

UNIVERSIDADE PRESBITERIANA MACKENZIE

Centro de Ciências Sociais e Aplicadas
Programa de Pós-Graduação em Administração de Empresas
Curso de Mestrado em Administração de Empresas

VÂNIA ROSATTI

UM MODELO DE *CREDIT SCORING* PARA MICROCRÉDITO: UMA
INOVAÇÃO NO MERCADO BRASILEIRO

São Paulo
2010

VÂNIA ROSATTI

**UM MODELO DE *CREDIT SCORING* PARA MICROCRÉDITO: UMA
INOVAÇÃO NO MERCADO BRASILEIRO**

Dissertação apresentada ao Programa de Pós-Graduação em Administração de Empresas da Universidade Presbiteriana Mackenzie, como requisito para a obtenção do título de Mestre em Administração de Empresas.

Orientador: Prof.Dr. Herbert Kimura

São Paulo
2010

R789m Rosatti, Vânia.

Um modelo de *credit scoring* para microcrédito: uma inovação no mercado brasileiro / Vânia Rosatti - 2010.

92 f. : il. ; 30 cm

Dissertação (Mestrado em Administração de empresas) –
Universidade Presbiteriana Mackenzie, São Paulo, 2010.

Bibliografia: f. 82-85.

1. Microcrédito. 2. *Credit scoring*. 3. Inovação. 4. *Bagging*. 5. *Boosting*. I. Título.

CDD 332.6737

VÂNIA ROSATTI

UM MODELO DE *CREDIT SCORING* PARA MICROCRÉDITO: UMA
INOVAÇÃO NO MERCADO BRASILEIRO

Dissertação apresentada ao Programa de Pós-Graduação em Administração de Empresas da Universidade Presbiteriana Mackenzie, como requisito para a obtenção do título de Mestre em Administração de Empresas.

Aprovada em 10 de Fevereiro de 2011.

BANCA EXAMINADORA

Prof.Dr. Herbert Kimura

Universidade Presbiteriana Mackenzie

Prof.Dr. Leonardo Fernando Cruz Basso

Universidade Presbiteriana Mackenzie

Prof.Dr. Fabiano Guasti Lima

Universidade de São Paulo

Ao meu filho Renan, eterna fonte de inspiração, energia e aprendizado.

Ao meu marido Carlos, pelo amor, compreensão e companheirismo.

Aos meus pais, pela vida.

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador o Prof. Dr. Herbert Kimura pela dedicação e paciência durante toda a elaboração do trabalho, que esteve sempre presente e que me motivou nos momentos mais difíceis e solitários desta jornada.

Agradeço ao Prof. Dr. Leonardo Fernando Cruz Basso que sempre esteve disposto a discutir meu trabalho e contribuir com seus comentários e sugestões.

Agradeço também ao Prof. Umberto Guarnier Mignozzetti que contribuiu e me ajudou nas dúvidas da ferramenta SPSS e no desenvolvimento do algoritmo de *boosting* no *software* R-Project.

E um último e especial agradecimento a pessoa que de forma anônima tornou possível o desenvolvimento do trabalho, pois cedeu a amostra das operações de microcrédito que foram utilizadas no trabalho.

“A coisa mais bela que o homem pode experimentar é o mistério. É esta a emoção fundamental que está na raiz de toda ciência e arte. O homem que desconhece esse encanto, incapaz de sentir admiração e estupefação, esse já está, por assim dizer, morto e tem os olhos extintos.”

Paráfrase de Albert Einstein

RESUMO

As experiências do Grameen Bank com operações de microcrédito têm sido reproduzidas em vários países, principalmente as relacionadas com as duas grandes inovações neste mercado: o papel do agente de crédito e o mecanismo de grupo solidário.

A massificação das operações e a redução de custos tornam-se imprescindíveis para que haja economia de escala e maior apetite para as IMFs ampliarem sua atuação neste mercado. Neste cenário, a implantação de modelos de *credit scoring* será a próxima inovação do microcrédito e proporcionará agilidade, redução de riscos e, conseqüentemente, redução dos custos.

Com base em informações históricas de operações de microcrédito foi elaborado um modelo de crédito. Foram identificadas variáveis chave que permitem distinguir os bons e maus pagadores. Os resultados mostram que, acoplando-se técnicas de linguagem de máquina – *bagging e boosting* – aos métodos tradicionais de análise de crédito – análise discriminante e regressão logística –, obtém-se melhora na *performance* dos modelos de *credit scoring* para microcrédito.

Palavras chave: microcrédito; *credit scoring*; inovação; *bagging*; *boosting*.

ABSTRACT

The Grameen Bank experiences with microcredit operations have been imitated in various countries, mainly the ones related to the two great innovations in this market: the credit agent's role and the solidary group mechanism.

The massification of the operations and the reduction in their costs become vital for economies of scale to be achieved, as well as a greater appetite for the MFIs to expand their activity in the microcredit market. In this context, the next great innovation in the microcredit market will be the introduction of credit scoring models in such operations. This will speed up the process, reduce the risks and consequently the costs.

Historical information about microcredit operations was taken into account for the creation of a credit model. It was then possible to identify key variables that help to distinguish between the good and the bad borrowers. The results show that as machine learning techniques – bagging and boosting – are added to the traditional methods of credit analysis – discriminant analysis and logistic regression –, an improvement in the performance of the credit scoring models for microcredit can be achieved.

Key words: microcredit; credit scoring; innovation; bagging; boosting.

LISTA DE FIGURAS

Figura 1: Valor da Exigibilidade em Milhões de R\$. Elaborado pela autora.....	18
Figura 2: Depósito interfinanceiro vinculado a operações de microfinanças. Elaborado pela.	18
Figura 3: Percentual cumprimento da exigibilidade pelas IMFs. Elaborado pela autora.....	19
Figura 4: Participação das IMFs por Região. Elaborado pela autora.	22
Figura 5: Evolução do Saldo da Carteira de Microcrédito. Elaborado pela autora.	24
Figura 6: Evolução Produção Mensal Operações de Microcrédito. Elaborado pela autora. ...	24
Figura 7: Evolução Quantidade Mensal Operações de Microcrédito. Elaborado pela autora..	25
Figura 8: Evolução Ticket Médio Operações de Microcrédito. Elaborado pela autora.	25
Figura 9: Modelo simples de aprendizagem de máquina - Haykin, 1999 - pág.61	41
Figura 10: Dados da base de contratos consolidado com a base de clientes. Elaborada pela autora com base no dicionário de dados das bases.	50
Figura 11: Percentual de acertos por Iterações. Elaborado pela autora.....	75
Figura 12: Percentual de acertos por Iterações. Elaborado pela autora.....	75
Figura 13: Percentual de acertos por Iterações. Elaborado pela autora.....	76

LISTA DE TABELAS

Tabela 1: Informações das MIFs, 2008. Elaborado pela autora.	22
Tabela 2: Evolução do mercado mundial de microcrédito. Elaborado pela autora.	22
Tabela 3: Resumo artigos acadêmicos sobre modelos de credit scoring de operações de microcrédito ou operações de crédito em países em desenvolvimento. Elaborado pela autora.	32
Tabela 4: Bons e Maus Pagadores por Predominância de Sexo nos Grupos Solidários. Elaborado pela autora.	59
Tabela 5: Percentuais de Bons e Maus Pagadores por Variável da Base. Elaborado pela autora.	61
Tabela 6: Variáveis da Base de Dados. Elaborado pela autora.	63
Tabela 7: Resultados dos Testes de Hipótese. Elaborado pela autora.	65
Tabela 8: Resultados das médias e desvios padrão de cada variável. Elaborado pela autora. .	67
Tabela 9: Resultados SPSS Análise Discriminante e Regressão. Elaborado pela autora.	70
Tabela 10: Resultado dos modelos nas amostras de calibração e validação. Elaborado pela autora.	72
Tabela 11: Resultados Bagging e Boosting. Elaborado pela autora.	74
Tabela 12: Resultados Bagging e Boosting com quantidades diferentes de iterações. Elaborado pela autora.	76

SUMÁRIO

1	INTRODUÇÃO.....	11
2	REFERENCIAL TEÓRICO.....	12
2.1	MICROCRÉDITO E MICROFINANÇAS	13
2.2	AS INOVAÇÕES NO MICROCRÉDITO	19
2.3	O HISTÓRICO E A EVOLUÇÃO DO MICROCRÉDITO.....	20
2.4	MICROCRÉDITO: ESTUDOS E SUAS CARACTERÍSTICAS.....	25
2.5	MODELOS DE ANÁLISE DE CONCESSÃO DE CRÉDITO.....	32
2.5.1.	Regressão Linear	35
2.5.2.	Análise Discriminante	36
2.5.3.	Regressão Logística.....	38
2.5.4.	Outros Modelos	39
2.6	TÉCNICAS DE APRENDIZAGEM DE MÁQUINA	40
2.6.1.	<i>Ensemble Methods</i>	42
2.6.2.	<i>Bagging</i>	44
2.6.3.	<i>Boosting</i>	45
3	OBJETIVO	48
4	METODOLOGIA.....	49
4.1	DESCRIÇÃO	49
4.2	BASE DE DADOS	50
4.3	O MICROCRÉDITO NA PRÁTICA	51
4.3.1	Levantamento Sócio-Econômico.....	52
4.3.2	Grupo Solidário	55
4.4	FORMATAÇÃO DA BASE DE DADOS	56
4.5	AS VARIÁVEIS E OS TESTES DE HIPÓTESE.....	62
4.6	O DESENVOLVIMENTO DO MODELO DE CRÉDITO	66
4.6.1	Calibração e Resultados.....	66
4.6.2	Validação	72
4.7	<i>BAGGING E BOOSTING</i>	73
5	CONCLUSÕES	78
6	REFERÊNCIAS BIBLIOGRÁFICAS	82
	APÊNDICE A: Algoritmo de <i>Boosting</i>	86
	APÊNDICE B: Algoritmo da Análise Discriminante	88
	APÊNDICE C: Algoritmo da Regressão Logística.....	88
	APÊNDICE D: Algoritmo de Geração de Amostras com Repetição.....	89
	APÊNDICE E: Algoritmo COMPLEX SAMPLE – SPSS	90

1 INTRODUÇÃO

A discussão sobre *credit scoring* em microcrédito é recente e poderá ajudar a reduzir os custos das operações de crédito para o público de baixa renda, se tornando a próxima inovação neste mercado.

Atualmente, a prática de análise de crédito das operações de microcrédito envolve análise individualizada de informações levantadas pelos agentes de crédito em entrevistas pessoais, que são conduzidas no próprio local onde a atividade dos empreendedores é exercida. Esses agentes recebem capacitação para efetuarem as entrevistas, serem consultores financeiros, acompanharem as operações durante toda a sua vida e analisarem o risco dos potenciais tomadores. A análise de crédito neste segmento ainda é efetuada de forma subjetiva.

O processo todo envolve um elevado custo, principalmente considerando os baixos volumes financeiros envolvidos em cada operação. Desta forma, uma estratégia sustentável para atuação e crescimento do microcrédito prescinde de ganhos de escala.

Este trabalho objetiva desenvolver modelos de classificação de risco de tomadores de operações de microcrédito, utilizando-se de uma amostra de operações de microcrédito realizadas por uma instituição financeira com forte atuação no mercado brasileiro. Aplicando-se técnicas estatísticas e computacionais baseadas na análise discriminante e na regressão logística, busca-se identificar variáveis e um modelo capaz de prever o risco de atraso de pagamento, distinguindo-se os bons e os maus pagadores em operações de microcrédito.

Neste contexto, o presente estudo visa enriquecer os trabalhos acadêmicos sobre o microcrédito, que constitui importante mecanismo de inclusão e ascensão social do público de baixa renda e cuja análise quantitativa ainda é pouco explorada do ponto de vista acadêmico, principalmente no contexto brasileiro.

A análise de concessão de crédito através de modelos quantitativos, automatizados e massificados pode ajudar a reduzir os custos das operações e tornar este mercado mais atrativo e competitivo e trazer benefícios na análise de risco de atraso e inadimplência.

2 REFERENCIAL TEÓRICO

Segundo relatório do MicroBanking Bulletin (2009), aproximadamente um bilhão de pessoas, perto de um sexto de toda a população do planeta, sobrevive com menos de um dólar por dia. Desta parte da população, 162 milhões vivem com menos de meio dólar por dia.

O mercado para o segmento de baixa renda tem crescido ao longo do tempo. De acordo com Prahalad (2009), quatro bilhões de pobres podem ser a engrenagem para a próxima rodada de comércio e prosperidade global. As empresas terão que trabalhar colaborativamente com a comunidade civil e os governos locais para o desenvolvimento do mercado para a população da base da pirâmide, criando milhões de novos empreendedores, desde mulheres trabalhadoras a distribuidores e microempresários (PRAHALAD, 2009).

Tradicionalmente, as pessoas de baixa renda têm pouco acesso ao crédito. Os altos custos, associados aos empréstimos de baixo valor financeiro, e a ausência de garantias para estas operações fazem com que o segmento de baixa renda não seja atrativo para a indústria financeira formal (MCKERNAN, 2002).

As altas taxas de juros envolvidas nas operações de crédito para o público de baixa renda dificultam ainda mais o acesso ao mercado financeiro e à obtenção de recursos e reduzem as oportunidades e possibilidades para transpor a linha da pobreza. Programas de microcrédito, como o do Grameen Bank, oferecem inovação e uma promessa de entrega de crédito para os pobres (MCKERNAN, 2002). O Grameen Bank e outros programas de microcrédito não disponibilizam somente crédito ao segmento de baixa renda; oferecem também outros serviços não relacionados ao crédito, como, por exemplo, programas de treinamento vocacional, ajuda no planejamento organizacional e desenvolvimento social (MCKERNAN, 2002).

Com o crescimento da demanda de crédito para o público de baixa renda e o aumento das oportunidades para as instituições que operam com microcrédito, várias inovações têm sido criadas nos programas de microcrédito, dentre elas, os programas de crédito baseados em grupos solidários: sucesso em países em desenvolvimento.

Programas que replicaram basicamente o modelo do Grameen Bank podem ser encontrados em vários países, como Estados Unidos, França, Itália e Canadá (RAHMAN, 1993 apud MCKERNAN, 2002). Em regiões carentes próximas às grandes cidades ou em áreas rurais pobres dos Estados Unidos, empréstimos baseados em grupos solidários são um caminho criativo para conceder crédito a microempreendedores – pessoas que possuem pequenos negócios, mas não tem acesso ao crédito em bancos e no mercado financeiro tradicional (LEDGERWOOD, 1999).

O modelo proposto pelo Grameen Bank reduz o custo das transações de empréstimos e oferece crédito a grupos solidários formados por pessoas pobres. A pressão do próprio grupo elimina a necessidade de garantias, encoraja o grupo e dá suporte emocional ao grupo de microempreendedores. Contratos com grupos solidários mitigam o risco de problemas criados entre credores e devedores. Vizinhos agora têm incentivos para monitorar uns aos outros, promovendo maior índice de repagamento e menor risco, mesmo com ausência de garantias (MORDUCH, 1999).

2.1 MICROCRÉDITO E MICROFINANÇAS

Crédito resume-se basicamente em captar recursos no mercado e utilizar estes recursos para financiar as necessidades de consumo e investimento (SILVA, 2006).

O crédito possibilita que as pessoas físicas tenham suas necessidades de consumo e investimento atendidas, quando lhes faltam recursos próprios disponíveis para este fim. Possibilita também que as empresas tenham uma alternativa para suprir suas necessidades de recursos para capital de giro e investimentos.

Os termos microcrédito e microfinanças são frequentemente utilizados como sinônimos (Estudo do Banco Central do Brasil, 2008). Cada qual possui uma definição teórica, mas o debate sobre os significados continuam.

Segundo Julie Drolet (2009), programas de microcrédito tendem a oferecer distribuição de crédito por organizações não governamentais (ONG), que são organizações voluntárias, sem fins lucrativos e que dependem de financiamento externo. Para a autora,

microfinanças são programas que fornecem uma gama de serviços financeiros, incluindo crédito, seguros, poupança e procuram gerar lucros para se tornarem auto-suficientes.

