JISTEM - Journal of Information Systems and Technology Management
*Revista de Gestão da Tecnologia e Sistemas de Informação*
Vol. 11, No. 2, May/Aug., 2014 pp. *379-396*
ISSN online: 1807-1775
DOI: 10.4301/S1807-17752014000200009

# ANÁLISE DE CRÉDITO POR MEIO DE MINERAÇÃO DE DADOS: APLICAÇÃO EM COOPERATIVA DE CRÉDITO

**Marcos de Moraes Sousa**
Instituto Federal Goiano Campus Ceres, Goiás, Brasil
**Reginaldo Santana Figueiredo**
Universidade Federal de Goiás, Goiás, Brasil

## RESUMO

A busca por eficiência no setor cooperativista de crédito tem levado as cooperativas a adotarem novas tecnologias e novos conhecimentos gerenciais. Dentre tais ferramentas, a Mineração de Dados tem-se destacado nos últimos anos como uma metodologia sofisticada na busca de conhecimento "escondido" nas bases de dados das organizações. Entende-se que o processo de concessão de crédito é uma operação central de uma cooperativa de crédito, assim, o uso de instrumentos que auxiliem são desejados e podem tornar-se fator-chave na gestão do crédito. Os passos utilizados na execução do processo de Descoberta de Conhecimento do presente estudo de caso foram: seleção dos dados; pré-processamento e limpeza dos dados; transformação dos dados; Mineração de Dados; interpretação e avaliação dos resultados. Os resultados foram avaliados por meio de validação cruzada em dez conjuntos e repetidos em dez simulações. Este estudo propôs-se a desenvolver modelos para analisar a capacidade dos cooperados de uma cooperativa de crédito de saldar seus compromissos, utilizando Árvore de Decisão – algoritmo C4.5 e Rede Neural Artificial – algoritmo *Multilayer Perceptrom*. Conclui-se que, para o problema proposto, os modelos tiveram resultados estatisticamente semelhantes e que podem auxiliar no processo decisório da cooperativa.

**Palavras-chave:** Cooperativismo de Crédito; Mineração de Dados; Árvore de Decisão; Rede Neural Artificial.

Manuscript first received/*Recebido em:* 17/07/2012 Manuscript accepted/*Aprovado em*: 21/03/2014

Address for correspondence / Endereço para correspondência

*Marcos de Moraes Sousa*, Professor in Administration, Federal Institute of Goiás; Ph.D. candidate, University of Brasília. Works on projects related to public administration, quantitative methods, and analysis of organizational performance. E-Mail: lceara@hotmail.com

*Reginaldo Santana Figueiredo*, Post-doctoral fellow in Modeling and Simulation, Department of Industrial and Systems Engineering, Texas A&M University, IE-TAMU, USA (2002), Ph.D., Industry Economics, Federal University of Rio de Janeiro (UFRJ); Associate Professor, Federal University of Goiás—UFG. He was also a consultant for the Brazilian government on the analysis of production chains and for the Department of Recreation, Park and Tourism Science of Texas A&M University, in the analysis of tourism's socioeconomic impact, using modeling and simulation techniques. E-mail: santanrf@uol.com.br

## 1. INTRODUÇÃO

O presente artigo trata do desenvolvimento de modelos para analisar a capacidade dos associados de uma cooperativa de crédito de saldar os seus compromissos. Para tal, foram utilizadas técnicas de Mineração de Dados (*Data Mining*).

Para a construção do modelo foi utilizada a base de dados real de cooperados tomadores de crédito de uma cooperativa de crédito do sistema SICOOB (Sistema Cooperativo Brasileiro). Ressalta-se que são dados de difícil acesso e coleta.

O cooperativismo de crédito é uma sociedade de pessoas e deve ser norteado por uma finalidade social. Entretanto, é também, uma instituição financeira e é regulamentada pelas normas impostas pelo Conselho Monetário Nacional e pelo Banco Central e, ademais, deve também ter o objetivo de permanência no mercado, o que impõe uma gestão eficiente dos recursos.

O cooperativismo pode ser classificado em duas vertentes: a doutrinária *rochdaleana*, que pretendia transformar a sociedade e reformar o homem; e a teórica, desenvolvida na Universidade de Münster (Alemanha), utilizando o instrumental da ciência da administração de empresa, vislumbrando a cooperativa como uma empresa moderna (Pinho, 2004).

Na perspectiva teórica, a Teoria de Münster é a que mais se desenvolveu, também conhecida por "Teoria Econômica da Cooperação Cooperativa", com origem no Instituto de Cooperativismo da Universidade de Münster, na Alemanha. Professores desta universidade, conjuntamente com pesquisadores latino-americanos, em oposição aos pressupostos doutrinários rochdaleanos, desenvolveram esta "Escola", cuja fundamentação metodológica advém do racionalismo crítico (Pinho, 2004).

Pinho (1982, p. 75) expõe, segundo as ideias de Boettcher, que o seguinte conceito de cooperativa baseado nos axiomas e pressupostos da Teoria de Münster: "as cooperativas são agrupamentos de indivíduos que defendem seus interesses econômicos individuais por meio de uma empresa que eles mantêm conjuntamente". Neste contexto, Frantz (1985:56) acrescenta que a cooperativa também pode ser compreendida como a definição de uma "[...] estratégia de competição com o objetivo de maximizar os resultados da ação econômica individual de cada produtor [...]".

Esta pesquisa vislumbra as cooperativas de crédito na ótica do cooperativismo teórico e, partindo dos pressupostos e axiomas desenvolvidos, a análise da informação para tomada de decisões é condição central. Ferramentas e metodologias que visam à análise de informações gerenciais têm evoluído muito nas últimas décadas.

A gestão de uma cooperativa de crédito é complexa, pois necessita manter o equilíbrio entre os anseios e necessidades dos cooperados e competir no mercado. As características de associação para os cooperados e de empresa para o mercado devem estar em certo equilíbrio.

O número de cooperados e de cooperativas vêm aumentando paulatinamente. Segundo dados da OCB (2014), existem hoje no Brasil, 1.047 cooperativas de crédito singular e 4.529 pontos de atendimento. O SICOOB é o maior sistema de crédito

cooperativo do Brasil, congrega 529 cooperativas singulares e 1.949 pontos de atendimento cooperativo (Portal do Cooperativismo de Crédito, 2014).

O ambiente dinâmico e competitivo do mercado financeiro brasileiro concomitante com mudanças na oferta de crédito nos últimos anos exige a adoção de uma postura profissional, o que conduz as cooperativas de crédito a adotarem o uso de novas tecnologias e conhecimentos gerenciais.

