UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO
Sistemas de Informação e Apoio à Decisão

Lisiane Priscila Roldão Selau

MODELAGEM PARA CONCESSÃO DE CRÉDITO A PESSOAS FÍSICAS EM EMPRESAS COMERCIAIS:
DA DECISÃO BINÁRIA PARA A DECISÃO MONETÁRIA

Porto Alegre
2012

Lisiane Priscila Roldão Selau

MODELAGEM PARA CONCESSÃO DE CRÉDITO A PESSOAS FÍSICAS EM EMPRESAS COMERCIAIS:

DA DECISÃO BINÁRIA PARA A DECISÃO MONETÁRIA

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Administração da Universidade Federal do Rio Grande do Sul, como requisito parcial para obtenção do título de Doutor em Administração.

Orientador: Prof. Dr. João Luiz Becker

Porto Alegre

2012

CIP - Catalogação na Publicação

Selau, Lisiane Priscila Roldão
Modelagem para concessão de crédito a pessoas físicas em empresas comerciais: da decisão binária para a decisão monetária / Lisiane Priscila Roldão Selau. -- 2012.
111 f.

Orientador: João Luiz Becker.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Escola de Administração, Programa de Pós-Graduação em Administração, Porto Alegre, BR-RS, 2012.

1. Análise de crédito. 2. Modelo de previsão. 3. Decisão monetária. I. Becker, João Luiz, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Lisiane Priscila Roldão Selau

MODELAGEM PARA CONCESSÃO DE CRÉDITO A PESSOAS FÍSICAS EM EMPRESAS COMERCIAIS:

DA DECISÃO BINÁRIA PARA A DECISÃO MONETÁRIA

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Administração da Universidade Federal do Rio Grande do Sul, como requisito parcial para obtenção do título de Doutor em Administração.

Conceito final:

Aprovado em:

BANCA EXAMINADORA

_____

Prof. Dr. João Zani - UNISINOS

_____

Prof. Dr. José Luis Duarte Ribeiro - PPGEP/UFRGS

_____

Profª. Drª. Denise Lindstrom Bandeira - PPGA/UFRGS

_____

Prof. Dr. Denis Borenstein - PPGA/UFRGS

_____

Orientador: Prof. Dr. João Luiz Becker - PPGA/UFRGS

RESUMO

A presente tese tem como objetivo propor um modelo de previsão para estimar o lucro médio esperado na concessão de crédito para pessoas físicas em empresas comerciais, obtendo assim uma medida monetária para dar suporte à tomada de decisão. O modelo proposto foi desenvolvido em três grandes etapas: 1) pré-processamento; 2) modelos de classificação; e 3) modelo de previsão do risco monetário. A primeira etapa inclui três passos: (i) delimitação da população, (ii) seleção da amostra, e (iii) análise preliminar. Na segunda etapa mais dois passos são necessários: (i) construção dos modelos, e (ii) qualidade dos modelos. Por fim, a última etapa trata das definições para construção do modelo de previsão do risco monetário propriamente dito, que utilizou os seguintes métodos: (i) *ensemble*, (ii) *hybrid*, e (iii) regressão linear múltipla. A exequibilidade do modelo proposto foi testada em dados reais de concessão de crédito. São avaliados os resultados de utilização do modelo de previsão, de forma a verificar o potencial aumento nos ganhos a partir da concessão do crédito, comparando quatro cenários: (i) sem utilizar nenhum modelo de previsão de risco de crédito; (ii) utilizando o modelo de classificação obtido com a regressão logística; (iii) utilizando o modelo de classificação obtido com a rede neural; e (iv) utilizando o modelo proposto para previsão do risco monetário. O modelo construído demonstrou resultados promissores na previsão do lucro médio esperado, apresentando um aumento estimado de 94,97% em comparação com o cenário sem uso de modelo de previsão, e um aumento de 26,08% quando comparado com o cenário de uso do modelo de classificação obtido com regressão logística. Uma análise de sensibilidade dos resultados com variações na margem de lucro por transação também foi realizada, evidenciando sua robustez. Nesse sentido, o modelo proposto se mostra efetivo como ferramenta de apoio para gestão no processo de decisão de concessão de crédito.

Palavras-chave: Análise de Crédito – Modelo de Previsão – Decisão Monetária

ABSTRACT

This thesis aims to propose a forecasting model to estimate the expected average profit in lending to individuals in commercial companies, thus obtaining a monetary measure to support decision making. The proposed model was developed in three major stages: 1) preprocessing, 2) classification models, and 3) model to forecast the currency risk. The first stage includes three steps: (i) delimitation of the population, (ii) sample selection, and (iii) preliminary analysis. In the second stage two more steps are necessary: (i) construction of models, and (ii) quality of the models. Finally, the last stage is regarding to the definitions for the construction of model prediction of the currency risk itself, which used the following methods: (i) ensemble, (ii) hybrid, and (iii) multiple linear regressions. The feasibility of the proposed model was tested on real data of grant credit. Results are evaluated using the prediction model in order to verify the potential increase in profits from the grant credit, comparing four scenarios: (i) without using any prevision model of credit risk, (ii) using the classification model obtained by logistic regression, (iii) using the classification model obtained with the neural network, and (iv) using the model to forecast the currency risk. The constructed model showed promising results in predicting the expected average profits, with an estimated increase of 94.97% compared to the scenario without the use of forecasting model, and an increase of 26.08% compared with the scenario of the classification model obtained by logistic regression. A sensitivity analysis of the results with variations in the profit margin per transaction was also performed, demonstrating its robustness. Accordingly, the proposed model proved effective as a support tool for management in the decision to grant credit.

Keywords: Credit Analysis – Forecasting Model – Monetary Decision

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

Tomar decisão é algo constante tanto na vida das pessoas quanto no contexto empresarial. Em vários momentos, é necessário avaliar e decidir, dentre várias alternativas, qual a melhor solução para os problemas. Simon (1960) define a tomada de decisão como o processo de pensamento e ação que se conclui com uma escolha. O modelo de tomada de decisão proposto pelo autor considera a dificuldade do indivíduo em realizar decisões puramente racionais e ótimas.

A avaliação das várias alternativas é realizada, muitas vezes, de forma empírica, considerando experiências e sentimentos. Segundo Tversky e Kahneman (1974), as decisões são tomadas tendo por base informações limitadas ou incompletas. Além disso, não é simples perceber quais informações relevantes estão faltando, o que pode levar a julgamentos equivocados.

Mais especificamente, no mercado de crédito à pessoa física, a correta decisão é essencial para a sobrevivência de empresas comerciais que utilizam o crédito como impulsionador de suas vendas. Steiner *et al.* (1999) ressaltam que qualquer erro na decisão de conceder crédito pode significar, em uma única operação, a perda do ganho obtido em dezenas de outras bem sucedidas.

Em algumas situações o crédito é concedido para compra de bens duráveis, caso em que a perda é um pouco inferior ao montante emprestado, pois o bem comprado pode ser retomado. Entretanto, a maioria das operações de crédito popular se dá para a compra de bens não duráveis, como roupas ou medicamentos, por exemplo, caso em que a perda é igual ao montante emprestado. Portanto, é importante prever e reduzir a inadimplência, pois os prejuízos com créditos mal sucedidos deverão ser cobertos com a cobrança de altas taxas de juros em novas concessões. Por outro lado, há a necessidade de reduzir também os erros relativos a não concessão de crédito aos potenciais bons pagadores (STEINER *et al.*, 1999).

O crédito é um dos principais meios disponíveis para que as pessoas possam adquirir a enorme gama de bens e serviços disponibilizados pela sociedade moderna. Na intenção de atender suas necessidades básicas e desejos, as pessoas utilizam a alternativa do crédito e acabam por comprometer boa parte do rendimento mensal (SILVA, 2008). Observa-se,

portanto, um contrassenso no mercado de crédito à pessoa física. De um lado, um grande número de pessoas recebe uma renda que compromete o equilíbrio orçamentário básico, e de outro, várias empresas oferecem promessa de crédito fácil e desburocratizado.

Nesse sentido, a crise mundial de crédito de 2008 trouxe a perspectiva de que só terão espaço aqueles mais adaptados para períodos difíceis. Garcia (2008) ressalta que os concessores de crédito, alarmados com o ocorrido no mercado imobiliário americano, estão adotando critérios mais seletivos para evitar problemas com inadimplência futuramente. Essa seletividade vem em boa hora, tendo em vista o crescimento mais lento das economias brasileira e mundial.

Em muitas empresas, a avaliação de crédito é feita com base em uma enorme variedade de informações vindas das mais diversas fontes. Os gerentes analisam estas informações de maneira subjetiva e muitas vezes não conseguem explicar os processos de tomada de decisões, embora consigam apontar os fatores que influenciam as decisões. Além disso, estes ambientes são dinâmicos, com constantes alterações, onde as decisões devem ser tomadas rapidamente (MENDES FILHO *et al.*, 1996).

Numa empresa comercial que concede crédito (como para qualquer emprestador), o objetivo da análise de crédito é identificar os riscos nas operações de empréstimo. Segundo Schrickel (1997), para a identificação dos riscos é necessário evidenciar conclusões quanto à capacidade de pagamento do tomador e fazer recomendações relativas à melhor estruturação e tipo de empréstimo a conceder, à luz das necessidades financeiras do solicitante, dos riscos identificados, sob a perspectiva de maximização dos resultados da instituição. Ainda de acordo com o mesmo autor, os instrumentos de análise variam com a situação peculiar que se tem à frente. Porém, tomar uma decisão dentro de um contexto incerto, em constante mutação, e tendo em mãos um volume de informações nem sempre suficiente, é extremamente difícil. Portanto, esta decisão será tanto melhor, quanto melhores forem as informações disponíveis.

Desta forma, se torna vital o uso dos modelos de previsão de risco (*credit scoring*), que baseados em dados recentes de clientes com a empresa, geram uma pontuação para as características, levando à criação de um padrão de comportamento em relação à inadimplência. Segundo Guimarães e Chaves Neto (2002), quando a empresa tem à sua disposição uma regra de reconhecimento de padrões e classificação que indique previamente a chance de inadimplência de um futuro cliente, a decisão de concessão de crédito fica

facilitada, podendo-se então utilizar argumentos quantitativos em substituição a argumentos subjetivos e decidir com maior confiança.

Segundo Queiroz (2006), após a estabilização da economia no Brasil, em 1994, grande parte das vendas do setor varejista foi impulsionada pela concessão de crédito ao consumidor, que muitas vezes é feita pela própria empresa, como forma de parcelar seus produtos e aumentar as vendas. Neste cenário, os modelos de *credit scoring* surgem como ferramentas de grande relevância para o processo decisório, permitindo avaliar o crescimento da carteira de clientes, do volume de vendas e transações, sem comprometer os níveis de rentabilidade das empresas.

O termo *credit scoring* é utilizado para descrever sistemas obtidos por meio de métodos estatísticos que geram uma pontuação que representa uma medida de risco associada ao tomador, utilizando para isso um conjunto de características deste. Segundo Sousa e Chaia (2000), em função dessa expectativa de risco gerada pelo modelo, o concessor avalia e decide se recusa ou aceita a solicitação de crédito. Uma representação do processo de concessão de crédito, utilizando os modelos de *credit scoring* é apresentada na Figura 1.

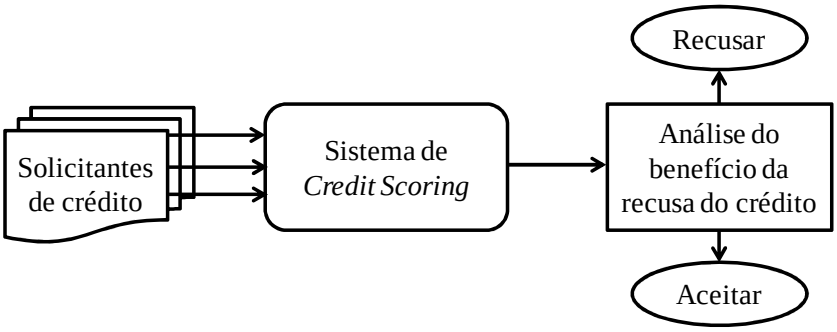

Figura 1 Processo de concessão de crédito com modelos de *credit scoring*
Fonte: Sousa e Chaia (2000)

Mais recentemente, uma linha de pesquisa promissora vem estendendo a utilização das pontuações fornecidas pelos modelos de *credit scoring*, alterando a maneira pela qual os candidatos de crédito são avaliados. O *profit scoring* propõe trabalhar com as pontuações além do aprovar ou negar, utilizando-as para atribuir limites, segmentar os clientes para campanha de marketing, entre outros (CROOK *et al.*, 2007).

Um aspecto relevante na construção dos modelos de previsão de risco de crédito diz respeito às técnicas quantitativas utilizadas. Os estudos relacionados à construção de tais modelos normalmente se preocupam em identificar, dentre as diversas técnicas quantitativas,

qual apresenta melhor poder de predição do risco. Porém, a utilização conjunta dessas técnicas possibilita a eliminação dos erros sistemáticos de previsão quando elas são utilizadas isoladamente no auxílio às decisões de concessão de crédito. A essa avaliação conjunta de técnicas para construção dos modelos de previsão de risco de crédito dá-se o nome de *ensemble* (TWALA, 2010; WANG *et al.*, 2011).

Tentando obter modelos de previsão de risco de crédito ainda mais eficientes, diversos autores (TSAI e CHEN, 2010; LEE *et al*., 2002; GHODSELAHI, 2011; LEE e CHEN, 2005) vêm estudando a utilização de modelos híbridos (*hybrid*) construídos por técnicas de naturezas diferentes e executado em dois estágios. No primeiro estágio, uma técnica é utilizada para construção de um modelo inicial para classificar os indivíduos. No segundo estágio, as previsões do primeiro estágio são adicionadas ao conjunto de dados servindo como informação de entrada para a criação do modelo final com uso da segunda técnica.

Hsieh e Hung (2010) afirmam que ainda há possibilidades de melhora na modelagem de crédito, tanto no que diz respeito à combinação de previsões, quanto numa melhor avaliação de variáveis contínuas. Autores como Tsai e Chen (2010) afirmam que deve haver ganho ao propor um modelo que ao invés de trabalhar com resposta binária (bom ou mau cliente) considere o lucro que pode ser obtido em uma concessão de crédito.

Nesse sentido, a presente tese tem como objetivo propor um modelo de previsão para estimar o lucro médio esperado na concessão de crédito para pessoas físicas em empresas comerciais, obtendo assim uma medida monetária para dar suporte à tomada de decisão. Para tanto, foram utilizados métodos emergentes em modelagem para previsão de risco de crédito (*ensemble* e *hybrid*), além de técnicas quantitativas para análise de dados (como, por exemplo, regressão logística, redes neurais e regressão linear).

1.2 **TEMA DA PESQUISA**

Esta pesquisa centra-se inicialmente em dois temas principais: processo decisório e risco da decisão. A tomada de decisão constitui-se num processo de escolha que conduz à alternativa que for considerada ótima para a organização, onde, por meio de regras e modelos, o tomador de decisão escolhe a melhor alternativa entre as existentes. O risco diz respeito à medida numérica associada à incerteza quanto ao futuro decorrente das possíveis alternativas de decisão.

Mais especificamente, no contexto desta pesquisa, têm-se os temas: decisão de concessão de crédito, risco de crédito e modelos de previsão de risco. A concessão de crédito é um importante instrumento para o desenvolvimento econômico e também se constitui uma importante atividade de empresas comerciais que tem no crédito um impulsionador das vendas. O risco de crédito pode ser definido como sendo a possibilidade do não cumprimento das obrigações contratuais relativas às transações financeiras. Os modelos de previsão têm a finalidade de estimar este risco com base nos dados cadastrais do cliente, utilizando um sistema de pontuação com uso de técnicas quantitativas para identificação de padrões de comportamento quanto à inadimplência. Na Figura 2 é representado o relacionamento desses temas para a pesquisa.

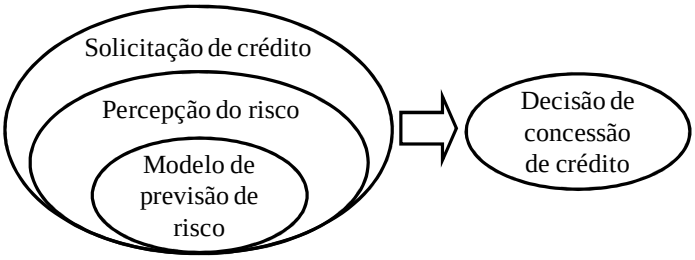

Figura 2 Relacionamento dos temas envolvidos na pesquisa

Em resposta ao crescimento recente da indústria de crédito, vários tipos de sistemas de pontuação têm sido desenvolvidos e aplicados com sucesso para apoiar as decisões de aprovação de crédito. Investigar modelos de crédito mais sofisticados é fundamental no fornecimento de resultados que atendam às necessidades de aplicações nas decisões de concessão de crédito (THOMAS, 2000).

1.3 **ESTRUTURA DO TRABALHO**

Esta tese está dividida em sete capítulos, incluindo este primeiro, conforme apresentado a seguir.

O segundo capítulo trata do referencial teórico que embasa o modelo proposto. São apresentados primeiramente os temas da tomada de decisão e os riscos envolvidos. Mais especificamente, são abordadas a importância e utilização da análise de crédito como ferramenta para controle de risco de inadimplência, bem como o uso de modelos para previsão de risco baseados em padrões de comportamento de pagamento e perfis dos

solicitantes de crédito. Também são revisados os métodos e técnicas quantitativas que são utilizados para construção do modelo proposto.

No terceiro capítulo está a metodologia de pesquisa com apresentação do problema a ser resolvido, das questões a serem respondidas e dos objetivos a serem alcançados. Também são ressaltadas a relevância e a justificativa do trabalho, finalizando com a caracterização do método de pesquisa utilizado.

No quarto capítulo é apresentado o modelo proposto para prever o risco monetário com a concessão do crédito, que constituiu a contribuição desta tese. As hipóteses do estudo também são expressas, evidenciando a originalidade da proposta e a expectativa de ganho em relação aos modelos tradicionais utilizados por empresas comerciais que concedem crédito próprio.

A descrição da modelagem para o desenvolvimento do estudo é exposta no quinto capítulo. Em resumo, o modelo proposto foi desenvolvido em três grandes etapas: 1) pré-processamento; 2) modelos de classificação; e 3) modelo de previsão do risco monetário. A primeira etapa inclui três passos: (i) delimitação da população, (ii) seleção da amostra, e (iii) análise preliminar. Na segunda etapa mais dois passos são necessários: (i) construção dos modelos, e (ii) qualidade dos modelos. Por fim, a última etapa trata das definições para construção do modelo de previsão do risco monetário propriamente dito, que utilizou os seguintes métodos: (i) *ensemble*, (ii) *hybrid*, e (iii) regressão linear.

O sexto capítulo traz os resultados da aplicação e validação de cada etapa do modelo proposto. São avaliados os resultados de utilização do modelo de previsão em dados reais de concessão de crédito, comparados em quatro cenários: (i) sem utilizar nenhum modelo de previsão de risco de crédito; (ii) utilizando o modelo de classificação obtido com a regressão logística; (iii) utilizando o modelo de classificação obtido com a rede neural; e (iv) utilizando o modelo proposto para previsão do risco monetário. Por fim, uma análise de sensibilidade dos resultados é apresentada, evidenciando a robustez do modelo.

O sétimo e último capítulo apresenta a verificação do alcance dos objetivos propostos com as principais conclusões do estudo e suas contribuições. Também são evidenciados os limites da pesquisa, sendo listadas algumas sugestões de trabalhos futuros.

2 ESQUEMA TEÓRICO-CONCEITUAL

2.1 TOMADA DE DECISÃO

De acordo com Simon (1960; 1970) e Tversky e Kahneman (1974), o processo decisório é um componente fundamental do comportamento humano, bem como do ambiente empresarial. Portanto, para os autores, não é surpresa que o assunto seja tratado em diversas áreas, como a matemática, a estatística, as ciências políticas e a economia, chegando até a sociologia e a psicologia.

De acordo com o modelo de escolha racional de tomada de decisão, os indivíduos tomam suas decisões visando à maximização de algo, adotando, para isto, um processo sequencial e linear. Nesses modelos, os tomadores de decisão identificam um problema, coletam e selecionam informações acerca das potenciais alternativas de solução do problema, comparam cada possibilidade de solução com alguns critérios pré-determinados, ordenam as soluções de acordo com preferências e selecionam a opção ótima (SIMON, 1970).

Segundo Baron (1994), a decisão baseia-se na escolha do que fazer ou não fazer, visando alcançar objetivos. São sustentadas em crenças a respeito de quais ações permitirão que se alcancem esses objetivos. Para o autor, uma estrutura de pensamento, denominada pesquisa-inferência, funciona como a base para a tomada de decisão. O processo inicia-se com uma questão ou um problema que, para ser resolvido, desencadeia-se uma pesquisa envolvendo as possibilidades de solução, objetivos e evidências. Depois desta etapa, é realizada a inferência ou uso das evidências, que serve para que cada alternativa seja fortalecida ou enfraquecida, permitindo condições para a tomada de decisão. A Figura 3 apresenta esse processo de forma esquemática.

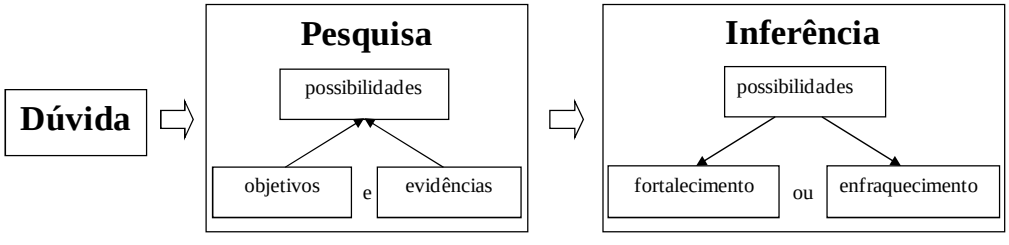

Figura 3 Estrutura de pensamento pesquisa-inferência

Cabe aqui esclarecer a relação entre tomada de decisão e julgamento, que envolvem o modo como as pessoas combinam desejos (valores pessoais, objetivos, utilidades) e crenças (conhecimentos, expectativas) na escolha de um curso de ação. Portanto, a tomada de decisão diz respeito à escolha de uma ação, e o julgamento se refere aos componentes do processo de tomada de decisão compreendidos na avaliação e estimação dos eventos que podem ocorrer, bem como dos aspectos cognitivos envolvidos (HASTIE, 2001).

Back (2002) faz uma reflexão acerca das influências no processo decisório e apresenta os elementos e suas inter-relações (Figura 4). Segundo o autor, a cultura está sempre presente e influencia os processos cognitivos, decisório e gerencial. As informações são extraídas por meio da análise dos dados e por um processo cognitivo que é influenciado pelos conhecimentos e modelos mentais do decisor. As decisões são tomadas por meio de um processo decisório, influenciado por seus valores, crenças e atitudes. No processo gerencial as decisões são transformadas em ação considerando as habilidades do decisor.

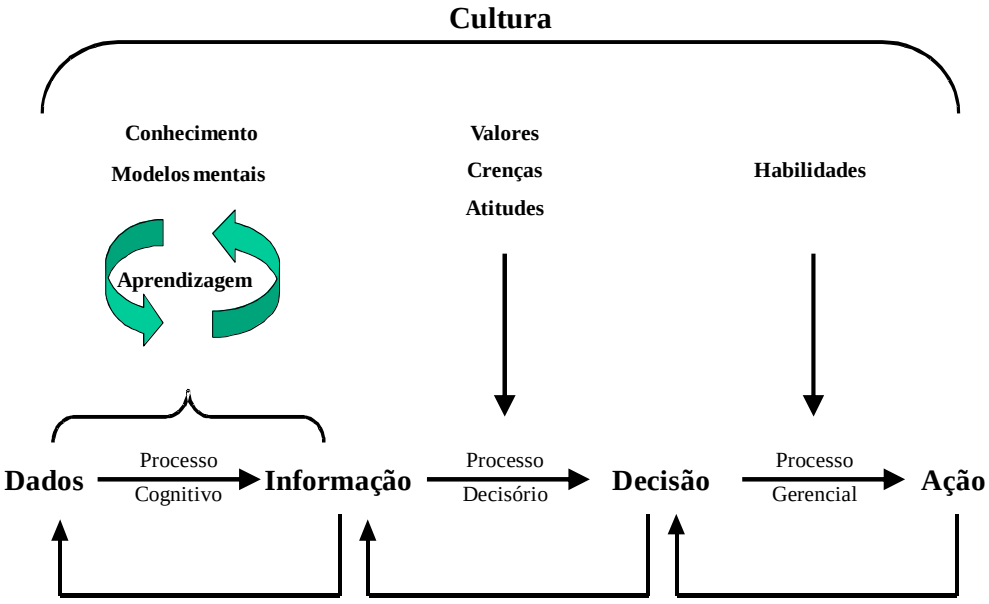

Figura 4 Influências no processo de decisão

Fonte: Back (2002)

Simon (1960) salienta que a solução de qualquer problema de decisão pode ser caracterizada em 4 etapas: (i) percepção da necessidade de decisão ou oportunidade, denominada de etapa da descoberta; (ii) formulação das possíveis alternativas de ação; (iii) avaliação das alternativas, considerando suas respectivas contribuições; e (iv) escolha de uma ou mais alternativas para fins de ação.

A tomada de decisão é uma questão central no âmbito organizacional. Shimizu (2001) sugere uma divisão para os tipos de decisões gerenciais: (i) estruturadas – envolvem um processo definido, sendo repetitivas e rotineiras; (ii) semi-estruturadas – envolvem um certo nível de previsibilidade; e (iii) não-estruturadas – decorrentes de alternativas de solução originais, sendo de caráter não rotineiro.

O processo de tomada de decisão não é tarefa simples na maioria das situações. Shimizu (2001) afirma que, com exceção dos problemas de rotina, bem conhecidos e com estrutura de opções bem definida, o processo de formular alternativas de decisão e escolher a melhor delas é quase sempre caótico e complexo.

Segundo Keller e Ho (1998), muitos modelos têm sido desenvolvidos com o objetivo de ajudar o decisor a escolher a melhor alternativa para algumas situações. Nesses casos, a estrutura do problema está bem determinada, o que permite o conhecimento das alternativas, resultados e as incertezas associadas a esses resultados. Isto ocorre com boa parte das decisões rotineiras das organizações.

De acordo com Clemen (1996), a tomada de decisão pode ser considerada uma tarefa difícil, devido a quatro aspectos envolvidos na decisão: (i) a complexidade do problema que envolve a decisão; (ii) a existência de múltiplos objetivos, algumas vezes conflitantes; (iii) as diferentes perspectivas do problema; e (iv) a incerteza inerente à tomada de decisão.

Outro aspecto importante da tomada de decisão diz respeito à qualidade. A qualidade da decisão é um assunto que tem sido tratado há algum tempo, como confirma Vlek (1984), que cita a conferência *Subjective Probability, Utility and Decision Making* ocorrida no ano de 1973, em Roma. De lá para cá, pesquisadores e estudiosos do tema (RUSSO e SCHOEMAKER, 2002) têm concordado que a adoção do processo decisório é o fator fundamental para definição da qualidade de uma decisão. Vlek (1984) destaca ainda que, a avaliação da qualidade de uma decisão poderia ser realizada tanto através de algum critério que tenha como base o resultado da decisão, como através de algum critério fundamentado no processo decisório.

Baron (1994) considera uma decisão boa quando se faz uso efetivo da informação disponível durante a tomada de decisão. McNeilly (2002) afirma que há três etapas básicas no processo de tomada de decisões estratégicas: obter a informação correta, tomar uma boa decisão e implantar esta decisão. Segundo o autor, o sucesso para obtenção de informação

correta está relacionado ao conhecimento dos tipos de informações necessárias para a tomada de decisão e a forma de encontrá-las e transmiti-las em tempo hábil.

Porém, Audy, Becker e Freitas (2001) afirmam que um dos grandes problemas enfrentados pelas organizações é a capacidade de obter o maior número de questões relacionadas ao processo decisório e processá-las de maneira objetiva visando uma maior eficiência na tomada de decisão. Um fator relevante é o estresse ao qual os decisores estão submetidos. A maioria das decisões é tomada sob elevado nível de pressão, o que interfere no comportamento presente e futuro dos decisores.

A sobrecarga de informação também é um problema enfrentado na tomada de decisão. Lurie (2004) comenta a respeito de um sensível equilíbrio entre a informação suficiente para que o indivíduo tome uma decisão, e o excesso de informação que provoca uma sobrecarga. Simon (1990) afirma que a informação consome a atenção de quem a utiliza e, desta forma, uma riqueza de informação poderia levar a uma pobreza de atenção.

De acordo com Dalfovo (1999), estar bem informado é imprescindível para os administradores, afinal a informação é a base para toda e qualquer tomada de decisão. Neste sentido, os sistemas de informação têm um papel fundamental e crescente em todas as organizações. Através destes, é possível alcançar melhores serviços, maior segurança, maior eficiência e eficácia, redução de despesas e um aperfeiçoamento no controle e na tomada de decisões.

Angeloni (2003) argumenta que a informação e o conhecimento devem ser vistos como uma cadeia de agregação de valor, sendo elementos essenciais à tomada de decisão. Portanto, tais elementos não devem ser limitados à cabeça dos indivíduos organizacionais, mas compartilhados por meio de um sistema de comunicação eficiente.

Segundo Freitas e Kladis (1995), as pessoas envolvidas no processo decisório precisam de suporte para que a decisão aconteça de forma satisfatória. Para os autores, o processo de tomada de decisão necessita ser bem compreendido e para isso se torna essencial que ferramentas, métodos e modelos estejam disponíveis no momento da decisão.

2.2 **RISCO NA TOMADA DE DECISÃO**

A habilidade de definir o que poderá acontecer no futuro e de escolher entre várias alternativas é central às sociedades contemporâneas. Segundo Bernstein (1997), a capacidade de administrar o risco e, com ele, a vontade de correr riscos e de fazer opções ousadas, são

elementos-chave da energia que impulsiona o sistema econômico. O autor ressalta ainda a importância do trabalho pioneiro de Graunt, que apresentou os conceitos teóricos básicos necessários à tomada de decisões sob condições de incerteza. Amostragens, médias e noções do que é normal compõem a estrutura que iria, mais à frente, abrigar a ciência da análise estatística, colocando a informação a serviço da tomada de decisão e influenciando nossos graus de crença sobre as probabilidades de eventos futuros.

Para Yates e Stone (1994), o risco pode ser considerado algo subjetivo, variando de uma pessoa para a outra. Riscos são sempre percebidos, e, portanto, sempre filtrados pelas limitações cognitivas dos indivíduos. Além disso, o fato de que o risco se apresenta nas mais diversas formas e situações possibilita criar a visão de que existem distintos conceitos para o termo risco. Neste momento, se faz necessária a diferenciação entre certeza, risco e incerteza, que pode causar alguma confusão em determinadas situações. Tais visões são formas de visualizar e simplificar a realidade com objetivo de chegar nos modelos de decisão. Turban e Meredith (1994) diferenciam os tipos de decisões tomadas associadas a essas condições:

a) Decisão tomada sob certeza: também conhecida como decisão determinística; neste caso é assumido que quem decide dispõe de informações completas, possibilitando o conhecimento do resultado de cada alternativa de ação que seja escolhida;

b) Decisão tomada sob risco: também chamada de decisão probabilística ou estocástica; aqui o decisor não tem controle sobre os estados futuros, podendo haver mais de um resultado possível para cada alternativa de ação. É dito que a decisão é tomada sob risco se for assumido que o decisor conhece ou pode estimar a probabilidade de ocorrência desses possíveis resultados;

c) Decisão tomada sob incerteza: neste caso também se assume que o decisor conhece os possíveis estados futuros para cada alternativa de ação, porém não conhece e nem tem condições de estimar a probabilidade de ocorrência desses possíveis resultados e, portanto, diz-se que decide sob incerteza.

Já Shaefer e Borcherding (1973) destacam que esta distinção entre risco e incerteza praticamente perdeu sentido, em função do pressuposto de Bayes, o qual afirma que toda probabilidade é subjetiva e que qualquer ato de previsão possuirá um grau de desinformação. Para Winterfeldt e Edwards (1986), toda incerteza é essencialmente do mesmo tipo, sendo que as probabilidades são números com os quais se mede a incerteza, ou seja, escalas de crenças pessoais sobre a incerteza de ocorrência de determinados eventos.

De acordo com Bernstein (1997), o tempo é o fator dominante na tomada de decisão. O risco e o tempo são as faces opostas da mesma moeda, afinal, sem amanhã não haveria risco. O tempo transforma o risco, e a natureza do risco é moldada pelo horizonte de tempo. O tempo é mais importante quando as decisões são irreversíveis, porém muitas decisões irreversíveis precisam ser tomadas com base em informações incompletas.

