



## O Uso de Data Mining na otimização da Análise de Concessão de Crédito

Eduardo Noronha de Andrade Freitas (UFG) [eduardonaf@gmail.com](mailto:eduardonaf@gmail.com)  
Carla Moema Moreira Duarte (UFG) [carlamoema@gmail.com](mailto:carlamoema@gmail.com)

### Resumo:

A Mineração de Dados é um processo de natureza interativa capaz de encontrar padrões em grandes conjuntos de dados, com o intuito de extrair conhecimento válido, útil e inovador. Mostra-se como importante instrumento de apoio à tomada de decisão em ambientes corporativos. Considerando a grande quantidade de informações armazenadas em bancos de dados de grandes corporações, e o investimento na descoberta de conhecimento a partir desse histórico, este artigo traz um estudo de caso da aplicação de técnicas de mineração de dados em uma base de operações de crédito. As técnicas foram empregadas com o auxílio da ferramenta Weka e os resultados demonstram seu potencial como elemento diferencial no mercado para análise e concessão de crédito.

**Palavras-chave** – classificação, descoberta do conhecimento em bases de dados, mineração de dados, risco, tomada de decisão.

## The Use of Data Mining in the optimization of Granting of Credit Analysis

### Abstract:

The data mining is a process of an interactive nature that is able to find patterns in large data sets in order to extract valid knowledge useful and innovative. It is shown as an important tool to support decision-making in corporate environments. Considering the large amount of information stored in databases of large corporations, and investment in knowledge discovery from this historical, this article presents a case of study of applying data mining techniques in a base of credit operations.

The techniques were employed with the help of the Weka tool and the results demonstrate its potential as a differential element in the market for review and credit granting.

**Key-words:** classification, knowledge discovery in databases, data mining, risks, decision-making.

## 1. Introdução

O volume de informações produzidas pelo mercado financeiro é surpreendente. Segundo a FEBRABAN (Federação Brasileira de Bancos - 2012) em abril de 2012 as operações de crédito representavam o equivalente a 49,6% do PIB (o Produto Interno Bruto registrado em 2011 foi de R\$ 4,143 trilhões conforme Instituto Brasileiro de Geografia e Estatística - IBGE), contra 43,1% em abril de 2010. No caso de Pessoa Jurídica os números demonstram um saldo registrado de operações até abril de 2012 de R\$674 bilhões, enquanto há dois anos este saldo era de R\$ 482 bilhões. Em 2010 o número de transações bancárias registradas foi cerca de 59,4 bilhões, já em 2011 esse total aumentou para 66,4 bilhões. O que representa uma média de, aproximadamente 2.106 operações por segundo, apenas no último ano. Diante do quadro, o grande desafio de instituições do mercado financeiro é tomar decisões otimizadas e efetivas, buscando maximizar os lucros perquiridos.

Algumas ferramentas e tecnologias comumente têm sido empregadas neste cenário, com o fito de auxiliar os profissionais da área, automatizando algumas atividades. O objetivo maior, é transformar dados em informações, e também em conhecimento. O conhecimento é o uso inteligente da informação, é a informação utilizada na prática (ALVARES, 2006).

Os bancos de dados relacionais permitem a extração de diversas informações usando consultas SQL (StructuredQueryLanguage ou Linguagem de Consulta Estruturada). Basta que as questões sejam definidas, sendo assim, as informações extraídas são na verdade respostas a questões existentes. A Mineração de Dados (Data Mining) consiste em descobrir informações totalmente novas tendo em mãos grandes quantidades de dados. Podendo estar na forma de Data Marts, arquivos texto (text mining), Banco de Dados Relacional entre outras. Possui uma vasta aplicação nos mais diferentes segmentos, tanto acadêmicos como corporativos, além de uma série de desafios relevantes que podem motivar trabalhos científicos. No entanto o avanço tecnológico e a oferta de ferramentas não dispensam o especialista do domínio minerado. A experiência profissional, a convivência com os processos e a leitura dos padrões descobertos são atributos que propiciam ao minerador amplas chances de sucesso no processo.

O potencial da mineração de dados pode ser aplicado em diversas áreas, inclusive no mercado financeiro.

A expansão do mercado de concessão de crédito leva as instituições financeiras a buscar garantir a competitividade dos seus produtos de financiamento (linhas de crédito), através de mecanismos de pesquisa e conhecimento, reduzindo o risco de inadimplência mas, mantendo a competitividade. Para atingir estes objetivos e atender a exigências de órgãos reguladores/fiscalizadores (Ex. Banco Central do Brasil através do Conselho Monetário Nacional), os bancos e cooperativas de concessão de crédito representam o principal setor usuário de tecnologia da informação no Brasil. De acordo com dados divulgados em maio deste ano pela Febraban, as despesas e investimentos - no ano de 2011 - das instituições financeiras em tecnologia registraram o valor de R\$ 18 bilhões, representando um crescimento de 27% em relação aos valores de 2009 (R\$ 14 bilhões). Em 2010 os investimentos foram de R\$ 16 bilhões.