Oliveira (2001) aponta a profissionalização de cooperados e de cooperativas como uma tendência relevante. O setor tem se desenvolvido de forma rápida, adotando uma estratégia de integração por meio de cooperativas centrais e necessitam, neste sentido, estar altamente em sintonia com o que há de mais eficiente em ferramentas de gestão.

Analisar o crédito constitui certamente um dos pontos mais importantes em instituições financeiras. Chaia (2003) destaca a importância da definição do tipo de análise a ser feito e da abrangência da mesma e ainda alerta para o perigo de copiar e utilizar modelos de outras instituições, resultando assim em avaliações inadequadas.

Um dos principais métodos de avaliação de crédito utilizado pelas instituições financeiras é o *credit scoring*. Chaia (2003, p. 23) define este modelo como o uso de ferramental estatístico na identificação dos fatores determinantes da probabilidade de o cliente tornar-se inadimplente, e aponta como principal vantagem o fato de que "[...] decisões sobre a concessão são tomadas com base em procedimentos impessoais e padronizados, gerando um maior grau de confiabilidade".

Com a concepção de cooperativa discutida anteriormente, torna-se altamente relevante o uso de tais metodologias objetivas na concessão de crédito, tal como o *credit scoring*. Evita-se que a decisão seja tomada somente pela avaliação em julgamentos subjetivos. Koh, Tan, e Goh (2006) condiciona o progresso do *credit scoring* ao aumento da competitividade, avanços na tecnologia computacional e no aumento exponencial de grandes bancos de dados.

Mester (1997) indica que a exatidão do modelo, a atualização dos dados e a avaliação e readequação dos modelos são alguns fatores críticos do *credit scoring*. Falhas nesses fatores limitam o uso de tal modelo. Tendo em vista que a concessão de crédito é um dos processos centrais das cooperativas de crédito, a análise de tal processo caracteriza-se como ponto fundamental para proteger o patrimônio coletivo da cooperativa.

A obtenção de ferramentas que classifiquem e ajudem a prever comportamentos de futuras concessões é fundamental para a gestão de crédito, com a vantagem de diminuir a subjetividade no processo, permitir a condução mais eficiente dos recursos e proporcionar maior celeridade nas propostas.

Estudos de verificação de análise de crédito por meio de mineração de dados aumentam a precisão dos modelos e foram realizados por vários autores nos últimos anos (Abellán & Mantas, 2014; Akkoç, 2012; Bhattacharyya, Jha, Tharakunnel, & Christopher, 2011; Chang & Yeh, 2012; Chen & Huang, 2011; Crone & Finlay, 2012; Cubiles-De-La-Vega, Blanco-Oliver, Pino-Mejías, & Lara-Rubio, 2013; García, Marqués, & Sánchez, 2012; Han, Han, & Zhao, 2013; Koh et al., 2006; Kruppa,

Schwarz, Arminger, & Ziegler, 2013; Lai, Yu, Wang, & Zhou, 2006; Lemos, Steiner, & Nievola, 2005; Majeske & Lauer, 2013; Marqués, García, & Sánchez, 2012; Nie, Rowe, Zhang, Tian, & Shi, 2011; Oreski & Oreski, 2014; Saberi et al., 2013; Wang, Ma, Huang, & Xu, 2012; Xiong, Wang, Mayers, & Monga, 2013; Yap, Ong, & Husain, 2011; Zhong, Miao, Shen, & Feng, 2014; Zhou, Jiang, Shi, & Tian, 2011; Zhu, Li, Wu, Wang, & Liang, 2013).

Apesar do crescente interesse, a aplicação dessas ferramentas em cooperativas ainda é pouco realizado. Khatchatourian e Treter (2010) aplicaram lógica *Fuzzy* na análise do desempenho financeiro em cooperativas de produção do Rio Grande do Sul. Zhu, Li, Wu, Wang, e Liang (2013) utilizaram *Support vector machine* - máquina de vetores de suporte – na análise de crédito em cooperativa de crédito de Barbados.

Há atualmente diversas técnicas de Mineração de Dados disponíveis. Assim, pretendeu-se examinar qual metodologia de mineração oferece melhores resultados na análise de crédito para Cooperativas de Crédito. Neste sentido, indaga-se, se um modelo de Mineração de Dados pode ter bom desempenho na classificação e previsão na gestão de crédito em cooperativas de crédito.

## 2. REFERENCIAL TEÓRICO

A terminologia "Descoberta de Conhecimento em Base de Dados" (*Knowledge Discorery in Databases – KDD)* foi utilizada pela primeira vez em 1989 para destacar que o conhecimento é o produto final do processo de descoberta em base de dados (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

Os termos *KDD* e Mineração de Dados foram entendidos por muitos pesquisadores como sinônimos até 1995 (Lemos et al., 2005). Fayyad et al. (1996) conceitua e distingue *KDD* e Mineração de Dados da seguinte forma: o primeiro refere-se ao processo geral de descobrir conhecimento útil dos dados e o segundo à aplicação específica de algoritmos para a extração de padrões e modelos dos dados. No conceito destes autores, Mineração de Dados seria, então, um passo no processo de *KDD*, consistindo de empregar análise de dados e algoritmos na produção de um conjunto particular de padrões e modelos.

Fayyad et al. (1996) denomina padrões como componentes dos modelos. Neste estudo foi utilizado o conceito de modelo definido por Pidd (1998, p. 23): "Modelo é uma representação externa e explícita de parte da realidade vista pela pessoa que deseja usar aquele modelo para entender, mudar, gerenciar e controlar parte daquela realidade".

Goldschmidt e Passos (2003, p. 6) dividem as atividades de *KDD* em três grupos: (i) desenvolvimento tecnológico – este grupo compreende "[...] as iniciativas de concepção, aprimoramento e desenvolvimento de algoritmos, ferramentas e tecnologias de apoio [...]" no processo de *KDD*; (ii) execução de *KDD* – este grupo inclui atividades relacionadas à utilização dos algoritmos, ferramentas e tecnologias desenvolvidas na procura de conhecimento; (iii) aplicação de resultados – com os modelos desenvolvidos na execução de *KDD*, "[...] as atividades se voltam à aplicação dos resultados no contexto em que foi realizado o processo de *KDD*".

