

Análise de Crédito Bancário de Pessoas Jurídicas

Fernanda Medeiros Assef (Universidade Federal do Paraná) fermassef@gmail.com
Maria Teresinha Arns Steiner (Pontifícia Universidade Católica do Paraná) maria.steiner@pucpr.br
Anis Assad Neto (Universidade Federal do Paraná) anis.assad@gmail.com
Cassius Tadeu Scarpin (Universidade Federal do Paraná) cassiusts@gmail.com
David Gabriel de Barros Franco (Pontifícia Universidade Católica do Paraná) davidgbf@gmail.com

Resumo: Este trabalho visa realizar a análise de crédito bancário de pessoa jurídica (em adimplentes; inadimplentes e temporariamente inadimplentes), de forma a auxiliar a tomada de decisão do analista deste ramo. Para tanto, fez-se uso de Redes Neurais Artificiais (RNAs), mais especificamente, do Perceptron de Múltiplas Camadas (Multiple Layer Perceptron; MLP). A classe “temporariamente inadimplente” representa os clientes que pagam seus empréstimos, mas com atraso. Para a implementação da RNA, fez-se uso do *software* MATLAB. O experimento realizado testou as redes em várias situações, principalmente no que diz respeito ao número de neurônios na camada oculta e às variações dos pesos iniciais. Para as simulações desenvolvidas, os dados de entrada da rede utilizados foram 15 informações (atributos) coletadas pelo especialista do banco analisado. A análise dos resultados foi realizada através da comparação entre as acurácias geradas pelas RNAs. Os resultados apresentados mostram que os clientes inadimplentes são facilmente identificáveis pela rede, mas quanto aos clientes adimplentes e os temporariamente inadimplentes a rede teve maior dificuldade na diferenciação, sugerindo que esses possuem perfil semelhante.

Palavras chave: Redes Neurais Artificiais, Análise de Crédito Bancário, MATLAB.

Bank Credit of Legal Entities Analysis

Abstract This work aims at analyzing bank credit of legal entity (in default, defaulters and temporarily defaulters), in order to assist decision making by the analyst of this branch. In order to do so, Artificial Neural Networks (RNAs), more specifically, the Multiple Layer Perceptron (MLP) were used. The class "temporarily defaulters" represents customers who repay their loans, but are late. For the implementation of RNA, we made use of the MATLAB software. The experiment carried out tested the networks in several situations, mainly regarding the number of neurons in the hidden layer and the variations of the initial weights. For the simulations developed, the input data of the network used were 15 information (attributes) collected by the expert of the analyzed bank. The analysis of the results was performed through the comparison between the accuracy generated by ANNs. The results show that the defaulting clients are easily identifiable by the network, but for the delinquent clients and the temporarily defaulters the network had greater difficulty in the differentiation, suggesting that they have a similar profile.

Key-words: Artificial Neural Networks, Analysis of Banking Credit, MATLAB.

1. Introdução

No ano de 2008, o Brasil atravessava o seu melhor momento econômico dos 30 anos que o precediam. Estabilidade econômica, expansão de renda e do consumo de investimentos crescentes aceleravam o crescimento econômico e assim reduziam as desigualdades sociais. De acordo com o Banco Nacional do Desenvolvimento Econômico e Social (BNDES, 2012), a economia brasileira chegou a alcançar um crescimento anual médio de 4% ao ano entre os

anos de 2000 e 2011, valor superior ao observado nas décadas anteriores, que foi cerca de 2% anuais.

De acordo com o Ministério de Minas e Energia (2015), realizando uma previsão da economia brasileira para os próximos 10 anos, o comércio mundial deverá agora entrar em uma trajetória mais vigorosa à medida que os níveis de atividade dos países desenvolvidos aumentem, de forma a favorecer a demanda por produtos de países emergentes. Com relação aos investimentos a serem realizados em instituições financeiras, considera-se que os ajustes necessários a serem feitos na economia devem estabelecer o ritmo das inversões nos primeiros anos. Porém, na segunda metade da década, a retomada de confiança dos agentes e a melhoria do ambiente macroeconômico impulsionarão os investimentos em importantes setores da economia.