Yates e Stone (1994) estabelecem um constructo com o objetivo de compreender as concepções aparentemente diferentes de risco. Tal construto é composto de três elementos essenciais: (i) perdas potenciais, caracterizando-se na privação do indivíduo do alcance de um resultado que já possuía ou que poderia obter; (ii) significância das perdas, baseada na relação direta entre o grau da perda potencial e o risco; e (iii) incerteza das perdas, fundamentada no entendimento que se os resultados são garantidos, o risco não existe.

Segundo Duarte Jr. (1996), o risco está presente em qualquer operação no mercado financeiro. Por essa razão, risco é um conceito multidimensional que cobre quatro grandes categorias: risco de mercado, risco operacional, risco de crédito e risco legal.

Risco de mercado depende de como os preços dos bens se comportam diante das condições de mercado. As volatilidades e correlações dos fatores que impactam a dinâmica dos preços devem ser quantificadas para que se possa entender e medir as possíveis perdas devido às flutuações do mercado.

Risco operacional está associado às perdas resultantes de controles inadequados, falhas de gerenciamento e erros humanos, podendo ser dividido em três grandes áreas: (i) risco organizacional, relacionado com uma administração ineficiente e sem objetivos de longo prazo bem definidos, fluxos de informações interno e externo deficientes e responsabilidades mal definidas; (ii) risco de operações, relacionado com processamento e armazenamento de dados passíveis de fraudes e erros; e (iii) risco de pessoal, relacionado com problemas como empregados não-qualificados ou pouco motivados.

Risco de crédito refere-se a possíveis perdas quando um dos contratantes não honra seus compromissos, podendo ser dividido em três grupos: (i) risco país, como no caso das moratórias de países latino-americanos; (ii) risco político, quando existem restrições ao fluxo livre de capitais; e (iii) risco da falta de pagamento, quando o contratante não tem condições de honrar os compromissos assumidos.

Por fim, o risco legal está relacionado a perdas quando um contrato não pode ser legalmente amparado, em decorrência de documentação insuficiente, insolvência, ilegalidade,

entre outros. Segundo Duarte Jr. (1996), nem sempre é fácil diferenciar que tipo de risco está presente em determinada situação, podendo variar de acordo com a ótica sob a qual o problema é analisado.

Quanto ao gerenciamento do risco, Duarte Jr. (1996) sugere alguns passos para sua implantação em uma organização. Através dessas recomendações, o autor ressalta a importância de envolvimento do pessoal que detêm o poder na organização até o investimento em sistemas de informação para facilitar a transmissão de conhecimento:

a) estabelecer o gerenciamento de risco: deve ser uma decisão de quem efetivamente detêm o poder decisório na instituição, de forma a obter resultados que tenham impacto imediato, com influência na rotina diária;

b) buscar profissionais qualificados e experientes para a tarefa: um mau gerenciamento de risco pode levar a uma falsa sensação de segurança, o que pode ser pior que desconhecer o risco;

c) ter bancos de dados e sistemas computacionais de boa qualidade: a confiabilidade da análise para o risco de uma instituição está diretamente relacionada à qualidade dos dados e dos procedimentos computacionais utilizados;

d) conferir independência e autoridade aos responsáveis: tornar o processo de gerenciamento de risco o mais transparente possível, já que os responsáveis exercem um papel de auditor interno, exigindo acesso a informações restritas;

e) investir em sistemas de informação e pessoal qualificado: requisito necessário para atingir um gerenciamento de risco satisfatório.

Segundo o autor, o maior prêmio por um bom gerenciamento do risco é uma instituição mais segura que conhece suas vantagens e desvantagens, em termos de retorno e risco, em relação aos seus concorrentes.

Em relação à medição do risco, Duarte Jr. (1996) afirma que não há muita uniformidade em seu cálculo para a tomada de decisão no ambiente organizacional. As metodologias para estimação do risco demandam conhecimentos sobre os mercados de interesse, alguma sofisticação matemática e sistemas de informações confiáveis. No caso do risco operacional e do risco legal, medir risco deve ser tratado como uma abordagem caso a caso. Porém, em relação ao risco de mercado e o risco de crédito, já se encontram disponíveis

algumas metodologias na bibliografia da área e, também, muitas delas, em uso frequente pelas empresas.

2.3 A DECISÃO DE CONCEDER CRÉDITO

A palavra crédito tem origem no latim "*credere*", que significa crer, confiar, acreditar, ou ainda do substantivo "*creditum*", que literalmente significa confiança. A palavra crédito pode ter vários significados, dependendo do contexto do qual se esteja tratando, mas em finanças, de acordo com Silva (2008), crédito é um instrumento de política financeira utilizado por empresas comerciais ou industriais na venda a prazo de seus produtos, ou por um banco comercial na concessão de empréstimo ou financiamento.

> Crédito é um conceito presente no dia-a-dia das pessoas e empresas mais do que possamos imaginar a princípio. Todos nós, tanto as pessoas quanto as empresas, estamos continuamente às voltas com o dilema de uma equação simples: a constante combinação de nossos recursos finitos com o conjunto de nossas imaginações e necessidades infinitas – existem mais maneiras de se gastar dinheiro, por exemplo, do que de ganhá-lo (SCHRICKEL, 1997, p. 11).

Define-se crédito como o ato de vontade ou disposição de alguém ceder, temporariamente, parte do seu patrimônio a um terceiro, com a expectativa de que esta parcela volte a sua posse integralmente depois de decorrido o tempo previamente estipulado. Esta parte do patrimônio pode ser materializada por dinheiro (empréstimo monetário) ou bens (venda com pagamento parcelado ou a prazo). Sendo um ato de vontade, sempre caberá ao cedente do patrimônio a decisão de cedê-lo ou não (SCHRICKEL, 1997). Tal relação entre as partes é ilustrada na Figura 5, identificando o significado restrito do crédito.

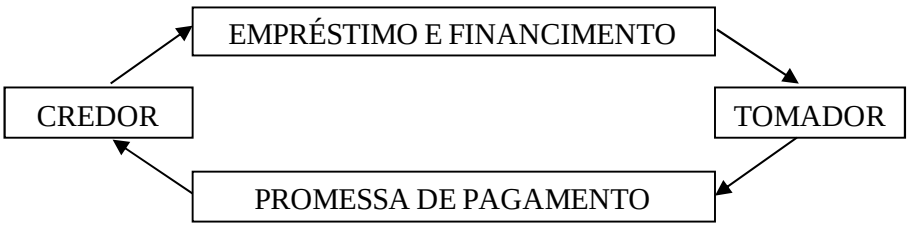

Figura 5 Ciclo da intermediação do crédito

Fonte: adaptado de Silva (2008)

No comércio, de um modo geral, o crédito assume o papel de facilitador da venda, possibilitando ao cliente adquirir o bem para atender sua necessidade, ao mesmo tempo em que incrementa as vendas do comerciante. Na indústria, o crédito também assume o papel de facilitador de venda. Silva (2008) afirma que sem a alternativa do crédito, a quantidade de compradores poderia ser muito menor e, por consequência, o lucro do fabricante também seria reduzido.

O crédito tem um papel econômico e social importante na vida das empresas e das pessoas. Possibilita às empresas aumentarem seu nível de atividade; estimula o consumo influenciando na demanda; ajuda as pessoas a obterem moradia, bens e até alimentos; e facilita a execução de projetos para os quais as empresas não disponham de recursos próprios suficientes. Por outro lado, cabe salientar que o uso inadequado do crédito pode levar uma empresa à falência ou um indivíduo à insolvência, assim como pode ser parte componente de um processo inflacionário (SILVA, 2008).

Segundo Schrickel (1997), a atividade de concessão de crédito, como tantas outras, baseia-se em informações e decisões. A esse respeito, Silva (2008) destaca que a obtenção de informações confiáveis e seu tratamento constituem a base para uma decisão de crédito segura. Antes da tomada de decisão, cabe ao credor realizar o processo de análise de crédito, buscando o maior número de informações relevantes, juntamente com o tratamento adequado destes dados.

2.4 **RISCO DE CRÉDITO**

Toda vez que uma empresa concede um crédito está automaticamente comprando um risco com todos os efeitos bons e ruins que a transação envolve. Diante dessa relação entre crédito e risco, alguns autores como Caouette *et al.* (1999) e Santos (2000) chegam a explicitar a existência do risco de crédito em suas definições de crédito, ao definir crédito como a troca de um valor presente por uma promessa de reembolso futuro, não necessariamente certo, em virtude do fator risco.

O risco de crédito está diretamente ligado ao mercado e suas mudanças. A gestão deve acompanhar essas flutuações para que a cultura do crédito e as estratégias de empréstimos possam ser repensadas e até redesenhadas (CAOUETTE *et al.*, 1999). Segundo Schrickel (1997), o risco sempre estará presente em qualquer empréstimo. Não há empréstimo

sem risco, porém o risco deve ser compatível com o negócio da empresa e à sua margem mínima almejada (receita).

Inadimplência, segundo Cia (2003), trata-se de um descumprimento por parte do devedor que acarrete alteração do montante (quanto) ou do momento (quando) do pagamento que é (eventualmente) feito ao credor, em relação ao que estava previsto em contrato. Há duas dimensões envolvidas na definição do autor: (i) quanto: que parcela do valor ou montante acordado para ser pago no vencimento é efetivamente paga; e (ii) quando: tempo decorrido entre o vencimento do contrato e o efetivo pagamento.

De acordo com Santos (2000), o risco de crédito pode ser determinado por fatores internos e externos. Fatores internos de risco são aqueles voltados à falta de experiência dos administradores no gerenciamento do crédito, a controles inadequados, à concentração de crédito nas mãos de clientes de alto risco, à política estratégica de crédito da instituição e à falta de modelagem estatística. Segundo o autor, são fatores controláveis, porém dependentes do nível de formação, da experiência adquirida e da especialização técnica dos decisores.

Os fatores externos de risco são de natureza macroeconômica, e, segundo Santos (2000), são os eventos que afetam o sistema econômico da empresa, não sendo controláveis por ela. Acontecimentos como ações do governo, desemprego, inflação, conjuntura econômica e recessão são alguns exemplos de fatores externos que podem influenciar a capacidade de pagamento.

Segundo Caouette *et al.* (1999), apesar do risco de crédito ser uma das formas mais antigas de risco nos mercados financeiros, executivos e acadêmicos ainda debatem, acaloradamente, sobre a melhor forma de sua apuração e administração. Conforme afirma Schrickel (1997), a análise de risco envolve a habilidade de tomar uma decisão de crédito, dentro de um cenário de incertezas, constantes mutações e informações incompletas. Esta habilidade depende da capacidade de analisar logicamente situações, não raro complexas, e chegar a uma conclusão prática e factível de ser estabelecida.

Corroborando essa dificuldade de obtenção de informações para a tomada de decisão quanto à concessão do crédito, estão os trabalhos de Joseph Stiglitz e George Akerlof, vencedores do Prêmio Nobel de Economia em 2001. Ocorre no mercado de crédito o fenômeno conhecido como informação assimétrica, cujas primeiras investigações são discutidas no artigo de Akerlof (1970) e cujas derivações teóricas mais conhecidas, aplicadas ao mercado de crédito, são frutos do trabalho seminal de Stiglitz e Weis (1981). Para

Akerlof (1970), o fenômeno ocorre no momento da análise de crédito, em que o concessor (que ao conceder o crédito está comprando o risco do não pagamento) tenta buscar o máximo de informações acerca do potencial cliente, e este, ainda que passe tais informações, naturalmente terá melhor conhecimento de seu comportamento como pagador do que o próprio concessor. Stiglitz e Weiss (1981) descrevem que os concessores de crédito, em decorrência da necessidade de se conhecer o cliente, procuram estabelecer mecanismos de modo a obter uma classificação de risco do empréstimo a ser efetuado.

Portanto, a atividade de concessão de crédito, como tantas outras, baseia-se em informações e decisões. A esse respeito, Silva (2008) destaca que a obtenção de informações confiáveis e seu tratamento constituem a base para uma decisão de crédito segura. Antes da tomada de decisão, cabe ao credor realizar o processo de análise de crédito, buscando o maior número de informações relevantes, juntamente com o tratamento adequado destes dados.

2.5 MODELOS DE PREVISÃO DE RISCO DE CRÉDITO

Ao longo de muitos anos, analistas eram os responsáveis pelas decisões de crédito. Thomas (2000) relata uma crise vivenciada pelas empresas que dependiam do trabalho destes analistas quando grande parte deles foi convocada para o serviço militar. Para contornar o problema, as empresas solicitaram aos seus analistas que informassem as regras utilizadas em suas decisões, passando a serem usadas por pessoas que não dominavam o assunto.

Somente após a 2ª Guerra Mundial que alguns gestores de crédito perceberam os benefícios que os modelos estatísticos poderiam trazer nas decisões de concessão de crédito. O aumento do número de pessoas que solicitavam crédito tornou economicamente impossível ter mão-de-obra suficiente para decisões que não fossem automatizadas. No fim dos anos 60 a evolução dos modelos de previsão de risco foi impulsionada pela chegada dos cartões de crédito (THOMAS, 2000).

Com o rápido desenvolvimento da informática, a partir dos anos 70, os sistemas de pontuação de crédito baseados na abordagem estatística surgiram no negócio de financiamento a pessoas físicas e jurídicas como um dos métodos mais importantes de suporte à tomada de decisão para grandes volumes de solicitações de crédito (SANTOS, 2000). No meio acadêmico, os estudos começaram no final da década de 60 e o modelo de Altman (1968) é considerado um marco teórico no estudo do risco de crédito.

Alguns autores como Silva (2008) e Caouette *et al.* (1999) têm citado a análise quantitativa como uma ferramenta poderosa na avaliação do risco de inadimplência presente na concessão de crédito. Uma das vantagens do uso de técnicas quantitativas para elaboração de sistemas de pontuação é que os pesos a serem atribuídos aos índices são determinados por cálculos e processos estatísticos, o que limita a subjetividade do analista no momento da análise, permanecendo ainda as subjetividades relacionadas às escolhas das técnicas estatísticas e dos processos de estimação.

Segundo Thomas (2000), tais ferramentas são usadas não somente para identificar os riscos dos clientes, mas também para monitorar o seu desempenho e para caracterizar seus diferentes padrões de comportamento, como, por exemplo, na decisão referente ao novo limite de crédito, na identificação da rentabilidade, nas ações de marketing, na oferta de novos produtos, na cobrança, na prevenção a fraude; enfim em todas as decisões relativas ao gerenciamento do crédito de clientes.

Dadas essas diversas decisões disponíveis através dos modelos de previsão de risco de crédito (*Credit scoring*), Paleologo *et al.* (2010) sugere uma classificação dos diferentes tipos de modelos que podem ser obtidos. Na presente tese, o foco principal está nos modelos para concessão de crédito (*Application scoring*).

a) *Application scoring*: consiste na estimativa da credibilidade de um novo candidato a crédito. Ele estima o risco de crédito em relação a condições sociais, demográficas e financeiras de um novo candidato para decidir se o crédito deve ou não ser concedido.

b) *Behavioral scoring*: similar ao *application scoring* com a diferença que envolve os clientes já existentes. Como consequência, o credor tem algumas evidências sobre o comportamento do cliente. Tais modelos analisam padrões de comportamento do cliente para apoiar os processos de gestão dinâmica da carteira.

c) *Collection scoring*: classifica os clientes em diferentes grupos de acordo com seus níveis de inadimplência, separando os clientes que precisam de ações mais decisivas daqueles que não necessitam ser atendidos imediatamente. Estes modelos permitem uma gestão dos clientes a partir dos primeiros sinais de inadimplência.

d) *Fraud detection*: classifica os candidatos de acordo com a probabilidade de que este seja um fraudador.

Segundo Mester (1997), o modelo de previsão pode ser definido como um método para avaliar o risco em concessões de crédito, construído por meio das características dos

proponentes, dados históricos e técnicas quantitativas. O método produz uma pontuação, utilizada pelos decisores para formar um ranking de risco dos clientes. Sua importância no apoio à tomada de decisão quanto à concessão de crédito já foi destacada por diversos autores (MESTER, 1997; THOMAS, 2000; PARK, 2004).

Segundo Silva (2006), os modelos de previsão de risco de crédito vêm como uma ferramenta de auxílio ao crédito massificado, que é caracterizado pela avaliação de um grande número de solicitações, já que a competitividade do mercado exige decisões rápidas. O analista informa os dados de seu potencial cliente no sistema de crédito e, imediatamente, o computador fornece a informação quanto à aprovação do crédito. O método estatístico utilizado para a construção do modelo leva em consideração o histórico da instituição com seus clientes, possibilitando a identificação e ponderação das características capazes de diferenciar o bom do mau pagador.

Para Silva (2008), as técnicas quantitativas têm sido consideradas como ferramentas poderosas na administração do risco de inadimplência existente na concessão de crédito. Para construção de um modelo de previsão de risco é importante identificar qual a técnica mais eficiente para modelar os dados de forma a conseguir a melhor previsão do comportamento dos clientes. Na Figura 6 é apresentada uma relação de alguns trabalhos sobre modelos de previsão de risco de crédito, mostrando as técnicas que foram utilizadas e as variáveis consideradas.

Após a atribuição de valores numéricos a cada característica selecionada do tomador, obtém-se uma pontuação, que indicará se o crédito pode ser concedido ou recusado, de maneira padronizada, consistente e objetiva, baseando-se nas probabilidades de reembolso. Tal ponderação é construída, buscando uma separação dos clientes segundo seu desempenho de crédito e características demográficas (SANTOS, 2000).

Pereira *et al.* (2002) definem como créditos bons os clientes que nunca atrasaram ou no pior dos casos foram moderadamente inadimplentes. Segundo Saunders (2000), a pontuação obtida com a utilização dos modelos de previsão pode ser usada de duas formas: interpretada literalmente como a probabilidade de ocorrência do não pagamento ou servir como uma medida de classificação para designar um potencial tomador em um grupo de bom ou mau cliente, comparando a pontuação obtida com um ponto de corte.

Autores	Técnicas	Variáveis
MENDES FILHO, E. F.; CARVALHO, A. C. P. L. F.; MATIAS, A. B. (1996)	Redes neurais	Sexo, idade, estado civil, tipo de residência, tempo de residência, profissão, renda e valor do patrimônio.
GUIMARÃES, I. A.; CHAVES NETO, A. (2002)	Análise discriminante e Regressão logística	Sexo, limite, tempo de residência, seguro automotivo, seguro residencial, seguro de vida, renda, idade, tempo no atual emprego, idade do cônjuge, telefone celular, estado civil, tipo do documento apresentado, escolaridade, tipo de residência, setor de atividade e CEP residencial.
GOUVÊA, M. A.; GONÇALVES, E. B. (2006)	Redes neurais e Algoritmos genéticos	Sexo, estado civil, fone residencial, fone comercial, tempo no emprego atual, salário, quantidade de parcelas a serem quitadas, primeira aquisição, tempo na residência atual, valor da parcela, valor total do empréstimo, tipo de crédito, idade, CEP residencial, CEP comercial, profissão, salário do cônjuge.
ANDREEVA, G.; ANSELLA, J.; CROOK, J. (2007)	Regressão logística e Análise de sobrevivência	Telefone de casa, tempo no emprego, tipo de residência, tipo de negócio, estado civil, telefone do empregador, ocupação, número de dependentes, idade, idade do cônjuge, tempo de residência.
DUTRA, M. S.; BIAZI, E. (2008)	Análise discriminante	Idade, sexo, se paga aluguel, tipo de profissão, tempo no emprego, número de contratos na instituição, região em que reside, renda, valor do empréstimo, quantidade de prestações e valor da prestação.
SEMEDO, D. P. V. (2009)	Regressão logística e Redes neurais	Sexo, estado civil, profissão, cargo, empresa, idade, escolaridade, nacionalidade, naturalidade, renda, valor solicitado e prestação mensal.
YAP, B.W.; ONG,S. H.; HUSAIN, N. H. M. (2011)	Regressão logística e Árvore de decisão	Sexo, idade, CEP residencial, tipo de ocupação, raça, estado civil, número de dependentes, número de carros, setor de trabalho.

Figura 6 Relação de técnicas e variáveis utilizadas em alguns modelos na literatura

Segundo Caouette *et al.* (1999), a adoção de modelos que visam estabelecer o risco de crédito em relação ao tomador não tornará os negócios isentos de risco. Os autores ainda fazem um alerta quanto a sua utilização: se os modelos forem manipulados sem o cuidado devido e sem a consciência sobre seu uso como ferramentas de apoio à análise, podem aumentar, e não minimizar, a exposição de uma instituição ao risco de crédito.

Santos (2000) salienta que, embora o uso de modelos de previsão seja de grande importância para a análise de crédito, há limitações para o seu uso: i) a determinação do risco de crédito deve considerar o impacto de fatores sistemáticos ou externos; ii) a utilização de informações imperfeitas dos tomadores e por dados disponíveis no mercado; iii) a ausência de banco de dados com informações dos tomadores em todo o mercado de crédito; e iv) as dificuldades de ajustamento de estratégias de diversificação em carteiras de empréstimos.

McAllister e Mingo (1994) destacam que o grande problema para implantação de sistemas de pontuação está na indisponibilidade e defasagem de informações completas e verídicas dos tomadores. Muitos concessores de crédito têm interesse e, mais ainda,

necessitam dos benefícios desses modelos, mas não têm um banco de dados com informações históricas de seus clientes, de forma a poder ponderar os fatores que levariam a obtenção da probabilidade de perdas com novos tomadores de crédito.

Sousa e Chaia (2000) ressaltam que, apesar de os sistemas de pontuação representarem um processo científico, não eliminam a possibilidade de se recusar um bom cliente ou que se aceite um mau pagador. Isto se deve ao fato de que nenhum modelo de previsão de risco de crédito alcança o total de características relevantes para discriminar e classificar os tipos de clientes. Mesmo que isso fosse possível, o custo de obtenção do modelo tornaria a análise economicamente inviável.

Parkinson e Ochs (1998) destacam algumas desvantagens do uso de modelos de previsão de risco de crédito: (i) custo de desenvolvimento; (ii) modelos com excesso de confiança; (iii) problemas de valores não preenchidos devido a cadastros incompletos; e (iv) interpretação equivocada dos escores. Ainda segundo os autores, as principais vantagens dos modelos são: (i) revisões de crédito consistentes; (ii) informações organizadas; (iii) eficiência no trato de dados fornecidos por terceiros; (iv) diminuição da metodologia subjetiva; (v) compreensão do processo; e (vi) maior eficiência do processo.

De acordo com Rosemberg e Gleit (1994), são muitas as vantagens da utilização de modelos de previsão de risco de crédito. Entre as principais vantagens, os autores destacam: se concede crédito (ou crédito adicional) aos melhores clientes (mais confiáveis), resultando aumento dos lucros; e nega-se crédito (ou diminui o crédito) aos piores clientes (menos confiáveis), gerando diminuição das perdas.

Caouette *et al.* (1999) defendem que os sistemas de pontuação de risco de crédito são importantes por dispor ao credor o conhecimento que não estaria, de outra maneira, prontamente disponível. Em sua contribuição, ou autores ainda acrescentam que há uma grande vantagem competitiva com a utilização dos modelos, pois com um sistema de pontuação integrado é possível operar em diversas regiões geográficas, envolvendo diversas pessoas e, mesmo assim, operar com elevado grau de objetividade nas decisões.

Nos últimos tempos, o uso de métodos para previsão de risco de crédito tem sido muito divulgado. Isto tem feito com que os concessores de crédito saiam numa corrida em busca dessas ferramentas. Contudo, esses métodos não podem ser entendidos como receitas milagrosas capazes de resolver todos os problemas relacionados ao risco de operações de crédito (SILVA, 2008).

Caouette *et al.* (1999) destacam que os modelos atuais de avaliação de risco de crédito estão mais para esforços pioneiros na busca de melhores opções do que para a solução final. Alguns dos resultados destes esforços podem vir a ser completamente descartados, mas a maioria será incorporada a modelos que ainda serão construídos. Neste sentido, todos os modelos de análise de crédito são pontes para o futuro.

2.6 MÉTODOS E TÉCNICAS EM MODELAGEM DE RISCO DE CRÉDITO

Na sequência são apresentados e descritos os métodos emergentes utilizados para construção de modelos de previsão de risco de crédito: *ensemble* e *hybrid*. Também são revisadas as técnicas estatísticas que serão utilizadas na aplicação do modelo proposto: regressão linear múltipla, regressão logística e redes neurais.

2.6.1 *Ensemble*

Muitas vezes, uma segunda opinião é considerada antes de tomar uma decisão, por vezes, uma terceira, e às vezes muito mais. Ao fazer isso, as opiniões individuais são pesadas e combinadas através de algum processo de pensamento para chegar a uma decisão final que é, presumivelmente, mais balizada. Segundo Polikar (2006), o processo de consulta a vários especialistas antes de tomar uma decisão final é natural; contudo, os benefícios de tal processo em aplicações de tomada de decisão automatizada só recentemente foram descobertos pela comunidade acadêmica.

Para Finlay (2011), não há consenso sobre qual técnica deve ser adotada para um dado problema ao construir um modelo de crédito. Dada esta incerteza, não é incomum que sejam construídos diversos classificadores utilizando diferentes técnicas, e depois se escolha o que produz a melhor solução para o problema. No entanto, ao comparar classificadores, não é garantido, necessariamente, que o melhor classificador supere todos os outros em todo o domínio do problema. Consequentemente, o autor sugere que as taxas de erro muitas vezes podem ser reduzidas através da combinação das saídas de vários classificadores.

Ao final, ao invés de escolher, dentre os modelos construídos a partir das diversas técnicas, o que alcança melhor desempenho, uma predição superior pode ser obtida a partir da avaliação conjunta das técnicas. Segundo Kittler *et al.* (1998), um modelo conjunto é obtido com uso de técnicas em paralelo, ou seja, várias técnicas são utilizadas em um mesmo

conjunto de dados e, posteriormente, suas classificações são combinadas com base em suas decisões.

Na literatura especializada em redes neurais, autores têm demonstrado que melhores resultados são conseguidos pela combinação de diferentes redes ao invés da seleção daquela que apresenta melhor desempenho individual (HAYKIN, 2001). Na Figura 7 é apresentado um esquema do método de combinação de previsões, também conhecido como *ensemble*.

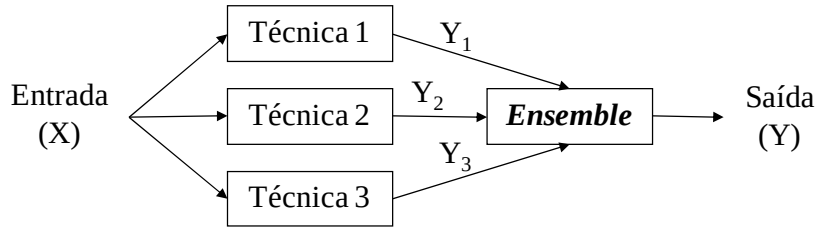

Figura 7 Representação do método de combinação de previsões (*ensemble*)

Werner e Ribeiro (2006) também tratam da combinação de previsões, propondo um modelo que utiliza mais de uma técnica ao mesmo tempo para o caso da previsão de demanda. Para os autores, como cada técnica individual pode captar diferentes características dos dados, diferentes comportamentos podem ser previstos. Assim, a combinação dos resultados pode reduzir os erros de previsão e obter melhores estimativas.

A combinação aparece como abordagem atraente para construção de modelos de previsão, pois ao invés de tentar escolher a melhor técnica a ser utilizada, formula-se o problema identificando que técnicas poderiam ajudar na melhora da previsão. Como os resultados podem ser afetados por diversas características, cada técnica pode contribuir capturando algum tipo de informação intrínseca aos dados (CLEMEN, 1989).

Dietterich (2000) aponta três razões – estatística, computacional, e representativa – para o funcionamento dos métodos de combinação de previsões.

A razão estatística surge quando o conjunto de dados é relativamente pequeno e, dessa forma, as amostras fornecem pouco poder de discriminação. Em decorrência disto, a combinação de diferentes soluções faz com que um classificador mais centrado seja obtido, que estatisticamente representa uma melhor aproximação para o verdadeiro classificador procurado.

Já a razão computacional revela-se em decorrência da constatação de que na prática é comumente inviável a obtenção de soluções ótimas, mesmo na disponibilidade de uma base

de treinamento suficientemente grande. Em decorrência disso e, supondo-se que as soluções sub-ótimas, em execuções independentes, sejam obtidas com relativa diversidade (sub-ótimos distintos), a combinação dessas soluções tem o objetivo de produzir uma solução mediana a estas, que tende a situar-se próxima à solução ótima.

Por fim, a terceira razão apontada considera casos em que a representação das soluções candidatas de um algoritmo de classificação pode, eventualmente, ser incapaz de descrever exatamente a real solução procurada, limitando-se à obtenção de soluções vizinhas (sub-ótimas). Nesse caso, com base no mesmo princípio das duas outras razões, a combinação de classificadores pode, por meio da ponderação das predições dessas soluções vizinhas, aproximar a predição da verdadeira solução.

Segundo Werner e Ribeiro (2006), para fazer uso da combinação e poder capturar os fatores que afetam as previsões, é preciso saber quais técnicas utilizar e como combiná-las. Nesse sentido, Flores e White (1988) propõem uma estrutura, estabelecendo, respectivamente, duas dimensões: (i) seleção das técnicas de previsão-base e (ii) seleção do método de combinação.

Como a principal motivação para a combinação de modelos é melhorar o desempenho da previsão, não há ganho nenhum em construir um modelo combinado composto por um conjunto de técnicas idênticas, que processe os dados de uma mesma maneira. A ideia é combinar técnicas que constituam soluções diferentes, obtendo diferentes padrões de erro quando apresentadas a um mesmo conjunto de dados (ZHOU *et al*., 2010).

A principal razão da combinação de modelos é a melhora da habilidade de generalização, permitindo que o modelo combinado minimize as falhas que possam ocorrer nos modelos individuais. Segundo Zhou *et al*. (2010), o objetivo é construir diferentes modelos de classificação e então combinar suas saídas de modo que o desempenho final seja melhor do que o desempenho individual. A diversidade de classificadores pode ser obtida de várias formas: (i) uso de diferentes conjuntos de dados; (ii) uso de diferentes parâmetros de treinamento; (iii) uso de diferentes técnicas de análise; e (iv) uso de diferentes conjuntos de características.

A construção de um modelo pela combinação de classificadores geralmente é feita em três etapas: (i) geração de um conjunto de classificadores candidatos; (ii) seleção dos candidatos que contribuem ao serem inseridos na combinação; e (iii) combinação das saídas geradas por esses elementos selecionados em uma única saída, que corresponderá ao resultado

final da classificação. Apesar de alguns autores não incluírem a etapa de seleção na construção dos modelos combinados, ela é de grande importância, uma vez que Zhou *et al.* (2002) mostraram que o uso de todos os candidatos disponíveis no *ensemble* pode degradar seu desempenho.

Polikar (2006) elenca diversos métodos de combinação: os métodos algébricos (média simples, média ponderada, soma simples, soma ponderada, produto, máximo, mínimo ou mediana) e os métodos baseados em votação (votação majoritária simples e votação majoritária ponderada). Bates e Granger (1969), em seu artigo, considerado seminal no estudo de combinação de previsões, propuseram que o método de combinar os resultados previstos por dois modelos individuais deveria constar de uma combinação linear, dando pesos inversamente proporcionais aos erros de previsão obtidos por cada modelo individual.

Granger e Ramanathan (1984) ressaltam que os métodos de combinação de previsões poderiam ser vistos como uma forma estruturada de regressão. Segundo os autores, os métodos são equivalentes ao Método de Mínimo Quadrados Ordinários, tendo como variável resposta, a previsão combinada, e como variáveis explicativas, as previsões do modelos individuais.

Para Clemen (1989), como as combinações de previsões têm se mostrado úteis nas mais variadas situações, o desafio não é justificar esta metodologia, mas, sim, encontrar maneiras fáceis e eficientes de implementá-la.

2.6.2 ***Hybrid***

Em mineração de dados, a modelagem sequencial ou híbrida tem sido uma área de pesquisa ativa para melhorar a classificação, a previsão e o desempenho de modelos de risco de crédito. Em geral, segundo Tsai e Chen (2010), o modelo híbrido é baseado na combinação de duas técnicas diferentes em sequência, conforme representado na Figura 8.

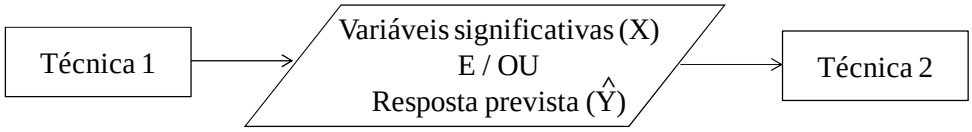

Figura 8 Representação da modelagem em sequência (*hybrid*)

Modelos híbridos vêm sendo utilizados principalmente para melhorar inconvenientes das técnicas de inteligência artificial (LEE *et al.*, 2002), já que a primeira técnica servirá para orientar o processamento da segunda (GHODSELAHI, 2011), diminuindo o tempo de processo e facilitando, através do primeiro método, a identificação da relevância das variáveis significativas (LEE e CHEN, 2005).