É comum a comparação de técnicas e modelos híbridos ou compostos. Dentre os modelos de análise de crédito e risco destacam-se o uso de: (a) regressão logística (Akkoç, 2012; Bhattacharyya et al., 2011; Cubiles-De-La-Vega et al., 2013; Han et al., 2013; Ju & Sohn, 2014; Koh et al., 2006; Kruppa et al., 2013; Nie et al., 2011; Wang et al., 2012; Yap et al., 2011); (b) árvores de decisão (Abellán & Mantas, 2014; Bhattacharyya et al., 2011; Chen & Huang, 2011; Crone & Finlay, 2012; Koh et al., 2006; Kruppa et al., 2013; Lemos et al., 2005; Nie et al., 2011; Wang et al., 2012; Yap et al., 2011); (c) redes neurais (Akkoç, 2012; Chen & Huang, 2011; Koh et al., 2006; Lai et al., 2006; Nie et al., 2011; Oreski & Oreski, 2014; Saberi et al., 2013; Wang et al., 2012); (d) *Support vector machine* - máquina de vetores de suporte (Bhattacharyya et al., 2011; Nie et al., 2011; Xiong et al., 2013; Zhong et al., 2014; Zhu et al., 2013); (e) métodos *ensemble* (Abellán & Mantas, 2014; García et al., 2012; Marqués et al., 2012; Nie et al., 2011; Wang et al., 2012).

O contexto real de aplicação dos estudos foram coletados em organizações do Canadá (Xiong et al., 2013), Alemanha (Han et al., 2013; Koh et al., 2006), Croácia (Oreski & Oreski, 2014), Peru (Cubiles-De-La-Vega et al., 2013), China (Nie et al., 2011), Turquia (Akkoç, 2012) e Barbados (Zhu et al., 2013). Análise de crédito por meio de mineração de dados ainda é escassa no Brasil, Lemos et al. (2005) verificaram a aplicação da análise de crédito bancário com a mesma metodologia e utilizaram como *lócus* uma agência do Banco do Brasil.

As árvores de decisão constituem um dos principais e mais populares métodos de Mineração de Dados (Wang et al., 2012). Este método, conforme Lemos et al. (2005, p. 229) é o único a exibir resultados em forma hierárquica, "[...] o atributo mais importante é apresentado na árvore como o primeiro nó, e os atributos menos relevantes são mostradas nos nós subsequentes".

Assim, árvore de decisão é uma estrutura usada para dividir grande quantidade de dados em sucessivos conjuntos menores pela aplicação de uma sequência de regras de decisão (Berry & Linoff, 2004).

A construção de árvores de decisão é especialmente atrativa no ambiente de *KDD*. As causas para tal propensão, abordadas por Gehrke (2003), são: resultado intuitivo e de fácil entendimento; árvores de decisão são não-paramétricas, aplicáveis, portanto, a tratamentos exploratórios; construção relativamente rápida comparada a outros métodos; a acurácia da árvore de decisão pode ser comparada com outros modelos.

É comum a transformação de uma árvore de decisão em regras de decisão. Árvore de Decisão pode ser compreendida como:

> [...] um grafo em que cada nó não folha representa um predicado (condição) envolvendo um atributo e um conjunto de valores. Os nós da folha correspondem à atribuição de um valor ou conjunto de valores a um atributo do problema (Goldschmidt & Passos, 2005, p. 57).

Neste sentido, os caminhos da árvore correspondem a regras do tipo "SE <condições> ENTÃO <conclusão>". Há muitos algoritmos desenvolvidos baseados na

indução de árvores de decisão, dentre os quais se destacam o C4.5, o *CART* (*Classification and Regression Trees*[1]), *QUEST (Quick, unbiased, efficient statistical tree*[2]*)* e *CHAID (chi-square automatic interaction detector*[3]*).*

Rede Neural Artificial – RNA - é um modelo matemático baseado na estrutura cerebral, ordenado em camadas e ligações. As RNAs têm origem em 1943, entretanto, é na década de 1980 que é despertado maior interesse pelo método, tendo como principal fator de desenvolvimento o avanço da tecnologia da informação (Braga, Carvalho, & Ludermir, 2000).

Na perspectiva de Goldschmidt e Passos (2005, p. 175), RNAs podem ser compreendidas como "[...] modelos matemáticos inspirados nos princípios de funcionamento dos neurônios biológicos e na estrutura do cérebro". Tais modelos, conforme os mesmos autores permitem simular capacidades humanas de aprender, generalizar, associar e abstrair.

Braga et al. (2000, p. 1) conceituam RNAs como "sistemas paralelos distribuídos compostos por unidades de processamento simples (nodos) que calculam determinadas funções matemáticas (normalmente não-lineares), [...] dispostas em uma ou mais camadas e interligadas por um grande número de conexões[...]"

A estrutura de uma RNA é, portanto, composta de camadas de neurônios e conexões, que são ponderadas por pesos. Conforme a Figura 01, os neurônios são representados pelos nodos e os pesos são representados pelas setas.

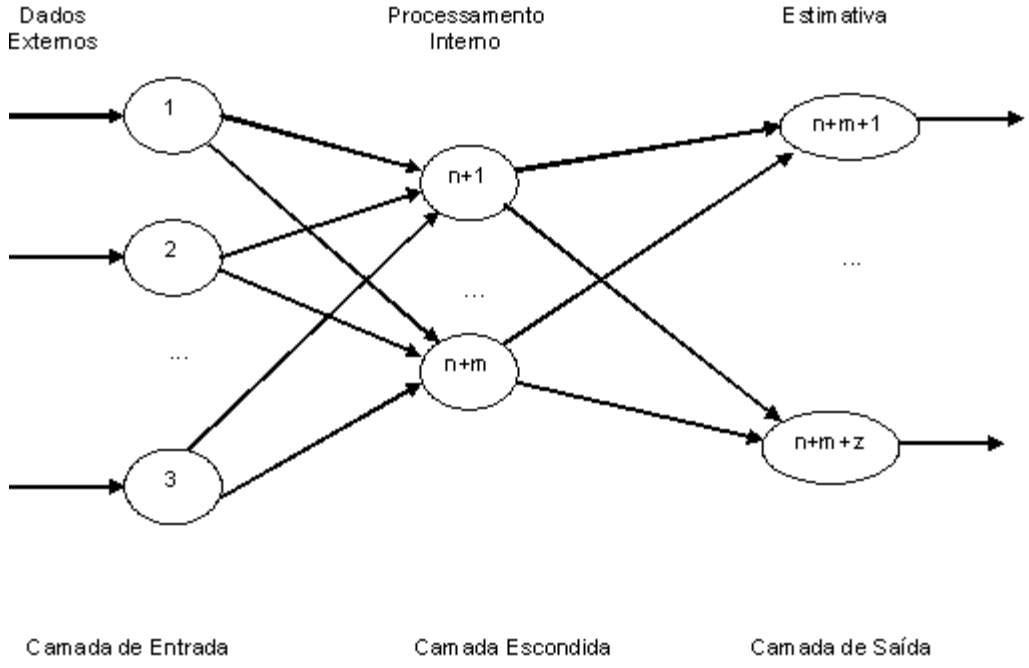


Figura 01 – Arquitetura de uma RNA.
Fonte: Goldschmidt & Passos (2005).

