1648	1,000
FUNC2	0	4268	,000
	1	1164	1,000
FUNC3	0	4137	,000
	1	1295	1,000
FUNC4	0	4107	,000
	1	1325	1,000
A	0	5422	,000
	1	10	1,000
B	0	5427	,000
	1	5	1,000
C	0	4575	,000
	1	857	1,000
D	0	5431	,000
	1	1	1,000
E	0	5409	,000
	1	23	1,000
F	0	5149	,000
	1	283	1,000
G	0	3145	,000
	1	2287	1,000
H	0	5103	,000
	1	329	1,000
I	0	5015	,000

	1	417	1,000
J	0	5273	,000
	1	159	1,000
K	0	5396	,000
	1	36	1,000
L	0	5390	,000
	1	42	1,000
M	0	5138	,000
	1	294	1,000
N	0	5102	,000
	1	330	1,000
O	0	5431	,000
	1	1	1,000
P	0	5326	,000
	1	106	1,000
Q	0	5314	,000
	1	118	1,000
R	0	5386	,000
	1	46	1,000
S	0	5344	,000
	1	88	1,000
CONSTITUT1	0	4074	,000
	1	1358	1,000
CONSTITUT2	0	4073	,000
	1	1359	1,000
CONSTITUT3	0	4074	,000
	1	1358	1,000
CONSTITUT4	0	4075	,000
	1	1357	1,000
DESDE1	0	4076	,000
	1	1356	1,000
DESDE2	0	4073	,000
	1	1359	1,000
DESDE3	0	4072	,000
	1	1360	1,000
DESDE4	0	4075	,000
	1	1357	1,000
PEQUENAEMPRESA	0	2681	,000
	1	2751	1,000
EMPRESA	0	4762	,000
	1	670	1,000
MICROEMPRESA	0	3522	,000
	1	1910	1,000
EMPRESARIALMIDDLE	0	5334	,000

	1	98	1,000
PJNÃOCLASSIFICÁVEL	0	5431	,000
	1	1	1,000
CORPORATE	0	5431	,000
	1	1	1,000
ESTADUAL	0	5431	,000
	1	1	1,000
RISCO_A	0	4369	,000
	1	1063	1,000
RISCO_B	0	3051	,000
	1	2381	1,000
RISCO_C	0	4438	,000
	1	994	1,000
RISCO_D	0	5336	,000
	1	96	1,000
RISCO_E	0	4534	,000
	1	898	1,000
AG.VAREJO-REDE CRÉDITO	0	5430	,000
Limite	1	2	1,000
AGÊNCIAS/ESCRIT.PRIVATE-CRÉDITO	0	3562	,000
Limite	1	1870	1,000
AGÊNCIA-SISTEMA-CRÉDITO	0	2643	,000
Limite	1	2789	1,000
DICRE/8554-COMITÊ LIM.CRÉDITO LARGE	0	5431	,000
CORPORATE Limite	1	1	1,000
DICRE/8918-LC.ESPEC./SUBCOM "B"	0	5429	,000
Limite	1	3	1,000
DICRE/8922-LIM.CRÉD.UPPER MID./MIDDLE-	0	5428	,000
SUBCOMITÊ "A" Limite	1	4	1,000
DICRE/8922-UPPER MIDDLE/MIDDLE-	0	5428	,000
SUBCOMITÊ "B" Limite	1	4	1,000
DICRE/8922-UPPER MIDDLE/MIDDLE-	0	5267	,000
SUBCOMITÊ "C" Limite	1	165	1,000
DICRE/9570-LIM.CRÉD.SUBCOMITÊ "A"	0	5430	,000
Limite	1	2	1,000
DICRE/9969-LC.VAR/MPE/SUBCOM A	0	5420	,000
Limite	1	12	1,000
DICRE/9969-LC.VAR/MPE/SUBCOM B	0	5402	,000
Limite	1	30	1,000
DICRE/9969-LC.VAR/MPE/SUBCOM C	0	4882	,000
Limite	1	550	1,000
RESTRIÇÃO FRACA/INFO	0	4342	,000
	1	1090	1,000
RESTRIÇÃO IMPEDITIVA	0	3427	,000

SEM RESTRIÇÕES	1	2005	1,000
	0	3095	,000
	1	2337	1,000

Fonte: A autora, 2017.

