

Classification Algorithms in Financial Application: Credit Risk Analysis on Legal Entities

F. Assef, M. Steiner, P. Steiner, and D. Franco

Abstract—This research aims at analyzing bank credit of legal entity (in non-default, default and temporarily default), for the purpose of assisting the decision made by the analyst of this area. For that, we used Artificial Neural Networks (ANNs), more specifically, the Multilayer Perceptron (MLP) and the Radial Basis Functions (RBF) and, also, the statistical model of Logistic Regression (LR). For the implementation of the ANNs and LR, the softwares MATLAB and SPSS were used, respectively. For the simulations developed 5.432 data with 15 attributes were collected by the experts of the institution bank (called “XYZ”). The results show that the default clients are easily identifiable, but for the non-delinquent clients and for the temporarily defaulters, the techniques had greater difficulty in the discrimination, suggesting that they are no so discriminants. The main contributions of this work are: the analysis of three classes of clients (non-default, default and temporarily default), rather than just two (non-default and default) as is usually done; the coding of variables (attributes) of the company XYZ aiming to maximize the accuracy of the techniques and the use of the one-against all method, little used by the researchers of this research area. This work presents new insights towards research over Credit Risk Assessment showing other possibilities of client classification and codification, allowing different types of studies to take place.

Index Terms— Artificial Neural Networks (ANNs), Credit Risk Assessment, Logistic Regression, Multilayer Perceptron, Radial Basis Functions.

I. INTRODUÇÃO

Avaliação de risco de crédito propõe um problema clássico de tomada de decisão, onde o mérito de determinado indivíduo de receber (ou não) o crédito é baseado em estimativas do potencial de devolução deste crédito acrescido pelas taxas bancárias [1].

Desde que foi introduzida na década de 50, a avaliação de risco tem sido amplamente utilizada [2]. Pesquisadores das mais diversas áreas têm procurado facilitar as negociações do setor financeiro, disponibilizando crédito entre outros produtos financeiros, prevendo tendências, simulando o comportamento tanto financeiro quanto o de investidores, além da 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 atividades.

F. M. Assef, Pontifícia Universidade Católica do Paraná (PUC-PR), Curitiba, Paraná, Brasil, fermassef@gmail.com.

M. T. A. Steiner, Pontifícia Universidade Católica do Paraná (PUC-PR), Curitiba, Paraná, Brasil, maria.steiner@pucpr.br.

P. J. Steiner Neto, Universidade Positivo (UP), Curitiba, Paraná, Brasil, pedrosteiner@up.edu.br.

D. G. B. Franco, Pontifícia Universidade Católica do Paraná (PUC-PR), Curitiba, Paraná, Brasil, davidgbf@gmail.com.

Nesse sentido, são utilizadas diferentes metodologias de solução para cada tipo de alternativa. Entre os métodos utilizados podemos 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 Linear - ADL e Regressão Logística-RL) e ainda abordagens denominadas de soft-computação (por exemplo, algoritmos da área da inteligência artificial (IA) e conjuntos brutos). Os métodos de IA, especialmente as Redes Neurais Artificiais (RNA), têm sido a ferramenta mais popular usada nos mercados financeiros [3].

Até a década de 70, as instituições financeiras focavam em metodologias semi-quantitativas para análise de medida e administração do crédito, que eram, sobretudo, avaliações subjetivas do crédito de determinado cliente por meio de análise de demonstrativos financeiros [4], [5].

A avaliação de risco de crédito é essencial para o controle dos próprios riscos da organização detentora do crédito. 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 [6].

Os métodos de avaliação de crédito podem ser classificados em seis categorias, sendo eles: Frameworks [7], [8]; Métodos Estatísticos [9], [10]; Teoria de Decisão [11], [12], [13]; Redes Neurais Artificiais [14], [15], [16]; Support Vector Machines [17], [18], [19]; e por fim, Análise por Envoltória de Dados [20], [21].

Em 19 de setembro de 2017 foi publicada uma reportagem no site da Rede Globo, relativa a uma matéria do “Jornal Nacional”, a respeito das motivações do Cadastro Positivo de Crédito não ser automatizado até o momento [22]. A medida ajudaria a reduzir os juros do consumidor, assim como a inadimplência nas instituições financeiras, uma vez que as taxas de juros seriam diferenciadas caso o consumidor pague em dia, com atraso ou então seja inadimplente [23].

Dentre as vantagens deste cadastro estariam inclusas: maior facilidade na aprovação de empréstimos e financiamentos e, também condições mais justas de pagamentos, com juros e prazos maiores ou menores, dependendo da referida classificação do cliente [24]. Desta forma, faz-se necessária uma ferramenta que leve em consideração os registros históricos de clientes, incluindo as suas classificações e que, por meio destas informações seja possível a classificação de novos clientes.

O desenvolvimento deste trabalho vem atender a esta necessidade. Os dados utilizados neste trabalho são reais e foram obtidos junto ao Banco XYZ, considerando os clientes

do município de Curitiba-PR, a respeito de pessoas jurídicas, no que diz respeito ao crédito bancário. Foi levantado o histórico de 8.358 clientes pessoa jurídica, observando apenas os que têm movimentação de empréstimos juntamente a instituição financeira analisada, sendo que foi constatado um total de 5.432 clientes. Seguindo a forma de classificação definida anteriormente, 2.600 foram classificados como adimplentes; 1.281, inadimplentes e 1.551 foram considerados pelo banco como temporariamente inadimplentes. Com o auxílio dos gerentes de crédito da instituição XYZ analisada ficaram definidos 15 atributos distintos para cada um dos 5.432 clientes, sendo eles: (1) Natureza Jurídica da Empresa; (2) Faturamento Bruto Anual; (3) Quantidade de Funcionários; (4) Atividade da Empresa; (5) Data de Constituição; (6) Data de Abertura da Empresa; (7) Segmento de Mercado; (8) Risco do Empréstimo; (9) Escala de Estabelecimento do Limite de Crédito; (10) Limite Estabelecido; (11) Aplicação; (12) Débito Vencido; (13) Débito Total; (14) Endividamento junto ao Sistema Financeiro Nacional; (15) Restrições ao Empréstimo.