De acordo com o *Microcredit Summit Campaign*, 2008, microcrédito é um programa de crédito de pequeno valor e outros serviços financeiros, como poupança, para pessoas muito pobres que desejam empreender em projetos, gerando renda para sua subsistência.

Microfinanças tem evoluído como uma abordagem de desenvolvimento econômico que pretende beneficiar homens e mulheres de baixa renda. O termo refere-se à oferta de serviços financeiros para clientes de baixa renda, incluindo autônomos. Serviços financeiros geralmente incluem crédito e poupança; todavia, algumas organizações de microfinanças também incluem seguro e serviços de pagamentos. Microfinanças não é simplesmente um banco, é uma ferramenta de desenvolvimento. Suas atividades usualmente envolvem: (i) pequenos empréstimos, geralmente para capital de giro; (ii) avaliação informal de tomadores e investimentos; (iii) grupo solidário ou poupança compulsória em substituição de garantias; (iv) acesso a empréstimos maiores baseados na *performance* de pagamento; (v) agilidade no desembolso de empréstimo e monitoramento; (vi) produtos de seguro e poupança (LEDGERWOOD, 1999).

No Brasil, é comum confundir o conceito de microcrédito com o conceito do crédito popular (MONZONI, 2007). Crédito popular são créditos de baixos valores oferecidos à população de baixa renda. Não possuem uma metodologia específica para sua geração e nem acompanhamento de pessoas capacitadas – os chamados agentes de crédito – para que estes recursos sejam utilizados para fins produtivos e para a geração de renda das famílias. Estes recursos geralmente são utilizados para consumo ou pagamento de outras dívidas (MONZONI, 2007).

Já o microcrédito produtivo orientado é o crédito concedido ao público de baixa renda – pessoas físicas ou jurídicas – empreendedor de atividades produtivas de pequeno porte. Este crédito visa o fomento da atividade produtiva com recursos para capital de giro ou investimento, considerando ainda que: (i) o atendimento seja efetuado por pessoa treinada para efetuar o levantamento sócio-econômico e prestar orientação para o negócio; (ii) o contato com o pequeno empreendedor deva ser mantido durante todo o período em que o

contrato de crédito esteja vigente, visando o melhor aproveitamento dos recursos na atividade fim; (iii) o valor e condições do crédito devam ser definidas após a avaliação da atividade do empreendedor e da sua capacidade de pagamento (MONZONI, 2007).

Microfinanças tem crescido por várias razões:

- (i) a promessa de estar ao alcance dos pobres: suporta geração de renda para o desenvolvimento operacional das famílias de baixa renda;
- (ii) a promessa de sustentabilidade financeira: pode ajudar a construir a independência financeira;
- (iii) o potencial de crescimento em sistema tradicional: microfinanças às vezes imita as atividades do sistema bancário tradicional, promovendo os serviços por meios similares, mas com maior flexibilidade e preços mais ajustados ao microempreendedor e com uma maior base de sustentabilidade. Isto pode tornar os serviços de microfinanças mais atrativos para um grande número de clientes de baixa renda;
- (iv) a contribuição para fortalecer e expandir um sistema financeiro formal: microfinanças pode fortalecer sistemas financeiros formais como poupança, cooperativas de crédito, redes de crédito e etc., expandindo o mercado potencial e lucratividade;
- (v) o aumento do número de histórias de sucesso: casos bem-sucedidos que contrastam com a dura realidade das organizações que falharam em termos de sustentabilidade e de combate à pobreza;
- (vi) a viabilidade de melhores produtos financeiros como resultado das experiências e da inovação: inovações têm se mostrado capazes de suprir a falta de garantias com grupos solidários e abordagens baseadas no caráter, resolvendo problemas de não pagamento, usando também pressão social e a promessa de valores de crédito crescente para os bons pagadores, baixando os custos de transação, aumentando a *performance* de pagamento e oferecendo serviços de poupança que vão ao encontro das necessidades dos pequenos poupadores (LEDGERWOOD, 1999).

Microfinanças surgiram para oferecer uma solução ganha-ganha, na qual as organizações financeiras e os clientes pobres lucram (MORDUCH, 1999).

No Brasil, para efeito da Resolução 3.310/05, são consideradas operações de microcrédito aquelas realizadas com: (i) pessoas físicas detentoras de contas de depósitos especiais de saldo médio inferior a R\$1.000,00, para viabilizar empreendimentos de natureza profissional, comercial ou industrial de pequeno porte, e com pessoas jurídicas classificadas como microempresas na forma da legislação em vigor; (ii) pessoas físicas de baixa renda detentoras ou não de contas de depósitos ou aplicações financeiras; (iii) pessoas físicas ou jurídicas empreendedoras de atividades produtivas de pequeno porte, com renda anual bruta de até R\$60.000,00.

As operações de microcrédito no mercado brasileiro devem observar ainda: (i) as taxas de juros efetivas não podem exceder 2% a.m. (dois por cento ao mês) e 4% a.m. (quatro por cento ao mês) nas operações de microcrédito produtivo orientado; (ii) o valor do crédito não pode ser superior a R\$600,00 (seiscentos reais) para pessoas físicas, R\$1.500,00 (um mil e quinhentos reais) para operações para microempreendedores, R\$5.000,00 (cinco mil reais) quando se tratar de microcrédito produtivo orientado; (iii) o prazo da operação não pode ser inferior a 120 dias; (iv) a taxa de abertura de crédito para pessoas físicas pode ser de até 2% (dois por cento) e de até 4% (quatro por cento) para pessoas jurídicas/microempresas.

Constituem-se garantias de operações de microcrédito, segundo a legislação brasileira vigente: (i) aval solidário em grupo com no mínimo três participantes; (ii) alienação fiduciária; (iii) fiança; (iv) outras garantias aceitas pela IMF.

Incluem-se no âmbito do Programa de Microcrédito Produtivo Orientado – PNMPO, instituídos pela Lei 11.110, de 25 de abril de 2005, as operações concedidas de acordo com as seguintes condições:

- Operações concedidas pelos bancos múltiplos com carteira comercial, bancos comerciais, Caixa Econômica Federal e pelas instituições definidas e compreendidas na referida Lei: (i) cooperativas singulares de crédito; (ii) agências de fomento; (iii) sociedades de crédito ao microempreendedor; (iv) organizações de sociedade civil de interesse público;
- Destinadas ao financiamento de bens, serviços e capital de giro, essenciais ao empreendimento;
- Utilize da metodologia baseada no relacionamento direto com o empreendedor no local onde é executada a atividade econômica, considerando: (i) atendimento por

pessoas treinadas para efetuar o levantamento sócio-econômico e prestar orientação educativa sobre o planejamento do negócio, para definição das necessidades de crédito e de gestão para o desenvolvimento do negócio; (ii) o contato com o tomador do recurso deve ser mantido durante todo o período do contrato, para acompanhamento e orientação, visando crescimento e sustentabilidade do negócio; (iii) os valores e condições do crédito devem ser definidos após a avaliação da atividade do empreendedor e da capacidade de endividamento do tomador.

A mesma resolução, 3.310/05, estabelece que os bancos múltiplos com carteira comercial, os bancos comerciais e a Caixa Econômica Federal devem manter aplicados em operações de microcrédito destinadas à população de baixa renda e a microempreendedores valor correspondente a, no mínimo, 2% dos saldos dos depósitos à vista captados pela instituição. Para o cumprimento da exigibilidade serão considerados: (i) os recursos repassados para outras instituições financeiras, por meio de depósito interfinanceiro vinculado a operações de microfinanças (DIM), exclusivamente para aplicações em operações de microcrédito, observadas as disposições da Resolução 3.399, de 29 de agosto de 2006, e regulamentação complementar; (ii) os créditos oriundos de operações de adiantamentos, empréstimos e financiamentos que atendam às condições estabelecidas nesta resolução, adquiridos de: (a) outras instituições financeiras; (b) organizações da sociedade civil de interesse público; (c) organizações não governamentais cujos estatutos prevejam a realização de operações de microcrédito; (d) entidades, fundos ou programas voltados para o microcrédito.

A figura 1 mostra o crescimento do montante de depósito à vista no mercado financeiro brasileiro, base para o cálculo da exigibilidade. Quanto maior o montante de depósito à vista, maior o montante exigido pelo Banco Central a ser destinado pelas IMFs nas operações de microcrédito.

A figura 2 mostra que os depósitos interfinanceiros que as instituições efetuam vinculados a operações de microcrédito (DIM) reportadas ao Banco Central, de acordo com as regras da Resolução 3.422, têm crescido desde 2004; mas não com a mesma velocidade do volume de depósito à vista do mercado financeiro.

A figura 3 mostra o histórico do total de recursos aplicados nas operações de microcrédito (DIM+Carteira de Microcrédito) sobre a exigibilidade, onde é possível observar que o percentual de cumprimento de exigibilidade se mantém próximo ao patamar de 40%. Para que os bancos estivessem utilizando toda a exigibilidade nestas operações, este percentual deveria estar num patamar de 100%. Em pouquíssimos períodos da serie estudada este percentual atingiu patamares superiores a 40%-50%. Isso mostra que apesar dos incentivos governamentais, o mercado de microcrédito ainda é incipiente e tem espaço para crescimento.

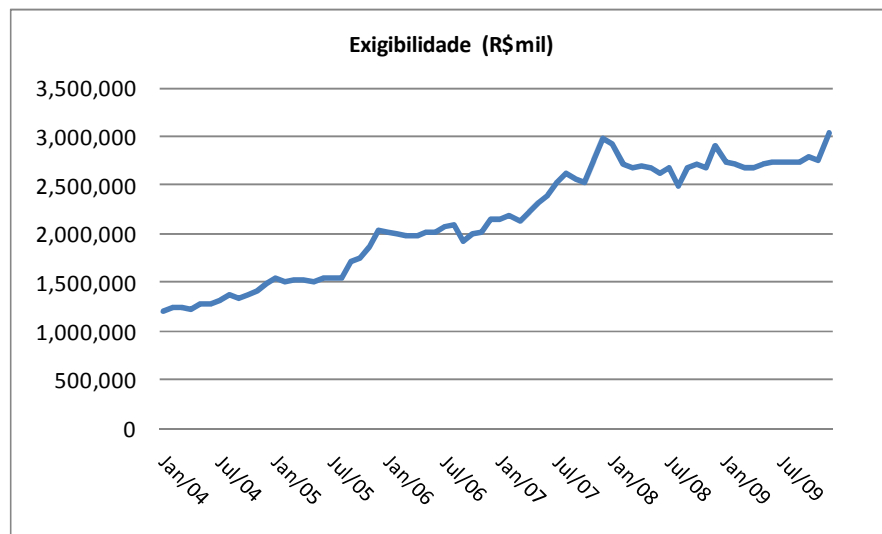


Figura 1: Valor da Exigibilidade em Milhões de R\$. Elaborado pela autora. Fonte: Banco Central do Brasil, site www.bcb.gov.br.

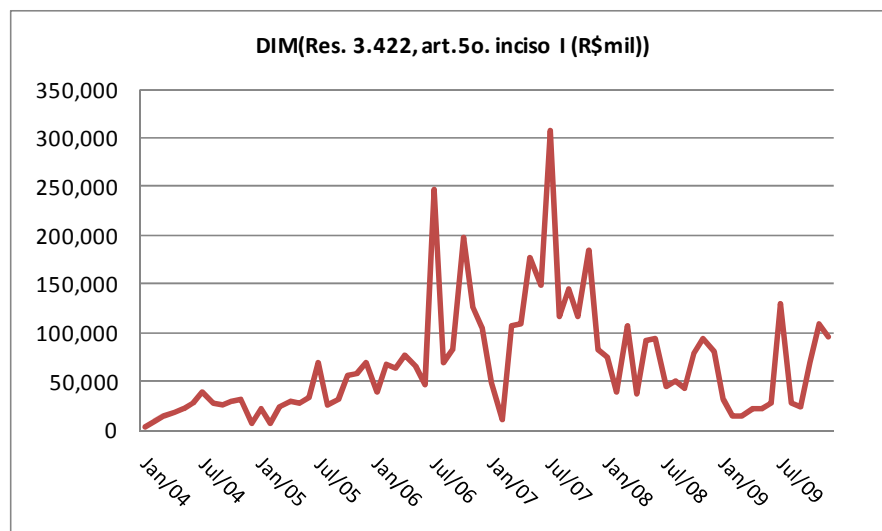


Figura 2: Depósito interfinanceiro vinculado a operações de microfinanças. Elaborado pela autora. Fonte: Banco Central do Brasil, site www.bcb.gov.br.

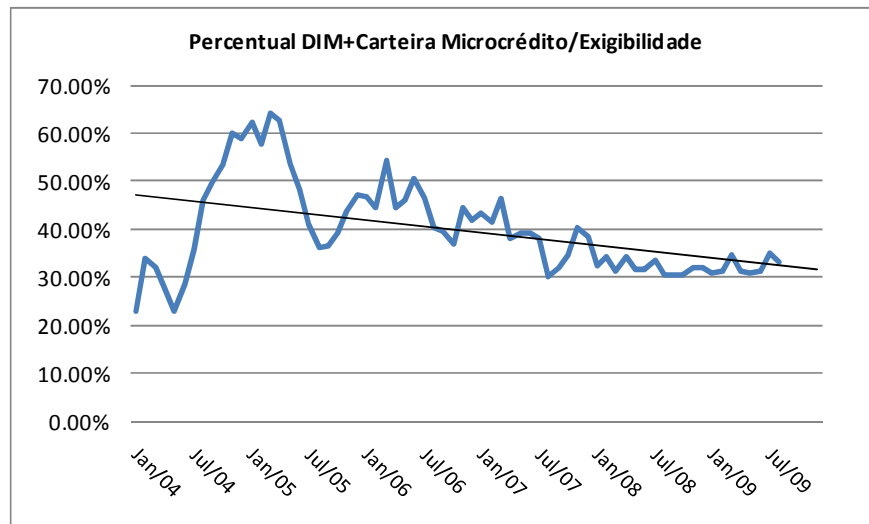


Figura 3: Percentual cumprimento da exigibilidade pelas IMFs. Elaborado pela autora.
 Fonte: Banco Central do Brasil, site www.bcb.gov.br.

2.2 AS INOVAÇÕES NO MICROCRÉDITO

O movimento de microfinanças tem mostrado que, apesar dos altos custos de transação e ausência de garantias, é possível emprestar de forma lucrativa para as famílias de baixa renda. Experiências têm mostrado que muitas famílias pobres podem poupar uma quantia quando lhes são dados os incentivos e meios para a poupança, o que sugere que um caminho para endereçar a restrição que estas famílias pobres têm ao mercado financeiro tradicional é incentivar e criar mecanismos de poupança em vez de focar apenas no crédito (MORDUCH, 1999).

Os grupos solidários do Grameen Bank formam-se voluntariamente; enquanto os empréstimos são efetuados individualmente, todos os membros do grupo são responsáveis pelo pagamento do empréstimo total do grupo. Cada grupo consiste em 5 devedores, com empréstimo primeiro para 2, depois para mais 2 e então para o 5º. membro do grupo. Estes grupos de 5 membros encontram-se semanalmente com outros 7 grupos. Representantes da organização se reúnem com 40 pessoas por vez. De acordo com as regras, se um dos membros entrar em *default*, todos os membros do grupo são negados para empréstimos subsequentes. (MORDUCH, 1999).

As datas de vencimentos dos empréstimos para a baixa renda costumam ser flexíveis: alguns devedores pagam semanalmente e outros, mensalmente. Os prazos dos empréstimos também são flexíveis e geralmente ficam entre 4 e 24 (MORDUCH, 1999).

Poucos programas requerem garantias; muitos têm substitutos, como por exemplo, programas que seguem o modelo do Grameen Bank e que requerem que os devedores contribuam com um fundo de emergência que representa 0.5% do valor emprestado. O fundo de emergência serve como seguro em casos de *default*, morte, incapacidade e etc., em valores proporcionais aos membros do grupo (MORDUCH, 1999).