**Tabela 24 - Variáveis Categóricas da RL no modelo: Temporariamente Inadimplentes ou Outra**

<i>Variáveis Categóricas</i>	<i>Código na Base de Dados</i>	<i>Frequência</i>	<i>Código do Parâmetro</i>
SOCIEDADE ANONIMA - CAP FECHADO	0	5418	,000
	1	14	1,000
SOCIEDADE CIVIL COM FINS LUCRATIVOS	0	5416	,000
	1	16	1,000
SOCIEDADE EM NOME COLETIVO	0	5431	,000
	1	1	1,000
SOCIEDADE LIMITADA	0	1320	,000
	1	4112	1,000
SOCIEDADE SIMPLES	0	5269	,000
	1	163	1,000
ASSOCIACAO CIVIL	0	5429	,000
	1	3	1,000
AUTARQUIA ESTADUAL	0	5431	,000
	1	1	1,000
EMPRESA INDIVIDUAL DE RESPONSABIL.LTDA	0	5113	,000
	1	319	1,000
EMPRESARIO INDIVIDUAL	0	4976	,000
	1	456	1,000
FIRMA INDIVIDUAL	0	5362	,000
	1	70	1,000
LTDA - SOC COTAS DE RESPONSABILIDADE LI	0	5214	,000
	1	218	1,000
MICROEMPREENDEDOR INDIVIDUAL - EI	0	5373	,000
	1	59	1,000
FAT1	0	4279	,000
	1	1153	1,000
FAT2	0	4054	,000
	1	1378	1,000
FAT3	0	3925	,000
	1	1507	1,000
FAT4	0	4038	,000
	1	1394	1,000
FUNC1	0	3784	,000
	1	1648	1,000

FUNC2	0	4268	,000
	1	1164	1,000
FUNC3	0	4137	,000
	1	1295	1,000
FUNC4	0	4107	,000
	1	1325	1,000
A	0	5422	,000
	1	10	1,000
B	0	5427	,000
	1	5	1,000
C	0	4575	,000
	1	857	1,000
D	0	5431	,000
	1	1	1,000
E	0	5409	,000
	1	23	1,000
F	0	5149	,000
	1	283	1,000
G	0	3145	,000
	1	2287	1,000
H	0	5103	,000
	1	329	1,000
I	0	5015	,000
	1	417	1,000
J	0	5273	,000
	1	159	1,000
K	0	5396	,000
	1	36	1,000
L	0	5390	,000
	1	42	1,000
M	0	5138	,000
	1	294	1,000
N	0	5102	,000
	1	330	1,000
O	0	5431	,000
	1	1	1,000
P	0	5326	,000
	1	106	1,000
Q	0	5314	,000
	1	118	1,000
R	0	5386	,000
	1	46	1,000
S	0	5344	,000
	1	88	1,000