1 Árvores de classificação e regressão.
2 Árvore estatística eficiente, rápida, sem viés.
3 Detector interativo automático qui-quadrado.

Há tipicamente três partes no processamento de RNA: a camada de entrada ou *input layer*, por onde são recebidos os dados; a camada interna, comumente chamada de "camada escondida" ou *hidden layer*, responsável pelo processamento dos dados, esta parte pode conter mais de uma camada e a camada de saída ou *output layer*, representando o resultado (Larose, 2005).

O primeiro passo da aplicação de um RNA é a etapa de aprendizagem da rede, onde há o ajuste dos parâmetros. Este aprendizado pode ser classificado em duas categorias: supervisionado e não-supervisionado, o primeiro ocorre quando é fornecido variáveis de saída, o segundo não necessita da variável alvo.

Braga et al. (2000, p. 227) indicam como pontos positivos que, suscitam interesse pelo método, a habilidade de aprendizado e posterior generalização, com a possibilidade de mapear funções multivariadas, a auto-organização, o processo de séries temporais, a possibilidade do uso de grande número de variáveis de entrada, a possibilidade do uso de amostragens e por ser caracterizada como um modelo não-paramétrico, portanto, e ainda ressaltam que, "[...] não há grande necessidade de se entender o processo propriamente dito". Entretanto, este último aspecto é também considerado pelos mesmos autores como a principal crítica, ou seja, a inabilidade do modelo em esclarecer de que maneira os resultados são gerados. Devido a esta especificidade, as RNAs são também denominadas de "caixas pretas".

## 3. METODOLOGIA

Para a realização da pesquisa optou-se pelo estudo de caso que, segundo Yin (2010), é oportuno para estudar acontecimentos contemporâneos em um contexto da vida real, quando o controle se torna mais difícil para o pesquisador. O estudo de caso é caracterizado como do tipo único, contemplando uma unidade de análise, envolvendo uma Cooperativa de Crédito do sistema SICOOB, estruturado de forma a contemplar a investigação da Cooperativa sob o enfoque do cooperativismo teórico.

A base de dados da cooperativa foi utilizada para avaliar o desempenho do sistema de análise de crédito. Essa base de dados corresponde aos dados históricos das análises de pessoas físicas de 2003 a 2007. Devido à mudança no sistema de informações, não é possível coleta dos dados anteriores a este período.

Os dados referentes à análise de crédito são altamente confidenciais e estratégicos devido ao sigilo bancário e também ao risco dos concorrentes adquirirem tais dados, torna-se, portanto, muito difícil de serem adquiridos por terceiros. A opção pela cooperativa foi, então, por sua disponibilidade em fornecer os dados.

Atualmente, a cooperativa utiliza como ferramenta para realizar sua análise de crédito, um aplicativo do sistema SICOOB chamado SisBr. É esse aplicativo que contém as informações que a gerência e diretoria se valem para fundamentar suas decisões de conceder ou não crédito.

A pesquisa utilizou os passos sugeridos por Fayyad, Piatetsky-Shapiro, e Smyth (1996) no processo de Descoberta de Conhecimento: seleção dos dados; pré-

processamento e limpeza dos dados; transformação dos dados; Mineração de Dados; interpretação e avaliação dos resultados.

Para a consecução dos objetivos, este estudo se apoiou nas atividades de execução de *KDD*, conforme discutido por Goldschmidt e Passos (2005). Dentre as técnicas de Mineração de Dados disponíveis, foram utilizadas Redes Neurais e Árvores de Decisão, ambas encontradas extensivamente nos estudos empíricos.

A coleta e seleção dos dados correspondem ao processo de captar, organizar e selecionar os dados disponíveis para a etapa da modelagem e Mineração de Dados, portanto, requer exame acurado. Dasu e Johnson (2003) apontam os seguintes elementos auxiliadores na análise de variáveis: a experiência anterior, o conhecimento, a quantidade e a qualidade dos dados.

## 3. RESULTADOS

Nesta parte do trabalho encontram-se os resultados das simulações das técnicas investigadas neste artigo, Árvore de Decisão e Redes Neurais Artificiais, bem como o teste estatístico comparativo entre ambas.

Não foram encontrados valores ausentes na base de dados estudada. O Quadro 01 apresenta a estrutura da base de dados construída, com as variáveis e valores possíveis.

Quadro 01 - Estrutura da base de dados.

| CÓDIGO | VARIÁVEIS | VALOR |
|---|---|---|
| 01 | Código do cooperado | Chave numérica, única para cada cooperado. |
| 02 | Gênero | 1= masculino<br>2= feminino |
| 03 | Idade | Valor numérico |
| 04 | Nível educacional | 1= pós-graduação<br>2= superior completo<br>3= superior incompleto<br>4= segundo grau completo<br>5= segundo grau incompleto<br>6= primeiro grau completo<br>7= primeiro grau incompleto |
| 05 | Cidade | Nome da cidade onde mora |
| 06 | Naturalidade | Nome da cidade onde nasceu. |
| 07 | Local de domicílio | 1= zona urbana<br>2= zona rural |
| 08 | Principal atividade | Nome da atividade |
| 09 | Estado civil | 1= casado<br>2= solteiro<br>3= viúvo<br>4= separado judicialmente<br>5= outros |
| 10 | Capital | Valor numérico |
| 11 | Relacionamento | 1= opera com a cooperativa há mais de 3 anos<br>2= opera com a cooperativa de 1 a 3 anos<br>3= opera com a cooperativa até 1 ano |

Quadro 01 - Estrutura da base de dados. (cont.)

