

Modelos de Análise de Crédito utilizando Técnicas de Aprendizado de Máquina

Oswaldo Luiz Humbert Fonseca

Universidade do Estado do Rio de Janeiro – UERJ
Rua Alberto Rangel, s/n, Vila Nova CEP 28630-050 Nova Friburgo, RJ
Oswaldo@bndes.gov.br

Francisco Duarte Mora Neto

Universidade do Estado do Rio de Janeiro – UERJ
Rua Alberto Rangel, s/n, Vila Nova CEP 28630-050 Nova Friburgo, RJ
fmoura@iprj.uerj.br

Flávio Joaquim de Souza

Universidade do Estado do Rio de Janeiro – UERJ
Av. São Francisco Xavier, 524 5. andar, Maracanã, Rio de Janeiro, RJ
fjsouza@eng.uerj.br

1. Introdução

Este trabalho se propõe a elaborar modelos de análise de crédito para micro, pequenas e médias empresas, utilizando técnicas de aprendizado de máquina, mais especificamente, serão analisados neste trabalho, modelos que utilizam a o método dos vizinhos mais próximos, as árvores de decisão, e um terceiro que mistura as técnicas de redes neurais com lógica *Fuzzy* (*Neuro-Fuzzy*).

O aprendizado de máquina é uma área da inteligência artificial cujo objetivo é o desenvolvimento de técnicas computacionais e de construções de sistemas capazes de adquirir conhecimento de forma automática e utilizável. A idéia básica é a poder obter conclusões genéricas gerais a partir de um conjunto particular de exemplos.

Neste trabalho foram selecionadas trezentas e dezoito solicitações de crédito retiradas aleatoriamente dos pedidos feitos ao Banco Nacional de Desenvolvimento Econômico e Social (BNDES). Estas solicitações dos cartões BNDES [1] foram analisadas por especialistas em crédito, dos bancos emissores, que decidiram por conceder ou não o crédito solicitado. Em duzentos e dez casos a houve a concessão, enquanto em cento e oito o crédito foi negado.

O trabalho está organizado em seis seções incluindo esta introdução. A seção 2 apresenta os conceitos básicos da análise de risco de crédito, a terceira os conceitos de aprendizado de máquina e na quarta o método utilizados são descritos. Um estudo de caso, com dados reais, é apresentado na quinta seção, enquanto a sexta seção mostra os resultados e conclusões do trabalho.

2. Análise do Risco de Crédito.

2.1. Risco de Crédito

No dicionário *Webster*, o termo risco é

definido da seguinte forma: “*um perigo; exposição à perda ou ao dano*”.

No sentido financeiro o termo risco assume dois sentidos. O primeiro é a possibilidade de prejuízo financeiro; e o segundo é a dispersão de resultados de um ativo em que pode ser associado numericamente com o desvio-padrão ou com a variância [2].

O risco nas operações de crédito existe porque o fluxo de caixa prometido pode ou não ser pago na sua totalidade pelos diversos tomadores. Se a instituição financeira recebesse tudo, obviamente não haveria risco de crédito.

Como os riscos de crédito são, em tese, não elimináveis, a incumbência das instituições é de estimar o risco de perda esperada e exigir prêmios pelo risco.

Risco de crédito significa o risco de perda em empréstimos ou em investimentos das mais variadas formas. Para melhor entender o risco de crédito, é necessário verificar o processo decisório, que incorpora a obtenção de um grande número de informações dos possíveis clientes.

As informações podem ser obtidas através de documentação oficial (demonstrativos contábeis, estatutos, cadastros, alterações contratuais, planejamento dos negócios etc.); através de visitas às empresas (aspectos administrativos, operacionais, tecnológicos, mercadológicos e estratégicos); e através de informações obtidas junto a terceiros (fornecedores, clientes, agências de informações, e outros bancos), [3].

Essas informações são processadas na etapa que se denomina análise de crédito. É onde, sob a macro-orientação da política de crédito da instituição, utilizam-se as mais diversas técnicas no sentido de se estabelecer o risco de crédito que a instituição estaria assumindo em negócios que viesse a realizar com o cliente em estudo.

