

**UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA E ENGENHARIA DE ALIMENTOS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONEGÓCIO – MESTRADO**

**MINERAÇÃO DE DADOS NA GESTÃO DE CRÉDITO EM COOPERATIVA DE
CRÉDITO**

Marcos de Moraes Sousa

GOIÂNIA
Dezembro, 2008

**UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA E ENGENHARIA DE ALIMENTOS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONEGÓCIO – MESTRADO**

**MINERAÇÃO DE DADOS NA GESTÃO DE CRÉDITO EM COOPERATIVA DE
CRÉDITO**

Orientador:

Dr. Reginaldo Santana Figueiredo

Dissertação apresentada ao Programa de Pós-Graduação em Agronegócio para a obtenção do título de Mestre em Agronegócio. Área de concentração: Sustentabilidade e competitividade dos sistemas agroindustriais.

GOIÂNIA

Dezembro, 2008

Dedico e agradeço a Deus, que permitiu a minha existência, e permitiu a realização do sonho de concluir este projeto.

À minha família, porto seguro onde sempre tenho apoio na hora em que tudo parece falhar.

À minha esposa, pela compreensão, paciência e cumplicidade neste projeto.

AGRADECIMENTOS

Ao Prof. Dr. Reginaldo Santana Figueiredo, um agradecimento especial se faz necessário, não apenas meu orientador, mas, um mestre no melhor sentido da palavra, me influenciou e me ajudou a compreender a vida acadêmica, me despertou para a necessidade de superação do indivíduo e certamente são aprendizados que me acompanharão pela minha vida.

Às secretárias do Mestrado, Cleonice e Tereza, pela forma carinhosa que tratam os mestrandos.

Aos colegas de Mestrado, por dividir os prazeres e obstáculos.

Aos professores do Mestrado, pelas discussões e orientações.

À Cooperativa de Crédito Rural de Rubiataba, na pessoa do presidente, Vanderval José Ribeiro, pela confiança e abertura às informações.

Aos membros da banca, Dr. Ricardo Avelino e Dr. José Antônio Zambon, pela disponibilidade e sugestões.

SUMÁRIO

| | |
|--|----|
| 1 INTRODUÇÃO | 12 |
| 2 REFERENCIAL TEÓRICO..... | 15 |
| 2.1 Vertentes do Cooperativismo..... | 15 |
| 2.2 Análise de Crédito..... | 17 |
| 2.3 Descoberta de Conhecimento e Mineração de Dados..... | 18 |
| 2.3.1 Redes Neurais..... | 20 |
| 2.3.2 Árvores de Decisão | 21 |
| 3 METODOLOGIA | 23 |
| 3.1 Descrição do Caso | 23 |
| 3.2 Coleta e Seleção dos Dados | 24 |
| 3.3 Pré-processamento e Limpeza dos Dados..... | 27 |
| 3.4 Transformação dos Dados | 30 |
| 3.5 Modelagem e Mineração dos Dados | 30 |
| 3.6 Avaliação dos Modelos..... | 31 |
| 4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS | 33 |
| 4.1 Pré-processamento e Limpeza dos Dados..... | 33 |
| 4.2 Modelagem e Mineração de Dados..... | 43 |
| 4.2.1 Árvore de Decisão | 45 |
| 4.2.2 Redes Neurais | 51 |
| 4.2.3 Modelo Empregado pela Cooperativa..... | 54 |
| 4.2.4 Avaliação Comparativa dos Modelos..... | 55 |
| 5 CONCLUSÕES E PROPOSTAS PARA ESTUDOS POSTERIORES..... | 57 |
| Referências..... | 59 |
| Apêndice I..... | 62 |
| Apêndice II..... | 65 |

LISTA DE FIGURAS

| | |
|--|----|
| Figura 01 – Arquitetura de uma RNA..... | 20 |
| Figura 02 – Boxplot da variável capital..... | 34 |
| Figura 03 – Histograma de freqüências da variável capital..... | 35 |
| Figura 04 – Boxplot da variável idade | 36 |
| Figura 05 – Histograma de freqüências da variável idade..... | 37 |
| Figura 06 – Gráfico de rede: variável de saída e liquidez das garantias | 39 |
| Figura 07 – Gráfico de rede: variável de saída e garantia das operações..... | 39 |
| Figura 08 – Gráfico de rede: variável de saída e risco atribuído..... | 40 |
| Figura 09 – Gráfico de rede: variável de saída e finalidade da operação | 41 |
| Figura 10 – Gráfico de rede: variável de saída e nível de comprometimento | 42 |
| Figura 11 – Gráfico de rede: variável de saída e principal atividade do cooperado..... | 43 |
| Figura 12 – Modelo de mineração de dados desenvolvido | 44 |
| Figura 13 – Árvore de decisão-avaliação da categoria garantia de média liquidez..... | 46 |
| Figura 14 – Árvore de decisão-avaliação da categoria garantia pessoal ou de liquidez..... | 47 |

LISTA DE QUADROS

| | |
|---|----|
| Quadro 01 – Estrutura da base de dados | 24 |
| Quadro 02 – Tipos de variáveis da base de dados | 27 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 01 – Parâmetros da variável discreta e contínua..... | 33 |
| Tabela 02 – Seleção das variáveis da base de dados..... | 37 |
| Tabela 03 – Avaliação do modelo desenvolvido pelo método Árvore de Decisão..... | 49 |
| Tabela 04 – Matriz dos valores preditos e reais da partição treinamento da Árvore de Decisão..... | 50 |
| Tabela 05 – Matriz dos valores preditos e reais da partição teste da Árvore de Decisão..... | 50 |
| Tabela 06 – Matriz dos valores preditos e reais da partição validação da Árvore de Decisão..... | 51 |
| Tabela 07 – Avaliação da importância relativa das variáveis pelo método RNA..... | 52 |
| Tabela 08 – Avaliação do modelo desenvolvido pelo método RNA..... | 52 |
| Tabela 09 – Matriz dos valores preditos e reais da partição treinamento da RNA..... | 53 |
| Tabela 10 – Matriz dos valores preditos e reais da partição teste da RNA..... | 53 |
| Tabela 11 – Matriz dos valores preditos e reais da partição validação da RNA..... | 54 |
| Tabela 12 – Avaliação do modelo atual da cooperativa..... | 54 |
| Tabela 13 – Matriz dos valores preditos e reais do modelo utilizado..... | 55 |
| Tabela 14 – Avaliação comparativa..... | 55 |

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. Este estudo propõe-se a desenvolver modelos para analisar a capacidade dos cooperados em saldar seus compromissos, utilizando Árvore de Decisão e Rede Neural Artificial e, ao mesmo tempo, avaliar o modelo utilizado pela cooperativa. Conclui-se que para o problema proposto, as técnicas tiveram bom desempenho e que podem auxiliar no processo decisório da cooperativa.

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

ABSTRACT

The search for efficiency in the cooperativist credit sector has led the cooperatives to adopt new technologies and new managerial knowledge. Among such tools, the Data Mining techniques have outstanding in recent years as a sophisticated methodology in the search of hidden "knowledge" in the databases of the organizations. The credit concession process is a central operation in a credit cooperative, in this way, the use of instruments that aid this operation are wanted and they can become a key factor in the credit management. This study has in view to develop models to analyze the capacity of the cooperated in liquidating the commitments, by using Decision Tree and Artificial Neural Net and, at the same time, to evaluate the model used by the cooperative. This study concluded that for the problem proposed, the techniques had good performances and that can aid the cooperative decision process.

Keywords: Credit Cooperative, Data Mining, Decision Tree, Artificial Neural Net.

1. INTRODUÇÃO

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

Concomitante ao desenvolvimento, de tal modelo, pretende-se fazer uma avaliação comparativa com o modelo utilizado pela cooperativa.

Para a construção do modelo foi utilizada a base de dados real de cooperados tomadores de crédito de uma cooperativa de crédito rural do sistema SICOOB (Sistema Cooperativo Brasileiro).

O cooperativismo de crédito surgiu historicamente para superar dificuldades agrícolas. Os alemães foram os primeiros a constituírem cooperativas de crédito. A primeira foi criada em 1850, justamente na modalidade rural.

A primeira cooperativa de crédito constituída no Brasil foi uma cooperativa rural em 1902, no Rio Grande do Sul. As cooperativas de crédito, desde seu início no Brasil, desempenharam papel complementar para outras cooperativas, principalmente cooperativas agrícolas.

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.

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 cooperativismo pode ser classificado em duas correntes, uma vertente doutrinária e uma teórica. A primeira tem sua origem na experiência dos pioneiros de Rochdale, Inglaterra. Essa corrente contribuiu no sentido de fornecer ao cooperativismo mundial o ideário e os princípios doutrinários que são referências utilizadas atualmente, entretanto, não se preocupou em analisar a competitividade das cooperativas. Neste sentido, a corrente teórica, mais precisamente a Teoria da

Cooperação de Münster, trouxe novas perspectivas, o que se torna altamente relevante tendo em vista o aumento da complexidade das organizações e dos mercados (PINHO, 1982, 2004; BIALOSKORSKI NETO, 2008; SOUSA; BUZIN; SILVA, 2008).

Neste contexto, entende-se cooperativa conforme o conceito de Frantz (1985, p. 56), “[...] estratégia de competição com o objetivo de maximizar os resultados da ação econômica individual de cada produtor [...]”.

O número de cooperados e de cooperativas vem aumentando paulatinamente. Em 2007 existiam no Brasil 1.412 cooperativas singulares. O SICOOB é hoje o maior sistema de crédito cooperativo do Brasil, congrega 654 cooperativas singulares. As cooperativas de crédito rurais correspondem a 31% do total de cooperativas (SOARES; SOBRINHO, 2007).

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 rural 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 integrativa através das centrais e necessitam, neste sentido, estar altamente em sintonia com o que há de mais eficiente em ferramentas de gestão.

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.

A combinação de técnicas computacionais com metodologias de análise de crédito tem sido realizada nos últimos anos com resultados satisfatórios. Dentre estas técnicas, destaca-se a Mineração de Dados (*data mining*) (GOLDSCHMIDT; PASSOS, 2005; KOH *et al*, 2006; LEMOS *et al*, 2005; LAI *et al*, 2006). Tal

ferramenta pode auxiliar na análise e interpretação dos dados históricos da cooperativa com o propósito de evitar riscos e ajudar no processo de tomada de decisão.

Dada a importância do processo de concessão de crédito em cooperativas de crédito, a alta competitividade do mercado financeiro e a atual tendência de profissionalização da gestão no setor, esta pesquisa procurou verificar como o uso de Mineração de Dados pode auxiliar na gestão de crédito.

Para a operacionalização de tal modelo, o estudo se apoiou nas atividades e estágios propostos pela literatura, que incluiu a coleta, seleção e preparação da base de dados; pré-processamento e limpeza dos dados; processamento, composto pela classificação e análise dos dados; elaboração de modelos de gestão de crédito baseado em Mineração de Dados para cooperativas de crédito; análise comparativa de ferramentas alternativas de mineração de dados na elaboração dos modelos e comparação do modelo desenvolvido com o modelo utilizado atualmente pela cooperativa.

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.

O trabalho está estruturado em quatro partes além desta. A primeira parte analisa as vertentes do cooperativismo, as principais correntes que fundamentam este tipo de organização. Posteriormente, faz uma breve apresentação da análise de crédito; do processo de *Knowledge Discovery in Databases – KDD*, descoberta de conhecimento e *Data Mining – DM*, mineração de dados. Por último, são apresentadas redes neurais e árvores de decisão, técnicas de mineração de dados que serão utilizadas na construção do modelo.

A segunda parte compõe-se das etapas realizadas na construção do modelo proposto. A terceira parte apresenta e discute o modelo em comparação com o atualmente utilizado pela cooperativa.

A última etapa corresponde às conclusões e propostas para estudos posteriores.

2. REFERENCIAL TEÓRICO

2.1 Vertentes do Cooperativismo

O cooperativismo pode ser classificado em duas vertentes, a doutrinária *rochdaleana*¹, 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, p. 299).

Apesar das iniciativas e empreendimentos² anteriores, a experiência dos tecelões de Rochdale é o marco histórico e institucional que, por mais de um século, representou as bases de discussão e de disseminação do cooperativismo. Essa “vertente” é marcada por uma visão doutrinária. Em oposição, ocasionado pelo desenvolvimento das ciências sociais, surgiram alguns estudos teóricos marcados por uma visão empresarial da cooperativa.

A Aliança Cooperativa Internacional³ reuniu e atualizou os princípios cooperativistas de Rochdale e que atualmente compõem o instrumental ideológico da doutrina cooperativista (PINHO, 1982). Essa característica é tão marcante que autores como Irion (1997) chega a considerar sinônimas as expressões cooperativismo e doutrina cooperativista.

Esse modelo *rochdaleano* inspirou a doutrina cooperativista e fez surgir novos enfoques contrapondo à este idealismo “romântico”, de mudança do homem. Estes novos olhares entendem que a cooperativa é também uma empresa e está inserida em um ambiente competitivo de negócios.

A vertente teórica procura “atender aos interesses individuais dos associados por meio da prática da solidariedade cooperativista”, afasta assim da “necessidade de transformação social e de mudanças da própria pessoa humana”. Ela tem como

¹ Vertente com origem na experiência dos tecelões do bairro de Rochdale, na cidade de Manchester, condado de Lancashire, Inglaterra.

² Colônias agrícolas; Falanstérios, associações mútuas, dentre outros.

³ Os princípios revistos e atualizados pela ACI em seus diversos congressos estão assim estabelecidos em Goiás (2002, p. 16): “Adesão livre (social, política e racial; distribuição de sobras e do público em geral, a) ao desenvolvimento da cooperativa, b) aos serviços comuns, c) aos associados “pro rata” das operações; taxa limitada de juros ao capital; constituição de um fundo para educação dos cooperados e do público em geral; ativa de um fundo entre as cooperativas, em plano local, nacional e internacional”.

precursores os trabalhos dos alemães Franz Oppenheimer e Robert Liefman (PINHO, 2004, p. 292).

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. Um grupo de professores desta universidade conjuntamente com pesquisadores latino americanos, opondo aos pressupostos doutrinários rochdaleanos, desenvolvem esta “Escola”, cuja fundamentação metodológica advém do racionalismo crítico (PINHO, 2004).

Os pressupostos fundamentais desta teoria consistem nos seguintes pontos (PINHO, 2004):

- inclusão do interesse pessoal e concorrência;
- avaliação dos associados na busca pelos interesses pessoais nas duas estratégias, atuação solidária ou individual;
- autonomia da cooperativa em relação às unidades econômicas dos sócios;
- busca pelos dirigentes em atender aos próprios interesses na medida em que promovem os interesses dos associados, o que ocasiona a necessidade de estabelecer controles contra ações negativas dos membros;
- solidariedade (consciente) entre associados e cooperativa é legitimada por normas contratuais ou estatutárias.

Percebe-se que há uma grande complexidade na gestão de uma cooperativa, que necessita atender às necessidades e anseios dos cooperados e ao mesmo tempo ser competitiva. Possui então, características de associação, na prestação dos serviços e, de empresa, na gestão racional dos recursos produtivos. Um equilíbrio entre estas duas características deve ser perseguido. A Teoria de Münster desenvolveu dois instrumentos para controlar essas duas facetas, a produtividade, em uma dimensão externa, para controlar o desempenho com o mercado, e a efetividade, em uma dimensão interna, para garantir que as vantagens retornem aos cooperados.

Frantz (1985, p. 57-58), afirma que, em longo prazo, a cooperativa consegue ser eficiente somente com o bom desempenho nos dois aspectos que a caracteriza,

na dimensão associativa e na dimensão empresarial, a primeira resulta na “[...] necessidade da participação política de seus associados na condução do empreendimento [...]” e a segunda da “[...] necessidade da participação econômica dos associados na cooperativa [...]”.

Da análise dos pressupostos preconizados por esta teoria, deriva dois axiomas que também são utilizados pela teoria econômica neoclássica: racionalidade ilimitada dos cooperados e da informação completa (MENEGÁRIO, 2000).

Boettcher *apud* Pinho (1982, p. 75) expõe o seguinte conceito de cooperativa baseado nos axiomas e pressupostos desta teoria: “as cooperativas são agrupamentos de indivíduos que defendem seus interesses econômicos individuais por meio de uma empresa que eles mantêm conjuntamente”.

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. As próximas seções tratam da análise de crédito e da mineração de dados.

2.2 Análise de Crédito

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 do cliente tornar-se inadimplente, e aponta como principal vantagem o fato de que “[...] decisões sobre a concessão são tomadas com base em procedimentos pessoais e padronizados gerando um maior grau de confiabilidade”.

Com a concepção de cooperativa discutida na seção anterior, 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 et al. (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.

A combinação de técnicas computacionais na análise de crédito tem sido realizada nos últimos anos, com resultados satisfatórios e a próxima seção irá examinar a Mineração de Dados.