| CÓDIGO | VARIÁVEIS | VALOR |
|---|---|---|
| 12 | Comportamento das operações | 1= normal<br>2= atrasos esporádicos<br>3= atrasos/renegociações constantes |
| 13 | Tempo de experiência na atividade/emprego | 1= mais de 5 anos<br>2= de 2 a 5 anos<br>3= até 3 anos |
| 14 | Consultas cadastrais | 1= ausência de restrições<br>2= com restrições irrelevantes justificadas<br>3= com restrições relevantes ou irrelevantes sem justificativas |
| 15 | Informações cadastrais na cooperativa | 1= cadastro atualizado e confiável<br>2= cadastro atualizado e não confiável<br>3= informações desatualizadas ou ausência de informações |
| 16 | Finalidade da operação | 1= custeio e investimento<br>2= financiamento de bens<br>3= crédito pessoal/cheque especial<br>4= renovações/composições de dívidas |
| 17 | Garantia das operações | 1= hipoteca – capital social<br>2= alienação fiduciária/warrants<br>3= penhor cedular/caução de títulos<br>4= pessoal |
| 18 | Liquidez das garantias | 1= garantia de fácil liquidez (venda até 6 meses)<br>2= garantia de média liquidez (venda de 6 a 12 meses)<br>3= garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) |
| 19 | Que frequência o cooperado opera (operações ativas) | 1= nunca<br>2= frequentemente<br>3= permanentemente |
| 20 | Valor da operação | 1= até 1% do Patrimônio Líquido Ajustado (PLA)<br>2= de 1,01% a 2% do PLA<br>3= de 2,01% a 3% do PLA<br>4= mais de 3% do PLA |
| 21 | Nível de comprometimento – prestações com renda Líquida do cooperado | 1= até 20% da renda média líquida<br>2= de 20% a 30% da renda média líquida<br>3= mais de 30% da renda líquida |
| 22 | Patrimônio líquido pessoal livre de relacionamento com endividamento total | 1= mais de 4 vezes<br>2= de 2 a 4 vezes<br>3= sem patrimônio pessoal ou até 2 vezes |
| 23 | Endividamento total em relação à renda líquida ano | 1= até 2 vezes<br>2= de 2 a 4 vezes<br>3= mais de 4 vezes |
| 24 | Total endividamento em relação ao capital integralizado | 1= até 4 vezes<br>2= de 4 a 8 vezes<br>3= de 8 a 12 vezes<br>4= acima de 12 vezes |
| 25 | Perfil da atividade econômica do associado | 1= ótima<br>2= boa<br>3= regular<br>4= ruim |

Quadro 01 - Estrutura da base de dados (Cont.)

| CÓDIGO | VARIÁVEIS | VALOR |
|---|---|---|
| 26 | Risco atribuído pela cooperativa | 1= AA<br>2= A<br>3= B<br>4= C<br>5= D<br>6= E<br>7= F<br>8= G<br>9= H |
| 27 | Resultado-Julho 2007 | 1= adimplente<br>2= inadimplente |
| 28 | Resultado-Agosto 2007 | 1= adimplente<br>2= inadimplente |
| 29 | Resultado-Setembro 2007 | 1= adimplente<br>2= inadimplente |
| 30 | Resultado-Outubro 2007 | 1= adimplente<br>2= inadimplente |
| 31 | Resultado-Novembro 2007 | 1= adimplente<br>2= inadimplente |
| 32 | Resultado-Dezembro 2007 | 1= adimplente<br>2= inadimplente |
| 33 | Resultado-Janeiro 2008 | 1= adimplente<br>2= inadimplente |
| 34 | Resultado-Fevereiro 2008 | 1= adimplente<br>2= inadimplente |
| 35 | Resultado-Março 2008 | 1= adimplente<br>2= inadimplente |
| 36 | Resultado-Abril 2008 | 1= adimplente<br>2= inadimplente |
| 37 | Resultado-Maio 2008 | 1= adimplente<br>2= inadimplente |
| 38 | Resultado-Junho 2008 | 1= adimplente<br>2= inadimplente |
| 39 | Resultado agregado | 1= adimplente<br>2= inadimplente |

Fonte: Dados da pesquisa, 2010.

A variável de saída é representada pelas variáveis 27 a 39, correspondentes ao período de julho de 2007 a junho de 2008. As variáveis 02 a 10 não constam na análise de crédito realizada pela cooperativa, assim, foram agregadas com a finalidade de ampliar a análise. A coleta destas variáveis foi realizada nos cadastros dos cooperados e representam os dados disponíveis pela cooperativa.

As variáveis 11 a 26 são utilizadas atualmente na análise de crédito e representam o comportamento histórico do cooperado na tomada de crédito. Os códigos 27 a 39 representam a variável de saída adotada pela pesquisa e retratam o período de julho de 2007 a junho de 2008. Estes são os dados disponibilizados pela cooperativa. Dados anteriores a estes não estão disponíveis.

A quantidade de variáveis utilizadas na análise enquadra-se dentro do utilizado em outras pesquisas. Por exemplo, Koh et al. (2006) utilizaram 20 variáveis e Lemos et al. (2005) utilizaram 24.

A variável, código do cooperado foi descartada, pois, foi útil somente para identificar o cooperado na coleta dos dados. A variável "risco atribuído" foi utilizada somente na etapa de pré-processamento e limpeza dos dados. Não foi utilizada nas etapas de transformação e modelagem porque esta variável se refere à saída do modelo utilizado pela cooperativa, portanto, representa o resultado do modelo atualmente empregado. A variável, resultado agregado, representa o resultado do período analisado (julho de 2007 a junho de 2008) e, segundo as regras do negócio da cooperativa e dos objetivos de estudo, constitui a variável-alvo de saída do modelo.

A transformação dos dados visa auxiliar a execução das técnicas de mineração de dados. Os dados foram agrupados segundo indicação de Goldschmidt & Passos (2005) em uma única tabela bidimensional.

Os dados foram coletados de duas fontes na cooperativa: da avaliação de crédito e do cadastro, de forma manual, registro a registro. Para a pesquisa foram utilizados dados históricos de 211 cooperados, pessoas físicas, sendo 22 inadimplentes e 189 adimplentes. Os dados representam o universo dos cooperados tomadores de crédito.

Dado o desbalanceamento entre inadimplentes e adimplentes, é possível incorrer em viés e problema de superajustamento (Chawla, 2005; Horta, Borges, Carvalho, & Alves, 2011). Para solucionar tal problema foi utilizada a técnica denominada SMOTE (*Synthetic Minority Oversampling Technique*) (Chawla, 2005) para inclusão de observações de cooperados inadimplentes.

Esse algoritmo é considerado um dos mais utilizados pela literatura (Horta et al., 2011). Dessa forma foram criadas 110 observações da classe minoritária, ou seja, inadimplentes.

A base ficou com 132 cooperados inadimplentes e 189 adimplentes, constituindo uma amostra de 321 observações. Posteriormente, a base de dados foi randomizada para evitar a concentração de mesmos valores em determinado conjunto de dados na validação cruzada e incorrer em superajustamento.

Para a implementação computacional das técnicas Árvore de Decisão e Redes Neurais, foi utilizada a base de dados descritos neste estudo, considerando-se para cada cooperado as variáveis já descritas anteriormente. Para exemplificar, foi escolhido um modelo gerado pela técnica Árvore de Decisão, para transformar em regras de decisão na fase pós-processamento. Optou-se por utilizar a ferramenta computacional de domínio público *WEKA* (*Waikato Environment for Knowledge Analysis)*.