Na etapa de decisão de crédito e, frente à possibilidade de uma estruturação de empréstimo com o cliente (montante, prazo, taxa, garantias e produto), toma-se uma decisão de negócios, que será sistematicamente utilizada como fator de realimentação da própria política de crédito da instituição.

A determinação do risco de crédito é realizada na etapa “análise de crédito” onde se procura determinar o risco de crédito do cliente. Segundo Schrickel [4],

“A análise de crédito envolve a habilidade de fazer uma decisão de crédito dentro de um cenário de incertezas e constantes mutações e transformações incompletas. Esta habilidade depende da capacidade de analisar logicamente situações, não raro, complexas, e chegar a uma conclusão clara, prática e factível, de ser implementada”.

Na sequência de análise que se utiliza para determinar o risco de crédito, toma-se como base as metodologias tradicionais, e, normalmente, levam-se em conta cinco aspectos do crédito, conhecidos como os cinco C’s do crédito.

2.2. Os cinco C’s do Crédito

Os bancos procuram distinguir entre os que provavelmente pagarão e os que não vão honrar os seus compromissos e, para isto, se utilizam de metodologias de análise de crédito. Os cinco “C’s” do crédito descritos são: caráter, capacidade, capital, condições e colateral. A análise destas características embasa metodologias utilizadas no processo de avaliação de crédito em todo o mundo.

O **CARÁTER** do tomador indica uma característica essencialmente subjetiva que trata da vontade do tomador do empréstimo de pagar suas contas, referindo-se à sua índole, ética e senso moral. [2]. Uma forma de um analista apurar estas características é obter informações junto a bancos e cartórios de títulos e protestos e a partir destas informações tirar conclusões sobre a conduta do cliente no que diz respeito à pontualidade e constância com que tem liquidado seus títulos e obrigações, ou seja, buscar dados históricos para verificar como o tomador de crédito honrou seus compromissos anteriores, constatando se tem restrições junto aos órgãos de proteção ao crédito e se foi pontual no pagamento de possíveis créditos anteriores.

Porém estas informações apenas servem como parâmetro de comportamento histórico, pois a pontualidade da liquidação dos compromissos pode ser fator importante no negócio do cliente, não exprimindo sua real vontade de pagar. [5].

A fidelidade do pretendente é outro fator, relativo ao caráter, a ser considerado para a concessão de crédito. Fidelidade é o compromisso do cliente em permanecer como parceiro da instituição que lhe prestou um atendimento para

suprir suas necessidades de crédito. Ela está diretamente relacionada com o tempo e com o nível de realização de negócios por parte do cliente.

A **CAPACIDADE** reflete a eficácia e eficiência de gestão dos administradores e o próprio grau de especialização da produção e comercialização da empresa. A capacidade de se honrar o pagamento de um determinado empréstimo pode ser analisada sob vários aspectos de forma a se mensurar as habilidades do devedor de pagar suas dívidas, no âmbito interno de suas atividades, [5].

A capacidade refere-se à competência empresarial do cliente e alguns estudos indicam que empresas mais tradicionais têm quadros gestores mais experientes no mercado e no setor em que atuam, ou seja, a idade da empresa pode, certamente, indicar um maior nível de experiência. Usualmente, o sucesso do negócio também está relacionado com o nível de escolaridade do empresário e dos demais membros da empresa.

Outras formas de análise de capacidade poderão também ser usadas, como, por exemplo, uma investigação, sobre aspectos como: as instalações da empresa, a estrutura organizacional (que pode revelar o grau de modernização de seus produtos e serviços); o sistema de informações gerenciais que pode prestar informações cada vez mais rápidas e seguras de forma a manter a competitividade da empresa; o sistema de marketing da empresa, (para ganhar e manter mercados e garantir um faturamento condizente com sua estrutura), assim como o grau de investimento em pesquisas para desenvolvimento de novos produtos e manutenção da lucratividade, [2].

Os aspectos relativos às **CONDIÇÕES** estão ligados aos fatores externos à empresa que podem de alguma maneira afetar seu desempenho nos negócios, prejudicando ao mesmo tempo sua capacidade de honrar os compromissos. A análise a ser feita considerando o ‘C’ de condições é a observação das variáveis micro e macro econômicas externas à atividade do tomador do empréstimo.