A avaliação de crédito tem sua complexidade devido à variedade de riscos envolvidos [25], [26]. O detentor do crédito, além de recusar o crédito a um cliente que seria merecedor (futuro cliente adimplente), optando por garantir a segurança do pagamento recusando os empréstimos mais arriscados, poderá conceder crédito a um cliente não merecedor (futuro cliente inadimplente) [4], acarretando em enormes prejuízos à instituição. A determinação do merecimento de crédito a um cliente não pode ser realizada de forma tão direta devido a 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 o próprio estilo de vida [27].

A presente pesquisa consiste basicamente na avaliação do risco associado a empréstimos bancários concedidos por uma organização financeira a uma determinada empresa (pessoa jurídica). Para tanto, fez-se uso das técnicas de RNAs, mais especificamente, das Redes *Multilayer Perceptron* (MLP), das Redes de Base Radial (RBF) e da técnica estatística RL, comparativamente.

Mais especificamente, objetiva-se com esta pesquisa, classificar os clientes de uma instituição financeira brasileira, aqui chamada de “XYZ”, em três diferentes categorias (classes). As classes dos “adimplentes” (A) e dos “inadimplentes” (I), conforme uma típica aplicação de avaliação de crédito e, além disso, faz a verificação de uma terceira classe, a dos “temporariamente inadimplentes” (TI), que se caracterizam pelo pagamento de suas dívidas em atraso, de 1 a 180 dias do prazo de pagamento. Nos casos de atrasos superiores a 180 dias, tais clientes são enquadrados como inadimplentes.

Para o desenvolvimento desse trabalho será utilizada uma metodologia de IA, mais especificamente, as RNA do tipo MLP (RNA-MLP). A RNA-MLP em questão será trabalhada para realizar a divisão das referidas classes segundo o modelo “um contra todos” (*one-against-all*) [28], da seguinte forma: “adimplentes e outros”; “inadimplentes e outros”; “temporariamente inadimplentes e outros”, conforme detalhamento mais adiante.

As principais contribuições deste trabalho são: a análise de três classes de clientes (A, I, e TI), ao invés de apenas duas (A e I) como em geral é feito; a codificação das variáveis (atributos) da empresa XYZ visando maximizar a acurácia das técnicas e a utilização do método *one-against all*, pouco utilizado pelos pesquisadores desta área de pesquisa.

Este artigo está organizado conforme: na seção II é apresentado um levantamento de trabalhos correlatos em análise de crédito bancário e diferentes metodologias; em III é descrita a metodologia aplicada nesta pesquisa; na seção IV observamos o procedimento de obtenção e análise dos resultados obtidos; e, finalmente, em V temos as conclusões obtidas por essa pesquisa.

II. TRABALHOS CORRELATOS

A avaliação de risco de crédito é um problema de mineração de dados muito desafiador e importante no domínio da análise financeira. Desde que foi introduzida na década de 1950, tem sido amplamente utilizada e, mais recentemente, para a concessão de empréstimo, especialmente para cartões de crédito e consentimento de crédito [29]. Esse problema consiste basicamente na avaliação do risco associado com o empréstimo de uma organização financeira para determinada empresa ou indivíduo.

Um estudo em larga escala realizado por Loterman *et al.* [30] comparou 24 técnicas para classificação de crédito bancário, incluindo heurísticas, RNA, *Support Vector Machines* (SVM) e modelos estatísticos. Essas técnicas foram avaliadas em seis bases de dados reais, sendo que duas técnicas tiveram desempenhos melhores em mais de uma base de dados em termos de *Mean Absolute Error* (MAE), sendo elas: as que utilizaram RNA (desempenho superior em três das bases de dados) e a que utilizou B-OLS (*Beta Ordinary Least Squares*). Utilizando uma extensa comparação de técnicas, Brown e Mues [31] fizeram uso de cinco bases de dados reais. As técnicas utilizadas foram, resumidamente, RL, RNA, Árvores de Decisão, SVM e Florestas Randômicas (FR). Os melhores desempenhos encontrados pelos autores foram para os métodos de FR e para a técnica de gradiente *boosting*.

Em Wang *et al.* [32], os autores compararam os desempenhos de RNA do tipo RBF, árvores de decisão e RL em duas bases de dados de crédito da UCI. Os melhores resultados obtidos foram por meio da técnica RL, com acurácia superior a 88% em todos os casos apresentados. Em Hens e Tiwari [33], a acurácia de um modelo híbrido proposto (SVM e AG – Algoritmo Genético), apesar de superior, ficou bem próxima a de RNA. Outro híbrido foi desenvolvido para o crédito bancário numa base de dados turca. Inicialmente três técnicas foram analisadas separadamente, sendo elas: ADL, RL e RNA e, em seguida, essas técnicas foram hibridizadas [34].

Nem sempre as RNA apresentam os melhores resultados; por exemplo, Marqués *et al.* [35] apresentaram 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.

Outra proposta de híbrido de AG com RNA foi desenvolvida 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, como observado em [36]. Já em Oreski e Oreski [37] é apresentado ainda um híbrido de AG com RNA, onde os resultados obtidos mostraram que o algoritmo proposto tem a 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. Outro modelo para classificação de clientes foi proposto com base em variáveis Bayesianas latentes. O método proposto foi comparado com técnicas como: ADL, RNA, SVM, *Classification and Regression Trees* (CART) e *Multivariate Adaptive Regression Splines* (MARS) [38], [39]. As RNA podem ser comparadas com técnicas estatísticas; um exemplo desta comparação é no momento que são comparadas as técnicas de RNA, ADL, análise discriminante quadrática (ADQ) e RL para a previsão de adimplência ou não de 5.500 clientes. Os resultados revelaram que as RNA tiveram desempenho superior às demais técnicas utilizadas possuindo menor custo de erro de classificação [40].