Segundo relatório de Coutinho (2009) para o BNDES, as instituições financeiras não bancárias responderam de 60% a 70% do crédito criado, levando bancos e mecanismos de precificação ao declínio. De acordo com o Serviço de Proteção ao Crédito (2017), em Janeiro de 2017, 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.

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 atividades. Nesse sentido, são utilizadas diferentes metodologias de solução para cada tipo de assunto. 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 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).

O objetivo deste trabalho é definir um modelo para a classificação de clientes “adimplentes” ou “inadimplentes” de pessoas jurídicas que fazem uso de serviços de crédito bancário. É apresentada, também, uma terceira classificação, “temporariamente inadimplentes”, referente aos clientes que atrasaram no pagamento, mas que conseguiram realizar o pagamento. Para tanto, foram considerados os dados disponibilizados por um Banco XYZ, que oferece uma gama de produtos bancários, tanto para setor público quanto para o privado.

Para o desenvolvimento desse trabalho será utilizada uma metodologia de IA, mais especificamente, as RNA do tipo *feedforward* de Múltiplas Camadas. A RNA em questão será trabalhada para realizar a divisão nas classes segundo o modelo um contra todos (*1-against-all*), da seguinte forma: adimplentes e outros; inadimplentes e outros; temporariamente inadimplentes e outros.

O trabalho está organizado da seguinte forma: na seção 2 é apresentada uma breve revisão sobre o funcionamento das RNA, além de uma revisão bibliométrica sobre este assunto. Na seção 3 é apresentada a metodologia utilizada para solucionar o problema em pauta; na seção 4 são apresentados os resultados obtidos e, finalmente, na seção 5, as conclusões.

2. Revisão da Literatura

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 elementos simples 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 efectoras, células que exercem funções imunológicas;
- c) Os interneurônios formam uma rede de comunicação e de integração entre os sensitivos e os motores.

Haykin (2009) descreve matematicamente o neurônio artificial pelas equações (1) e (2) a seguir.

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

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

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; θ_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 .

A função de ativação $\varphi(\cdot)$ é muito importante para o comportamento de uma RNA uma vez que define a saída do neurônio artificial e, dessa forma, o caminho pelo qual a informação será conduzida. Quanto às funções de ativação das RNA, Hagan *et al.* (2014) final apresentam as três mais utilizadas, sendo elas: Função de Limite Rígido (Função Limiar ou Função Passo); Função Linear; e a Função Sigmoidal.

Comumente um neurônio, mesmo com muitos insumos, pode não ser suficiente para a obtenção da solução desejada. Pode ser necessária uma quantidade maior de neurônios, operando em paralelo, no que é chamado de "camada". A composição de camadas de uma RNA é chamada de arquitetura, ou topologia da rede.

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 chamada de rede *feedforward* de camada única.

A segunda classe de uma rede neuronal *feedforward* distingue-se pela presença de uma ou mais camadas ocultas, operando em paralelo, cujos nós que realizam os cálculos 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.

A terceira topologia é denominada de RNA Recorrentes. Segundo Haykin (2009), essa arquitetura distingue-se das demais pelo fato de possuir pelo menos um *loop* de retroalimentação.

Depois da decisão da arquitetura da RNA a ser utilizada deve ser definido como funcionará o aprendizado da estrutura para que possa lidar com os dados. Dada uma RNA, o aprendizado ocorrerá à proporção que os pesos forem ajustados com base em alguma regra pré-estabelecida, como, por exemplo, a regra delta (ou regra de Widrow-Hoff).

Matematicamente, podemos definir a atualização dos pesos da rede na iteração n pela equação apresentada em (3).