Lee *et al.* (2002) construíram um modelo de pontuação de crédito integrando as redes neurais com a análise discriminante, incluindo o resultado de classificação de crédito da análise discriminante de forma a simplificar a estrutura da rede e melhorar a precisão de classificação de crédito. Os resultados revelaram que a abordagem híbrida proposta converge mais rápido que o modelo único de redes neurais. Além disso, a precisão de classificação aumentou e superou as abordagens tradicionais de análise discriminante e regressão logística.

Tsai e Chen (2010) inovaram testando várias combinações de modelos híbridos, enquanto outros pesquisadores mostravam a eficiência desses modelos em comparação às técnicas individuais (LEE, *et al.*, 2002; HSIEH, 2005; CHEN, *et al.*, 2009). Seu estudo indicou que a análise de regressão logística utilizada como primeiro componente combinado com redes neurais como o segundo componente foi superior aos outros modelos. Neste tipo de modelagem as variáveis significativas obtidas na regressão logística são utilizadas como nós de entrada do modelo de redes neurais, a fim de melhorar a decisão da estrutura da rede e dar suporte às dificuldades de interpretação dos resultados obtidos.

Hsieh (2005) propõe um modelo híbrido em que a primeira técnica serve para a retirada de dados atípicos, deixando o conjunto de dados reduzido com os indivíduos bem classificados pela primeira técnica que é “mais limpo” que o conjunto de dados original pronto para uso da segunda técnica. A proposta é descrita em três etapas: (i) utilizam-se os n indivíduos da amostra de treinamento com a primeira técnica. Nesta etapa m indivíduos ($m<n$) serão bem classificados por esta técnica. (ii) com os m indivíduos bem classificados, utiliza-se agora a segunda técnica. O resultado é um classificador obtido da modelagem em sequência entre as duas técnicas. (iii) com um novo conjunto de dados, a amostra de teste, (representando novos clientes), espera-se que esse classificador seja mais preciso que o primeiro que foi utilizado individualmente.

2.6.3 **Regressão Linear Múltipla**

A análise de regressão é, sem dúvida, uma das técnicas mais utilizadas para analisar dados (CHATTERJEE *et al.*, 2000). Dentro da estatística, a maioria dos métodos de análise utiliza a teoria de regressão em sua fundamentação. Esta teoria é bem descrita em obras clássicas como Draper e Smith (1981), Montgomery e Peck (1982) e Neter *et al.* (1996), entre outros.

Regressão linear múltipla é uma técnica estatística utilizada para predizer valores de uma variável resposta, baseando-se em um conjunto de valores de várias variáveis explicativas. O objetivo é descrever esta relação por meio de uma equação (DRAPER e SMITH , 1981). Assim, o modelo de regressão linear pode ser descrito como na Equação 01:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon \qquad (01)$$

onde Y representa a variável resposta e X_1 a X_k as variáveis explicativas. As constantes β_j (j = 1, ..., k) são denominadas parâmetros e representam o modo e a magnitude da influência de cada X_j sobre Y. O termo β_0 é denominado intercepto e representa o valor de Y quando todos os valores de X_j são nulos. O termo ε representa o componente aleatório de Y, ou seja, Y é previsivelmente afetado pelas variáveis explicativas, mas elas sozinhas não explicam a totalidade da variação dos seus valores.

O método mais utilizado para estimação dos parâmetros β_j do modelo de regressão é conhecido como método de mínimos quadrados, desenvolvido por Gauss e Legendre. Segundo Memória (2004), a primeira publicação do método deu-se em 1805 por Adrien-Marie Legendre, mas Carl Gauss sustentava que já o utilizava desde 1794. O princípio fundamental deste método consiste em determinar estimativas dos parâmetros que minimizem o quadrado das diferenças entre os valores observados e os valores estimados pelo modelo proposto (DRAPER e SMITH, 1981).

Hair *et al.* (2005) estabelecem um processo de construção do modelo de regressão múltipla (Figura 9) que consiste de seis estágios com recomendações para construir, estimar, interpretar e validar a análise.

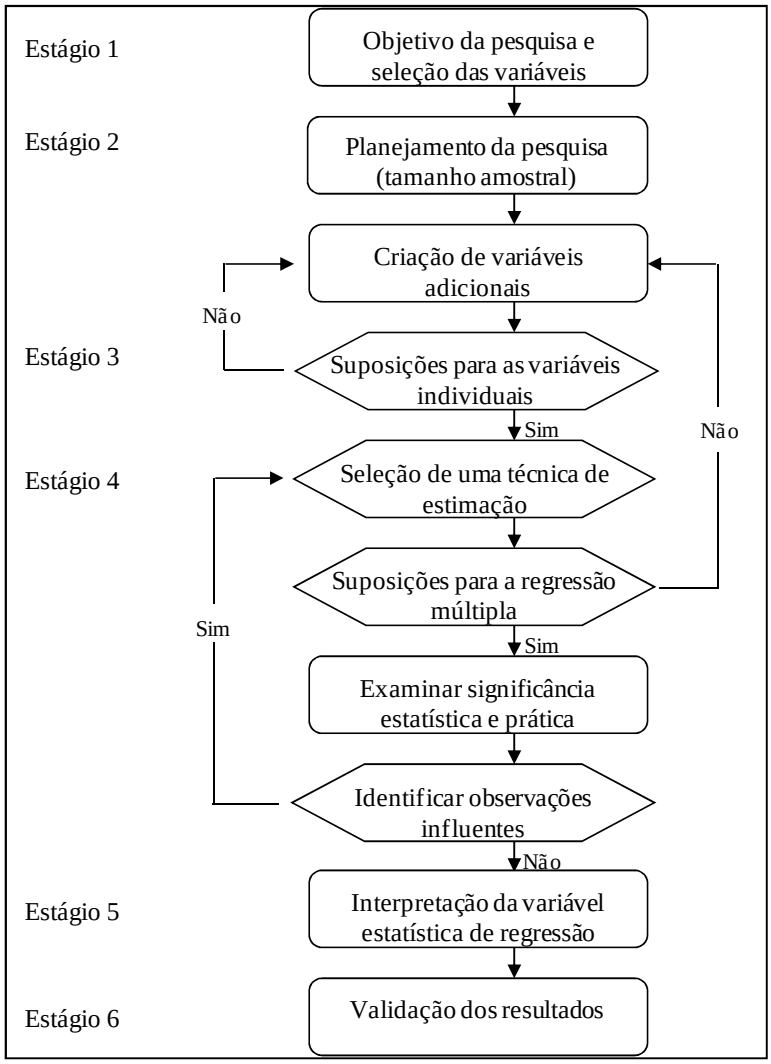

Figura 9 Processo de decisão para a regressão linear múltipla

Fonte: Hair *et al.* (2005)

Estágio 1: objetivos da regressão múltipla

Na análise de regressão múltipla, o objetivo é encontrar uma relação entre uma única variável resposta e várias explicativas, sendo que todas as variáveis envolvidas devem, a princípio, ser métricas. Ao selecionar aplicações adequadas para a técnica, o pesquisador deve considerar três pontos importantes: (i) adequação do problema de pesquisa; (ii) especificação de uma relação estatística; e (iii) seleção das variáveis resposta e explicativas.

Quanto ao problema de pesquisa apropriado para a regressão múltipla, há duas grandes classes para aplicação, podendo ser escolhidos concomitantemente: previsão e explicação. Um dos propósitos fundamentais da regressão múltipla é prever a variável resposta a partir das variáveis explicativas. A regressão múltipla pode servir ainda para verificar no conjunto de variáveis explicativas quais as que mais influenciam a variável resposta.

É apropriado também que o pesquisador especifique uma relação estatística entre as variáveis estudadas e não uma relação funcional. Em uma relação funcional o valor calculado é exato, não existindo erro algum na previsão. Porém, como ao estimar um modelo de regressão são utilizados dados amostrais, sempre haverá alguma componente aleatória na relação em estudo. Utilizando uma relação estatística, o valor previsto para a variável resposta com base nas variáveis explicativas será sempre um valor médio, nunca exato.

A qualidade do modelo ajustado é, em grande parte, explicado por uma boa seleção das variáveis a serem usadas na análise. A seleção da variável resposta muitas vezes é ditada pelo problema de pesquisa definido pelo pesquisador, sendo sempre recomendado que um embasamento teórico dirija a seleção das variáveis explicativas.

Estágio 2: planejamento de pesquisa de uma análise de regressão múltipla

No planejamento de uma análise de regressão múltipla, o pesquisador deve considerar questões como: (i) o tamanho da amostra; (ii) a natureza das variáveis explicativas; e (iii) a possível criação de novas variáveis para representar relações especiais entre as variáveis.

O tamanho da amostra é talvez o elemento mais influente no controle do pesquisador. Os efeitos do tamanho da amostra são diretamente observados no que diz respeito ao poder estatístico dos testes de significância e na generalização dos resultados para a população. O tamanho da amostra tem impacto direto e considerável sobre o poder de detectar diferenças significativas nos testes para o coeficiente de determinação e para os coeficientes de regressão das variáveis explicativas. Como regra geral, deve haver pelo menos 5 observações para cada variável explicativa considerada no modelo final, sendo que o ideal é ter entre 15 e 20 observações por variável. O requisito mais crítico é encontrado para o procedimento *stepwise* em que é recomendado pelo menos 50 observações por variável explicativa.

As variáveis explicativas podem ter efeitos fixos e aleatórios. Os efeitos fixos são definidos pelo pesquisador no momento do planejamento, como por exemplo, fixar o estudo de três tipos de adoçante e avaliar a preferência por cada refrigerante, usando o nível de doçura como variável explicativa. Um exemplo de efeito aleatório poderia ser a realização de uma pesquisa para ajudar a avaliar a relação entre idade dos respondentes e frequência de visitas ao médico, e nesse caso a variável idade seria aleatoriamente selecionada na população. Se os efeitos são aleatórios, então existe uma parcela de erro associado à amostragem, mas podem-se utilizar procedimentos para efeitos fixos, pois estes são bastante robustos e os resultados são confiáveis.

A relação básica representada na regressão múltipla é a associação linear entre as variáveis resposta e explicativas. A falta de habilidade da técnica de modelar relações não lineares pode restringir o pesquisador, sendo necessária nesses casos alguma transformação de efeitos não-lineares nas variáveis, isto é, a aplicação de funções matemáticas como logaritmo, por exemplo, aos valores da série original para criar outra série com características mais adequadas. Outro interesse do pesquisador pode ser pela inclusão de dados não-métricos como, por exemplo, sexo ou profissão, na equação de regressão. Nesse caso, é necessária a criação de variáveis adicionais chamadas de *dummies* (ou dicotômicas), onde cada variável dicotômica representa uma categoria da variável não-métrica original. A criação de variáveis adicionais fornece ao pesquisador grande flexibilidade na representação de uma vasta gama de relações em modelos de regressão.

Estágio 3: suposições em análise de regressão múltipla

Um ponto importante e muitas vezes negligenciado é a verificação das suposições para a aplicação da regressão múltipla. Tais suposições se aplicam tanto às variáveis individualmente (resposta e explicativas) como à relação como um todo. As suposições a serem examinadas são as seguintes: (i) linearidade do fenômeno; (ii) variância constante; (iii) independência; e (iv) normalidade da distribuição dos erros.

Para se avaliar a adequação do modelo de regressão múltipla é necessário analisar os resíduos, já que estes devem refletir as propriedades do erro aleatório da população, se o modelo for apropriado. A linearidade do fenômeno é avaliada graficamente, cruzando os resíduos contra as variáveis explicativas, devendo comportar-se aleatoriamente, sem tendência e ao redor de zero. A variância constante dos termos de erro (homoscedasticidade) deve ser avaliada também a partir do gráfico dos resíduos contra a variável resposta, onde se deve observar uma distribuição aleatória em torno de zero. A independência é verificada analisando os resíduos ao longo do tempo em que as observações foram obtidas, obtendo-se também uma distribuição aleatória. Por fim, a normalidade da distribuição dos erros pode ser observada fazendo um histograma dos resíduos ou mesmo através de um teste estatístico apropriado. As verificações e as ações corretivas (obtida via transformação dos dados) para essas violações são, portanto, muito recomendadas, aumentando qualidade nas interpretações e previsões da regressão múltipla.

Estágio 4: estimação do modelo de regressão e avaliação do ajuste geral

Com o objetivo da análise de regressão múltipla especificado, as variáveis selecionadas, a execução da pesquisa planejada, o pesquisador está pronto para estimar o modelo de regressão e avaliar a capacidade preditiva das variáveis explicativas. São três as tarefas básicas nesse estágio: (i) selecionar um método para especificar o modelo de regressão a ser estimado; (ii) avaliar a significância do modelo geral na previsão da variável resposta; e (iii) determinar se algumas das observações exercem influência indevida nos resultados.

Ao selecionar o método de estimação, o pesquisador pode especificar um modelo (método confirmatório), escolhendo as variáveis explicativas que devem fazer parte da regressão final, ou utilizar um procedimento de regressão que selecione automaticamente as variáveis explicativas, como os métodos de busca sequencial (*forward*, *backward* e *stepwise*), ou ainda a abordagem combinatória, onde são montados todos os subconjuntos possíveis das variáveis. Independente do método de estimação escolhido, o critério mais importante será o bom conhecimento do pesquisador sobre o contexto da pesquisa, que permita uma perspectiva objetiva e fundamentada quanto às variáveis a serem incluídas e aos sinais e magnitudes esperados de seus coeficientes. Sem esse conhecimento, os resultados da regressão podem ter elevada precisão preditiva sem qualquer relevância prática ou teórica.

O processo recomenda ainda testar se a variável estatística, isto é, a combinação linear das variáveis explicativas, satisfaz as suposições de regressão (linearidade, homocedasticidade, independência e normalidade). Esta investigação é conduzida pela análise dos resíduos. Também devem ser analisadas as significâncias dos coeficientes das variáveis explicativas, através do teste t, e do modelo como um todo, através do teste F. Por fim, é também realizado o exame do coeficiente de determinação (R^2), que indica a porcentagem de variação total da variável resposta que é explicada pelo modelo.

Por fim, a atenção é para a identificação de observações que estão fora dos padrões gerais do conjunto de dados ou que influenciam fortemente os resultados de regressão: observações influentes. As observações influentes são baseadas em uma das quatro condições: (1) um erro em observações ou entrada de dados; (2) uma observação válida, mas excepcional, explicável por uma situação extraordinária; (3) uma observação excepcional sem explicação convincente; ou (4) uma observação comum em suas características individuais, mas excepcional em sua combinação de características. Em todas as situações, o pesquisador é encorajado a eliminar observações verdadeiramente excepcionais, mas ainda assim evitar a eliminação daquelas que, apesar de diferentes, são representativas da população.

Estágio 5: interpretação da variável estatística de regressão

A tarefa do pesquisador agora consiste em interpretar a variável estatística analisando os coeficientes estimados pela regressão em termos da sua explicação da variável resposta. Também deve ser analisado o impacto potencial das variáveis omitidas em razão de multicolinearidade para garantir que a significância prática seja avaliada juntamente com a significância estatística. Nesse estágio é avaliada a: (i) utilização dos coeficientes de regressão; (ii) padronização dos coeficientes de regressão; e (iii) análise da multicolinearidade.

Os coeficientes de regressão são utilizados para calcular os valores previstos para a variável resposta e também para expressar a variação esperada na variável resposta para cada variação unitária nas variáveis explicativas. Pode ser de interesse também a avaliação do impacto de cada variável explicativa na previsão da resposta. Infelizmente, em muitos casos, os coeficientes de regressão não fornecem esta informação por serem fortemente influenciados pela unidade de medida das variáveis. Esse problema pode ser resolvido utilizando um coeficiente de regressão modificado, o coeficiente beta. Coeficientes beta são os coeficientes de regressão padronizados, não sendo influenciados pelas unidades de medida das variáveis explicativas ou resposta e, portanto, podem ser usados para comparações.

Uma questão-chave na interpretação da regressão linear múltipla é a correlação entre as variáveis explicativas. A multicolinearidade pode trazer problemas de interpretação de relações entre a variável resposta e as explicativas. Uma medida comumente usada para avaliar a multicolinearidade é a tolerância ($1-R_i^2$), onde R_i^2 é o coeficiente de determinação de uma variável explicativa X_i com as outras variáveis explicativas restantes. Assim, se a tolerância é pequena, a variável explicativa X_i pode ser predita pelas outras com eficiência e não precisa entrar no modelo. Um valor de referência para a tolerância é 0,10. Como solução à multicolinearidade, o pesquisador pode: (1) eliminar uma ou mais variáveis explicativas altamente correlacionas da equação; (2) verificar a matriz de correlações antes da regressão; (3) usar o modelo apenas para previsões; ou (4) usar análise de componentes principais para refletir mais claramente os efeitos das variáveis explicativas.

Estágio 6: validação dos resultados

Após identificar o melhor modelo de regressão, o último passo é garantir que ele possa ser generalizado para a população. A melhor orientação e verificar se o modelo encontrado se ajusta a um modelo teórico. Porém, nem sempre existem resultados anteriores sobre o estudo, sejam teóricos ou empíricos. Nesse caso, sugere-se a avaliação do modelo através das

seguintes abordagens alternativas: (i) amostras adicionais ou particionadas; (ii) cálculo da estatística PRESS; (iii) comparação de modelos de regressão; e (iv) previsão com o modelo.

A mais apropriada abordagem empírica de validação é testar o modelo de regressão em uma nova amostra retirada da população. Com uma nova amostra é possível garantir a representatividade do modelo, prevendo novos valores e testando seu poder preditivo. É possível também utilizar a nova amostra para criar um novo modelo e então comparar com a equação original, comparando as variáveis significantes incluídas e seus coeficientes.

Nem sempre é possível a seleção de uma nova amostra para a validação do modelo. Uma abordagem alternativa é a utilização da amostra original de uma forma especializada, calculando a estatística PRESS, uma medida semelhante ao R^2. A abordagem consiste na estimação de n-1 modelos de regressão, onde a cada estimação do modelo, n-1 observações são utilizadas para sua construção e é então utilizado para prever a observação omitida. Os resíduos da previsão das observações omitidas a cada rodada são utilizados para avaliar o ajuste preditivo do modelo.

Quando se comparam modelos de regressão, o critério mais comum é o ajuste preditivo geral. O coeficiente de determinação fornece essa informação, porém tem a desvantagem de ser influenciado pelo número de variáveis explicativas no modelo. Portanto, para comparação de modelos de regressão com diferentes números de variáveis explicativas, utiliza-se o R^2 ajustado, que é igualmente útil na comparação de modelos com diferentes conjuntos de dados, já que faz uma compensação para os diferentes tamanhos de amostras.

Por fim, previsões com o modelo construído podem ser feitas aplicando-o a um novo conjunto de dados, utilizando os valores das variáveis explicativas para calcular o valor da variável resposta. Ao utilizar o modelo é necessário considerar alguns fatores que podem impactar na qualidade das previsões: (1) calcular os intervalos de confiança das previsões para avaliar a amplitude dos valores da variável resposta; (2) atentar para que as condições e relações medidas sejam mantidas como quando se obteve a amostra; e (3) utilizar o modelo dentro da faixa de valores observados das variáveis explicativas.

2.6.4 **Regressão Logística**

A utilização da técnica de regressão logística é adequada em muitas situações porque permite que se analise o efeito de uma ou mais variáveis explicativas (discretas ou contínuas) sobre uma variável resposta dicotômica, representando a presença (1) ou ausência (0) de uma característica (HOSMER e LEMESHOW, 1989).

A técnica de regressão logística foi desenvolvida por volta de 1960 em resposta ao desafio de realizar predições ou explicar a ocorrência de determinados fenômenos em que a variável resposta fosse de natureza binária. Um dos estudos pioneiros que mais contribuíram para o avanço da técnica foi o famoso *Framingham Heart Study*, realizado com a colaboração da Universidade de Boston. O objetivo principal do estudo foi identificar os fatores que contribuíam para a ocorrência de doenças cardiovasculares (CORRAR *et al.*, 2007).

A regressão logística é uma técnica que se caracteriza por descrever a relação entre várias variáveis explicativas (X_j) e uma variável resposta binária (Y), codificada como 1 ou 0 (KLEINBAUM, 1996). Este modelo descreve o valor esperado de Y por meio da expressão apresentada na Equação 02.

$$E(Y) = \frac{1}{1+\exp\left[-\left(\beta_0 + \sum_{j=1}^{k} \beta_j X_j\right)\right]} \tag{02}$$

O objetivo na análise de regressão logística é descrever o modelo matemático de Y em função dos valores de X_j e de β_j. Assim, utilizando o método de estimação da máxima verossimilhança, os parâmetros do modelo são ajustados (HOSMER e LEMESHOW, 1989). A expressão geral do modelo logístico é dada pelas Equações 03 e 04.

$$f(z) = \frac{1}{1+e^{-z}} \tag{03}$$

$$z = \beta_0 + \sum_{j=1}^{k} \beta_j X_j \tag{04}$$

em que z é conhecido com *log odds*, variando de -∞ a +∞ como se observa na Figura 10. Assim, a função logística $f(z)$ normaliza a saída do modelo para o intervalo [0,1], informando a probabilidade de ocorrência do evento de interesse.

De acordo com Hosmer e Lemeshow (1989), a regressão logística tornou-se, portanto, um método padrão de análise de regressão para variável resposta medida de forma dicotômica. Assim, a diferença principal da regressão logística quando comparada ao modelo linear clássico é que a distribuição da variável resposta segue uma distribuição binomial, e não uma distribuição normal.

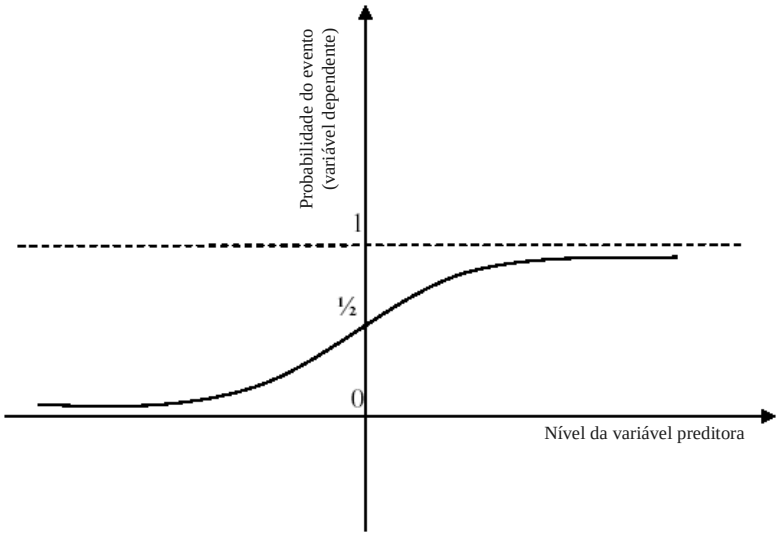

Figura 10 Forma da relação logística entre as variáveis
Fonte: Hair *et al.* (2005)

Hair *et al.* (2005) afirmam que a regressão logística se assemelha em muitos pontos à regressão linear, mas difere basicamente no sentido de prever a probabilidade de um evento ocorrer. Para obter um valor previsto delimitado entre zero e um, usa-se uma relação assumida entre as variáveis explicativas e a variável resposta que lembra uma curva em forma de 'S' (como observado na Figura 10), a distribuição sigmóide.

Os modelos lineares de regressão não podem acomodar tal relação entre as variáveis, já que ela é inerentemente não-linear. Por isso a regressão logística foi desenvolvida para lidar especificamente com essas questões. A regressão logística deriva seu nome justamente dessa transformação logística utilizada com a variável resposta (HAIR *et al.*, 2005). Ainda de acordo com esses autores, devido à natureza não-linear da transformação logística, para a estimação do modelo é necessária a utilização de um procedimento que, de forma iterativa, encontra as melhores estimativas para os coeficientes: o método da máxima verossimilhança. Isso resulta na análise do valor de verossimilhança ao invés da soma de quadrados, utilizada na regressão linear, como medida do ajuste geral do modelo.

Como o uso do modelo linear poderia conduzir a predições de valores menores que zero e maiores que um, torna-se necessário converter as observações em razão de chances (*odds ratio*) e submetê-las a uma transformação logarítmica. Com isso, o modelo passa a evidenciar mudanças nas inter-relações dos *logs* da variável explicativa. O modelo de regressão logística é obtido pelo procedimento de comparação da probabilidade de um evento

ocorrer com a probabilidade de não ocorrer. De acordo com Hair *et al.* (2005), esta razão pode ser expressa segundo a Equação 05.

$$\frac{\text{Prob(evento ocorrer)}}{\text{Prob(evento não ocorrer)}} = e^{B_0+B_1X_1+...+B_kX_k} \quad (05)$$

Os coeficientes estimados (B_0, B_1, ..., B_k) são medidas das variações na proporção das probabilidades, chamada de razão de desigualdade. São expressos em logaritmos, necessitando serem transformados para facilitar a interpretação. Um coeficiente positivo revela que aquela variável aumenta a probabilidade de ocorrência do evento, enquanto que um valor negativo diminui a probabilidade prevista.

Ao utilizar a técnica de regressão logística, o interesse pode estar na identificação do efeito de um fator de risco específico ou em determinar quais são os vários fatores associados com a variável resposta. Segundo Hosmer e Lemeshow (1989), a função logística vem sendo utilizada não apenas pela simplicidade de suas propriedades teóricas, mas, principalmente, devido a sua simples interpretação como o logaritmo da razão de chances (*odds ratio*).

Para testar a significância dos coeficientes, Hair *et al.* (2005) sugerem o uso da estatística de Wald. Ela fornece a significância estatística para cada coeficiente estimado, de modo que o teste de hipóteses pode ocorrer como acontece na regressão múltipla. Outra semelhança com a regressão múltipla está no fato de que dados nominais e categóricos podem ser tomados como variáveis explicativas do modelo por meio de codificação dicotômica (variável *dummy*).

Segundo Corrar *et al.* (2007), um dos motivos pelos quais a regressão logística tem sido muito utilizada é o pequeno número de suposições. Com esta técnica, o pesquisador consegue contornar certas restrições encontradas em outros modelos multivariados. A regressão logística não depende de suposições rígidas, tais como a normalidade das variáveis independentes e a igualdade das matrizes de covariância nos grupos. De acordo com Hair *et al.* (2005), essas suposições geralmente não são válidas em muitas situações práticas, principalmente quando há variáveis explicativas de natureza não métrica.

Apesar de sua flexibilidade, existe o pressuposto importante da baixa correlação entre as variáveis explicativas, já que o modelo de regressão logística é sensível à colinearidade entre as variáveis (HAIR *et al.*, 2005). A utilização de variáveis altamente correlacionadas para a estimação do modelo pode ocasionar estimativas extremamente inflacionadas dos coeficientes de regressão (HOSMER e LEMESHOW, 1989).

Segundo Corrar *et al*. (2007), o método *stepwise* para escolha de variáveis para compor o modelo é considerado como uma das ações corretivas para os problemas de multicolinearidade. O procedimento de avaliação das variáveis explicativas desconsidera variáveis que apresentem sinais de multicolinearidade, optando por manter no modelo apenas aquelas de maior significância estatística.

Portanto, o pesquisador que tem um problema que envolva uma variável resposta dicotômica não precisa apelar para métodos elaborados para suprir as limitações da regressão múltipla, nem precisa forçar-se a usar a análise discriminante, principalmente se suas suposições estatísticas não são satisfeitas. A regressão logística aborda satisfatoriamente esses problemas e oferece um método de análise desenvolvido especialmente para lidar com esse tipo de situação da forma mais eficiente possível (HAIR *et al.*, 2005).

2.6.5 **Rede neural**

Rede neural é uma das técnicas de tratamento de dados mais recentes e que tem despertado grande interesse tanto de pesquisadores da área de tecnologia quanto da área de negócios (CORRAR *et al*., 2007). Segundo Kovács (2002), dependendo do problema para o qual são aplicadas, as redes neurais têm apresentado desempenho superior a outros métodos de análise estatística. Por exemplo, em Subramanian *et al*. (1993), é apresentada a comparação da técnica com outros métodos estatísticos, em que os autores concluíram que as redes neurais apresentaram melhores resultados, em diversas circunstâncias, principalmente para análises de maior complexidade.

De acordo com Hair *et al*. (2005), redes neurais são uma abordagem diferente em relação a outras técnicas estatísticas. A diferença não está somente na estrutura, mas também no processo, já que as redes neurais têm um elemento-chave: a aprendizagem. Essa é outra analogia com o cérebro humano, pela qual os erros de saída são retornados ao início da rede, sendo o modelo ajustado adequadamente.

A estrutura e a operação da rede podem ser descritas por quatro conceitos: (1) o tipo de modelo de rede neural; (2) as unidades de processamentos (nós) que coletam informações, processam e criam um valor de saída; (3) o sistema de nós arranjados para transferir sinais dos nós de entrada para os nós de saída, por meio dos nós intermediários; e (4) o aprendizado pelo qual o sistema ‘retorna’ erros na previsão para ajustar o modelo (HAIR *et al*., 2005).

Haykin (2001) apresenta o elemento mais básico de uma rede neural (Figura 11). O nó é análogo ao neurônio do cérebro humano, recebendo informações de entrada e criando resultados de saída. O processamento dessa informação acontece pela criação de um valor somado no qual cada entrada é multiplicada por seu respectivo peso. Esse valor é então processado por uma função de ativação, gerando uma saída que é enviada para o nó seguinte. Em geral, a função de ativação é não linear, como a função sigmóide, da classe geral de curvas em forma de 'S' que incluiu a função logística. Outro elemento de entrada dos nós, chamado bias funciona como uma constante da função aditiva.

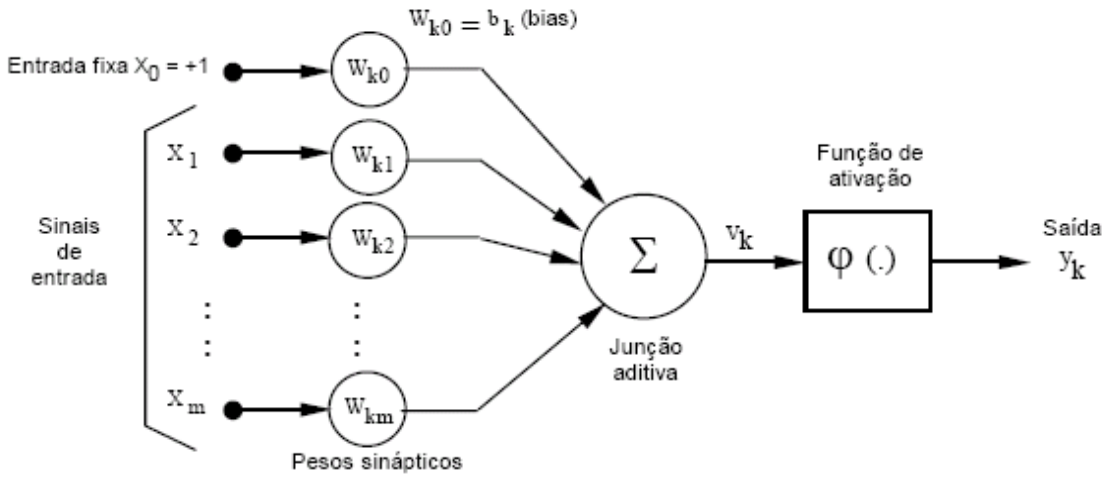

Figura 11 Modelo não linear de um nó de uma rede neural.

Fonte: Haykin (2001).

Uma rede neural é um arranjo sequencial de três tipos de nós: de entrada, de saída e intermediários (ocultos ou escondidos). Os nós de entrada recebem os dados de cada caso e os transmitem para o restante da rede. Variáveis métricas necessitam apenas um nó para cada variável, já variáveis não métricas precisam ser codificadas, de forma que cada categoria é representada por uma variável binária (HAIR *et al*., 2005). Na Figura 12 é apresentado um modelo representativo do arranjo de uma rede neural.

Segundo Hair *et al*. (2005), um nó de saída recebe entradas e obtém um valor de saída, sendo este o resultado da previsão. É por meio das camadas ocultas e da função de ativação que a rede neural consegue representar as relações não lineares entre as variáveis.

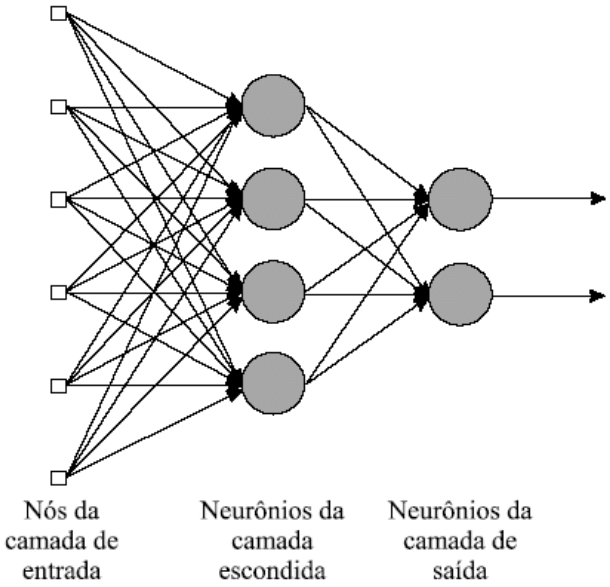

Figura 12 Modelo estrutural de uma rede neural.