CONSTITUT1	0	4074	,000
	1	1358	1,000
CONSTITUT2	0	4073	,000
	1	1359	1,000
CONSTITUT3	0	4074	,000
	1	1358	1,000
CONSTITUT4	0	4075	,000
	1	1357	1,000
DESDE1	0	4076	,000
	1	1356	1,000
DESDE2	0	4073	,000
	1	1359	1,000
DESDE3	0	4072	,000
	1	1360	1,000
DESDE4	0	4075	,000
	1	1357	1,000
PEQUENAEMPRESA	0	2681	,000
	1	2751	1,000
EMPRESA	0	4762	,000
	1	670	1,000
MICROEMPRESA	0	3522	,000
	1	1910	1,000
EMPRESARIALMIDDLE	0	5334	,000
	1	98	1,000
PJNÃOCLASSIFICÁVEL	0	5431	,000
	1	1	1,000
CORPORATE	0	5431	,000
	1	1	1,000
ESTADUAL	0	5431	,000
	1	1	1,000
RISCO_A	0	4369	,000
	1	1063	1,000
RISCO_B	0	3051	,000
	1	2381	1,000
RISCO_C	0	4438	,000
	1	994	1,000
RISCO_D	0	5336	,000
	1	96	1,000
RISCO_E	0	4534	,000
	1	898	1,000
AG.VAREJO-REDE CRÉDITO	0	5430	,000
Limite	1	2	1,000
AGÊNCIAS/ESCRIT.PRIVATE-CRÉDITO	0	3562	,000
Limite	1	1870	1,000

AGÊNCIA-SISTEMA-CRÉDITO	0	2643	,000
Limite	1	2789	1,000
DICRE/8554-COMITÊ LIM.CRÉDITO LARGE	0	5431	,000
CORPORATE Limite	1	1	1,000
DICRE/8918-LC.ESPEC./SUBCOM "B"	0	5429	,000
Limite	1	3	1,000
DICRE/8922-LIM.CRÉD.UPPER MID./MIDDLE-	0	5428	,000
SUBCOMITÊ "A" Limite	1	4	1,000
DICRE/8922-UPPER MIDDLE/MIDDLE-	0	5428	,000
SUBCOMITÊ "B" Limite	1	4	1,000
DICRE/8922-UPPER MIDDLE/MIDDLE-	0	5267	,000
SUBCOMITÊ "C" Limite	1	165	1,000
DICRE/9570-LIM.CRÉD.SUBCOMITÊ "A"	0	5430	,000
Limite	1	2	1,000
DICRE/9969-LC.VAR/MPE/SUBCOM A	0	5420	,000
Limite	1	12	1,000
DICRE/9969-LC.VAR/MPE/SUBCOM B	0	5402	,000
Limite	1	30	1,000
DICRE/9969-LC.VAR/MPE/SUBCOM C	0	4882	,000
Limite	1	550	1,000
RESTRIÇÃO FRACA/INFO	0	4342	,000
	1	1090	1,000
RESTRIÇÃO IMPEDITIVA	0	3427	,000
	1	2005	1,000
SEM RESTRIÇÕES	0	3095	,000
	1	2337	1,000

Fonte: A autora, 2017.

## APÊNDICE B – VARIÁVEIS PRESENTES NA REGRESSÃO

A seguir são apresentadas as variáveis presentes nas equações da regressão dos modelos: “Adimplentes ou Outra”, “Inadimplentes ou Outra”, e “Temporariamente Inadimplentes ou Outra”, respectivamente (Tabela 25, Tabela 26 e Tabela 27, respectivamente)

Com relação à Tabela 25, observa-se na coluna “Sig.” valores inferiores a 0,05 para o intervalo de Faturamento Bruto Anual 1, seção do código CNAE H, seção do código CNAE I, seção do código CNAE Q, e para todos os tipos de riscos considerados na regressão. Desta forma, essas foram as variáveis que contribuem significativamente para a capacidade de previsão do modelo. Ou seja, no momento que o faturamento entre R\$ 0,00 e R\$536200,00; as seções do CNAE sejam H, I ou Q; e sejam apresentados riscos A, B, C ou D de concessão de empréstimos; o modelo proposto é mais facilmente influenciado na classificação para as classes Adimplentes ou Outros.

Realizando o mesmo procedimento para a Tabela 26, os valores de “Sig.” inferiores a 0,05, ou seja, variáveis que mais contribuem para o modelo, são as seções de CNAE C, F, H, J, L; os riscos associados ao empréstimo considerados na regressão e ainda as restrições de empréstimo dadas ao cliente. Assim, empresas que demonstrem as condições demonstradas fazem com que o modelo proposto seja influenciado para esse tipo de classificação (Inadimplentes ou Outros).