A promessa das microfinanças está calcada na inovação: novas estruturas de gestão, novos contratos e novas atitudes. Enquanto os contratos com grupos solidários são a inovação mais comemorada em microfinanças, todos os programas usam uma variedade de outras inovações que podem ser muito importantes – especialmente pelas formas dinâmicas de incentivos ao pagamento e datas de pagamentos ajustadas as capacidades de pagamento dos empreendedores (MORDUCH, 1999).

O scoring é provavelmente a próxima importante inovação tecnológica em microfinanças. Porém, o scoring não substituirá os grupos solidários ou os agentes de crédito, e nunca terá a mesma eficiência que nos países ricos, pois muito do risco do microcrédito não está relacionado com características que podem ser quantificadas de forma barata. Muitos dos maiores oferecedores de microcrédito terão o *credit scoring* como a mais importante ferramenta de decisão (SCHREINER, 2000).

2.3 O HISTÓRICO E A EVOLUÇÃO DO MICROCRÉDITO

Muitos autores relatam a iniciativa de Muhammad Yunus – o Grameen Bank, em 1976 – como sendo a primeira em microfinanças. Outros autores relatam experiências muito anteriores como sendo as pioneiras.

Segundo Monzoni (2007), em 1864 houve a primeira iniciativa de microfinanças no sul da Alemanha, quando um inverno rigoroso tornou os agricultores dependentes de agiotas para a obtenção de recursos financeiros. Um pastor chamado Raiffeinsen passou a fornecer

aos fazendeiros farinha de trigo, o que possibilitou a produção e comercialização de pão, gerando capital de giro para os fazendeiros. Essa iniciativa cresceu, formando a chamada “Associação do Pão”, que se tornou uma cooperativa de crédito para a população carente.

Várias outras iniciativas são conhecidas; porém, a mais divulgada, copiada e reconhecida no mundo todo, sem dúvida é a do professor Muhammad Yunus, o Grameen Bank.

A Microfinance Information eXchange (The MIX) é uma organização não-governamental sediada em Washington, DC, nos Estados Unidos, fundada em 2002, cujo principal objetivo é promover o intercâmbio de informações no mercado de microfinanças. O The MIX tem dois produtos muito conhecidos internacionalmente: o MIX MarketTM e o MicroBanking Bulletin (MBB).

No site <http://www.mixmarket.org> é possível obter o MicroBanking Bulletin e encontrar informações sobre o mercado de microfinanças mundial. Esta é uma importante ferramenta de disseminação de informações que contribui para o fortalecimento do mercado (MONZONI, 2007).

De acordo com o Mix Market, o mercado de microfinanças tinha em 2008 (último dado disponível) 1.781 IMFs com um total 76 MM de clientes. Estes dados consideram apenas as instituições que reportam os dados de suas operações para o para o Mix Market ou MicroBanking Bulletin.

A figura 4 mostra a participação das instituições de microcrédito por região, de acordo com a classificação da Mix Market.

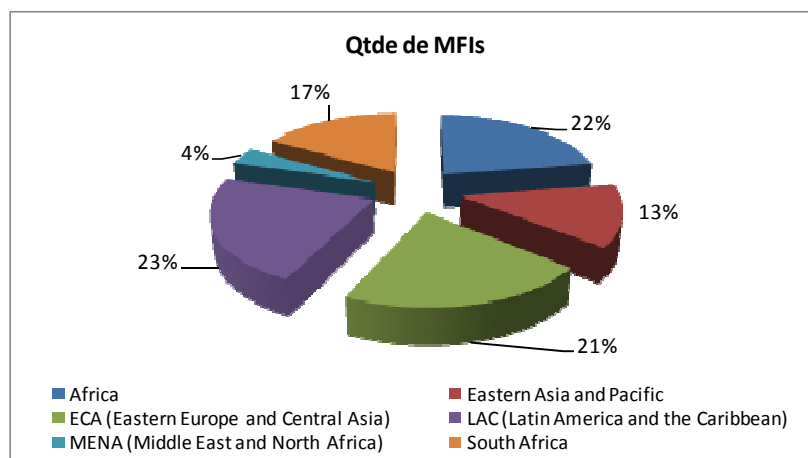


Figura 4: Participação das IMFs por Região. Elaborado pela autora.

Fonte: The *MIX Market*TM, site www.mixmarket.gov.

A tabela 1 mostra um resumo do mercado mundial de microfinanças, de acordo com a posição de 2008 do Mix Market. A partir destes dados é possível observar que o Oriente Médio e o norte da África são as regiões com menor quantidade de IMFs, enquanto que o Caribe e a América Latina possuem a maior quantidade de IMFs.

	Africa	Eastern Asia and Pacific	ECA (Eastern Europe and Central Asia)	LAC (Latin America and the Caribbean)	MENA (Middle East and North Africa)	South Africa	Total
MFIs	399	237	381	400	66	298	1781
Gross Loan (bilhões USD)	0.31	4.60	10.00	15.40	1.30	4.70	36.31
Ativos (bilhões USD)	5.20	5.40	13.40	19.40	1.60	7.00	52.00
Ticket Médio (USD)	307.30	316.00	2,002.40	838.20	358.00	126.20	-
Qtde de Clientes (milhões)	6.70	10.90	3.00	12.80	2.70	40.20	76.30
Depósitos (bilhões USD)	2.80	2.00	6.00	9.80	0.07	1.90	22.57
Qtde de depositantes (milhões)	16.80	4.30	5.20	13.30	0.10	30.80	70.50

Tabela 1: Informações das MIFs, 2008. Elaborado pela autora.

Fonte: The *MIX Market*TM, site www.mixmarket.gov.

Os dados de evolução do mercado mundial, segundo a mesma fonte, mostram que a quantidade de tomadores de microcrédito cresceu, de 2003 a 2008, à taxa média de 20% ao ano. Se considerarmos a carteira de microcrédito, o crescimento médio foi de 34.2% ao ano, conforme tabela 2:

% de Crescimento anual do mercado	2003	2004	2005	2006	2007	2008	Média
Qtde de tomadores (%)	14.50	21.00	26.50	22.00	22.00	20.00	21.00
Gross Loan Portfolio (%)	32.80	37.80	32.40	38.10	45.40	20.10	34.20
Qtde de poupadores (%)	13.20	11.20	12.10	11.70	1.50	25.90	12.40
Volume de Poupança Voluntária	33.90	22.90	13.80	29.10	31.90	16.80	24.50

Tabela 2: Evolução do mercado mundial de microcrédito. Elaborado pela autora.

Fonte: The *MIX Market*TM, site www.mixmarket.gov.

Segundo Ribeiro e Carvalho (2004), o Brasil não teve a mesma expansão no setor de microfinanças que ocorreu na Ásia, África e outros países da América Latina. O mercado brasileiro, apesar do seu grande potencial, tem crescido de forma lenta.

O mercado brasileiro teve sua primeira iniciativa de microcrédito registrada em 1973. Esta iniciativa, chamada Projeto UNO (União Nordestina de Assistência a Pequenas Organizações), não alcançou grandes proporções, mas foi o ponto de partida para a criação da rede Ceape (Centro de Apoio ao Pequeno Empreendedor) (RIBEIRO e CARVALHO, 2004).

A maior iniciativa brasileira foi o Programa CrediAmigo, que surgiu em 1997, pelo Banco do Nordeste (estatal federal), com apoio do Banco Mundial e a assistência técnica da Áccion Internacional (RIBEIRO e CARVALHO, 2004).

No final dos anos 1990, notou-se um maior envolvimento governamental no sentido de estabelecer um arcabouço legal e institucional para o setor, viabilizando sua expansão. Como principais legislações, podemos citar também a lei 9.790, de 23 de março de 1999 – que regulamenta o terceiro setor – e a lei 10.194, de 14 de fevereiro de 2001 – que institui um novo modelo institucional para as Sociedades de Crédito ao Microempreendedor, as SCMs (RIBEIRO e CARVALHO, 2004).

Segundo dados extraídos do site do Banco Central do Brasil, é possível observar que, em dezembro de 2009, o Brasil possuía saldo da carteira de microcrédito da ordem de R\$ 1.7 bilhões, considerando as operações de microcrédito para consumo (56% da carteira) e microcrédito empreendedor (44% da carteira). A produção de operações de microcrédito consumo e empreendedor em dezembro de 2009 foi de R\$ 318 milhões, 65% dessa oriundos do microcrédito empreendedor. No mesmo período, foram produzidos 912.000 contratos com ticket médio de R\$ 348,00.

As figuras de 5 a 8, mostram a evolução do mercado de microcrédito no Brasil. As figuras permitem concluir que há um maior crescimento das operações de microcrédito produtivo, ou seja, o crédito destinado ao empreendedorismo e geração de renda. O mercado de operações de microcrédito com a finalidade de consumo está estabilizado no mesmo patamar desde 2006.

Outra questão relevante é que há um consistente aumento do ticket médio das operações de microcrédito. Este mesmo crescimento não é observado nas quantidades de operações de microcrédito, o que demonstra que ainda há no Brasil uma grande parte da população de baixa renda e de microempreendedores que não têm acesso às operações de microcrédito. Esta parte da população também não é atendida pelo mercado financeiro tradicional.

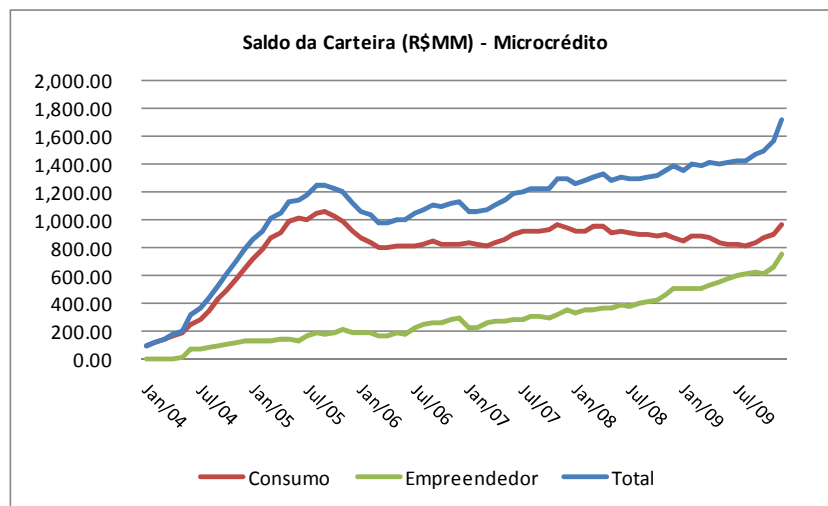


Figura 5: Evolução do Saldo da Carteira de Microcrédito. Elaborado pela autora.
Fonte: Banco Central do Brasil, site www.bcb.gov.br.

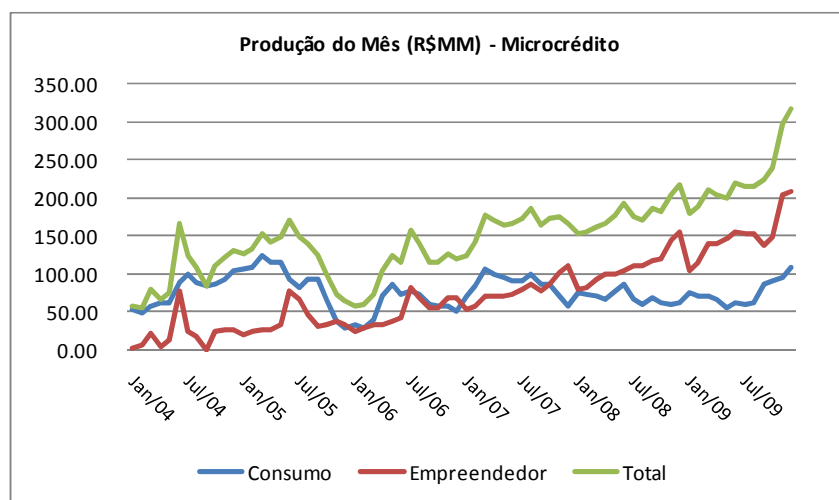


Figura 6: Evolução Produção Mensal Operações de Microcrédito. Elaborado pela autora.
Fonte: Banco Central do Brasil, site www.bcb.gov.br.

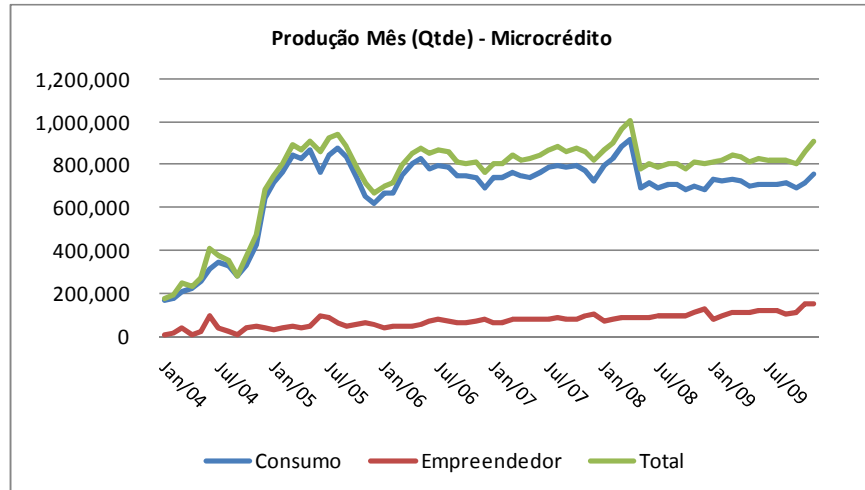


Figura 7: Evolução Quantidade Mensal Operações de Microcrédito. Elaborado pela autora.
Fonte: Banco Central do Brasil, site www.bcb.gov.br.

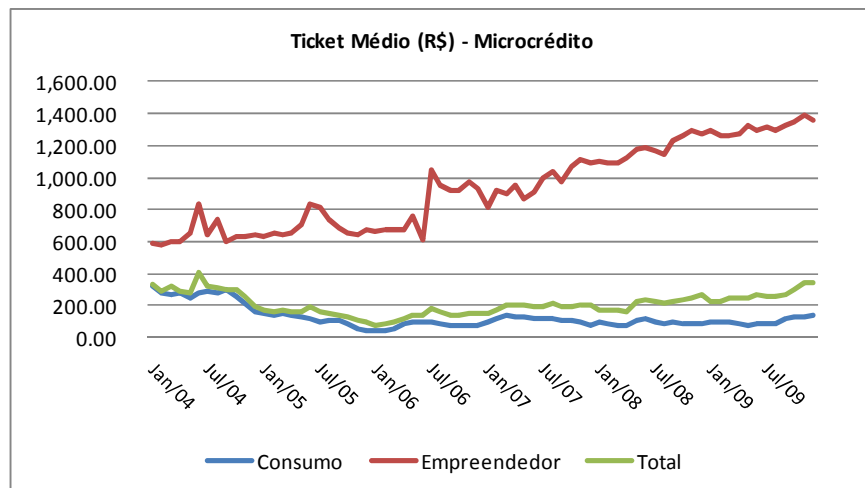


Figura 8: Evolução Ticket Médio Operações de Microcrédito. Elaborado pela autora.
Fonte: Banco Central do Brasil, site www.bcb.gov.br.

2.4 MICROCRÉDITO: ESTUDOS E SUAS CARACTERÍSTICAS

Nos anos 80 e 90, as políticas deram um grande salto, argumentando que as novas organizações de microfinanças deveriam ser lucrativas – ou usando um código de conduta – deveriam ser sustentáveis. Este argumento tem 3 etapas: (i) é esperado que pequenos empréstimos sejam mais caros para os bancos administrarem e as famílias pobres devam pagar altas taxas de juros. O argumento é que acesso ao crédito é mais importante que preço para a população de baixa renda; (ii) os subsídios são a raiz do problema nos bancos públicos e para as organizações não governamentais. Subsídios podem enfraquecer incentivos para a

inovação e redução de custos; (iii) os subsídios não estão disponíveis em quantia necessária para ser o combustível de crescimento do setor. O argumento que microfinanças deve procurar lucratividade sem perder o foco na geração de renda é o que tem ressonância com o apelo ganha-ganha, admitindo um pequeno *trade-off* entre os objetivos sociais e econômicos (CULL, DERNIGUÇ-KUNT e MORDUCH, 2009).