Fonte: Haykin (2001).

A forma mais comum de treino da rede é a retropropagação. As variáveis de entrada são apresentadas aos nós e seu efeito se propaga através da rede, camada por camada, obtendo-se uma saída para a rede, sem alteração dos pesos sinápticos. O valor do erro é calculado, comparando-se a saída da rede com a saída esperada, e os pesos sinápticos são ajustados, tentando reduzir esse erro. Depois de treinada, a rede neural está apta a associar um conjunto de valores que são apresentados em suas entradas a um resultado de saída (HAYKIN, 2001).

Para Yu *et al*. (2002), o algoritmo de retropropagação do erro é essencial para muitos trabalhos atuais sobre aprendizado em redes neurais. Segundo Loesch e Sari (1996), o algoritmo pode ser dividido em cinco passos: (i) apresentação de um padrão de entrada e da saída desejada; (ii) cálculo dos valores de saída; (iii) ajuste dos pesos da camada de saída; (iv) ajuste de pesos das camadas escondidas; e (v) verificação da magnitude do erro

Ainda que o modelo de rede neural possa ser utilizado em situações que outras técnicas estatísticas, tais como regressão múltipla, análise discriminante e regressão logística seriam também indicadas, ele não informa sobre a importância relativa das variáveis independentes na predição devido à combinação não linear de pesos que ocorre na camada oculta. Nesse contexto, Hair *et al*. (2005) indicam a aplicação de redes neurais em problemas de previsão e classificação quando o interesse está na precisão de classificação e não na interpretação das variáveis explicativas.

Saunders (2000) argumenta que a aplicação de técnicas não lineares, como redes neurais, à análise de risco de crédito promete uma melhora sobre os modelos mais antigos de pontuação de crédito com uso de técnicas lineares de estatística. Segundo Lee e Chen (2005), com a utilização de redes neurais consegue-se poder explicativo adicional, haja vista as complexas correlações e interações entre as variáveis explicativas, que muitas vezes são realmente não lineares.

3 METODOLOGIA DE PESQUISA

Neste capítulo, são apresentados os principais elementos que conduzem a realização desta tese.

3.1 PROBLEMA E QUESTÕES DE PESQUISA

Um problema na construção dos modelos de previsão de risco de crédito está no uso de informação de qualidade para identificar o perfil de pagamento dos clientes. Os modelos atuais de *credit scoring* utilizados no mercado de concessão de crédito se limitam a uma classificação binária dos clientes (bom ou mau pagador), o que se constitui em perda de informação (TSAI e CHEN, 2010). A perda de informação começa com a definição do desempenho dos clientes quanto à inadimplência, utilizando-se um escala dicotômica: bom e mau. Ao invés disso, um melhor aproveitamento da informação poderia ser obtido ao utilizar uma escala contínua para determinar o comportamento de pagamento, desde o cliente com maior risco até o cliente com menor risco.

Segundo Gouvêa e Gonçalves (2006), na definição da variável resposta para construção de um modelo de previsão de risco de crédito, as instituições consideram apenas os clientes bons e maus para a construção do modelo devido à maior facilidade de trabalhar com modelos de resposta binária. Esta tendência também é observada em trabalhos acadêmicos e reflete dupla perda de informação pelo fato de que tanto o valor quanto o tempo de atraso do cliente não são incluídos no modelo.

Corroborando, Caouette *et al.* (1999) afirmam que modelos de previsão que medem o risco de crédito ainda estão em etapa de desenvolvimento; desta forma, são úteis, porém ainda imperfeitos. Portanto, uma melhora nesses modelos poderia ser obtida através de um método mais abrangente para o preparo das variáveis a serem utilizadas na sua construção. De acordo com Broady-Preston e Hayward (1998), a empresa com vantagens informacionais tem a habilidade de armazenar e recuperar a informação, utilizando-a em processos estratégicos específicos.

Sistemas de pontuação de crédito não são perfeitos e, todos os anos, uma proporção significativa da dívida dos clientes não é reembolsada devido ao fracasso de tais sistemas em identificar indivíduos que posteriormente se tornarão inadimplentes em seus empréstimos. Consequentemente, há um interesse considerável na melhora da capacidade dos modelos de

previsão de risco de crédito, porque mesmo pequenas melhoras no desempenho de previsão podem render benefícios financeiros significativos (FINLAY, 2011).

Desta forma, o problema de pesquisa pode ser resumido a partir das seguintes questões:

- Como incorporar em um modelo de previsão de risco de crédito informações perdidas com a dicotomia do desfecho de pagamento de um cliente (bom e mau), usualmente utilizada nos modelos tradicionais?
- Como prever o lucro esperado com um cliente após a concessão de crédito, baseando, desta forma, a decisão de crédito em uma medida monetária de risco?

Com isso, não se considera apenas o desfecho binário (bom e mau pagador) que utiliza somente a informação do tempo de atraso do cliente, mas avalia-se também o quanto cada cliente paga. Assim, reconhece-se ainda que um cliente mau pagador possa fazer alguns pagamentos antes de se tornar inadimplente, possibilitando a geração de lucro e não, necessariamente, prejuízo completo, como se admite nos modelos tradicionais. Desta forma, é possível prever o valor esperado com cada cliente e decidir sobre a concessão do crédito com maior segurança, tendo em mãos uma medida monetária de risco.

3.2 **OBJETIVOS**

O objetivo geral desta tese é propor um modelo de previsão para estimar o lucro médio esperado na concessão de crédito para pessoas físicas em empresas comerciais, obtendo assim uma medida monetária para dar suporte à tomada de decisão. Para o alcance do objetivo geral, foram estabelecidos os seguintes objetivos específicos:

- classificar os clientes, utilizando modelos de classificação obtidos com diferentes técnicas quantitativas;
- combinar os resultados previstos pelos modelos anteriores para obter uma previsão conjunta;
- prever o risco monetário para os clientes utilizando as previsões dos modelos de classificação como variáveis explicativas;
- validar o modelo proposto aplicando-o em dados reais de concessão de crédito de uma empresa comercial que utiliza o crédito próprio como forma de impulsionar suas vendas;

- avaliar o potencial ganho de utilização do modelo proposto, comparando-o com cenários de não utilização de qualquer modelo de previsão de risco de crédito e ainda com o uso de modelos tradicionais de classificação;
- propor uma fórmula para cálculo do limite de crédito a ser concedido, utilizando o valor monetário previsto pelo modelo proposto.

3.3 RELEVÂNCIA E JUSTIFICATIVA

3.3.1 Contribuição Teórica e Prática

O tema se encaixa no contexto das decisões de qualquer tipo de empresa (financeiras, industriais, comerciais ou de serviços), já que a concessão de crédito é uma forma de empréstimo ou de financiamento que muitas empresas fornecem aos seus clientes. Segundo Silva (2003), a arte de bem conceder crédito é fundamental tanto para quem concede como para quem recebe. No primeiro caso, o retorno dos recursos emprestados é fator determinante para novas concessões e, muitas vezes, para a sobrevivência do próprio negócio. No segundo, poderá ser a solução para adquirir bens de consumo de forma parcelada ou mesmo resolver uma situação financeira desfavorável.

Schrickel (1997) afirma que perder dinheiro faz parte do negócio de crédito, mas o que jamais deve ocorrer é que a perda tenha ocorrido por informações que não foram devidamente ponderadas, embora previstas ou previsíveis. A esta perda o autor dá o nome de “perda mal perdida” ou “perda burra”. Assim, segundo Steiner *et al.* (1999), os modelos de previsão de risco são muito utilizados para auxílio na análise de crédito, tendo como vantagens o aumento do número de merecedores que terão o crédito aprovado, aumentando os lucros; o aumento do número de não merecedores que não receberão o crédito, diminuindo as perdas; as solicitações de crédito podem ser analisadas rapidamente; os critérios subjetivos são substituídos pelas decisões objetivas; e por fim um menor número de pessoas será necessário para administrar o crédito.

Para Vasconcellos (2002), saber se um cliente provavelmente honrará seus compromissos apresenta-se como uma informação imprescindível na hora de tomar uma decisão com vistas à concessão de crédito. Com isso, é possível demonstrar que os concessores de crédito poderiam ter um acréscimo nos lucros se, na construção do modelo de

previsão de risco de crédito, os critérios fossem mais abrangentes. De posse da previsão fornecida pelo modelo, o concessor pode ter um diagnóstico preliminar do provável comportamento do novo cliente, concedendo ou não o crédito.

3.3.2 **Ineditismo da Proposta**

Saunders (2000) constata que as metodologias utilizadas para análise de risco ainda estão em fase de aprimoramento. Existem muitas lacunas na busca de uma gestão adequada de risco de crédito, em meio a uma geração de profissionais da área financeira (analistas, matemáticos, estatísticos, administradores, gestores, etc.) que vêm aplicando seus conhecimentos e habilidades em construção de modelos para análise desta área. Portanto, um modelo que vá além da classificação binária (conceder ou não conceder o crédito), estabelecendo uma medida monetária de risco para cada cliente, traz subsídios para a tomada de decisão com maior segurança.

Nesse sentido, esta pesquisa mostra-se original e inédita no que diz respeito a três fatores: a) ir além da decisão binária de classificar os clientes em bons ou maus e visualizar a possibilidade de obtenção de lucro com clientes antes da inadimplência; b) propor um modelo para prever o lucro médio esperado com os clientes após a concessão do crédito, tomando esse valor como uma medida monetária do risco para a tomada de decisão de crédito; e c) utilizar de forma conjunta métodos emergentes para a construção de modelos de previsão de risco de crédito (*ensemble* e *hybrid*).

Os estudos sobre modelos de previsão de risco de crédito que, ao invés de trabalhar com uma resposta binária (bom ou mau cliente), consideram o ganho que pode ser obtido com uma concessão de crédito, ainda são escassos, sendo, portanto, uma oportunidade de inovação na modelagem de risco de crédito. Tsai e Chen (2010) confirmam essa lacuna ao afirmar que os trabalhos atuais, acadêmicos ou aplicados, se preocupam em estudar o desempenho dos modelos, avaliando sua precisão e taxa de erro. Segundo os autores, nenhum estudo ainda examinou a possibilidade de estimar o lucro utilizando os modelos de crédito.

3.4 **CARACTERIZAÇÃO DA PESQUISA**

O verbo pesquisar pode ser definido como por em marcha um conjunto de ações propostas para encontrar a solução para um problema, tendo por base procedimentos racionais

e sistemáticos. Realiza-se pesquisa quando se tem um problema e não se têm informações para solucioná-lo. Os tipos de pesquisa podem ser classificados em função de sua natureza, de sua forma de abordagem, de seus objetivos e de seus procedimentos técnicos (GIL, 1999).

O método de pesquisa proposto para o desenvolvimento deste estudo pode ser classificado, quanto aos seus objetivos, como uma pesquisa explicativa, pois visa descobrir os fatores (características dos clientes) que contribuem para a ocorrência de um fenômeno (a inadimplência). Do ponto de vista da natureza da pesquisa, trata-se de uma pesquisa aplicada, pois tem por objetivo buscar conhecimento para aplicação prática com vistas à solução de problemas específicos. Quanto à abordagem do problema, esta tese pode ser vista como uma pesquisa essencialmente quantitativa, já que se propõe estudar modelos para previsão de risco de crédito e redução de inadimplência com uso de técnicas quantitativas.

Do ponto de vista dos procedimentos técnicos, o método de pesquisa empregado compreende os seguintes pontos: (i) revisão e entendimento da teoria; (ii) proposição de um modelo para previsão do risco monetário de crédito; (iii) validação do modelo em um caso aplicado; e (iv) avaliação dos resultados do modelo proposto.

O presente trabalho é desenvolvido através das etapas descritas a seguir:

- buscar referencial teórico sobre a importância e utilização da análise de crédito como medida para controle do risco de inadimplência, bem como o uso de modelos de previsão de risco de crédito;
- propor um modelo para previsão do valor esperado com cada cliente após a concessão do crédito, baseando a tomada de decisão em uma medida monetária de risco;
- detalhar os passos para o desenvolvimento do modelo proposto para previsão do risco monetário, desde a coleta de dados até a construção do modelo final;
- aplicar e validar o modelo proposto em dados reais de concessão de crédito, verificando seu desempenho através da amostra de teste, que simula a situação de análise e concessão de crédito a novos clientes;
- avaliar o potencial aumento nos lucros, comparando quatro cenários: (i) sem utilizar nenhum modelo de previsão de risco de crédito; (ii) utilizando o modelo de classificação obtido com a regressão logística; (iii) utilizando o modelo de classificação obtido com a rede neural; e (iv) utilizando o modelo proposto para previsão do risco monetário.

4 MODELO PROPOSTO

Os trabalhos encontrados atualmente na literatura se restringem em prever a probabilidade de ocorrência da inadimplência, ou seja, se um cliente pode ser considerado bom ou mau pagador (TSAI e CHEN, 2010). Com isso, é ignorada a possibilidade de que alguns clientes inadimplentes podem fazer alguns pagamentos, dando inclusive lucro para a empresa, antes de se tornarem maus pagadores. A proposta deste trabalho é, portanto, identificar uma forma de melhor prever a inadimplência após a concessão de crédito, estabelecendo uma medida monetária de risco para a tomada de decisão.

Modelos de previsão de risco de crédito vêm sendo amplamente estudados e ganhando forças devido a sua importância para a saúde financeira das empresas que concedem crédito aos seus clientes. Afinal, o sucesso das empresas está diretamente relacionado à sua capacidade de gerir os riscos (GHODSELAHI, 2011). Para lidar com estes desenvolvimentos, estão sendo usadas ferramentas matemáticas e estatísticas cada vez mais sofisticadas, sendo que, conforme Tsai e Chen (2010), uma pequena melhora na precisão da previsão de risco de crédito pode resultar numa grande redução do risco e gerar significativa economia para a instituição.

A gestão do risco de crédito vem alcançando posição de destaque, o que se reflete num maior interesse por modelos de previsão de risco de crédito. Porém, de acordo com Dutra e Biazi (2008), esses modelos dificultam em parte a gestão de riscos, por oferecerem apenas duas opções, rejeição ou aceitação da operação, não permitindo o controle do nível de risco. Com essa informação mais detalhada, os concessores poderiam ser mais ou menos agressivos na concessão de crédito.

A partir da revisão bibliográfica apresentada é possível verificar que os modelos atuais para previsão de risco de crédito têm deficiências quanto ao tratamento da variável resposta. Nesse sentido, percebe-se a possibilidade de inovação quanto à modelagem de risco, definindo, assim, a duas hipóteses a serem testadas neste estudo.

H_1 – Um cliente considerado mau pagador pode fazer alguns pagamentos antes de se tornar inadimplente, podendo inclusive trazer lucro para a empresa.

H_2 – O valor previsto para o lucro com cada cliente serve como uma medida monetária de risco para a tomada de decisão de crédito.

Utilizando modelos de crédito que apresentam resposta binária têm-se as seguintes situações: ou os clientes são previstos como bons, e deveriam ter o crédito aceito; ou os clientes são previstos como maus, e deveriam ter o crédito rejeitado. Assim, ao trabalhar com esse tipo de modelo, concluir-se-ia que o cliente bom sempre dá lucro e o cliente mau sempre dá prejuízo, o que pode não ser verdade. Um modelo para previsão do risco monetário, que vá além da dicotomia do desfecho de crédito de um cliente, deve conduzir a resultados superiores aos modelos tradicionais.

De acordo com a literatura, os estudos sobre esses modelos que, ao invés de trabalhar com uma resposta binária, consideram o ganho que pode ser obtido com a concessão de crédito, ainda são escassos, sendo, portanto, uma oportunidade de inovação na modelagem de risco de crédito. Tsai e Chen (2010) confirmam essa lacuna ao afirmar que os trabalhos atuais se preocupam em estudar o desempenho dos modelos, avaliando sua precisão e taxa de erro.

Lahsasna *et al.* (2010) afirmam que uma tendência recente no sistema de avaliação de risco de crédito é avaliar o lucro que pode ser adquirido com o cliente antes de sua inadimplência em vez de medir a probabilidade de inadimplência em suas obrigações financeiras. Portanto, mais atenção deve ser dada ao desenvolvimento do interesse da concessão de crédito e mais técnicas devem ser desenvolvidas para enfrentar os desafios mais recentes de modelagem.

Diante do exposto, o modelo proposto tem a finalidade de ser mais eficiente na utilização das informações dos clientes para mensurar o risco da concessão de crédito. Para levar a cabo os objetivos propostos, é estabelecido um conjunto de etapas para obtenção do modelo para prever o lucro esperado de um cliente, contribuição principal da tese, apresentado na Figura 13.

O modelo proposto contém três etapas. A primeira etapa é o pré-processamento, em que são definidas questões quanto à seleção da amostra e categorização das variáveis. Na segunda etapa são desenvolvidos os modelos de classificação utilizando diferentes técnicas quantitativas. A última etapa trata da construção do modelo final de previsão do risco monetário, utilizando conjuntamente (*ensemble*) as previsões das técnicas de classificação como variáveis explicativas na modelagem sequencial (*hybrid*). As duas etapas iniciais do modelo foram adaptadas da sistemática proposta em Selau (2008).

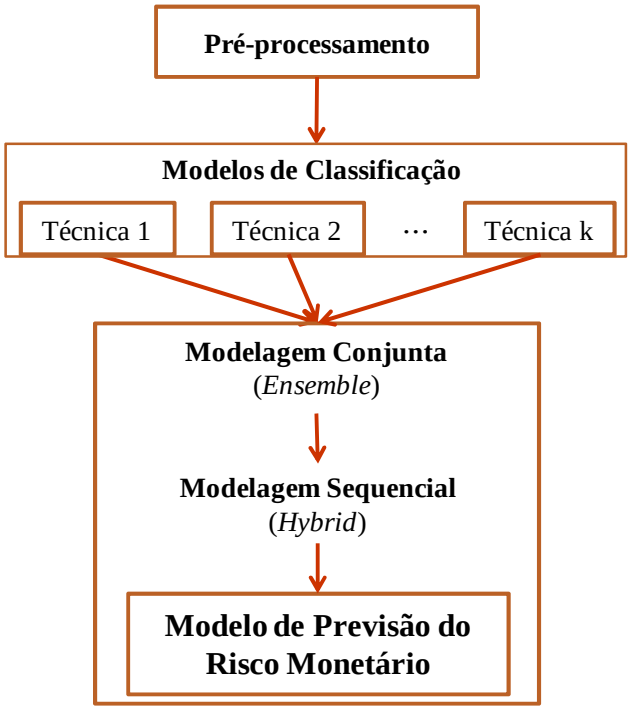

Figura 13 Modelo proposto para previsão de risco monetário

4.1 PRÉ-PROCESSAMENTO

Na etapa de pré-processamento são feitas delimitações iniciais para a seleção da amostra como, por exemplo, a definição do que é considerado um cliente bom ou mau pagador. Também são determinadas questões quanto ao tamanho e período da amostra e identificação de que informações devem ser coletadas. De posse da amostra, após uma validação dos dados, é realizado o agrupamento e seleção das variáveis que serão utilizadas na próxima etapa, da construção dos modelos de classificação. A Figura 14 apresenta os passos que contemplam esta primeira etapa do modelo proposto.

Delimitação da População	• existência de histórico de crédito; • seleção da população alvo para a qual se dirige o modelo; • definição de desempenho satisfatório e insatisfatório.
Seleção da Amostra	• identificação de variáveis disponíveis no sistema da empresa; • definição do período e tamanho da amostra; • validação da consistência e preenchimento dos dados; • separação da amostra para análise e teste.
Análise Preliminar	• escolha de variáveis para entrar na modelagem; • agrupamento de atributos de variáveis; • criação das variáveis *dummies*.

Figura 14 Etapa de pré-processamento

4.1.1 **Delimitação da População**

Antes da definição dos parâmetros para a seleção da amostra, é necessário decidir para qual segmento da população o modelo construído vai ser utilizado. Em empresas de pequeno e médio porte, nas quais há somente um produto de crédito, pode ser toda a população, ou seja, todos os clientes do negócio. Já para empresas grandes, nas quais são oferecidos diversos produtos de crédito, a população para o estudo e, por conseguinte, o modelo construído, devem ser limitados por tipo de produto (GOUVÊA e GONÇALVES, 2006).

Para a construção do modelo, é imprescindível que existam créditos concedidos pelo negócio e que os resultados da concessão tenham sido avaliados. Portanto, para que se possa desenvolver o modelo, inicialmente se deve definir os conceitos de desempenho aceitável e inaceitável. A orientação para essa definição deve ser feita pelo concessor, que definirá os atrasos que podem ser aceitos pelo negócio. A classificação de clientes quanto à inadimplência é uma etapa chave do processo de desenvolvimento de um modelo de previsão de risco de crédito. Sem dúvida que o que pode ser um bom cliente para uma empresa, poder ser mau para outra. Por esta razão, Wynn e McNab (2008) mencionam que a definição de desempenho satisfatório deve refletir a experiência da própria instituição e, por isso, são definidos como parâmetros do modelo.

4.1.2 **Seleção da Amostra**

Dentre as possíveis informações a serem selecionadas, chamadas também de variáveis demográficas, pode-se citar: sexo, idade, escolaridade, estado civil, tipo de ocupação, tipo de residência, tempo no emprego atual, entre outras (MESTER, 1997). Além destas variáveis, Hand e Henley (1997) citam outras que costumam ser consideradas na construção de modelos de *credit scoring*: tempo no endereço atual, CEP residencial e comercial, se tem telefone, renda, se tem cartões de crédito, tipo de conta bancária e objetivo do empréstimo. Nesse sentido, Lewis (1992) sugere que seja avaliada a inclusão de informações na proposta de crédito consideradas importantes para uma avaliação futura do modelo construído.

Na utilização de técnicas estatísticas, o tamanho da amostra depende do número de variáveis explicativas que farão parte do estudo para construção do modelo final. Desta forma, Hair *et al.* (2005) sugerem a utilização de uma proporção de pelo menos 20 observações para cada variável explicativa. Quanto ao período para extração da amostra, Lawrence (1992)

sugere observar um tempo de 12 a 18 meses após a concessão do crédito para que se verifique o desempenho de pagamento dos clientes.

Um último ponto a ser considerado quanto à amostra é a questão da divisão entre análise (ou treinamento) e teste, a fim de evitar qualquer tipo de viés. De acordo com Hair *et al.* (2005), não existem regras fixas quanto à partição da amostra. Devido à importância que a construção do modelo tem em relação ao seu teste, Haykin (2001) sugere a seguinte divisão: 80% da amostra total para análise e 20% para teste do modelo.

4.1.3 **Análise Preliminar**

As variáveis das quais será obtida a predição são normalmente contínuas ou categóricas. As contínuas podem ser agrupadas em intervalos, tornando-as também variáveis categóricas. Apesar disso não ser necessário para vários procedimentos estatísticos, o agrupamento oferece a vantagem de colocar todas as variáveis sobre a mesma forma, além de vantagens interpretativas. Para que as variáveis sejam categorizadas é preciso anteriormente escolher o número de categorias e as posições dos pontos de corte (HAND, 2001). Cada grupo formado dará origem a uma variável *dummy* para construção do modelo, assumindo valores 1 ou 0, indicando presença ou ausência da característica em questão.

Para o agrupamento, calcula-se o risco relativo (RR), conforme Equação 06, associado aos diferentes atributos (níveis) das variáveis explicativas. Quanto mais os percentuais de bons e maus diferirem para os atributos de uma mesma variável, maior será a utilidade dessa variável para o prognóstico de desempenho futuro (LEWIS, 1992).

$$Bons_k(\%) = \frac{b_k}{b} \text{x} 100\,, \quad Maus_k(\%) = \frac{m_k}{m} \text{x} 100\,, \quad e \quad RR_k = \frac{Bons_k(\%)}{Maus_k(\%)} \tag{06}$$

onde

k: é a k-ésima categoria da variável (k=1, 2, 3, ..., K);

b_k: número de clientes bons na k-ésima categoria;

b: total de clientes bons observados para a variável;

m_k: número de clientes maus na k-ésima categoria;

m: total de clientes maus observados para a variável;

RR_k: risco relativo de um cliente bom presente na k-ésima categoria em relação a um cliente mau.

Thomas (2000) ressalta que a designação dos atributos que formarão as categorias é considerada uma etapa crítica na construção de modelos de previsão de risco, tendo influência decisiva no desempenho final. Os grupos devem ser formados de modo que o risco seja homogêneo dentro de cada categoria e heterogêneo entre elas, seguindo a escala apresentada na Figura 15. Esse método de agrupamento é explicado por Lewis (1992) e Hand e Henley (1997).

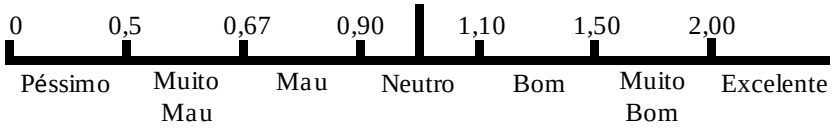

Figura 15 Classes de risco relativo para agrupamento.

Gouvêa e Gonçalves (2006) ressaltam três principais razões para se realizar este método de agrupamento das variáveis categóricas: (i) evitar categorias com um número pequeno de observações, o que pode levar a estimativas pouco robustas dos parâmetros; (ii) reduzir o número de parâmetros do modelo, pois se duas categorias apresentam risco próximo, é razoável agrupá-las numa única classe; e (iii) identificar previamente se a categoria em questão está mais ligada a clientes bons ou maus.

4.2 MODELOS DE CLASSIFICAÇÃO

A etapa de obtenção dos modelos de classificação compreende a construção em si dos modelos e a avaliação da sua qualidade na classificação, conforme apresentado na Figura 16.

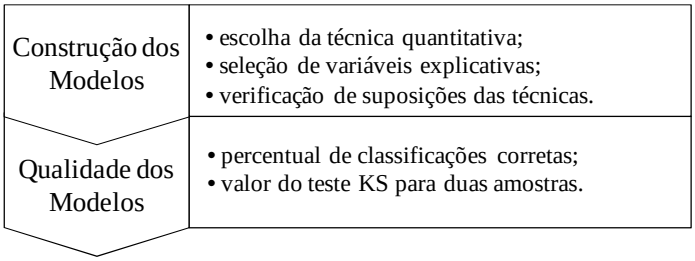

Figura 16 Etapa de obtenção dos modelos de classificação

4.2.1 Construção dos Modelos

São diversas as técnicas que têm sido utilizadas para a construção dos modelos de previsão de risco de crédito. De acordo com Keramati e Yousefi (2011), dentre os métodos

mais adotados estão: regressão linear, programação linear, algoritmos genéticos, análise de agrupamento, árvore de decisão, análise discriminante, regressão logística, redes neurais e a análise de sobrevivência.

A construção de um modelo é uma tarefa complexa. É necessária, por exemplo, a avaliação de variáveis que devem entrar ou sair da análise para evitar problemas de multicolinearidade. Esse cuidado é importante porque, muitas vezes, devido à presença de alta correlação entre as variáveis explicativas, podem ocorrer trocas de sinais dos pesos no modelo final (HAIR *et al*., 2005).

Gauchi e Chagnon (2001) comparam 20 métodos de seleção de variáveis baseados em diferentes critérios de avaliação, incluindo ajuste do modelo e capacidade de predição. Dentre os métodos, destacam-se o BCOR (*backward correlations*), BQ (*backward* Q^2_{cum}) e algoritmo genético (AG). O método BCOR usa os parâmetros da regressão PLS para rodar uma sequência de eliminação de variáveis a partir da significância dos coeficientes de correlação de cada variável explicativa. O método BQ, por sua vez, sistematicamente elimina a variável associada ao menor coeficiente da regressão PLS, registrando o valor Q^2_{cum} para medir a qualidade da predição a cada eliminação. Por fim, o conjunto de variáveis que maximiza o Q^2_{cum} é escolhido. Já o AG, utilizado para identificar as variáveis mais relevantes a serem utilizadas na regressão PLS, retém um número reduzido de variáveis e conduz a bons resultados na predição, porém apresenta alta variabilidade e requer demasiado processamento computacional.

Zimmer e Anzanello (2011) sugerem um método para seleção de variáveis em que os parâmetros gerados por regressão PLS (*Partial Least Squares*) dão origem a índices de importância das variáveis explicativas, identificando as variáveis mais relevantes para explicação da variabilidade na variável de resposta. Inicia-se então um processo de eliminação de variáveis do tipo *backward*, sendo a ordem de eliminação definida pelo índice de importância. O desempenho do modelo resultante após cada eliminação de variável é avaliado pelo erro quadrático médio.

Segundo Hand e Henley (1997) existem três formas mais comuns de se escolher quais características serão utilizadas: (i) baseando-se no conhecimento de especialistas, os quais por trabalharem com as informações diariamente, tem uma boa visão de quais dados serão mais úteis; (ii) usando o procedimento chamado *stepwise*, o qual consiste na adição de novas características ou grupo delas a cada passo, de forma a se achar o conjunto que obtém melhor resultado; (iii) analisando cada característica individualmente usando uma função para

determinar o poder discriminante da característica, sendo que as que tivessem baixo poder discriminante não seriam consideradas para inclusão no modelo. Vale lembrar que normalmente são usados os três métodos em conjunto.

4.2.2 **Qualidade dos Modelos**

Geralmente, duas medidas de desempenho são utilizadas para avaliar a qualidade dos modelos de classificação construídos: (i) o percentual de classificações corretas; e (ii) o valor do teste de Kolmogorov-Smirnov (KS) para duas amostras.

O percentual de acerto nas classificações deve ser avaliado através do cruzamento dos resultados observados e previstos pelo modelo, conforme apresentado na Tabela 1. Na diagonal principal ficam os casos classificados corretamente, clientes maus que foram previstos como maus e clientes bons previstos como bons. Desta forma, a taxa de acerto é medida pela divisão da quantidade de clientes corretamente classificados pelo total de clientes que fizeram parte da análise. Especialistas consideram satisfatórios os modelos com taxa de acerto superior a 65% (PICININI *et al.*, 2003).

Tabela 1 Verificação de acerto nas classificações do modelo

Observado \ Previsto	MAU	BOM	TOTAL
MAU	**Acertos na classificação dos maus**	Maus classificados como bons	Total de maus na amostra
BOM	Bons classificados como maus	**Acertos na classificação dos bons**	Total de bons na amostra
TOTAL	Total de maus previsto pelo modelo	Total de bons previsto pelo modelo	**Total de clientes**

Já o teste de KS tem como característica a simplicidade. O que se busca é determinar a diferença máxima entre duas distribuições acumuladas. As duas sub-populações (bons e maus), traduzidas pelos seus respectivos resultados previstos pelo modelo, são dispostas em distribuição cumulativas de frequências. Determinam-se as diferenças entre as distribuições de bons e maus para cada resultado previsto, sendo o valor do teste de KS a maior dessas diferenças em módulo. Segundo Picinini *et al.* (2003), obtendo-se uma diferença maior que 30%, pode-se considerar que o modelo é eficiente na predição dos dois grupos.

4.3 MODELO DE PREVISÃO DE RISCO MONETÁRIO

O objetivo do modelo proposto é ir além da previsão binária baseada no comportamento de bom ou mau pagador, mas sem perder esta informação. Na Figura 17 é apresentado um esquema de análise para construção do modelo proposto.

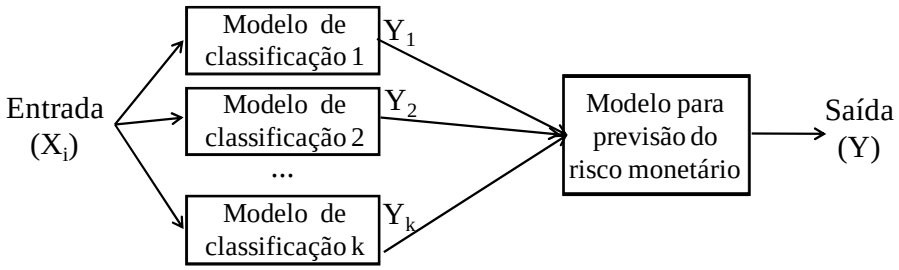

Figura 17 Esquema de análise para construção do modelo proposto

As características dos clientes (sexo, idade, etc.) são as informações de entrada (X_i) que são ponderadas na construção dos modelos de classificação. Os resultados dos modelos de classificação são as probabilidades estimadas de inadimplência (Y_1, Y_2, ..., Y_k). Estes resultados previstos pelos modelos de classificação serão avaliados combinadamente, conforme o método de avaliação conjunta (*ensemble*), e serão considerados como variáveis explicativas para construção do modelo de previsão do risco monetário, conforme o método da modelagem sequencial (*hybrid*), dando como saída o valor esperado com o cliente (Y).