2.3 Descoberta de Conhecimento e Mineração de Dados

A terminologia “Descoberta de Conhecimento em Base de Dados” (*Knowledge Discovery 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 *et al.*, 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 modelos⁴.

Goldschmidt e Passos, (2005, p. 6) dividem as atividades de *KDD* em três grupos:

1. Desenvolvimento tecnológico – este grupo compreende “[...] as iniciativas de concepção, aprimoramento e desenvolvimento de algoritmos, ferramentas e tecnologias de apoio [...]” no processo de *KDD*.

⁴ Fayyad et al. (1996) denomina padrões como componentes dos modelos. Neste estudo será 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”.

2. Execução de KDD – este grupo inclui atividades relacionadas à utilização dos algoritmos, ferramentas e tecnologias desenvolvidas na procura de conhecimento.
3. 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”.

Para a consecução dos objetivos, este estudo se apoiará nas atividades de execução de KDD.

A Mineração de Dados tem obtido resultados positivos em diversas áreas como, por exemplo: vendas, finanças, seguros, planos de saúde, transporte, medicina, telecomunicações, mercado financeiro (LEMOS *et al.* 2005); gestão do conhecimento (CARDOSO; MACHADO, 2008); aplicações na medicina (FERRO; LEE, 2001); marketing, investimentos, detecção de fraudes, manufatura, internet, dentre outros (FAYYAD *et al.*,1996). Ferreira (2005) analisou o uso de Mineração de Dados na retenção de clientes em telefonia celular. Fayyad *et al.* (1996) foi um dos pioneiros com a análise de 3 terabytes (10^{12} bytes) de imagens do Skycat, sistema utilizado por astrônomos para analisar, classificar e catalogar imagens de objetos do espaço.

Estudos de verificação de análise de crédito foram realizados por vários autores, destacam-se Koh *et al.* (2006), Lemos *et al.* (2005) e Lai *et al.* (2006). Os primeiros autores investigaram os passos necessários para a construção de modelos de *Credit Scoring* com técnicas de Mineração de Dados, como unidade empírica, eles utilizaram um banco alemão para ilustrar o uso de tal metodologia. Lemos *et al.* (2005) verificaram a aplicação da análise de crédito bancário com a mesma metodologia e utilizaram como objeto uma agência do Banco do Brasil. Lai *et al.* (2006) utilizaram redes neurais para projetar sistemas de risco de crédito.

Piatetsky-Shapiro *et al.* (1994) apontam a crescente disponibilidade comercial em ferramentas para aplicações genéricas e específicas como fator de desenvolvimento de tarefas de *KDD*. Por outro lado, mencionam como dificuldades a falta de rigor estatístico, a abundância de padrões e a fraca integração dos sistemas.

Há muitas técnicas disponíveis e diferentes tipos de métodos específicos para aplicações particulares. Entre estas técnicas, duas são comuns e destacadas pelos autores que abordaram a análise de crédito: as Árvores de decisão e as Redes Neurais.

2.3.1 Redes Neurais

As Redes Neurais Artificiais RNA tem 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) RNA's 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; Carvalho e Ludermir (2000, p. 1) conceituam RNA's 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.

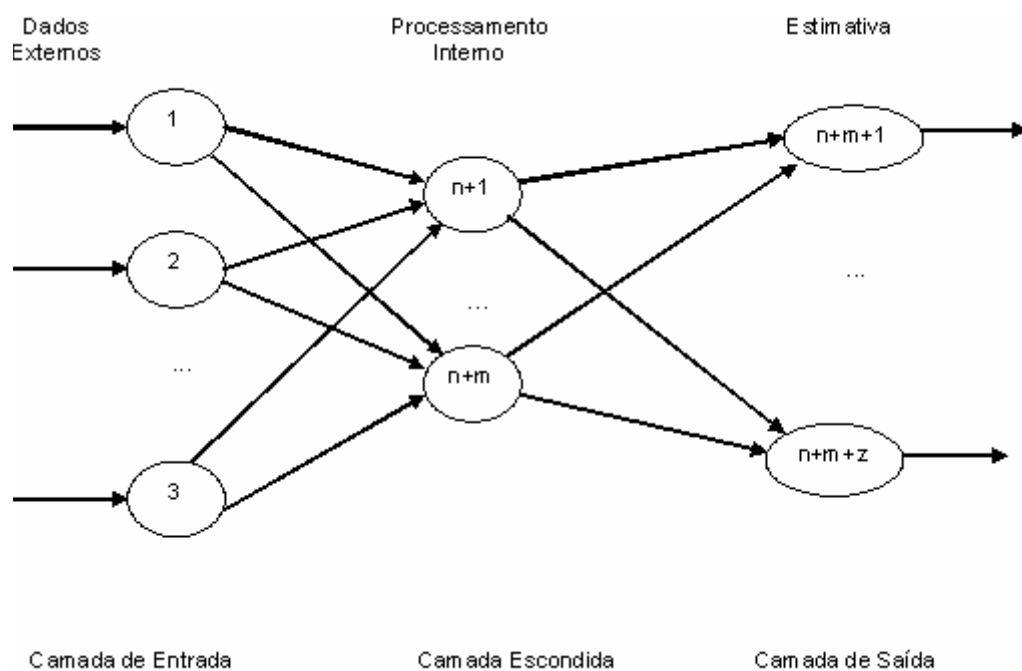


Figura 01 – Arquitetura de uma RNA.
Fonte: Goldschmidt e Passos, 2005.

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 layers*, 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; Carvalho e Ludermir (2000) 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, Braga; Carvalho e Ludermir (2000, p. 227) 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 incapacidade do modelo em esclarecer de que maneira os resultados são gerados. Devido a esta especificidade, as RNA's são também denominadas de “caixas pretas”.

2.3.2 Árvores de Decisão

Árvores de decisão constituem um dos principais métodos de Mineração de Dados. 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 seqüência 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 E 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 e o CART (*Classification and Regression Trees*⁵). Optou-se, nesta pesquisa, por utilizar o algoritmo C4.5 que segundo Goldschmidt e Passos (2005) é amplamente utilizado e aceito. Para este estudo utilizou-se uma versão mais atual do algoritmo, denominada C5.0.

⁵ Árvores de classificação e regressão.

3. METODOLOGIA

3.1. Descrição do Caso

Para a realização da pesquisa optou-se pelo estudo de caso, que, segundo Yin (2003) 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 a Cooperativa de Crédito Rural de Rubiataba – SICCOOB DO VALE, 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, há a impossibilidade de 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 SICCOOB 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 *et al.* (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.

Dentre as técnicas de Mineração de Dados disponíveis, foram utilizadas duas, Redes Neurais e Árvores de Decisão, ambas utilizadas nos trabalhos de Koh *et al.*, (2006) e LEMOS *et al.* (2005) na análise de crédito.

3.2. Coleta e Seleção dos Dados

Essa etapa corresponde 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 Johson (2003), apontam os seguintes elementos auxiliares na análise de variáveis: a experiência anterior, o conhecimento, a quantidade e a qualidade dos dados.

Os dados foram extraídos de duas fontes na cooperativa: da análise de crédito dos cooperados tomadores de crédito e do cadastro.

Para a pesquisa foram utilizados dados históricos de 211 cooperados pessoas físicas. Os dados representam o universo dos cooperados tomadores de crédito. De cada cooperado foram obtidas as 39 informações especificadas no quadro 01.

A variável de saída é representada pelas variáveis 27 a 39, correspondentes ao período de julho de 2007 a junho de 2008.

Quadro 01 - Estrutura da base de dados.

| CÓDIGO | VARIÁVEIS | VALOR |
|--------|---|--|
| 01 | Código do cooperado | Chave numérica, única para cada cooperado. |
| 02 | Gênero | 1= masculino 2= feminino |
| 03 | Idade | Valor numérico |
| 04 | Nível educacional | 1= pós-graduação 2= superior completo 3= superior incompleto 4= segundo grau completo 5= segundo grau incompleto 6= primeiro grau completo 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 2= zona rural |
| 08 | Principal atividade | Nome da atividade |
| 09 | Estado civil | 1= casado 2= solteiro 3= viúvo 4= separado judicialmente 5= outros |
| 10 | Capital | Valor numérico |
| 11 | Relacionamento | 1= opera com a cooperativa há mais de 3 anos 2= opera com a cooperativa de 1 a 3 anos 3= opera com a cooperativa até 1 ano |
| 12 | Comportamento das operações | 1= normal 2= atrasos esporádicos 3= atrasos/re negociações constantes |
| 13 | Tempo de experiência na atividade/emprego | 1= mais de 5 anos 2= de 2 a 5 anos 3= até 3 anos |

Fonte: Dados da pesquisa, 2008.

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

| CÓDIGO | VARIÁVEIS | VALOR |
|--------|--|--|
| 14 | Consultas cadastrais | 1= ausência de restrições 2= com restrições irrelevantes justificadas 3= com restrições relevantes ou irrelevantes sem justificativas |
| 15 | Informações cadastrais na cooperativa | 1= cadastro atualizado e confiável 2= cadastro atualizado e não confiável 3= informações desatualizadas ou ausência de informações |
| 16 | Finalidade da operação | 1= custeio e investimento 2= financiamento de bens 3= crédito pessoal/cheque especial 4= renovações/composições de dívidas ⁶ |
| 17 | Garantia das operações | 1= hipoteca – capital social 2= alienação fiduciária/warrants 3= penhor cedular/caução de títulos 4= pessoal |
| 18 | Liquidez das garantias | 1= garantia de fácil liquidez (venda até 6 meses) 2= garantia de média liquidez (venda de 6 a 12 meses) 3= garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) |
| 19 | Que freqüência o cooperado opera (operações ativas) | 1= nunca 2= frequentemente 3= permanentemente |
| 20 | Valor da operação | 1= até 1% do PLA ⁷ 2= de 1,01% a 2% do PLA 3= de 2,01% a 3% do PLA 4= mais de 3% do PLA |
| 21 | Nível de comprometimento – prestações com rend. Liq. ⁸ | 1= até 20% da renda média líquida 2= de 20% a 30% da renda média líquida 3= mais de 30% da renda líquida |
| 22 | Patrimônio líquido pessoal livre relac. Endivid. tot. ⁹ | 1= mais de 4 vezes 2= de 2 a 4 vezes 3= sem patrimônio pessoal ou até 2 vezes |
| 23 | Endividamento total em relação a renda líquida ano ¹⁰ | 1= até 2 vezes 2= de 2 a 4 vezes 3= mais de 4 vezes |
| 24 | Total endividamento relação ao capital integralizado ¹¹ | 1= até 4 vezes 2= de 4 a 8 vezes 3= de 8 a 12 vezes 4= acima de 12 vezes |
| 25 | Perfil da atividade econômica do associado | 1= ótima 2= boa 3= regular 4= ruim |

Fonte: Dados da pesquisa, 2008.

⁶ Alongamento do prazo da dívida.

⁷ Patrimônio Líquido Ajustado compreende: PLA=Capital Social+Reservas+Receitas-Despesas. Também conhecido como Patrimônio de Referência (PR).

⁸ Esta variável refere-se ao valor percentual da prestação da dívida em relação à renda líquida mensal do cooperado.

⁹ Esta variável refere-se ao valor do patrimônio líquido do cooperado que não está comprometido com outra dívida, livre de ônus.

¹⁰ Esta variável refere-se ao valor do endividamento contraído pelo cooperado em relação à sua renda líquida ano.

¹¹ Esta variável refere-se ao valor do endividamento contraído pelo cooperado em relação ao seu capital integralizado na cooperativa.

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

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

Fonte: Dados da pesquisa, 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 95 variáveis de saída adotadas pela pesquisa. 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. Koh *et al*, (2006) utilizaram 20 variáveis e Lemos *et al*, (2005) utilizaram 24.

3.3 Pré-processamento e Limpeza dos Dados

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 (junho 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.

As variáveis do conjunto de dados coletados na cooperativa podem ser classificadas conforme o quadro 02.

Quadro 02 – Tipos de variáveis da base de dados.

| Código | Variáveis | Tipo de variável | |
|---------------|---|-------------------------|----------|
| 01 | Código do cooperado | Nominal | Discreta |
| 02 | Sexo | Nominal | Discreta |
| 03 | Idade | Escalar | Contínua |
| 04 | Nível educacional | Nominal | Discreta |
| 05 | Cidade | Nominal | Discreta |
| 06 | Naturalidade | Nominal | Discreta |
| 07 | Local de domicílio | Nominal | Discreta |
| 08 | Principal atividade | Nominal | Discreta |
| 09 | Estado civil | Nominal | Discreta |
| 10 | Capital | Escalar | Contínua |
| 11 | Relacionamento | Nominal | Discreta |
| 12 | Comportamento das operações | Nominal | Discreta |
| 13 | Tempo de experiência na atividade/emprego | Nominal | Discreta |
| 14 | Consultas cadastrais | Nominal | Discreta |
| 15 | Informações cadastrais na cooperativa | Nominal | Discreta |
| 16 | Finalidade da operação | Nominal | Discreta |
| 17 | Garantia das operações | Nominal | Discreta |
| 18 | Liquidez das garantias | Nominal | Discreta |

Fonte: Dados da pesquisa, 2008.

Quadro 02 – Tipos de variáveis da base de dados. (Continuação)

| | | | |
|----|---|---------|----------|
| 19 | Que freqüência o cooperado opera (operações ativas) | Nominal | Discreta |
| 20 | Valor da operação | Nominal | Discreta |
| 21 | Nível de comprometimento – prestações com rend. Liq. | Nominal | Discreta |
| 22 | Patrimônio líquido pessoal livre relac. Endivid. tot. | Nominal | Discreta |
| 23 | Endividamento total em relação a renda líquida ano | Nominal | Discreta |
| 24 | Total endividamento relação ao capital integralizado | Nominal | Discreta |
| 25 | Perfil da atividade econômica do associado | Nominal | Discreta |
| 26 | Risco atribuído pela cooperativa | Nominal | Discreta |
| 27 | Resultado-Julho 2007 | Nominal | Discreta |
| 28 | Resultado-Agosto 2007 | Nominal | Discreta |
| 29 | Resultado-Setembro 2007 | Nominal | Discreta |
| 30 | Resultado-Outubro 2007 | Nominal | Discreta |
| 31 | Resultado-Novembro 2007 | Nominal | Discreta |
| 32 | Resultado-Dezembro 2007 | Nominal | Discreta |
| 33 | Resultado-Janeiro 2008 | Nominal | Discreta |
| 34 | Resultado-Fevereiro 2008 | Nominal | Discreta |
| 35 | Resultado-Março 2008 | Nominal | Discreta |
| 36 | Resultado-Abril 2008 | Nominal | Discreta |
| 37 | Resultado-Maio 2008 | Nominal | Discreta |
| 38 | Resultado-Junho 2008 | Nominal | Discreta |
| 39 | Resultado agregado | Nominal | Discreta |

Fonte: Dados da pesquisa, 2008.

Para a avaliação dos valores discrepantes ou *outliers* das variáveis do tipo escalar e contínua, foram utilizados os métodos sugeridos por Triola (2005) para encontrar *outliers* extremos e suaves.

O cálculo para encontrar *outliers* extremos estão apresentados nas fórmulas 1 e 2, onde são valores de x tais que:

$$X < Q_1 - 3 //Q \quad (1)$$

$$X > Q_3 + 3 //Q \quad (2)$$

Onde:

//Q representa o intervalo interquartil e é definido por:

$$//Q = Q_3 - Q_1$$

Q₃: representa o terceiro quartil;

Q₁: representa o primeiro quartil.

Para a variável idade, optou-se pela utilização do cálculo para encontrar *outliers* suaves, indicados pelas fórmulas 3 e 4, onde são valores de x , tais que:

$$Q_1 - 3 // Q \leq x < Q_1 - 1,5 // Q \quad (3)$$

ou

$$Q_3 + 1,5 // Q < x \leq Q_3 + 3 // Q \quad (4)$$

Para avaliar as variáveis por ordem de importância, utilizou-se a função *Pearson* para classificar as variáveis.

A medida usada para a ordem de importância depende do tipo de variável. No caso deste estudo, onde há uma variação entre tipos diferentes de variáveis, ou seja, a variável de saída é categórica e há variáveis de entrada categóricas, é utilizada a função *Pearson*. Lapponi (2005, p. 176) apresenta a fórmula seguinte:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad (5)$$

Considerou-se como importante as variáveis que obtiveram valores acima de 0,9 na função *Pearson*.

O modelo utilizou todas as variáveis que alcançaram algum valor pela função *Pearson*, as variáveis que não atingiram nenhum valor não foram utilizadas no modelo.

Foram utilizados também gráficos de rede para avaliar a força das relações entre as variáveis importantes. Este tipo de gráfico apresenta linhas indicando a força do relacionamento entre variáveis, quanto mais freqüente a relação, mais espessa é a linha.

Por último foi verificada a existência de valores ausentes (*missing*), na base de dados. Pela predominância de atributos categóricos presentes na base de dados da cooperativa, foi utilizada a medida estatística “moda”, conforme sugerida por Goldschmidt e Passos (2005).