Utilizando os métodos de classificação: ADL, ADQ, RL, RNA-MLP, SVM, árvores de classificação, e métodos de ensemble como o *bagging* e o algoritmo de *boosting*, Cubiles-de-la-Veja *et al.* [41] analisaram o risco de crédito em uma instituição financeira peruana. Os resultados obtidos pelos autores apresentaram que o modelo de MLP reduziu em 13,7% os custos de erro de classificação em relação à aplicação dos demais métodos testados. Em Abedini *et al.* [42], técnicas de ensemble foram utilizadas comparativamente a RNA-MLP, SVM, RL, e ainda a RL com *bagging* e as MLP utilizando RF. Os resultados mostraram que o modelo proposto possui desempenho superior à aplicação das técnicas por si só.

Semelhante à pesquisa aqui apresentada, foi proposta uma metodologia que dividia os clientes em “bons pagantes”, “não podem pagar” e “não vão pagar”. Em Bravo *et al.* [43] foi utilizado um 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 apenas em “pagantes” ou “não pagantes”. Em Koutanaei *et al.* [44], os autores também desenvolveram modelos híbridos para o pré-processamento das informações utilizando a Análise de Componentes Principais (ACP) e AG. Em seguida, estes dados foram 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*) apresentou acurácia superior às demais testadas.

Existem trabalhos desenvolvidos acerca da melhoria de desempenho das RNA, como em Zhao *et al.* [45], onde foram utilizadas diferentes estratégias: seja otimizando a distribuição das informações das bases de dados utilizando um método denominado de *Average Random Choosing*; ou comparando os efeitos da quantidade de instâncias de treinamento-validação-teste ou, ainda, encontrando 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.

São muitos os trabalhos com aplicações em crédito bancário com abordagens comparativas com diferentes classificadores conhecidos para o problema de risco de crédito, como em Ala'raj e Abbod [46], onde foram comparadas as técnicas RL, RNA, SVM, RF, árvores de decisão e Redes Bayesianas *naive*; e Ala'raj e Abbod [47] que diferiram nas comparações realizadas (algoritmos individualmente, híbridos, entre outros), ou até ainda acrescentando técnicas de ensemble como gradiente *boosting* [48], [49].

Uma proposta alternativa seria um modelo heurístico de *Particle Swarm Optimization* (PSO), mas utilizando funções de multiobjetivo para a classificação de clientes, comparando-o com as redes Bayesianas *naive*, RL, Otimização Mínima Sequencial, Árvores de Decisão e RNA, em três bases de dados. Neste caso, 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, o melhor desempenho foi para a RL com acurácia de 76,70% [50].

III. METODOLOGIA APLICADA

Nesta seção do artigo, são explicadas e exemplificadas as técnicas utilizadas para a resolução do Problema de Risco de Crédito da instituição XYZ. A Fig. 1 a seguir apresenta de forma detalhada as etapas utilizadas no desenvolvimento deste trabalho.

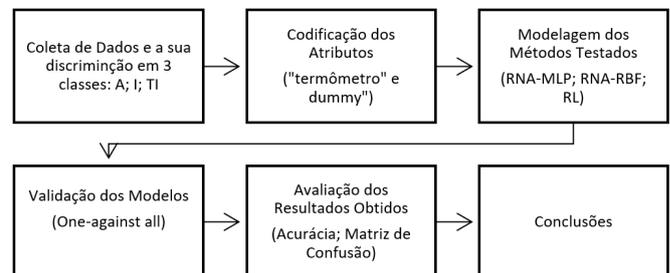


Fig. 1. Etapas da Pesquisa.

A. Redes Neurais Artificiais

As RNAs são modelos computacionais baseados na estrutura e no funcionamento do cérebro biológico que, apesar de não “rodar programas”, capacita o indivíduo, sendo responsável por seus pensamentos, percepções, cognições e emoções tendo, portanto, as mais diversas aplicações [51] e [52]. 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 [53]: i. ter capacidade de aprender através de exemplos, generalizando o aprendizado para reconhecer novas instâncias similares; ii. apresentar bom desempenho em tarefas onde o conhecimento do especialista não está disponível; iii. ser adequada a problemas onde não existe conhecimento a respeito dos modelos matemáticos dos domínios de aplicação; iv. possuir elevada “imunidade” ao ruído, sendo capaz de manter o desempenho em

presença de algumas informações falsas ou ausentes; v. apresentar a possibilidade de simulação de raciocínio “a priori” e impreciso.

Em resumo, as RNAs codificam informações utilizando a taxa de disparo (nível de atividade) dos neurônios biológicos, fornecendo os *outputs* [54].

Os modelos clássicos de RNAs, *Perceptron* e o já mencionado MLP, adotam a aprendizagem supervisionada por correção de erro. Elas se diferenciam pelo número de camadas, sendo que a rede *Perceptron* possui uma única camada e a MLP, como o próprio nome diz, possui duas ou mais camadas. Ambas são adequadas a problemas de classificação, mas as RNAs *Perceptron* resolvem somente problemas linearmente separáveis, sendo uma representação mais simples de uma RNA utilizada para a classificação de padrões. Já uma MLP, nada mais é que uma generalização do *Perceptron* de camada única. As redes neurais MLP também utilizam a regra de aprendizagem por correção de erro do paradigma supervisionado [55].

O treinamento RNA-MLP ocorre em duas etapas: a propagação *forward* que consiste na “alimentação” da rede com as instâncias (padrões de entrada) para frente e a propagação *backward* que consiste no ajuste dos pesos da rede de acordo com o erro obtido (diferença entre os valores obtidos e os valores desejados para as instâncias) [56].