Freqüentemente, as instituições concedem os primeiros créditos com menores valores, que vão aumentando com tempo. Este é um mecanismo que reduz o risco da instituição financeira e que cria um incentivo para que os clientes efetuem o pagamento de suas obrigações, com a promessa de maiores valores nos empréstimos futuros, fornecendo também a instituição um histórico de crédito do cliente de baixa renda (LEDGERWOOD, 1999).

Os prazos dos empréstimos afetam a *performance* do pagamento, as receitas das instituições de microcrédito, os custos financeiros para o cliente e, por último, o uso adequado do empréstimo. Quanto maior a proximidade entre o prazo de pagamento e a necessidade do cliente, maior a facilidade para o cliente efetuar o pagamento das parcelas e a quitação total em dia (LEDGERWOOD, 1999).

O prazo dos empréstimos deve ser ajustado de forma a minimizar as necessidades de pré-pagamento. Quando um empréstimo é liquidado antecipadamente, a receita obtida com a operação de empréstimo é menor que a projetada inicialmente para aquela operação, resultando num decréscimo de receita para a IMF. Adicionalmente, uma liquidação antecipada pode indicar que o cliente esteja recebendo empréstimos de outro credor, que deve estar lhe oferecendo taxas de juros menores, melhores serviços, valores superiores ou melhores prazos. Cabe à IMF examinar o produto e as condições de crédito que está oferecendo aos clientes (LEDGERWOOD, 1999).

Uma das características dos contratos de microcrédito é que o primeiro vencimento ou pagamento deve ser imediatamente após o desembolso. Em contratos tradicionais de empréstimos, primeiro o tomador faz o investimento para depois efetuar o pagamento, que pode ser em parcelas ou em um pagamento único. Algumas IMFs, que adotam o estilo do Grameen Bank, programam o primeiro pagamento em uma semana após o desembolso (MORDUCH, 1999).

Algumas IMFs emprestam dinheiro baseadas na boa reputação das pessoas na comunidade em que vivem. Agentes de crédito fazem visitas na comunidade e perguntam sobre o caráter e o comportamento do potencial cliente (LEDGERWOOD, 1999).

Os efeitos da vizinhança são definidos como qualquer característica que pode afetar o retorno ou o resultado econômico de um indivíduo. Características da vizinhança são normalmente definidas como cidade e/ou nível da comunidade que consegue capturar a natureza ou qualidade da comunidade em que os indivíduos residem. Medidas comuns incluem renda média, educação, etnia, chefe de família, etc.

Se os planos de reduzir a pobreza dos movimentos de microfinanças não forem abandonados, algumas escolhas deverão ser feitas. Uma possibilidade é ter outro olhar para as estruturas de gestão e desenho dos processos visando reduzir os custos e manter os resultados. Isso ainda está longe de ser simples, e é difícil pensar neste progresso sem uma segunda onda de inovação. Doadores podem contribuir para encorajar experiências e evolução, e não somente a réplica das melhores práticas de modelos já existentes (MORDUCH, 1999).

Diferentemente do que ocorre no crédito tradicional onde há vasta utilização de modelos de *credit scoring* para operações de crédito de varejo, cartão de crédito, crédito imobiliário e etc. Os modelos de *credit scoring* em microfinanças ainda são poucos.

Os trabalhos mais conhecidos e referenciados em estudos acadêmicos são os de Viganó (1993), que construiu um modelo a partir de um estudo de caso na África e Schreiner (2004), com um estudo de microcrédito na Bolívia.

Os trabalhos seguem praticamente a mesma linha utilizando dados cadastrais dos tomadores de crédito, informações das próprias operações de crédito e outros dados financeiros e de perfil sócio-econômico dos tomadores.

Abaixo segue resumo elaborado com os artigos estudados sobre *credit scoring* em microcrédito ou operações de crédito de países em desenvolvimento.

<i>Autor</i>	<i>Local</i>	<i>Técnica</i>	<i>Dados Utilizados</i>	<i>Positivamente Correlacionado</i>	<i>Negativamente Correlacionado</i>	<i>Não Correlacionado</i>
Laura Viganó - 1993	CNCA Burkina Faso	Análise Discriminante	Dados quantitativos de 100 operações acrescidas de dados qualitativos obtidos em pesquisa. Foram colhidos: - dados pessoais, como: idade, sexo, etc. - dados da empresa: atividade, habilidades, estágio da empresa e etc. - lucratividade - ativos - Situação Financeira - Planos de Investimento - Relacionamento com o CNCA - Dados da operação: prazo, valor, etc. No modelo foram utilizadas 53 variáveis.	Sexo (mulheres são melhores clientes) Ser casado (melhor) Idade (+idade é melhor) Instabilidade do negócio (melhor) Grau de tecnologia e capacidade de se adaptar as mudanças (melhor) Aposentado (melhor) Proximidade com o banco (melhor)	Ativos Telefone Quantidade de terra produtiva	Estabilidade de caixa e lucratividade (resultados controversos).
Araújo, Carmona e Deboça	Cred Cidadania Brasil	Regressão Logística	200 clientes - 100 adimplentes (sem atraso ou atraso inferior a 30 dias) e 100 inadimplentes (atraso superior a 60 dias em pelo menos uma parcela). Dados utilizados: estado civil, sexo, atividade econômica, local de residência, idade, qtde dependentes, receita familiar bruta, despesa familiar bruta, receita familiar líquida, despesa bruta do negocio, receita bruta do negócio, resultado líquido do	Clientes com maior renda têm maior probabilidade de inadimplência (clientes declaram aos agentes de crédito renda superior a verdadeira). Solteiros são mais propensos a inadimplência. Empréstimos com maior prazo têm maior probabilidade de se tornarem inadimplentes. Indivíduos com maior número de dependentes têm maior probabilidade de se tornarem inadimplentes.	Maiores valores tendem a ser mais adimplentes. Está ligado a metodologia do Cred Cidadania que concede valores maiores depois que o cliente tem histórico com o banco - empréstimos com valores crescente. Empreendimentos com maior horizonte de tempo são mais adimplentes. A atuação de alguns agentes de crédito melhora a adimplência. Histórico positivo com a instituição melhora a	

			<p>negócio, resultado líquido total, tempo de funcionamento do negócio, receita bruta do avalista, despesa bruta do avalista, resultado líquido do avalista, valor do último empréstimo, parcela do último empréstimo, qtde de parcelas do último empréstimo, percentual de endividamento, qtde de créditos anteriores com a instituição, histórico do titular e do avalista com a instituição e o agente de crédito. Operações de outubro a dezembro de 2004. Probabilidade inferior a 0,50 são adimplentes e com probabilidade superior são inadimplentes.</p>		adimplência.	
<p>Thi Huyen Thanh Dinh e Stefanie Kleimeier - 2007</p>	Vietnã	<p>Regressão Logística</p>	<p>56.000 operações (de 1992 a 2005) de pequenas empresas, crédito pessoal, crédito imobiliário e cartão de crédito, de um país em desenvolvimento.</p>	<p>Residência - há regiões com maior número de bons pagadores. Empréstimos anteriores reduzem o risco de default. Pessoas sem telefone têm maior risco de default, que pode estar relacionada com a dificuldade de contato com o cliente. Pessoas entre 18 e 24 são melhores pagadores. Casados têm maior risco de default, devido ao</p>	<p>Residência - há regiões com maior número de maus pagadores.</p>	<p>Educação. Tempo de relacionamento com o banco. Prazo do empréstimo - no Vietnã o cliente escolhe o prazo e não é determinado pelo banco. Conta corrente com o banco (apenas 34% da amostra era correntista do banco).</p>

				<p>maior número de dependentes. Status da residência - só reduz risco, se a residência é usada como garantia, pois há o risco de perda do imóvel. Tempo de residência - mais tempo na mesma residência, menor risco de default.</p>	
<p>Mark Schreiner - 2004</p>	Bolívia	<p>Regressão Logística</p>	<p>Amostra separada entre clientes com atraso superior e inferior a 15 dias. Dados utilizados: data do desembolso, valor desembolsado, tipo de garantia, agência, gerente de crédito, sexo, ramo de atividade, atraso, Base de 39.956 operações de 1986 a 1996. 5% da amostra tem atraso superior a 15 dias.</p>	<p>Risco de atraso aumenta a medida que cliente toma novos empréstimos. Risco aumenta para operações contratadas antes da época do Natal. Risco de atraso aumenta quando o officer se torna mais experiente, provavelmente porque há um aumento na sua carteira de clientes. Em microcrédito relacionamento é importante e por isso o risco é afetado pelo perfil do officer. Há relação de maior risco em algumas agências, o que pode apontar para a necessidade de treinamento ou perfil das pessoas. Comércio apresenta menor risco que produção. Atrasos anteriores reduzem o risco de atrasos futuros (pode estar relacionado com uma situação de choque específica da amostra).</p>	<p>Número de atrasos traz pouca informação sobre períodos de atrasos maiores. Os atrasos menores podem estar ligados a falta de ajuste entre o vencimento da parcela e o caixa do tomador. Diferenças de risco entre os gêneros é praticamente nula na amostra. Gênero associado à outras variáveis como garantias e ramo de atividade podem ser significantes, mas sozinha não traz resposta para um maior ou menor risco. Risco não aumenta com o aumento do valor do desembolso. Os aumentos dos desembolsos são pequenos para o público estudado. Na amostra a ausência de garantia não traz mais risco.</p>

				Os dois resultados anteriores são diferentes do que comumente aparece em estudos similares.		
Kristina Meier e Dan Balke - 2006	MFBA Azerbaijão	Regressão Logística	<p>Base de dados acumulada desde o início das operações do MFBA em dezembro de 2002.</p> <p>Dados de entrevista: Tempo do negócio, receita de vendas mensal, patrimônio líquido, histórico de crédito, fluxo de caixa líquido.</p> <p>Dados MFBA: estado civil, sexo, idade, número de membros da família, valor do crédito, ramos de atividade, vendas mensais, balanço total, receita familiar, despesa familiar, margem bruta, margem líquida, rotação, dívidas de curto prazo, liquidez, número de pontos de venda, número de empregados, crédito anterior com MFBA, prazo crédito anterior com MFBA, atraso, anos no negócio, ativos correntes, caixa, recebíveis, estoque, ativos fixos, débitos de curto e longo prazo, outras receitas, capacidade de repagamento, garantia. Foram consideradas 10.053 observações (nem</p>	<p>Quanto menor a renda maior a probabilidade de atraso.</p> <p>Foram analisados 7 ramos de atividade: comércio, serviços, produção, transportes, agricultura, construção e outros. Destes, produção é o mais arriscado.</p> <p>Quantidade de empregados tem relação positiva com a probabilidade de atraso.</p> <p>Probabilidade de atraso aumenta com as despesas familiares.</p> <p>Ausência de outras rendas aumenta a probabilidade de atraso.</p> <p>Risco de atraso aumenta com o índice total da dívidas/patrimônio.</p>	<p>Clientes mais velhos têm menor probabilidade de atraso.</p> <p>Tempo de negócio reduz o risco de atraso.</p>	<p>Valor do crédito sem efeito na probabilidade de atraso.</p> <p>Variáveis sem resultado significativo: dívidas, histórico de crédito, ativos correntes/total de ativos, total de dívidas/total de ativos e patrimônio/ativos correntes.</p> <p>Vendas mensais sem efeito na probabilidade de atraso.</p> <p>Capacidade de repagamento sem influência na probabilidade de atraso.</p> <p>Patrimônio não tem relação com atraso.</p>

	todas com dados completos). Variável dependente - atraso maior que 2 dias.	
--	---	--

Tabela 3: Resumo artigos acadêmicos sobre modelos de *credit scoring* de operações de microcrédito ou operações de crédito em países em desenvolvimento. Elaborado pela autora.

2.5 MODELOS DE ANÁLISE DE CONCESSÃO DE CRÉDITO

Segundo o Dicionário Aurélio de Língua Portuguesa, crédito significa:

Confiança ou segurança na verdade de alguma coisa; crença. / O que é devido a alguém. / Em contabilidade, o que representa saldo, por oposição a débito. / Reputação de solvência, boa fama: ter crédito na praça. / Fig. Influência, valimento, consideração: ter crédito com os outros. / Fig. Autoridade, valia, importância: homem de crédito. // Comprar a crédito, comprar fiado, receber o objeto comprado sem o pagar logo. // Crédito comercial, industrial, agrícola etc., facilidade de ter adiantamentos de dinheiro para a prosperidade do comércio, da indústria, da agricultura etc.

Em poucas palavras, crédito é comprar agora e pagar depois. A palavra crédito vem do latim *credo* e que significa: crer ou acreditar em.

Se alguém tem crédito, então é porque alguém tem a crença de que ele ou ela honrará sua obrigação.

Muitas pessoas atualmente vêem o crédito como um direito, mas na verdade nele está implícita uma obrigação. Os tomadores pagam um preço para que: (i) os credores acreditem nele; (ii) paguem de acordo com as condições estabelecidas; (iii) paguem um prêmio de risco pela possibilidade de não pagamento (ANDERSON, 2007).

O mercado de crédito é um mercado imperfeito. Segundo Stiglitz e Weiss (1981), o mercado de crédito pode ter um ponto de excesso de demanda, o que torna as taxas de juros mais elevadas, o que pode atrair tomadores de maior risco de inadimplência. Os tomadores mais arriscados são menos sensíveis a taxa de juros elevadas, além da existência de assimetria de informações entre o tomador e o credor.

Na relação entre tomador e credor é comum haver dois tipos de problemas: seleção adversa e risco moral. A seleção adversa consiste na possibilidade do credor emprestar os recursos para os tomadores de maior risco. O risco moral consiste na possibilidade do tomador, após a concessão do empréstimo, mudar seu comportamento, reduzindo a probabilidade de pagamento (LEDGERWOOD, 1999).

Credit Scoring é simplesmente o uso de modelos estatísticos para transformar dados relevantes em medidas numéricas que guiam as decisões de crédito. O *credit scoring* foi usado primeiramente nos anos 60 para determinar se as pessoas que estavam adquirindo crédito seriam capazes de pagar suas dívidas e honrar suas obrigações (ANDERSON, 2007). Em financiamento imobiliário seu uso foi reduzido devido à necessidade de avaliação individual do imóvel e do tomador de crédito, porém cresceu no mercado de cartões principalmente devido à agilidade, a redução de perdas e crescimento do mercado (ANDERSON, 2007).

Naquela época, *credit scoring*, estava associado somente a aceitar ou rejeitar decisões de novos negócios. Todavia, no século XXI, passou a ser mais usado para descrever qualquer modelo estatístico capaz de conceder e gerenciar crédito. Isto inclui medidas de risco, retorno, reação e retenção (os quatro Rs) quer seja para marketing, processos de novos negócios, gestão de contas, cobrança, recuperação e também para a gestão do ciclo de risco de crédito (ANDERSON, 2007).

No crédito tradicional é efetuada uma avaliação de um tomador potencial de acordo com os 5 Cs do crédito: caráter, capacidade, capital, colateral e condições (SILVA, 2006). Estas avaliações são baseadas na própria experiência do analista, do que ele aprendeu com sua experiência passada, com pessoas que lhes transmitiram conhecimento e considerando além das informações históricas, uma predição de comportamento dos potenciais tomadores. A chave deste processo é a obtenção de informações através do relacionamento com o cliente.

O uso do *credit scoring* tem trazido uma mudança no relacionamento entre o tomador de uma operação de empréstimo e o credor. O ponto é que tomadores e credores têm conhecimentos particulares de si mesmo, mas que se tornam ineficientes numa era onde existe alta mobilidade dos tomadores e uma quantidade cada vez maior de instituições financeiras

dispostas a conceder crédito. Os Cs ainda são usados, mas atualmente o *credit scoring* pode capturar melhor as informações e extrair delas o máximo de valor, trazendo economia de escala, vantagem em custos, aprovação massificada para grande quantidade de clientes e agilidade, especialmente nos créditos de varejo (ANDERSON, 2007). A análise individual não foi eliminada e ainda existe principalmente, nos casos de empréstimos de grandes valores (bancos e clientes do atacado) e com garantias diferenciadas.