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

Onde w_{kj} e w_{kj} representam os pesos do neurônio k da camada j e Δw_{kj} são os ajuste destes pesos, todos com $0 < j < m$ (onde, $j = 0$ representa o *bias* e m representa a m -ésima informação (entrada) da instância analisada no momento). Os algoritmos de aprendizado, a serem discutidos a seguir, diferem na forma de como o Δw_{kj} é calculado. De acordo com a regra delta, o ajuste Δw_{kj} aplicado ao peso sináptico w_{kj} é definido pela fórmula apresentada em (4).

$$\Delta w_{kj}(n) = \gamma \cdot \delta_k(n) \cdot x_j(n) \quad (4)$$

Onde γ é a taxa de aprendizado $0 < \gamma < 1$, δ_k é o erro obtido, ou seja, a diferença entre o valor esperado d_j e o sinal efetivo de saída da rede a_k , conforme apresentado pela equação (5) e, finalmente, x_j é o sinal de entrada da instância analisada.

$$\delta_k(n) = d_j(n) - a_k(n) \quad (5)$$

Além do treinamento, é necessário realizar um diagnóstico, ou seja, uma avaliação sobre uma RNA proposta com o objetivo de evitar o vício (ou seja, a “memorização” ao invés do “aprendizado”) dos dados. Existem basicamente quatro tipos de validação (diagnóstico) para as RNA, sendo eles método *hold-out*, validação cruzada (*cross-validation*), deixe um fora (*leave-one-out*) e *bootstrap*.

O *hold-out* é um método típico de validação cruzada, ele é amplamente utilizado para avaliar previsões por computador baseado na aprendizagem. Para esse método utiliza-se uma divisão fixa de aproximadamente 2/3 dos casos para treinamento e 1/3 para teste (SOUZA, 2011). Neste método, é necessário assegurar que a amostragem randômica seja feita de tal forma que possa garantir que cada classe está propriamente representada em ambos os conjuntos (treinamento e de teste). Este procedimento é chamado de estratificação e desta forma tem-se, então, *holdout* estratificado (STEINER, 2012). Assim, o número de instâncias nos conjuntos de treinamento e de teste deve ser proporcional ao número de instâncias representativas de cada classe do conjunto original.

De acordo com Amari *et al.* (1997) para utilizar o método de *cross-validation* é necessário dividir todos os exemplos (instâncias) disponíveis em conjuntos disjuntos. Um desses conjuntos é utilizado para treinar as RNA, enquanto os demais serão utilizados para a validação e treinamento. De acordo com Steiner (2012), no *cross-validation*, decidimos por um número fixo de *folds* (partições dos dados), por exemplo, três *folds*. Então os dados são divididos em aproximadamente 3 partições iguais, sendo que usamos 2/3 para treinamento e 1/3 para teste e repetimos o procedimento 3 vezes tal que, ao final, todo exemplo/padrão/instância tenha sido usado exatamente uma vez para teste. Da mesma forma que o método anterior, deve-se trabalhar com o *cross-validation* estratificado.

Conforme Rodrigues, Macrini e Monteiro (2008), quando o número de exemplos rotulados disponíveis, n , for severamente limitado, pode-se usar a forma de *cross-validation* conhecida como *leave-one-out*. Para este método, $(n-1)$ exemplos são usados para treinar a RNA, e o modelo, então, é validado testando-o sobre o exemplo “deixado de fora”. O experimento é repetido para um total de n vezes, cada vez deixando de fora um exemplo diferente para a validação. De acordo com Steiner (2012), este procedimento é atrativo por duas razões: primeiramente, a maior quantidade possível de dados é usada para treinamento em cada caso que, presumidamente, aumenta a chance do classificador ser mais acurado; em segundo lugar, o procedimento é determinístico, ou seja, não envolve amostragem randômica. Este procedimento possui duas principais desvantagens, em primeiro lugar é o alto custo computacional, uma vez que deverá ser executado n vezes, o que o torna inviável para bases de dados grandes. A segunda desvantagem deste procedimento é que a base não pode ser estratificada.

Por fim, o *bootstrap* é um método baseado em dados para inferência estatística. A ideia desta técnica é amostrar a base de dados com substituição para formar o conjunto de treinamento. Uma base de dados de n exemplos é amostrada n vezes, com substituição, para dar outra base de dados com n exemplos. Como alguns elementos na segunda base de dados serão repetidos, haverá alguns exemplos da base de dados original que não terão sido escolhidos, dessa forma, esses dados não escolhidos serão os utilizados como exemplos para teste (STEINER, 2012).