Através do processo de Descoberta do Conhecimento em Bases de Dados (DCBD) a instituição pode aprender com o passado, analisando a carteira de crédito e identificando operações que se mostraram rentáveis assim como outras que geraram prejuízos.

Este ano o PlanoReal completa 18 anos, nesse período a economia brasileira mudou de maneira drástica, e as empresas tiveram que se adaptar a um novo panorama com a moeda estável, e a inflação, de certa forma, controlada. O cenário favoreceu o crescimento do mercado interno e da demanda por crédito facilitado. Este histórico de operações de crédito

serve como base para extração de informações relevantes, através dos métodos de mineração de dados.

O presente artigo tem como objetivo demonstrar de que forma sistemas de apoio à decisão dão suporte a uma instituição financeira concessora de crédito a clientes Pessoa Jurídica, cobrindo os aspectos práticos e fornecendo a base teórica para um melhor entendimento do assunto. Visando um enfoque prático e aplicado, a atividade de DCBD será descrita, proporcionando uma melhor compreensão, bem como a exploração de parte do potencial de aplicação desta área de pesquisa.

Logo, é demonstrado como criar e aplicar um classificador em uma base de dados, para caracterizar clientes que pagam em dia (bons clientes) e clientes que não pagam suas parcelas, considerando o histórico de pagamento de clientes da instituição financeira. Esta classificação fornecerá apoio à análise de novas propostas de crédito. Este artigo está organizado da seguinte forma. A Sessão 2 descreve o panorama atual do crédito no Brasil. A Sessão 3 traz uma síntese de pesquisas semelhantes embora, usando outras técnicas. A Sessão 4 apresenta o caso prático e as informações extraídas. Nas Seções 5 e 6 são apresentados os resultados e a conclusão, respectivamente.

## **2. Fundamentação teórica**

O risco de crédito foi definido pelo Bacen (Banco Central do Brasil) na Resolução 3.721/2009 em seu Art. 2º como:

(...) a possibilidade de ocorrência de perdas associadas ao não cumprimento pelo tomador ou contraparte de suas respectivas obrigações financeiras nos termos pactuados, à desvalorização do contrato de crédito decorrente da deterioração na classificação de risco do tomador, à redução de ganhos ou remunerações, às vantagens concedidas na renegociação e aos custos de recuperação.

Ou seja, a mensuração do risco de inadimplência nas operações de crédito e sua manutenção em níveis aceitáveis não são só esperadas das instituições financeiras como também exigidas, tudo visando à saúde do Sistema Financeiro Nacional. Esta mensuração é apenas uma das faces que compõe a gestão do risco de crédito.

O BACEN também determina que a estrutura de Gerenciamento do Risco de Crédito seja capaz de simular como ficaria a carteira de operações de crédito em momentos de crise (englobando ciclos econômicos, alteração das condições de mercado e de liquidez) através de testes de estresse. Os resultados destas simulações devem ser usados como apoio na revisão de procedimentos e normativas voltadas à concessão de crédito da instituição.

Dessa forma, as operações de crédito são classificadas em categorias, tendo por base: a situação econômico-financeira dos clientes, informações cadastrais atualizadas e o período de atraso no pagamento de prestações, de acordo com a Resolução nº 3.721 de 30/04/2009 do BACEN. Inclusive novas modalidades de operação (linhas de crédito) devem ser avaliadas levando em conta os aspectos do risco de crédito.

A concessão de crédito com qualidade está relacionada ao risco de inadimplência daquele cliente ou, o não cumprimento das obrigações nos termos pactuados. Segundo Sicsú (2010) é uma decisão sob condições de incerteza, quer o crédito seja solicitado ou quer seja oferecido pelo credor sempre existe a possibilidade de perda. Podendo o credor estimar a probabilidade de que essa perda ocorra, sua decisão será mais confiável.

Para Sicsú (2010) o objetivo dos modelos de credit scoring é prever, na data da decisão, a probabilidade de que o crédito, se concedido, incorra em perda para o credor. O risco de crédito corresponde à possibilidade de perda em uma operação de crédito. O credit score é então uma medida do risco de crédito. Modelos de credit scoring são formulas de cálculo dos

escores de crédito, denominação genérica dada no mercado, ainda segundo Sicsú (2010).

Consequentemente, o resultado da análise utilizando os modelos é a resposta à questão de conceder ou não o crédito, mediante a definição anterior de um nível aceitável de risco, um ponto de corte. Os sistemas de credit scoring se baseiam em informações estatísticas em cima de dados históricos que são cruzados utilizando métodos objetivos, estes ao final fornecem uma “nota” àquela operação ou ao cliente. O que reduz o aspecto subjetivo da análise de crédito e o tempo destinado a esta etapa, além de fornecer segurança e demonstrar imparcialidade ao responsável pela decisão.

Além disso, é importante a participação dos analistas de crédito na definição e na validação de um modelo de credit scoring. Também se faz necessário o conhecimento em tecnologia para dar suporte a este modelo usando o resultado das técnicas de DCBD com mineração de dados - gerando um sistema de apoio à decisão do crédito.