Credit scoring traz valor para o negócio quando é usado para guiar as decisões que afetam o cliente. São modelados cenários e políticas que guiam as ações, como por exemplo, aceitar ou rejeitar uma operação, a definição do valor máximo do empréstimo, o prazo de pagamento, a taxa de juros e etc. O custo-benefício da automação do processo tem pressionado as organizações a adotarem cada vez mais os modelos de *credit scoring* e usarem analistas apenas para operações de exceção ou para créditos de grandes volumes.

Um modelo de *credit scoring* é uma fórmula que coloca pesos em diferentes características dos tomadores, dos credores e da operação de crédito. A fórmula produz uma estimativa da probabilidade de um evento ocorrer. Os pesos da fórmula são calculados com ferramentas estatísticas. A matemática é a parte simples, o difícil é coletar os dados das características e *performance* das operações de empréstimos concedidas no passado (SCHREINER, 2000).

Credit scoring é altamente dependente de informações, portanto a credibilidade do modelo varia de acordo com a qualidade das informações. Os modelos podem ser desenvolvidos a partir de dados proprietários ou recorrendo à utilização de dados de *bureaux* de crédito ou compartilhamento de informações com outras organizações.

As ferramentas de *credit scoring* são baseadas em estatística e técnicas de pesquisa operacional e são algumas das mais rentáveis e bem sucedidas aplicações da teoria estatística nos últimos 20 anos (ANDERSON, 2007).

Cada técnica de modelagem tem suas próprias forças e fraquezas que variam de acordo com as circunstâncias. Serão descritas 6 técnicas principais: (i) regressão linear; (ii) análise discriminante; (iii) regressão logística; (iv) árvores de decisão; (v) redes neurais; (vi)

programação linear. Atualmente a regressão logística é usada entre 80% e 90% dos desenvolvimentos de modelos de escore. Os 10%-20% restantes ainda utilizam as técnicas lineares devido a sua flexibilidade e facilidade de uso (ANDERSON, 2007).

Instituições de microfinanças que pretendem usar modelos de *credit scoring* no futuro devem iniciar a coleta de dados já. Sem uma base histórica das características e *performance* das operações de microcrédito, além da necessidade de registrar as informações coletadas pelos agentes de crédito, será impossível desenvolver e implementar uma ferramenta de *credit scoring* (SCHREINER, 2000).

2.5.1. Regressão Linear

Regressão é um método estatístico, que relaciona duas ou mais variáveis quantitativas, buscando uma possível relação de dependência entre as variáveis. Uma regressão pressupõe que as variáveis estudadas tenham um relação linear de dependência entre si, e que pode ser representada pela seguinte equação linear:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in} + \varepsilon_i$$

Onde $\beta_0, \beta_1, \dots, \beta_n$ são os parâmetros da relação linear entre a variável dependente y_i e a variáveis independentes x_i , que são as variáveis e observações que tentam explicar as variações de y_i .

Várias premissas (HAIR JR., 2006) são utilizadas em uma regressão linear: (i) linearidade; (ii) homocedasticidade; (iii) distribuição normal do erro (ε); (iv) erro independente; (v) aditividade; (vi) não correlação dos estimadores; (vii) uso de variáveis relevantes.

Em *credit scoring* ou outra aplicação onde o resultado pode ser binário, a regressão linear é transformada em modelagem de probabilidade linear, onde o resultado final é uma estimativa do $p(\text{Good})$, a fórmula é representada por:

$$p(\text{Good})_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \varepsilon_i$$

A probabilidade de cada registro i , é a soma de uma constante e o produto de uma série de pesos β_j e variáveis x_{ij} , onde cada variável assume um valor diferente para cada registro e pesos diferentes para cada ocorrência da variável j (HAIR JR., 2006).

2.5.2. Análise Discriminante

A análise discriminante é uma técnica estatística que permite que pesquisadores estudem as diferenças entre dois ou mais grupos de objetos respeitando várias características/variáveis simultaneamente. Em ciências sociais, há um grande número de situações que esta técnica pode ser usada. A análise discriminante ajuda a analisar as diferenças entre os grupos e/ou prover os estudiosos com significados que classificam qualquer caso dentro dos grupos que mais possuem similaridades. Os grupos devem ser mutuamente excludentes. As características/variáveis usadas para distinguir os grupos são chamadas de variáveis discriminantes. Estas variáveis devem ser mensuradas com um intervalo ou índice de dimensionamento, o que significa que as variâncias podem ser calculadas e podem, legitimamente, ser aplicadas em equações matemáticas (KLECKA, 1980).

A análise discriminante linear multivariada permite identificar, através de uma função discriminante, variáveis que possibilitem segregar ou distinguir grupos de observações. No caso de *credit scoring*, a análise discriminante gera uma função ou mais funções que auxiliam na classificação do bom e do mau pagador. É possível também classificar os tomadores por faixas de rating.

Do ponto de vista matemático, para o caso de análise de dois grupos, por exemplo, adimplentes e inadimplentes, pode-se obter uma função discriminante expressa por:

$$Y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4, \text{ onde:}$$

- Y é a variável dependente, representa a pontuação ou escore obtido por uma determinada observação;
- $a_1, a_2, a_3, a_4, \dots$ são os coeficientes que indicam a influência de cada variável independente na classificação de uma observação nos grupos considerados;
- $x_1, x_2, x_3, x_4, \dots$ são as variáveis independentes, representativas dos valores referentes às variáveis de uma dada observação.

Assim, a função discriminante implica no cálculo de um escore, advindo dos coeficientes e das variáveis independentes, que indicam maior proximidade a um determinado grupo. Na análise de crédito, são utilizadas variáveis que se referem aos dados de cadastro ou a outras informações ou características do potencial tomador e, a partir destes, calculado o escore Y . Indivíduos com maior escore semelhante são classificados na mesma faixa indicativa da qualidade de crédito ou no mesmo *rating*.

Abaixo estão relacionadas premissas para o uso da análise discriminante, onde:

g = quantidade de grupos

p = quantidade de variáveis discriminantes

n_i = quantidade de observações no grupo i

n = total de quantidades de casos que cobre todos os grupos

Premissas (KLECKA, 1980):

- Amostra deve ser separada entre dois ou mais grupos: $g \gg 2$;
- Deve haver no mínimo duas observações por grupo: $n_i \gg 2$;
- Pode haver qualquer quantidade de variáveis discriminantes, desde que seja menor que o total de observações menos 2: $0 < p < (n - 2)$;
- A variável não-discriminante deve ser uma combinação de outras variáveis discriminantes;
- A matriz de covariância de cada grupo deve ser (aproximadamente) igual, a menos que fórmulas especiais sejam usadas;
- A população de cada grupo deve ter uma distribuição normal.

A análise discriminante foi escolhida como técnica a ser utilizada no estudo de Altman de 1968, apesar de não ser tão popular quanto à regressão logística, a análise discriminante tem sido utilizada desde os anos 30 (ALTMAN, 1968). A análise discriminante foi usada naquela época principalmente em biologia e ciências comportamentais. Depois passou a ser utilizada também em pesquisas financeiras e avaliação de crédito e investimento (ALTMAN, 1968).

2.5.3. Regressão Logística

Muitos fenômenos sociais são discretos ou qualitativos ao invés de terem uma natureza quantitativa ou contínua. Este tipo de evento pode ocorrer ou não ocorre – uma pessoa faz uma escolha ao invés da outra, uma pessoa passa de um estado para outro e etc (PAMPEL, 2000).

Um fenômeno binário discreto geralmente toma a forma de um indicador dicotômico ou uma variável dummy. É possível representar os dois valores com qualquer número, por facilidade de representação empregam-se os valores 0 ou 1. O coeficiente de regressão da variável dependente pode prever a probabilidade de um evento ocorrer devido às mudanças das variáveis dependentes (PAMPEL, 2000).

Na regressão logística, diferentemente da análise discriminante, analisam-se apenas dois grupos, assim caso haja a necessidade de estudos de mais de dois grupos, a análise discriminante é mais indicada (PAMPEL, 2000).

Na regressão logística, a variável dependente Y é definida como uma variável binária, que assume os valores 0 ou 1 e as variáveis independentes X_1, X_2, \dots, X_n , estão associadas às características e ou eventos associados ao grupo. A função de distribuição logística é dada por:

$$P_i = [Y = 1 | X = X_j] = \frac{1}{1 + e^{-z}}, \text{ onde:}$$

P_i , representa a probabilidade de um evento ocorrer, sendo:

$$Z = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n$$

Por exemplo, sendo $Y = 1$ o caso em que o indivíduo é mau pagador, isto é, inadimplente, então o modelo logístico indica a probabilidade de um indivíduo ser inadimplente condicionada às variáveis dependentes X .

Quando as premissas da análise discriminante e da regressão logística são obedecidas, os dois métodos apresentam resultados e classificação comparáveis, contudo quando a premissa da normalidade das variáveis não é observada, a regressão logística tem maiores vantagens (HAIR JR., 2006).

De forma geral, a regressão logística apresenta uma menor quantidade de premissas restritivas para a utilização da metodologia, o que facilita a sua utilização.

2.5.4. Outros Modelos

Árvore de decisão é uma ferramenta gráfica com estrutura de linhas e quadros que mostram as possibilidades de ligação entre eventos que podem ou não ser controlados. As árvores de decisão também são usadas para a visualização de dados e previsões de problemas. Pode haver dois ou mais ramos para cada informação e vários níveis, dependendo da complexidade da árvore. O ramo final pode ser usado para estimar a pontuação final. Para um resultado binário, o valor é uma probabilidade e todos os ramos, além da probabilidade, terão os pontos de corte acima da média $p(\text{Bad})$ (ANDERSON, 2007).

Redes neurais podem ser descritas como redes computacionais que respondem aos dados de entrada e que se adaptam ao ambiente. São capazes de simular o processo do cérebro humano, principalmente a respeito da sua própria organização e aprendizado. O resultado final é similar a uma árvore de decisão, exceto no detalhamento, que é muito mais refinado e complexo. Tem como vantagem a capacidade de: (i) processar uma enorme quantidade de dados; (ii) descobrir e rastrear os relacionamentos entre os dados, especialmente suas interações; (iii) lidar com relacionamentos não-lineares entre os dados; (iv) aprender, baseado nas diferenças entre os dados observados e os dados atuais (ANDERSON, 2007). Uma rede neural é um processador paralelamente distribuído constituído de unidade de processamento

simples, que tem a propensão de armazenar conhecimento experimental e torná-lo disponível para uso. Ela se assemelha ao cérebro por dois principais motivos: (i) o conhecimento é adquirido pela rede a partir do seu ambiente através de um processo de aprendizagem; (ii) forças de conexões entre neurônios, conhecidas como sinapses, são utilizadas para armazenar o conhecimento (HAYKIN, 1999).

O modelo neural é determinístico. Para algumas aplicações de redes neurais é necessário que a análise seja baseada num modelo estocástico, mais precisamente, necessita-se que o neurônio assuma dois estados $+1$ ou -1 . A decisão de disparar o neurônio (ligado ou desligado) é probabilística. Considere que x represente o estado do neurônio e $P(v)$ a probabilidade de disparar, v é o campo local induzido do neurônio, podemos então escrever que:

$$\begin{cases} x = +1 \text{ com probabilidade } P(v) \\ x = -1 \text{ com probabilidade } 1 - P(v) \end{cases}$$

2.6 TÉCNICAS DE APRENDIZAGEM DE MÁQUINA

Segundo Witten e Frank, 2005, alguns dicionários definem “aprender” como:

- Adquirir conhecimento pelo estudo, experiência ou ser ensinado;
- Tornar-se consciente;
- Memorizar;
- Ser informado a cerca de alguma coisa;
- Receber instruções.

Estas definições apresentam algumas deficiências quando falamos em máquinas. No caso das duas primeiras definições, é virtualmente impossível testar se o conhecimento foi adquirido ou não. Como se sabe se uma máquina conseguiu obter conhecimento sobre alguma coisa? Provavelmente não se podem fazer apenas perguntas para ela, mas sim testar sua habilidade em responder a questões. Como sabemos se uma máquina se tornou consciente de algo?

As três últimas definições, embora colocadas em termos humanos, memorizar e receber instrução parecem estar bem próximas do que se pode entender como aprendizagem de máquina. Máquinas são passivas, e para os computadores estas tarefas são triviais. Ao invés disso, aprendizagem de máquina se refere a melhorar a *performance* ou pelo menos descobrir a potencial resposta em novas situações. Memorizar ou ser informado de algo não significa que há a capacidade de aplicação deste novo conhecimento para novas situações.

Uma definição de aprendizagem de máquina poderia ser que máquinas aprendem quando elas mudam seu comportamento, de tal maneira, que elas executam a tarefa melhor no futuro do que efetuam hoje. Isso relaciona aprendizagem a *performance* e não apenas a conhecimento. Pode-se testar aprendizagem observando o comportamento atual e comparando-o com o comportamento passado. Este é um tipo de definição mais objetiva e bem mais satisfatória (WITTEN E FRANK, 2005).

A aprendizagem de máquina não é mera aplicação de procedimentos estatísticos. Há várias técnicas e modelos onde se pode aplicar a aprendizagem de máquina (WITTEN E FRANK, 2005).

O procedimento utilizado no processo de aprendizagem é chamado de algoritmo de aprendizagem, cuja função é melhorar a *performance* para alcançar os objetivos desejados (HAYKIN, 1999).

No modelo simples de aprendizagem de máquina, o ambiente fornece alguma informação para um elemento de aprendizagem, que utiliza esta informação para aprimorar a base de conhecimento, e finalmente o elemento de desempenho utiliza a base de conhecimento para executar a tarefa. A máquina opera inicialmente por suposição e depois recebe realimentação, permitindo que a máquina avalie hipóteses e as revise, se necessário (HAYKIN, 1999).

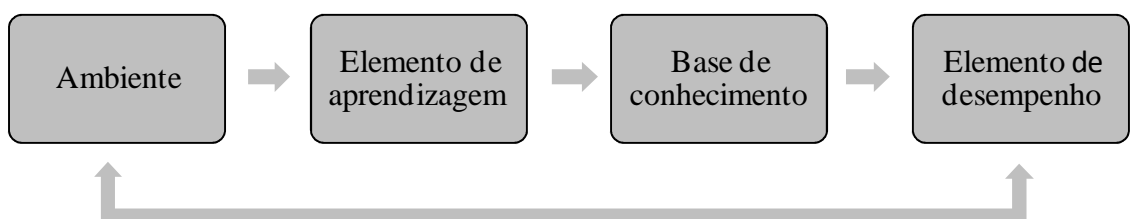


Figura 9: Modelo simples de aprendizagem de máquina - Haykin, 1999 - pág.61

2.6.1. *Ensemble Methods*

De acordo com Opitz e Maclin (1999), um *ensemble* consiste em um conjunto de funções treinadas individualmente cujas previsões são combinadas na classificação de novas observações. Ou seja, a idéia básica da metodologia *ensemble* é a construção de um mecanismo preditivo global, através da integração de múltiplos modelos, que possibilite estimativas mais acuradas e confiáveis (ROKACH, 2009).

Bühlmann e Yu (2003) citam o trabalho de Tukey (1977) contendo um modelo de regressão linear aplicado inicialmente aos dados originais e posteriormente aos resíduos como origem dos métodos *ensemble*. Assim, uma aplicação sucessiva de uma mesma técnica aos dados e aos resíduos constitui um exemplo de método *ensemble*. Considerando o desenvolvimento da teoria de estatística e o aumento do poder computacional das máquinas, mecanismos de combinação de modelos puderam ser implementados em aplicações mais complexas.

Diversos autores como, por exemplo, Breiman (1996), Bauer e Kohavi (1999), Maclin e Opitz (1997) identificam melhorias substanciais em classificação através de métodos *ensemble*. Considerando seus ganhos de desempenho em classificação, os métodos *ensemble* ou métodos de aprendizagem *ensemble* constituem uma das linhas de pesquisa mais ativas na área de aprendizagem supervisionada (MOKEDDEM e BELBACHIR, 2009).