Goldschmidt e Passos (2005, p. 50) argumentam que, para melhor fidedignidade da avaliação do modelo de conhecimento, "[...] os dados utilizados na construção do modelo não devem ser os mesmos utilizados na avaliação desse modelo". Os mesmos autores ainda afirmam que deve haver no mínimo duas partições: a partição de treinamento e a partição de teste. A primeira inclui os dados para a construção do modelo e a segunda, os dados para avaliação.

Dividir o conjunto de dados tem o propósito de simplificar, sumarizar e reduzir a variabilidade e tamanho da base de dados, resultando na seleção de modelos mais sofisticados e acurados (Dasu & Johnson, 2003).

Neste estudo, para melhor isenção da avaliação, tanto para a árvore de decisão quanto para a RNA, foi utilizada a validação cruzada com K conjuntos (*K-Fold Cross-Validation).* Segundo Goldschmidt e Passos (2005, p. 51), neste método, a base de dados é dividida aleatoriamente com N elementos em K subconjuntos separados, "cada um dos K subconjuntos é utilizado como conjunto-teste e os (K-1) demais subconjuntos são reunidos em um conjunto de treinamento. O processo é repetido K vezes, sendo gerados e avaliados K modelos [...]". Os dados foram divididos em dez conjuntos e repetidos em dez simulações conforme proposto por Witten e Frank (2005). A validação cruzada é encontrada em vários estudos de análise de crédito (Akkoç, 2012; Chang & Yeh, 2012; Han et al., 2013)

Para este estudo, decidiu-se por utilizar RNA de múltiplas camadas, *Multilayer Perceptron* (*MLP*), com o algoritmo de aprendizagem *backpropagation*. A quantidade de neurônios da camada de entrada foi de 66, da camada intermediária 2 e a quantidade de neurônios na camada de saída foi igual a 2. Em todos os testes foi utilizada taxa de aprendizagem igual a 0.01, dado que segundo as simulações foi a melhor taxa que melhorou a classificação e também foi utilizada por Lemos et al., (2005). Optou-se por não utilizar a taxa *momentum* conforme Lemos et al. (2005). Além disso, o acréscimo dessa taxa não melhorou o desempenho da classificação.

O aprendizado da RNA é do tipo supervisionado. Ferreira (2005, p. 37) descreve este tipo da seguinte forma: "[...] a rede é treinada através do fornecimento dos valores de entrada e dos respectivos valores de saída [...]".

Para a avaliação comparativa dos modelos foi utilizado o parâmetro percentual total de valores preditos corretamente com o teste estatístico *t* modificado (*corrected resample t-test*) com nível de significância de 0.05 (ou 5%) em duas caudas e nove graus de liberdade, conforme proposto por Witten e Frank (2005) de acordo com a fórmula 1, apresentada abaixo:

$$t = \frac{\bar{d}}{\sqrt{\left(\frac{1}{k} + \frac{n_2}{n_1}\right)\sigma_d^2}} \quad (1)$$

Em que:

$\bar{d}$ diferença das médias de acerto de valores preditos entre os modelos desenvolvidos.

$k$ número de conjuntos vezes o número de repetições.

$n_2$ número de amostras para teste.

$n_1$ número de amostras para treinamento.

$\sigma_d^2$ variância da diferença das médias.

### 3.1 Árvore de Decisão

Optou-se, nesta pesquisa, por utilizar a ferramenta J4.8, que é a implementação do *software WEKA* do algoritmo da árvore de decisão C4.5. Segundo Goldschmidt e Passos (2005), esta árvore é amplamente utilizada e aceita. Tomou-se um modelo gerado pela técnica Árvore de Decisão para exemplificar as regras e matriz de confusão. O modelo abaixo gerou 41 folhas, ou seja, conjuntos de regras de decisão do tipo **se-então**. Algumas regras do primeiro conjunto estão apresentadas abaixo.

- **Se** Liquidez das garantias = garantia de fácil liquidez (venda até 6 meses) **então** inadimplente.
- **Se** Liquidez das garantias = garantia de média liquidez (venda de 6 a 12 meses) **e** Nível de comprometimento = até 20% da renda média líquida **então** adimplente.
- **Se** Liquidez das garantias = garantia de média liquidez (venda de 6 a 12 meses) **e** Nível de comprometimento = de 20% a 30% da renda média líquida **então** adimplente.

O Quadro 02 mostra a matriz de confusão gerada pelo conjunto de teste da árvore avaliada. Esta matriz apresenta as instâncias classificadas em previstos e reais para avaliar o tipo de acerto e erro dos modelos. A diagonal principal indica os valores corretamente classificados. Os valores estão expressos em termos absolutos.

Quadro 02: Matriz de confusão de um modelo desenvolvido pelo método Árvore de Decisão.

| | | **Previstos** | |
|---|---|---|---|
| **Reais** | | **Inadimplente** | **Adimplente** |
| | **Inadimplente** | 121 | 11 |
| | **Adimplente** | 8 | 181 |

Fonte: Dados da pesquisa, 2010.

Neste conjunto, o modelo baseado no algoritmo C4.5 de árvore de decisão classificou 302 registros corretamente, representando assim uma taxa de acerto de 94,08% e 19 observações incorretas, representando 5,92%.

### 3. 2. Rede Neural

A RNA deste estudo, conforme já discutido no referencial teórico, foi constituída por três camadas: entrada, intermediária e saída. O aprendizado da rede foi do tipo supervisionado, pois foram indicados os valores de saída do modelo.

Da mesma forma que a Árvore de Decisão, foi tomado um modelo para exemplificar os resultados gerados pelo *software WEKA.* O Quadro 03 apresenta a matriz de confusão para um modelo desenvolvido.

Quadro 03 – Matriz de confusão de um modelo desenvolvido pelo método RNA.

| | | **Previstos** | |
|---|---|---|---|
| | | **Inadimplente** | **Adimplente** |
| **Reais** | **Inadimplente** | 118 | 14 |
| | **Adimplente** | 13 | 176 |

Fonte: Dados da pesquisa, 2010.

Este modelo baseado em RNA, com o uso do algoritmo *Multilayer Perceptrom*, classificou 294 registros corretamente, representando uma taxa de acerto de 91,59% e 27 observações incorretas, representando 8,41%.

## 3. 3. Avaliação dos Modelos

Nesta seção são avaliados comparativamente os modelos desenvolvidos neste estudo: RNA (algoritmo *MultilayerPerceptrom*) e Árvore de Decisão (algoritmo C4.5). Para a realização desta avaliação, foi utilizado o percentual total de valores preditos corretamente.