A Serasa Experian é um bureau de crédito muito utilizado por instituições financeiras e não-financeiras para conhecer o cliente. A empresa utiliza a solução SAS Enterprise Miner da norte-americana SAS, especializada em Business intelligence e soluções analíticas integradas. Com a ferramenta, a Serasa Experian disponibiliza às empresas informações sobre a saúde financeira e creditícia de seus clientes. A ferramenta é capaz de minerar informações a partir de dados estruturados e não estruturados. Este ano o grupo anunciou o uso das redes sociais também como fonte de pesquisa, buscando agregar ainda mais conteúdo as suas bases de dados (CIAB, 2012).

O banco Bradesco é outro caso de sucesso em análise de dados da carteira de clientes. As informações extraídas baseadas no histórico de clientes e em indicadores de inadimplência são enviadas às agências para dar suporte à decisão. A organização também usa a tecnologia para atender às regulamentações e normas mundiais e nacionais de procedimentos bancários (HAUBER, 2010).

### **3. Trabalhos correlatos**

Conseguir prever se um tomador de crédito cumprirá ou não com suas obrigações tem sido tema de diversos trabalhos, envolvendo estudos na área de mineração de dados e finanças. Tais estudos buscam identificar modelos capazes de classificar clientes como prováveis bons e maus pagadores. Para tanto, dentro do cadastro de clientes existem características comuns que, quando identificadas no conjunto de dados, auxiliam a traçar o perfil ideal.

#### **3.1 Taiwan, RC**

A pesquisa feita por Feng-Chia Li, Peng-Kai Wang e Gwo-En Wang fez uma comparação entre três classificadores muito usados em modelos de credit scoring:

- Support Vector Machine (SVM - Máquina de Vetores de Suporte) – Vapnik (1998) citado por Lima (2009, p.22) definiu como uma máquina de aprendizado estatístico linear cujo método de treinamento, baseado no princípio da minimização do risco estrutural – SRM, do inglês StructuralRiskMinimization – é capaz de encontrar um hiperplano ótimo de maximização da margem de separação entre duas classes. Ainda de acordo com Lima (2009, p.23) as máquinas de vetores de suporte são vistas como uma das mais estudadas ferramentas de aprendizado de máquina da atualidade;
- K-nearestneighbor(KNN – K-vizinhos mais próximos) – O classificador KNN primeiro produz os vetores de características e rótulos das instâncias de treinamento, e, em seguida, as mesmas características serão calculadas para a amostra de testes cujas classes não são conhecidas. Distâncias entre os pontos de dados são computadas e as amostras ‘K mais próximas’ serão selecionadas. O novo ponto de dados está previsto para ser um membro do

grupo mais numeroso;

- Extreme Learning Machine (ELM - Máquina de Aprendizado Extremo) – De acordo com os pesquisadores o modelo ELM escolhe aleatoriamente nós ocultos e determina os pesos de saída, o que proporciona velocidade na aprendizagem. Este algoritmo tem estrutura mais simples do que a maioria dos algoritmos de aprendizagem com redes neurais.

Foram utilizadas duas bases de dados distintas da Universidade da Califórnia, contendo dados de crédito de Alemães e Australianos. O estudo concluiu que o algoritmo KNN apresentou maior taxa de precisão (apesar de pequena a diferença) quando aplicado ao conjunto de dados australianos. O algoritmo SVM teve melhor taxa de precisão quando aplicado ao conjunto de dados alemães em relação aos demais. Contudo o algoritmo ELM foi o classificador mais rápido nas duas bases de dados. Os autores sugerem uma abordagem híbrida da mineração de dados na concepção de um modelo aprimorado de credit scoring.

### **3.2 Malásia**

Três métodos de mineração de dados usados no desenvolvimento de modelos de credit scoring foram alvo da pesquisa dos autores Yap Bee Wah, Irma Rohaiza Ibrahim. São eles: Modelos de Regressão Logística, de Classificação e Regressão (CART) e modelos de Rede Neural.

O objetivo foi desenvolver um modelo de credit scoring para concessão de cartão de crédito examinando a capacidade preditiva dos métodos. Os dados foram disponibilizados pelo Departamento de Cartão de Crédito do Centre Bank. Dentre o conjunto de dados existiam indicadores demográficos além do histórico de pagamentos de empréstimos anteriores.

O modelo de rede neural teve maior taxa de precisão identificada (RN=76,46%; LR=74,56%; CART=73,66%) apesar de todos os métodos terem selecionado variáveis de entrada semelhantes.

De acordo com a pesquisa, os métodos baseados em redes neurais e árvores de decisão podem modelar complexas relações não-lineares melhores que os tradicionais métodos estatísticos (análise discriminante, análise de probit). Contudo, em aplicações práticas, técnicas sofisticadas como Redes Neurais, MARS (Técnica de regressão multivariada não paramétrica com seleção de variáveis) e SVM (Máquina de Vetores de Suporte) demonstraram pequena melhoria na precisão dos resultados, quando comparadas às árvores de decisão e modelos de regressão logística (técnicas de fácil compreensão para usuários).

### **3.3 Reino Unido**

O estudo desenvolvido por Jia Wu, Sunil Vadera, Karl Dayson, Diane Burrridge e Ian Clough comparou quatro modelos de classificação de informações. O intuito era desenvolver um sistema de avaliação de risco em empréstimos para a uma empresa concessora de crédito do Reino Unido, que não visa lucro.