A identificação de tendências setoriais, crescimento e recessão de mercados relacionados com o ramo de atividade da empresa, sua dependência do governo, do mercado externo, informações sobre concorrência do setor e políticas econômicas que possam vir a alterar as condições de comercialização de produtos relacionados com sua operacionalização, são alguns fatores que auxiliam a identificação das condições que podem ser favoráveis ou não ao retorno do crédito concedido, [5].

Segundo Silva [2], um outro fator pode ser importante é a sazonalidade do produto, uma vez que existem produtos que apenas são consumidos em uma determinada época do ano como é o caso,

por exemplo, do sorvete.

O **CAPITAL**, considerado como um dos C's do crédito, refere-se ao estudo do patrimônio, da solidez da empresa, ou à estrutura de composição da mesma, no sentido de ter recursos próprios que aplicados na atividade produtiva, geram resultados que permitem arcar com o ônus dos créditos conseguidos junto a terceiros. Assim, para uma análise de crédito, é importante verificar o montante de capital próprio que é empregado em uma empresa e também sua estrutura, comparativamente, ao capital de terceiros, que deve ser capaz de gerar receita que permita saldar os empréstimos realizados, [5].

O 'C' de capital, no âmbito da análise do crédito, procura identificar se o tomador do crédito, pela quantidade e forma de aplicação de recursos na empresa, tem condições financeiras de saldar seus empréstimos no prazo combinado.

COLATERAL significa garantia. Esse 'C' do crédito deve ser, como os demais, sempre analisado no conjunto da situação em que deve ocorrer o crédito. Procura averiguar a capacidade do cliente em oferecer garantias colaterais, espécie de segurança adicional necessária à concessão do crédito que atenua o risco da operação, [5].

As garantias são divididas em reais e pessoais. As garantias reais são o direito especial de garantia sobre móveis ou imóveis, além da promessa de contra-prestação, tais como Alienação Fiduciária, Hipoteca, Penhor Mercantil, Caução, Ações, Cédula Hipotecária, Certificado de Depósito, Debêntures, Duplicatas, Letras de Câmbio, Letra Hipotecária, Nota Promissória, Título de Dívida etc. As garantias pessoais se resumem à promessa de contra-prestação. Neste caso, o credor se contenta com a garantia comum representada pelo patrimônio presente e futuro do devedor, avalista ou fiador. Como exemplos de garantias pessoais, tem-se: aval, carta de crédito e carta de fiança, [2].

É natural que, se o pretendente ao crédito, não se mostrar, por exemplo, uma pessoa idônea, mesmo oferecendo suficientes garantias, o risco desse empréstimo não voltar no prazo combinado será muito grande. A garantia nunca deve ser o motivo para se efetuar o crédito. Somente ela não faz com que o crédito retorne no prazo combinado. Ações judiciais para cobrança do crédito inadimplente podem ser muito demoradas e na maioria das vezes serão questionadas muitas cláusulas do acordo previamente firmado entre as partes. A intenção de qualquer instituição de crédito é o retorno dos empréstimos concedidos e que seja no prazo combinado para que possam ser realizados novos negócios com rendimento melhor que pendências nos tribunais de cobrança de devedores inadimplentes, que podem se arrastar por anos sem solução.

3. Aprendizado de Máquina e Classificadores

O Aprendizado de máquina, como foi dito na introdução, é uma área da Inteligência artificial que procura tirar conclusões a partir sobre exemplos fornecidos por um processo externo a aprendizagem. O aprendizado pode ser dividido em supervisionado e não supervisionado. No aprendizado supervisionado é fornecido ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido. [12].

Já no aprendizado não supervisionado, o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou clusters. Após a determinação dos agrupamentos formados, normalmente, é necessária uma análise para determinar o que cada agrupamento significa no contexto do problema que está sendo analisado [12].

Nos métodos utilizados neste estudo, aprendizagem supervisionada, cada exemplo, é descrito por um vetor de valores de características, ou atributos, e o rótulo da classe associada. O objetivo do algoritmo de indução é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados, ou seja, exemplos que não tenham o rótulo da classe.