3.4 Transformação dos Dados

Esta etapa visa auxiliar a execução das técnicas de mineração de dados. Os dados foram agrupados segundo indicação de Goldschmidt e 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. Desta coleta foi criada uma tabela em duas dimensões. Os atributos selecionados encontram-se nas colunas e os registros individuais nas linhas.

O modelo gerado pela técnica Árvore de Decisão, foi transformado em regras de decisão na fase pós-processamento.

3.5 Modelagem e Mineração de Dados

Para a modelagem e mineração do problema, a base de dados foi dividida em três partições: treinamento, teste e validação dos modelos.

Utilizou-se as seguintes ferramentas no processo de construção do modelo:

- Nódulo de importação de planilha eletrônica.
- Avaliação do tipo de variável, valores, campos, valores ausentes e classificação das variáveis de entrada, saída, entrada e saída, partição ou ainda cancelamento de variáveis.
- Classificação por ordem de importância – pertence à categoria modelagem, no caso do problema deste estudo foi utilizada a função Pearson.
- Gráfico de Rede – este gráfico avalia a força da relação entre as variáveis utilizando linhas para indicar o relacionamento, quanto mais grossa a linha, mais forte é a relação.
- Partição – são utilizados no processo de construção do modelo para dividir a base de dados nos estágios de treinamento, teste e validação.
- Algoritmo C5.0 – este nódulo permite a configuração da construção da árvore, podendo gerar duas classes de modelos, árvores ou regras de decisão. É também verificada a acurácia de cada folha da árvore ou regra, caso não contribuam de forma significativa são removidas ou “podadas”.
- Algoritmo Rede Neural – usado para configurar a criação e treinamento da Rede Neural Artificial. Verifica também a importância relativa das variáveis, a acurácia do modelo bem como a quantidade de neurônios utilizados nas três camadas, de entrada, escondida e saída.

- Análise – este nódulo é altamente relevante na avaliação dos modelos. Proporciona a investigação dos resultados dos modelos gerados. Esta ferramenta mostra o percentual dos valores corretos e erros. Mostra também uma matriz indicando a combinação de valores preditos e valores verdadeiros.

3.6 Avaliação dos Modelos

Nesta seção é apresentada como foi realizada a avaliação dos resultados encontrados com a implementação computacional das metodologias Árvores de Decisão e Redes Neurais e do problema descrito anteriormente, bem como a comparação entre elas e com o modelo utilizado pela cooperativa.

Para a implementação computacional das técnicas Árvore de Decisão e Redes Neurais, foi utilizada a base de dados descritos neste estudo, composta por 211 cooperados pessoa física, considerando-se para cada cooperado as variáveis já descritas anteriormente.

Goldschmidt e Passos, (2005, p. 50) argumenta 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 e JOHNSON, 2003).

Neste estudo, para melhor isenção da avaliação, os dados foram divididos em três partições no processo de modelagem: 33% para treinamento; 33% para teste e 34% para validação.

Para a avaliação da precisão das regras geradas pela Árvore de Decisão foi empregada a acurácia como medida. Goldschmidt e Passos (2005, p. 56) entendem acurácia ou precisão da regra com a seguinte definição: “[...] consideremos um modelo de conhecimento composto por regras da forma: $X \rightarrow Y$, onde X e Y são predicados (condições que podem se tornar verdadeiras ou falsas em função de cada registro da base de dados)” e acrescentam que “é o percentual de registros da base

de dados que ao satisfazerem ao antecedente da regra, satisfazem também ao conseqüente”.

$$Acc = \frac{|X \wedge Y|}{X} \quad (6)$$

Ferreira (2005) explica que, SE X ENTÃO Y onde X é a quantidade de observações que cumprem a condição SE e Y é a quantidade de observações que cumprem a afirmação ENTÃO.

Para este estudo decidiu-se por utilizar RNA de múltiplas camadas. A quantidade de neurônios da camada de entrada foi de 66, da camada intermediária 3 e a quantidade de neurônios na camada de saída foi igual a 3.

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 [...]”.

A importância relativa das variáveis de entrada da RNA contém os resultados da análise de sensibilidade. As variáveis são listadas em ordem de importância, do mais importante para o menos importante. O valor varia entre 0 (campo que não tem efeito na previsão) e 1 (campo que determina completamente a previsão).

Para a avaliação do modelo empregado pela cooperativa, foi verificada qual o percentual de acerto que o modelo classifica corretamente os cooperados tomadores de crédito. Foi verificada também se a variável “risco atribuído” é considerada importante para prever o comportamento dos cooperados na tomada de crédito. A variável “risco atribuído” corresponde ao resultado final do modelo utilizado. Nela contém a ponderação das variáveis 11 a 26 expostas no quadro 01. A organização não permite a divulgação dos pesos e dos pontos atribuídos a cada variável.

Foram empregadas matrizes de coincidência 3 por 3 dos valores classificados em preditos e reais para avaliar o tipo de acerto e erro dos modelos.

Para a avaliação comparativa dos modelos foi utilizado o percentual total de valores preditos incorretamente.

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Baseado nas etapas apresentada no tópico 3.1, este capítulo explora as variáveis e a estrutura da base de dados estudada. Busca-se, portanto, apoiado na compreensão do ambiente de negócios da cooperativa e na problemática proposta, analisar os dados e informações para o desenvolvimento e avaliação dos modelos.

Esta parte do estudo está dividida em duas etapas. A primeira, denominada pré-processamento e limpeza dos dados, ocorre antes da aplicação dos algoritmos e tem por finalidade examinar a importância das variáveis para posterior seleção, bem como investigar valores discrepantes e a força da relação entre as variáveis. A segunda corresponde, ao processo de modelagem, aplicação dos algoritmos de Mineração de Dados e avaliação comparativa entre os modelos desenvolvidos e o empregado atualmente pela cooperativa.

4.1 Pré-processamento e Limpeza dos Dados

Para avaliar a existência de valores discrepantes, ou seja, valores que se encontram distantes dos demais na distribuição, os dados da variável discreta e contínua foram submetidos à análise gráfica *boxplot* e histograma de frequência, com o objetivo de visualizar tais valores.

Os parâmetros da variável discreta idade e da variável contínua capital estão apresentados na tabela 01.

Tabela 01 – Parâmetros da variável discreta e contínua.

| Variável | Média | Mínimo | Máximo | Desvio Padrão |
|----------|--------------|------------|----------------|---------------|
| Idade | 49 | 20 | 86 | 14 |
| Capital | R\$ 9.854,97 | R\$ 155,17 | R\$ 107.082,72 | R\$ 14.646,04 |

Fonte: Dados da pesquisa, 2008.

A figura 02 mostra o gráfico *boxplot* da variável capital. A visualização apresenta valores discrepantes suaves e extremos.

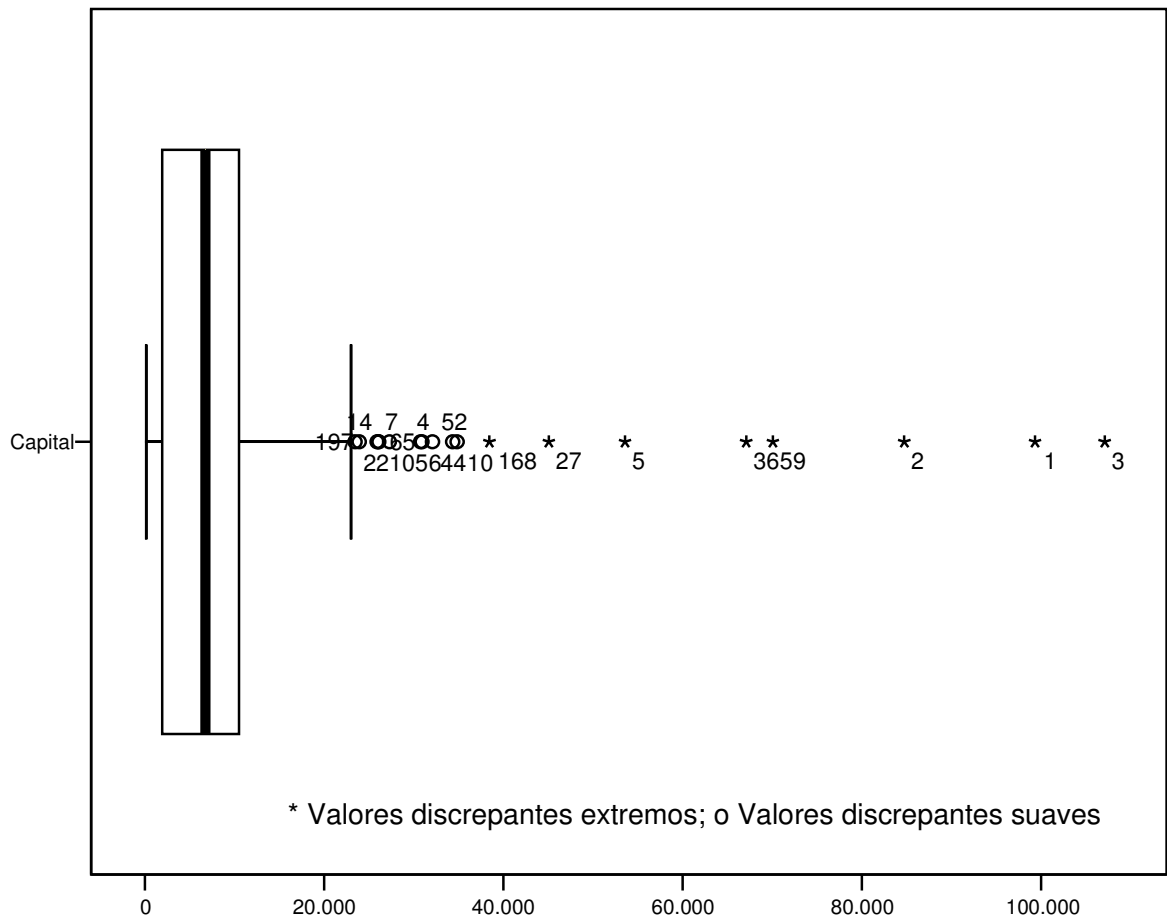


Figura 02 – Boxplot da variável capital.

Fonte: Dados da pesquisa, 2008.

Os valores representados por círculos, são categorizados como *outliers* (valores discrepantes) suaves, por sua vez, os valores representados por asteriscos são identificados como *outliers* extremos. Os números que aparecem ao lado dos círculos e estrelas representam as linhas onde se encontram na base de dados analisada. Por exemplo, o número três, mostrado como último ponto no gráfico, representa o valor R\$ 107.082,72, este registro encontra-se na linha 3 na base de dados da cooperativa, a estrela ao lado do número indica que é também um *outlier* extremo.

A figura 03 expõe a distribuição de frequência da variável capital, comprovando os *outliers* encontrados no gráfico *boxplot* da figura 02.

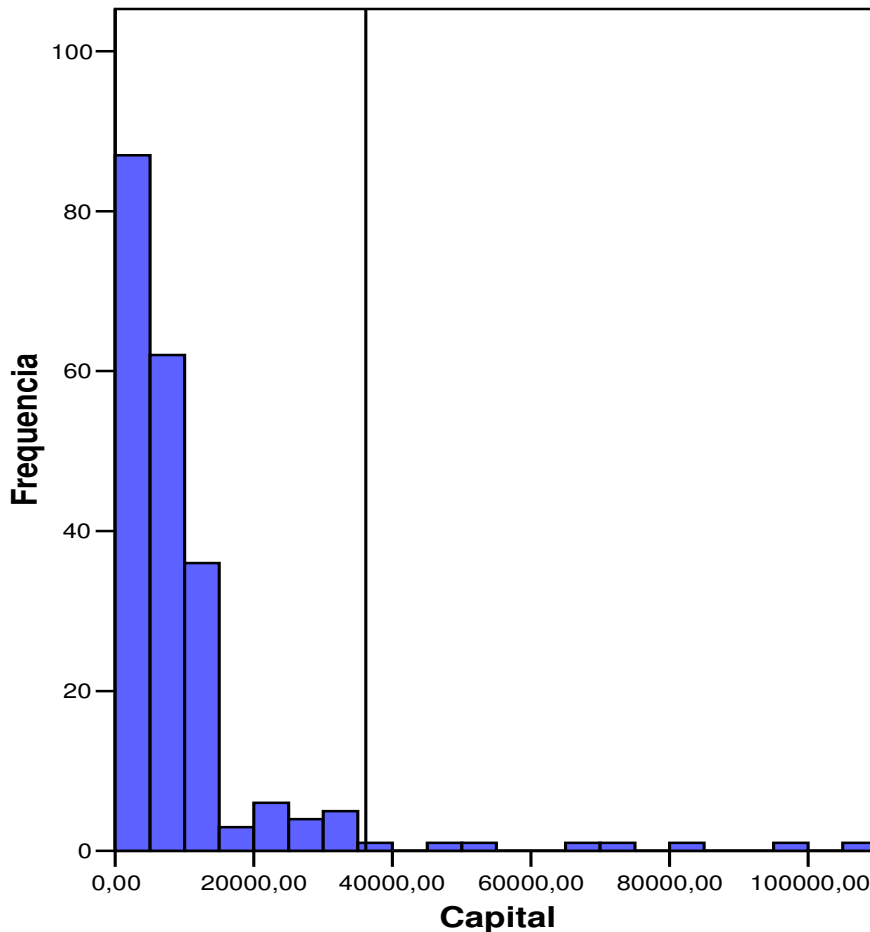


Figura 03 – Histograma de frequências da variável capital.

Fonte: Dados da pesquisa, 2008.

Com o cálculo aplicado para detectar valores discrepantes extremos encontraram-se os seguintes resultados: (inferior: -R\$22.737,63); (superior: R\$36.165,05). A linha que corta o valor 36.165,05 no eixo x evidencia o ponto que separa os valores considerados *outliers*.

Pela análise da distribuição, pelo boxplot e pelos cálculos de valores discrepantes, percebe-se que há alguns valores muito altos, entretanto decidiu-se por deixá-los, haja vista que há inadimplentes e a avaliação de tais cooperados é importante para a construção do modelo. A existência destes valores *outliers* pode ser em decorrência de vários fatores, como por exemplo, maior tempo de cooperação e/ou concentração das movimentações financeiras na cooperativa. Uma discussão maior deste aspecto foge aos objetivos estabelecidos neste trabalho.

A figura 04 apresenta o gráfico *boxplot* da variável idade. Conforme a visualização do gráfico, não foram encontrados valores discrepantes.

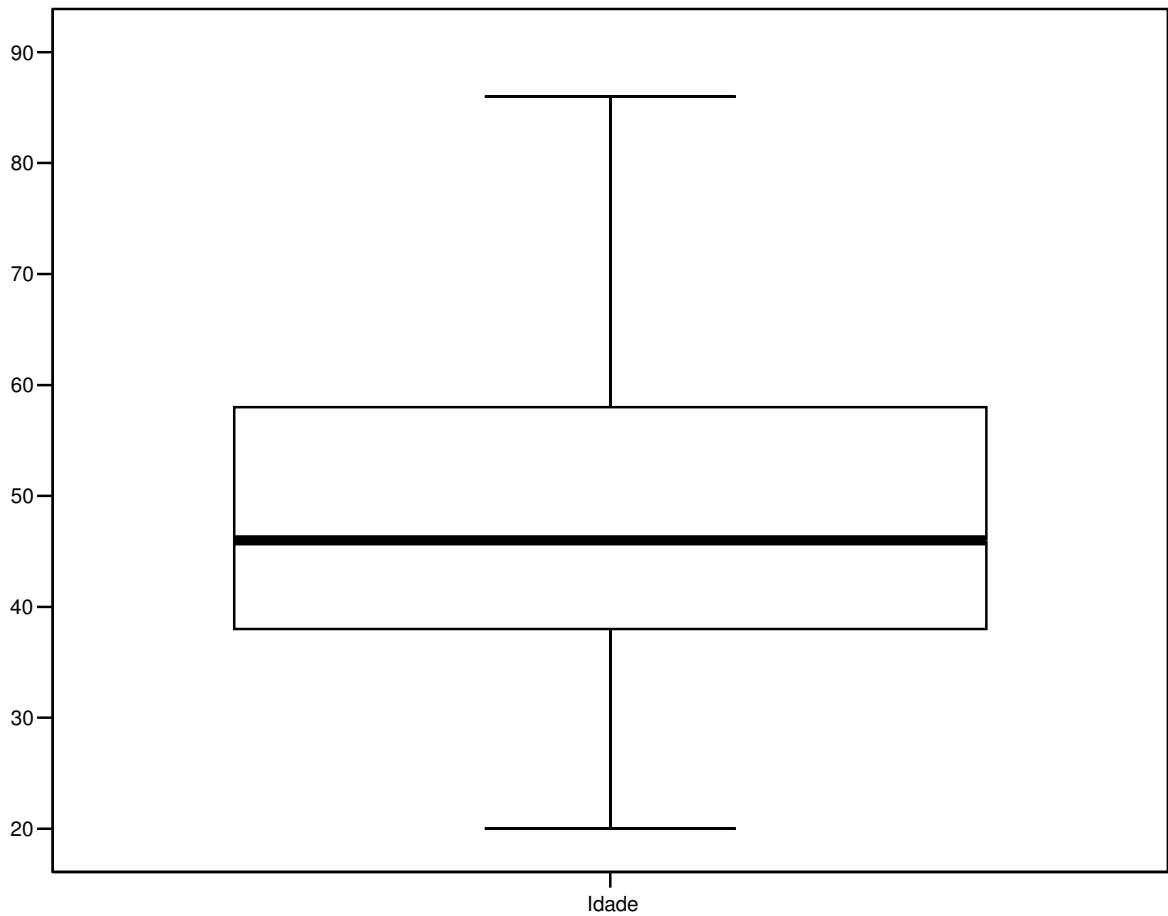


Figura 04 – Boxplot da variável idade.