Os pesquisadores avaliaram o desempenho dos quatro algoritmos levando em conta aspectos técnicos e gerenciais do ambiente em que seriam aplicados. A qualidade dos dados foi levantada e juntamente a importância da fase de pré-processamento destes.

Os métodos comparados são:

- Árvore de decisão – É um método de aprendizado supervisionado e segundo Goldschmidt e Passos (2005) em geral a construção de árvores de decisão é realizada segundo uma abordagem recursiva de particionamento da base de dados. Como um arranjo de testes, que prescreve um teste adequado a cada passo em uma análise. O algoritmo mais comum de decisão é C4.5, inspirado no ID3 (Iterative Dichotomiser 3). É capaz de lidar com falta de

valores em atributos e também permite podar, fazendo um ajuste dos dados;

- Clustering – Este é um método de aprendizado não supervisionado, pois não há categoria ou classe pré-definida no conjunto de dados. O objetivo da clusterização é agrupar objetos para que, aqueles objetos de um cluster sejam mutuamente similares, mas muito diferentes de objetos em outros clusters, conforme pesquisadores acrescentam. Os algoritmos mais comuns a esse método são COBWEB e k-means;

- Bayesiano – É um método estatístico baseado no teorema de Bayes e está relacionado ao cálculo de probabilidades condicionais (GOLDSCHMIDT E PASSOS, 2005). Os dados são representados por uma rede bayesiana simples;

- FReBE – Também tem seus fundamentos em redes bayesianas, identificando exemplares deste o conceito com uma medida de semelhança de família, de acordo com os autores da pesquisa o método foi originalmente criado por Wittgenstein (1973). Usa exemplos para descrever os dados;

Os testes foram feitos em duas bases de dados distintas, a primeira com 109 registros e a segunda com 1000 registros. Os resultados na amostra menor não foram satisfatórios, pois apresentaram taxas de precisão abaixo de 40%. No entanto no conjunto de dados maior esta taxa subiu para em torno de 75% com os algoritmos de árvore de decisão, redes bayesianas e FReBE. Este com uma diferença de 4% em relação ao primeiro.

A pesquisa concluiu que o método mais viável para a instituição é o de árvores de decisão, por ser de fácil compreensão, não exigindo profundo conhecimento técnico da equipe. Foi considerada pequena a diferença da taxa de precisão em relação aos demais métodos.

#### 4. Modelo proposto

A base de dados utilizada na pesquisa consiste em operações realizadas nos últimos 04 anos. As operações estão classificadas em níveis de risco que vão de A até H (A, B, C, D, E, F, G, H). A Resolução nº 2.682 de 21 de dezembro de 1999 determina os critérios dessa classificação objetivamente, de acordo com a pontualidade dos pagamentos das parcelas (Tabela 1).

Nível	Atraso	% provisionado
AA	Em dia	0
A	Em dia	0,5%
B	15 dias < atraso < 30 dias	1%
C	31 dias < atraso < 60 dias	3%
D	61 dias < atraso < 90 dias	10%
E	91 dias < atraso < 120 dias	30%
F	121 dias < atraso < 150 dias	50%
G	151 dias < atraso < 180 dias	70%
H	Atraso acima de 180 dias	100%

Fonte: Bacen (1999)

Tabela 1 – Provisões para Créditos de Liquidações Duvidosas

Os níveis de risco determinam o percentual daquela operação que deve ser provisionado (Provisões para Créditos de Liquidações Duvidosas - PCLD). Quanto maior o atraso na operação, maior será o valor reservado pela instituição em decorrência daquele contrato. Esta reserva financeira fica estagnada, não podendo ser aplicada, nem direcionada a outros fins.

Todas as propostas de crédito recebidas passam pela Análise de Crédito baseada nos 6c's do crédito (Capacidade, caráter, condições, capital, colateral e conglomerado) e recebem uma nota, chamada de risco inicial (Tabela 2) que determinará a aceitação ou não da proposta, fazendo com que estase torne uma operação.

<b>Risco Inicial</b>	<b>Enquadramento</b>
<b>AA</b>	Baixíssimo Risco
<b>A</b>	Baixo Risco
<b>B</b>	Médio Risco

Tabela 2 – Classificação de risco usada na Análise de Crédito

A operação inicia com o risco trazido da etapa de Análise de Crédito, que representa na verdade o risco do cliente. A partir da efetiva relação de negócio (assinatura do contrato, geração de parcelas para pagamento), a operação passa a ser monitorada conforme o cumprimento das obrigações estipuladas no contrato (pagamento de parcelas), de acordo a Resolução nº 2.682 de 21 de dezembro de 1999.

O estudo foi realizado a partir de uma amostra da base de dados, com informações da carteira de crédito, alguns dados foram manipulados para preservar a identidade da instituição.