2.1. Revisão Bibliométrica

Para fazer uma análise a respeito da evolução de publicações com utilização de RNA foi utilizada a base Scopus, onde foi pesquisado o termo *Artificial Neural Networks*, presente em título, resumo ou como palavra-chave. Em seguida filtrou-se apenas artigos publicados em periódicos e conferências e por fim, foi aplicado um novo filtro onde a pesquisa seria refinada para que os artigos em questão possuísem todos os termos *Artificial Neural Networks* nas palavras-chave. A partir destas restrições, a base de dados retornou 43.591 artigos dentro deste tema e o volume de publicações por ano, desde 1988 que foi o primeiro artigo encontrado pela base. A Figura 1 apresenta a evolução no número de publicações sobre as RNA.

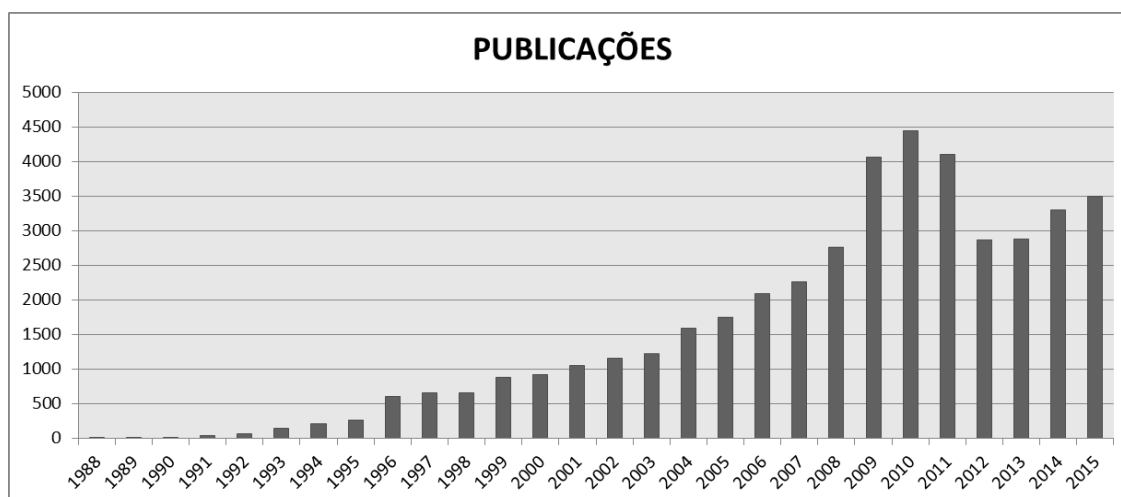


Figura 1 – Evolução no número de publicações ligadas às Redes Neurais Artificiais

É possível perceber que o tema teve constante aumento no volume de publicações tendo uma baixa apenas no ano de 2012, mas já apresentando novo crescimento nos anos seguintes. Buscando verificar a distribuição destas publicações entre os países, foi gerado o gráfico da

Figura 2, a seguir, que apresenta também o acumulado por país, necessário para as análises seguintes.