Fonte: Dados da pesquisa, 2008.

Com o emprego do cálculo para detectar os valores discrepantes suaves, encontraram-se os seguintes resultados: (-22; 8; 88; 118), assim, nota-se que não há *outliers* nesta variável. Estes resultados são comprovados pelo histograma de freqüências na figura 05, onde nota-se que estão normalmente distribuídos.

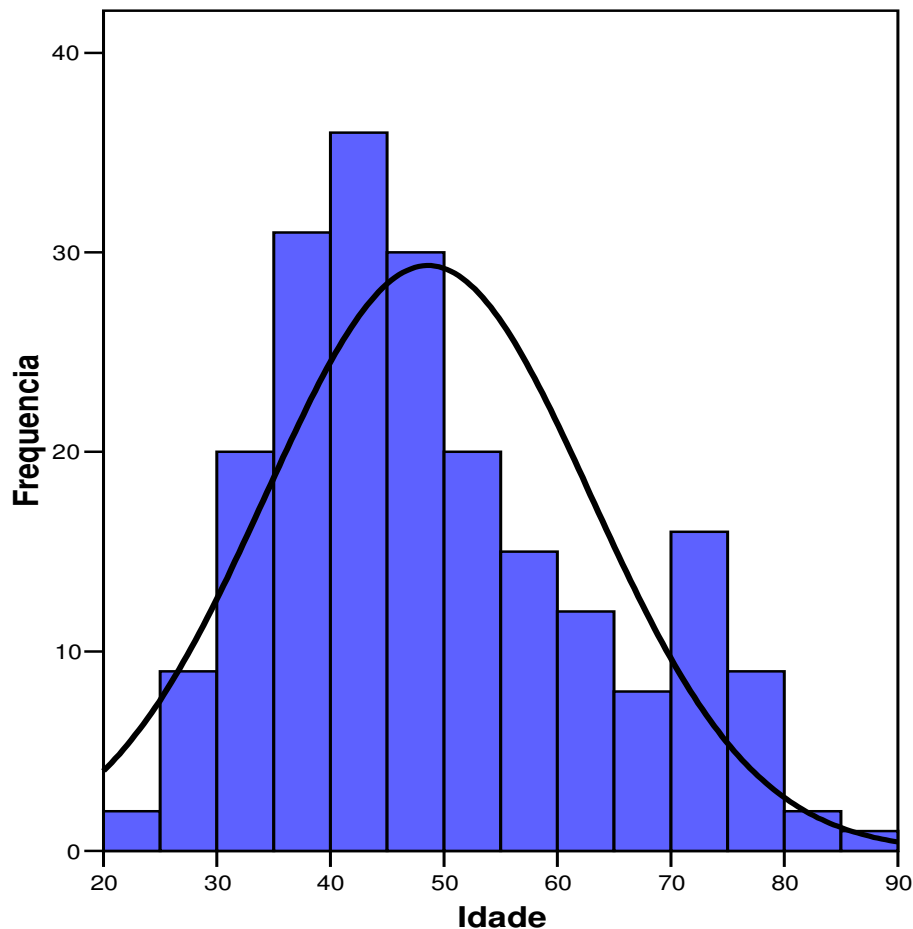


Figura 05 – Histograma de frequências da variável idade.

Fonte: Dados da pesquisa, 2008.

A seleção das variáveis é uma etapa importante no processo de *KDD*. Consiste na identificação de quais variáveis devem ser empregadas no modelo.

A tabela 04 apresenta as variáveis utilizadas neste estudo consideradas importantes pela aplicação da função de correlação de *Pearson*.

Tabela 02 – Seleção das variáveis da base de dados.

| Rank | Código | Variável | Importância | Valor |
|------|--------|--------------------------|-------------|-------|
| 1 | 18 | Liquidez das garantias | Sim | 1,000 |
| 2 | 17 | Garantia das operações | Sim | 1,000 |
| 3 | 26 | Risco atribuído | Sim | 1,000 |
| 4 | 16 | Finalidade da operação | Sim | 1,000 |
| 5 | 21 | Nível de comprometimento | Sim | 0,967 |
| 6 | 08 | Principal atividade | Sim | 0,912 |
| 7 | 10 | Capital | Não | 0,845 |
| 8 | 24 | Total endividamento | Não | 0,510 |
| 9 | 09 | Estado civil | Não | 0,387 |
| 10 | 02 | Gênero | Não | 0,243 |
| 11 | 19 | Operações ativas | Não | 0,191 |

Fonte: Dados da pesquisa, 2008.

Tabela 02 – Seleção das variáveis da base de dados. (Continuação)

| Rank | Código | Variável | Importância | Valor |
|------|--------|---|-------------|-------|
| 12 | 23 | Endividamento total em relação a renda l. ano | Não | 0,186 |
| 13 | 11 | Relacionamento | Não | 0,174 |
| 14 | 03 | Idade | Não | 0,166 |
| 15 | 07 | local de domicílio | Não | 0,127 |
| 16 | 04 | Nível educacional | Não | 0,104 |
| 17 | 06 | Naturalidade | Não | 0,000 |

Fonte: Dados da pesquisa, 2008.

Conforme descrito no tópico 3.3, todas as variáveis que não obtiveram algum resultado na função de *Pearson* foram descartadas, assim, foram utilizadas no processo de modelagem as variáveis de *rank* 1 a 16 da tabela 02.

A variável de código 08 não é utilizada no modelo de análise atual, mas foi julgada importante. As variáveis de código 10, 09, 02, 03 07 e 04 não foram consideradas importantes, porém, serão utilizadas para a modelagem do problema.

Os gráficos de rede a seguir mostram a força entre as variáveis de entrada importantes, mostradas na tabela 04 e a variável saída do modelo. As tabelas mostram a frequência absoluta e relativa da força da relação entre as variáveis e estão listadas no anexo 02.

Conforme já mencionado, este tipo de gráfico relaciona a quantidade de vezes que ocorreu tal relação. A força entre as variáveis é indicada através de linhas. A espessura da linha indica tal relação, no nosso estudo, mais espesso significa mais forte.

A figura 06 mostra a relação entre a variável resultado e a variável liquidez das garantias. A variável resultado, exibida no gráfico pela cor azul, são apresentadas nas três possibilidades apontadas neste estudo, adimplente, inadimplente ou adimplente com atrasos/renegociações constantes. A figura 06 indica uma forte relação entre a categoria adimplente e a categoria garantia pessoal ou de difícil liquidez. Essa garantia é representada por bens imóveis. Há uma relação unidirecional entre a categoria adimplente com atrasos/renegociações constantes e a categoria garantia de fácil liquidez.

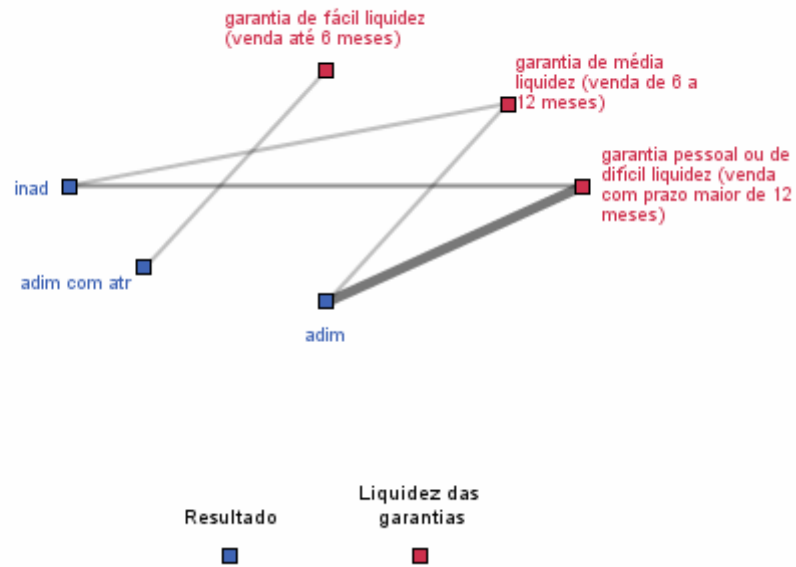


Figura 06- Gráfico de rede: variável de saída e liquidez das garantias.

Fonte: Dados da pesquisa, 2008.

A figura 07 mostra a relação entre a variável resultado e a variável garantia das operações.

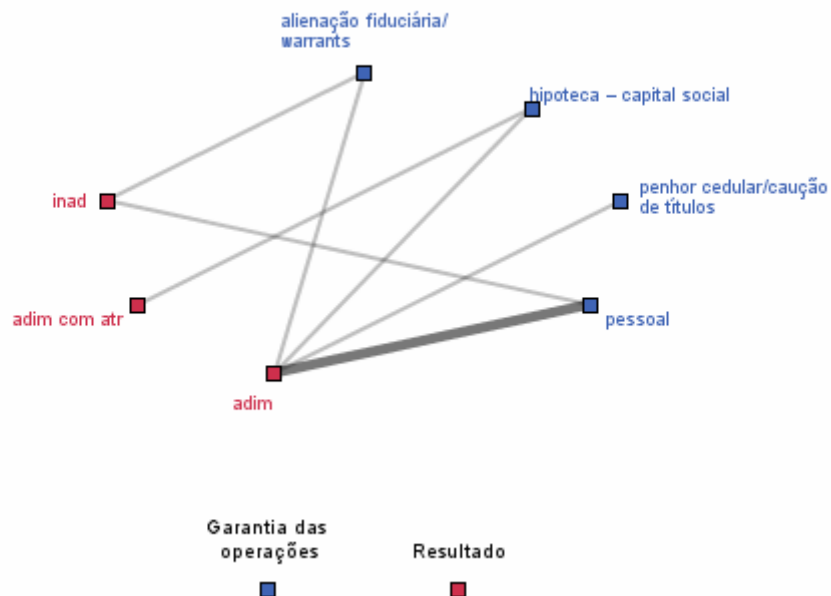


Figura 07- Gráfico de rede: variável de saída e garantia das operações.

Fonte: Dados da pesquisa, 2008.

Há uma forte relação entre a categoria adimplente e a categoria garantia pessoal. É importante ressaltar também que cooperados que apresentem hipoteca e penhor cedular/caução de títulos não são configurados como inadimplentes.

A figura 08 mostra a relação entre a variável risco atribuído e a variável resultado. O gráfico mostra que há uma forte relação entre a classificação “C” e a categoria adimplente, mostra também uma média relação entre a classificação “B” e a categoria adimplente. Os cooperados classificados como “A” são sempre adimplentes. Adimplentes com atrasos/renegociações constantes são categorizados como “D”. Este gráfico mostra evidências que o modelo da cooperativa consegue classificar os cooperados com certa eficiência.

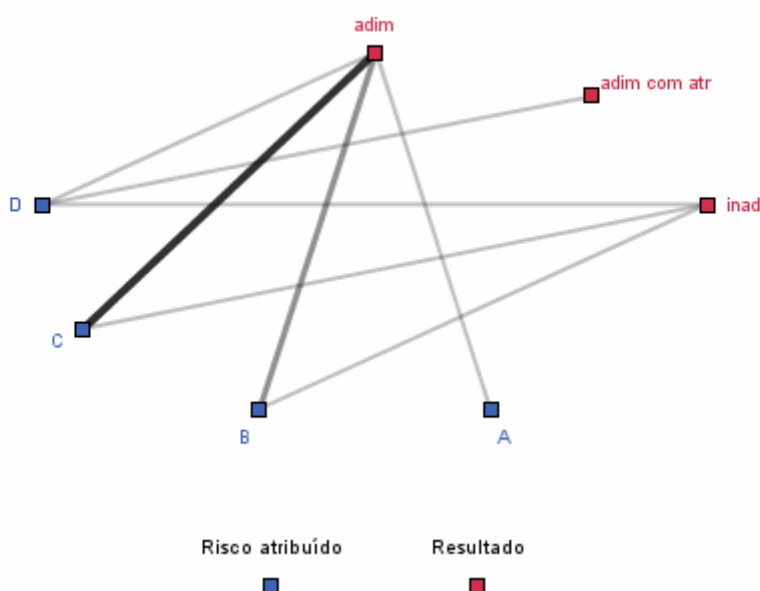


Figura 08- Gráfico de rede: variável de saída e risco atribuído.

Fonte: Dados da pesquisa, 2008.

A figura 09 apresenta a relação entre a variável resultado e a variável finalidade da operação. Percebe-se uma forte relação entre a categoria adimplente e duas modalidades de crédito: “crédito pessoal e cheque especial” e uma relação ainda mais forte com a categoria “custeio e investimento”, utilizado para investimento em melhorias e manutenção no processo de produção dos cooperados. Há uma força média entre a categoria adimplente e a categoria financiamento de bens. A categoria adimplente com atrasos/renegociações constantes, somente possui relação com a categoria renovação/composições de dívidas.

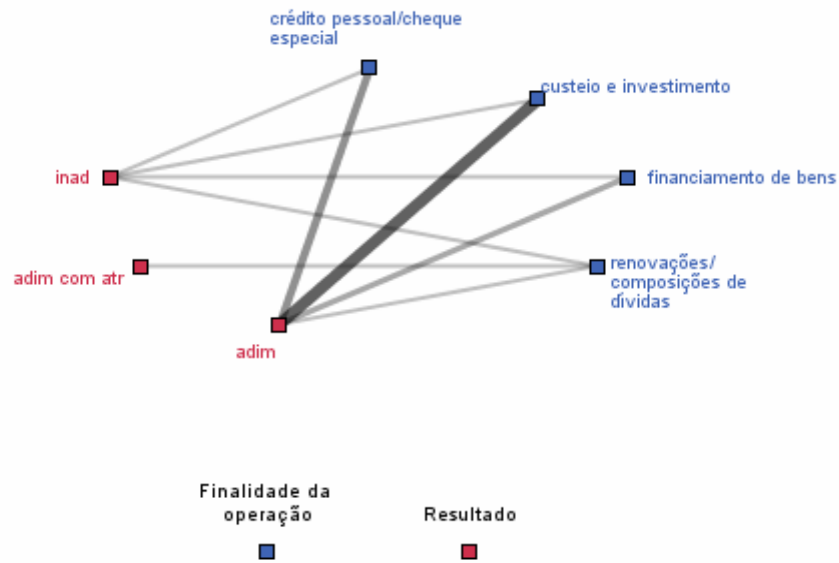


Figura 09- Gráfico de rede: variável de saída e finalidade da operação.

Fonte: Dados da pesquisa, 2008.

A figura 10 mostra o relacionamento entre a variável resultado e a variável nível de comprometimento com a renda média líquida do cooperado. Há uma forte relação entre a categoria adimplente e a categoria nível de comprometimento “de 20% a 30% da renda média líquida”. Portanto, existe um forte indício que o cooperado adimplente compromete mais de 20% e menos de 30% da renda, ou seja, não é muito agressivo e tampouco muito conservador. Há uma força média entre a categoria adimplente e a categoria nível de comprometimento “até 20% da renda média líquida”.

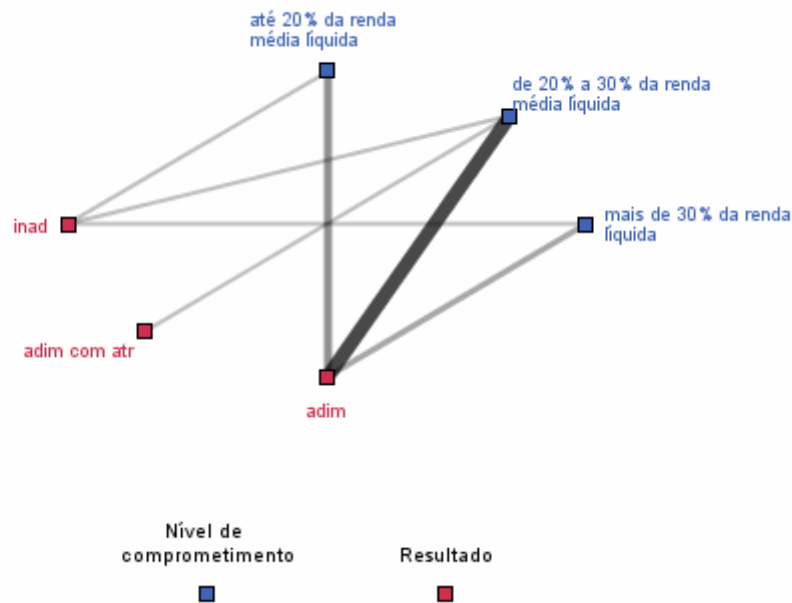


Figura 10- Gráfico de rede: variável de saída e nível de comprometimento.