A ferramenta de mineração de dados Weka - desenvolvida em linguagem Java – foi utilizada para descobrir padrões nestas informações que sejam úteis a novas operações. É uma coleção de algoritmos muito usada para o aprendizado de tarefas de mineração de dados. Os algoritmos podem ser aplicados diretamente a um conjunto de dados ou chamados a partir do seu próprio código Java. Também contém ferramentas para pré-processamento, classificação, regressão, clustering, regras de associação e visualização (WEKA, 2012). A tarefa utilizada foi a de Classificação com o algoritmo J4.8 (implementação do algoritmo C4.5 na ferramenta).

Segundo Goldschmidt (2005), uma árvore de decisão é um modelo de conhecimento em que cada nó interno da árvore representa uma decisão sobre um atributo que determina como os dados estão particionados pelos seus nós filhos. Inicialmente a raiz da árvore contém toda a base de dados. Um predicado, denominado ponto de separação, é escolhido como sendo a condição que melhor separa ou discrimina as classes.

Os atributos preditivos considerados no estudo e que representam os nós na árvore de decisão foram:

- Tipo de crédito: Produtos disponíveis aos clientes, de acordo com o propósito do crédito, podendo ser empréstimo (capital de giro) ou financiamento (investimento em máquinas, equipamentos, obras civis, compra de matéria prima);
- Tipo de Garantia: Fidejussória (aval dos sócios e/ou de terceiros) ou Real (hipoteca e/ou penhor);
- Cidade: Filtradas aquelas com mais de 100 operações ativas;
- Enquadramento: Empresas com mais de um ano, menos de um ano e clientes preferenciais (aqueles com bom histórico de pagamentos);
- Valor Contratado: Segmentado em até 30 mil, entre 31 e 60 mil, entre 61 e 90 mil, maior que 90 mil;
- Segmento: Divididos em Comércio, Serviço, Indústria e Turismo.

O atributo classificador escolhido é o risco da operação. O nível de risco foi agrupado em Baixo ou Alto. As operações com níveis de risco AA, A, B e C foram classificadas como de Baixo risco. Os níveis AA e A não possuem atraso e são créditos rentáveis para a instituição. Contratos enquadrados no nível de risco B podem representar atraso ou não, já que é um dos níveis aceitos na Análise de Crédito. Contratos enquadrados no nível de risco C estão em atraso gerando o pagamento de encargos de mora e juros e ainda estão numa faixa de provisão aceitável (3%). Operações classificadas em D, E, F, G e H são prejudiciais à carteira de crédito e foram classificadas como risco Alto.

Antes de iniciar o processo de mineração foi preciso fazer a limpeza dos dados, eliminando

ruídos que poderiam alterar o resultado do processamento. Amostra contendo 839 registros compostos por sete atributos.

Na Figura 1 está ilustrada a Árvore de decisão criada a partir das regras encontradas pelo algoritmo J4.8 (taxa de precisão= 72,7%):

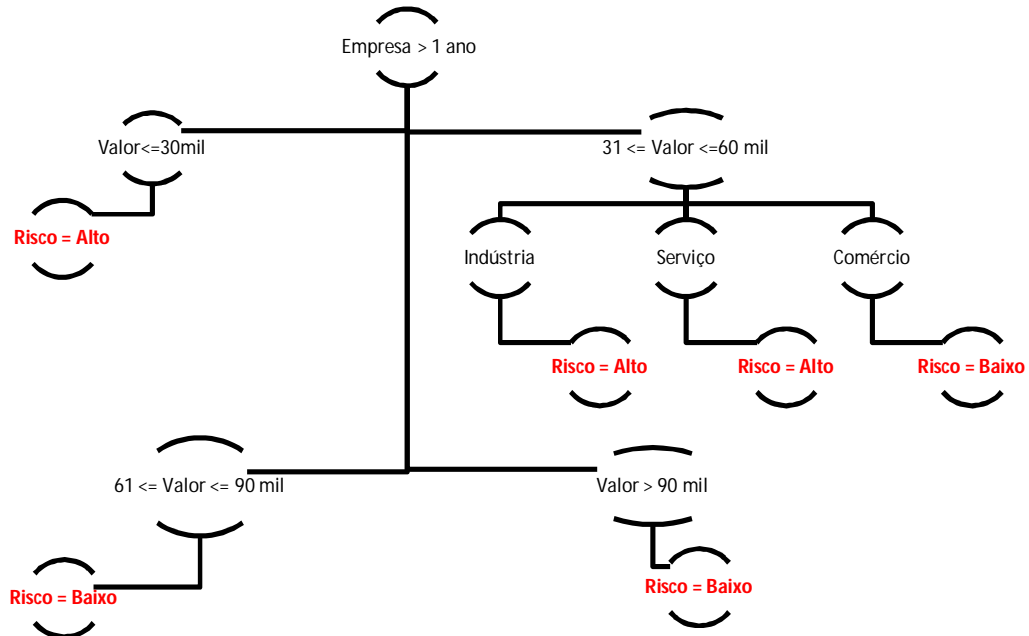


Figura 1 – Ilustração da árvore de decisão para empresas com mais de um ano de atividade