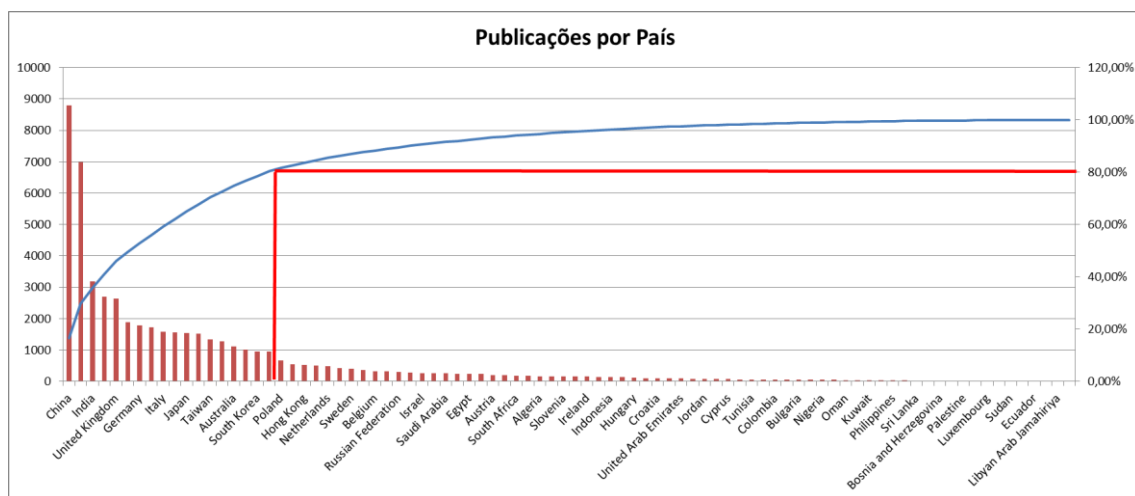


Figura 2 - Pareto publicações por país

Para verificar melhor a divisão de artigos por países, foram selecionados os mesmos responsáveis por 80,00% das publicações levantadas anteriormente, que podem ser observados na Figura 3.



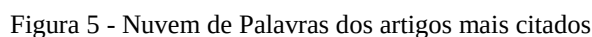
Figura 3 - Países com maior número de publicações

O gráfico da Figura 3 apresenta os países com uma maior representatividade nas publicações geradas pela pesquisa na base de dados. Podemos perceber que a China e os Estados Unidos da América são responsáveis por quase metade das publicações levantadas (46,41% do total de publicações).

Quanto às áreas de pesquisa mais publicadas, percebe-se que as três principais são *Computer Science*, *Engineering* e *Medicine*. Sendo estas as que possuem mais publicações ligadas às RNA, como podemos ver no gráfico da Figura 4 a seguir.



O mais citado foi o Barabási e Albert (1999) sob o título: “Emergence of scaling in random networks”, onde foi proposto um modelo com duas mecânicas genéricas: a primeira que as redes se expandem continuamente com a adição de novos vértices, e a segunda que novos vértices entram preferencialmente em locais que já estejam bem conectados. Com o objetivo de verificar quais são as palavras chaves mais utilizadas nos artigos mais influentes levantados foi construída uma nuvem de palavras, apresentada na Figura 5.



Outra análise realizada foi feita de forma a verificar as tendências de publicações ligadas às RNA, desejando investigar quais as áreas de pesquisas que as 2.000 publicações mais recentes sobre este assunto atuam. De forma análoga às publicações mais citadas, foram filtradas as

[illegible]

Desta vez, percebemos que as palavras estão ligadas às formas de utilização das RNA mais do que os campos de aplicação que elas estão sendo empregadas. O que nos mostra que este assunto se tornou admissível para diversos campos de estudo, agora sendo diferenciado apenas na forma de utilização do algoritmo e o que é combinado com o mesmo, sejam técnicas estatísticas, metaheurísticas, entre outros.

A presente pesquisa se restringe a encontrar uma solução realizável para o problema de pontuação de crédito voltado a clientes em forma de pessoas jurídicas com a aplicação de diferentes topologias de RNA do tipo *Multi-Layer Perceptron*. As eventuais decisões gerenciais consequentes da categorização concluída não serão responsabilidade do trabalho aqui exposto.

Os dados utilizados neste artigo são reais e foram obtidos junto ao Banco XYZ, considerando os clientes de todo o Estado do Paraná, 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 empresas.

Atualmente, o Banco XYZ utiliza como ferramenta para realizar sua análise de crédito, um aplicativo interno chamado Análise de Crédito (ANC). É desse aplicativo, que contém as informações cadastrais e contábeis das empresas, que a gerência do Banco se vale para apoiar suas decisões de conceder (ou não) crédito bancário.

Neste trabalho, pretende-se utilizar os dados históricos de um total de 8.358 clientes pessoa jurídica e de cada um destes clientes foram considerados 15 atributos, listados a seguir, definidas pelo próprio especialista do referido banco:

- a) Natureza Jurídica (Sociedade Limitada; Ltda – Sociedade Cotas de Responsabilidade; Empresa Individual de Responsabilidade Ltda; Firma Individual; Empresário Individual; Sociedade Simples; Microempreendedor Individual);