A Tabela 01 apresenta o resultado do percentual dos valores preditos corretamente e o respectivo desvio padrão das simulações feitas com os modelos em estudo.

Tabela 01 – Avaliação comparativa.

| **Algoritmo** | **C4.5** | ***MultilayerPerceptrom*** |
|---|---|---|
| Percentual correto | 97,07% | 95,58% |
| Desvio padrão | 2,76 | 3,47 |

Fonte: Dados da pesquisa, 2010.

As simulações feitas na ferramenta *Experimenter* do *WEKA* indicam que a técnica Árvore de Decisão, com a implementação do algoritmo C4.5, é estatisticamente semelhante ao nível de significância de 0.05 em duas caudas que a RNA com a implementação do algoritmo *MultilayerPerceptrom*. A Figura 02 apresenta a saída do *software WEKA* das simulações feitas do problema previamente descrito.

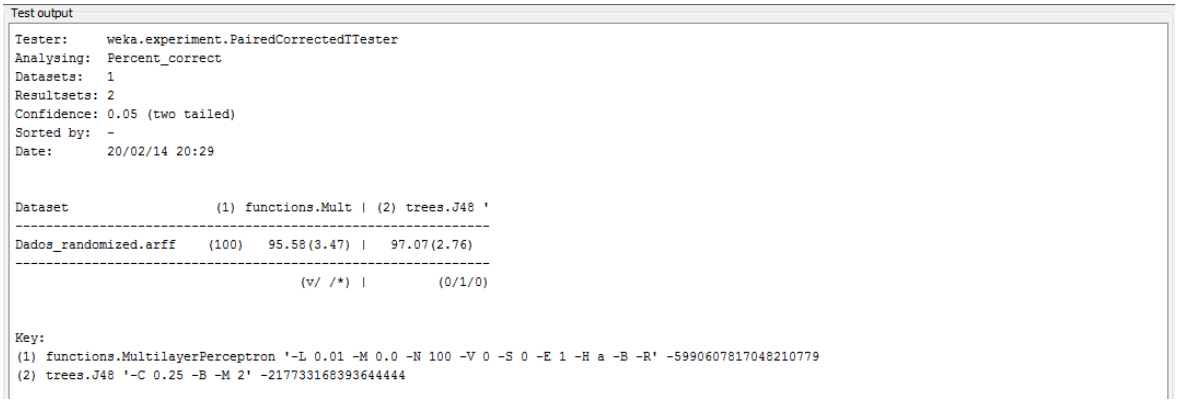


Figura 02: Caixa de saída das simulações da ferramenta *experimenter* do *WEKA*.

Fonte: Dados da pesquisa, 2010.

O desempenho da árvore de decisão para o presente problema foi superior ao encontrado por Yap et al. (2011) que obteve uma taxa de erro de 28,1%.

O estudo de Lemos et al. (2005), apesar de não realizar teste estatístico, encontrou taxa de acerto maior da rede neural em comparação com a árvore de decisão. Apesar da semelhança estatística, as árvores de decisão são consideradas de fácil uso (Lemos et al., 2005).

Os resultados indicam que os modelos de classificação por meio de mineração de dados desenvolvidos podem ser úteis na avaliação pela cooperativa e, dessa forma, melhorar o desempenho, conforme já encontrado na análise em organização de microcrédito (Cubiles-De-La-Vega et al., 2013) e cooperativa de crédito (Zhu et al., 2013).

## 4. CONCLUSÕES

O objetivo deste estudo foi desenvolver e avaliar modelos de Mineração de Dados para classificar e prever o comportamento dos cooperados em saldar os compromissos contraídos. Para o desenvolvimento do modelo foram utilizadas Árvore de Decisão e RNA, ambas, técnicas de Mineração de Dados.

O processo de preparação e modelagem dos dados seguiu os passos sugeridos pela literatura: seleção dos dados; pré-processamento e limpeza dos dados; transformação dos dados; Mineração de Dados; interpretação e avaliação dos resultados. Os dados foram particionados em conjuntos de treinamento e teste.

Embora a Árvore de Decisão tenha obtido 97,07% de acerto nas simulações feitas e a RNA 95,58%, o algoritmo baseado em Árvore de Decisão C4.5 obteve resultado estatisticamente semelhante ao modelo baseado em Redes Neurais Artificiais *MultilayerPerceptrom*.

O processo de descoberta de conhecimento e o uso dos modelos baseados em Mineração de Dados desenvolvidos podem trazer vantagens práticas para a cooperativa. A compreensão das variáveis e seus relacionamentos podem ajudar a melhor classificar e prever o comportamento dos cooperados.

A avaliação mais profunda das variáveis pode ainda ajudar a incluir variáveis que sejam importantes e excluir variáveis que não se mostrem relevantes, com vantagens de proporcionar modelos de gestão de crédito mais sucintos e precisos, com economia de tempo na execução e melhor acurácia nas decisões. A análise de casos discrepantes, ou que estão fora do padrão, pode ser relevante para a formação de uma nova classificação ou, inversamente, compor padrões indesejados.

Pode-se enumerar como limitação deste trabalho: a ausência de bases de dados de outras cooperativas, para comparar, avaliar e validar o modelo; limitações do sistema de informações utilizado pela cooperativa que impossibilita a coleta de dados das variáveis de entrada anteriores a 2003 e limita a fornecer os dados referentes à variável de saída há apenas seis meses; a falta de integração de alguns módulos da base de dados com planilhas eletrônicas.

Sugerem-se as seguintes propostas para estudos posteriores: a utilização de diferentes bases de dados para validação do modelo de análise de crédito; o uso de outras técnicas de Mineração de Dados; o uso de modelos híbridos com a combinação de diferentes técnicas para aperfeiçoar e melhorar o desempenho na classificação e previsão; análises de investimentos com a avaliação do tipo de erro e qual o impacto financeiro que o modelo apresenta para a lucratividade e rentabilidade da cooperativa; avaliação dos casos discrepantes, principalmente para a variável capital, com o objetivo de verificar a existência de novos padrões e classificações.

## REFERÊNCIAS

Abellán, J., & Mantas, C. J. (2014). Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, *41*(8), 3825–3830.

Akkoç, S. (2012). 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*, *222*(1), 168–178.

Berry, M. J. A., & Linoff, G. (2004). *Data mining techniques: For marketing, sales and customer relationship management* (2nd ed.). Indianapolis: wiley Publishing.

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Christopher, J. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, *50*(3), 602–613.

Braga, A. de P., Carvalho, A. P. de L. F., & Ludermir, T. B. (2000). *Redes neurais artificiais: Teoria e aplicações*. Rio de Janeiro: LTC.