Dado que o modelo proposto tem como objetivo prever o valor monetário esperado com a concessão de crédito a pessoas físicas em empresas comerciais, esse valor de cada cliente pode ser medido através da receita de vendas ou ainda por meio do lucro obtido com as vendas. As diversas formas de mensuração do valor monetário que podem ser utilizadas são expostas na Figura 18.

	Receita de vendas
-	Custos dos produtos vendidos
=	**Lucro bruto**
-	Despesas operacionais
=	**Lucro Operacional**
-	Despesas financeiras
=	**Lucro líquido antes do IR**
-	Imposto de renda
-	Participações e contribuições
=	**Lucro líquido do exercício**

Figura 18 Demonstrativo de resultado resumido

Fonte: Adaptado de Gitman (2001, p.103)

Após a estimação do modelo, o resultado será uma equação que servirá para previsão do valor esperado com cada cliente. Esta medida servirá como uma ferramenta adicional para a tomada de decisão, baseando a decisão de crédito em uma medida monetária de risco. O resultado para o modelo será uma função dos escores dos modelos de classificação, como a apresentada na Equação 07.

$$\text{E(Lucro)} = B_0 + B_1(\text{Escore modelo1}) + B_2(\text{Escore modelo2}) + ... + B_k(\text{Escore modelok}) \qquad (07)$$

Os resultados da utilização do modelo de previsão do risco monetário serão avaliados de forma a verificar o potencial aumento nos ganhos a partir da concessão do crédito. A decisão monetária será confrontada com a decisão binária, dos modelos de classificação, para evidenciar o ganho esperado com o uso do modelo proposto.

5 DESCRIÇÃO DA MODELAGEM

Neste capítulo são apresentados em detalhes todos os passos adotados para a aplicação do modelo proposto no capítulo anterior em dados reais de concessão de crédito. Na sequência são descritas as três etapas para o desenvolvimento do modelo: pré-processamento, modelos de classificação e modelo de previsão do risco monetário.

5.1 PRÉ-PROCESSAMENTO

A etapa de pré-processamento contempla os passos: (i) delimitação da população; (ii) seleção da amostra; e (iii) análise preliminar. Nesta etapa são coletados e tratados os dados que serão utilizados na etapa seguinte, de construção dos modelos de classificação.

5.1.1 Delimitação da População

A delimitação da população compreende: (i) existência de histórico de crédito; (ii) seleção da população alvo para a qual se dirige o modelo; e (iii) definição de desempenho satisfatório e insatisfatório.

A suposição básica para construir um modelo de classificação é que o padrão de comportamento dos clientes se mantém ao longo do tempo. Portanto, considerando que a construção do modelo é exclusivamente baseada na experiência do uso do crédito pela empresa, todos os dados utilizados no desenvolvimento são oriundos dos registros deste negócio. Os dados da amostra têm que constituir toda a informação conhecida dos clientes na hora da concessão do crédito e também seus estados subsequentes como bons ou maus pagadores.

Para a construção do modelo, é imprescindível que existam créditos concedidos pelo negócio e que os resultados da concessão tenham sido avaliados. Portanto, para que se possa desenvolver o modelo, inicialmente se deve definir os conceitos de desempenho aceitável e inaceitável. Quatro grupos devem ser separados no total de créditos concedidos: (i) os clientes que nunca utilizaram o crédito – sem uso; (ii) os clientes com pouco ou nenhum atraso – bons; (iii) os clientes em faixas de atrasos intermediárias – indeterminados; e (iv) os clientes com atrasos consideráveis – maus. A definição de atrasos consideráveis deve ser feita pelo concessor, que definirá os atrasos que podem ser aceitos pelo negócio. Na construção do

modelo, somente são utilizados os grupos de clientes bons e maus para acentuar a separação de perfis.

5.1.2 **Seleção da Amostra**

A seleção da amostra é realizada a partir da: (i) identificação de variáveis disponíveis no sistema da empresa; (ii) definição do período e tamanho da amostra; (iii) validação da consistência e preenchimento dos dados; e (iv) separação da amostra para análise e teste.

Inicialmente é necessário fazer um levantamento de quais informações podem ser boas variáveis explicativas para o modelo, que podem ser obtidas por meio de indicações da literatura, consulta a especialistas ou ainda pela observação detalhada da proposta de crédito da empresa. Desta forma, as variáveis a serem coletadas estão limitadas à disponibilidade da informação na base de dados da empresa.

Para definição do período para extração da amostra é necessário observar um tempo entre a concessão do crédito e a verificação de seu desempenho de pagamento. Considerar 12 a 18 meses após a concessão do crédito é usualmente suficiente para que se verifique a ocorrência de pagamentos ou de parcelas em aberto com muitos dias de atraso, definindo o cliente mau, e também de se consolidar o comportamento de pagamento do bom cliente. O tamanho da amostra geralmente não é um problema, pois quando se trata de empresas que concedem crédito à pessoa física, há abundância de dados históricos.

Definidas as informações a serem consideradas, antes de começar as análises, é necessário efetuar uma exploração prévia dos dados, em que todos os campos são avaliados quanto ao seu conteúdo. Devem ser verificadas questões relacionadas à qualidade de preenchimento, consistência dos campos e presença de observações faltantes (*missing*), eliminando dados inconsistentes ou atípicos.

Ao se testar o modelo com a mesma amostra utilizada para sua construção, pode-se concluir que o seu desempenho é bom quando, na verdade, ele pode funcionar bem apenas para estas observações. Portanto, para verificar se o poder preditivo do modelo é mantido para outras amostras provenientes da mesma população, são necessários testes para a sua validação. A separação das amostras será feita na proporção de 80% para análise e 20% para teste, através de rotina computacional, gerando-se uma variável aleatória uniformemente distribuída utilizada para alocar, ao acaso, os casos às respectivas amostras.

5.1.3 **Análise Preliminar**

A análise preliminar contempla: (i) escolha de variáveis para entrar na modelagem; (ii) agrupamento de atributos de variáveis; e (iii) criação das variáveis *dummies*.

O primeiro passo, antes da análise das informações do banco de dados, trata-se da escolha das variáveis que entrarão na análise, podendo vir a integrar o modelo final. Através da análise da associação entre as variáveis explicativas e a variável resposta (tipo de cliente) é possível selecionar quais poderão entrar na fase seguinte (construção do modelo).

Com o uso de tabelas de contingência, será calculado o risco relativo (RR), conforme Equação 06, apresentada na Seção 4.1.3, dividindo-se o percentual de bons clientes pelo percentual de maus de cada atributo. Por exemplo, se a mesma fração de bons e maus clientes tem casa própria ou alugada, essa variável não provê nenhuma informação que ajude estabelecer a probabilidade de um cliente vir a se tornar bom ou mau pagador.

Após a categorização dos atributos, seguindo a escala proposta (péssimo – RR<0,50; muito mau – RR entre 0,50 e 0,67; mau – RR entre 0,67 e 0,90; neutro – RR entre 0,90 e 1,10; bom – RR entre 1,10 e 1,50; muito bom – RR entre 1,50 e 2,00; e excelente – RR maior que 2,00), passa-se para a criação de uma variável *dummy* para cada agrupamento. Essa variável só assume dois valores: 1 ou 0 (o cliente possui tal característica ou não). Com esse artifício evitam-se problemas decorrentes da não linearidade dos atributos.

5.2 **MODELOS DE CLASSIFICAÇÃO**

5.2.1 **Construção dos Modelos**

A construção dos modelos compreende: (i) escolha da técnica quantitativa; (ii) seleção de variáveis explicativas; e (iii) verificação de suposições das técnicas.

Depois da definição dos agrupamentos e da criação das respectivas variáveis *dummies*, o analista deve escolher a técnica a ser utilizada para a modelagem. Neste estudo sugere-se a utilização da regressão logística e da rede neural. Tais métodos encontram-se entre os mais utilizados para a construção de modelos de crédito (THOMAS, 2000).

Hair *et al.* (2005) sugerem que a inclusão das variáveis explicativas no modelo aconteça conforme sua associação com a variável resposta. Num segundo momento, se o modelo ainda não atingir um nível de desempenho satisfatório, passa-se para a inclusão das variáveis com menor grau de explicação. Portanto, o método *stepwise*, incorporado em muitos pacotes estatísticos, será adotado no desenvolvimento dos modelos de classificação principalmente pela simplicidade de seu algoritmo, que automaticamente seleciona a melhor combinação de variáveis explicativas para entrada no modelo, além de ser, possivelmente, o mais amplamente difundido (ZIMMER; ANZANELLO, 2011).

Para seguir com a avaliação e utilização dos modelos construídos, é necessária a observação dos pressupostos para utilização das técnicas. Na regressão logística, o único pressuposto a ser verificado é o da ausência de multicolinearidade, que pode ser atendida com a utilização do método *stepwise* para a seleção das variáveis explicativas. A técnica de redes neurais é bastante flexível, sendo que nenhum pressuposto precisa ser verificado. As redes neurais não pressupõem um modelo ao qual os dados devem ser ajustados, já que o modelo é gerado pelo processo de aprendizagem (CORRAR *et al.*, 2007).

5.2.2 **Qualidade dos Modelos**

Duas medidas de desempenho serão utilizadas para avaliar a qualidade dos modelos: (i) percentual de classificações corretas; e (ii) o valor do teste de Kolmogorov-Smirnov (KS) para duas amostras.

O percentual de acerto nas classificações será avaliado pelo cruzamento dos resultados observados e previstos pelo modelo. Desta forma, a taxa de acerto é medida pela divisão da quantidade de clientes corretamente classificados pelo total de clientes que fizeram parte da análise. Especialistas consideram satisfatórios os modelos com taxa de acerto superior a 65%. Com o cálculo do teste de KS, o que se busca é determinar a diferença máxima entre duas distribuições acumuladas. Obtendo-se uma diferença maior que 30 entre as distribuições, pode-se considerar que o modelo é eficiente na separação dos grupos de bons e maus pagadores.

5.3 MODELO DE PREVISÃO DE RISCO MONETÁRIO

Com o intuito de ir além da classificação dos clientes em grupos e, portanto, prever o lucro esperado com cada cliente após a concessão do crédito, neste trabalho é investigada a construção de um modelo de previsão para o risco monetário com cada cliente, gerado a partir de um sistema híbrido composto por técnicas de naturezas diferentes e executado em dois estágios. No primeiro estágio tem-se o resultado das previsões dos modelos de classificação (regressão logística e rede neural), construídos tendo por base informações de perfil dos clientes; e no segundo estágio é criado um modelo de previsão do lucro esperado com cada cliente, em que aquelas previsões dos modelos de classificação são utilizadas como variáveis explicativas para previsão do risco monetário. O modelo proposto para previsão do risco monetário será construído utilizando regressão linear múltipla. O objetivo do uso da técnica é a previsão da variável resposta (lucro) em função dos escores previstos com os modelos de classificação (logístico e neural).

Dada a limitação das informações disponíveis na base de dados cedida pela empresa, dentre as possíveis formas de medir o valor monetário de cada cliente, optou-se por adotar o cálculo do lucro bruto, que é função da margem bruta obtida com as vendas. Segundo Gitman (2001), a margem bruta mensura a percentagem de cada unidade monetária de vendas que sobra após a empresa ter pago seus produtos. A margem bruta é calculada conforme a Equação 08.

$$\text{Margem Bruta} = \frac{\text{Vendas} - \text{Custo dos produtos}}{\text{Vendas}} = \frac{\text{Lucro Bruto}}{\text{Vendas}} \qquad (08)$$

Assim, definidos a margem bruta obtida com a venda e o valor total de venda dos clientes, se obtém o lucro bruto a partir da Equação 09.

$$\text{Lucro Bruto} = \text{Margem Bruta} \times \text{Vendas} \qquad (09)$$

A margem bruta representa o percentual das vendas que fica na empresa para cobertura de suas despesas operacionais, sendo uma medida de lucratividade das vendas (lucro sobre as vendas). Avalia o ganho operacional da empresa em relação a seu faturamento, mostrando o lucro disponível por unidade de venda. O lucro bruto é um relevante indicador para viabilidade de qualquer negócio (GITMAN, 2001).

Após a estimação do modelo o resultado será uma equação de regressão que servirá para previsão do lucro esperado com cada cliente. Esta medida servirá como uma ferramenta

adicional para a tomada de decisão, baseando a decisão de crédito em uma medida monetária de risco. O resultado para o modelo será uma função como a apresentada na Equação 10.

$$\text{E(Lucro)} = B_0 + B_1\,(\text{Escore Logístico}) + B_2\,(\text{Escore Neural}) \tag{10}$$

Obtidos segundo as metodologias tradicionais de modelos classificadores, os escores logístico e neural são apresentados na Equação 11 e na Figura 19, respectivamente:

$$\text{Escore Logístico} = \frac{1}{1+\exp(B_{L0} + B_{L1}X_1 + B_{L2}X_2 + ... + B_{Lk}X_k)} \tag{11}$$

onde X_1, X_2, ..., X_k representam as variáveis com características de perfis dos clientes, transformadas em variáveis *dummies*, e os coeficiente B_{L1}, B_{L2}, ..., B_{Lk} são os coeficientes estimados pela regressão logística, representando a influência de cada variável explicativa sobre o valor do escore.

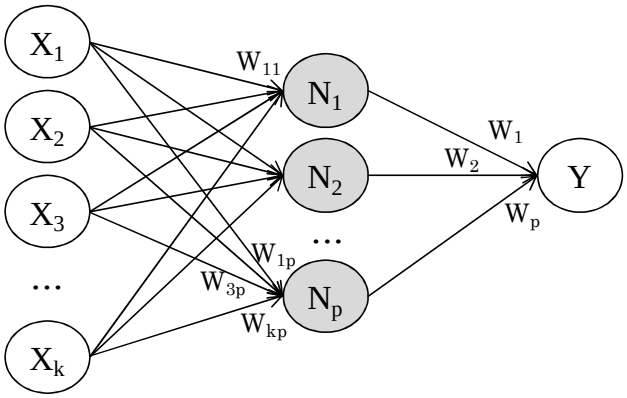

Figura 19 Representação de obtenção do escore neural

Assim como na Equação 11, X_1, X_2, ..., X_k também são as variáveis explicativas com características de perfis dos clientes, transformadas em variáveis *dummies*. Já W_{11}, ..., W_{kp} representam os pesos sinápticos que relacionam os valores das k variáveis explicavas com os p neurônios escondidos (N_1, N_2, ..., N_p) da rede neural, e W_1, W_2, ..., W_p são os pesos sinápticos que relacionam o resultados dos p neurônios escondidos com a camada de saída de rede em que é apresentada a previsão (Y), o escore neural.

Antes da utilização do modelo construído é importante avaliar as suposições para a regressão linear múltipla (linearidade, variância constante, independência e normalidade da distribuição dos erros). Hair *et al.* (2005) salientam que o atendimento das suposições é extremamente sensível ao tamanho da amostra. Porém, mesmo quando se verifica as

suposições de forma aproximada, a análise do modelo obtido para fins de previsão não fica prejudicada.

Os resultados de utilização do modelo de previsão do lucro são avaliados de forma a verificar o potencial aumento nos ganhos da empresa. A previsão de lucro é comparada em quatro cenários: (i) sem utilizar nenhum modelo de previsão de risco de crédito; (ii) utilizando o modelo de classificação obtido com a regressão logística; (iii) utilizando o modelo de classificação obtido com a rede neural; e (iv) utilizando o modelo proposto para previsão do risco monetário.

Obtendo-se o valor de lucro médio esperado com cada cliente a partir do modelo proposto, é possível também sugerir um limite de crédito tomando por base o risco monetário envolvido na concessão. Assim, o limite será atribuído em função da margem de lucro bruta assumida pela empresa e do valor do lucro médio previsto pelo modelo proposto, a partir da inversão da Equação 09, anteriormente apresentada. Na Equação 12 é apresentada a forma de cálculo do limite sugerido.

$$\text{Limite sugerido} = \frac{\text{E(Lucro)}}{\text{Margem de lucro}} \tag{12}$$

A proposta aqui é estabelecer o limite de crédito de forma objetiva, considerando a informação do risco de cada cliente, medida através da estimativa de lucro dada pelo modelo proposto. Desta forma, a empresa que concede o crédito não corre o risco de ser liberal, aumentando sua exposição ao risco da inadimplência; nem de ser rigorosa, de forma que sua exposição de risco será reduzida, porém podendo provocar uma redução no nível de negócios do cliente.

6 RESULTADOS DA APLICAÇÃO

Depois de proposto o modelo para previsão do risco monetário, cabe neste momento a validação deste por meio da aplicação em dados de concessão de crédito. Utilizando um banco de dados real de concessão de crédito, realiza-se a validação do modelo segundo seu desempenho na previsão do risco, bem como a discussão dos resultados encontrados.

6.1 PRÉ-PROCESSAMENTO

Nesta seção são apresentados os resultados relacionados às definições iniciais para delimitação da população, seleção da amostra e análises preliminares necessárias antes da construção dos modelos de classificação.

6.1.1 Delimitação da População

No desenvolvimento desta pesquisa, foram utilizadas informações sobre os clientes de uma rede de farmácias com unidades em todo o Rio Grande do Sul. É oferecido aos clientes um cartão de crédito próprio como forma de facilitar o pagamento das compras, sendo este o único produto de crédito da empresa. O pagamento pode, ainda, ser parcelado em até três vezes, com vencimento único da fatura, enviada ao endereço do cliente, semelhante a um cartão de crédito convencional. A empresa ainda não utiliza nenhum modelo para previsão de risco de crédito, concedendo para todos os solicitantes que não tiverem qualquer tipo de restrição em listas como SPC e SERASA.

De acordo com a determinação da qualidade do crédito desejada pelo concessor, chega-se à definição dos grupos de clientes segundo os atrasos. Para a empresa, o cliente bom é definido como aquele que tem atrasos de até 30 dias e os clientes maus são aqueles com pelo menos um atraso superior a 90 dias. Como indefinidos são classificados os clientes com atrasos entre 31 e 90 dias. No desenvolvimento dos modelos, este último grupo de clientes é excluído da amostra de forma a conseguir maior poder de discriminação. Além destes três grupos de clientes, foi separado um quarto grupo composto dos clientes que não tiveram nenhuma compra com o cartão no período em estudo.

6.1.2 **Seleção da Amostra**

A identificação das informações disponíveis no sistema da empresa, que serviram como variáveis explicativas para a análise, foi feita a partir da proposta que é preenchida pelos clientes no momento da solicitação do crédito. Nesta etapa, foram selecionadas 16 características, listadas na Figura 20.

Variável	**Descrição**
Sexo	Feminino ou masculino
Idade	Idade do cliente no dia do cadastro (em anos)
Estado Civil	Casado, solteiro, divorciado, viúvo, etc.
Escolaridade	Fundamental, médio ou superior
Renda	Valor da renda (R$)
Tipo de Renda	Renda declarada ou comprovada
Profissão	Profissão ou cargo do cliente
Tipo Ocupação	Assalariado, autônomo, profissional liberal, etc.
CEP Residencial	CEP do local onde reside
CEP Comercial	CEP do local onde trabalha
Tempo Serviço	Tempo no emprego atual (em meses)
Crédito 3ºs	Tem crédito em outros estabelecimentos?
Tipo Residência	Própria, alugada, cedida ou com pais
Cidade Nascimento	Cidade de naturalidade do cliente
Filho	Tem filhos?
Pensão	Paga pensão alimentícia?

Figura 20 Variáveis identificadas para criação do modelo

Fonte: proposta de crédito da empresa

Nesta aplicação contemplou-se a maioria das variáveis mais importantes para a formulação dos modelos de *credit scoring* (CAOUETTE, *et al.*, 1999). Outras variáveis (por exemplo: número de dependentes, profissão do cônjuge, valor pago de aluguel ou de prestação do imóvel, outros comprometimentos financeiros, tempo na residência atual, informação de outros bens, etc.) também poderiam ser consideradas na construção do modelo e, portanto, foram sugeridas para inclusão futura na proposta de crédito.

O período de cadastro constante na amostra compreende informações de clientes aprovados de dezembro de 2005 a junho de 2006. A amostra foi retirada em junho de 2007, atendendo ao período de 12 a 18 meses após a concessão para verificar a definição de seu desempenho como pagador. O total de clientes na amostra e a quantidade por tipo de cliente são apresentados na Tabela 2.

Tabela 2 Total de clientes por tipo

Tipo Cliente	Quantidade	%
Mau	3.368	19,8
Bom	8.026	47,2
Indefinido	1.082	6,4
Sem compra	4.529	26,6
Total	**17.005**	**100**

Como somente os grupos bons e maus são considerados para o desenvolvimento do modelo, a amostra se reduz a 11.394 clientes. Antes de passar para a separação da amostra total em duas partes (análise e teste) é importante uma avaliação geral do preenchimento dos dados para eliminar dados inconsistentes ou discrepantes. Analisando as informações de idade e tempo de serviço, verificou-se a ocorrência de valores negativos e também casos de clientes com menos de 18 anos, demonstrando erros de preenchimento nos campos. Na informação de CEP residencial, observaram-se casos de CEP geral (por exemplo: 90000-000 ou 91000-000), zerados ou de outros estados, não sendo utilizados na análise.

De forma aleatória, foram separadas as amostras de análise e teste, na proporção de 80% e 20%, respectivamente. A amostra de análise ficou formada por 6.395 clientes bons e 2.720 maus; um total de 9.115 observações, e a amostra de teste, reservada para a posterior validação dos modelos construídos ficou composta de 1.631 clientes bons e 648 maus, num total de 2.279 observações. Esta quantidade está além do proposto por Lewis (1992) de 1.500 casos para cada grupo de cliente.

6.1.3 **Análise Preliminar**

A análise para escolha das possíveis variáveis explicativas foi realizada por meio do cálculo do risco relativo, dividindo-se o percentual de bons clientes pelo percentual de maus em cada atributo, conforme apresentado na Seção 4.1.3, na Equação 06. Nesta fase inicial, três variáveis (tipo de renda, crédito em outros estabelecimentos e valor da renda) foram excluídas da análise por terem poder de discriminação baixo, em que o risco relativo dos seus atributos foi próximo de 1. A informação de pagamento de pensão alimentícia também foi excluída da análise por ter muitos dados faltantes, tendo somente 2,13% dos casos (194 clientes) preenchidos.

Para incluir as variáveis profissão, cidade de nascimento, CEP residencial e CEP comercial na análise, foi necessário agrupá-las, dado o grande número de atributos. Para tal

agrupamento foi utilizada a escala de risco relativo que foi apresentada também na Seção 4.1.3, na Figura 15 (péssimo – RR<0,50; muito mau – RR entre 0,50 e 0,67; mau – RR entre 0,67 e 0,90; neutro – RR entre 0,90 e 1,10; bom – RR entre 1,10 e 1,50; muito bom – RR entre 1,50 e 2,00; e excelente – RR maior que 2,00). Os resultados dos agrupamentos para as variáveis em questão são apresentados na Tabela 3.

Tabela 3 Agrupamento e *dummies* para profissões e cidades de nascimento e CEP

VARIÁVEL	*Dummy*	Mau	Bom	Total	RR	Classe de Risco
Grupos de Profissões						
G_PROF1	DGPROF1	90	83	173	0,39	Péssimo
G_PROF2	DGPROF2	364	502	866	0,59	Muito Mau
G_PROF3	DGPROF3	883	1.623	2.506	0,78	Mau
G_PROF4	DGPROF4	254	566	820	0,95	Neutro
G_PROF5	DGPROF5	184	505	689	1,17	Bom
G_PROF6	DGPROF6	92	375	467	1,73	Muito Bom
G_PROF7	DGPROF7	282	1.528	1.810	2,30	Excelente
Não informado		276	634	910	-	-
Não classificado		295	579	874	-	-
Grupos de Cidades de Nascimento						
G_CID.NA1	DCIDNA1	73	79	152	0,46	Péssimo
G_CID.NA2	DCIDNA2	1.075	1.478	2.553	0,58	Muito Mau
G_CID.NA3	DCIDNA3	387	733	1.120	0,81	Mau
G_CID.NA4	DCIDNA4	253	569	822	0,96	Neutro
G_CID.NA5	DCIDNA5	181	543	724	1,28	Bom
G_CID.NA6	DCIDNA6	127	507	634	1,70	Muito Bom
G_CID.NA7	DCIDNA7	72	585	657	3,46	Excelente
Não informado		173	574	747	-	-
Não classificado		379	1.327	1.706	-	-
Grupos de CEP Residencial						
G_CEP.RE1	DGCEPRE1	18	17	35	0,40	Péssimo
G_CEP.RE2	DGCEPRE2	559	824	1383	0,63	Muito Mau
G_CEP.RE3	DGCEPRE3	916	1.709	2625	0,79	Mau
G_CEP.RE4	DGCEPRE4	237	544	781	0,98	Neutro
G_CEP.RE5	DGCEPRE5	513	1.536	2049	1,27	Bom
G_CEP.RE6	DGCEPRE6	80	339	419	1,80	Muito Bom
G_CEP.RE7	DGCEPRE7	88	718	806	3,47	Excelente
Não informado		6	3	9	-	-
Não classificado		303	705	1.008	-	-
Grupos de CEP Comercial						
G_CEP.CO1	DGCEPCO1	40	38	78	0,40	Péssimo
G_CEP.CO2	DGCEPCO2	328	472	800	0,61	Muito Mau
G_CEP.CO3	DGCEPCO3	520	958	1478	0,78	Mau
G_CEP.CO4	DGCEPCO4	56	140	196	1,06	Neutro
G_CEP.CO5	DGCEPCO5	218	639	857	1,25	Bom
G_CEP.CO6	DGCEPCO6	24	89	113	1,58	Muito Bom
G_CEP.CO7	DGCEPCO7	102	695	797	2,90	Excelente
Não informado		1.228	2.939	4.167	-	-
Não classificado		204	425	629	-	-
Totais		2.720	6.395	9.115	-	-

Como se observa, o agrupamento das variáveis com muitos atributos, segundo a análise do risco relativo, levou a criação de sete grupos, do risco péssimo ao excelente. Cada grupo de risco é transformado em uma variável *dummy* (0 ou 1) que serão, portanto, as variáveis explicativas para a construção dos modelos de classificação. As linhas de 'não informado' referem-se aos clientes que não tinham informação em seu cadastro para aquela

variável, e as linhas de ‘não classificado’ referem-se aos casos em que os atributos das variáveis tinham menos de 30 observações e, portanto, não foram classificados pela pouca representatividade.

Para entender o cálculo do risco relativo (RR), toma-se, por exemplo, o primeiro grupo de profissões em que se obteve RR = 0,39. Este valor é resultado da razão do percentual de bons pelo percentual de maus naquela classe, ou seja, 1,30% de bons (83/6.395) e 3,31% de maus (90/2.720). As relações dos atributos classificados em cada um dos grupos de profissões, cidades de nascimentos, CEP residencial e CEP comercial são apresentadas no Apêndice A, Apêndice B, Apêndice C e Apêndice D, respectivamente.

Para as demais informações do banco de dados também foram criadas variáveis *dummies* e os resultados são apresentados na Tabela 4.

Tabela 4 Criação de variáveis *dummies* para demais variáveis categóricas

VARIÁVEL	***Dummy***	**Mau**	**Bom**	**Total**	**RR**	**Classe de Risco**
Sexo						
Masculino	DSEXOM	972	1.845	2.817	0,81	Mau
Feminino	DSEXOF	1.748	4.550	6.298	1,11	Bom
Escolaridade						
Fundamental	DPRI	1.344	3.206	4.550	1,01	Neutro
Médio	DSEC	1.217	2.631	3.848	0,92	Neutro
Superior	DSUP	159	558	717	1,49	Bom
Estado Civil						
Solteiro	DSOLTE	1.593	2.500	4.093	0,67	Muito Mau
Separado	DSEPARA	74	163	237	0,94	Neutro
Concubinato	DCONCUB	20	45	65	0,96	Neutro
Outros	DOUTR	161	375	536	0,99	Neutro
Divorciado	DDIVOR	148	445	593	1,28	Bom
Casado	DCASADO	585	2.193	2.778	1,59	Muito Bom
Viúvo	DVIUVO	139	674	813	2,06	Excelente
Tem Filho?						
Sim	DFILHO	1.037	2.038	3.075	0,84	Mau
Não	DNFILHO	1.683	4.357	6.040	1,10	Bom
Tipo de Ocupação						
Autônomo	DOCUP_AU	1.122	2.009	3.131	0,76	Mau
Assalariado	DOCUP_AS	1.153	2.350	3.503	0,87	Mau
Profissional liberal	DOCUP_PL	43	101	144	1,00	Neutro
Aposentado	DOCUP_AP	342	1.645	1.987	2,05	Excelente
Funcionário público	DOCUP_FP	60	290	350	2,06	Excelente
Tipo de Residência						
Reside com pais	DRES_PAI	53	62	115	0,50	Péssimo
Alugada	DRES_ALU	325	531	856	0,69	Mau
Outras	DRES_OUT	308	565	873	0,78	Mau
Cedida	DRES_CED	181	339	520	0,80	Mau
Própria	DRES_PRO	1.853	4.898	6.751	1,12	Bom
Totais		2.720	6.395	9.115	-	-

As variáveis numéricas, idade e tempo de serviço, foram segmentadas em classes, também segundo o risco relativo, e os resultados são apresentados na Tabela 5. Considerando os agrupamentos realizados e as segmentações das variáveis numéricas, foram construídas ao

todo 69 variáveis *dummies*, que serão utilizadas como possíveis variáveis explicativas para construção dos modelos de classificação na seção seguinte.

Tabela 5 Categorização e criação de variáveis *dummies* para variáveis numéricas

VARIÁVEL	*Dummy*	Mau	Bom	Total	RR	Classe de Risco
Idade						
até 20 anos	DIDAD1	457	406	863	0,38	Péssimo
21 a 25 anos	DIDAD2	421	656	1.077	0,66	Muito Mau
26 a 30 anos	DIDAD3	438	629	1.067	0,61	Muito Mau
31 a 35 anos	DIDAD4	344	604	948	0,75	Mau
36 a 40 anos	DIDAD5	283	656	939	0,99	Neutro
41 a 50 anos	DIDAD6	411	1.261	1.672	1,30	Bom
51 a 60 anos	DIDAD7	206	1.035	1.241	2,13	Excelente
acima de 60 anos	DIDAD8	143	1.108	1.251	3,26	Excelente
Não informado		17	40	57	-	-
Tempo de Serviço						
até 3 meses	DTSERV1	43	94	137	0,93	Neutro
4 a 6 meses	DTSERV2	94	123	217	0,56	Muito Mau
7 a 18 meses	DTSERV3	262	412	674	0,67	Muito Mau
19 a 24 meses	DTSERV4	75	131	206	0,74	Mau
25 a 36 meses	DTSERV5	105	192	297	0,78	Mau
37 a 60 meses	DTSERV6	79	227	306	1,22	Bom
61 a 90 meses	DTSERV7	39	158	197	1,72	Muito Bom
91 a 120 meses	DTSERV8	13	93	106	3,04	Excelente
acima de 120 meses	DTSERV9	45	264	309	2,50	Excelente
Não informado		1.965	4.701	6.666	-	-
Totais		2.720	6.395	9.115	-	-

A informação do risco relativo calculado já adianta o que esperar para os coeficientes dessas variáveis nos modelos construídos. Por exemplo, a *dummy* DIDAD1 (clientes com idade até 20 anos), por ter classificação de risco péssimo, deverá aparecer com o maior coeficiente negativo, enquanto que a variável DIDAD8 (clientes com idade acima dos 60 anos), por ter classificação de risco excelente, deverá aparecer com o maior coeficiente positivo, dentre as variáveis *dummies* relacionadas à idade.

6.2 MODELOS DE CLASSIFICAÇÃO

Os resultados da construção e avaliação da qualidade dos modelos de classificação, utilizando as técnicas de regressão logística e de rede neural, são apresentados na sequência.

6.2.1 Construção dos Modelos

Para a construção do modelo com a regressão logística, foi utilizado o SPSS versão 18.0 (*Statistical Package for Social Science*). O *software* utilizado para treinamento e teste das redes neurais, empregada para a construção do segundo modelo de classificação, foi o *BrainMaker Professional* versão 3.7.

Nos testes iniciais para construção do modelo logístico, através do método *stepwise*, utilizaram-se níveis de significância para a entrada e saída de variáveis do modelo de 5% e 10%, respectivamente. Para que algumas variáveis tivessem significância para entrar no modelo final, foi necessário fazer o agrupamento de *dummies* próximas, como por exemplo, os grupos de cidade de nascimento 1 e 2 (péssimo e muito mau) e os grupos de profissões 6 e 7 (muito bom e excelente), entre outros.