E, finalmente, observando a Tabela 27, verificamos que os valores de “Sig.” inferiores a 0,05 estão presentes para as variáveis: Seções CNAE (C; F; G; H; I; J; L; M; N; P; Q), para todos os riscos de empréstimos utilizados pelo modelo proposto, e para a restrição impeditiva especificamente. Desta forma, os clientes do banco que demonstrarem estas condições influenciam o modelo para a classificação do tipo “Temporariamente Inadimplentes ou Outra”.







	DICRE9570LIM.CRÉD.SUBCOMITÊAL imite(1)	-42.412	36105,038	,000	1	,999	,000	,000	.
	DICRE9969LC.VARMPESUBCOMALi imite(1)	-21.906	22723,699	,000	1	,999	,000	,000	.
	DICRE8922UPPERMIDDLEMIDDLESU BCOMITÊBLimite(1)	-20.388	22723,699	,000	1	,999	,000	,000	.
	AG.VAREJOREDECRÉDITOLimite(1)	-41.272	35381,802	,000	1	,999	,000	,000	.
Restrição do Cliente	RESTRICÃOIMPEDITIVA(1)	-1,105	,095	134,263	1	,000	,331	,275	,399
	SEMRESTRICÕES(1)	,053	,083	,415	1	,520	1,055	,897	1,241
	Constante	40,924	46094,025	,000	1	,999	5,93E+17		

Fonte: A autora, 2017.

Tabela 26 - Variáveis da Regressão do modelo: Inadimplentes ou Outra

Atributo	Variável da Equação	Coeficiente na Regressão	S.E.	Wald	Graus de Liberdade	Sig.	95% I.C para EXP(B)		
							Exp(B)	Lower	Upper
Natureza Jurídica	SOCIEDADELIMITADA(1)	19,770	60794,621	,000	1	1,000	3,85E+08	,000	,
	LTDASOCOTASDERESPONSABILIDADELI(1)	20,277	60794,621	,000	1	1,000	6,40E+08	,000	,
	MICROEMPREENDEDORINDIVIDUALEI(1)	20,141	60794,621	,000	1	1,000	5,59E+08	,000	,
	SOCIEDADESIMPLES(1)	19,220	60794,621	,000	1	1,000	2,22E+08	,000	,
	EMPRESAINDIVIDUALDERESPONSABIL.LTDA(1)	19,760	60794,621	,000	1	1,000	3,82E+08	,000	,
	EMPRESARIOINDIVIDUAL(1)	19,565	60794,621	,000	1	1,000	3,14E+08	,000	,
	FIRMAINDIVIDUAL(1)	18,615	60794,621	,000	1	1,000	1,22E+08	,000	,
	SOCIEDADEANONIMACAPFECHADO(1)	18,786	60794,621	,000	1	1,000	1,44E+08	,000	,
	SOCIEDADECIVILCOMFINSLUCRATIVOS(1)	17,069	60794,621	,000	1	1,000	2,59E+07	,000	,
	SOCIEDADEENOMEINCOLETIVO(1)	3,275	72879,770	,000	1	1,000	2,64E+01	,000	,
ASSOCIACAOCIVIL(1)	4,195	64963,791	,000	1	1,000	6,63E+01	,000	,	
Faturamento Bruto Anual	FAT1(1)	-,178	,271	,431	1	,511	8,37E-01	,492	1,423
	FAT2(1)	-,314	,207	2,300	1	,129	7,31E-01	,487	1,096
	FAT3(1)	,145	,165	,777	1	,378	1,16E+00	,837	1,598
# de Funcionários	FUNC1(1)	,314	,164	3,656	1	,056	1,37E+00	,992	1,888
	FUNC2(1)	,128	,171	,563	1	,453	1,14E+00	,813	1,589
	FUNC3(1)	,088	,154	,326	1	,568	1,09E+00	,808	1,475