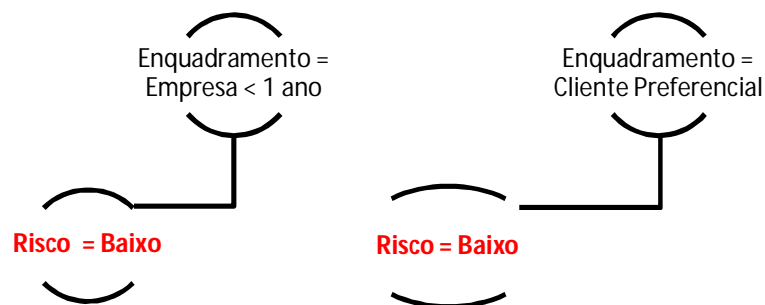


Figura 2 – Ilustração da árvore de decisão para empresas com menos de um ano de atividade e empresas enquadradas como Cliente Preferencial



O algoritmo foi aplicado novamente usando como filtro cidades com mais de 100 operações ativas:

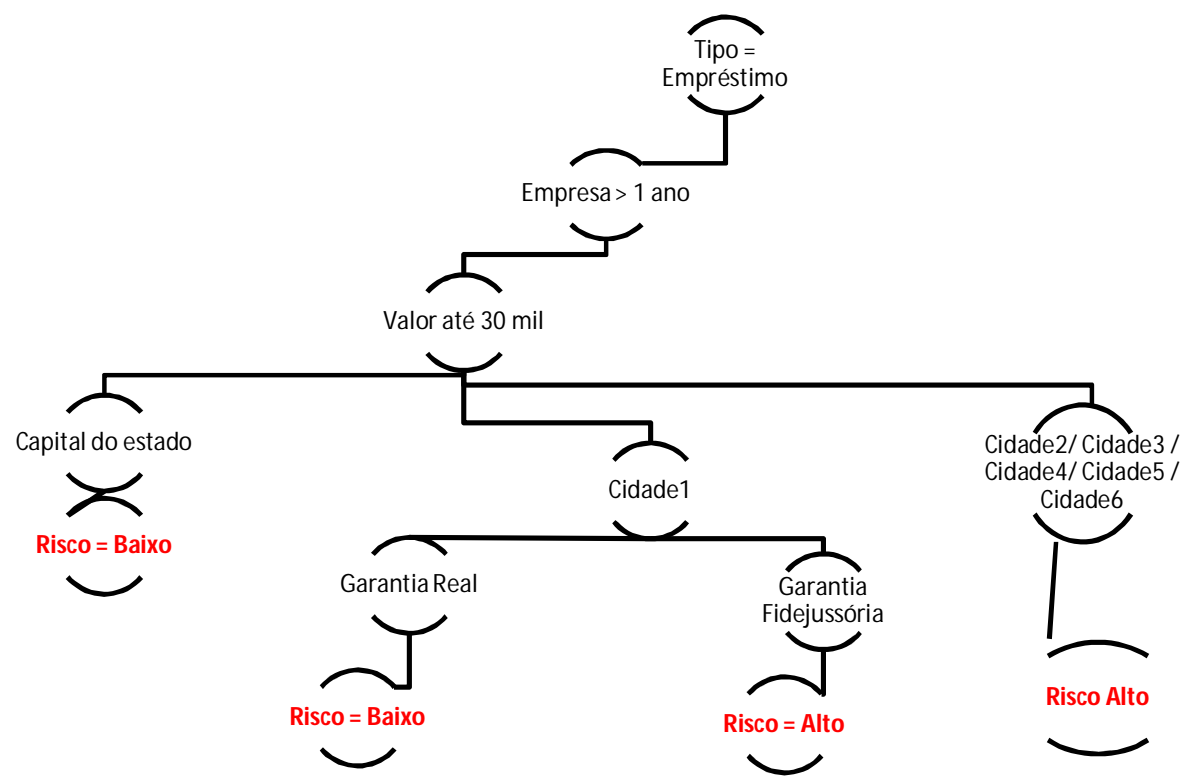


Figura 3– Ilustração da árvore de decisão para linhas de empréstimo com empresas com mais de um ano de atividade de acordo com a cidade

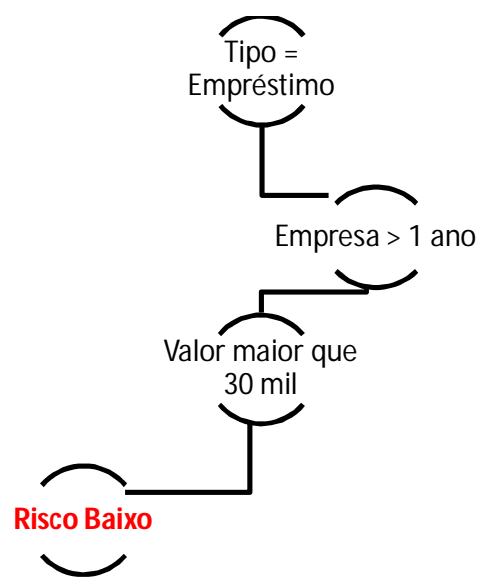


Figura 4– Ilustração da árvore de decisão para linhas de empréstimo com empresas com mais de um ano de atividade e valor contratado maior que 30 mil