Hsieh e Hung (2010) citam que a metodologia *ensemble* tem sido utilizada em diversas áreas do conhecimento. Por exemplo, Tan, Gilbert e Deville (2003) aplicam métodos *ensemble* em bioinformática, em problemas de classificação de proteínas em diversas classes. Em geografia e sociologia, Bruzzone, Cossu e Vernazza (2004) detectam a cobertura terrestre através combinação de função de classificação de imagens. Maimon e Rokach (2004) aplicam técnicas de *ensemble* atreladas a árvores de decisão para mineração de dados de manufatura.

Estudos usando os métodos *ensemble* em finanças também começam a se difundir. Por exemplo, Leigh, Purvis e Ragusa (2002) buscam realizar previsões sobre o valor do índice da bolsa de Nova Iorque através de análise técnica reconhecimento de padrões, redes neurais e algoritmos genéticos. Lai, He e Yen (2007) estudam o *value-at-risk* de posições em óleo cru

juntando em métodos *ensemble* com análise de ondaletas e redes neurais artificiais. Especificamente em aplicações de métodos *ensemble* para a área de crédito, podem ser citados os trabalhos de Lai et al. (2006) que usam redes neurais baseadas em confiabilidade, Alfaro et al. (2008) que agregam redes neurais na análise de falência, e de Hsieh e Hung (2010) que avaliam escores de crédito combinando redes neurais, redes bayesianas e *support vector machines*.

Conforme Dietterich (2000), dentre os diversos métodos para construir ensembles duas das técnicas mais populares são o *bagging* ou *bootstrap aggregating* desenvolvido por Breiman (1996) e o *boosting*, proposto inicialmente por Freund e Schapire (1996), cujos algoritmos mais conhecidos baseiam-se na família de algoritmos Adaboost. O *boosting* é comumente denominado de *Arcing* (*Adaptative resampling and combining*), em função de um trabalho de Breiman (1998) que trouxe novas formas de se compreender e utilizar o algoritmo.

Dentro do contexto de ensemble methods, o *bagging* e o *boosting* são dois mecanismos gerais para aprimorar o desempenho de um determinado algoritmo de aprendizagem, chamado de algoritmo base (FREUND e SCHAPIRE, 1998). Esses métodos permitem a redução de variâncias de erros de estimação (TUMER e GHOSH, 1999) sem necessariamente aumentar o viés (ROKACH, 2005), proporcionando ganhos tanto do ponto de vista de teoria estatística quanto sob uma perspectiva de aplicabilidade prática. Bartlett e Shawe-Taylor (1998) citam, inclusive, que em determinadas situações, os métodos podem até mesmo reduzir vieses.

De acordo com Freund e Schapire (1998), *bagging* e *boosting* são algoritmos bastante similares no sentido de que envolvem a incorporação de versões modificadas, do conjunto de dados de treinamento no algoritmo de aprendizagem base, combinando as regras resultantes a partir de uma maioria de votos, por exemplo. Assim, ambos os métodos estão fundamentados em técnicas de re-amostragem para obter diferentes conjuntos de dados para treinamento de cada um dos classificadores do modelo (OPTIZ e MACLIN, 1999).

A despeito de similaridades, existem também diferenças relevantes entre os algoritmos. A principal diferença baseia-se no fato de que no *bagging*, as amostras são

escolhidas de maneira aleatória, independente e com repetição dos registros, enquanto que no *boosting*, as amostras são escolhidas de forma serial e determinística, através da qual uma dada amostra depende fortemente de todas as outras previamente geradas (FREUND e SCHAPIRE, 1998).

A seguir, são apresentados os fundamentos dos métodos de *bagging* e *boosting* que serão estudados neste trabalho para classificação de crédito. É importante destacar que aplicações similares de métodos ensemble já foram estudados como, por exemplo, em Paleologo, Elisseeff e Antonini (2010) que estudam o *bagging* para escore de crédito e em Xie et al. (2009) que analisam o *boosting* aplicado junto a regressão logística.

2.6.2. *Bagging*

Bagging (*Bootstrap Aggregating*) quando os estudos possuem dados altamente relacionados e uma amostra pequena se comparada com a dimensão da amostra universal, criar uma boa e simples regra de classificação dos dados é tarefa difícil. Geralmente o classificador é enviesado e com grande variância, o que gera uma baixa *performance* do modelo. Com o objetivo de aprimorar o modelo é possível utilizar classificadores fracos e combiná-los entre si a fim de tornar a regra desenvolvida mais poderosa. Algumas técnicas têm sido desenvolvidas e as mais populares são *bagging* e *boosting* (SKURICHINA e DUIN, 2002).

Bagging é um *ensemble method* para melhorar a estimativa de um conjunto instável ou sistema de classificação. O *bagging* foi definido como uma técnica de redução de variância para um processo de análise de dados, como árvores de decisão ou regressão linear. Este método tem atraído atenção, devido principalmente, a sua simplicidade de execução e pela popularidade da metodologia *bootstrap*.

Considere uma dada regressão ou classificação. Temos pares (X_i, Y_i) ($i = 1, 2, 3, \dots, n$), onde $X_i \in \mathbb{R}^d$ e d é a variável dimensional preditora e a resposta $Y_i \in \mathbb{R}$ (regressão) ou $Y_i \in \{0, 1, \dots, J - 1\}$ (classificação com J classes). A função alvo de interesse é usualmente $E[Y | X = x]$ para regressão ou função multivariada $P[Y = j | X = x]$ ($j = 0, 1, \dots, J - 1$)

para classificação. O estimador da função, que é o resultado de um processo para a base dada é:

$\hat{g}(\cdot) = h_n((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$, onde a função defini o estimador da função dos dados.

O algoritmo do *bagging* é definido a seguir:

1ª. Etapa: construir uma amostra *bootstrap* aleatória: $(X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*)$ com substituição dos dados por n vezes: $(X_1, Y_1), \dots, (X_n, Y_n)$

2ª. Etapa: Calcular o estimador *bootstrap* $\hat{g}^*(\cdot)$ pelo princípio do plug-in:

$$\hat{g}^* = h_n((X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*))(\cdot).$$

3ª. Etapa: Repetir a 1ª e a 2ª etapa por M vezes, frequentemente escolhido entre 50 ou 100 vezes, com $\hat{g}^{*k}(\cdot)$ ($k = 1, \dots, M$).

O estimador *bagged* é: $\hat{g}_{Bag}(\cdot) = M^{-1} \sum_{k=1}^M \hat{g}^{*k}(\cdot)$.

Em teoria o estimador *bagged* é: $\hat{g}_{Bag}(\cdot) = E^* [\hat{g}^*(\cdot)]$.

2.6.3. *Boosting*

Boosting, proposto por Freund e Schapire (1996), é uma técnica que combina classificadores fracos, que apresentam baixa *performance*, gerando uma combinação que tem uma regra de classificação com uma melhor *performance*. No *boosting*, classificadores e amostra de treinamento são obtidos sequencialmente, de forma estritamente determinística. A cada passo, os dados de treinamento são re-balanceados de forma que objetos classificados com pesos errados sejam modificados (SKURICHINA e DUIN, 2002).

Boosting é um método para melhorar a *performance* de qualquer algoritmo de aprendizagem. Em teoria, *boosting* pode ser usado para reduzir erros significantes em um

algoritmo de aprendizagem fraco que gera classificadores que precisam ser um pouco melhores que os encontrados de forma randômica. Apesar dos benefícios potenciais encontrados nos resultados teóricos, na prática o verdadeiro valor do *boosting* pode somente ser avaliado no teste de um problema real na aprendizagem de máquina (FREUND e SCHAPIRE, 1996).

O AdaBoost, algoritmo proposto em 1995 por Freund e Schapire, resolveu muitas das dificuldades de ordem prática dos algoritmos *boosting* anteriores. O algoritmo tem como entrada um conjunto de treinamento $(x_1, y_1), \dots, (x_m, y_m)$ onde para cada x_i pertence a algum domínio ou exemplo do espaço X , e para cada y_i do espaço Y .

A seguir, é realizada uma descrição do algoritmo AdaBoost, baseada em Freund e Schapire, 1995. $Y = \{-1, +1\}$ é uma premissa do modelo de Freund e Schapire. A AdaBoost chama um conjunto fraco de aprendizagem por repetidas vezes em um série de rodadas $t = 1, 2, \dots, T$. Uma das principais idéias deste algoritmo é manter uma distribuição ou um conjunto de pesos sobre o conjunto de treinamentos. O peso da distribuição da amostra de treinamento de i até t significa $D_t(i)$. Inicialmente todos os pesos são iguais, mas em cada rodada, os pesos incorretamente classificados são acrescidos de aprendizagens fracas de modo que a amostra tenha um aprendizado.

O trabalho é encontrar a hipótese fraca $h_t: X \rightarrow \{-1, +1\}$ apropriada para a distribuição D_t . A melhor das hipóteses fracas é medida por um erro:

$$\varepsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i] = \sum_{i: h_t(x_i) \neq y_i} D_t(i)$$

Observe que o erro é medido com relação à distribuição D_t onde cada aprendizagem fraca foi treinada.

Sejam x_i as observações de uma amostra que está sendo estudada e que tem chance e probabilidade de se repetir de acordo com dados do passado e y_i os resultados escolhidos. As hipóteses fracas são as regras rejeitadas que são fornecidas pelo modelo que está sendo

testado onde os subgrupos que foram examinados pelo modelo foram escolhidos de acordo com uma distribuição D_t .

- Uma vez que a hipótese fraca h_t tenha sido recebida, o AdaBoost escolhe um parâmetro α_t como um algoritmo. Intuitivamente, α_t mede a importância que é transferida para h_t . Note que, $\alpha_t \geq 0$ se $\varepsilon_t \leq \frac{1}{2}$ (que pode ser assumido sem perda de generalidade), e α_t maior que ε_t .

A distribuição D_t é novamente atualizada usando a regra mostrada pelos números. O efeito desta regra é aumentar o peso das amostras desclassificadas por h_t e diminuir os pesos das amostras corretamente classificadas. Então, o peso tende a se concentrar numa amostra forte.

A hipótese final H é o voto da maioria ponderada pelos T das hipóteses fracas, onde α_t é o peso escolhido para h_t .

Dados: $(x_i, y_i), \dots, (x_m, y_m)$, onde $x_i \in X, y_i \in Y = \{-1, +1\}$;

Inicializa-se: $D_1(i) = \frac{1}{m}$

Para $t = 1, \dots, T$:

- Treinar uma aprendizagem fraca usando a distribuição D_t .
- Obter a hipótese fraca $h_t: X \rightarrow \{-1, +1\}$, com erro: $\varepsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i]$
- Escolha $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$

$$\begin{aligned} \text{Atualizando: } D_{t+1}(i) &= \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{se } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{se } h_t(x_i) \neq y_i \end{cases} \\ &= \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t} \end{aligned}$$

Onde Z_t é um fator de normalização (escolhido de forma que D_{t+1} seja uma distribuição).

O resultado final da hipótese: $H(x) = \text{sign}(\sum_{t=1}^T \alpha_t h_t(x))$

3 OBJETIVO

A pesquisa tem por objetivos principais (i) o desenvolvimento teórico de modelos de análise de risco de crédito de operações de microcrédito; (ii) a identificação de variáveis chave para a análise de crédito e probabilidade de default dos empreendedores de baixa renda.

Como objetivos específicos, dentro do levantamento bibliográfico, a pesquisa busca: (i) discutir o impacto em renda do microcrédito produtivo orientado; (ii) apresentar contribuições e estudos já realizados sobre as variáveis chave para a análise de crédito e probabilidade de default dos empreendedores de baixa renda; (iii) discutir o estágio do microcrédito no Brasil; (iv) comparar o microcrédito no Brasil e no mundo; (v) descrever o modelo de avaliação individual de crédito efetuada pelo agente de crédito.

Objetivos específicos ligados à modelagem quantitativa para análise de crédito envolvem: (i) o entendimento das premissas dos modelos de avaliação do risco de crédito de tomadores de recursos; (ii) a aplicação de técnicas estatísticas para análise da concessão de crédito; (iii) o teste de adequação dos modelos de previsão de risco de crédito.

O trabalho busca também investigar se técnicas tradicionais de análise discriminante e regressão logística fornecem resultados substancialmente melhores quando acoplados a técnicas de aprendizagem de máquina, baseada em *bagging* e *boosting*.

4 METODOLOGIA

4.1 DESCRIÇÃO

O procedimento metodológico da pesquisa consiste basicamente da utilização de métodos quantitativos para o desenvolvimento de um modelo de crédito aplicado as operações de microcrédito.

Além do processo quantitativo, a pesquisa apresenta a descrição do processo de concessão e acompanhamento do microcrédito relatado a partir de visita efetuada pela autora nos locais onde as operações da instituição financeira são realizadas.

Os modelos quantitativos a serem utilizados envolvem as técnicas tradicionais de análise discriminante e regressão logística, considerando as variáveis disponíveis. Ambas as técnicas serão aplicadas a uma amostra de uma carteira de microcrédito que contém uma parte das operações de microcrédito concedidas entre 2004 e 2010 por uma instituição financeira.

Para a estimação de parâmetros dos modelos de análise discriminante e regressão logística, será utilizada a ferramenta SPSS – Statistical Package for the Social Science e R-Project.

Uma parte da amostra será utilizada para elaboração e calibração dos modelos e outra parte para validação e teste do modelo. Desta forma, pode-se testar, em uma amostra de validação, a capacidade preditiva do modelo criado com base na amostra de calibração.

Após a elaboração dos modelos serão utilizadas as técnicas de aprendizagem de máquina descritas anteriormente, *boosting* e *bagging*, com o objetivo de verificar potenciais ganhos de classificação.

4.2 BASE DE DADOS

A amostra utilizada é composta de amostras de duas bases de dados: (i) base de clientes que já realizaram uma operação de microcrédito com a instituição financeira; (ii) base de dados de parcelas contendo informações dos contratos de microcrédito concedidos pela instituição financeira no período de 2004 a 2010. As amostras fornecidas são suficientes para representar as operações de microcrédito.

A base de clientes possui as informações cadastrais dos clientes, as informações financeiras (informações sócio-econômicas que foram obtidas por meio de entrevista com o agente de crédito) e as informações básicas da operação contratada. Caso a instituição financeira tenha concedido mais de uma operação de microcrédito para o mesmo cliente, o cliente estará registrado na base tantas vezes quantas forem às operações contratadas.

Abaixo está demonstrada a figura que representa a base de dados que contempla todas as informações disponíveis:

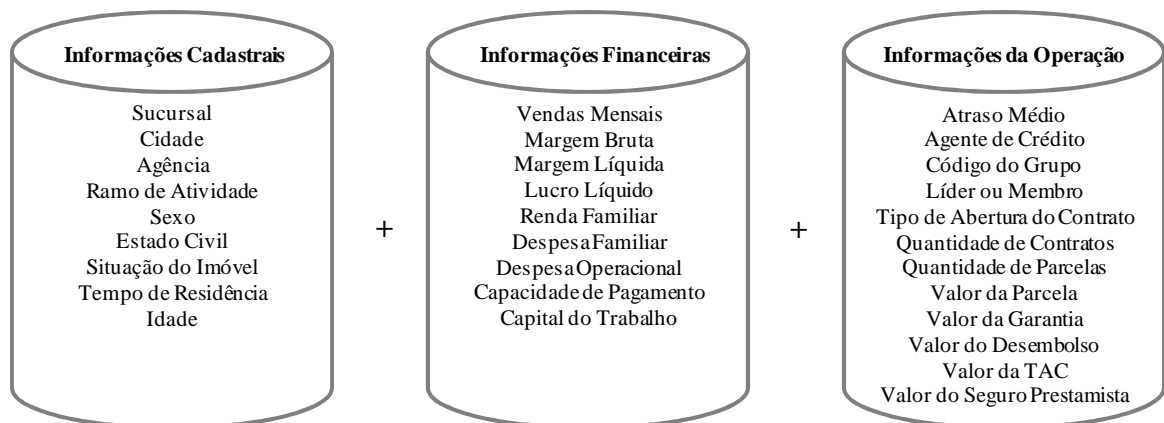


Figura 10: Dados da base de contratos consolidado com a base de clientes. Elaborada pela autora com base no dicionário de dados das bases.