A(1)	-,637	1,456	,192	1	,662	5,29E-01	,030	9,169
B(1)	-,249	1,732	,021	1	,886	7,80E-01	,026	23,227
C(1)	1,018	,453	5,039	1	,025	2,77E+00	1,138	6,727
D(1)	-17,520	40192,970	,000	1	1,000	2,46E-08	,000	.
E(1)	,631	,777	,660	1	,417	1,88E+00	,410	8,617
F(1)	1,062	,486	4,767	1	,029	2,89E+00	1,115	7,501
G(1)	,822	,443	3,451	1	,063	2,28E+00	,956	5,419
H(1)	1,038	,474	4,786	1	,029	2,82E+00	1,114	7,155
I(1)	,388	,478	,661	1	,416	1,47E+00	,578	3,762
J(1)	1,013	,506	4,006	1	,045	2,75E+00	1,021	7,424
K(1)	-,090	,857	,011	1	,917	9,14E-01	,170	4,901
L(1)	1,629	,694	5,515	1	,019	5,10E+00	1,309	19,861
M(1)	,685	,496	1,902	1	,168	1,98E+00	,749	5,247
N(1)	,820	,478	2,943	1	,086	2,27E+00	,890	5,799
P(1)	,875	,573	2,331	1	,127	2,40E+00	,780	7,375
Q(1)	-,207	,644	,104	1	,748	8,13E-01	,230	2,874
R(1)	,814	,677	1,444	1	,229	2,26E+00	,599	8,501
CONSTITUT1(1)	-,028	,186	,022	1	,882	9,73E-01	,675	1,401
CONSTITUT2(1)	,102	,172	,352	1	,553	1,11E+00	,791	1,550
CONSTITUT3(1)	,304	,155	3,856	1	,050	1,36E+00	1,001	1,835
DESDE1(1)	-,204	,193	1,117	1	,291	8,16E-01	,559	1,190
DESDE2(1)	-,171	,169	1,027	1	,311	8,43E-01	,605	1,173
DESDE3(1)	-,102	,144	,507	1	,476	9,03E-01	,681	1,196
PEQUENAEMPRESA(1)	,742	45718,636	,000	1	1,000	2,10E+00	,000	.
EMPRESA(1)	,314	45718,636	,000	1	1,000	1,37E+00	,000	.
MICROEMPRESA(1)	,324	45718,636	,000	1	1,000	1,38E+00	,000	.
Seção do Código CNAE								
	Data de Constituição							
	Cliente Desde							
	Segmento							

	EMPRESARIALMIDDLE(1)	-,593	45718,636	,000	1	1,000	5,53E-01	,000	.
	PINÃOCLASSIFICAVEL(1)	-20,096	60874,202	,000	1	1,000	1,87E-09	,000	.
	RISCO_A(1)	-3,932	,233	284,442	1	,000	1,96E-02	,012	,031
	RISCO_B(1)	-3,027	,137	487,124	1	,000	4,85E-02	,037	,063
	RISCO_C(1)	-2,772	,148	349,117	1	,000	6,25E-02	,047	,084
	RISCO_D(1)	-2,703	,272	98,382	1	,000	6,70E-02	,039	,114
	AGÊNCIASISTEMACRÉDITOLimite(1)	18,987	21614,126	,000	1	,999	1,76E+08	,000	.
	AGÊNCIASESCRIT.PRIVATECRÉDITOLimite(1)	19,351	21614,126	,000	1	,999	2,54E+08	,000	.
	DICRE9969LC.VARMPESUBCOMCLimite(1)	19,637	21614,126	,000	1	,999	3,38E+08	,000	.
	DICRE8922LIM.CRÉD.UPPERMID.MIDDLESUBCOMITÉALimite(1)	16,606	21614,126	,000	1	,999	1,63E+07	,000	.
	DICRE9969LC.VARMPESUBCOMBLimite(1)	18,736	21614,126	,000	1	,999	1,37E+08	,000	.
	DICRE8922UPPERMIDDLEMIDDLESUBCOMITÉLImite(1)	19,970	21614,126	,000	1	,999	4,71E+08	,000	.
	DICRE9570LIM.CRÉD.SUBCOMITÉALimite(1)	2,549	35468,771	,000	1	1,000	1,28E+01	,000	.
	DICRE9969LC.VARMPESUBCOMALimite(1)	20,414	21614,126	,000	1	,999	7,34E+08	,000	.
	DICRE8922UPPERMIDDLEMIDDLESUBCOMITÉBLimite(1)	18,377	21614,126	,000	1	,999	9,57E+07	,000	.
	AG.VAREJOREDECRÉDITOLimite(1)	40,955	35219,894	,000	1	,999	6,11E+17	,000	.
	RESTRICÃOIMPEDITIVA(1)	3,817	,295	167,219	1	,000	4,55E+01	25,505	81,132
	SEMRESTRICÕES(1)	1,281	,307	17,392	1	,000	3,60E+00	1,972	6,573
	Constante	-41,866	45527,390	,000	1	,999	6,58E-19		