Com o intuito de obter o escore logístico, conforme apresentado anteriormente na Seção 5.3, na Equação 11, foram avaliadas as 69 variáveis *dummies*, relacionadas como possíveis variáveis explicativas, sendo que apenas 29 delas foram significativas para compor o modelo final, como é apresentado na Equação 13, que retorna a probabilidade de um proponente vir a ser um bom pagador. A especificação das variáveis utilizadas no modelo é apresentada na Figura 21.

$$Y = \frac{1}{1 + \exp(0{,}876 - 0{,}829\,DIDAD1 - 0{,}409\,DIDAD23 - 0{,}252\,DIDAD4 + 0{,}232\,DIDAD6 + 0{,}644\,DIDAD7 + 1{,}047\,DIDAD8 + 0{,}327\,DSEXOF - 0{,}287\,DPRIM + 0{,}270\,DSUP + 0{,}410\,DCASADO + 0{,}340\,DTSERV6 + 0{,}627\,DTSERV7 + 0{,}792\,DTSERV89 - 0{,}293\,DFILHO - 0{,}547\,DRES\_ALU - 0{,}392\,DGCEPR12 - 0{,}172\,DGCEPRE3 + 0{,}197\,DGCEPRE5 + 0{,}328\,DGCEPRE6 + 0{,}608\,DGCEPRE7 - 0{,}768\,DGCEPCO1 + 0{,}218\,DGCEPC56 + 0{,}472\,DGCEPCO7 - 0{,}718\,DGPROF1 - 0{,}318\,DGPROF2 + 0{,}283\,DGPROF67 - 0{,}449\,DCIDNA12 - 0{,}328\,DCIDNA3 + 0{,}592\,DCIDNA7)} \quad (13)$$

Algumas variáveis tiveram que ser agrupadas para que se verificasse a significância estatística e elas fizessem parte do modelo final. A interpretação da equação resultante demonstra a ponderação atribuída a cada variável para a separação dos clientes nos grupos. O sinal dos coeficientes de cada uma das variáveis indica o sentido para a classificação do tipo de cliente, sendo um indicativo de uma característica para um cliente mau pagador o sinal negativo, e de um cliente bom o sinal positivo. Ou seja, um proponente que tem idade acima de 41 anos (DIDADE6), tem curso superior (DSUP), é casado (DCASADO), entre outras informações, tem maior probabilidade de ser um bom pagador. A variável DTSERV89 indica os clientes que possuem tempo de serviço maior do que 90 meses e, uma vez que o sinal também é positivo, o modelo considera que as pessoas que estão em seus empregos há mais tempo possuem menor risco de inadimplência.

Na medida em que as variáveis explicativas estão padronizadas no intervalo 0 e 1, o efeito de cada variável pode ser avaliado diretamente pelo valor absoluto do respectivo coeficiente. Pode-se observar que, de acordo com o modelo construído, as variáveis que exercem maior influência sobre o risco de crédito são DIDAD1, DIDAD8, DTSERV89, DGCEPC01, DGCEPC07 e DGPROF1.

Y = propensão de vir a ser um bom cliente	DRES_ALU = tipo de residência alugada
DIDAD1 = idade até 20 anos	DGCEPR12 = CEP residencial com péssimo ou muito mau desempenho
DIDAD23 = idade entre 21 e 30 anos	DGCEPRE3 = CEP residencial com mau desempenho
DIDAD4 = idade entre 31 e 35 anos	DGCEPRE5 = CEP residencial com bom desempenho
DIDAD6 = idade entre 41 e 50 anos	DGCEPRE6 = CEP residencial com muito bom desempenho
DIDAD7 = idade entre 51 e 60 anos	DGCEPRE7 = CEP residencial com excelente desempenho
DIDAD8 = idade superior a 60 anos	DGCEPCO1 = CEP comercial com péssimo desempenho
DSEXOF = sexo feminino	DGCEPC56 = CEP comercial com bom ou muito bom desempenho
DPRIM = possui escolaridade primária (fundamental)	DGCEPCO7 = CEP comercial com excelente desempenho
DSUP = possui curso superior	DGPROF1 = profissão com péssimo desempenho
DCASADO = é casado	DGPROF2 = profissão com muito mau desempenho
DTSERV6 = tempo de serviço entre 37 e 60 meses	DGPROF67 = profissão com muito bom ou excelente desempenho
DTSERV7 = tempo de serviço entre 61 e 90 meses	DCIDNA12 = cidade de nascimento com péssimo ou muito mau desempenho
DTSERV89 = tempo de serviço superior a 90 meses	DCIDNA3 = cidade de nascimento com mau desempenho
DFILHO = tem filhos	DCIDNA7 = cidade de nascimento com excelente desempenho

Figura 21 Especificação das variáveis utilizadas no modelo

No que diz respeito à verificação do atendimento de suposições para utilização da técnica, somente a baixa multicolinearidade deve ser confirmada. Com a utilização do método *stepwise* para escolha das variáveis explicativas para compor o modelo, garantiu-se o atendimento deste pressuposto.

O *software* utilizado para a construção das redes neurais não tem nenhum método de seleção de variáveis como, por exemplo, o *stepwise* que seleciona automaticamente as variáveis com maior poder de explicação. Portanto, utilizou-se, para o estudo das redes, as variáveis que foram indicadas previamente através do modelo logístico.

Para a obtenção das redes neurais utilizou-se a função de ativação sigmóide e o algoritmo de aprendizado supervisionado de retropropagação de erro, com somente uma camada oculta. Várias redes foram criadas com diferentes quantidades de neurônios nesta camada para verificar o desempenho quanto à predição dos bons e maus clientes. Para avaliar o desempenho das redes compararam-se os valores do teste de KS para as amostras de análise e de teste. Os resultados das redes construídas são apresentados na Tabela 6.

Tabela 6 Comparação dos melhores modelos neurais construídos

Modelo	Nº Neurônios camada oculta	KS	
		Análise	Teste
RN1	35	38,3	34,0
RN2	30	39,2	33,9
RN3	30	39,3	35,1
RN4	35	41,0	32,7
RN5	27	40,3	34,6
RN6	35	40,3	35,4

O modelo neural escolhido, com melhor desempenho para previsão do risco de crédito, foi o modelo RN6. O modelo RN4 teve melhores resultados para a amostra de análise, porém o desempenho na amostra de teste é bem menor, evidenciando o excesso de encaixe da rede. Os pesos obtidos para os neurônios da camada oculta e da camada de saída, para obtenção do escore neural, são apresentados no Apêndice E.

Uma avaliação dos modelos de classificação é apresentada na Figura 22 e na Figura 23 com a distribuição dos bons e maus pagadores e a taxa de sinistro, que corresponde ao percentual de maus pagadores sobre o total de clientes.

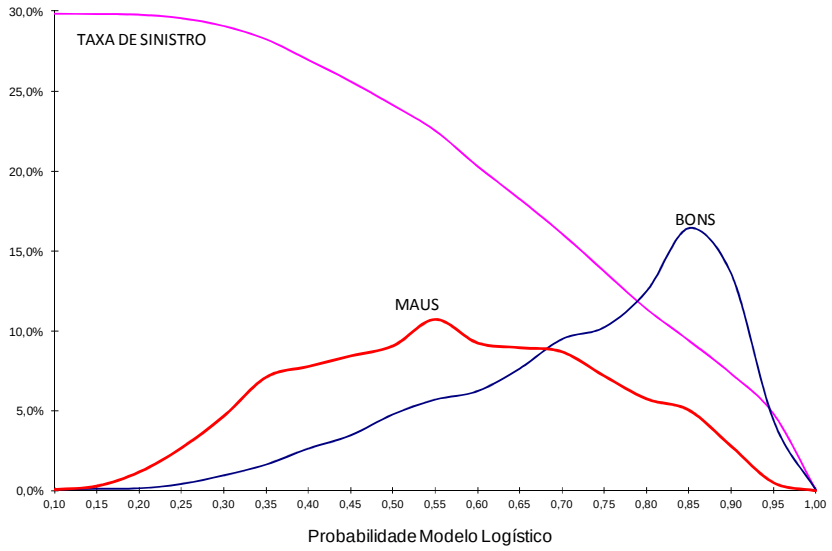

Figura 22 Distribuição dos clientes e taxa de sinistro com modelo logístico

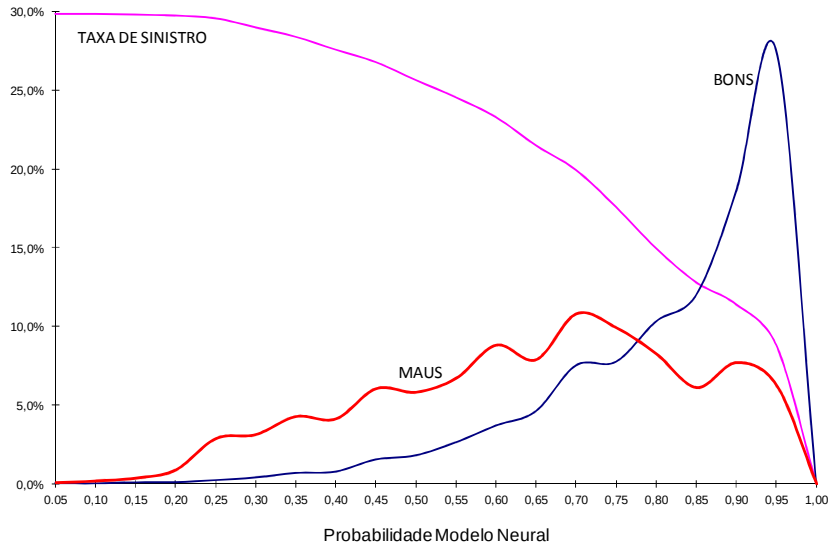

Figura 23 Distribuição dos clientes e taxa de sinistro com modelo neural

Analisando o comportamento das curvas de distribuição dos bons e maus pagadores, verifica-se que os modelos conseguem separar razoavelmente os dois grupos de clientes, já que é possível observar a tendência de que os maus se concentram à esquerda da escala e os bons à direita. O resultado desta separação é visualizado também com a queda na taxa de sinistro conforme se avança na direção dos maiores escores, partindo de 30% e chegando a quase 0%, nos maiores valores de probabilidade. É importante salientar que a separação obtida entre as duas curvas não é tão expressiva e que tal resultado é devido ao número e tipo de variáveis utilizadas na análise. Incorporar outras variáveis em uma avaliação futura do modelo deve melhorar o afastamento dos perfis.

6.2.2 **Qualidade dos Modelos**

Um meio adequado de verificar o poder de previsão dos modelos construídos é a medição do percentual de classificações corretas. Para avaliar a classificação dos clientes é necessário definir um ponto de corte na escala de escores, acima do qual os clientes são classificados como bons e, abaixo do qual, como maus pagadores. Utilizando-se um ponto de corte de 0,5 (ou seja, se a probabilidade de pagamento for menor que 50%, não é concedido o crédito) tem-se, na Tabela 7, a matriz das classificações dos clientes nos grupos originais e sua comparação com as classificações segundo o modelo logístico, e na Tabela 8 a matriz de classificação para o modelo neural, para as amostras de análise e teste.

Tabela 7 Percentuais de acerto do modelo logístico

AMOSTRA	GRUPOS ORIGINAIS	CLASSIFICAÇÃO		
		MAU	**BOM**	**TOTAL**
ANÁLISE	**MAU**	875 (32,2%)	1.845 (67,8%)	2.720
	BOM	591 (9,2%)	5.804 (90,8%)	6.395
	TOTAL	**1.466 (16,1%)**	**7.649 (83,9%)**	**9.115**
TESTE	**MAU**	180 (27,8%)	468 (72,2%)	6.48
	BOM	165 (10,1%)	1.466 (89,9%)	1.631
	TOTAL	**345 (15,1%)**	**1.934 (84,9%)**	**2.279**

O percentual de acerto total do modelo logístico foi de 73,3% para a amostra de análise, com 875 maus e 5.804 bons corretamente classificados. Na amostra de teste esse percentual foi de 72,2%, sendo 180 maus e 1.466 bons corretamente classificados. Os percentuais de aprovação para as duas amostras também foram bem próximos, com 83,9% para a amostra de análise e de 84,9% para a amostra de teste.

Tabela 8 Percentuais de acerto do modelo neural

AMOSTRA	GRUPOS ORIGINAIS	CLASSIFICAÇÃO		
		MAU	**BOM**	**TOTAL**
ANÁLISE	**MAU**	936 (34,4%)	1.784 (65,6%)	2.720
	BOM	515 (8,1%)	5.880 (91,9%)	6.395
	TOTAL	**1.451 (15,9%)**	**7.664 (84,1%)**	**9.115**
TESTE	**MAU**	240 (37,0%)	408 (63,0%)	648
	BOM	215 (13,2%)	1.416 (86,8%)	1.631
	TOTAL	**455 (20,0%)**	**1.824 (80,0%)**	**2.279**

Para o modelo neural, o percentual de acerto total na classificação para a amostra de análise foi de 74,8%, com 936 maus e 5.880 bons corretamente classificados, chegando a uma aprovação de 84,1% do total de clientes. Na amostra de teste o percentual de classificações corretas foi de 72,7%, sendo 240 maus e 1.416 bons corretamente classificados, atingindo uma aprovação total de 80%.

Outra forma de avaliar a qualidade de um modelo de classificação é através do teste não-paramétrico de Kolmogorov-Smirnov (KS). Este teste tem por objetivo determinar se duas amostras provêm de uma mesma população. No presente caso, espera-se provar que as duas amostras de clientes (bons e maus) provêm de populações distintas o que significaria que os modelos conseguem separar os dois grupos. A representação dos valores do teste de KS para os modelos construídos é apresentada na Figura 24.

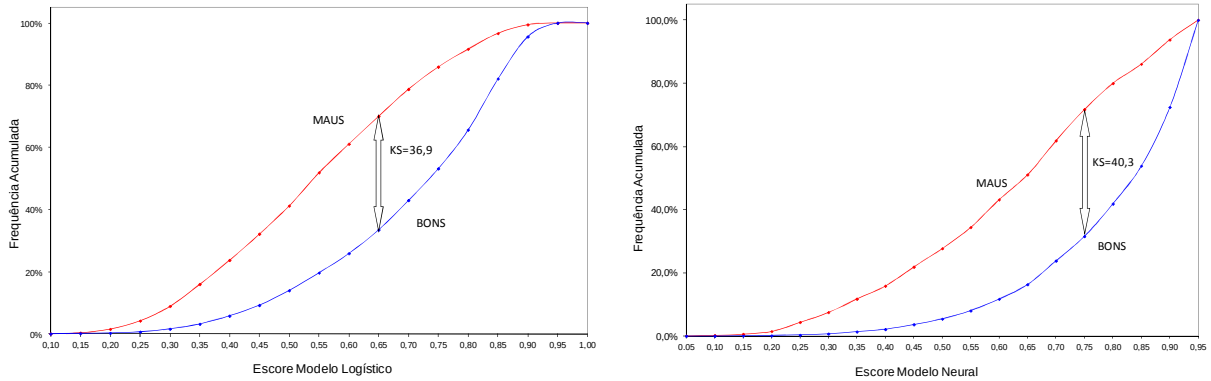

Figura 24 Representação do valor de KS para os modelos construídos

Através dos gráficos é possível visualizar o grau de separação entre os clientes maus e bons através dos escores dos modelos. Na Tabela 9 é apresentado um resumo com as duas medidas de qualidade utilizadas para comparar os resultados obtidos com os dois modelos de classificação construídos, para as amostras de análise e de teste.

Tabela 9 Medidas de desempenho dos modelos construídos

	Percentual de Acerto		**Valor de KS**	
Modelo	Análise	Teste	Análise	Teste
Logístico	73,3	72,2	36,9	31,7
Neural	74,8	72,7	40,3	35,4

Tanto na amostra de análise como na amostra de teste, os percentuais de acerto total encontrados para os modelos são superiores a 65% e os valores para o teste KS são maiores que 30, valores mínimos para considerar um modelo com bom poder de separação (PICININI *et al.*, 2003). Ao analisar os valores das duas medidas, observa-se que os resultados obtidos com o modelo neural foram ligeiramente superiores ao obtido com a regressão logística, o que pode ser explicado por sua abordagem diferenciada no relacionamento das variáveis.

6.3 **MODELO DE PREVISÃO DO RISCO MONETÁRIO**

Neste momento o interesse é a construção do modelo final que servirá para fazer a previsão do lucro médio (variável resposta) com os clientes em função dos escores (variáveis explicativas) previstos pelos modelos de classificação obtidos com regressão logística e rede neural. Além de prever o lucro médio, o objetivo com este modelo também é que essa previsão sirva como um balizador para a concessão do crédito, fornecendo uma informação de risco monetário esperado com cada cliente.

Portanto, para a obtenção da variável resposta, o lucro bruto, é necessária a determinação da margem bruta. A empresa estima sua margem na ordem de 30%, de forma que o lucro bruto foi calculado a partir da Equação 14.

$$\text{Lucro bruto} = 0{,}3(\text{valor de compras} - \text{saldo sinistrado}) \qquad (14)$$

Os valores de compras e atrasos são observados no período de 12 a 18 meses após a concessão do crédito, que é o período da amostragem, descrito na etapa de seleção da amostra. O valor das compras representa a soma de tudo que foi comprado por cada cliente desde sua concessão até a extração da amostra. Já o saldo sinistrado denota a parte do valor das compras que não foi paga e estava em atraso há mais de 90 dias, também no momento da seleção da amostra. Somente foi considerado como saldo sinistrado os atrasos superiores a 90 dias em função da definição da empresa do que seria considerado um cliente mau pagador.

Quanto ao tamanho da amostra recomendado para utilização da regressão linear múltipla, como são apenas duas variáveis explicativas, qualquer recomendação de tamanho

mínimo para a amostra é atendida com folga. A mesma partição em amostra de análise e teste, utilizada para criação dos modelos de classificação com regressão logística e rede neural, é conservada. Ao avaliar os resultados pretende-se validar as previsões encontradas com a amostra de análise, que foi utilizada para a construção do modelo, com a amostra de teste para verificar se o poder de previsão é mantido com indivíduos que não foram considerados inicialmente. Essa validação com a amostra de teste serve, portanto, para dar indicativos de como o modelo de previsão do lucro se comportará quando for utilizado na prática.

De forma a avaliar o ganho de lucro da empresa com a utilização do modelo proposto, serão comparados os valores para a variável lucro médio em quatro cenários: (i) sem utilizar nenhum modelo de previsão de risco de crédito; (ii) utilizando o modelo de classificação obtido com a regressão logística; (iii) utilizando o modelo de classificação obtido com a rede neural; e (iv) utilizando o modelo proposto para previsão do risco monetário. As medidas descritivas para o primeiro caso, sem utilização de nenhum modelo de previsão, são apresentadas na Tabela 10.

Tabela 10 Medidas de lucro bruto para os grupos de clientes observados

Amostra	Tipo	N	Média	Desvio-padrão	Soma
Análise	Mau	2.720	-210,74	197,99	-573.219,17
	Bom	6.395	139,93	131,71	894.843,79
	Total	**9.115**	**35,29**	**222,74**	**321.624,63**
Teste	Mau	648	-212,18	210,74	-137.491,57
	Bom	1.631	135,54	127,68	221.059,07
	Total	**2.279**	**36,67**	**221,12**	**83.567,50**
Total	Mau	3.368	-211,02	200,47	-710.710,74
	Bom	8.026	139,04	130,91	1.115.902,86
	Total	**11.394**	**35,56**	**222,41**	**405.192,13**

Sem a utilização de nenhum modelo de previsão de risco de crédito, a empresa concedeu crédito para todos os proponentes e, desta forma, teve um lucro médio por cliente no valor de R$35,56. Com cada cliente mau pagador perdeu em média R$211,02 e com os bons pagadores ganhou em média R$139,04. Portanto, o valor total de ganho com os 11.394 clientes foi de R$ 405.192,13.

Considerando que a empresa adotasse um modelo de classificação obtido com a regressão logística ou com a rede neural seriam observadas as medidas apresentadas na Tabela 11 e Tabela 12, respectivamente.

Tabela 11 Medidas de lucro para os grupos de clientes previstos pela regressão logística

Amostra	Decisão	N	Média	Desvio-padrão	Soma
Análise	Negar	1.466	-74,39	244,01	-109.061,20
	Conceder	**7.649**	**56,31**	**212,06**	**430.685,83**
	Total	9.115	35,29	222,74	321.624,63
Teste	Negar	345	-36,89	220,11	-12.726,21
	Conceder	**1.934**	**49,79**	**218,77**	**96.293,71**
	Total	2.279	36,67	221,12	83.567,50
Total	Negar	1.811	-67,25	240,03	-121.787,41
	Conceder	**9.583**	**54,99**	**213,44**	**526.979,54**
	Total	11.394	35,56	222,41	405.192,13

Assim, utilizando um modelo de previsão de risco que classifica os clientes, como o logístico, por exemplo, a empresa esperaria um lucro por cliente no valor de R$54,99, e aprovando 9.583 clientes do total de 11.394 (84,1%) o ganho total seria de R$526.979,54.

Tabela 12 Medidas de lucro para os grupos de clientes previstos pela rede neural

Amostra	Decisão	N	Média	Desvio-padrão	Soma
Análise	Negar	1.451	-90,34	244,50	-131.087,32
	Conceder	**7.664**	**59,07**	**210,10**	**452.711,94**
	Total	9.115	35,29	222,74	321.624,63
Teste	Negar	345	-67,43	235,69	-23.261,76
	Conceder	**1.934**	**55,24**	**213,20**	**106.829,25**
	Total	2.279	36,67	221,12	83.567,50
Total	Negar	1.796	-85,94	242,93	-154.349,07
	Conceder	**9.598**	**58,30**	**210,72**	**559.541,20**
	Total	11.394	35,56	222,41	405.192,13

Com a utilização do modelo de classificação neural o lucro esperado por cliente já seria um pouco melhor, como já era esperado em função de seu maior acerto na predição em comparação com o modelo logístico. O lucro esperado por cliente seria de R$58,30, e aprovando 9.598 clientes do total de 11.394 (84,2%) o ganho total seria de R$559.541,20.

Seguindo com a construção do modelo proposto para previsão do lucro, foi realizada a avaliação do comportamento das variáveis explicativas para verificar a necessidade de alguma transformação nos dados ou de criação de alguma variável adicional. Foram utilizados os escores previstos com os modelos de classificação com regressão logística e rede neural e não as classificações, observando-se um relacionamento aproximadamente linear entre as variáveis explicativas (escores) e a variável resposta (lucro bruto). Desta forma, não houve

necessidade de aplicação de qualquer tipo de transformação nos dados ou de criação de variável *dummy* para trabalhar com problemas de não-linearidade.

O modelo obtido teve significância estatística ($F = 520,377$; $p < 0,001$), porém o valor do coeficiente de determinação foi baixo ($R^2 = 0,103$), que indica a porcentagem de variação total da variável resposta (lucro bruto) explicada pelo modelo. O motivo de um valor de explicação tão baixo já era, em parte, esperado, como já foi na construção dos modelos de classificação com regressão logística e rede neural, em função das variáveis de perfis consideradas (sexo, idade, escolaridade, etc.) não terem um grande poder de predição da condição de pagamento dos clientes.

O valor dos coeficientes do modelo, conforme proposto anteriormente na Equação 10, e os testes para sua significância são apresentados na Tabela 13.

Tabela 13 Coeficientes do modelo de previsão do lucro

Variáveis	B	Erro padrão	Beta	t	Significância
Constante	-273,904	9,835		-27,849	< 0,001
Escore Logístico	85,668	23,859	0,070	3,591	< 0,001
Escore Neural	315,789	23,750	0,258	13,297	< 0,001

Assim a função obtida para previsão do lucro médio é a representada na Equação 15.

$$E(\text{Lucro}) = -273,904 + 85,668\ \text{Escore Logístico} + 315,789\ \text{Escore Neural} \quad (15)$$

Analisando os coeficientes obtidos, verifica-se que a cada ponto percentual obtido na probabilidade de pagamento com o modelo logístico espera-se um aumento marginal de R$0,85 no lucro médio com o cliente. Da mesma forma, a cada ponto percentual obtido na probabilidade de pagamento com o modelo neural espera-se um aumento marginal de R$3,16 no lucro com o cliente. O coeficiente para o escore neural foi maior que o coeficiente para o escore logístico em função do erro de predição menor obtido com as redes neurais, como observado na Seção 6.2.2, em que foram avaliadas as medidas de qualidade dos modelos de classificação construídos, indo também ao encontro do que propõem Bates e Granger (1969).

As suposições para utilização da técnica de regressão linear múltipla (linearidade, variância constante, independência e normalidade da distribuição dos erros) foram verificadas e atendidas de forma aproximada, em função do tamanho da amostra utilizada. Quanto à identificação de informações fora dos padrões dos dados, uma observação foi retirada da análise por ter um valor discrepante para a variável resposta, representando um lucro muito alto em relação ao restante dos clientes.

Obtido o modelo e os valores previstos para o lucro médio a partir da aplicação da equação de regressão estimada, parte-se para sua avaliação e validação. Na Tabela 14 são apresentadas as distribuições dos clientes bons, maus e total de acordo com o lucro previsto pelo modelo e também as taxas de sinistro e aprovação.

Tabela 14 Distribuição dos clientes e taxa de sinistro do modelo de previsão do lucro

Lucro médio	Maus			Bons			Total			Taxa de sinistro		Taxa de
Esperado (R$)	#	%	% acum.	#	%	% acum.	#	%	% acum.	Classe	Total	Aprovação
até -200	8	0,29	0,29	3	0,05	0,05	11	0,12	0,12	72,73	29,84	100,00
de -200 a -180	8	0,29	0,58	5	0,08	0,13	13	0,14	0,26	61,54	29,79	99,88
de -180 a -160	53	1,95	2,53	8	0,13	0,26	61	0,67	0,93	86,89	29,74	99,74
de -160 a -140	91	3,35	5,88	17	0,27	0,53	108	1,18	2,11	84,26	29,36	99,07
de -140 a -120	89	3,27	9,15	31	0,48	1,01	120	1,32	3,43	74,17	28,69	97,88
de -120 a -100	157	5,77	14,92	65	1,02	2,03	222	2,44	5,87	70,72	28,07	96,57
de -100 a -80	163	5,99	20,91	94	1,47	3,50	257	2,82	8,69	63,42	26,97	94,13
de -80 a -60	170	6,25	27,16	111	1,74	5,24	281	3,08	11,77	60,50	25,84	91,31
de -60 a -40	227	8,35	35,51	194	3,03	8,27	421	4,62	16,39	53,92	24,63	88,23
de -40 a -20	217	7,98	43,49	246	3,85	12,12	463	5,08	21,47	46,87	23,02	83,61
de -20 a 0	220	8,09	51,58	312	4,88	17,00	532	5,84	27,31	41,35	21,47	78,53
de 0 a 20	334	12,28	63,86	494	7,72	24,72	828	9,08	36,39	40,34	19,88	72,69
de 20 a 40	255	9,38	73,24	568	8,88	33,60	823	9,03	45,42	30,98	16,95	63,61
de 40 a 60	203	7,46	80,70	579	9,05	42,65	782	8,58	54,00	25,96	14,63	54,58
de 60 a 80	183	6,73	87,43	793	12,40	55,05	976	10,71	64,71	18,75	12,52	46,00
de 80 a 100	195	7,17	94,60	1.119	17,50	72,55	1.314	14,42	79,13	14,84	10,63	35,29
de 100 a 120	141	5,18	99,78	1.589	24,85	97,40	1.730	18,98	98,11	8,15	7,72	20,88
acima de 120	6	0,22	100,00	167	2,61	100,00	173	1,90	100,00	3,47	3,47	1,90
TOTAL	2.720	100,00	-	6.395	100,00	-	9.115	100,00	-	-	-	-

A coluna da taxa de sinistro total informa o percentual de maus clientes que seriam aprovados sobre o total de clientes aprovados se fosse utilizado como corte o limite inferior da classe. Por exemplo, se fosse utilizado como corte o valor de lucro 0 (zero), esperar-se-ia 19,88% de maus clientes no total aprovado. A taxa de sinistro na classe informa ainda que na classe de lucro entre 0 e 20 têm 40,34% de clientes maus pagadores. O que se observa é um decréscimo nesta taxa que é explicado pelo poder de explicação do modelo, e pode também ser observado na Figura 25.

A última coluna da Tabela 14 é a taxa de aprovação, ou seja, se fosse utilizado como corte a aprovação somente de clientes com lucro esperado positivo, 72,69% dos proponentes teriam o crédito aprovado. Deste total, aproximadamente 52% dos maus clientes teriam o crédito negado e 83% dos bons seriam aprovados.

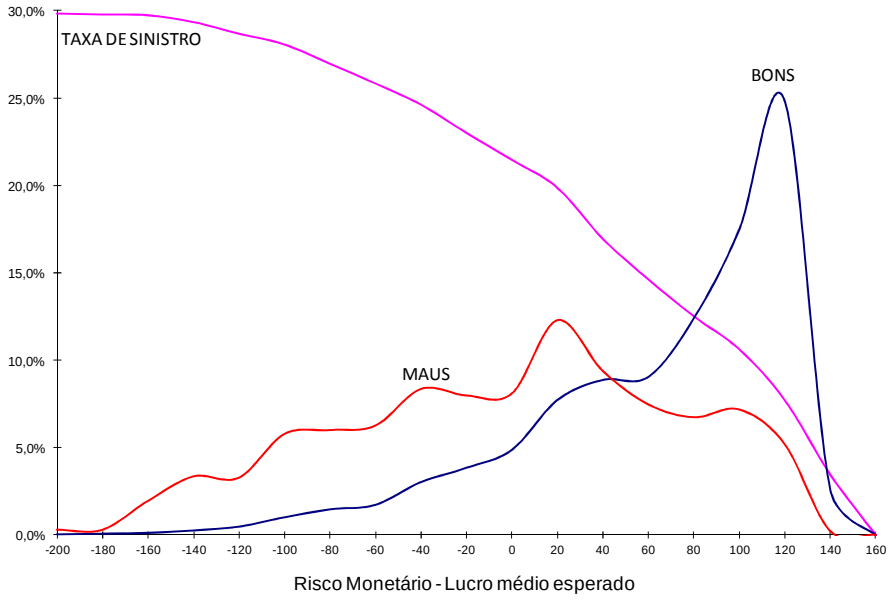

Figura 25 Distribuição dos clientes e taxa de sinistro com modelo de previsão do lucro

A distribuição dos escores do modelo permite identificar a tendência de que os clientes maus tendem a ter valores de lucro médio esperado negativo enquanto que os bons tendem a ter valores de lucro esperado positivos. O resultado desta separação é visualizado com a queda na taxa de sinistro conforme se avança na direção dos maiores lucros. É importante salientar que a separação obtida entre as duas curvas não é tão expressiva, como já foi evidenciado pelo baixo valor obtido para o coeficiente de determinação (R^2) e que tal resultado pode ser devido ao número e tipo de variáveis utilizadas na análise. Incorporar outras possíveis variáveis em uma avaliação futura do modelo pode melhorar a separação dos perfis.

Na Tabela 15 é apresentada uma análise dos resultados de lucro previsto pelo modelo em comparação com o lucro total observado com cada grupo de clientes, na amostra de análise. Analisando os resultados, observa-se que, com os 2.720 clientes maus pagadores, a empresa obteve um prejuízo de R$573.219,17 e com os 6.395 clientes bons pagadores obteve um lucro de R$894.843,80 resultando que ao todo o lucro da empresa com esses clientes foi de R$321.624,63. Avaliando os valores de lucro médio previsto, apresentados na primeira coluna, a empresa poderia agora decidir com base nesta medida monetária de risco e com isso, só conceder, por exemplo, quando o lucro esperado for positivo. Assim a taxa de aprovação seria de 72,69%, como já visto anteriormente na análise da Tabela 14, aumentando o lucro para R$ 466.812,83, considerando os dados da amostra de análise.