- b) Faturamento Bruto Anual (apresenta uma variação de 0 a R\$769.665.000,00);
- c) Quantidade de Funcionários (apresenta uma variação de 0 a 5.113);
- d) Atividade (apresenta uma infinidade de variações);
- e) Data da Constituição;
- f) Data de abertura;
- g) Segmento (Micro; Pequena; Corporate; Empresa; *Empresarial Middle*; *Empresarial Upper Middle*; Estadual; PJ – não classificável / sem perfil de segmentação);
- h) Risco (variação de A a E);
- i) Escala de Estabelecimento de Limite de Crédito;
- j) Limite Estabelecido (variação de 0 a R\$ 35.000.000,00);
- k) Aplicação (variação de 0 a R\$ 76.147.526,00);
- l) Débito Vencido (variação de 0 a R\$ 12.777.748,00);
- m) Débito Total (variação de 0 a R\$ 32.123.211,00);
- n) Endividamento Sistema Financeiro Nacional (SFN) Total (variação de 201 a R\$ 355.563.956,00);
- o) Restrições (Restrições Fracas / Informações; Restrições Impeditivas; Sem Restrições).

Para a implementação computacional da RNA foi utilizado o *MATLAB 2017a*. Quanto aos parâmetros da RNA, foram testadas as topologias contendo de 1 a 20 neurônios na camada oculta; a função de ativação utilizada foi a função tangente-logística, o algoritmo de treinamento utilizado foi o *Levenberg-Marquardt* descrito em THE MATHWORKS INC. Assim como os métodos quase-Newton, o algoritmo de *Levenberg-Marquardt* foi projetado para se aproximar, em velocidade, aos métodos de treinamento de segunda ordem sem ter que computar a matriz Hessiana. Em relação aos pesos utilizados foram testados cinco conjuntos de pesos iniciais.

4. Resultados

Aplicando-se a metodologia da forma que foi apresentada anteriormente foram obtidos os seguintes resultados após o treinamento e teste da RNA conforme qual o modelo que está sendo analisado.

A seguir, na Tabela 1, os resultados dos testes da RNA para a classificação das instâncias “Adimplentes e Outros”. Percebemos que a melhor topologia para esta rede foi de dezenove neurônios na camada oculta, utilizando os pesos considerados *default* pelo *software*, e o resultado foi de 73,90% de acurácia.

NÚMERO DE NEURÔNIOS	ADIMPLENTES				
	Pesos - Default	Pesos - 1	Pesos - 2	Pesos - 3	Pesos - 4
1	71,10%	71,10%	71,10%	71,30%	71,30%
2	70,10%	70,70%	71,30%	71,90%	71,90%
3	71,60%	72,20%	71,40%	71,30%	71,30%
4	72,10%	70,20%	70,80%	71,60%	71,60%
5	71,50%	71,40%	71,50%	72,00%	72,00%
6	71,80%	72,30%	72,40%	72,20%	72,20%
7	71,80%	71,50%	72,40%	71,90%	71,90%
8	72,10%	72,80%	70,50%	72,60%	72,60%
9	71,50%	73,10%	71,40%	72,00%	72,00%
10	69,80%	71,10%	71,80%	72,40%	72,40%
11	71,70%	72,60%	71,50%	71,60%	71,60%
12	71,40%	71,70%	71,10%	71,80%	71,80%
13	70,60%	72,10%	71,00%	70,10%	70,10%
14	73,80%	72,70%	71,00%	70,40%	70,40%
15	71,40%	70,70%	70,70%	71,40%	71,40%
16	71,20%	71,30%	71,80%	73,00%	73,00%
17	73,10%	71,90%	71,90%	72,80%	72,80%
18	71,40%	71,90%	73,40%	71,90%	71,90%
19	73,90%	71,60%	71,10%	71,40%	71,40%
20	72,20%	72,20%	73,00%	69,00%	69,00%

Tabela 1 - Resultados da RNA “Adimplentes e Outros”

A seguir, na Tabela 2, temos de forma análoga, os resultados da RNA para a classificação das instâncias “Inadimplentes e Outros”. Percebemos que a melhor topologia para esta rede foi, também de dezenove neurônios na camada oculta e pesos do conjunto *default*, resultando numa acurácia de 90,30%