Fonte: Dados da pesquisa, 2008.

A figura 11 mostra a relação entre a variável resultado e a variável principal atividade desenvolvida pelo cooperado. Percebe-se que há uma forte relação entre a categoria adimplente e a categoria pecuarista. Este resultado também deve ser analisado com cuidado, pois, a atividade de pecuária é majoritária, representando 55,9% dos cooperados. Algumas atividades não apresentaram relação com inadimplência, como as variáveis: autônomo, auxiliar de escritório, contador, estudante, mecânico, pensionista, professor do ensino superior e tabelião. A atividade profissional liberal só apresentou relação com a categoria inadimplente. A atividade agropecuária foi a única que apresentou relação com a categoria adimplente com atrasos/renegeações constantes.

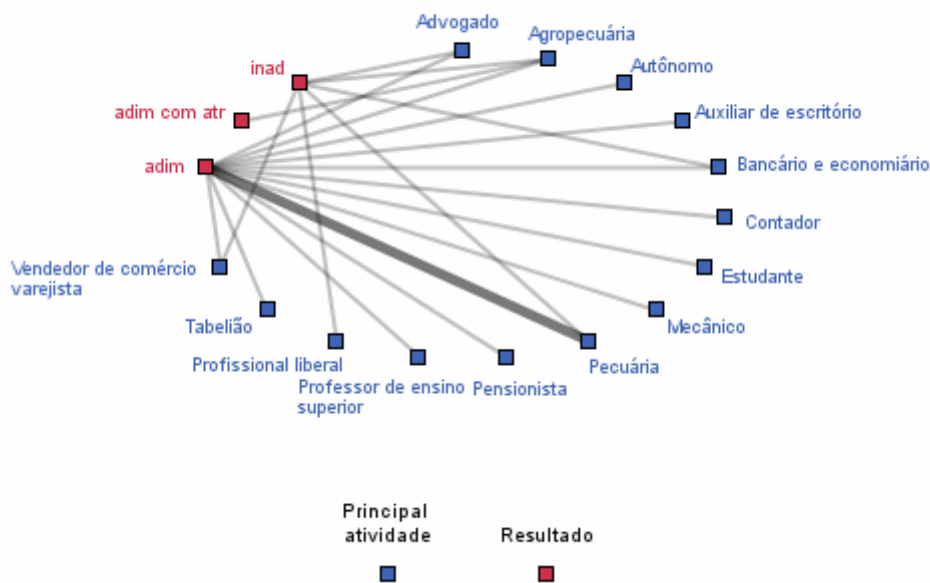


Figura 11-Gráfico de rede: variável de saída e principal atividade do cooperado.

Fonte: Dados da pesquisa, 2008.

Não foram encontrados valores ausentes na base de dados estudada, portanto, a limpeza dos dados através do preenchimento com a medida estatística “moda” não foi necessária.

4.2 Modelagem e Mineração de Dados

A figura 12 mostra o diagrama do modelo de mineração de dados desenvolvido para o problema proposto¹².

¹² Maior discussão pode ser encontrada na seção 3.5 onde há detalhamento sobre a função dos nós.

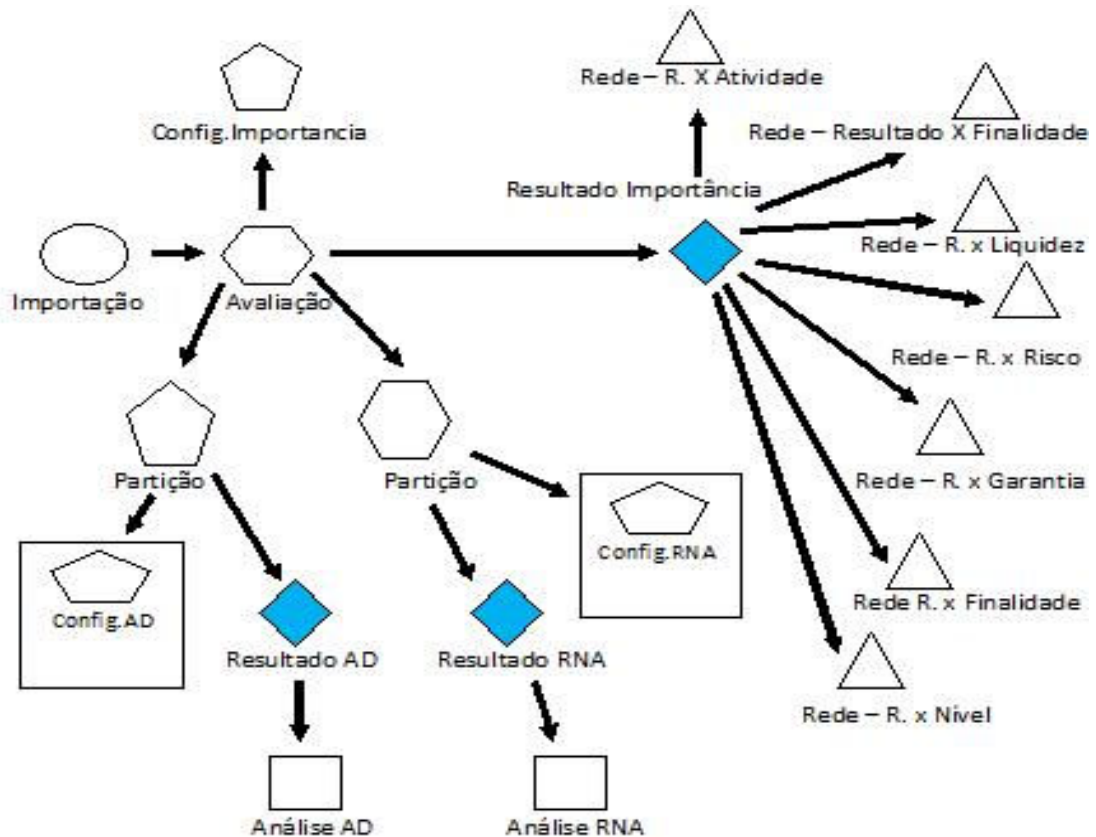


Figura 12- Modelo de mineração de dados desenvolvido.

Fonte: Dados da pesquisa, 2008.

Conforme demonstra o diagrama da figura 12, o fluxo inicia-se com a importação dos dados de planilha eletrônica, com o nódulo importação. A próxima etapa consiste no nó de avaliação da base de dados, com a verificação do tipo de variável, valores, valores ausentes e a definição das variáveis em: entrada, saída, entrada e saída, partição ou a exclusão de variáveis.

Acima do nódulo de avaliação encontra-se o nódulo de configuração da classificação por ordem de importância das variáveis. Este nódulo permite verificar a importância da variável, auxilia, portanto, na decisão de incluir ou eliminar variáveis.

À frente do nó de avaliação encontra-se o nó resultado do nó de classificação por ordem de importância. As variáveis de entrada consideradas importantes e incluídas nos modelos foram submetidas ao gráfico de rede, para avaliar o relacionamento com a variável de saída. No diagrama, os gráficos de rede estão situados após o nó resultado da classificação por ordem de importância.

Antes de configurar os modelos, foi realizado a partição dos dados em treinamento, teste e validação. No diagrama, este nó encontra-se abaixo do nó de avaliação da base de dados.

O nó de partição origina-se dois nós: nó de configuração dos algoritmos Árvore de Decisão C5.0 e Rede Neural e nó resultado dos modelos. À frente do nó de resultado dos modelos encontra-se o nó de análise, para examinar a acurácia do modelo em classificar e prever os modelos bem como investigar o tipo de acerto e erro do modelo.

4.2.1 Árvore de Decisão

Árvores de decisão são estruturas hierárquicas de decisão que podem, a partir destas, induzir regras de decisão.

Para melhor apresentação, o modelo Árvore de Decisão foi dividido em duas partes, com as variáveis mais significativas. As figuras 13 e 14 apresentam as Árvores de Decisão geradas.

A variável liquidez das garantias foi considerada a variável mais importante pelo algoritmo c5.0. A figura 13 mostra a avaliação da Árvore de Decisão a partir da categoria garantia de média liquidez (venda de 6 a 12 meses), com evidência de comportamento adimplente. Deste nó, a árvore aprofunda a avaliação com a análise da variável nível de comprometimento. Estas análises apontam comportamento inadimplente a categoria “mais de 30% da renda líquida”, as categorias “até 20% da renda média líquida” e “de 20% a 30% da renda média líquida” apresentam comportamento adimplente.

A categoria garantia de fácil liquidez (venda até 6 meses) apresenta comportamento adimplente com atrasos/renegociações contantes.

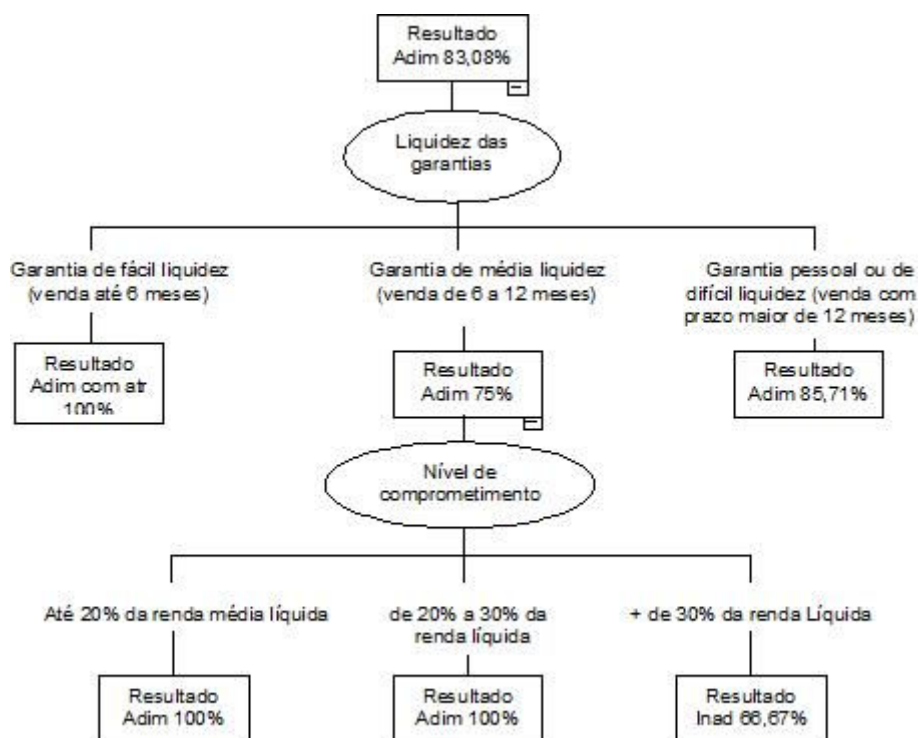


Figura 13 - Árvore de decisão – avaliação da categoria garantia de média liquidez.

Fonte: Dados da pesquisa, 2008.

A figura 14 analisa os resultados a partir da categoria garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses), com forte evidência de adimplência.

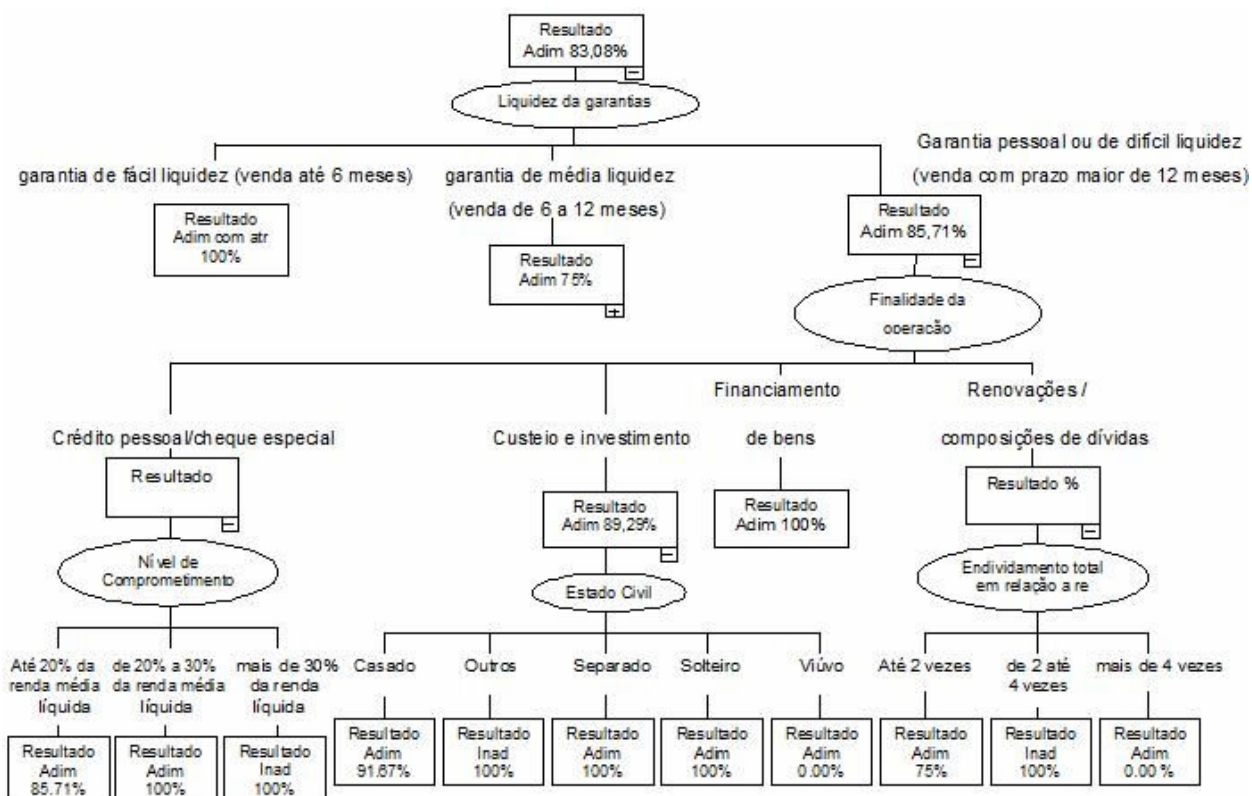


Figura 14 - Árvore de decisão – avaliação da categoria garantia pessoal ou de difícil liquidez.
 Fonte: Dados da pesquisa, 2008.

As variáveis mais importantes a partir deste nó que representa foram: nível de comprometimento, estado civil e endividamento total em relação a renda líquida ano.

A categoria crédito pessoal/cheque especial origina a avaliação da variável nível de comprometimento; a categoria custeio e investimento origina a avaliação da variável estado civil e a categoria renovações/composições de dívidas origina a avaliação da variável endividamento total em relação a renda líquida ano.

O modelo de Árvore de Decisão gerou 10 conjuntos de regras de decisão do tipo **se-então**. O primeiro conjunto está apresentado abaixo, o conjunto completo está exposto no apêndice 02.

- Conjunto de regra 1 – acurácia estimada 92,31%
Se Liquidez das garantias = garantia de fácil liquidez (venda até 6 meses) **então** adimplente com atrasos/renegociações constantes.
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.

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

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = crédito pessoal/cheque especial **e** Nível de comprometimento = até 20% da renda média líquida **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = crédito pessoal/cheque especial **e** Nível de comprometimento = de 20% a 30% da renda média líquida **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = crédito pessoal/cheque especial **e** Nível de comprometimento = mais de 30% da renda líquida **então** inadimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = custeio e investimento **e** Estado civil = Casado **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = custeio e investimento **e** Estado civil = Outros **então** inadimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = custeio e investimento **e** Estado civil = Separado **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = custeio e investimento **e** Estado civil = Solteiro **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = custeio e investimento **e** Estado civil = Viuvo **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com

prazo maior de 12 meses) e Finalidade da operação = financiamento de bens **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) e Finalidade da operação = renovações/composições de dívidas e Endividamento total em relação a re = até 2 vezes **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) e Finalidade da operação = renovações/composições de dívidas e Endividamento total em relação a re = de 2 a 4 vezes **então** inadimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) e Finalidade da operação = renovações/composições de dívidas e Endividamento total em relação a re = mais de 4 vezes **então** adimplente.

A tabela 03 mostra o resultado da avaliação das três partições do modelo Árvore de Decisão para o problema.

Tabela 03 – Avaliação do modelo desenvolvido pelo método árvore de decisão.

| Partição | Treinamento | | Teste | | Validação | |
|----------|-------------|------|-------|--------|-----------|--------|
| | n | % | n | % | n | % |
| Correto | 65 | 100% | 62 | 88,57% | 70 | 92,11% |
| Erro | 0 | 0% | 8 | 11,43% | 6 | 7,89% |
| Total | 65 | | 70 | | 76 | |

Fonte: Dados da pesquisa, 2008.