Tabela 15 Análise comparativa do lucro esperado com lucro observado

Lucro médio	Maus		Bons		Total		Aprovação Acumulada	
Esperado (R$)	#	R$	#	R$	#	R$	%	R$
até -200	8	-1.137,12	3	792,83	11	-344,29	100,00%	321.624,63
de -200 a -180	8	-1.609,86	5	700,35	13	-909,51	99,88%	321.968,92
de -180 a -160	53	-11.962,87	8	1.283,82	61	-10.679,05	99,74%	322.878,43
de -160 a -140	91	-20.258,74	17	2.824,40	108	-17.434,34	99,07%	333.557,48
de -140 a -120	89	-20.295,81	31	4.738,61	120	-15.557,21	97,88%	350.991,82
de -120 a -100	157	-37.486,75	65	7.877,97	222	-29.608,77	96,57%	366.549,02
de -100 a -80	163	-32.560,58	94	14.088,69	257	-18.471,89	94,13%	396.157,80
de -80 a -60	170	-38.356,04	111	16.231,80	281	-22.124,24	91,31%	414.629,69
de -60 a -40	227	-48.568,45	194	26.647,90	421	-21.920,55	88,23%	436.753,93
de -40 a -20	217	-38.723,72	246	34.701,27	463	-4.022,45	83,61%	458.674,48
de -20 a 0	220	-45.773,65	312	41.657,75	532	-4.115,90	78,53%	462.696,93
de 0 a 20	334	-75.722,87	494	66.508,71	828	-9.214,16	72,69%	466.812,83
de 20 a 40	255	-53.555,51	568	80.841,39	823	27.285,88	63,61%	476.026,99
de 40 a 60	203	-40.292,27	579	80.411,69	782	40.119,42	54,58%	448.741,11
de 60 a 80	183	-38.466,65	793	112.349,04	976	73.882,39	46,00%	408.621,69
de 80 a 100	195	-40.683,14	1.119	163.109,42	1.314	122.426,28	35,29%	334.739,30
de 100 a 120	141	-26.721,37	1.589	218.921,76	1.730	192.200,39	20,88%	212.313,02
acima de 120	6	-1.043,78	167	21.156,40	173	20.112,62	1,90%	20.112,62
TOTAL	2.720	-573.219,17	6.395	894.843,80	9.115	321.624,63	--	--

Observa-se que o máximo lucro (R$ 476.026,99) é obtido na faixa de 20 a 40, sinalizando um possível ponto de corte a ser adotado pela empresa. Este ponto de máximo é também observado na Figura 26, que apresenta o comparativo de prejuízos e lucros totais esperados com a aprovação, considerando cada faixa de previsão do modelo.

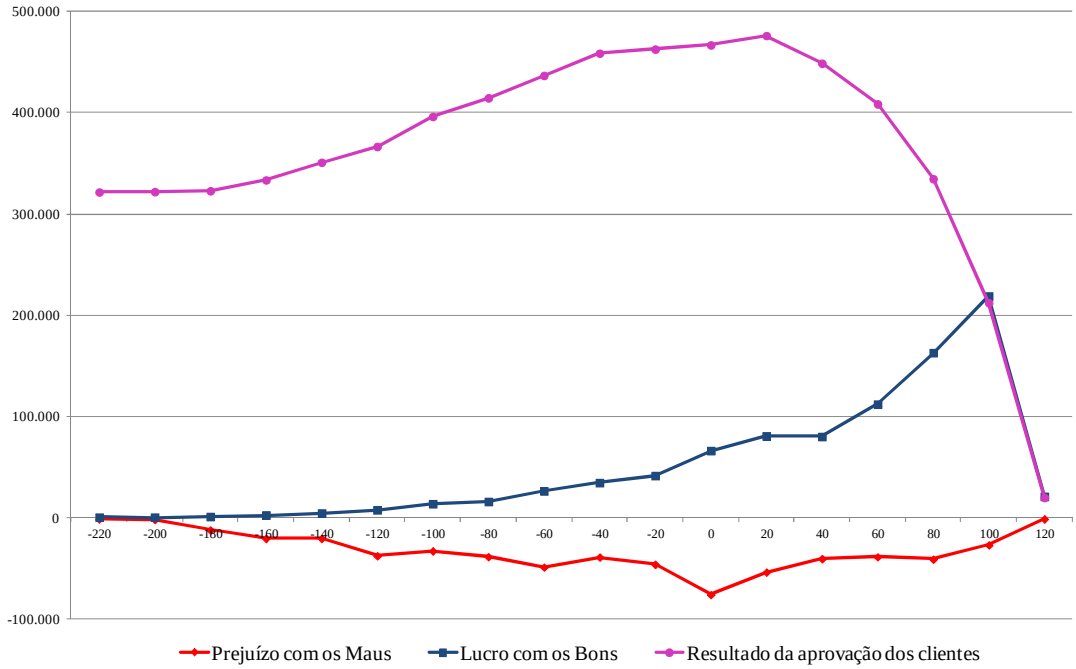

Figura 26 Prejuízos e lucros em cada grupo e total com a aprovação de crédito

Considerando que o valor previsto pelo modelo é o lucro médio esperado com cada cliente, quando o modelo indicar uma expectativa de lucro negativo (prejuízo), a decisão seria, por exemplo, a não concessão do crédito. Assim, as medidas descritivas para o lucro bruto observado para as duas amostras, de análise e de teste, são apresentadas na Tabela 16.

Tabela 16 Medidas de lucro para os grupos de clientes previstos pelo modelo proposto

Amostra	Decisão	N	Média	Desvio-padrão	Soma
Análise	Negar	2.489	-58,33	237,25	-145.188,20
	Conceder	**6.626**	**70,45**	**206,36**	**466.812,83**
	Total	9.115	35,29	222,74	321.624,63
Teste	Negar	600	-42,40	234,06	-25.442,38
	Conceder	**1.679**	**64,93**	**209,25**	**109.009,88**
	Total	2.279	36,67	221,12	83.567,50
Total	Negar	3.089	-55,24	236,68	-170.630,58
	Conceder	**8.305**	**69,33**	**206,95**	**575.822,71**
	Total	11.394	35,56	222,41	405.192,13

Observa-se, portanto, que, com a utilização do modelo proposto para previsão do risco monetário, se conseguiria um ganho maior (R$575.822,71 em comparação com R$405.192,13 sem a utilização do modelo). Considerando que só seriam aprovados os clientes que tenham previsão de lucro médio positivo pelo modelo, somente 8.305 (72,9%) dos 11.394 clientes da amostra teriam sido aprovados. Com isso, o lucro esperado por cliente seria de R$69,33, e o ganho total seria de R$575.822,71, maior que o obtido com os modelos de classificação construídos com a regressão logística e rede neural.

Além deste ganho adicional, deve-se levar em consideração que haveria uma economia no que diz respeito aos custos de concessão, pois a empresa teria conseguido um lucro maior concedendo para menos clientes, 8.305 contra 9.583 do modelo logístico e 9.598 do modelo neural. Por outro lado, se fosse possível a empresa intensificar sua captação de forma a obter candidatos ao crédito com perfil de clientes lucrativos e, com isso, conceder para a mesma quantidade de clientes (aproximadamente 9.500 clientes, o que equivale a 84% da base de dados), se esperaria um lucro total na ordem de R$658.635,00. Desta forma, conseguir-se-ia aproximadamente R$100.000,00 a mais de lucro com os mesmos clientes previstos com qualquer um dos modelos de classificação e pelo menos R$150.000,00 a mais de lucro quando comparado com o cenário sem o uso de nenhum modelo de previsão.

Um resumo das medidas de lucro estimadas para os quatro cenários que foram comparados é apresentado na Tabela 17. Avaliando estes ganhos em termos relativos,

verifica-se que, utilizando apenas o modelo de classificação logístico, o aumento estimado no lucro foi de 54,64% e, utilizando o modelo de classificação neural, o lucro estimado teve um aumento de 63,95% em relação ao cenário sem utilização de nenhum método para previsão de risco de crédito. Já utilizando o modelo proposto para previsão de risco monetário, o aumento estimado foi de 94,97%, em termos de lucro médio.

Tabela 17 Resumo das medidas de lucro para os quatro cenários

Ganho Estimado	Sem modelo de previsão	Classificador Logístico	Classificador Neural	Modelo para Risco Monetário
Média (R$)	35,56	54,99	58,30	69,33
Clientes aprovados	11.394	9.583	9.598	8.305
Total (R$)	405.192,13	526.979,54	559.541,20	575.822,71

Uma última questão importante quanto à validação do modelo diz respeito a observar o comportamento dos escores nas duas amostras (análise e teste) e fazer a verificação de adequação do modelo encontrado. Espera-se que as distribuições das previsões de lucro nas duas amostras não sejam diferentes. Para testar essa hipótese utiliza-se o mesmo teste KS para duas amostras que foi útil para verificar a qualidade dos modelos de classificação com regressão logística e rede neural. Os resultados desta comparação são apresentados na Tabela 18.

Tabela 18 Validação do modelo proposto, comparando amostra de análise e teste

Lucro médio esperado	Amostra de análise				Amostra de teste				Teste KS
	Total		Acumulado		Total		Acumulado		
R$	#	%	#	%	#	%	#	%	%
até -200	11	0,12%	11	0,12%	1	0,04%	1	0,04%	-0,08%
de -200 a -180	13	0,14%	24	0,26%	3	0,13%	4	0,18%	-0,09%
de -180 a -160	61	0,67%	85	0,93%	7	0,31%	11	0,48%	-0,45%
de -160 a -140	108	1,18%	193	2,12%	19	0,83%	30	1,32%	-0,80%
de -140 a -120	120	1,32%	313	3,43%	34	1,49%	64	2,81%	-0,63%
de -120 a -100	222	2,44%	535	5,87%	57	2,50%	121	5,31%	-0,56%
de -100 a -80	257	2,82%	792	8,69%	54	2,37%	175	7,68%	-1,01%
de -80 a -60	281	3,08%	1.073	11,77%	78	3,42%	253	11,10%	-0,67%
de -60 a -40	421	4,62%	1.494	16,39%	107	4,70%	360	15,80%	-0,59%
de -40 a -20	463	5,08%	1.957	21,47%	102	4,48%	462	20,27%	-1,20%
de -20 a 0	532	5,84%	2.489	27,31%	138	6,06%	600	26,33%	-0,98%
de 0 a 20	828	9,08%	3.317	36,39%	200	8,78%	800	35,10%	-1,29%
de 20 a 40	823	9,03%	4.140	45,42%	220	9,65%	1.020	44,76%	-0,66%
de 40 a 60	782	8,58%	4.922	54,00%	196	8,60%	1.216	53,36%	-0,64%
de 60 a 80	976	10,71	5.898	64,71%	248	10,88	1.464	64,24%	-0,47%
de 80 a 100	1.314	14,42	7.212	79,12%	333	14,61	1.797	78,85%	-0,27%
de 100 a 120	1.730	18,98	8.942	98,10%	440	19,31	2.237	98,16%	0,06%
acima de 120	173	1,90%	9.115	100,00	42	1,84%	2.279	100,00	0,00%
TOTAL	9.115	100%	-	-	2.279	100%	-	-	1,29%

A avaliação dos percentuais de clientes em cada classe de lucro revela uma distribuição dos valores previstos muito semelhante nas duas amostras, já que as diferenças são pequenas. O reflexo disso está no baixo valor de KS encontrado (1,29) que é menor que o valor crítico do teste (3,19), o que leva a não rejeitar a hipótese de igualdade das distribuições. Com isso, o modelo construído a partir da amostra de análise pode ser utilizado para previsão, pois não apresentou diferenças significativas quando utilizado na amostra de teste.

De posse da estimativa de lucro médio dada pelo modelo proposto e, sabendo que a margem bruta admitida pela empresa é de 30%, pode-se obter uma sugestão para o limite de crédito a ser atribuído para os clientes que tiverem o crédito concedido. O cálculo do limite sugerido é obtido pela Equação 16.

$$\text{Limite sugerido} = \frac{\text{E(Lucro)}}{0{,}3} \tag{16}$$

Na Tabela 19 são apresentados os resultados para os valores de limite sugerido (médio e total) em cada faixa de lucro estimado pelo modelo comparando com os valores reais atribuídos pela empresa.

Tabela 19 Limite atribuído e limite sugerido pelo modelo

Lucro médio esperado (R$)	Clientes	Limite atribuído		Limite sugerido	
		Médio	Total	Médio	Total
de 0 a 20	1.028	157,88	162.300,00	32,39	33.294,93
de 20 a 40	1.043	155,71	162.410,00	102,88	107.305,73
de 40 a 60	978	159,31	155.805,00	168,20	164.499,81
de 60 a 80	1.224	166,58	203.890,00	234,93	287.548,85
de 80 a 100	1.647	175,51	289.060,00	304,64	501.738,44
de 100 a 120	2.170	175,93	381.760,00	365,97	794.152,40
acima de 120	215	206,00	44.290,00	406,93	87.490,97
Total	8.305	168,51	1.399.515,00	237,93	1.976.031,12

Os resultados demonstram que os limites médios atuais concedidos aos clientes são praticamente os mesmos para todas as faixas de lucro médio esperado. Adotando o método sugerido para atribuição de limite, esta disparidade foi alterada, tornando os limites dos clientes mais alinhados com o seu risco monetário previsto. Dessa forma, a utilização do modelo proposto proporcionou uma melhor adequação do limite de crédito ao perfil de risco do cliente, proporcionando ao decisor uma métrica para decisão objetiva quanto ao limite a ser concedido.

6.4 ANÁLISE DE SENSIBILIDADE DO MODELO

Para avaliar o comportamento do modelo em outras situações, foi realizada uma análise de sensibilidade, alterando um parâmetro e verificando seu comportamento quanto à previsão do lucro médio. Foram adotados diversos valores para a margem bruta (de 5% a 60%) para comparar com o valor considerado pela empresa em estudo (30%). Os resultados obtidos são apresentados na Tabela 20.

Tabela 20 Medidas de lucro para os quatro cenários variando a margem bruta

Margem Bruta	Ganho Estimado	Sem modelo de previsão	Classificador Logístico	Classificador Neural	Modelo para Risco Monetário
5%	Média (R$)	-64,71	-48,4802	-45,567	4,83
	Clientes aprovados	11.394	9.583	9.598	1.032
	Total (R$)	-737.269,96	-464.586,13	-437.352,38	4.980,74
10%	Média (R$)	-44,65	-27,79	-24,7941	14,86
	Clientes aprovados	11.394	9.583	9.598	3.689
	Total (R$)	-508.777,54	-266.272,99	-237.973,67	54.835,9
20%	Média (R$)	-4,55	13,60	16,75	45,70
	Clientes aprovados	11.394	9.583	9.598	6.226
	Total (R$)	-51.792,71	130.353,27	160.783,77	284.553,97
30%	Média (R$)	35,56	54,99	58,30	69,33
	Clientes aprovados	11.394	9.583	9.598	8.305
	Total (R$)	405.192,13	526.979,54	559.541,20	575.822,71
40%	Média (R$)	75,67	96,38	99,84	101,70
	Clientes aprovados	11.394	9.583	9.598	9.443
	Total (R$)	862.176,96	923.605,80	958.298,63	960.381,46
50%	Média (R$)	115,78	137,77	141,39	134,63
	Clientes aprovados	11.394	9.583	9.598	10.181
	Total (R$)	1.319.161,80	1.320.232,07	1.357.056,06	1.370.700,57
60%	Média (R$)	155,88	179,16	182,94	168,97
	Clientes aprovados	11.394	9.583	9.598	10.668
	Total (R$)	1.776.146,64	1.716.858,34	1.755.813,50	1.802.533,88

Analisando os resultados obtidos, observa-se que o modelo proposto apresenta estimativas de lucro total positivas para todos os percentuais de margem bruta considerados. As estimativas de lucro total foram maiores para o modelo proposto quando comparada com os outros cenários: sem uso de nenhum modelo de previsão de risco, utilizando apenas o modelo de classificação logístico ou utilizando apenas o modelo de classificação neural.

Na Figura 27 observa-se o comportamento assintótico da previsão de lucro total dada pelo modelo proposto, sempre com resultado superior aos outros cenários comparados. Os resultados apresentados confirmam a robustez do modelo, ao variar a margem bruta de lucro

admitida, pois mesmo que a margem de lucro da empresa fosse menor que a admitida (30%), o modelo proposto apresentaria resultados melhores que os demais cenários.

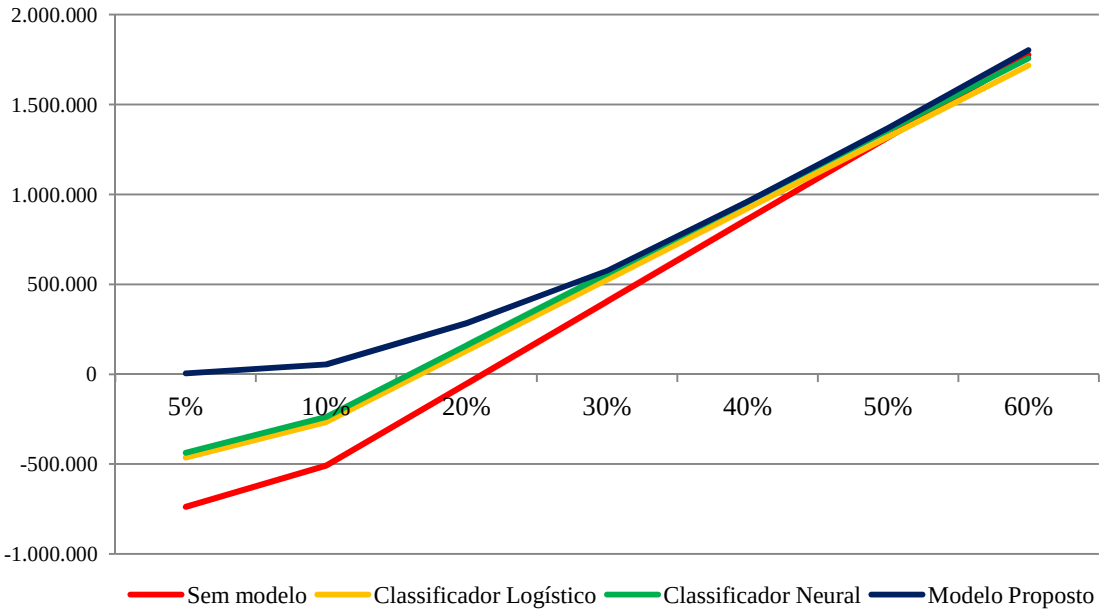

Figura 27 Análise de sensibilidade da previsão do lucro modificando a margem

7 CONSIDERAÇÕES FINAIS

Neste capítulo são apresentadas as principais conclusões e contribuições obtidas com o estudo e as limitações da pesquisa com sinalização de algumas sugestões para trabalhos futuros.

7.1 CONCLUSÕES E CONTRIBUIÇÕES

O crédito é um ativo valioso dentro de muitas empresas, sendo um dos principais recursos para ampliar sua rentabilidade. Nesse sentido, é necessário que este tenha uma política bem estruturada de forma a alocar eficientemente este ativo objetivando com isso a sua maximização de valor e, consequentemente, a maximização da riqueza dos acionistas da empresa. Porém, decisões quanto à concessão de crédito não são fáceis de serem tomadas.

Foi justamente esta dificuldade de tomada de decisão o incentivo para este estudo: construir um modelo para previsão do risco monetário associado à concessão de crédito. O modelo foi construído para prever o lucro médio esperado, utilizando as previsões obtidas com os modelos de classificação, com regressão logística e rede neural, como variáveis explicativas. Para construção do modelo proposto fez-se uso do método *ensemble*, que propõe a utilização conjunta dos classificadores (logístico e neural) para uma melhor predição, e do método *hybrid*, que propõe a modelagem sequencial, onde os resultados de uma técnica são utilizados como variáveis explicativas para a outra.

De posse da previsão de lucro dada pelo modelo proposto, a empresa passa a ter uma escala para tomada de decisão. No momento que a previsão indica maior prejuízo do que ganho em virtude da aceitação dos proponentes, obtém-se o corte na escala para a definição de sua política de crédito. Assim, essa análise permite que a empresa consiga diminuir suas perdas relativas à aprovação de potenciais inadimplentes, além de obter uma melhor forma de atribuir os limites de crédito.

O modelo foi construído e validado, demonstrando resultados promissores na previsão do lucro esperado com cada cliente. Para demonstrar a efetividade desta proposta foram avaliados os ganhos esperados em quatro cenários: (i) sem utilização de nenhum modelo de previsão de risco de crédito; (ii) utilizando um modelo de classificação obtido com regressão logística; (iii) utilizando um modelo de classificação obtido com rede neural; e (iv) utilizando o modelo proposto para previsão do risco monetário.

O lucro médio esperado com cada cliente sem a utilização de nenhum modelo foi de R$35,56 enquanto que utilizando o modelo de classificação obtido com regressão logística esse valor subiu para R$54,99, o que representa 54,64% de aumento. Já utilizando o modelo de classificação obtido com a rede neural o lucro médio estimado por cliente foi de R$58,30, representando um aumento de 63,95% em relação ao cenário sem utilização de nenhum método para previsão de risco de crédito.

Utilizando o modelo proposto para previsão de risco monetário o lucro médio ficou em R$69,33 por cliente, denotando um aumento estimado de 94,97% em comparação com o cenário sem uso de modelo de previsão, e um aumento de 26,08% quando comparado com o cenário de uso do modelo de classificação obtido com regressão logística. Uma análise de sensibilidade dos resultados a variações na margem de lucro por transação também foi realizada, evidenciando sua robustez. Nesse sentido, o modelo proposto se mostra eficiente como ferramenta de apoio para gestão no processo de decisão de concessão de crédito.

A originalidade da pesquisa está em obter um modelo para prever o valor médio esperado de lucro bruto com um cliente após a concessão do crédito. Com isso, a tomada de decisão de crédito é balizada por um valor monetário associado à operação de crédito e não apenas na determinação de um comportamento binário de bom ou mau pagador, como nos modelos tradicionais de previsão de risco de crédito. Tsai e Chen (2010) afirmam que os trabalhos atuais se preocupam em estudar o desempenho dos modelos de classificação, avaliando sua precisão e taxa de erro, sendo que nenhum estudo explorou a possibilidade de estimar o lucro utilizando modelos de risco de crédito.

A relevância deste trabalho deve-se à importância do microcrédito no atual contexto econômico e financeiro do país, constituindo modalidade de crédito maciçamente utilizada por muitas empresas. Além disso, cabe mencionar a contribuição que os resultados encontrados na aplicação do modelo proposto podem fornecer ao processo de concessão e análise do crédito da empresa em estudo, bem como para outras instituições de microcrédito, carentes de instrumentos metodológicos para auxílio à gestão do risco.

7.2 **LIMITAÇÕES DA PESQUISA**

Segundo Silva (2008), o uso de modelos de previsão presta uma grande contribuição à análise de crédito. Contudo, o uso desses modelos não elimina a necessidade de que os concessores de crédito tenham definições políticas e estratégicas claras e que seus

profissionais estejam treinados para uma boa análise de crédito. Deste modo, os modelos devem ser entendidos como um instrumental complementar para o analista e não uma solução.

Cabe lembrar que o modelo proposto nesta tese se restringe a concessão de crédito para pessoas físicas em empresas comerciais que utilizam o crédito próprio como uma forma de impulsionar suas vendas. Qualquer aplicação em um contexto diferente do proposto aqui deve ter suas devidas adaptações analisadas. Tanto as características dos clientes, como as definições de comportamento quanto à inadimplência ou de lucro por parte de quem concede, teriam modificações substanciais e precisariam ser melhor avaliadas.

Uma limitação do modelo construído está na baixa capacidade preditiva. Porém, independentemente do método estatístico utilizado ou do número de variáveis consideradas, uma previsão perfeita quanto à inadimplência seria impossível, até porque, muitas vezes, clientes lucrativos têm as mesmas características dos clientes inadimplentes. Thomas (2000) afirma que o importante é buscar um modelo que classifique erroneamente tão poucos proponentes quanto possível e que demonstre melhora nos resultados de inadimplência.

Cabe ressaltar que um modelo de previsão de risco de crédito não é permanente. Pereira *et al.* (2002) sugerem que, após um ano de utilização, uma revisão seja feita, seguindo os mesmos passos para a construção do modelo original. Os autores afirmam também que a revisão se torna necessária também se houver mudança significativa na inadimplência, na lucratividade, nos prazos ou condições do negócio e, principalmente, no perfil da população. Tais alterações devem ser monitoradas através de relatórios de acompanhamento para o modelo.

Uma limitação existente na formulação de modelos de previsão de risco de crédito está relacionada à obtenção de uma amostra com viés de seleção, já que a população estudada refere-se somente a créditos concedidos. Os clientes que foram negados por algum motivo (por exemplo, por estarem incluídos em listas do SPC ou SERASA) e que, portanto, não fazem parte da amostra, serão potenciais clientes para solicitações de crédito futuras, mas seu perfil pode não estar contemplado no modelo. Vasconcellos (2002) apresenta um estudo sobre o efeito desse viés em modelos de previsão, mostrando que o uso de amostras restritas aos créditos aprovados gera resultados com vieses, mas que o tamanho e direção do viés do modelo nem mesmo podem ser conhecidos. O autor alerta também que não seria interessante uma empresa incorrer nos custos de conceder crédito a todos os clientes, pois ainda assim não se chegaria à população de todos os potenciais tomadores de crédito, já que poderia ocorrer

um viés por parte dos próprios clientes que podem escolher a empresa para solicitar crédito de acordo com seus critérios. Feelders (2000), em seu artigo, também relata estudos para tentar inferir o comportamento dos proponentes rejeitados pela instituição e que, portanto, não fazem parte da amostra. O método utilizado pelo autor, denominado por inferência dos rejeitados, consiste em inferir o comportamento dos proponentes rejeitados, caso tivessem sido aprovados.

Os principais modelos de previsão de risco de crédito estão sustentados basicamente em duas categorias: modelos de aprovação de crédito (*application scoring*) e modelos de comportamento de crédito (*behavioural scoring*). Este trabalho limitou-se a aplicação do modelo proposto a um estudo para aprovação de crédito a novos clientes de uma empresa que concede crédito próprio para financiamento dos produtos vendidos em suas lojas. Portanto, uma sugestão de pesquisa futura seria avaliar a aplicação do modelo proposto para construção de um *behavioural scoring*, avaliando, desta forma, ganhos relativos a clientes que já tenham um histórico de crédito com a empresa.

Em decorrência de algumas limitações desta pesquisa, outras possibilidades de trabalhos futuros emergem:

- no intuito de obter uma maior capacidade preditiva do modelo, buscar por variáveis explicativas que vão além das características demográficas dos clientes como, por exemplo, informações sobre pagamentos com a própria empresa ou com outros estabelecimentos, ou informações fornecidas por órgão de proteção ao crédito, como o SPC e o SERASA;
- construir um modelo condicionado ao valor de limite de crédito atribuído, pois é plausível supor que o risco de um cliente se altere dependendo do valor de crédito a ele concedido;
- estudar uma melhor medição do lucro/prejuízo com os clientes, considerando pagamento de juros e taxas em horizontes futuros (safras de pagamentos) e, ainda, considerar o desconto de custos operacionais e administrativos que não foram utilizados neste trabalho pelo fato de o caso em estudo não ter estas informações disponibilizadas;
- avaliar a aplicação do modelo proposto no ambiente empresarial para mensurar longitudinalmente os resultados percebidos;

- avaliar os efeitos nos resultados estimados pelo modelo com a alteração da definição de bom e mau pagadores;
- testar a construção do modelo com a inclusão dos clientes indefinidos (com atrasos entre 31 e 90 dias) de forma a verificar a existência de efeitos não lineares nas relações entre as variáveis explicativas e o risco monetário estimado.

REFERÊNCIAS BIBLIOGRÁFICAS

AKERLOF, G. A. The market for lemmons: quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, v.84, n.3, p.488-500, 1970.

ALTMAN, E.L. Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance*, v.23, n.4, p.589-609, 1968.

ANDREEVA, G.; ANSELLA, J.; CROOK, J. Modelling profitability using survival combination scores. *European Journal of Operational Research*, v.183, n.3, p.1537-1549, 2007.

ANGELONI, M. T. Elementos intervenientes na tomada de decisão. *Ciência da Informação*, v.32, n.1, p.17-22, 2003.

AUDY, J. L. N.; BECKER, J. L.; FREITAS, H. Modelo de planejamento estratégico de sistemas de informações: a visão do processo decisório e o papel da aprendizagem organizacional. In: VANTI, A. A. *Gestão da tecnologia empresarial e da informação*: *conceitos e estudos de casos*. São Paulo: Internet, 2001.

BACK, R. S. *Um método para definição de indicadores de desempenho aplicado à gestão de projetos de sistema de informação*. Dissertação de Mestrado, Programa de Pós-Graduação em Administração, Universidade Federal do Rio Grande do Sul, 2002.

BARON, J. *Thinking and deciding*. 2. ed. London: Cambridge University Press, 1994.

BATES, J. M.; GRANGER, C. W. J. The combining of forecasts. *Operational Research Quarterly*, v.20, n.4, p.451-468, 1969.

BERNSTEIN, P. L. *Desafio aos deuses: a fascinante história do risco*. 2 ed. Rio de Janeiro: Campus, 1997.

BROADY-PRESTON, J.; HAYWARD, T. Turbulent change: strategy and information flow in UK retail banks. *Journal of Information Science*, v.24, n.6, p.395-408, 1998.

CAOUETTE, J. B.; ALTMAN, E. I.; NARAYANAN, P. *Gestão do risco de crédito: o próximo grande desafio financeiro.* São Paulo: Qualitymark, 1999.

CHATTERJEE, S.; HADI, A.; PRICE, B. *Regression analysis by example*. New York: Wiley, 2000.

CHEN, W.; MA, C.; MA, L. Mining the customer credit using hybrid support vector machine technique. *Expert Systems with Applications*, v.36, n.4, p.7611-7616, 2009.

CIA, J. C. Propostas de medidas de inadimplência para o mercado brasileiro. In: *ENANPAD*, Rio de Janeiro, Anais..., 2003.

CLEMEN, R. T. Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting*. v.5, p.559-583, 1989.

CLEMEN, R. T. *Making hard decision: an introduction to decision analysis*. 2 ed. Belmont: Duxbury, 1996.

CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. *Análise multivariada: para cursos de Administração, Ciências Contábeis e Economia*. São Paulo: Atlas, 2007.

CROOK, J. N., EDELMAN, D. B., THOMAS, L. C. Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, v.183, n.3, p.1447-1465, 2007.

DALFOVO, O. Sistema de informação executiva auxilia a tomada de decisão. *Revista Developers*, n.40, p.28-32, 1999.

DIETTERICH, T. G., Ensemble methods in machine learning. *Lecture Notes in Computer Science*, v.1857, p.1-15, 2000.

DRAPER, N. R., SMITH, H. *Applied regression analysis*. New York: John Wiley & Sons. 1981.

DUARTE Jr., A. M. Risco: definições, tipos, medição e recomendações para seu gerenciamento. *Revista Resenha BM&F*, n.114, p.25-33, 1996.

DUTRA, M. S.; BIAZI, E. Uma abordagem alternativa de credit scoring usando análise discriminante: eficiência na concessão de crédito para o segmento de pessoas físicas no Brasil. In: *SPOLM*, Rio de Janeiro, Anais... 2008.

FEELDERS, A. J. Credit scoring and reject inference with mixture models, *International Journal of Intelligent Systems in Accounting Finance and Management*, v.9, n.1, p.1-8, 2000.

FINLAY, S. Multiple classifier architectures and their application to credit risk assessment. *European Journal of Operational Research*, v.210, n.2, p.368-378, 2011.

FLORES, B. E.; WHITE, E. M. A framework for the combination of forecasts. *Journal Academic Marketing Science*, v. 16, n.3-4, p.95-103, 1988.

FREITAS, H. M. R.; KLADIS, C. M. O processo decisório: modelos e dificuldades. *Revista Decidir*, n.8, p.30-36, 1995.

GARCIA, L. Crescimento do crédito pode ser prejudicial ao consumidor. *Jornal O Debate*, 10 de julho, 2008.

GAUCHI, J. P.; CHAGNON, P. Comparison of selection methods of explanatory variables in PLS regression with application to manufacturing process data. *Chemometrics Intelligent Laboratory Systems*, v.58, n.2, p.171-193, 2001.

GHODSELAHI, A. A hybrid support vector machine ensemble model for credit scoring. *International Journal of Computer Applications*, v.17, n.5, 2011.

GIL, A. C. *Métodos e técnicas de pesquisa social*. 5.ed. São Paulo: Atlas, 1999.

GITMAN, L. J. *Princípios de administração financeira – essencial*. Porto Alegre: Bookman, 2001.

GOUVÊA, M. A.; GONÇALVES, E. B. Análise de risco de crédito com o uso de modelos de redes neurais e algoritmos genéticos. In: *IX SEMEAD* – Seminários em Administração FEA-USP, São Paulo. Anais, 2006.

GRANGER, C. W. J.; RAMANATHAN, R. Improved methods of forecasting. *Journal of Forecasting*, v.3, p.197-204, 1984.

GUIMARÃES, I. A.; CHAVES NETO, A. Reconhecimento de padrões: metodologias estatísticas em crédito ao consumidor. *RAE Eletrônica EAESP/FGV*, v.1, n.1, 2002.

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. *Análise multivariada de dados*. 5.ed. Porto Alegre: Bookman, 2005.

HAND, D. J.; HENLEY, W. E. Statistical classification methods in consumer credit scoring: a review. *Journal of Royal Statistical Society, Series A*, v.160, n.3, p.523-541, 1997.

HAND, D. J. Modeling consumer credit risk. *IMA Journal of Management Mathematics*, v.12, p.139-155, 2001.

HASTIE, R. Problems for judgment and decision making. *Annual Review of Psychology*, v.52, p.653-683, 2001.

HAYKIN, S. *Redes neurais: princípios e prática*. Trad. Paulo Martins Engel. 2.ed. Porto Alegre: Bookman, 2001.

HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression*. New York: John Wiley & Sons, 1989.

HSIEH, N. C. Hybrid mining approach in the design of credit scoring models. *Expert Systems with Applications*, v.28, p.655-665, 2005.