Classificador é uma função que possui como entrada vetores de características, e como saída, rótulos que identificam a classe que tais vetores provavelmente pertencem [12].

Em aprendizado de máquina existem muitos algoritmos de aprendizado que induzem classificadores. Este trabalho concentra-se em três métodos que são descritos no item a seguir.

4. Métodos Utilizados

4.1. Método dos vizinhos mais próximos

Trata-se de um método de classificação que não possui processamento na fase de treinamento. Para cada padrão de teste, é medida a distância entre ele e todos os padrões de treinamento. Verifica-se a quais classes pertencem os "K" padrões mais próximos e a classificação é feita associando-se o padrão de teste à classe que for predominante.. A principal desvantagem está na complexidade do teste, visto que, para cada padrão de teste, além de ser necessário realizar N medições de distância, sendo N o número de padrões de treinamento, é necessário ordenar os resultados.

4.2. Árvore de Decisão

Uma árvore de decisão consiste em uma estrutura de dados em que:

- um nó folha, corresponde a uma classe, ou

- um nó de decisão que contém um teste sobre algum atributo e para cada resultado deste teste há uma aresta para uma subárvore. Cada subárvore terá a mesma estrutura da árvore. [12].

Para se construir uma árvore de decisão a partir de um conjunto de treinamento T , deve-se primeiramente adotar as classes denotadas de $\{C_1, C_2, \dots, C_k\}$. A partir daí, as etapas seguintes são:

- Se T contém um ou mais exemplos e todos pertencem à classe C_j , logo a árvore de decisão para T é um nó folha identificando a classe C_j ;
- Se T não contém exemplos, então a árvore é uma folha, porém a classe associada à esta folha deve ser determinada a partir de informação além de T ;
- Se T contém exemplos que pertencem à várias classes, então refina-se T em subconjuntos de exemplos que são conjuntos de exemplos pertencentes a uma única classe;
- Deve-se promover a poda da árvore de decisão, que consiste em solucionar um superajuste de dados. Após a construção de uma árvore de decisão pode ocorrer do classificador induzido ser muito específico para o conjunto de treinamento, isto quer dizer que o classificador superajustou os dados de treinamento, ou ocorreu um *overfitting*. Neste caso promove-se a poda da árvore de decisão após induzi-la, processo denominado de pós-poda, processo no qual reduz-se o número de nós internos, reduzindo a complexidade da árvore, provocando um desempenho melhor que da árvore original. É possível também realizar a pré-poda na árvore de decisão, que se realiza enquanto a árvore é induzida.

Uma árvore de decisão, depois de ser construída, pode ser utilizada com a finalidade de classificar novos exemplos. Inicia-se, então pela raiz e caminha através de cada nó de decisão até encontrar uma folha. Feito isso, a classe do novo exemplo é dada pela classe daquela folha.

4.3 Modelos Neuro-Fuzzy Hierárquicos

Sistemas neuro-fuzzy (SNF) [6,8,10] são sistemas híbridos pois utilizam mais de uma técnica de identificação de sistemas para a solução de um problema de modelagem. Essa mistura de técnicas reflete na obtenção de um sistema mais poderoso (em termos de interpretação, de aprendizado, de estimativa de parâmetros, de generalização, etc). Os sistemas neuro-fuzzy combinam a capacidade de

aprendizado das redes neurais artificiais (RNAs), [7], com o poder de interpretação lingüístico dos sistemas de inferência fuzzy (SIFs), [9].

A idéia básica de um sistema neuro-fuzzy é implementar um sistema de inferência fuzzy numa arquitetura paralela distribuída de tal forma que os paradigmas de aprendizado comuns às redes neurais possam ser aproveitados nessa arquitetura híbrida.

Os sistemas neuro-fuzzy e os sistemas fuzzy realizam, internamente, um mapeamento entre regiões do espaço de entrada em regiões fuzzy do espaço de saída. Este mapeamento é feito através de regras fuzzy do sistema. As variáveis de entrada e saída dos sistemas neuro-fuzzy e dos sistemas fuzzy são divididas em vários termos lingüísticos (por exemplo: baixo, alto) que são utilizados pelas regras fuzzy.