As propagações *forward* e *backward* ocorrem com a aplicação do algoritmo de treinamento “retropropagação do erro” (ou *backpropagation error*). Este algoritmo tem a sua atuação nos pesos sinápticos da RNA, de forma a minimizar o erro, fazendo com que as saídas obtidas para cada uma das instâncias fiquem as mais próximas dos valores desejados para cada instância; isto é realizado por meio da técnica do gradiente descendente [57]. O erro obtido após a convergência do algoritmo *backpropagation* poderá ser considerado satisfatório (ou não). Deve-se aplicar o referido algoritmo para vários conjuntos de pesos iniciais, já que os mesmos são cruciais para uma aprendizagem adequada da RNA. A Tabela I a seguir apresenta, de forma resumida, os principais aspectos utilizados para o treinamento das MLPs.

Para a aplicação da RNA, optou-se por fixar 31 sementes diferentes (ou seja, valores aleatórios para que o próprio *software* criasse os pesos), sendo o primeiro teste realizado com o conjunto de pesos *default* e os demais sendo denominados de conjunto de pesos 1, conjunto de pesos 2 e assim por diante até o conjunto de pesos 30 conforme estipulado pelos autores. A RNA possui três camadas incluindo camada de entrada, camada oculta e camada de saída, onde foram testados diferentes quantidades de neurônios para a camada oculta, variando de 1 a 20. O algoritmo de treinamento escolhido foi o de *Levenberg-Marquardt*, parte da biblioteca de redes neurais do MATLAB projetado para se aproximar, em velocidade, aos métodos de treinamento de segunda ordem sem ter que computar a matriz Hessiana e, por fim, para a função de ativação, tanto da camada de entrada para a camada oculta quanto da camada oculta para a camada de saída, foi escolhida a Tagente Sigmoidal. Os demais parâmetros utilizados para uma RNA MLP foram considerados os padrões da biblioteca utilizada no *software*.

O segundo modelo de RNA explorado nesta pesquisa, também muito utilizado na literatura, foram as redes de Funções de Bases Radiais, *Radial Basis Functions* (RBF). As RBF também são redes multicamadas onde a ativação de um neurônio ocorre em função da distância entre as instâncias de entrada e dos pesos. Vale destacar que o nome, RBF, se deve ao fato da utilização, pelos neurônios da camada oculta, de funções de bases radiais.

TABELA I
TOPOLOGIA DAS REDES NEURAIAS MLP

PARÂMETROS	MLP
Total de Instâncias	5.432
Instâncias utilizadas no Treinamento	3.639
Instâncias utilizadas no Teste	1.793
Treinamento	Supervisionado
Quantidade de Conjuntos de Pesos Sinápticos testados	31 conjuntos testados
Número de Camadas	3 camadas (entrada-oculta-saída)
Varição dos Neurônios na Camada Oculta	1 - 20 neurônios
Algoritmo de Treinamento	Levenberg-Marquardt
Função de Ativação (1) (camada de entrada-camada oculta)	Tangente Sigmoidal
Função de Ativação (2) (camada oculta-camada de saída)	Tangente Sigmoidal

TABELA II
PARÂMETROS DA RBF

INFORMAÇÕES	RBF
Total de Instâncias	5.432
Instâncias utilizadas no Treinamento	3.639
Instâncias utilizadas no Teste	1.793
Número de Camadas	3 camadas (entrada-oculta-saída)
Numero Máximo de Neurônios na Camada Oculta	3.639
<i>Spread</i> utilizado	1

A Tabela II apresenta, de forma resumida, os principais aspectos utilizados para o treinamento das RBFs. A forma de avaliação para ambas as RNA testadas foi o método do holdout estratificado, dividindo as 5.432 instâncias em 3.639 para treinamento e 1.793 para teste, por 3.639 iterações.

B. 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 RL é o utilizado muito frequentemente para análise desse tipo de dados aqui abordados [58].

O que distingue a RL de modelos de Regressão Lineares é que a saída do primeiro é saída binária (dicotômica). A variável binária que é modelada é a variável resposta (variável dependente). Para que um modelo se adapte bem aos dados deve-se garantir que: os preditores não sejam correlacionados e, além disso, sejam significativamente relacionados com a

resposta [59]. Nesta pesquisa não se utilizou uma configuração específica para este método.

O modelo estatístico de RL fornece as respostas em termos de probabilidade de $Y \in [0, 1]$ dado um vetor X (instâncias) de preditores. A probabilidade da RL é dada pela equação (1) a seguir,

$$Prob\{Y = 1|X\} = [1 + e^{-X\beta}]^{-1} \tag{1}$$

onde a função (1) é denominada de função logística. Uma vez que a distribuição de uma variável binária aleatória Y é completamente definida pela verdadeira probabilidade de $Y = 1$, o expoente beta dá a razão ímpar da variável dependente e, sendo que o cálculo não faz premissas sobre a distribuição dos preditores, o modelo logístico não faz suposições relacionadas à distribuição [60].

Após aplicar a metodologia apresentada nas seções anteriores, fazendo uso do referido método *one-against-all*, os modelos foram analisados em termos de acurácia. Vale destacar que no caso das RNAs (MLP e RBF) busca-se maximizar o seu desempenho, ou seja, procura-se pela minimização do erro quadrático médio (MSE) apresentado na equação (2) [61],

$$mse = \frac{1}{n} * \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \tag{2}$$

onde, n é o tamanho do vetor de previsões realizadas, Y é o vetor dos valores observados da variável prevista e \hat{Y} é estimador da variável.

IV. OBTENÇÃO E DISCUSSÃO DOS RESULTADOS

Inicialmente, os 15 atributos citados na seção III foram transformados em valores binários tendo em vista que estes estavam apresentados de duas diferentes formas (ordinais ou nominais) e, também, as diferentes escalas (atributos com valores pequenos e outros com valores muito grandes). Para os atributos ordinais (2; 3; 9; 10; 11; 12; 13 e 14) foi aplicada a “codificação termômetro” e para os atributos nominais (1; 4; 5; 6; 7; 8 e 15) foi aplicada a “codificação *dummy*” [62]. As duas codificações estão exemplificadas nas Tabelas III e IV, respectivamente.