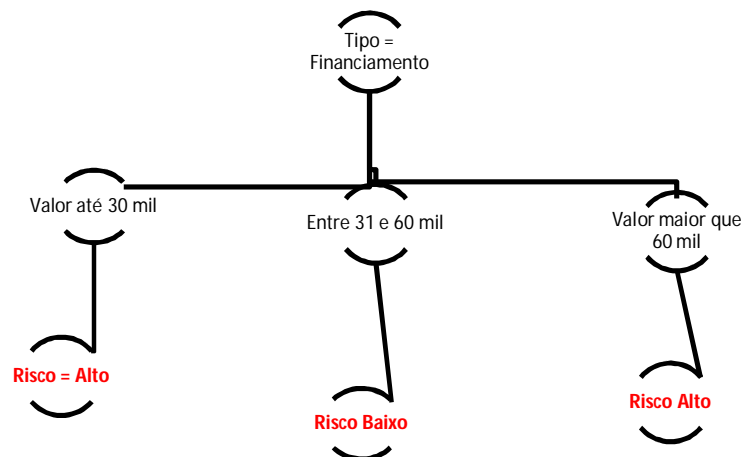


Figura 5 - Ilustração da árvore de decisão para linhas de financiamento levando em consideração o valor contratado

## 5. Análise dos resultados

Os resultados demonstraram que, analisando a carteira de forma ampla, há uma faixa de baixo risco em contratos com valores superiores a 60 mil em empresas com mais de um ano de atividade. Podem ser desenvolvidos produtos voltados especificamente para estes clientes.

Empresas com menos de um ano foram classificadas com risco baixo. Atualmente a instituição tem rígidas exigências de garantias e políticas cautelosas ao conceder crédito a empresas com essa característica, devido à instabilidade apresentada no primeiro ano de atividade. Em contrapartida, foi criado o enquadramento Cliente Preferencial voltado para empresas que possuem bom histórico de operações. Estes clientes têm taxas de juros mais baixas, prazos mais longos e limites de crédito diferenciados dos demais. A árvore demonstrou como a experiência de fidelizar clientes tem sido bem sucedida.

Focando nas cidades foi observada a necessidade de intensificar as operações de cobrança no interior do estado e aumentar a rigidez da análise de crédito. Também se torna necessário expandir a carteira para outros municípios, pois há uma concentração em apenas sete municípios de um total de 246 existentes no estado. A inadimplência registrada fora da capital acaba por contaminar a carteira como um todo.

Além disso, com a análise das informações a instituição percebeu a necessidade de diversificar o crédito, promovendo ações de incentivo às linhas de investimento. Estas representam um risco mais baixo de desvio da finalidade, em relação ao capital de giro (empréstimo). Os setores de cobrança e análise de crédito passam a ter informações importantes sobre determinados setores do mercado e regiões do estado, podendo otimizar o uso de recursos em seus processos.

## 6. Conclusão

A Mineração de Dados surgiu com o principal objetivo de dar suporte à tomada de decisões. É necessário que o usuário de um sistema de descoberta tenha um grande conhecimento do negócio em que está inserido assim como da própria base de dados que está lidando, para ser capaz de selecionar corretamente os conjuntos de dados e as classes de padrões relevantes.

Para a aplicação desta técnica de mineração foi exigido o estudo da fundamentação teórica

(envolvendo conceitos, técnicas e algoritmos) aliada à experiência na vivência prática do negócio utilizado, apesar disso este artigo traz apenas uma amostra de como transformar as informações extraídas em conhecimento. Todo o processo de Descoberta é complexo, envolvendo a execução de tarefas, configuração de parâmetros e grande interação com o usuário. No entanto, o sucesso do procedimento pode trazer uma recompensa de grande valia para as organizações. A inclusão de mais variáveis ao modelo e a comparação dos resultados fica como um potencial trabalho futuro.

## Referências

ÁLVARES, R. V. **Mineração de Dados: Introdução e Aplicações**. SQL Magazine, São Paulo: Editora DevMedia, p. 30-36, Abril, 2006.

BACEN. **Relatório de Inclusão financeira**. 2010, ano 1. [http://www.bcb.gov.br/Nor/relincfin/relatorio\\_inclusao\\_financeira.pdf](http://www.bcb.gov.br/Nor/relincfin/relatorio_inclusao_financeira.pdf). Acesso em junho de 2012.

BRASIL, Banco Central do. **Resolução nº.2.682, 21 de dezembro de 1999**. Disponível em: [http://www.bcb.gov.br/pre/normativos/res/1999/pdf/res\\_2682.pdf](http://www.bcb.gov.br/pre/normativos/res/1999/pdf/res_2682.pdf). Acesso em junho de 2012.

BRASIL, Banco Central do. **Resolução nº.3.721, 30 de abril de 2009**. Disponível em: [http://www.bcb.gov.br/pre/normativos/res/2009/pdf/res\\_3721.pdf](http://www.bcb.gov.br/pre/normativos/res/2009/pdf/res_3721.pdf). Acesso em junho de 2012.

CARVALHO, L.A.V.. **A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração**. 1edition, 2005.

FAYAD, U, M. ET AL. **From data mining to knowledge Discovery and data mining**. Menlo Park, California, 1996.