O particionamento do espaço de entrada indica a forma como as regras fuzzy estão relacionadas no espaço. O particionamento mais comum é o em grade que, apesar de ser o mais simples, possui restrições com relação ao número possível de variáveis de entrada. O sistema Neuro-Fuzzy Hierárquico Binário (NFHB) [10] utiliza um particionamento recursivo, denominado BSP (Binary Space Partitioning), com o objetivo de reduzir essa limitação, além de possuir capacidade ilimitada de criar e expandir sua estrutura.

O particionamento BSP é flexível e minimiza o problema do crescimento exponencial do número de regras, pois só cria novas regras localmente, conforme o conjunto de treinamento. Sua principal vantagem é permitir a criação de sistemas que constroem sua própria estrutura de forma automática. Este tipo de particionamento é considerado recursivo pois emprega processos recursivos em sua geração, o que acarreta a criação de modelos com hierarquia na estrutura e, conseqüentemente, hierarquia nas regras. O sistema NFHB é constituído de células que produzem esta forma de particionar espaços.

O sistema NFHB possui um algoritmo de treinamento, baseado no método do gradiente descendente [7,10], para o aprendizado da estrutura do modelo e, conseqüentemente, das regras lingüísticas. Os parâmetros que definem os perfis das funções de pertinência dos antecedentes e conseqüentes são encarados como os pesos fuzzy do sistema neuro-fuzzy.

Para limitar o crescimento indefinido da estrutura do sistema, criou-se um parâmetro de aprendizado denominado taxa de decomposição (δ). Este parâmetro é adimensional e atua impedindo que o processo de decomposição seja realizado indefinidamente. Maiores detalhes sobre o funcionamento deste algoritmo podem ser encontrados em [10-12].

5. Estudo de Caso.

Neste trabalho foram selecionadas trezentas e dezoito solicitações de crédito retiradas aleatoriamente dos pedidos feitos ao BNDES. Estas solicitações foram analisadas por especialistas em crédito, dos bancos emissores, que decidiram por conceder ou não o crédito solicitado. Em duzentas e dez casos a houve a concessão, enquanto em cento e oito o crédito foi negado.

Um total de onze variáveis foi utilizado neste trabalho para caracterizar a empresa que fez a solicitação do crédito. Estes atributos foram agrupados de acordo com os C's da análise de crédito e são apresentadas abaixo:

Caráter

- Existência de pendências financeiras
- Existência de Protestos ou ações judiciais
- Presença no cadastro de cheque sem fundos
- Existência de restrições com os sócios

Capacidade

- Tempo de atividade da empresa

Condições

- Setor de Atividade
- Unidade da Federação
- Código de endereçamento postal
- Natureza jurídica
- Banco Emissor

Capital

- Faturamento Bruto Anual

Com isso, o movimento selecionado a partir dos pedidos, formou uma matriz de trezentos e dezoito linhas e doze colunas, sendo que as onze primeiras são as variáveis e a décima segunda coluna representa a opinião do Banco emissor quanto à concessão ou não do crédito.

Os dados dessa matriz foram separados em seis conjuntos de dados, cada um contendo trinta e cinco casos onde o crédito foi concedido e dezoito onde o crédito foi negado, totalizando 53 empresas em cada um dos seis conjuntos de dados.

Para se formar o conjunto de treinamento foram selecionados cinco conjuntos de cinquenta e três empresas, totalizando duzentos e sessenta e cinco exemplos, e este conjunto serviu como massa de dados para a realização do aprendizado do indutor.

O sexto conjunto, que não participou da fase de treinamento, foi utilizado como conjunto de teste, para verificar se o programa estaria classificando corretamente os exemplares não utilizados no processo de aprendizagem. Ao todo foram criados, então, seis conjuntos de treinamento assim nomeados: A1, B1, C1, D1, E1, F1 e seis conjuntos de teste: a2, b2, c2, d2, e2, f2. Foram feitas seis execuções para cada um dos três métodos

analisados: A Árvore de Decisão; O Método dos K vizinhos mais próximos (K-NN); O Sistema Neuro-Fuzzy Hierárquico Binário. Nos resultados das 18 execuções os indutores acertaram todos os 53 exemplos apresentados.