Dos registros classificados nas três partições, 197 foram classificados corretamente, representando 93,36% e 14 incorretos, representando 6,64%.

A tabela 04 apresenta a matriz de coincidência 3 por 3 dos resultados preditos e reais da partição “treinamento” do modelo árvore de decisão.

Tabela 04 – Matriz dos valores preditos e reais da partição treinamento da árvore de decisão.

| Valores preditos | | | |
|------------------------------|-------------------|----------------------------|---------------------|
| Valores reais | Adimplente | Adimplente com atr. | Inadimplente |
| Adimplente | 54 | 0 | 0 |
| Adimplente com atraso | 0 | 1 | 0 |
| Inadimplente | 0 | 0 | 10 |

Fonte: Dados da pesquisa, 2008.

A diagonal principal da matriz representa os valores preditos que foram classificados corretamente, os valores fora desta diagonal representam valores preditos incorretamente classificados. Na tabela 04, a diagonal principal é constituída pelos valores 54, 1 e 10.

Dos registros classificados corretamente na partição “treinamento”, 54 são adimplentes, 1 é adimplente com atrasos/renegociações constantes e 10 inadimplentes. A Árvore de Decisão não classificou incorretamente na partição “treinamento”, atingindo 100% de acerto.

A tabela 05 apresenta a matriz de coincidência dos valores preditos e reais na partição “teste” do modelo Árvore de Decisão.

Tabela 05 – Matriz dos valores preditos e reais da partição teste da árvore de decisão.

| Valores preditos | | | |
|------------------------------|-------------------|----------------------------|---------------------|
| Valores reais | Adimplente | Adimplente com atr. | Inadimplente |
| Adimplente | 62 | 0 | 1 |
| Adimplente com atraso | 0 | 0 | 0 |
| Inadimplente | 7 | 0 | 0 |

Fonte: Dados da pesquisa, 2008.

A Árvore de Decisão classificou somente adimplentes na partição “teste”, destes, 62 foram classificados corretamente. Dos registros classificados incorretamente, 7 foram classificados como adimplentes quando na verdade eram inadimplentes e 1 foi classificado como inadimplente quando era adimplente.

A tabela 06 apresenta a matriz de coincidência dos valores preditos e reais na partição “validação” do modelo Árvore de Decisão.

Tabela 06 – Matriz dos valores preditos e reais da partição validação da árvore de decisão.

| Valores preditos | | | |
|------------------------------|-------------------|----------------------------|---------------------|
| Valores reais | Adimplente | Adimplente com atr. | Inadimplente |
| Adimplente | 70 | 0 | 2 |
| Adimplente com atraso | 0 | 0 | 0 |
| Inadimplente | 4 | 0 | 0 |

Fonte: Dados da pesquisa, 2008.

Dos registros da partição “validação”, 70 adimplentes foram classificados corretamente. Da mesma forma que a partição “teste”, a Árvore de Decisão somente classificou corretamente cooperados adimplentes. Dos registros classificados incorretamente, 4 foram classificados como adimplentes quando na verdade eram inadimplentes e 2 foram classificados como inadimplentes quando na verdade eram adimplentes.

4.2.2 Rede Neural

RNA é um modelo matemático baseado na estrutura cerebral, ordenado em camadas e ligações.

A RNA deste estudo, conforme já discutido, foi constituída por três camadas: entrada, intermediária e saída. A camada de entrada foi constituída de 66 neurônios, a camada intermediária de 3 e a de saída 3. O aprendizado da rede foi do tipo supervisionado, pois, foi indicado os valores de saída do modelo.

A importância relativa das variáveis de entrada está apresentada na tabela 07. As variáveis estão listadas em ordem de importância, do mais importante para o menos importante, segundo o modelo gerado método RNA.

Varáveis que não são utilizadas atualmente no modelo empregado pela cooperativa foram consideradas importantes, como: idade, nível educacional, estado civil, gênero, capital e local de domicílio.

Tabela 07 – Avaliação da importância relativa das variáveis pelo método RNA.

| Variável | Importância relativa |
|-------------------------------------|----------------------|
| Finalidade da operação | 0,21018000 |
| Principal atividade | 0,01575530 |
| Nível de comprometimento | 0,01549780 |
| Patrimônio líquido pessoal | 0,00975731 |
| Relacionamento | 0,00600704 |
| Total endividamento | 0,00498873 |
| Endividamento total em relação a re | 0,00410150 |
| Idade | 0,00140974 |
| Operações ativas | 8,13580E-4 |
| Nível educacional | 8,07122E-4 |
| Estado civil | 6,57478E-4 |
| Liquidez das garantias | 6,46681E-4 |
| Gênero | 3,70123E-4 |
| Garantia das operações | 2,97283E-4 |
| Capital | 1,26708E-4 |
| Local de domicílio | 5,05494E-4 |

Fonte: Dados da pesquisa, 2008.

O resultado da avaliação das três partições do modelo RNA está apresentado na tabela 08.

Tabela 08 – Avaliação do modelo desenvolvido pelo método RNA.

| Partição | Treinamento | | Teste | | Validação | |
|----------|-------------|------|-------|--------|-----------|-------|
| | n | % | n | % | n | % |
| Correto | 73 | 100% | 61 | 91,04% | 59 | 83,1% |
| Erro | 0 | 0% | 6 | 8,96% | 12 | 16,9% |
| Total | 73 | | 67 | | 71 | |

Fonte: Dados da pesquisa, 2008.

Dos registros classificados nas três partições, 193 foram classificados corretamente, representando 91,47% e 18 incorretos representando 8,53%.

A tabela 09 apresenta a matriz de coincidência dos valores preditos e reais na partição “treinamento” da RNA.

Tabela 09 – Matriz dos valores preditos e reais da partição treinamento da RNA.

| Valores preditos | | | |
|------------------------------|-------------------|----------------------------|---------------------|
| Valores reais | Adimplente | Adimplente com atr. | Inadimplente |
| Adimplente | 70 | 0 | 0 |
| Adimplente com atraso | 0 | 0 | 0 |
| Inadimplente | 0 | 0 | 3 |

Fonte: Dados da pesquisa, 2008.

Dos registros classificados corretamente na partição “treinamento”, 70 são adimplentes e 3 são inadimplentes. Da mesma forma que a Árvore de Decisão, a RNA não classificou incorretamente na partição “treinamento”, atingindo também, 100% de acerto.

A tabela 10 apresenta a matriz de coincidência dos valores preditos e reais na partição “teste” da RNA.

Tabela 10 – Matriz dos valores preditos e reais da partição teste da RNA.

| Valores preditos | | | |
|------------------------------|-------------------|----------------------------|---------------------|
| Valores reais | Adimplente | Adimplente com atr. | Inadimplente |
| Adimplente | 58 | 0 | 1 |
| Adimplente com atraso | 0 | 0 | 1 |
| Inadimplente | 4 | 0 | 3 |

Fonte: Dados da pesquisa, 2008.

Dos registros classificados corretamente na partição “teste”, 58 foram adimplentes e 3 inadimplentes. Dos registros incorretos, 4 foram classificados como adimplentes quando na verdade eram inadimplentes, 1 foi classificado como inadimplente quando na verdade era adimplente e 1 foi classificado como adimplente quando na verdade era adimplente com atrasos/renegeiações constantes.

A tabela 11 apresenta a matriz de coincidência dos valores preditos e reais na partição “validação” da RNA.

Tabela 11 – Matriz dos valores preditos e reais da partição validação da RNA.

| Valores preditos | | | |
|------------------------------|-------------------|----------------------------|---------------------|
| Valores reais | Adimplente | Adimplente com atr. | Inadimplente |
| Adimplente | 58 | 0 | 2 |
| Adimplente com atraso | 0 | 0 | 0 |
| Inadimplente | 10 | 0 | 1 |

Fonte: Dados da pesquisa, 2008.

Dos registros da partição “validação”, 58 adimplentes e 1 inadimplente foram classificados corretamente. Dos registros classificados incorretamente, 10 foram classificados como adimplentes quando na verdade eram inadimplentes e 2 foram classificados como inadimplentes quando na verdade eram adimplentes.

4.2.3 Modelo Empregado pela Cooperativa

Esta seção avaliou o modelo atualmente empregado pela cooperativa na classificação dos cooperados. Apresenta, assim como nos demais modelos analisados, o tipo de erro gerado pelo modelo.

A tabela 12 apresenta o percentual de classificações corretas e erros do modelo atual empregado pela cooperativa.

Tabela 12 – Avaliação do modelo atual da cooperativa.

| Avaliação | n | % |
|------------------|----------|----------|
| Correto | 194 | 91,94% |
| Erro | 17 | 8,06% |
| Total | 211 | |

Fonte: Dados da pesquisa, 2008.

O modelo atualmente empregado pela cooperativa obteve 91,94% de acerto na classificação dos cooperados da base de dados em adimplentes e inadimplentes.

A tabela 13 apresenta a matriz de coincidência dos valores preditos e reais do modelo empregado.

Tabela 13 – Matriz dos valores preditos e reais do modelo utilizado.

| | | | |
|----------------------------|-------------------|----------------------------|---------------------|
| Valores preditos | | | |
| Valores reais | Adimplente | Adimplente com atr. | Inadimplente |
| Adimplente | 185 | 0 | 4 |
| Adimplente com atr. | 0 | 0 | 1 |
| Inadimplente | 12 | 0 | 9 |

Fonte: Dados da pesquisa, 2008.

O modelo da cooperativa previu corretamente 185 adimplentes e 9 inadimplentes. Dos registros incorretos, 12 foram classificados como adimplentes quando na verdade eram inadimplentes, 4 foram classificados como inadimplentes quando eram adimplentes e 1 foi classificado como inadimplente quando era adimplente com atrasos/renegociações constantes.

4.2.4 Avaliação Comparativa dos Modelos

Nesta seção são avaliados comparativamente os modelos desenvolvidos neste estudo: RNA e Árvore de Decisão (algoritmo C5.0) com o modelo atualmente empregado pela cooperativa. Para a realização desta avaliação, foi utilizado o percentual total de valores preditos incorretamente.

A tabela 14 apresenta o resultado do percentual dos erros dos modelos em estudo.

Tabela 14 – Avaliação comparativa.

| Erro Total de Classificação | | |
|------------------------------------|--------------------|---------------------|
| Árvore de Decisão | Rede Neural | Modelo atual |
| % | % | % |
| 6,64 | 8,53 | 8,06 |

Fonte: Dados da pesquisa, 2008.

O melhor desempenho foi da técnica Árvore de Decisão, seguido pelo modelo utilizado pela cooperativa. A RNA obteve pior desempenho entre os modelos avaliados, contudo, atingiu 91,47% de previsões corretas.

A variável “risco atribuído” representa a classificação gerada pelo modelo utilizado pela cooperativa. Esta variável foi categorizada como importante na

aplicação da função de correlação de *Pearson*. Demonstra assim que, à luz deste método, o modelo tem um bom desempenho.

O modelo utilizado pela cooperativa obteve mais de 90% de acerto, entretanto, o modelo baseado em Árvore de Decisão obteve melhor desempenho, mostrando que podem ser de grande valia para a organização.

5 CONCLUSÕES E PROPOSTA PARA ESTUDOS POSTERIORES

O objetivo deste estudo foi desenvolver um modelo 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. Concomitantemente ao desenvolvimento dos modelos foi avaliado o modelo atualmente utilizado pela cooperativa.

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, teste e validação.

O modelo baseado em Mineração de Dados, Árvore de Decisão obteve melhor resultado que o modelo empregado pela cooperativa, todavia, todos os modelos tiveram resultados acima de 90% de acerto. A Árvore de Decisão obteve erro total de classificação de 6,64%, a RNA 8,53% e o modelo empregado atualmente empregado 8,06%.

Não obstante o bom resultado obtido pelo modelo da cooperativa, o processo de descoberta de conhecimento e o uso dos modelos baseados em Mineração de Dados desenvolvidos podem trazer algumas vantagens práticas para a cooperativa. A compreensão mais profunda das variáveis e seus relacionamentos podem ajudar a melhor classificar e prever o comportamento dos cooperados. Pela avaliação da importância 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 de cooperados *outliers* ou que estão fora do padrão, pode ser relevante para a formação de uma nova classificação ou inversamente, compor padrões indesejados.

Além desses fatores, a diferença entre o modelo de melhor desempenho, que foi a Árvore de Decisão e o modelo atualmente empregado é de 1,42%. Esta quantidade de erro a mais pode implicar em operações indesejáveis, tendo em vista que a cooperativa pode não conceder crédito incorretamente ou realizar operações que resultem em inadimplências.

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 *outliers*, principalmente para a variável capital, com o objetivo de verificar a existência de novos padrões e classificações; análise da divergência de desempenho entre as partições da base de dados.

REFERÊNCIAS

Banco Central do Brasil. **Relatório anual 2006**. v. 42. Brasília: BCB. 2007.

BERRY, Michael J. A.; LINOFF, Gordon. **Data mining techniques: for marketing, sales, and customer relationship management**. 2nd. ed. Indianapolis: Wiley Publishing. 2004.

BIALOSKOSRKI NETO, Sigismundo. Um ensaio sobre desempenho econômico e participação em cooperativas agropecuárias. **RER**. v. 45, n. 1, p. 119-138, jan/mar. 2008.

CARDOSO, Olinda Nogueira Paes; MACHADO, Rosa Teresa Moreira. Gestão do conhecimento usando *data mining*: estudo de caso na Universidade Federal de Lavras. **Revista de Administração Pública**. V. 42, n. 3, p. 495-528, maio/jun. 2008.

CHAIA, Alexandre Jorge. **Modelos de gestão do risco de crédito e sua aplicabilidade ao mercado brasileiro**. Dissertação de Mestrado. FEA/USP, 2003.

DASU, Tamraparni; JOHNSON, Theodore. **Exploratory data mining and data cleaning**. New Jersey: John Wiley & Sons, 2003.

FAYYAD, Usama M.; PIATETSKY-SHAPIRO, Gregory.;SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. **American Association for Artificial Intelligence**. CA, USA: AAAI/MIT. p. 37-54. 1996.

FERREIRA, Jorge Brantes. **Mineração de dados na retenção de clientes em telefonia celular**. Dissertação de Mestrado. PUC-RIO. Rio de Janeiro/RJ, 2005.

FERRO, Mariza; LEE, Huei Diana. O processo de *KDD – Knowledge Discovery in Databases* para aplicações na medicina. In: SEMANA DE INFORMÁTICA DE CASCAVEL, 2001, Cascavel. **Anais do SEMINC 2001**. p. 57-62. 2001.

FRANTZ, Walter. A cooperação e a prática cooperativa. **Perspectiva econômica**. v. 151, n. 51, p. 53-70. 1985. (série coop. n. 16).

GEHRKE, Johannes. Decision tree. In: YE, Nong. (ed.) **The handbook of data mining**. New Jersey: Lawrence Erlbaum Associates, p. 3-23. 2003.

GOIÁS (Estado). Sindicato e Organização das Cooperativas do Estado de Goiás – OCG. **Cooperativismo passo a passo**: manual de orientação. Goiânia, 2002, 5 ed.

GOLDSCHMIDT, Ronaldo; EMMANUEL, Passos. **Data mining**: uma guia prático. Rio de Janeiro: Elsevier, 2005.

IRION, João Eduardo Oliveira. **Cooperativismo e economia social**. São Paulo: STS, 1997.

KOH, Hian Chye; TAN, Wei Chin; GOH, Chwee Peng. A two-step method to construct credit scoring models with data mining techniques. **International Journal of Business and Information**. v. 1, n.1, pp. 96-118. 2006.

LAI, Kin Keung; YU, Lean; WANG, Shouyang; ZHOU, Ligang. Credit risk analysis using a reliability-based neural network ensemble model. **ICANN 2006**, Part II, Incs 4132. pp. 682-690. 2006.

LAPPONI, Juan Carlos. **Estatística usando excel**. Rio de Janeiro: Elsevier, 2005.

LAROSE, T. Daniel. **Discovering Knowledge in Data: an introduction to data mining**. New Jersey: John Wiley & Sons, 2005.

LEMOS, Eliane Prezepiorski; STEINER, Maria Teresinha Arns; NIEVOLA, Julio César. Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining. **Revista de Administração**. v. 40, n.3, p. 225-234, jul/ago/set. 2005.

MENEGÁRIO, Alexandre Hattner. **Emprego de indicadores sócio-econômicos na avaliação financeira de cooperativas agropecuárias**. Dissertação de Mestrado. Escola Superior de Agricultura Luiz de Queiroz. Piracicaba/SP, 2000.

MESTER, Loretta J. What's the point of credit scoring? **Business Review**. Sep/Oct. pp. 3-16. 1997.

OLIVEIRA, Djalma de Pinho Rebouças de. **Manual de gestão de cooperativas: uma abordagem prática**. São Paulo: Atlas, 2001.