TABELA III
EXEMPLO DO PROCEDIMENTO “CODIFICAÇÃO TERMÔMETRO”

Faturamento Bruto Anual (FBA) x 1.000	Entradas		
$FBA \leq 1.000$	0	0	0
$1.000 < FBA \leq 2.000$	0	0	1
$2.000 \leq FBA \leq 3.000$	0	1	1
$FBA \geq 3.000$	1	1	1

Optou-se por utilizar 3 bits nesta codificação pois foi observado que a divisão dos dados em quartis possuía melhor configuração que demais opções (decis, percentis).

Assim, se determinada instância (empresa) possui o FBA de R\$ 2.800,00, as entradas correspondentes a esta instância nas técnicas utilizadas serão (0, 1, 1), ou seja, teremos três entradas binárias ao invés de apenas uma com o seu valor original (Tabela IV).

TABELA IV
EXEMPLO DO PROCEDIMENTO “CODIFICAÇÃO DUMMY”

Segmento	Entradas	
Microempresa	0	0
Pequena Empresa	0	1
Empresa	1	0

Neste caso, se a empresa analisada é uma microempresa, as suas entradas correspondentes serão (0, 0). Da mesma forma, tem-se a codificação e a interpretação das demais 13 variáveis.

Ambas as RNA (MLP e RBF) foram geradas no ©Matlab versão 2017a, quanto à RL, foi utilizado o *software* SPSS, em mesmo processador e sistema operacional.

Para as RNAs do tipo MLP, a melhor arquitetura, em termos de acurácia e de MSE, para cada classificação (Adimplente contra todos, Inadimplentes contra todos e Temporariamente Inadimplente contra todos) é apresentada pela Tabela V.

TABELA V
RESULTADOS REDES MLP

CLASSE	CONJUNTO DE PESOS	NEURÔNIOS	MSE	ACURÁCIA
ADIMPLENTES (A)	18	20	0,1926	74,70%
	13	13	0,07884	91,40%
INADIMPLENTES (I)	26	14	0,07878	91,40%
	28	13	0,08762	91,40%
TEMPORARIAMENTE INADIMPLENTES (TI)	15	19	0,1883	74,60%

Analisando os resultados apresentados nesta Tabela V percebe-se que a classificação de maior acurácia é representada pela classe dos Inadimplentes (ou seja, para a rede que classifica “Inadimplentes contra todos”; de 91,40% que ocorreu em três topologias, com 13 e 14 Neurônios na camada oculta, dependendo do conjunto de pesos utilizado), mostrando que os clientes inadimplentes têm características que os definem os discriminam relativamente bem das demais classes. Verifica-se também que a acurácia dos Adimplentes (74,70%) e da rede dos Temporariamente Inadimplentes (74,60%) é relativamente próxima, podendo sugerir assim que os perfis destas duas classes são relativamente semelhantes, dificultando a sua discriminação. Ainda, para ser possível fazer a análise de falsos positivos são também apresentadas as matrizes de confusão para as redes MLP (Tabelas VI, VII e VIII).

TABELA VI
MATRIZ DE CONFUSÃO MLP – ADIMPLENTES SIMULADO

		1	0	
REAL	1	701	157	926
	0	297	638	867
		998	795	74,7%

TABELA VII
MATRIZ DE CONFUSÃO MLP – INADIMPLENTES SIMULADO

		1	0	
REAL	1	354	69	423
	0	85	1.285	1.370
		439	1.354	91,4%

Já os resultados obtidos com a aplicação das RNAs do tipo RBF estão contidos na Tabela IX. Diferentemente das MLP, as RBF precisaram ter análise do desvio padrão uma vez que não

foram fixados valores aleatórios, etapa realizada para a obtenção de resultados das MLP conforme apresentado na Tabela I.

Para esta pesquisa, os tempos computacionais não foram considerados na análise uma vez que este custo foi irrelevante à análise ao comparar uma técnica com a outra.

Foi possível observar que a quantidade de neurônios para se obter a melhor acurácia por meio das RBFs apresentou grande variação. Da mesma forma que para as MLP verifica-se que o melhor modelo é o de classificação dos clientes Inadimplentes (80,72%). No entanto, para as demais classes, as acurácias médias foram de 58,84% para a classe de Adimplentes e de 70,60% para a classe de Temporariamente Inadimplentes.

TABELA VIII
MATRIZ DE CONFUSÃO MLP – TEMPORARIAMENTE
INADIMPLENTES
SIMULADO

		1	0	
REAL	1	361	151	512
	0	304	977	1.281
		665	1.128	74,60%

TABELA IX
RESULTADOS REDES RBF ACURÁCIA MSE

CLASSE	MÉDIA	DESVIO-PADRÃO	MÉDIA	DESVIO-PADRÃO
ADIMPLENTES (A)	58,84%	3,09%	0,32453	0,39883
INADIMPLENTES (I)	80,72%	2,93%	0,21608	0,393438
TEMPORARIAMENTE INADIMPLENTES (TI)	70,60%	0,74%	0,28952	0,386438

Por fim, o último método aplicado, foi o da RL, cujos resultados são apresentados na Tabela X a seguir. Novamente observa-se que a classificação dos Inadimplentes (88,50%) é realizada de forma mais discriminatória do que as demais e, novamente, a classificação dos Adimplentes e a dos Temporariamente Inadimplentes apresentaram resultados semelhantes (71,50% e 71,60%, respectivamente).

TABELA X. RESULTADOS RL	
CLASSE	ACURÁCIA
ADIMPLENTES (A)	71,50%
INADIMPLENTES (I)	88,50%
TEMPORARIAMENTE INADIMPLENTES (TI)	71,60%

V. CONCLUSÃO

O objetivo da presente pesquisa foi definir três modelos para a classificação de clientes “adimplentes” ou “inadimplentes” ou “temporariamente inadimplentes” de pessoas jurídicas que fazem uso de serviços de crédito bancário. Para tanto, foram considerados os registros históricos disponibilizados por uma instituição financeira, aqui denominada de Banco XYZ.