Fonte: A autora, 2017.





PEQUENAEMPRESA(1)	,090	46227,887	,000	1	1,000	1,09E+00	,000	.
EMPRESA(1)	-1,115	46227,887	,000	1	1,000	8,92E-01	,000	.
MICROEMPRESA(1)	-2,262	46227,887	,000	1	1,000	7,70E-01	,000	.
EMPRESARIALMIDDLE(1)	-3,303	46227,887	,000	1	1,000	7,39E-01	,000	.
PJNÃOCLASSIFICÁVEL(1)	-20,189	61257,590	,000	1	1,000	1,71E-09	,000	.
RISCO_A(1)	1,083	,163	44,223	1	,000	2,95E+00	2,146	4,064
RISCO_B(1)	1,713	,145	140,009	1	,000	5,55E+00	4,175	7,365
RISCO_C(1)	1,987	,150	174,510	1	,000	7,29E+00	5,430	9,791
RISCO_D(1)	2,090	,252	68,643	1	,000	8,09E+00	4,931	13,257
AGÊNCIASISTEMACRÉDITOLimite(1)	20,356	22852,690	,000	1	,999	6,93E+08	,000	.
AGÊNCIASESCRIT.PRIVATECRÉDITOLimite(1)	20,505	22852,690	,000	1	,999	8,04E+08	,000	.
DICRE9969LC.VARMPESUBCOMCLimite(1)	20,729	22852,690	,000	1	,999	1,01E+09	,000	.
DICRE8922LIM.CRÉD.UPPERMID.MIDDLESUBCOMITÊALimite(1)	22,862	22852,690	,000	1	,999	8,49E+09	,000	.
DICRE9969LC.VARMPESUBCOMBLimite(1)	21,673	22852,690	,000	1	,999	2,58E+09	,000	.
DICRE8922UPPERMIDDLEMIDDLESUBCOMITÊCLimite(1)	20,281	22852,690	,000	1	,999	6,43E+08	,000	.
DICRE9570LIM.CRÉD.SUBCOMITÊALimite(1)	41,867	36374,148	,000	1	,999	1,52E+18	,000	.
DICRE9969LC.VARMPESUBCOMALimite(1)	20,705	22852,690	,000	1	,999	9,82E+08	,000	.
DICRE8922UPPERMIDDLEMIDDLESUBCOMITÊBLimite(1)	22,105	22852,690	,000	1	,999	3,98E+09	,000	.
AG.VAREJOREDECRÉDITOLimite(1)	,685	36306,236	,000	1	1,000	1,98E+00	,000	.
RESTRICÃOIMPEDITIVA(1)	-4,478	,097	24,321	1	,000	6,20E-01	,513	,750
SEMRESTRICÕES(1)	-1,143	,083	2,961	1	,085	8,67E-01	,737	1,020
Constante	-42,184	45978,375	,000	1	,999	4,78E-19		

Fonte: A autora, 2017.