PIATETSKY-SHAPIRO, Gregory *et al.* Progress and challenges in knowledge discovery in databases. **American Association for Artificial Intelligence**. CA, USA: AAAI. pp.77-82. 1994.

PIDD, Michael. **Modelagem empresarial: ferramentas para tomada de decisão**. Porto Alegre: Artes Médicas, 1998.

PINHEIRO, Marcos Antonio Henriques. **Cooperativas de crédito: história da evolução normativa no Brasil**. Brasília: BCB, 2007. 94 p.

PINHO, Diva Benevides. **O Pensamento cooperativo e o cooperativismo**. São Paulo: CNPq, 1982, v. 1.

O cooperativismo no Brasil: Da Vertente Pioneira à Vertente Solidária. São Paulo: Saraiva, 2004.

SOARES, Marden Marques; SOBRINHO, Abelardo Soares de Melo. **Microfinanças: o papel do Banco Central do Brasil e a importância do cooperativismo de crédito**. Brasília: BCB, 2007.

SOUSA, Marcos de Moraes; BUZIN, Estevão Júlio W. Keglevich de; SILVA, Walter Sales. Cooperativismo de crédito na dinâmica da agricultura familiar. **Revista da FACER**. v. 1, n. 1, p. 69-79, fev. 2008.

TRIOLA, Mário, F. **Introdução à estatística**. 9 ed. Rio de Janeiro: LTC, 2005.

YIN, Robert K. **Estudo de Caso: Planejamento e Métodos**. 2 ed. Porto Alegre: Bookman, 2001.

APÊNDICE I

| Força da relação ¹³ | Q | % | Variável |
|--------------------------------|-----|-------|--|
| Forte | 169 | 80,09 | Liquidez das garantias = "garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses)" Resultado = "adim" |
| Média | 20 | 9,48 | Liquidez das garantias = "garantia de média liquidez (venda de 6 a 12 meses)" Resultado = "adim" |
| Média | 19 | 9,00 | Liquidez das garantias = "garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses)" Resultado = "inad" |
| Fraca | 2 | 0,95 | Liquidez das garantias = "garantia de média liquidez (venda de 6 a 12 meses)" Resultado = "inad" |
| Fraca | 1 | 0,47 | Liquidez das garantias = "garantia de fácil liquidez (venda até 6 meses)" Resultado = "adim com atr" |

Dados relativos ao gráfico de rede: variável de saída e liquidez das garantias

| Força da relação | Q | % | Relação |
|------------------|-----|-------|---|
| Forte | 162 | 76,78 | Garantia das operações = "pessoal" Resultado = "adim" |
| Média | 23 | 10,90 | Garantia das operações = "alienação fiduciária/warrants" Resultado = "adim" |
| Média | 19 | 9,00 | Garantia das operações = "pessoal" Resultado = "inad" |
| Fraca | 3 | 1,42 | Garantia das operações = "penhor cedular/caução de títulos" Resultado = "adim" |
| Fraca | 2 | 0,95 | Garantia das operações = "alienação fiduciária/warrants" Resultado = "inad" |
| Fraca | 1 | 0,47 | Garantia das operações = "hipoteca – capital social" Resultado = "adim com atr" |
| Fraca | 1 | 0,47 | Garantia das operações = "hipoteca – capital social" Resultado = "adim" |

Dados relativos ao gráfico de rede: variável de saída e garantia das operações

| Força da relação | Q | % | Variável |
|------------------|-----|-------|--|
| Forte | 123 | 58,29 | Risco atribuído = "C" Resultado = "adim" |
| Forte | 56 | 26,54 | Risco atribuído = "B" Resultado = "adim" |
| Fraca | 9 | 4,27 | Risco atribuído = "D" Resultado = "inad" |
| Fraca | 7 | 3,32 | Risco atribuído = "C" Resultado = "inad" |
| Fraca | 7 | 3,32 | Risco atribuído = "A" Resultado = "adim" |
| Fraca | 5 | 2,37 | Risco atribuído = "B" Resultado = "inad" |
| Fraca | 3 | 1,42 | Risco atribuído = "D" Resultado = "adim" |
| Fraca | 1 | 0,47 | Risco atribuído = "D" Resultado = "adim com atr" |

Dados relativos ao gráfico de rede: variável de saída e risco atribuído

¹³ Foram considerados os seguintes valores da força das relações: $0 \leq \text{fraca} \leq 15$; $15 < \text{média} < 35$; $\text{forte} \geq 35$.

| Força da relação | Q | % | Variável |
|-------------------------|----------|----------|---|
| Forte | 91 | 43,13 | Finalidade da operação = "custeio e investimento" Resultado = "adim" |
| Forte | 61 | 28,91 | Finalidade da operação = "crédito pessoal/cheque especial" Resultado = "adim" |
| Média | 33 | 15,64 | Finalidade da operação = "financiamento de bens" Resultado = "adim" |
| Fraca | 7 | 3,32 | Finalidade da operação = "renovações/composições de dívidas" Resultado = "inad" |
| Fraca | 7 | 3,32 | Finalidade da operação = "crédito pessoal/cheque especial" Resultado = "inad" |
| Fraca | 5 | 2,37 | Finalidade da operação = "custeio e investimento" Resultado = "inad" |
| Fraca | 4 | 1,90 | Finalidade da operação = "renovações/composições de dívidas" Resultado = "adim" |
| Fraca | 2 | 0,95 | Finalidade da operação = "financiamento de bens" Resultado = "inad" |
| Fraca | 1 | 0,47 | Finalidade da operação = "renovações/composições de dívidas" Resultado = "adim com atr" |

Dados relativos ao gráfico de rede: variável de saída e finalidade da operação

| Força da relação | Q | % | Variável |
|-------------------------|----------|----------|---|
| Forte | 114 | 54,03 | Nível de comprometimento = "de 20% a 30% da renda média líquida" Resultado = "adim" |
| Média | 46 | 21,8 | Nível de comprometimento = "até 20% da renda média líquida" Resultado = "adim" |
| Fraca | 29 | 13,74 | Nível de comprometimento = "mais de 30% da renda líquida" Resultado = "adim" |
| Fraca | 8 | 3,79 | Nível de comprometimento = "mais de 30% da renda líquida" Resultado = "inad" |
| Fraca | 8 | 3,79 | Nível de comprometimento = "até 20% da renda média líquida" Resultado = "inad" |
| Fraca | 5 | 2,37 | Nível de comprometimento = "de 20% a 30% da renda média líquida" Resultado = "inad" |
| Fraca | 1 | 0,47 | Nível de comprometimento = "de 20% a 30% da renda média líquida" Resultado = "adim com atr" |

Dados relativos ao gráfico de rede: variável de saída e nível de comprometimento

| Força da relação | Q | % | Variável |
|-------------------------|----------|----------|---|
| Forte | 165 | 78,20 | Principal atividade = "Pecuária" Resultado = "adim" |
| Fraca | 15 | 7,11 | Principal atividade = "Pecuária" Resultado = "inad" |
| Fraca | 8 | 3,79 | Principal atividade = "Agropecuária" Resultado = "adim" |
| Fraca | 5 | 2,37 | Principal atividade = "Bancário e economiário" |

| | | | |
|-------|---|------|---|
| | | | Resultado = "adim" |
| Fraca | 2 | 0,95 | Principal atividade = "Agropecuária" Resultado = "inad" |
| Fraca | 2 | 0,95 | Principal atividade = "Contador" Resultado = "adim" |
| Fraca | 1 | 0,47 | Principal atividade = "Profissional liberal" Resultado = "inad" |
| Fraca | 1 | 0,47 | Principal atividade = "Agropecuária" Resultado = "adim com atr" |
| Fraca | 1 | 0,47 | Principal atividade = "Vendedor de comércio varejista" Resultado = "adim" |
| Fraca | 1 | 0,47 | Principal atividade = "Vendedor de comércio varejista" Resultado = "inad" |
| Fraca | 1 | 0,47 | Principal atividade = "Advogado" Resultado = "inad" |
| Fraca | 1 | 0,47 | Principal atividade = "Professor de ensino superior" Resultado = "adim" |
| Fraca | 1 | 0,47 | Principal atividade = "Tabelião" Resultado = "adim" |
| Fraca | 1 | 0,47 | Principal atividade = "Autônomo" Resultado = "adim" |
| Fraca | 1 | 0,47 | Principal atividade = "Estudante" Resultado = "adim" |
| Fraca | 1 | 0,47 | Principal atividade = "Auxiliar de escritório" Resultado = "adim" |
| Fraca | 1 | 0,47 | Principal atividade = "Mecânico" Resultado = "adim" |
| Fraca | 1 | 0,47 | Principal atividade = "Advogado" Resultado = "adim" |
| Fraca | 1 | 0,47 | Principal atividade = "Pensionista" Resultado = "adim" |
| Fraca | 1 | 0,47 | Principal atividade = "Bancário e economiário" Resultado = "inad" |

Dados relativos ao gráfico de rede: variável de saída e principal atividade

APÊNDICE II

- Conjunto de regra 1 – acurácia estimada 92,31%
 - Se** Liquidez das garantias = garantia de fácil liquidez (venda até 6 meses) **então** adimplente com atrasos/renegociações constantes.
 - 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.
 - Se** Liquidez das garantias = garantia de média liquidez (venda de 6 a 12 meses) **e** Nível de comprometimento = mais de 30% da renda líquida **então** inadimplente.
 - Se** Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = crédito pessoal/cheque especial **e** Nível de comprometimento = até 20% da renda média líquida **então** adimplente.
 - Se** Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = crédito pessoal/cheque especial **e** Nível de comprometimento = de 20% a 30% da renda média líquida **então** adimplente.
 - Se** Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = crédito pessoal/cheque especial **e** Nível de comprometimento = mais de 30% da renda líquida **então** inadimplente.
 - Se** Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = custeio e investimento **e** Estado civil = Casado **então** adimplente.
 - Se** Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = custeio e investimento **e** Estado civil = Outros **então** inadimplente.
 - Se** Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com

prazo maior de 12 meses) **e** Finalidade da operação = custeio e investimento **e** Estado civil = Separado **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = custeio e investimento **e** Estado civil = Solteiro **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = custeio e investimento **e** Estado civil = Viuvo **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = financiamento de bens **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = renovações/composições de dívidas **e** Endividamento total em relação a re = até 2 vezes **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = renovações/composições de dívidas **e** Endividamento total em relação a re = de 2 a 4 vezes **então** inadimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Finalidade da operação = renovações/composições de dívidas **e** Endividamento total em relação a re = mais de 4 vezes **então** adimplente.

- Conjunto de regra 2 – acurácia estimada 91,83%

Se Principal atividade = ["Advogado" "Autônomo" "Auxiliar de escritório" "Mecânico" "Pensionista" "Professor de ensino superior" "Profissional liberal"] **então** adimplente.

Se Principal atividade = ["Agropecuária"] **então** adimplente.

Se Principal atividade = ["Bancário e economiário"] **então** adim

Se Principal atividade = ["Contador"] **então** adimplente.

Se Principal atividade = ["Estudante"] **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = até 20% da renda média líquida **e** Nível educacional = ["1 grau"] **e** Idade <= 39 **então**

inadimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = até 20% da renda média líquida **e** Nível educacional = ["1 grau"] **e** Idade > 39 **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = até 20% da renda média líquida **e** Nível educacional = ["1 grau inc" "Pós graduação" "superior completo"] **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = até 20% da renda média líquida **e** Nível educacional = ["2 grau"] **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = até 20% da renda média líquida **e** Nível educacional = ["Superior inc"] **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = até 20% da renda média líquida **e** Nível educacional = ["superior completo"] **então** inadimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Operações ativas = frequentemente **e** Capital <= 11518,240 **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Operações ativas = frequentemente **e** Capital > 11518,240 **então** inadimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Operações ativas = nunca **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Operações ativas = permanentemente **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = mais de 30% da renda líquida **e** local de domicílio = Zona urbana **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Nível de comprometimento = mais de 30% da renda líquida **e** local de domicílio = Zona rural **então** inadimplente.

Se Principal atividade = ["Tabelião"] **então** adimplente.

Se Principal atividade = ["Vendedor de comércio varejista"] **então** inadimplente.

- Conjunto de regra 3 – acurácia estimada 83,30%

Se Garantia das operações = alienação fiduciária/warrants **e** Idade <= 39 **então** inadimplente.

Se Garantia das operações = alienação fiduciária/warrants **e** Idade > 39 **então** adimplente.

Se Garantia das operações = hipoteca – capital social **então** adimplente com atrasos/renegociações constantes.

Se Garantia das operações = penhor cedular/caução de títulos **então** adimplente.

Se Garantia das operações = pessoal **e** Principal atividade = ["Advogado" "Autônomo" "Auxiliar de escritório" "Contador" "Mecânico" "Pensionista" "Professor de ensino superior" "Profissional liberal" "Vendedor de comércio varejista"] **então** adimplente.

Se Garantia das operações = pessoal **e** Principal atividade = ["Agropecuária"] **então** adimplente.

Se Garantia das operações = pessoal **e** Principal atividade = ["Bancário e economiário"] **então** adimplente.

Se Garantia das operações = pessoal **e** Principal atividade = ["Estudante"] **então** adimplente.

Se Garantia das operações = pessoal **e** Principal atividade = ["Pecuária"] **e** Estado civil = Casado **e** Idade <= 47 **e** Finalidade da operação = crédito pessoal/cheque especial **então** inadimplente.

Se Garantia das operações = pessoal **e** Principal atividade = ["Pecuária"] **e** Estado civil = Casado **e** Idade <= 47 **e** Finalidade da operação = custeio e investimento **e** Idade <= 41 **então** adimplente.

Se Garantia das operações = pessoal **e** Principal atividade = ["Pecuária"] **e** Estado civil = Casado **e** Idade <= 47 **e** Finalidade da operação = custeio e investimento **e** Idade > 41 **e** Total endividamento = acima de 12 vezes **então** inadimplente.

Se Garantia das operações = pessoal **e** Principal atividade = ["Pecuária"] **e** Estado civil = Casado **e** Idade <= 47 **e** Finalidade da operação = custeio e investimento **e** Idade > 41 **e** Total endividamento = até 4 vezes **então** inadimplente.

Se Garantia das operações = pessoal **e** Principal atividade = ["Pecuária"] **e**

Estado civil = Casado e Idade <= 47 e Finalidade da operação = custeio e investimento e Idade > 41 e Total endividamento = de 4 a 8 vezes **então** inadimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Pecuária"] e Estado civil = Casado e Idade <= 47 e Finalidade da operação = custeio e investimento e Idade > 41 e Total endividamento = de 8 a 12 vezes **então** adimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Pecuária"] e Estado civil = Casado e Idade <= 47 e Finalidade da operação = financiamento de bens **então** adimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Pecuária"] e Estado civil = Casado e Idade <= 47 e Finalidade da operação = renovações/composições de dívidas e Relacionamento = opera com a cooperativa até 1 ano **então** inadimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Pecuária"] e Estado civil = Casado e Idade <= 47 e Finalidade da operação = renovações/composições de dívidas e Relacionamento = opera com a cooperativa de 1 a 3 anos **então** inadimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Pecuária"] e Estado civil = Casado e Idade <= 47 e Finalidade da operação = renovações/composições de dívidas e Relacionamento = opera com a cooperativa há mais de 3 anos **então** adimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Pecuária"] e Estado civil = Casado e Idade > 47 **então** adimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Pecuária"] e Estado civil = Outros **então** inadimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Pecuária"] e Estado civil = Separado **então** inadimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Pecuária"] e Estado civil = Solteiro **então** adimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Pecuária"] e Estado civil = Viuvo **então** adimplente.

Se Garantia das operações = pessoal e Principal atividade = ["Tabelião"] **então** adimplente.

- Conjunto de regra 4 – acurácia estimada 87,12%
 - Se** Garantia das operações = alienação fiduciária/warrants **então** adimplente.
 - Se** Garantia das operações = hipoteca – capital social **então** adimplente com atrasos/renegociações constantes.
 - Se** Garantia das operações = penhor cedular/caução de títulos **então** adimplente.
 - Se** Garantia das operações = pessoal **e** Finalidade da operação = crédito pessoal/cheque especial **e** Nível de comprometimento = até 20% da renda média líquida **e** Capital \leq 14874,290 **então** adimplente.
 - Se** Garantia das operações = pessoal **e** Finalidade da operação = crédito pessoal/cheque especial **e** Nível de comprometimento = até 20% da renda média líquida **e** Capital $>$ 14874,290 **então** inadimplente.
 - Se** Garantia das operações = pessoal **e** Finalidade da operação = crédito pessoal/cheque especial **e** Nível de comprometimento = de 20% a 30% da renda média líquida **então** adimplente.
 - Se** Garantia das operações = pessoal **e** Finalidade da operação = crédito pessoal/cheque especial **e** Nível de comprometimento = mais de 30% da renda líquida **então** inadimplente.
 - Se** Garantia das operações = pessoal **e** Finalidade da operação = custeio e investimento **e** local de domicílio = Zona urbana **então** adimplente.
 - Se** Garantia das operações = pessoal **e** Finalidade da operação = custeio e investimento **e** local de domicílio = Zona rural **então** inadimplente.
 - Se** Garantia das operações = pessoal **e** Finalidade da operação = financiamento de bens **então** adimplente.
 - Se** Garantia das operações = pessoal **e** Finalidade da operação = renovações/composições de dívidas **então** inadimplente.
- Conjunto de regra 5 – acurácia estimada 89,45%
 - Se** Garantia das operações = alienação fiduciária/warrants **e** Relacionamento = opera com a cooperativa até 1 ano **então** inadimplente.
 - Se** Garantia das operações = alienação fiduciária/warrants **e** Relacionamento = opera com a cooperativa de 1 a 3 anos **então** inadimplente.
 - Se** Garantia das operações = alienação fiduciária/warrants **e** Relacionamento = opera com a cooperativa há mais de 3 anos **então** adimplente.