Fez-se uso das RNA do tipo MLP, com o algoritmo de aprendizagem *levenberg-marquardt* e com uma topologia com apenas uma camada escondida na qual o número de neurônios variou de 1 a 20, fazendo uso do modelo *one-against-all*. O mesmo teste foi realizado 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. E o terceiro método foi a RL de forma a comparar o modelo estatístico com os métodos de IA testados.

Quanto aos métodos em si, os melhores resultados em termos de acurácia foram gerados pelas RNAs do tipo MLP (74,70%; 91,40% e 74,60% para as classes adimplentes, inadimplente e temporariamente inadimplentes, respectivamente), como visto na Tabela III e nas matrizes de confusão representadas pelas Tabelas IV, V e IV. O segundo melhor resultado foi apresentado pela RL, onde para a classe dos adimplentes obteve-se acurácia de 71,50%, para os inadimplentes 88,50% e para os temporariamente inadimplentes, 71,60% de acurácia. A Tabela VII, resultados da RBF, mostra uma redução na acurácia do conjunto de teste para todas as classificações testadas (em média, 58,84%; 80,72% e 70,60% para as classes adimplentes, inadimplente e temporariamente inadimplentes, respectivamente).

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. Isso mostra que a classe dos Inadimplentes está mais bem discriminada do que as demais. Este fato pode ser considerado extremamente valioso, já que são exatamente as Inadimplentes 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, Inadimplentes e Temporariamente Inadimplentes; pesos que forneceram os resultados da Tabela III) e observar em qual classe o percentual foi mais alto. Caso esta nova instância seja enquadrada na classe “Temporariamente Inadimplentes”, evidenciando uma empresa que provavelmente pagará, mas com atraso de até 180 dias, os gerentes da instituição terão que decidir se concederão o crédito ou não.

O presente trabalho proporciona uma diferente análise dos tipos de clientes a serem avaliados para empréstimo bancário acrescentando uma terceira classificação: clientes temporariamente inadimplentes. Esta nova classe proporciona novas possibilidades de julgamentos tanto para pesquisas futuras na área de crédito bancário, assim como na maneira de avaliar os clientes na prática. Devido a introdução deste terceira classe, fez-se uso do método *one-against all*, pouco utilizado pelos pesquisadores. Além disso, são apresentadas duas formas de se codificar os atributos (codificações “termômetro” e “dummy”), visando a obtenção de uma maior acurácia das técnicas.

Como trabalhos futuros podem ser listados a utilização de modelos híbridos combinando diferentes técnicas para aperfeiçoamento do desempenho da classificação, um exame de outros métodos de avaliação de performance também pode ser sugerida (utilizando AUC – *área under the curve* – ou Fscore, por exemplo), ainda, uma análise sobre o impacto financeiro que o modelo apresenta para a lucratividade da instituição bancária também seria uma sugestão, bem como a utilização de técnicas de ensemble como visto em Addo *et al.* [63]. Outras alternativas de pesquisas podem ser a comparação dos resultados das mesmas técnicas utilizadas por demais autores para a base de dados em questão ou ainda uma análise dos

clientes que foram classificados positivamente em mais de um grupo, buscando talvez identificar algum atributo comum que possa causar esse tipo de situação.

AGRADECIMENTOS

O 1º. e 4º. autores deste trabalho são gratos à CAPES pelas bolsas de Mestrado e Doutorado, respectivamente, e o 2º. autor, ao CNPq, pela Bolsa Produtividade à Pesquisa.