Algumas das variáveis disponíveis na base de clientes e na base de contratos já foram utilizadas em outros estudos acadêmicos de modelos de *credit scoring* para microcrédito ou para operações de crédito em países em desenvolvimento. Dentre estas informações pode-se destacar:

- Dados cadastrais como sexo, idade e região de residência, que são usadas em estudos para países em desenvolvimento, mas são proibidas de serem utilizadas como variáveis classificatórias nos países desenvolvidos (DINH e KLEIMEIER, 2007);
- Atraso de pagamento das parcelas: é comum nos estudos de microcrédito utilizar a variável quantidade de dias de atraso para a classificação de bons e maus pagadores, uma vez que os índices de inadimplências deste segmento são muito baixos (SCHREINER, 2004);
- O ramo de atividade, valor de vendas mensais e outros indicadores econômicos são variáveis utilizadas em estudos de *credit scoring* (MEIER e BALKE, 2006);
- Em alguns estudos, além de dados quantitativos, foram levantadas informações qualitativas em entrevistas. Essa técnica não será utilizada no estudo. Todas as informações dos clientes foram coletadas pelos agentes de crédito em entrevistas para levantamento de dados sócio-econômicos.

4.3 O MICROCRÉDITO NA PRÁTICA

Todo o processo de concessão de microcrédito produtivo orientado se dá com base no relacionamento entre o microempreendedor, candidato a tomador do empréstimo, e o agente de crédito.

Este relacionamento se inicia com a prospecção do cliente e entrevista onde é efetuado o levantamento sócio-econômico do microempreendedor e se mantém durante toda a vida da operação.

O agente de crédito tem uma área geográfica definida de atuação e é responsável basicamente por: (i) prospecção de clientes; (ii) levantamento dos dados sócio-econômicos por meio de entrevista; (iii) reuniões com os grupos solidários para explicar as obrigações do grupo e acompanhá-los até a liquidação da operação; (iv) entrega da proposta para seu gerente/supervisor; (v) participação do comitê de crédito da agência para defesa e aprovação das propostas; (vi) formalização contratual das operações; (vii) envio dos contratos para processamento; (viii) entrega dos carnês aos tomadores; (ix) acompanhamento da operação durante toda a sua vida; (x) repactuações, quando necessário.

Durante a visita que foi efetuada com o agente de crédito, numa comunidade carente da cidade de São Paulo, foi possível acompanhar o processo de levantamento das informações sócio-econômicas e uma reunião com um grupo solidário. A visita foi de extrema importância para o estudo, pois possibilitou a autora vivenciar alguns processos, como por exemplo, o papel do agente e o papel do grupo solidário.

A seguir estão descritos os processos considerados mais importantes, sendo o primeiro o levantamento sócio-econômico e o segundo, a formação e o papel do grupo solidário.

4.3.1 Levantamento Sócio-Econômico

A concessão do crédito se inicia com o levantamento sócio-econômico que é efetuado por meio de perguntas muito simples e que foram adequadas à linguagem e realidade do microempreendedor. Esta adequação é de fundamental importância para o processo, pois permite ao microempreendedor compreender mais facilmente o que está sendo perguntado e que assim sejam fornecidas as informações necessárias para o processo de avaliação e aprovação de crédito.

Esta pequena e simples entrevista é que permite ao agente de crédito elaborar um balanço simples do negócio do cliente. O agente inicia a entrevista explicando que fará algumas perguntas para colher informações que serão utilizadas pela instituição financeira na análise do crédito. No exemplo, não foram levantadas as informações cadastrais do cliente, pois estas haviam sido levantadas numa primeira visita. Era a segunda visita ao microempreendedor que havia sido agendada com o cliente.

A seguir estão descritas as principais perguntas e a sequência do levantamento sócio-econômico efetuado, sendo elas:

1 – Quanto o senhor (a) vende num mês muito bom para o seu estabelecimento? E num mês regular? E num mês ruim, quanto você vende? – a partir destes dados é calculada a média do volume de vendas mensal do cliente.

2 – Qual o produto mais vendido na sua loja? Quantas unidades deste produto você vende por mês? Qual o valor de compra deste produto? E o valor de venda? Quantas unidades

deste produto você tem no estoque? – a partir destes dados é calculado o valor total de venda dos produtos, o valor de custo, o valor em estoque e o líquido das vendas deste produto.

3 – Em seguida, repete as mesmas perguntas e o mesmo processo para os outros dois produtos mais vendidos da loja. E, depois, efetua o mesmo cálculo.

4 – Qual percentual de suas vendas é pago à vista? E a prazo?

5 – Quantos funcionários você tem?

6 – Quanto paga mensalmente pela(s) maquina(s) com cartão de crédito (um dos estabelecimentos visitados tinha maquina(s) de cartão)?

7 – Quanto você paga de aluguel do estabelecimento?

8 – Quanto paga de conta de água e luz do estabelecimento?

9 – Quanto paga de conta de telefone do estabelecimento?

10 – Qual a despesa mensal que tem para ir comprar os produtos vendidos (despesas com gasolina, condução para deslocamento, refeição e etc.)?

Os dados levantados nos itens de 5 a 10 são a base para o cálculo das despesas operacionais do negócio e do lucro líquido do negócio.

11 – O senhor(a) possui despesas financeiras ou dívidas com bancos?

12 – Qual a despesa que tem na sua residência com supermercado (compra de arroz, feijão, carne, verduras, legumes, material de limpeza e etc.)?

13 – Quanto paga de aluguel da residência, caso não seja própria?

14 – Qual a despesa mensal com água e luz na residência?

15 – Quanto paga de conta de gás e de telefone por mês?

16 – Paga algum tipo de plano de saúde ou tem despesa freqüente com remédios?

17 – Paga escola e/ou transporte escolar para os filhos?

18 – O senhor(a) tem crediário em alguma loja (menciona o nome de algumas das lojas de varejo mais conhecidas da cidade de São Paulo)? – vale ressaltar que as pessoas não classificam este tipo de dívida como despesa financeira e outras dívidas, que foram o objeto da pergunta de número 11.

19 – O senhor(a) tem financiamento de automóvel? E dívida com cartão de crédito? – as pessoas também não classificam este tipo de dívida como despesa financeira e dívidas, que foram o objeto da pergunta de número 11.

20 – Qual a renda familiar?

Os dados levantados nos itens de 11 a 20 são usados para calcular as despesas totais familiares. A renda familiar (informada no item 20) é somada ao lucro líquido do negócio e subtraída das despesas totais familiares, obtendo-se a capacidade de pagamento do cliente.

21 – Quanto o senhor(a) tem em dinheiro no caixa do estabelecimento?

22 – Quanto tem a receber das vendas que fez a fiado? E das vendas efetuadas com cartão de crédito (um dos estabelecimentos visitados tinha maquina(s) de cartão)?

23 – Quanto o senhor(a) tem de estoque, a preço de custo, na sua loja (o agente de crédito aponta para os itens que são considerados estoque e exemplifica, para facilitar o entendimento do cliente do que deve ser considerado pelo cliente na sua estimativa)?

A soma dos dados levantados nos itens de 21 a 23 são considerados como o ativo circulante.

24 – Quanto o senhor(a) tem investido na loja (novamente o agente aponta e exemplifica quais os itens devem ser considerados no cálculo – balcão, prateleiras, etc.)?

25 – O senhor(a) tem veículo próprio? E o estabelecimento é próprio?

Os valores levantados nos itens de 24 e 25, se existirem, são somados e classificados como ativo fixo. O valor do ativo fixo é somado ao valor do ativo circulante, sendo calculado o valor do ativo total.

26 – Quanto o senhor(a) tem a receber de compras já efetuadas no estabelecimento e que ainda não foram pagas? Quanto das compras é pago à vista e a prazo?

27 – O senhor(a) possui outras dívidas ainda não mencionadas?

Os valores levantados nos itens 26 e 27 são somados para o cálculo do passivo circulante.

29 – O agente questiona o cliente se alguma das dívidas, financiamentos e crediários informados anteriormente tem mais que 12 prestações a serem pagas e, em caso afirmativo, considera a soma destes valores como passivo de longo prazo. O passivo total é a soma do passivo circulante com o passivo de longo prazo.

Nesta fase, estão encerradas as questões, e o agente finaliza o preenchimento do formulário de levantamento sócio-econômico, calculando o patrimônio líquido do estabelecimento e os índices de endividamento.

As parcelas de microcrédito do caso estudado têm vencimento mensal e o agente ajusta a data de vencimento que melhor atenda ao grupo. Os levantamentos sócio-econômicos são efetuados separadamente com cada membro do grupo, mas o valor sugerido, a data de vencimento e as negociações/formalizações finais são efetuadas na presença de todos os membros do grupo. O valor emprestado e a data de vencimento devem ser acordados e aceitos por todos os membros do grupo.

O levantamento sócio-econômico, além de ser um material rico para a análise de crédito e obtenção de informações financeiras do cliente, pode ser uma ferramenta valiosa para que o empreendedor passe a conhecer conceitos básicos de gestão do negócio. A simples consultoria financeira prestada ajuda o microempreendedor a efetuar melhor gestão do seu negócio, potencializando seu crescimento. Esta também é uma função do agente de crédito.

4.3.2 Grupo Solidário

Os grupos são formados pelos próprios microempreendedores, não há indicação da instituição financeira para que um ou outro membro participe do grupo. Os grupos do caso estudado são formados por no mínimo 3 e no máximo 4 pessoas. A instituição financeira em questão não concede microcrédito de forma individual, exceto a destinação do crédito for investimento no negócio. Também não é concedido crédito individual para clientes novos.

A razão pela qual a escolha dos membros seja do próprio grupo, se fundamenta na necessidade de haver confiança entre os membros do grupo, pois em caso de inadimplência de um dos membros, todos serão impactados. É comum que o grupo se ajude mutuamente, sendo a forma mais comum de ajuda o pagamento de uma parcela da dívida de um dos membros por outro(s) membro(s) até que este possa se ajustar financeiramente e efetuar o pagamento ao seu colega.

Desde o processo inicial de levantamentos dos dados sócio-econômicos até formação do grupo, as pessoas sugerem, discutem e acordam as datas de vencimento e valores adequados as necessidades e possibilidade de pagamento de cada um dos microempreendedores.

4.4 FORMATAÇÃO DA BASE DE DADOS

Conforme já descrito anteriormente, o levantamento de dados inclui a extração de informações de duas bases de dados:

- Base clientes: registros com as informações sócio-econômicas, cadastrais dos clientes e informações básicas dos contratos de microcrédito. Caso o cliente tenha efetuado mais de uma operação com o IMF, o mesmo constará na base tantas vezes quantas forem as operações já realizadas pelo cliente na instituição. No caso de operações com grupo solidário, cada membro do grupo é um cliente e possui suas informações neste mesmo arquivo. A base não representa a totalidade de clientes de microcrédito da instituição financeira, sendo apenas uma amostra a ser utilizada no estudo;
- Base de contratos: registros com as informações das parcelas de cada contrato de microcrédito. A base não representa a totalidade de contratos de microcrédito da instituição financeira, sendo apenas uma amostra a ser utilizada no estudo.

Para estruturação da base consolidada com as informações para o estudo, foi efetuado o cruzamento das duas bases de forma a obter uma base única contendo:

- Informações cadastrais dos líderes, membros do grupo ou dos indivíduos (no caso dos empréstimos não solidários) que efetuaram operações de microcrédito com a IMF;
- Informações sócio-econômicas dos clientes;
- Informações das operações de microcrédito.

Foi constatado que as bases apresentavam várias inconsistências sendo as mais comuns: (i) contratos sem as informações de todas as parcelas da operação; (ii) operações sem o número do contrato de empréstimo, sendo impossível identificar as respectivas parcelas; (iii) registros duplicados; (iv) registros com mesmo número de contrato, mas para clientes distintos.

A qualidade das informações é condição indispensável para a criação de modelos de crédito consistentes, robustos e que possibilitem a mitigação do risco de crédito das instituições. Qualquer instituição que pretende desenvolver um modelo de *credit scoring* precisa iniciar o armazenamento das informações históricas de qualidade para que estas sejam fonte para a construção dos modelos. Segundo Mark Schreiner (2000), este é um dos principais desafios para as instituições de microcrédito.

Após os cruzamentos das bases de clientes e de contratos, foram excluídas:

- Operações de Pessoa Jurídica, pois a quantidade era muito pequena;
- Operações que todas as parcelas ainda se encontravam a vencer;
- Contratos com atraso superior a 360 dias (considerados como prejuízo).

Foram criadas algumas variáveis na base, entre elas: (i) dias de atraso médio: soma dos dias de atraso de cada parcela sobre a quantidade de parcelas pagas em atraso; (ii) dias de atraso da primeira parcela; (iii) somatória dos dias atrasos das parcelas; (iv) identificação de bom (0) ou mau (1) pagador; (v) percentual de garantia sobre o valor do desembolso; (vi) faixa de tempo de residência; (vii) parcela sobre vendas; (viii) despesa operacional sobre vendas; (ix) percentual de garantia.

Após todas as consistências e refinamentos na base de dados foi formatada uma base final com 102.777 contratos de clientes de microcrédito. Todos os contratos utilizados no estudo possuem pelo menos uma parcela já vencida, podendo a mesma ter sido: paga em dia, paga em atraso ou não paga. Deve ser ressaltado que a análise deste trabalho esta baseada nos clientes que tiveram suas operações aprovadas pela instituição financeira, não incluindo na base os clientes e operações que porventura tenham sido rejeitados pela instituição, pois os dados e os motivos de rejeições de clientes e operações não foram disponibilizados.

Para modelos de *credit scoring* novas variáveis explicativas são necessárias. Também para a variável dependente na análise de crédito, muitos autores precisam criar suas próprias variáveis, quando os dados não estão totalmente disponíveis. Neste trabalho, a variável dependente foi criada com base na média de dias de atraso de pagamento, de forma que seja possível classificar os contratos em bons e maus pagadores, sendo para tal utilizado o seguinte critério:

- Bom pagador(0): todo contrato cuja média de dias de atraso seja menor ou igual a 10 dias;
- Mau pagador(1): todo contrato cuja média de dias de atraso seja superior a 10 dias.

Da base final de contratos não foram excluídos os registros que porventura tenham algum campo inconsistente, como por exemplo: data de nascimento inválida, renda familiar não informada, ausência de situação do imóvel ou estado civil e etc. No caso de inconsistência de informação, o procedimento adotado foi o da exclusão da informação para o estudo, e não a exclusão do registro.

Sabemos que são considerados *outliers* as observações que ficam completamente fora da realidade estudada. Nos casos onde as informações apresentavam esta característica, as mesmas foram excluídas da análise, visando minimizar possíveis distorções dos resultados. Nos casos de *outliers*, o procedimento adotado também foi o da exclusão da informação para o estudo e não a exclusão do registro.

Após a classificação dos contratos em bons e maus pagadores, conforme critério descrito acima, análises numéricas foram elaboradas, visando o detalhamento e o melhor entendimento da base e do público tomador de microcrédito no cenário estudado, sendo possível verificar que:

- Dos 102.777 contratos, 10.6% foram classificados como contratos de maus pagadores e 89.4% contratos de bons pagadores;
- 67.2% dos contratos foram realizados com clientes mulheres e 32.8% com homens, confirmando a maior penetração de operações de microcrédito entre as mulheres de baixa renda;
- 10% dos contratos realizados com mulheres foram classificados como maus pagadores contra 11.8% dos contratos de homens;
- A maioria dos contratos, 69.5% deles não possuem a informação de estado civil, o que pode distorcer a análise. Dos contratos que possuem a informação de estado civil, 31.5% das operações foram contratadas por clientes solteiros ou casados (casado + marital). Nos contratos de operações de clientes casados foi constatado um menor percentual de maus pagadores;

- A maioria dos contratos, 69.5% deles não possuem a informação de situação do imóvel disponível, o que pode distorcer a análise. A maioria dos contratos que possuem a informação disponível, 26.5% se referem a operações contratadas por clientes com casa própria;
- 92.8% das operações são contratadas por grupos solidários. Pelas informações disponíveis não foi possível afirmar que as operações com grupos solidários possuem maior probabilidade de pagamento que as operações individuais, pois o percentual de operações individuais na base é muito pequeno, razão pela qual decidiu-se pela exclusão destes registros da análise;
- 98.2% das operações são efetuadas por grupos solidários representando um total de 26.084 grupos. Na média, cada grupo possui 3,86 membros. Conforme já mencionado anteriormente, a instituição financeira em questão forma grupos solidários de no mínimo 3 e no máximo 4 membros, escolhidos pelos próprios participantes do grupo, sem que a IMF interfira nesta formação.
- 54.9% dos grupos são mistos, formados por homens e mulheres. Os grupos formados apenas por homens (grupos masculinos) são a minoria, 8.9%. O percentual de maus pagadores nos grupos femininos é menor que em grupos mistos ou só masculinos. O percentual de maus pagadores entre os grupos mistos e só masculinos é praticamente o mesmo.