HSIEH, N. C.; HUNG, L. P. A data driven ensemble classifier for credit scoring analysis. *Expert Systems with Applications*, v.37, n.1, p. 534-545, 2010.

KELLER, L. R.; HO, J. L. Decision problem structuring: generating options. *IEEE Transactions on Systems, Man and Cybernetics*, v.18, n.4, p.715-728, 1998.

KERAMATI, A.; YOUSEFI, N. A proposed classification of data mining techniques in credit scoring. In: *International Conference on Industrial Engineering and Operations Management*, Kuala Lumpur, Malaysia, 2011.

KITTLER, J.; HATEF, M.; DUIN, R. P. W.; MATAS, J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.20, n.3, p.226-239, 1998.

KLEINBAUM, D. G. *Logistic regression*: *a self-learning text*. New York: Springer; 1996.

KOVÁCS, Z. L., *Redes neurais artificiais: fundamentos e aplicações*. São Paulo: Livraria da Física, 2002.

LAHSASNA, A.; AINON, R. N.; WAH, T. Y. Credit scoring models using soft computing methods: a survey. *The International Arab Journal of Information Technology*, v.7, n.2, p.115-123, 2010.

LAWRENCE, J. *Introduction to neural networks – design, theory and applications*. Nevada: California Scientific Software, 1994.

LAWRENCE, D. *Handbook of consumer lending*. New Jersey: Prentice Hall. 1992

LEE T.-S.; CHIU, C.-C.; LU, C.-J.; CHEN, I-F. Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, v.23, p.245-254, 2002.

LEE, T.-S.; CHEN, I.-F. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, v.28, p.743-752, 2005.

LEWIS, E. M. *An introduction to credit scoring*. San Rafael: Fair, Isaac and Co., Inc. 1992.

LOESCH, C.; SARI, S.T. *Redes neurais artificiais*: *fundamentos e modelos*. Blumenau, FURB, 1996.

LURIE, N. H. Decision making in information-rich environments: the role of information structure. *Journal of Consumer Research*, v.30, n.4, p.473-486, 2004.

McALLISTER, P. H.; MINGO, J. J. Commercial loan risk management, credit-scoring and pricing: The need for a new shared data base. *Journal of Commercial Bank Lending*,

v.76, n.9, p.6-22, 1994.

MCNEILLY, M. Gathering information for strategic decisions, routinely. *Strategy & Leadership*, v.30, n.5, p. 29-34, 2002.

MEMÓRIA, J. M. P. *Breve história da estatística*. Brasília - DF: Embrapa Informação Tecnológica. 2004.

MENDES FILHO, E. F.; CARVALHO, A. C. P. L. F.; MATIAS, A. B. Utilização de redes neurais artificiais na análise de risco de crédito a pessoas físicas. In: *III Simpósio Brasileiro de Redes Neurais*, Recife, Anais..., 1996.

MESTER, L. J. What's the point of credit scoring? *Business Review*, p.3-16, set.-oct.1997.

MONTGOMERY, D. C.; PECK, E. A. *Introduction to linear regression analysis*. New York: John Wiley & Sons. 1982.

NETER, J., KUTNER, M. H., NACHTSHEIM, C. J., WASSERMAN, W. *Applied linear statistical models*. London: Irwin. 1996.

PALEOLOGO, G., ELISSEEFF, A., ANTONINI, G., Subagging for credit scoring models. *European Journal of Operational Research*, v.201, n.2, p.490-499, 2010.

PARK, S. Solving the mystery of credit scoring models. *Business Credit*, v.106 n.3, p.43-47, 2004.

PARKINSON, K. L.; OCHS, J. R. Using credit screening to manage credit risk. *Business Credit*, v.100, n.3, p.23-27, 1998.

PEREIRA, A. P. F.; BARROSO, M. H.; NEPOMUCENO FILHO, F. Uso do credit score na análise de crédito de pessoa física. In: *Congresso Nacional de Excelência em Gestão*, Niterói, RJ, 2002.

PICININI, R.; OLIVEIRA, G. M. B.; MONTEIRO, L. H. A. Mineração de critério de credit scoring utilizando algoritmos genéticos. In: *VI Simpósio Brasileiro de Automação Inteligente*, Bauru, SP, 2003.

POLIKAR, R. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, v.6, n.3, p.21-45, 2006.

QUEIROZ, R. S. B. A importância dos modelos de credit scoring na concessão de crédito ao consumidor no varejo. In: *Anais do IX SEMEAD – FEA/USP*, São Paulo - SP. 2006.

ROSEMBERG, E.; GLEIT, A. Quantitative methods in credit management: a survey. *Operations Research*, v.42, n.4, p.589-613, 1994.

RUSSO, J. E.; SCHOEMAKER, P. J. H. *Decisões vencedoras: como tomar a melhor decisão, como acertar na primeira tentativa*. Rio de Janeiro: Campus, 2002.

SANTOS, J. *Análise de crédito: empresas e pessoas físicas*. São Paulo: Atlas, 2000.

SAUNDERS, A. *Medindo o risco de crédito: novas abordagens para value at risk e outros paradigmas*. Rio de Janeiro: Qualitymark, 2000.

SCHAEFER R.E.; BORCHERDING, K. A note on the consistency between two approaches to incorporate date from unreliable sources in bayesian analysis. *Organizational Behavior and Human Performance*, n. 9, p.504-508, 1973.

SCHRICKEL, W. K. *Análise de crédito: concessão e gerência de empréstimos*. 3.ed. São Paulo: Atlas, 1997.

SELAU, L. P. R. *Construção de modelos de previsão de risco de crédito*. Dissertação de

Mestrado, Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, 2008.

SEMEDO, D. P. V. *Credit Scoring: aplicação da regressão logística vs redes neuronais artificiais na avaliação do risco de crédito no cercado Cabo-Verdiano*. Dissertação de Mestrado, Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa, 2009.

SHIMIZU, T. *Decisão nas organizações*. São Paulo: Atlas, 2001.

SILVA, J. P. Os dois lados do crédito. *GV Executivo*, v.5, n.3, p.68-72, 2006.

SILVA, J. P. *Gestão e análise de risco de crédito*. São Paulo: Atlas, 2008.

SILVA, M. A. *Elaboração de um modelo de análise e concessão de créditos para pessoas físicas em um banco.* Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, 2003.

SIMON, H. A. *The new science of management decision*. New York: Harper, 1960.

SIMON, H. A. *Comportamento administrativo: estudo dos processos decisórios nas organizações administrativas*. Rio de Janeiro: Fundação Getúlio Vargas, 1970.

SIMON, H. A. Invariants of human behavior. *Annual Review of Psychology*, v.41, p.1-19. 1990.

SOUSA, A. F.; CHAIA, A. J. Política de crédito: uma análise qualitativa dos processos em empresas. *Caderno de Pesquisas em Administração*, São Paulo, v.7, n.3, 2000.

STEINER, M. T. A.; CARNIERI, C.; KOPITTKE, B. H.; STEINER NETO, P. J. Sistemas especialistas probabilísticos e redes neurais na análise do crédito bancário. *Revista de Administração*, São Paulo, v.34, n.3, 1999.

STIGLITZ, J.; WEISS, A. Credit rationing in markets with imperfect information. *American Economic Review*, v.71, n.3, p.393-410, 1981.

SUBRAMANIAN, V., HUNG, M. S., HU, M. Y. An experimental evaluation of neural networks for classification. *Computers & Operations Research*, v.20, n.7., p.769-782, 1993.

THOMAS, C. L. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, v.16, n.2, p.149-172, 2000.

TSAI, C.-F.; CHEN, M. L. Credit rating by hybrid machine learning techniques. *Applied Soft Computing*, v.10, p.374-380, 2010.

TURBAN, E.; MEREDITH, J. R. *Fundamentals of management science*. 6 ed. Boston: Irwin, 1994.

TVERSKY, A.; KAHNEMAN, D. Judment under uncertainty: Heuristics and biases. *Science*, n.185, p.1124-1131, 1974.

TWALA, B. Multiple classifier application to credit risk assessment. *Expert Systems with Applications*, v.37, p.3326-3336, 2010.

VASCONCELLOS, M. S. *Proposta de método para análise de concessões de crédito a pessoas físicas.* Dissertação de Mestrado, Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, 2002.

VLEK, C. What constitutes a good decision? *Acta Psychological*, v.56, p.5-27, 1984.

VON WINTERFELDT, D.; EDWARD, W. *Decision analysis and behavioral research*.

Cambridge University Press, 1986.

WANG, G.; HAO, J.; MA, J.; JIANG, H. A comparative assessment of ensemble learning for credit scoring. *Expert systems with appl*ications, v.38, p.223-230, 2011.

WERNER, L.; RIBEIRO, J. L. D. Modelo composto para prever demanda através da integração de previsões. *Produção*, v.16, n.3, p.493-509, 2006.

WYNN, A.; McNAB, H. *Principles and practice of consumer credit risk management*. 3.ed. Canterbury: CIB Publishing, 2008.

YAP, B.W.; ONG,S. H.; HUSAIN, N. H. M. Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications*, v.38, n.10, p.13274-13283, 2011.

YATES, J. F.; STONE, E. R. The risk construct. In: YATES, J. F. *Risk-taking behavior*. England: John Wiley & Sons, 1994.

YU, X.; EFE, M. O.; KAYNAK, O. A general backpropagation algorithm for feedforward neural networks learning, *IEEE Trans. on Neural Networks*. v.13, n.1, p.251–254. 2002.

ZHOU, Z. H.; WU, J.; TANG, W. Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, v.137, p.239-263, 2002.

ZHOU, L., LAI, K. K., YU, L. Least squares support vector machines ensemble models for credit scoring. *Expert Systems with Applications*, v.37, p.127-133, 2010.

ZIMMER, J.; ANZANELLO, M. J. Um novo método para seleção de variáveis preditivas com base em índices de importância das variáveis. In: *XLIII Simpósio Brasileiro de Pesquisa Operacion*al, Ubatuba, 2011.

APÊNDICE A - AGRUPAMENTO DE PROFISSÕES

Péssimo Desempenho	BABA	PROMOTOR VENDAS
	COZINHEIRO	ALMOXARIFE
	PINTOR	
Muito Mau Desempenho	AUX PRODUCAO	PEDREIRO
	CABELEIREIRO	PORTEIRO
	CONFEITEIRO	RECEPCIONISTA
	GERENTE	VENDEDOR
	PADEIRO	
Mau Desempenho	AUTONOMO	MANICURE
	AUX ADMINISTRATIVO	MECANICO
	AUX COZINHA	TEC ENFERMAGEM
	AUX SERVICOS GERAIS	VIGILANTE
	COMERCIANTE	
Desempenho Neutro	ATENDENTE	INDUSTRIARIO
	COMERCIARIO	MOTORISTA
	DOMESTICA	
Bom Desempenho	CAIXA	SECRETARIA
	DO LAR	SERVENTE
	PENSIONISTA	
Muito Bom Desempenho	AGRICULTOR	OPERADOR
	BALCONISTA	METALUGICO
	COSTUREIRO	AUX ENFERMAGEM
	DIARISTA	
Excelente Desempenho	APOSENTADO	PROFESSOR

APÊNDICE B - AGRUPAMENTO DE CIDADES DE NASCIMENTO

Péssimo Desempenho	ALVORADA	
Muito Mau Desempenho	CRUZ ALTA	RIO GRANDE
	ESTEIO	TRAMANDAI
	PORTO ALEGRE	
Mau Desempenho	CANOAS	SAO BORJA
	GRAVATAI	SAO GABRIEL
	IJUI	SAPIRANGA
	NOVO HAMBURGO	SAPUCAIA DO SUL
	PELOTAS	URUGUAIANA
Desempenho Neutro	ALEGRETE	SANTO ANGELO
	CAMAQUA	SAO FRANCISCO DE PAULA
	CANELA	SAO LOURENCO DO SUL
	SANTANA DO LIVRAMENTO	VIAMAO
Bom Desempenho	BAGE	PASSO FUNDO
	BUTIA	RIO PARDO
	CACAPAVA DO SUL	SANTA CRUZ DO SUL
	GUAIBA	SAO JERONIMO
	MONTENEGRO	SAO LEOPOLDO
	OSORIO	SAO LUIZ GONZAGA
Muito Bom Desempenho	CACHOEIRA DO SUL	SANTA ROSA
	CAXIAS DO SUL	SANTA VITORIA DO PALMAR
	PALMEIRA DAS MISSOES	TAQUARA
	SANTA MARIA	
Excelente Desempenho	CANGUCU	SAO SEPE
	ENCRUZILHADA DO SUL	TAPES
	GIRUA	TORRES
	HORIZONTINA	TRES DE MAIO
	ROLANTE	TRIUNFO
	SANTO ANTONIO DA PATRULHA	

APÊNDICE C - AGRUPAMENTO DE CEP RESIDENCIAL

	2 PRIMEIRAS POSIÇÕES	3 PRIMEIRAS POSIÇÕES	4 PRIMEIRAS POSIÇÕES
Péssimo Desempenho			**9670** SÃO JERÔNIMO
Muito Mau Desempenho			**9175** Aberta dos Morros - POA **9179** Restinga - POA **9191** Camaquã - POA **9192** Cavalhada/Camaquã - POA **9440** Águas Claras - VIAM **9449** NS Aparecida/Pq Índio Jari - VIAM **9481** Formozo/Passo Feijó - ALVO **9493** Dist Industrial/Cohab - CACH
Mau Desempenho		**902** Farrapos/Navegantes/Humaitá - POA **906** Partenon/Jardim Botânico - POA **912** Protásio Alves/Rubem Berta - POA **915** Lomba do Pinheiro/Agronomia - POA **923** Mathias Velho/Harmonia - CANO **934** Lomba Grande/Santo Afonso - NH **941** GRAVATAÍ **945** Vila Augusta/Jd Universit. - VIAM	**9117** Rubem Berta - POA **9172** Nonoai/Teresópolis - POA **9174** Cavalhada/Vila Nova - POA **9190** Tristeza/Vila Assunção - POA **9326** Centro/Vila Teópolis - EST **9329** Pq Primavera/Pq St Inácio - EST **9353** São Jorge/Vila Diehl - NH **9400** GRAVATAÍ **9441** Centro/Tarumã - VIAM **9442** Vila Elsa/Estalagem - VIAM **9444** Jd Krahe/St Onofre - VIAM **9482** Maria Regina/Sumaré - ALVO **9483** Tijuca/Piratini - ALVO **9485** Aparecida/Jd Algarve - ALVO **9490** Jardim América/Vila City - CACH **9607** Porto/Três Vendas - PEL **9618** CAMAQUÃ **9750** URUGUAIANA
Desempenho Neutro		**908** Santa Tereza/Medianeira - POA **913** Vila Jardim/Vila Ipiranga - POA **914** Protásio Alves/Jardim Carvalho - POA	**9332** Industrial/Ouro Branco - NH **9445** São Lucas/Florescente - VIAM **9447** St Cecília/Viamópolis - VIAM **9480** Maringá/Sumaré - ALVO **9494** Vila Vista Alegre - CACH **9495** Vila Bom Princípio/Pq Matriz - CACH
Bom Desempenho		**922** Fátima/Rio Branco - CANO **924** Igara/São José/Guajuviras - CANO **925** GUAÍBA **930** SÃO LEOPOLDO **938** NOVA HARTZ, SAPIRANGA **955** OSÓRIO, CAPÃO DA CANOA **956** TAQUARA, CANELA, GRAMADO **957** BENTO GONÇALVES, GARIBALDI	**9178** Lami/Belém Novo - POA **9328** Vila Esperança/Pq Amador - EST **9330** Centro - NH **9333** Liberdade/Ideal - NH **9334** Primavera/Petrópolis - NH **9354** Canudos/Mauá - NH **9443** Jardim Krahe/Sítio S.José - VIAM **9496** Pq Granja Esperança - CACH **9601** Centro - PEL **9617** SÃO LOURENÇO DO SUL **9674** ARROIO RATOS e CHARQUEADAS **9754** ALEGRETE
Muito Bom Desempenho	**99** PASSO FUNDO	**950** CAXIAS DO SUL **962** RIO GRANDE, STA VITÓRIA PALMAR **965** CACHOEIRA DO SUL, CAÇAPAVA **970** SANTA MARIA	**9407** GRAVATAÍ
Excelente Desempenho	**98** CRUZ ALTA	**937** CAMPO BOM **958** ESTRELA, TAQUARI, VENÂNCIO AIRES **964** BAGÉ, DOM PEDRITO **966** RIO PARDO, PÂNTANO GRANDE **968** SANTA CRUZ DO SUL **971** SANTA MARIA, ITAARA **973** SÃO GABRIEL, LAVRAS DO SUL	**9352** Guarani/Vila Nova - NH

APÊNDICE D - AGRUPAMENTO DE CEP COMERCIAL

	2 PRIMEIRAS POSIÇÕES	3 PRIMEIRAS POSIÇÕES	4 PRIMEIRAS POSIÇÕES
Péssimo Desempenho		**912** Protásio Alves/Rubem Berta - POA	**9670** SÃO JERÔNIMO
Muito Mau Desempenho		**900** Centro/Farroupilha/Bom Fim - POA **906** Partenon/Jardim Botânico - POA **932** ESTEIO, SAPUCAIA DO SUL **948** ALVORADA	**9174** Cavalhada/Vila Nova - POA **9190** Tristeza/Vila Assunção - POA **9192** Cavalhada/Camaquã - POA
Mau Desempenho		**901** Azenha/Menino Deus/Praia Belas - POA **902** Farrapos/Navegantes/Humaitá - POA **908** Santa Tereza/Medianeira - POA **910** Passo D'Areia/Jardim Lindóia - POA **911** Sarandi/Rubem Berta - POA **913** Vila Jardim/Vila Ipiranga - POA **915** Lomba do Pinheiro/Agronomia - POA **933** Rio Branco/Primavera/Industrial - NH **940** GRAVATAÍ **949** CACHOEIRINHA **961** CAMAQUÃ, CAPÃO DO LEÃO	**9175** Aberta dos Morros - POA **9179** Restinga - POA **9191** Camaquã - POA **9351** Centro/Hamburgo Velho - NH **9353** São Jorge/Vila Diehl - NH **9602** Fragata/Três Vendas - PEL
Desempenho Neutro		**904** Auxiliadora/Petrópolis - POA	**9380** SAPIRANGA
Bom Desempenho		**905** São João/Floresta/Higienópolis - POA **925** GUAÍBA **934** Lomba Grande/Santo Afonso - NH **955** OSÓRIO, CAPÃO DA CANOA **956** TAQUARA, CANELA, GRAMADO **957** BENTO GONÇALVES, GARIBALDI	**9178** Lami/Belém Novo - POA **9389** NOVA HARTZ **9441** Centro/Tarumã - VIAM **9601** Centro - PEL **9674** ARROIO RATOS e CHARQUEADAS
Muito Bom Desempenho	**99** PASSO FUNDO	**959** LAJEADO, ENCANTADO, PROGRESSO **962** RIO GRANDE, STA VITÓRIA PALMAR	
Excelente Desempenho	**97** SANTA MARIA **98** CRUZ ALTA	**937** CAMPO BOM **950** CAXIAS DO SUL **958** ESTRELA, TAQUARI, VENÂNCIO AIRES **964** BAGÉ, DOM PEDRITO **965** CACHOEIRA DO SUL, CAÇAPAVA **966** RIO PARDO, PÂNTANO GRANDE **968** SANTA CRUZ DO SUL	

APÊNDICE E – PESOS DOS NEURÔNIOS DA REDE NEURAL

Neurônios da camada oculta

	DIDAD1	DIDAD4	DIDAD6	DIDAD7	DIDAD8	DIDAD23	DSEXOF	DPRIM	DSUP	DCASADO	DTSERV67	DTSERV89	DFILHO	DRES_ALU	DGCEPR12	DGCEPRE3	DGCEPCO1	DGCEPCO7	DGCEPC56	DGPROF1	DGPROF2	DGPROF5	DGPROF67	DCIDNA12	DCIDNA3	DCIDNA7	BIAS
N1	-2,925	-2,5252	-4,0002	-0,6846	0,8474	1,1644	-3,584	-3,927	0,579	-1,2564	1,9402	0,6662	3,9942	0,2876	2,0246	-4,0002	-4,8116	2,6726	1,9492	-3,648	-2,059	-3,7584	-3,9146	-3,6566	-3,3194	3,3562	-4,0006
N2	-3,0646	-0,218	0,0642	-0,751	3,2992	-1,626	0,805	-1,1606	-2,0424	0,203	-3,1566	1,8194	-0,8622	-1,8808	-4,0002	-3,002	-6,4764	0,9822	-1,375	-4,0002	-1,4114	-2,384	-0,239	-3,6976	-1,9184	-2,698	-0,838
N3	-3,9966	-2,41	-1,881	1,6852	3,5002	-1,0476	-3,2766	-3,8996	-2,4322	3,9916	2,763	2,6086	2,2892	0,9862	-3,3252	-4,0002	-6,5076	1,538	3,209	-3,0946	-1,8812	-2,0916	-2,4008	-3,999	-3,9532	2,9852	-3,993
N4	-2,9292	2,9462	1,9722	0,5394	2,72	2,6272	-3,9746	-0,5004	-3,997	0,6276	2,7586	-3,9934	0,894	-4,0014	-0,9714	0,9692	-2,4556	3,919	1,9472	-3,9852	-3,9834	0,817	0,7542	-3,2202	-2,4112	-0,3512	-2,3806
N5	-3,6464	0,3126	0,861	2,1596	3,8944	0,5086	-2,123	-1,4386	-3,1446	2,9252	2,996	1,8806	0,214	-1,7108	-0,766	-4,0002	-7,0522	1,5802	-1,3164	-3,2236	-0,3834	-2,335	-1,848	-3,8908	-4,0002	3,3296	-3,9952
N6	-3,4756	0,719	1,3594	-0,7494	4	-1,258	-0,172	-1,4682	-1,7408	1,1904	-2,2546	1,6616	-0,4954	-1,0044	-4,0002	-2,7308	-6,3974	2,4744	-1,215	-2,986	-2,0466	-2,3036	-0,414	-3,5624	-2,3482	-2,9666	-1,3684
N7	-1,4446	-4,0002	-3,9992	-4,051	-4,2094	-3,997	-3,9994	-3,9992	-3,676	-3,9982	-3,9994	2,8702	-3,9992	2,2014	-3,999	-4	-3,9852	-4,001	-1,8696	-2,3136	-3,83	-4,0002	-4,0362	-3,9984	-4,0002	-4,0584	-3,9986
N8	-3,6252	-1,5474	-1,2434	1,5054	3,139	-0,31	0,7776	-1,411	2,6886	3,3634	3,2474	2,238	-1,6172	1,0776	0,453	0,7934	0,3256	1,0382	-0,0826	-1,0512	-0,7302	-3,0706	-2,2292	-3,5816	0,467	3,9722	-3,5082
N9	-4	2,651	3,6372	-3,4236	3,6322	1,8644	-0,3934	-0,9746	-5,3776	-0,103	-5,2416	3,1046	-0,5604	-0,3074	-3,9966	-3,391	-7,6176	0,6386	-3,6142	-3,2236	-3,991	-5,4444	-3,4272	-2,9052	-3,8092	-5,9544	-4,0026
N10	3,1862	-3,1744	-2,4036	-1,6732	-0,4234	-2,3734	-1,3422	1,5362	1,875	0,0226	-0,891	2,6642	1,4716	2,115	1,2852	-0,462	-0,9992	-3,9966	0,5772	-2,9524	0,037	0,8442	-2,633	-1,6224	-3,0666	-3,9826	-3,9916
N11	-3,5442	-0,1872	-1,4632	-1,301	3,996	1,5762	-1,9894	-2,1894	-2,6212	3,6312	1,8212	2,1452	2,4004	-0,6102	-1,9322	-3,8894	-7,1984	2,117	1,0444	-0,667	-0,9032	-3,4476	-2,2276	-3,8994	-2,252	1,93	-1,8016
N12	-0,9586	0,66	-2,2632	-3,3062	-2,86	1,0116	0,5766	-0,0916	-3,2786	-3,5194	-2,556	-3,9876	0,437	1,484	1,3594	1,2884	-0,045	0,2546	-0,7396	-3,449	0,9394	0,5112	0,0614	2,612	0,2408	-3,3796	-3,9694
N13	-2,8212	-1,0232	2,8492	0,4566	0,04	-3,3256	0,0916	-1,6106	0,484	-0,9734	-3,3994	1,2854	-0,6232	0,017	-2,2296	-1,8084	-3,6996	-2,63	-1,4812	-3,961	1,0854	0,9356	0,1524	-1,5546	-1,051	-2,136	-0,0226
N14	-3,9536	2,0852	2,6752	3,9932	3,8064	-2,1766	-0,761	0,298	-0,7016	2,3512	2,0046	3,3036	-1,342	-3,9644	-3,2444	-0,328	-2,1604	3,8802	3,6642	1,5442	-2,669	3,5076	3,9436	-3,872	0,3208	1,1434	-3,5212
N15	-3,118	-3,989	0,2412	3,7236	3,9026	-2,218	1,0894	-1,124	3,7294	3,9192	2,1162	3,9832	-0,7982	-3,896	-1,7906	-2,4116	-2,1384	3,9806	-0,4306	-3,9942	-0,328	-1,7174	-3,9502	-3,942	-0,4608	0,5622	-1,5806
N16	1,4716	-2,382	-0,671	-0,9484	-2,7924	-3,2692	2,1866	0,5104	2,0794	0,618	-3,9986	-3,992	-0,248	-3,9684	-2,2596	1,93	-2,095	0,2876	-1,9254	3,9656	-2,1142	0,6664	2,4164	-1,5792	-0,837	-1,5474	-3,8934
N17	1,4956	-3,984	-3,5526	-2,397	0,17	-3,9106	-3,9674	-3,979	3,9546	-3,9446	0,0226	3,513	-3,914	0,811	-3,9802	-3,9446	-3,9692	1,2504	3,6004	3,9612	-3,0052	-3,9714	-3,9616	-3,9702	-3,9906	-1,3634	-3,9486
N18	-5,4514	-7,9994	-7,9994	-7,9994	-2,145	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	4,0002	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-3,6374	-7,9994	-7,9994
N19	-0,3516	0,0546	-1,856	-3,7396	3,9606	-1,1956	-3,6864	-3,9582	-3,9586	3,9936	3,0014	3,076	2,0586	1,185	-3,976	-3,9416	-3,998	3,407	3,9672	0,9136	-2,2904	-3,689	-3,956	-3,9644	-3,9864	0,6072	-3,8434
N20	-0,8702	-3,0316	-2,093	-2,358	-2,0284	-3,304	-0,1354	-0,5312	-2,835	-1,0426	-0,0908	3,655	-1,3496	1,2922	-2,911	-2,0246	-3,9474	2,281	-1,9764	-1,5624	-3,2926	-3,4216	-1,1626	-2,983	-3,1154	-3,2662	1,0974
N21	-1,8384	-0,4526	-1,47	1,101	-2,1466	0,3312	-0,0962	-0,1512	0,9956	-0,2472	0,937	-3,9894	0,5414	-3,9762	-0,723	0,0382	3,983	-3,9674	-3,9666	1,9444	-1,4572	1,404	-0,9802	1,4454	0,685	-1,4676	-3,9652
N22	-2,4022	-3,9984	-3,978	-3,978	0,06	1,5176	-3,9724	-3,9808	-3,553	-3,9504	3,3226	3,9906	-3,9276	-1,005	-3,983	-3,9484	-3,8316	3,9984	-0,215	3,9742	-3,866	-3,98	-0,1552	-3,9774	-3,9954	-1,1922	-3,944
N23	-1,727	2,4464	-1,445	2,5176	2,692	-2,737	-2,1024	1,3054	1,9656	-2,6542	-1,1422	-0,1284	-0,5734	-3,9694	-1,0324	-0,004	3,4902	-3,7408	1,0432	0,6852	1,589	2,009	-0,3212	-0,7056	3,6892	3,155	-3,9182
N24	1,9816	-2,2406	-0,4786	-1,0242	-3,8864	0,5816	-2,6232	1,821	-3,687	0,8732	-2,2056	-1,8662	-1,672	3,8912	1,6744	1,655	-2,0208	0,0314	-2,6636	0,6836	0,8844	-2,7732	-2,8166	2,8102	3,9386	-2,1714	-3,9432
N25	-3,9546	-2,6532	-2,1402	-0,1714	3,0576	-0,869	-3,7276	-3,827	-2,4852	3,9946	3,8134	2,4194	0,158	1,148	-2,9616	-4,0006	-4,0002	3,918	3,9936	-1,6962	-2,2234	1,9772	-3,3106	-3,997	-3,921	3,267	-3,9226
N26	3,0524	0,205	3,0354	3,8176	-0,1182	-0,4646	0,4266	-1,1532	-2,7874	3,728	3,9546	2,3616	0,63	-2,89	-0,1646	-3,7234	-1,049	2,7572	3,4136	2,633	-1,4776	-3,9876	-1,5824	-0,7366	-0,6606	-0,84	-3,4244
N27	-1,677	-1,5812	-1,6852	-3,9972	-4,0034	-2,9772	-3,992	-3,9954	-3,4516	-3,9882	0,3642	0,9004	-3,9762	-1,2004	-3,9934	-3,9936	-3,9552	-3,9922	-2,5486	0,0408	-3,5756	-3,994	-3,9982	-3,9926	-3,984	-3,9986	-3,9882
N28	-2,9514	3,9944	-1,957	-1,5976	-0,6622	2,1096	0,1108	2,8706	3,86	-0,714	-3,9964	1,7602	-1,1256	-1,4512	0,9632	-1,7342	-2,7782	-3,4572	1,0922	2,3224	-2,8164	0,7508	0,5264	-2,8366	-2,6416	-4,0026	-3,9064
N29	-1,1712	-2,606	-2,8854	2,187	4,0002	-0,1946	1,8534	-0,8492	-1,173	1,693	-2,5242	-0,836	1,7884	-4,0006	0,3214	-3,2492	-3,7662	1,237	-2,2992	-0,726	2,785	-3,0922	-2,5644	-2,6002	0,2708	-2,7404	-2,351
N30	0,347	-3,998	-3,4556	-4,0008	-3,6192	-3,6842	-3,9954	-4,0008	-3,6262	-3,9922	-1,9892	2,8412	-3,9872	1,9584	-3,9954	-3,9992	-3,9008	-1,2254	0,6606	3,4732	-3,8656	-3,9966	-0,282	-3,995	-3,999	-2,4196	-3,995
N31	-3,9214	2,0316	2,9646	1,0574	4,0002	0,9312	0,2406	-1,044	-4,1112	1,3712	-3,201	2,4974	-1,154	-1,4212	-4,0002	-1,2766	-7,3694	1,7444	-2,7802	-4,0002	-3,3872	-4,5014	-1,1672	-2,7134	-2,235	-5,368	-2,8442
N32	-3,9274	1,9106	2,6894	2,0864	3,941	1,1066	0,3086	-0,9304	-3,491	1,472	-2,715	2,3736	-1,1796	-1,3296	-4,0002	-1,202	-7,2304	1,5034	-2,7882	-3,7326	-3,0366	-3,6324	-0,9866	-2,5576	-2,2354	-5,108	-2,7084
N33	-3,142	-3,7164	-2,9042	-3,991	-4,003	-3,253	-3,9884	-3,994	-2,9696	-3,9806	-3,9864	2,373	-3,971	0,3182	-3,9924	-3,9826	-4,0004	-3,986	-2,6916	-2,4532	-3,557	-3,9924	-3,9974	-3,9902	-3,996	-3,9974	-3,9882
N34	-0,4614	-3,8054	-3,503	-3,7746	-3,9536	-3,678	-2,6674	-3,3232	-3,8146	-3,421	1,8674	3,771	-3,4536	1,358	-3,7904	-3,598	-3,9806	0,333	-2,0072	-0,8572	-3,8904	-3,906	-3,538	-3,6836	-3,782	-3,903	-2,4162
N35	-3,9402	1,6962	2,5824	-0,147	3,9974	0,5964	0,334	-1,1782	-3,786	0,9108	-3,732	2,2344	-1,1296	-1,567	-4,0002	-1,1706	-7,3224	1,725	-2,3692	-4,0002	-2,8314	-3,9574	-0,613	-2,6696	-2,1342	-5,239	-2,6566

Neurônios da camada de saída

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17
tipo_cli	0,0408	0,0164	0,079	-1,2922	0,0564	0,006	0,1072	-1,042	-0,0114	-2,4544	0,0936	-2,089	-1,4422	1,9152	2,321	-1,586	1,599

N18	N19	N20	N21	N22	N23	N24	N25	N26	N27	N28	N29	N30	N31	N32	N33	N34	N35	BIAS
-0,0324	1,876	0,0592	-2,114	1,3852	-1,987	-1,2264	0,1434	-0,9456	0,4626	-1,3094	-0,8252	0,3744	-0,0106	0,0146	0,5862	-0,1596	0,0024	2,3756