Se Garantia das operações = hipoteca – capital social **então** adimplente com atraso/renegociação constante.

Se Garantia das operações = penhor cedular/caução de títulos **então** adimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = de 2 a 4 vezes **então** inadimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = até 20% da renda média líquida **e** Relacionamento = opera com a cooperativa até 1 ano **então** inadimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = até 20% da renda média líquida **e** Relacionamento = opera com a cooperativa de 1 a 3 anos **então** adimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = até 20% da renda média líquida **e** Relacionamento = opera com a cooperativa há mais de 3 anos **então** adimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Gênero = m **e** Nível educacional in ["1 grau"] **então** adimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Gênero = m **e** Nível educacional in ["1 grau inc"] **então** adimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Gênero = m **e** Nível educacional in ["2 grau"] **e** Idade <= 44 **e** Endividamento total em relação a re = até 2 vezes **então** adimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Gênero = m **e** Nível educacional in ["2 grau"] **e** Idade <= 44 **e** Endividamento total em relação a re = de 2 a 4 vezes **então** inadimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Gênero = m **e** Nível educacional in ["2 grau"] **e** Idade <= 44 **e** Endividamento total em relação a re = mais de 4 vezes **então** inadimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Gênero = m **e** Nível educacional in ["2 grau"] **e** Idade > 44 **então** adimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Gênero = m **e** Nível educacional = ["Pós graduação" "Superior inc" "superior"] **então** adimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Gênero = m **e** Nível educacional = ["superior completo"] **então** adimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = de 20% a 30% da renda média líquida **e** Gênero = f **então** adimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Nível de comprometimento = mais de 30% da renda líquida **então** inadimplente.

Se Garantia das operações = pessoal **e** Patrimônio líquido pessoal = sem patrimônio pessoal ou até 2 vezes **então** adimplente.

- Conjunto de regra 6 – acurácia estimada 88,69%

Se Principal atividade = ["Advogado" "Autônomo" "Auxiliar de escritório" "Mecânico" "Pensionista" "Professor de ensino superior" "Profissional liberal"] **então** adimplente.

Se Principal atividade = ["Agropecuária"] **então** adimplente.

Se Principal atividade = ["Bancário e economiário"] **então** adimplente.

Se Principal atividade = ["Contador"] **então** adimplente.

Se Principal atividade = ["Estudante"] **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Gênero = m **e** Nível de comprometimento = até 20% da renda média líquida **e** Capital <= 14874,290 **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Gênero = m **e** Nível de comprometimento = até 20% da renda média líquida **e** Capital > 14874,290 **então** inadimplente.

Se Principal atividade = ["Pecuária"] e Gênero = m e Nível de comprometimento = de 20% a 30% da renda média líquida e Endividamento total em relação a re = até 2 vezes **então** adimplente.

Se Principal atividade = ["Pecuária"] e Gênero = m e Nível de comprometimento = de 20% a 30% da renda média líquida e Endividamento total em relação a re = de 2 a 4 vezes e Finalidade da operação = crédito pessoal/cheque especial **então** adimplente.

Se Principal atividade = ["Pecuária"] e Gênero = m e Nível de comprometimento = de 20% a 30% da renda média líquida e Endividamento total em relação a re = de 2 a 4 vezes e Finalidade da operação = custeio e investimento **então** adimplente.

Se Principal atividade = ["Pecuária"] e Gênero = m e Nível de comprometimento = de 20% a 30% da renda média líquida e Endividamento total em relação a re = de 2 a 4 vezes e Finalidade da operação = financiamento de bens **então** adimplente.

Se Principal atividade = ["Pecuária"] e Gênero = m e Nível de comprometimento = de 20% a 30% da renda média líquida e Endividamento total em relação a re = de 2 a 4 vezes e Finalidade da operação = renovações/composições de dívidas **então** inadimplente.

Se Principal atividade = ["Pecuária"] e Gênero = m e Nível de comprometimento = de 20% a 30% da renda média líquida e Endividamento total em relação a re = mais de 4 vezes **então** adimplente.

Se Principal atividade = ["Pecuária"] e Gênero = m e Nível de comprometimento = mais de 30% da renda líquida e Finalidade da operação = crédito pessoal/cheque especial **então** inadimplente.

Se Principal atividade = ["Pecuária"] e Gênero = m e Nível de comprometimento = mais de 30% da renda líquida e Finalidade da operação = custeio e investimento **então** adimplente.

Se Principal atividade = ["Pecuária"] e Gênero = m e Nível de comprometimento = mais de 30% da renda líquida e Finalidade da operação = financiamento de bens **então** inadimplente.

Se Principal atividade = ["Pecuária"] e Gênero = m e Nível de comprometimento = mais de 30% da renda líquida e Finalidade da operação = renovações/composições de dívidas **então** inadimplente.

Se Principal atividade = ["Pecuária"] **e** Gênero = f **então** inadimplente.

Se Principal atividade = ["Tabelião"] **então** adimplente.

Se Principal atividade = ["Vendedor de comércio varejista"] **então** inadimplente.

- Conjunto de regra 7 – acurácia estimada 89,32%

Se Principal atividade in ["Advogado" "Autônomo" "Auxiliar de escritório" "Mecânico" "Pensionista" "Professor de ensino superior" "Profissional liberal"]
[Mode: adim] => adim

Se Principal atividade = ["Agropecuária"] **então** adimplente com atrasos/renegociações constantes.

Se Principal atividade = ["Bancário e economiário"] **então** adimplente.

Se Principal atividade = ["Contador"] **então** adimplente.

Se Principal atividade = ["Estudante"] **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa até 1 ano **e** Capital <= 1901,140 **e** Nível educacional = ["1 grau"] **então** inadimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa até 1 ano **e** Capital <= 1901,140 **e** Nível educacional = ["1 grau inc"] **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa até 1 ano **e** Capital <= 1901,140 **e** Nível educacional = ["2 grau"] **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa até 1 ano **e** Capital <= 1901,140 **e** Nível educacional = ["Pós graduação" "Superior inc" "superior" "superior completo"] **então** inad

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa até 1 ano **e** Capital > 1901,140 **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa de 1 a 3 anos **e** Finalidade da operação = crédito pessoal/cheque especial **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa de 1 a 3 anos **e** Finalidade da operação = custeio e investimento **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa de 1 a 3 anos **e** Finalidade da operação = financiamento de bens **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa de 1 a 3 anos **e** Finalidade da operação = renovações/composições de dívidas **então** inadimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** local de domicílio = Zona urbana **e** Total endividamento = acima de 12 vezes **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** local de domicílio = Zona urbana **e** Total endividamento = até 4 vezes **e** Finalidade da operação = crédito pessoal/cheque especial **então** inadimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** local de domicílio = Zona urbana **e** Total endividamento = até 4 vezes **e** Finalidade da operação = custeio e investimento **e** Operações ativas = frequentemente **então** inadimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** local de domicílio = Zona urbana **e** Total endividamento = até 4 vezes **e** Finalidade da operação = custeio e investimento **e** Operações ativas = nunca **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** local de domicílio = Zona urbana **e** Total endividamento = até 4 vezes **e** Finalidade da operação = custeio e investimento **e** Operações ativas = permanentemente **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** local de domicílio = Zona urbana **e** Total endividamento = até 4 vezes **e** Finalidade da operação = financiamento de bens **então** adim

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** local de domicílio = Zona urbana **e** Total endividamento = até 4 vezes **e** Finalidade da operação = renovações/composições de dívidas **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a

cooperativa há mais de 3 anos **e** local de domicílio = Zona urbana **e** Total endividamento = de 4 a 8 vezes **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** local de domicílio = Zona urbana **e** Total endividamento = de 8 a 12 vezes **então** adimplente.

Se Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** local de domicílio = Zona rural **então** inadimplente.

Se Principal atividade = ["Tabelião"] **então** adimplente.

Se Principal atividade = ["Vendedor de comércio varejista"] **então** inadimplente.

- Conjunto de regra 8 – acurácia estimada 92,24%

Se Liquidez das garantias = garantia de fácil liquidez (venda até 6 meses) **então** adimplente com atrasos/renegociações constantes.

Se Liquidez das garantias = garantia de média liquidez (venda de 6 a 12 meses) **e** Idade \leq 38 **então** inadimplente.

Se Liquidez das garantias = garantia de média liquidez (venda de 6 a 12 meses) **e** Idade $>$ 38 **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Advogado" "Autônomo" "Auxiliar de escritório" "Contador" "Estudante" "Mecânico" "Pensionista" "Professor de ensino superior" "Profissional liberal" "Vendedor de comércio varejista"] **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Agropecuária"] **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Bancário e economiário"] **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa até 1 ano **e** Gênero = m **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa até 1 ano **e** Gênero = f **então** inadimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa de 1 a 3 anos **e** Finalidade da operação = crédito pessoal/cheque especial **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa de 1 a 3 anos **e** Finalidade da operação = custeio e investimento **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa de 1 a 3 anos **e** Finalidade da operação = financiamento de bens **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa de 1 a 3 anos **e** Finalidade da operação = renovações/composições de dívidas **então** inadimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** Endividamento total em relação a re = até 2 vezes **e** Capital \leq 14874,290 **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** Endividamento total em relação a re = até 2 vezes **e** Capital $>$ 14874,290 **então** inadimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) **e** Principal atividade = ["Pecuária"] **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** Endividamento total em relação a re = de 2 a 4 vezes **e** local de domicílio =

Zona urbana e Operações ativas = frequentemente **então** inadimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) e Principal atividade = ["Pecuária"] e Relacionamento = opera com a cooperativa há mais de 3 anos e Endividamento total em relação a re = de 2 a 4 vezes e local de domicílio = Zona urbana e Operações ativas = nunca **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) e Principal atividade = ["Pecuária"] e Relacionamento = opera com a cooperativa há mais de 3 anos e Endividamento total em relação a re = de 2 a 4 vezes e local de domicílio = Zona urbana e Operações ativas = permanentemente **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) e Principal atividade = ["Pecuária"] e Relacionamento = opera com a cooperativa há mais de 3 anos e Endividamento total em relação a re = de 2 a 4 vezes e local de domicílio = Zona rural **então** inadimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) e Principal atividade = ["Pecuária"] e Relacionamento = opera com a cooperativa há mais de 3 anos e Endividamento total em relação a re = mais de 4 vezes **então** adimplente.

Se Liquidez das garantias = garantia pessoal ou de difícil liquidez (venda com prazo maior de 12 meses) e Principal atividade in ["Tabelião"] **então** adimplente.

- Conjunto de regra 9 – acurácia estimada 90,29%

Se Garantia das operações = alienação fiduciária/warrants e Idade <= 39 **então** inadimplente.

Se Garantia das operações = alienação fiduciária/warrants e Idade > 39 **então** adimplente.

Se Garantia das operações = hipoteca – capital social **então** adimplente com atrasos/renegociações constantes.

Se Garantia das operações = penhor cedular/caução de títulos **então** adimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a

cooperativa até 1 ano e Finalidade da operação = crédito pessoal/cheque especial **então** adimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a cooperativa até 1 ano e Finalidade da operação = custeio e investimento **então** adimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a cooperativa até 1 ano e Finalidade da operação = financiamento de bens **então** adimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a cooperativa até 1 ano e Finalidade da operação = renovações/composições de dívidas **então** inadimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a cooperativa de 1 a 3 anos e Finalidade da operação = crédito pessoal/cheque especial **então** adimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a cooperativa de 1 a 3 anos e Finalidade da operação = custeio e investimento **então** adimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a cooperativa de 1 a 3 anos e Finalidade da operação = financiamento de bens **então** adimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a cooperativa de 1 a 3 anos e Finalidade da operação = renovações/composições de dívidas **então** inadimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a cooperativa há mais de 3 anos e Patrimônio líquido pessoal = de 2 a 4 vezes **então** inadimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a cooperativa há mais de 3 anos e Patrimônio líquido pessoal = mais de 4 vezes e Total endividamento = ["acima de 12 vezes" "de 8 a 12 vezes"] **então** adimplente.

Se Garantia das operações = pessoal e Relacionamento = opera com a cooperativa há mais de 3 anos e Patrimônio líquido pessoal = mais de 4 vezes e Total endividamento = ["até 4 vezes"] e Endividamento total em relação a re = até 2 vezes **então** adimplente.

Se Garantia das operações = pessoal **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Total endividamento = ["até 4 vezes"] **e** Endividamento total em relação a re = de 2 a 4 vezes **então** inadimplente.

Se Garantia das operações = pessoal **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Total endividamento = ["até 4 vezes"] **e** Endividamento total em relação a re = mais de 4 vezes **então** adimplente.

Se Garantia das operações = pessoal **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** Patrimônio líquido pessoal = mais de 4 vezes **e** Total endividamento = ["de 4 a 8 vezes"] **então** adimplente.

Se Garantia das operações = pessoal **e** Relacionamento = opera com a cooperativa há mais de 3 anos **e** Patrimônio líquido pessoal = sem patrimônio pessoal ou até 2 vezes **então** adimplente.

- Conjunto de regra 10 – acurácia estimada 87,90%

Se Garantia das operações = alienação fiduciária/warrants **então** adimplente.

Se Garantia das operações = hipoteca – capital social **então** adimplente com atrasos/renegociações constantes.

Se Garantia das operações = penhor cedular/caução de títulos **então** adimplente.

Se Garantia das operações = pessoal **e** Relacionamento = opera com a cooperativa até 1 ano **e** Nível educacional in ["1 grau"] **e** Total endividamento = acima de 12 vezes **então** inadimplente.

Se Garantia das operações = pessoal **e** Relacionamento = opera com a cooperativa até 1 ano **e** Nível educacional in ["1 grau"] **e** Total endividamento = até 4 vezes **então** inadimplente.

Se Garantia das operações = pessoal **e** Relacionamento = opera com a cooperativa até 1 ano **e** Nível educacional in ["1 grau"] **e** Total endividamento = de 4 a 8 vezes **então** adimplente.

Se Garantia das operações = pessoal **e** Relacionamento = opera com a cooperativa até 1 ano **e** Nível educacional in ["1 grau"] **e** Total endividamento = de 8 a 12 vezes **então** inadimplente.

Se Garantia das operações = pessoal **e** Relacionamento = opera com a

- cooperativa até 1 ano e Nível educacional in ["1 grau inc"] **então** adimplente.
- Se** Garantia das operações = pessoal e Relacionamento = opera com a cooperativa até 1 ano e Nível educacional in ["2 grau"] **então** adimplente.
- Se** Garantia das operações = pessoal e Relacionamento = opera com a cooperativa até 1 ano e Nível educacional in ["Pós graduação" "superior" "superior completo"] **então** inadimplente.
- Se** Garantia das operações = pessoal e Relacionamento = opera com a cooperativa até 1 ano e Nível educacional in ["Superior inc"] **então** adimplente.
- Se** Garantia das operações = pessoal e Relacionamento = opera com a cooperativa de 1 a 3 anos **então** adimplente.
- Se** Garantia das operações = pessoal e Relacionamento = opera com a cooperativa há mais de 3 anos e Nível de comprometimento = até 20% da renda média líquida **então** adimplente.
- Se** Garantia das operações = pessoal e Nível de comprometimento = de 20% a 30% da renda média líquida e Operações ativas = frequentemente e Idade ≤ 47 **então** inadimplente.
- Se** Garantia das operações = pessoal e Nível de comprometimento = de 20% a 30% da renda média líquida e Operações ativas = frequentemente e Idade > 47 **então** adimplente.
- Se** Garantia das operações = pessoal e Nível de comprometimento = de 20% a 30% da renda média líquida e Operações ativas = nunca **então** adimplente.
- Se** Garantia das operações = pessoal e Nível de comprometimento = de 20% a 30% da renda média líquida e Operações ativas = permanentemente **então** adimplente.
- Se** Garantia das operações = pessoal e Nível de comprometimento = mais de 30% da renda líquida e local de domicílio = Zona urbana **então** adimplente.
- Se** Garantia das operações = pessoal e Nível de comprometimento = mais de 30% da renda líquida e local de domicílio = Zona rural **então** inadimplente.