REFERÊNCIAS

- [1] G. Fahner, “Estimating causal effects of credit decisions”, *Int. J. of Forecasting*, vol. 28, no. 1, pp. 248 – 260, 2012.
- [2] S. Luo, X. Kong and T. Nie, “Spline Based Survival Model for Credit Risk Modelling”, *Eur. J. Oper. Res.*, v. 253, n. 3, pp. 869 – 879, 2016.
- [3] A. Bahrammirzaee, “A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems”, *Neural Comput. Appl.*, v. 19, pp. 1165 – 1195, 2010.
- [4] E. I. Altman and A. Saunders, “Credit Risk Measurement: developments over the 20 years”, *J. of Banking and Finance*, vol. 21, pp. 1721 – 1744, 1998.
- [5] X. Chen, X. Wang and D. D. Wu, “Credit risk measurement and early warning of SMEs: An empirical study of listed SMEs in China”, *Decis. Support Syst.*, vol. 49, pp. 301 – 310, 2010.
- [6] F. Zhang, P. R. Tadikamalla and J. Shang, “Corporate credit-risk evaluation system: Integrating explicit and implicit financial performances”, *Int. J. of Prod. Economics*, vol. 177, pp. 77 – 100, 2016.
- [7] C. Bolton, “Logistic Regression and its Application in Credit Scoring”, *Universidade de Pretoria*, vol.238, 2009.
- [8] J. M. Redondo and F. Ortin, “A Saas Framework for Credit Risk Analysis Services”. *IEEE América Lat.*, vol. 15, no. 3, pp. 474-481, 2017.
- [9] Y. V. N. Acevedo, J. F. L. Quintero, C. E. M. Marín, C. González, “Business Rules Model for the Automation in the Receipt of Credit Applications by Financial Institutions Based on ArchiMate”. *IEEE América Lat.*, vol. 14, no. 6, pp. 2801-2806, 2016.
- [10] D. Diakoulaki, G. Mavrotas and L. Papayannakis, “A multicriteria approach for evaluating the performance of industrial firms”, *Omega*, vol. 20, no. 4, pp. 467 – 474, 1992.
- [11] W. F. Messier and J. V. Hansen, “Inducing Rules for expert system development an example using default and bankruptcy data”, *Management Sci.*, vol. 9, pp. 253 – 266, 1985.
- [12] K. Bryant, “ALEES: an agricultural loan evaluation expert system”, *Expert Syst. Appl.*, vol. 21, pp. 75 – 85, 2001.
- [13] D. Zhang, X. Zhou, S. C. Leung and J. Zheng, “Vertical bagging decision trees model for credit scoring”, *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7838 – 7843, 2010.
- [14] V. Desai, J. Crook and G. Overstreet Jr, “A comparison of neural networks and linear scoring models in credit union environment”, *Eur. J. Oper. Res.*, vol. 95, pp. 24 – 37, 1996.
- [15] D. West, “Neural Network Credit Scoring Models”, *Comput. Opns. Res.*, vol. 27, no. 11, pp. 1131 – 1152, 2000.
- [16] T. Pavlenko and O. Chernyak, “Credit risk modeling using Bayesian networks”, *Int. J. of Intelligent Systems*, vol. 25, no. 4, pp. 326 – 344, 2010.
- [17] T. V. Gestel, B. Baesens, J. Garcia and P. V. Dijke. “A support vector machine approach to credit scoring”, *J. of Banking & Finance*, vol. 2, pp. 73 – 82, 2003.
- [18] T. Bellotti and J. Crook, “Support Vector Machines for credit scoring and discovery of significant features”, *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3302 – 3308, 2009.
- [19] Z. Zhang, G. Gao and Y. Shi, “Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzyfication and penalty factors”, *Eur. J. Oper. Res.*, vol. 237, no. 1, pp. 335 – 348, 2014.
- [20] R. Shanmugam and C. Johnson. “At a crossroad of data envelopment and principal component analysis”, *Omega*, vol. 35, no. 4, pp. 351 – 364, 2007.
- [21] G. Iazzolino, M. E. Bruni and P. Beraldi, “Using DEA and financial ratings for credit evaluation: an empirical analysis”, *Appl. Econ. Let.*, vol. 20, no. 14, pp. 1310 – 1317, 2013.
- [22] Jornal Nacional do Brasil, “Após um ano, cadastro positivo de crédito ainda não é automático”. *Rede Globo de Telecomunicações*. Disponível em: <<https://goo.gl/VpaUdk>>. Acesso em: 21/09/2017.
- [23] Brasil, “Adesão ao Cadastro Positivo será automática”. *Palácio do Planalto*. Disponível em: <<https://goo.gl/jCKfwE>>. Acesso em: 21/09/2017.
- [24] Portal Brasil, “Cadastro Positivo entra em vigor e facilita concessão de crédito”. Disponível em: <<https://goo.gl/kgatW6>>. Acesso em: 21/09/2017.
- [25] Z. Babic and N. Plazibat, “Ranking of Enterprises based on multi-criterial analysis”, *Int. J. of Prod. Economics*, vol. 56/57, pp. 29 – 35, 1998.
- [26] P. Selwyn, “Financial credit-risk evaluation with neural and neuralfuzzy systems”, *Eur. J. Oper. Res.* vol. 112, pp. 310 – 321, 1999.
- [27] J. Ignatius, A. Hatami-Marbini, A. Rahman, L. Dharmotharan and P. Khoshnevis, “A fuzzy decision support system for credit scoring”, *Neural Comput. Appl.*, pp. 1 – 17, 2016.
- [28] L. Zhou and H. Fujita, “Posterior probability-based ensemble strategy using optimizing decision directed acyclic graph for multi-class classification”, *Inf. Sci.*, vol. 400-401, pp. 142-156, 2017.
- [29] S. Luo, X. Kong, T. Nie, “Spline Based Survival Model for Credit Risk Modelling”, *Eur. J. Oper. Res.*, v. 253, no. 3, pp. 869 – 879, 2016.
- [30] G. Loterman, I. Brown, D. Martens, C. Mues and B. Baesens, “Benchmarking regression algorithms for loss given default modeling”, *Int. J. of Forecasting*, vol. 28, pp. 161 – 170, 2012.
- [31] I. Brown and C. Mues. “An experimental comparison of classification algorithms for imbalanced credit scoring data sets”, *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3446-3453, 2012.
- [32] J. Wang, A.-R. Hedar, S. Wang and J. Ma, “Rough set and scatter search metaheuristic-based feature selection for credit scoring”, *Expert Syst. Appl.*, vol. 39, pp. 6123 – 6128, 2012.
- [33] A. B. Hens and M. K. Tiwari, “Computational time reduction for credit scoring: An integrated approach based on support vector machine and stratified sampling method”, *Expert Syst. Appl.*, vol. 39, no. 8, pp. 6774 – 6781, 2012.
- [34] S. Akkoç, “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”, *Eur. J. Oper. Res.*, vol. 222, pp. 168 – 178, 2012.
- [35] A. I. Marqués, V. García and J. S. Sánchez, “Exploring the behavior of base classifiers in credit scoring ensembles”, *Expert Syst. Appl.*, vol. 39, pp. 10244 – 10250, 2012.
- [36] S. Oreski, D. Oreski and G. Oreski, “Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment”, *Expert Syst. Appl.*, vol. 39, pp. 12605 – 12617, 2012.
- [37] S. Oreski and G. Oreski, “Genetic algorithm-based heuristic for feature selection in credit risk assessment”, *Expert Syst. Appl.*, vol. 41, no. 4, pp. 2052 – 2064, 2014.
- [38] L. -J. Kao, C. -C. Chiu and F. -Y. Chiu, “A Bayesian latent variable model with classification and regression tree approach