Um outro produto do modelo é o fato deste apontar as variáveis mais relevantes para o resultado ou a análise, propriamente dita. O sistema NFHB indicou a seguinte sequência de variáveis de entrada mais relevantes: Existência de Protestos ou Ações judiciais; Unidade da Federação; Código de Endereçamento Postal; Existência de pendências com os sócios; Presença do Cadastro de Cheques sem Fundos; Natureza Jurídica; Banco Emissor; Faturamento e Tempo de atividade da empresa.

Este fato destaca a importância da localização geográfica da empresa na decisão da concessão. Um outro fato, que também merece destaque, é o de que apesar de nenhuma empresa com pendências financeiras ter o seu crédito aprovado, o indutor não usou esta informação para chegar às mesmas conclusões do analista do banco emissor.

6. Conclusões

Os resultados alcançados, pelo classificador, foram considerados plenamente satisfatórios pelos especialistas, pois na totalidade dos casos que foram apresentados para serem avaliados pelo indutor, o modelo apresentou a mesma opinião que o banco emissor havia dado.

Tanto o modelo NFHB quanto o produzido com a técnicas de árvore de decisão apontam as variáveis mais relevantes para o resultado ou a análise, propriamente dita..

Ambos os modelos destacaram a importância da localização geográfica da empresa na decisão da concessão.

Um outro fato, que também merece destaque, é o de que apesar de nenhuma empresa com pendências financeiras ter o seu crédito aprovado, os indutores não usaram esta informação para chegar às mesmas conclusões do analista de crédito

A análise automática do crédito é uma tarefa de fundamental importância para solução do problema do relacionamento entre o BNDES e o micro, pequeno e médio empresário. O trabalho apresentado mostra que os métodos mostraram-se bastante promissor na solução deste problema.

Referências

- [1] Portal cartão BNDES. Virtual www.cartaobndes.gov.br (2005).
- [2] Silva, J. A. Análise do Estabelecimento de Limite de Crédito – Um estudo de Caso. Universidade de Taubaté, São Paulo (2002).
- [3] Bueno, V. F. F. Avaliação de Risco na Concessão de Crédito Bancário para Micro e

- Pequenas Empresas. Universidade Federal de Santa Catarina, Santa Catarina (2003)
- [4] Schrickel, W K. Análise de Crédito, Concessão e Gerência de Empréstimos. São Paulo, Atlas (1998)
 - [5] Guimarães, J. B. Financiamento de Micros e Pequenas Empresas em uma Instituição Pública de Crédito, PUC/MG, . Minas Gerais, (2002).
 - [6] Halgamuge, S. K., Glesner, M.: Neural Networks in Designing Fuzzy Systems for Real World Applications. Fuzzy Sets and Systems N0.65, pp. 1-12. (1994).
 - [7] Haykin, S.: Neural Networks – A Comprehensive Foundation. Macmillan College Publishing Company, Inc.(1994).
 - [8] Kruse, R., Nauck, D.: NEFFCLASS – A Neuro-Fuzzy Approach for the Classification of Data. Proc. Of the 1995 ACM Symposium on Applied Computing, Nashville. (1995).
 - [9] Mendel, J.: Fuzzy Logic Systems for Engineering: A Tutorial. Proceedings of the IEEE, Vol.83, n.3, pp.345-377. (1995).
 - [10] Souza, Flávio Joaquim de: Modelos Neuro-Fuzzy Hierárquicos. Tese de Doutorado. Puc-Rio (1999).
 - [11] Souza, F.J., Neto, L.S. , Vellasco, M.M.B.R.: Electric Load Forecasting: Evaluating the Novel Neuro-Fuzzy BSP Model. Electrical Power Energy Systems. Vol 26, nº 2 . pp 131-142, Ed. Elsevier. ISSN 0142-0615. Holland. (2004).
 - [12] REZENDE, Solange Oliveira, et al. Sistemas Inteligentes. Fundamentos e Aplicações. São Paulo, 2003: Editora Manole Ltda.