Chaia, A. J. (2003). *Modelos de gestão de risco de crédito e sua aplicabilidade ao mercado brasileiro*. Dissertação de Mestrado. FEA/USP.

Chang, S.-Y., & Yeh, T.-Y. (2012). An artificial immune classifier for credit scoring analysis. *Applied Soft Computing*, *12*(2), 611–618.

Chawla, N. V. (2005). Data mining for imbalanced datasets: An overview. In *Data mining and knowledge discovery handbook* (pp. 853–867). New Jersey: Springer.

Chen, S. C., & Huang, M. Y. (2011). Constructing credit auditing and control & management model with data mining technique. *Expert Systems with Applications*, *38*(5359-5365).

Crone, S. F., & Finlay, S. (2012). Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting*, *28*(1), 224–238.

Cubiles-De-La-Vega, M.-D., Blanco-Oliver, A., Pino-Mejías, R., & Lara-Rubio, J. (2013). Improving the management of microfinance institutions by using credit scoring models based on Statistical Learning techniques. *Expert Systems with Applications*, *40*(17), 6910–6917.

Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning*. New Jersey: John Wiley & Sons.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *A I Magazine*, *17*(3), 37–54.

Ferreira, J. B. (2005). *Mineração de dados na retenção de clientes em telefonia celular*. Dissertação de Mestrado. PUC-RIO.

García, V., Marqués, A. I., & Sánchez, J. S. (2012). On the use of data filtering techniques for credit risk prediction with instance-based models. *Expert Systems with Applications*, *39*(18), 13267–13276.

Gehrke, J. (2003). Decision tree. In *The handbook of data mining* (pp. 3–23). New Jersey: Lawrence Erlbaum Associates.

Goldschmidt, R., & Passos, E. (2005). *Data mining: Um guia prático*. Rio de Janeiro: Elsevier.

Han, L., Han, L., & Zhao, H. (2013). Orthogonal support vector machine for credit scoring. *Engineering Applications of Artificial Intelligence*, *26*(2), 848–862.

Horta, R. A. M., Borges, C. C. H., Carvalho, F. A. A., & Alves, F. J. S. (2011). Previsão de insolvência: Uma estratégia para balanceamento da base de dados utilizando variáveis contábeis de empresas brasileiras. *Sociedade, Contabilidade E Gestão*, *6*(2), 21–36.

Ju, Y. H., & Sohn, S. Y. (2014). Updating a credit-scoring model based on new attributes without realization of actual data. *European Journal of Operational Research*, *234*(1), 119–126.

Khatchatourian, O., & Treter, J. (2010). APLICAÇÃO DA LÓGICA FUZZY PARA AVALIAÇÃO ECONÔMICO-FINANCEIRA DE COOPERATIVAS DE PRODUÇÃO. *Revista de Gestão Da Tecnologia E Sistemas de Informação*, *7*(1), 141–162.

Koh, H. C., Tan, W. C., & Goh, C. P. (2006). A two-step method to construct credit scoring models with data mining techniques. *International Journal of Business and Information*, *1*(1), 96–118.

Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, *40*(13), 5125–5131.

Lai, K. K., Yu, L., Wang, S., & Zhou, L. (2006). Credit risk analysis using a reliability-based neural network ensemble model. In *Artificial Neural Networks-ICANN 2006* (pp. 682–690). Springer Berlin Heidelberg.

Larose, T. D. (2005). *Discovering knowledge in data: An introduction to data mining*. New Jersey: John Wiley & Sons.

Lemos, E. P., Steiner, M. T. A., & Nievola, J. C. (2005). Análise de crédito bancário por meio de redes neurais e árvore de decisao: Uma aplicação simples de data mining. *Revista de Administração Da Universidade de São Paulo*, *40*(3), 225–234.

Majeske, K. D., & Lauer, T. W. (2013). The bank loan approval decision from multiple perspectives. *Expert Systems with Applications*, *40*(5), 1591–1598.

Marqués, A. I., García, V., & Sánchez, J. S. (2012). Two-level classifier ensembles for credit risk assessment. *Expert Systems with Applications*, *39*(12), 10916–10922.

Mester, L. J. (1997). What's the point of credit scoring? *Business Review*, *3*, 3–16.

Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*, *38*(12), 15273–15285.

OCB. (2014). Organização das Cooperativas Brasileiras. *Números*. Retrieved February 20, 2014, from http://www.ocb.org.br/site/ramos/credito_numeros.asp

Oliveira, D. P. R. (2001). *Manual de gestão de cooperativas: Uma abordagem prática*. São Paulo: Atlas.

Oreski, S., & Oreski, G. (2014). Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications*, *41*(4), 2052–2064.

Pidd, M. (1998). *Modelagem empresarial: Ferramentas para tomada de decisão*. São Paulo: Atlas.

Pinho, D. B. (1982). *O pensamento cooperativo e o cooperativismo brasileiro*. CNPq/BNCC.

Pinho, D. B. (2004). *O cooperativismo no Brasil: Da vertente pioneira à vertente solidaria*. São Paulo: Saraiva.

Portal do Cooperativismo de Crédito. (2014). *Dados consolidados dos sistemas cooperativos*. Retrieved February 20, 2014, from http://cooperativismodecredito.coop.br/cenario-brasileiro/dados-consolidados-dos-sistemas-cooperativos/

Saberi, M., Mirtalaie, M. S., Hussain, F. K., Azadeh, A., Hussain, O. K., & Ashjari, B. (2013). A granular computing-based approach to credit scoring modeling. *Neurocomputing*, *122*(25), 100–115.

Wang, G., Ma, J., Huang, L., & Xu, K. (2012). Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems*, *26*, 61–68.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco: Elsevier.

Xiong, T., Wang, S., Mayers, A., & Monga, E. (2013). Personal bankruptcy prediction by mining credit card data. *Expert Systems with Applications*, *40*(2), 665–676.

Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness. *Expert Systems with Applications*, *38*(10), 13274–13283.

Yin, Robert, K. (2010). *Estudo de caso: planejamento e métodos* (4th ed.). Porto Alegre: Bookman.

Zhong, H., Miao, C., Shen, Z., & Feng, Y. (2014). Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings. *Neurocomputing*, *128*(27), 285–295.

Zhou, X., Jiang, W., Shi, Y., & Tian, Y. (2011). Credit risk evaluation with kernel-based affine subspace nearest points learning method. *Expert Systems with Applications*, *38*(4), 4272–4279.

Zhu, X., Li, J., Wu, D., Wang, H., & Liang, C. (2013). Balancing accuracy, complexity and interpretability in consumer credit decision making: A C-TOPSIS classification approach. *Knowledge-Based Systems*, *52*, 258–267.