- for behavior and credit scoring”, *Knowl. Based Syst.*, vol. 36, pp. 245 – 252, 2012.
- [39] H. A. Abdou, M. D. D. Tsafack, C. G. Ntim and R. D. Baker, “Predicting creditworthiness in retail banking with limited scoring data”, *Knowl. Based Syst.*, vol. 103, pp. 89 – 103, 2016.
- [40] A. Blanco, R. Pino-Mejías, J. Lara and S. Rayo, “Credit scoring models for the microfinance industry using neural networks: Evidence from Peru”, *Expert Syst. Appl.*, vol. 40, pp. 356 – 364, 2013.
- [41] M. -D. Cubiles-de-la-Vega, A. Blanco-Oliver, R. Pino-Mejías and J. Lara-Rubio, “Improving the management of microfinance institutions by using credit scoring models based on Statistical Learning techniques”, *Expert Syst. Appl.*, vol. 40, pp. 6910 – 6917, 2013.
- [42] M. Abedini, F. Ahmadyadeh and R. Noorossana, “Customer credit scoring using a hybrid data mining approach”, *Kybernetes*, vol. 45, no. 10, 2016.
- [43] C. Bravo, L. C. Thomas and R. Weber, “Improving credit scoring by differentiating defaulter behavior”, *J. Oper. Res. Soc.*, vol. 66, pp. 771 – 781, 2015.
- [44] F. N. Koutanaei, H. Sajedi and M. Khanbabaei, “A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring”, *J. Retail Consum Serv.*, vol. 27, pp. 11 – 23, 2015.
- [45] Y. Zhao, S. Xu, B. H. Kang, M. M. J. Kabir, Y. Liu and R. Wasinger, “Investigation and improvement of multi-layer perception neural networks for credit scoring”, *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3508 – 3516, 2015.
- [46] M. Ala’raj and M. F. Abbod, “Classifiers consensus system approach for credit scoring”, *Knowl. Based Syst.*, vol. 104, pp. 89 – 105, 2016.
- [47] M. Ala’raj and M. F. Abbod, “A new hybrid ensemble credit scoring model based on classifiers consensus system approach”, *Expert Syst. Appl.*, vol. 64, pp. 36 – 55, 2016.
- [48] K. Andric and D. Kalpic, “The effect of class distribution on classification algorithms in credit risk assessment”, *39th International Convention on Information and Communication Technology, Electronics and Microelectronics*, MIPRO 2016 – Proceedings, 2016.
- [49] A. Bequé and S. Lessmann, “Extreme learning machines for credit scoring: An empirical evaluation”, *Expert Syst. Appl.*, vol. 86, pp. 42 – 53, 2017.
- [50] Y. Guo and C. Dong, “A novel intelligent credit scoring method using MOPSO”, *Proceedings of the 29th Chinese Control and Decision Conference*, CCDC 2017, 2017.
- [51] R. L. U. Cazarez and C. L. Martín, “Neural Networks for predicting student performance in online education”, *IEEE América Lat.*, vol. 16, no. 7, pp. 2053-2060, 2018.
- [52] E. F. Felix and E. Possan, “Modeling the Carbon Front of Concrete Structures in the Marine Environment through ANN”, *IEEE América Lat.*, vol. 16, no. 6, 2018.
- [53] G. Bittencourt, “Inteligência Artificial: Ferramentas e Teorias”. 3ª. ed. Florianópolis: Editora da UFSC, 2006. ISBN 85-328-0138-2.
- [54] X. Lin, X. Wang and Z. Hao, “Supervised learning in multilayer spiking neural networks with inner products of spike trains”, *Neurocomputing*, vol. 237, pp. 59 – 70, 2017.
- [55] S. Haykin, “Redes neurais: princípios e prática”. 2ª. ed. Porto Alegre: Bookman, 2001.
- [56] L. V. Fausett, “Fundamentals of neural networks: architectures, algorithms, and applications”. 1a. ed. USA: Prentice Hall, 1994.
- [57] S. Haykin. *Neural Networks and Learning Machines*. 3a. ed. USA: Prentice Hall, 2009. 906p.
- [58] R. A. L. F. P. P. Costa Jr. and F. R. B. Cruz, “Minimização do erro no algoritmo back-propagation aplicado ao problema de manutenção de motores”, *Pesq. Oper.*, vol. 18, no. 1, pp. 21 – 36, 1998.
- [59] D. W. Hosmer, S. Lemeshow and R. X. Sturdivant, “Applied Logistic Regression”. 3a. ed. New Jersey: John Wiley & Sons Inc., 2013.
- [60] J. M. Hilbe, “Practical Guide to Logistic Regression”. USA: Taylor & Francis, 2016.
- [61] F. E. Harrell Jr, “Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis”. Springer, 2015.
- [62] M. T. A. Steiner, J. C. Nievola, N. Y. Soma, T. Shimizu & P. J. Steiner Neto. Extração de regras de classificação a partir de redes neurais para auxílio à tomada de decisão na concessão de crédito bancário. *Pesquisa Operacional*, vol. 27, n. 3, pp.407-426, 2007.
- [63] P. M. Addo, D. Guegan and B. Hassani, “Credit Risk Analysis Using Machine Learning and Deep Learning Models”, *Risks*, v. 6, no. 2. 2018.



Fernanda Medeiros Assef possui Graduação e Mestrado em Engenharia de Produção pela Universidade Federal do Paraná (UFPR; 2015 e 2018, respectivamente) e está cursando o Doutorado pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas (PPGEPS) pela Pontifícia Universidade Católica do Paraná (PUCPR), com principal atuação na área de Pesquisa Operacional.



Maria Teresinha Arns Steiner possui Pós-Doutorado pelo Instituto Tecnológico da Aeronáutica (ITA; 2005) e pelo Instituto Superior Técnico de Lisboa (IST; 2014). Atuou na Universidade Federal do Paraná (UFPR) de 1978 a 2010. Desde 2011 vem atuando na PUCPR, junto ao PPGEPS. Tem experiência em temas de Pesquisa Operacional.



Pedro José Steiner Neto possui Pós-Doutorado pelo IST de Lisboa (2014). Atuou na UFPR de 1980 a 2014. Desde 2015 vem atuando na Universidade Positivo (UP), junto ao Programa de Pós-Graduação em Administração (PPAD). Tem experiência em temas relacionados à Tomadas de Decisão.



David Gabriel de Barros Franco possui graduação em Engenharia de Produção pela PUCPR (2012) e Mestrado em Engenharia de Produção e Sistemas (PPGEPS) também pela PUCPR (2015). Atualmente é doutorando do PPGEPS pela PUCPR. É especialista no *software* MATLAB.