FEBRABAN. **Pesquisa CIAB FEBRABAN 2012: O setor Bancário em números**. Congresso e Exposição de Tecnologia da Informação das Instituições Financeiras. São Paulo. Disponível em: [http://www.ciab.com.br/\\_pdfs/publicacoes/Pesquisa2012.pdf](http://www.ciab.com.br/_pdfs/publicacoes/Pesquisa2012.pdf). Acesso em junho de 2012.

GOLDSCHMIDT, Ronaldo; PASSOS Emmanuel. **Data Mining Um guia prático**. Rio de Janeiro: Editora Campus, 2005.

GONÇALVES, Eduardo Corrêa. **Extração de Árvores de Decisão com a Ferramenta de Data Mining Weka**. Disponível em: <http://www.devmedia.com.br/extracao-de-arvores-de-decisao-com-a-ferramenta-de-data-mining-weka/3388>. Acesso em março de 2012.

HAUBER, Brad. **Wearing the Customer's Shoes: Brazil's Banco Bradesco gains perspective and profitability through CRM**. Teradata magazine online. Q4; 2010. Disponível em: <http://www.teradatamagazine.com/v10n04/Features/Wearing-the-Customers-Shoes/>. Acesso em junho de 2012.

IBGE. **Sala de Imprensa : Contas Nacionais Trimestrais-Indicadores de Volume e Valores Correntes**. Disponível em: [http://www.ibge.gov.br/home/presidencia/noticias/noticia\\_visualiza.php?id\\_noticia=2093](http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=2093). Acesso em junho de 2012.

LI, Feng-Chia; WANG, Peng-Kai; WAN, Gwo-En. **Comparison of the Primitive Classifiers with Extreme Learning Machine in Credit Scoring**. Department of Information Management, Jen-Teh Junior College, Miaoli city, Hwa Hsia College, Taipei city, Taiwan. Inglês. Disponível em: <http://ieeexplore.ieee.org/Xplore/guesthome.jsp>. Acesso em: Junho de 2012.

LIMA, Naiyan Hari Cândido. **Comitê de máquinas SVM utilizando Reforço adaptativo**. Disponível em: <http://www.engcomp.ufm.br/publicacoes/ECC-2009-1-7.pdf>. Acesso em junho de 2012.

MALHOTRA, Naresh K. **Pesquisa de mercado: uma orientação aplicada**. Trad. de Nivaldo Montingelli Jr. e Alfredo Alves de Farias. Porto Alegre: Ed. Bookman, 2001.

MARQUES, LFB. **Gerenciamento do Risco de Crédito: cálculo do risco de crédito para a carteira de um banco de varejo**. PPGA – UFRGS, 2002. Dissertação disponível em: <http://hdl.handle.net/10183/2031>. Acesso em Março de 2012.

SANTOS, R. - **Weka na Munheca - Um guia para uso do Weka em scripts e integração com aplicações em Java**. Disponível em: <http://www.lac.inpe.br/~rafael.santos/Docs/CAP359/2005/weka.pdf>. Acesso em março de 2012.

SANTOS, R.S. - **Aplicação de um modelo preditivo de mineração de dados para apoio à decisão de crédito.** UFMG, 2006. Dissertação disponível em: [http://www.bibliotecadigital.ufmg.br/dspace/bitstream/1843/VALA-6T7R6V/1/mestrado\\_roberto\\_de\\_souza\\_santos.pdf](http://www.bibliotecadigital.ufmg.br/dspace/bitstream/1843/VALA-6T7R6V/1/mestrado_roberto_de_souza_santos.pdf). Acesso em Março de 2012.

SARDENBERG, Rubens. **Panorama do Mercado de Crédito – FEBRABAN.** 2012. Disponível em: [http://www.febraban.org.br/acervo.asp?id\\_pagina=193&id\\_paginaDe=78&id\\_texto=0](http://www.febraban.org.br/acervo.asp?id_pagina=193&id_paginaDe=78&id_texto=0). Acesso em junho de 2012.

SCHRINCKEL, W. **Análise de Crédito: concessão e gerência de empréstimos.** São Paulo. Ed. Atlas, 1997.

SICSÚ, Abraham Laredo. **Credit Scoring Desenvolvimento – Implantação - Acompanhamento.** Editora: Edgar Blucher, 2010.

WAH, Yap Bee; IBRAHIM, Irma Rohaiza. **Using Data Mining Predictive Models to Classify Credit Card Applicants.** Faculty of Computer and Mathematical Sciences; Universiti Teknologi MARA; Selangor, Malaysia. Inglês. Disponível em: <http://ieeexplore.ieee.org/Xplore/guesthome.jsp>. Acesso em junho de 2012.

Weka. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>. Acesso em março de 2012.

WU, Jia; VADERA, Sunil; DAYSON, Karl; BURRIDGE, Diane; CLOUGH, Ian. **A Comparison of Data Mining Methods in Microfinance.** University of Salford; East Lans Moneyline; United Kingdom. Inglês. Disponível em: <http://ieeexplore.ieee.org/Xplore/guesthome.jsp>. Acesso em junho de 2012.