NÚMERO DE NEURÔNIOS	INADIMPLENTES				
	Pesos - Default	Pesos - 1	Pesos - 2	Pesos - 3	Pesos - 4
1	88,80%	88,80%	88,50%	88,60%	88,60%
2	88,50%	89,30%	89,20%	88,60%	88,60%
3	89,40%	89,00%	88,70%	88,60%	88,60%
4	88,90%	88,70%	87,70%	88,50%	88,50%
5	76,40%	89,10%	89,20%	89,80%	89,90%
6	89,40%	88,40%	90,00%	88,80%	88,80%
7	89,70%	88,60%	90,10%	90,70%	90,70%
8	88,90%	89,70%	88,50%	89,50%	89,50%
9	89,70%	89,90%	88,80%	89,80%	89,80%
10	89,20%	88,90%	89,30%	89,80%	89,80%
11	89,50%	89,10%	88,80%	90,10%	90,10%
12	89,20%	88,70%	89,50%	89,60%	89,60%
13	89,20%	88,40%	90,10%	87,60%	87,60%
14	89,60%	89,40%	89,90%	89,80%	89,80%
15	89,70%	88,70%	90,60%	88,40%	88,40%
16	91,00%	89,20%	89,70%	89,40%	89,40%
17	89,20%	89,10%	89,70%	89,70%	89,70%
18	89,80%	87,90%	88,90%	89,60%	89,60%
19	90,30%	89,70%	88,70%	88,70%	88,70%
20	89,60%	88,90%	90,20%	89,70%	89,70%

Tabela 2 - Resultados da RNA “Inadimplentes e Outros”

Por fim, na Tabela 3, temos a classificação para as instâncias “Temporariamente Inadimplentes e Outros”. Neste caso, a melhor arquitetura de rede envolvem quinze neurônios na camada oculta tanto para os conjuntos de pesos 3 e 4, resultando em 73,50% de acurácia.

TEMPORARIAMENTE INADIMPLENTES					
NÚMERO DE NEURÔNIOS	Pesos - Default	Pesos - 1	Pesos - 2	Pesos - 3	Pesos - 4
1	71,60%	71,40%	71,40%	72,10%	72,10%
2	71,40%	71,50%	71,60%	71,90%	71,90%
3	71,70%	72,00%	71,90%	71,70%	71,70%
4	72,10%	71,60%	71,20%	71,90%	71,90%
5	72,30%	71,60%	71,40%	72,30%	72,30%
6	72,30%	72,30%	73,00%	72,20%	72,20%
7	71,50%	71,80%	71,80%	72,20%	72,20%
8	71,60%	71,60%	71,80%	71,90%	71,90%
9	72,10%	72,50%	72,10%	71,80%	71,80%
10	72,20%	71,80%	72,30%	71,60%	71,60%
11	72,30%	71,90%	72,00%	72,10%	72,10%
12	72,70%	72,20%	73,20%	71,70%	71,70%
13	72,00%	72,50%	71,60%	72,60%	72,60%
14	72,30%	72,10%	72,40%	72,30%	72,30%
15	72,40%	72,10%	71,80%	73,50%	73,50%
16	72,10%	72,50%	72,20%	71,20%	71,20%
17	72,90%	73,10%	71,80%	72,00%	72,00%
18	71,60%	70,00%	72,60%	72,10%	72,10%
19	71,90%	72,00%	73,20%	72,70%	72,70%
20	72,00%	72,40%	72,20%	71,90%	71,90%

Tabela 3 - Resultados da RNA “Temporariamente Inadimplentes e Outros”

Como a melhor opção para o Banco XYZ analisado é reduzir os erros para os testes contidos nas Tabelas 1 e 3 (“Adimplentes e Outros” e “Temporariamente Inadimplentes” e Outros). Podemos ainda fazer uma análise quanto aos valores que a referida instituição perde concedendo crédito para os clientes inadimplentes e quanto que a instituição perderia caso utilizasse esta RNA para a classificação do cliente.

5. Conclusão

O objetivo do presente trabalho foi definir um modelo 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 dados disponibilizados por uma instituição financeira, aqui denominada de Banco XYZ. Para cada instância foram considerados 15 atributos listados na seção 2 deste trabalho.

Para tanto fez-se uso das RNA do tipo *feedforward*, com o algoritmo de aprendizagem *back-propagation* e com uma topologia com apenas uma camada escondida na qual o número de neurônios variou de 1 a 20, fazendo uso do modelo *1-against-all*.

Por meio dos resultados obtidos, foi possível verificar que a classificação com maior acurácia para a RNA foi a dos “Inadimplentes e Outros” (Tabela 2; 91,00% de acurácia), sugerindo que esta classe (“Inadimplentes”) possui um perfil mais distinto que as demais.

Já a classificação entre “Adimplentes e Outros” e “Temporariamente Inadimplentes e Outros” ficou com acurácias de 73,90% e 73,50%, respectivamente, apontando perfis não tão distintos para estas duas classes de clientes.

Apesar da acurácia não ter sido tão significativa, a RNA reduziria significativamente os custos que hoje o Banco XYZ possui com as dívidas dos clientes inadimplentes, tornando esta ferramenta um grande auxílio na tomada de decisão na concessão (ou não) de crédito bancário para determinado cliente.

Referências

- AMARI, S.; MURATA, N.; MÜLLER, K. R.; FINKE, M.; YANG, H. H.** Asymptotic Statistical Theory of Overtraining and Cross-Validation. *IEEE Transactions on Neural Networks*, v. 8, n. 5, p. 985 – 996, 1997.
- 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.
- BANCO NACIONAL DE DESENVOLVIMENTO ECONÔMICO E SOCIAL (Brasil).** *BNDES 60 anos: perspectivas setoriais*. Rio de Janeiro: Banco Nacional de Desenvolvimento Econômico e Social, p. 12 – 41, 2012.
- BARABÁSI, A. L.; ALBERT, R.** Emergence of Scaling in Random Networks. *Science*, v. 286, 1999.
- COUTINHO, L. G.** *A crise financeira internacional, os impactos sobre a economia brasileira e o papel do BNDES*, 13 maio 2009. 39 slides.
- GURNEY, K.** *An Introduction to neural networks*, England: Taylor & Francis e-library, p. 17, 2004.
- HAGAN, M. T.; DEMUTH, H. B.; BEALE, M. H.; JESÚS, O.** *Neural Network Design*, 2ª Edição, p. 3 – 5, 2014.
- HAYKIN, Simon.** *Neural Networks and Learning Machines*. 3a. ed. New Jersey: Prentice Hall, 2009.
- MINISTÉRIO DE MINAS E ENERGIA.** *Caracterização do Cenário Econômico para os próximos 10 anos (2015 – 2024)*. Disponível em: <[http://www.epe.gov.br/mercado/Documents/Série Estudos de Energia/DEA 12-15 NT Cenário econômico 2015-2024vf.pdf](http://www.epe.gov.br/mercado/Documents/Série%20Estudos%20de%20Energia/DEA%2012-15%20NT%20Cenario%20economico%202015-2024vf.pdf)>. Acesso em: 08 de março de 2017.
- PAWLINA, W.** *Ross Histologia Texto e Atlas: Correções com Biologia Celular e Molecular*, 7ª Edição, Rio de Janeiro: Guanabara Koogan, p. 361 – 362, 2016.
- RODRIGUES, T. B.; MACRINI, J. L. R.; MONTEIRO, E. C.** *Seleção de variáveis e classificação de padrões por redes neurais como auxílio ao diagnóstico de cardiopatia isquêmica*. *Pesquisa Operacional*, v. 28, n. 2, p. 285 – 302, 2008.
- 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.
- SOUZA, M. B.** *Rede de aprendizado supervisionado como método de auxílio na detecção do ceratocone*. São Paulo: USP, 2011. 111 p. Programa de Pós-Graduação em Oftalmologia, Universidade de São Paulo, São Paulo, 2011.
- STEINER, M. T. A.** *Redes neurais artificiais*. 2012. 42 slides.
- THE MATHWORKS INC.** *Levenberg-Marquardt Backpropagation*. Natick, 2014. Disponível em: <<http://www.mathworks.com/help/nnet/ref/trainlm.html>>. Acesso em: 05 set. 2017.