	Total Geral do Grupos		Total
	% Bom Pagador	% Mau Pagador	
Grupos Femininos	92.2%	7.75%	9,431
Grupos Masculinos	88.7%	11.30%	2,328
Grupos Mistos	91.2%	8.79%	14,325

Tabela 4: Bons e Maus Pagadores por Predominância de Sexo nos Grupos Solidários.
Elaborado pela autora.

- Praticamente 50% das operações de microcrédito da base estudada são concedidas a clientes novos e os outros 50% das operações são concedidas a clientes que já possuem histórico com a IMF e contrataram outra operação anteriormente. O percentual de maus pagadores nos contratos novos é 37.4% superior a média da base e o de bons pagadores 30% inferior a média da base;
- O prazo médio das operações da base analisada é de 5,94 meses, desta forma para efetuar a análise do prazo das operações e o percentual de maus pagadores por prazo das operações, foi efetuada a categorização das operações em 3 faixas de

prazo: (i) até 6 meses; (ii) de 7 a 12 meses; (iii) acima de 12 meses. Com esta classificação é possível observar que a maioria das operações, 86.7%, são efetuadas com prazo até 6 meses. O percentual de maus pagadores no universo de operações de microcrédito com prazo até 6 meses é similar ao da totalidade das bases, porém as operações com prazo superior a 6 meses possuem um percentual de maus pagadores bem superior a média da carteira, 24.4% e 35.4%, respectivamente para prazo de 7 a 12 meses e acima de 12 meses;

- A base possui operações de microcrédito efetuadas por 187 agentes de crédito diferentes. Na média cada agente efetuou 587 operações, dos quais 103 agentes apresentam em sua carteira percentual de maus pagadores abaixo da média da base e 13 agentes acima da média. Analisando as distribuições dos agentes que ficam acima ou abaixo da média de maus pagadores, é possível observar que a maior parte dos agentes de crédito, não possui grande variação em torno da média.

Após o levantamento inicial dos dados foi efetuada uma análise dos percentuais de bons e maus pagadores em cada uma das variáveis. No caso de variáveis com mais que dois valores possíveis foram realizados testes estatísticos, visando identificar qual dos valores seria o mais significativo. Para a classificação dos resultados de maior significância foi utilizado o valor de 1,96 (duas vezes a distância da média).

A figura abaixo contém a relação de todas as variáveis estudadas com seus respectivos valores possíveis e percentuais de bons e maus pagadores. As variáveis e valores significantes estão marcados com **, conforme a seguir:

VARIÁVEL	%BONS		%MAUS	
Total base	89,4%		10,6%	
SEXO				
Feminino	90,0%		10,0%	
Masculino	88,2%		11,8%	
ESTADO CIVIL				
Casado	91,0%	**	9,0%	**
Desquitado	84,8%		15,2%	
Divorciado	88,2%		11,8%	
Marital	89,8%		10,2%	
Separado	88,8%		11,1%	
Solteiro	87,4%	**	12,6%	**
Viúvo	90,9%		9,1%	
SITUAÇÃO IMÓVEL				
Alugada	85,7%		14,3%	**
Outros	86,2%		13,8%	**
Parentes	86,6%		13,4%	**
Própria	89,5%		10,5%	**
Própria financ.	86,1%		13,9%	
TIPO DE EMPRÉSTIMO				
Individual	80,4%		19,6%	
Solidário	89,4%		10,6%	
TIPO RENOVAÇÃO				
Novo	85,4%		10,2%	
Renovação	92,6%		7,4%	
PRAZO OPERAÇÃO				
Até 6m	89,8%		14,6%	**
De 7 a 12m	86,8%	**	13,2%	**
Acima 12m	75,0%		25,0%	**
FAIXA ETÁRIA				
De 18-28anos	86,4%	**	13,9%	**
De 28-38anos	89,6%		10,4%	
De 38-48anos	90,6%		9,4%	**
De 48-58anos	91,3%	**	8,7%	**
De 58-68anos	91,5%		8,5%	**
De 68-78anos	92,8%		7,2%	**
De 78-88anos	92,0%		8,0%	
Acima 88 anos	75,0%		25,0%	
TEMPO RESIDÊNCIA				
Até 1ano	89,3%		10,7%	
De 1-6anos	89,0%		11,0%	
De 6-11anos	88,9%		11,1%	
De 11-16anos	90,9%		9,1%	**
De 16-21anos	91,8%		8,2%	**
De 21-26anos	88,8%		11,2%	
De 26-31anos	90,5%		9,5%	
Acima 31anos	92,2%		7,8%	**

Tabela 5: Percentuais de Bons e Maus Pagadores por Variável da Base. Elaborado pela autora.

4.5 AS VARIÁVEIS E OS TESTES DE HIPÓTESE

Um total de 33 variáveis foi considerado no estudo. Muitas destas variáveis também podem ser encontradas nos estudos de modelos de *credit scoring* para microcrédito realizados por Schreiner (2004), Viganó (1993) e Diniz e Kleimeier (2007). Os dados basicamente podem ser classificados em três categorias: informações cadastrais, informações das operações de microcrédito e informações financeiras (sócio-econômicas) dos tomadores.

A seguir, as variáveis consideradas no estudo:

<i>VARIÁVEL</i>	<i>DESCRIÇÃO DA VARIÁVEL</i>	<i>TIPO</i>
Atraso Médio	Quantidade média de dias de atraso = somatória de dias de atraso / quantidade de parcelas pagas com atraso.	Calculada
Atraso da primeira Parcela	Quantidade de dias de atraso da primeira parcela.	Calculada
Total de dias de atraso	Quantidade total de dias de atraso = somatória de dias de atraso de todas as parcelas.	Calculada
Sucursal	Código da Sucursal que efetuou a venda da operação. As operações estão registradas em 11 sucursais: São Paulo, Rio de Janeiro, Pernambuco, Paraíba, Maranhão, Alagoas, Rio Grande do Norte, Sergipe, Santa Catarina, Piauí e Bahia.	Informada
Região	Código da Região que efetuou a venda da operação. As operações estão registradas em 5 (cinco) regiões: Região Sul /Sudeste, Região Norte I, II e III e Rio.	Informada
Agência	Nome da agência bancária onde a operação foi efetuada.	Informada
Agente	Código do Agente de Crédito responsável pela operação.	Informada
Ramo de Atividade	Código e Descrição do Ramo de Atividade do microempreendedor.	Informada
Código do Grupo	Código do Grupo Solidário (se operação em Grupo).	Informada
Indicativo de Líder	Indicativo de líder ou membro do grupo solidário (se operação em Grupo).	Informada
Sexo	Sexo do tomador do empréstimo.	Informada
Situação do Imóvel	Indicativo do tipo de imóvel do cliente: próprio, alugado, parentes, financiado ou outros.	Informada
Tempo de Residência	Tempo de residência do cliente no mesmo imóvel.	Informada
Tipo Abertura Contrato	Indicativo de operação com cliente novo ou já cliente.	Informada
Quantidade de Contratos	Quantidade de contratos que o microempreendedor possui com a instituição quando efetuou a operação que está registrada com esta	Informada

	quantidade.	
Quantidade de Parcelas	Quantidade de parcelas do empréstimo.	Informada
Valor da Parcela	Valor da parcela da operação de crédito.	Informada
Valor da Garantia	Valor da garantia dada para a operação. A garantia que foi aceita nos casos registrados, foi bem e o valor discriminado é o valor de avaliação do bem.	Informada
Valor do Desembolso	Valor total desembolsado na operação de crédito.	Informada
Valor da TAC	Valor da tarifa de contratação da operação de microcrédito que foi cobrada do tomador do empréstimo.	Informada
Valor do Seguro Prestamista	Valor do prêmio de contratação de seguro prestamista cobrado do tomador do microcrédito, caso ele tenha optado pelo seguro na contratação da operação.	Informada
Margem Líquida	Margem líquida do negócio	Informada
Vendas Mensais	Valor médio das vendas mensais em reais das vendas do microempreendedor (de acordo com o questionário de levantamento sócio-econômico).	Informada
Margem Bruta	Valor médio em reais da margem bruta obtida pelo microempreendedor com as vendas mensais = vendas mensais – custo das vendas mensais.	Informada
Lucro Líquido	Lucro líquido mensal do negócio = margem bruta – despesa operacional.	Informada
Renda Familiar	Valor da renda familiar mensal. (de acordo com o questionário de levantamento sócio-econômico).	Informada
Despesa Familiar	Somatória do valor das despesas mensais da família do microempreendedor: despesa de aluguel, água, luz, telefone, plano de saúde e etc.	Informada
Despesa Operacional	Somatória do valor das despesas mensais do microempreendedor com o negócio: despesa de aluguel, água, luz, telefone, funcionários e etc. (de acordo com o questionário de levantamento sócio-econômico).	Informada
Capacidade de Pagamento	Capacidade de Pagamento total do microempreendedor = Lucro Líquido + Renda Familiar – Despesa Familiar.	Informada
Capital do Trabalho	Excedente de ativos de curto prazo sobre passivo de curto prazo	Informada
Parcela sobre Vendas	Valor da Parcela/Vendas	Calculada
Parcelas sobre ML	Valor da Parcela/ML	Calculada
Despesa Operacional sobre Vendas	Despesa Operacional/Vendas	Calculada

Tabela 6: Variáveis da Base de Dados. Elaborado pela autora.

Nem todas as variáveis contidas na base serão significantes para distinguir os bons dos maus pagadores. A próxima etapa é identificar quais variáveis serão relevantes no estudo. Em

alguns estudos de *credit scoring* observa-se que após as análises de eficiência e aplicabilidade as variáveis iniciais são reduzidas para no máximo 10 ou 15.

O painel de estudos sobre *credit scoring* em operações de microcrédito, apresentado na página 28, aponta os resultados obtidos nos respectivos estudos e as variáveis que diferenciam os bons dos maus pagadores. Nestes estudos é possível observar que a maioria das variáveis chave para distinguir os bons dos maus pagadores são as variáveis que dizem respeito às informações cadastrais dos clientes. Visando comparar os resultados deste trabalho com estudos análogos, foram realizados testes de hipóteses com as variáveis cadastrais dos clientes tomadores de microcrédito que estão disponíveis na base de dados.

Para os testes de hipótese foi utilizado o software SPSS 17 seguindo o procedimento de: (i) elaboração de gráfico que permite a comparação percentual dos maus e bons pagadores dentro da variável estudada; (ii) teste de estatística descritiva - *crosstabs* ou *compare means (Test-T)*.

A tabela 7 apresenta as hipóteses nulas e as hipóteses alternativas testadas e seus respectivos resultados, conforme a seguir:

<i>Variável</i>	<i>Hipótese nula e alternativa</i>	<i>Res.</i>
Sexo	H ₀ : mulheres e homens possuem o mesma média de atraso de pagamento H ₁ : mulheres e homens não possuem a mesma média de atraso de pagamento	Rejeita H ₀
Faixa Etária	H ₀ : a média de idade entre bons e maus pagadores é a mesma H ₁ : a média de idade entre bons e maus pagadores é diferente	Rejeita H ₀
Faixa tempo de residência	H ₀ : a média de tempo de residência entre bons e maus pagadores é a mesma H ₁ : a média de tempo de residência entre bons e maus pagadores é diferente	Rejeita H ₀
Prazo	H ₀ : o prazo médio das operações de bons e maus pagadores é a mesma H ₁ : o prazo médio de operações de bons e maus pagadores é diferente	Rejeita H ₀
Quantidade de contratos	H ₀ : a média de quantidade de operações de bons e maus pagadores é a mesma H ₁ : a média de quantidade de operações de bons e maus pagadores é diferente	Rejeita H ₀

Sucursal	H ₀ : a média de quantidade de operações de bons e maus pagadores é a mesma H ₁ : a média de quantidade de operações de bons e maus pagadores é diferente	Rejeita H ₀
Região	H ₀ : a média de quantidade de operações de bons e maus pagadores é a mesma H ₁ : a média de quantidade de operações de bons e maus pagadores é diferente	Rejeita H ₀
Cidade	H ₀ : a média de quantidade de operações de bons e maus pagadores é a mesma H ₁ : a média de quantidade de operações de bons e maus pagadores é diferente	Rejeita H ₀
Agência	H ₀ : a média de quantidade de operações de bons e maus pagadores é a mesma H ₁ : a média de quantidade de operações de bons e maus pagadores é diferente	Rejeita H ₀
Estado Civil	H ₀ : a média de quantidade de operações de bons e maus pagadores é a mesma H ₁ : a média de quantidade de operações de bons e maus pagadores é diferente	Rejeita H ₀
Situação do Imóvel	H ₀ : a média de quantidade de operações de bons e maus pagadores é a mesma H ₁ : a média de quantidade de operações de bons e maus pagadores é diferente	Rejeita H ₀

Tabela 7: Resultados dos Testes de Hipótese. Elaborado pela autora.

Estes resultados são similares aos encontrados nos estudos de Laura Viganó (1993) que identificou as variáveis: sexo, estado civil e idade, como estatisticamente significantes e o trabalho de Diniz e Kleimeier (2007) que identificou as variáveis: local de residência, idade, estado civil e tempo de residência, como variáveis significantes.

Estes resultados ainda não são conclusivos sobre a determinação das variáveis chave para a classificação e distinção dos bons e maus pagadores. Por ora os resultados apresentados contemplam apenas testes isolados de cada uma das variáveis e que apontam para uma significância estatística para as variáveis.

4.6 O DESENVOLVIMENTO DO MODELO DE CRÉDITO

4.6.1 Calibração e Resultados

Modelos de *credit scoring* são comumente estruturados no modelo Z-score de Altman (1968). O modelo se utiliza de dados históricos onde é aplicada uma técnica estatística como, por exemplo, análise discriminante ou regressão logística, a fim de identificar quais variáveis chave são capazes de distinguir os adimplentes dos inadimplentes de operações de crédito. Após a definição destas variáveis, o modelo cria uma pontuação para cada tomador de crédito capaz de distinguir o adimplente do inadimplente. Esta pontuação deve ser comparada com uma frequência de corte para determinar se o tomador foi aceito, rejeitado ou se necessita de uma análise mais profunda.

O modelo definido com base nas informações históricas é utilizado para as novas operações podendo, desta forma, ajudar a classificar os tomadores e dando indicações para rejeitar, aceitar ou indicar a operação de crédito para uma análise mais detalhada, geralmente efetuada por analista e não pelos modelos automáticos. Uma calibração do modelo periodicamente é necessária para que as variáveis sejam retroalimentadas e que uma ótima frequência de corte seja determinada.

A primeira etapa do desenvolvimento do modelo será escolher as variáveis chave que serão utilizadas. Como nos estudos de Viganó (1993) e Dinh e Kleimeier (2007) inicia-se a modelagem com um número grande de variáveis e a primeira etapa consiste em identificar quais variáveis são as chaves para a construção do modelo. Primeiramente as variáveis numéricas foram avaliadas de acordo com os valores médios, excluindo-se os *missing values*. A base de dados possui um grande número de dados inválidos. Mesmo com a exclusão destes dados, algumas informações não parecem consistentes quando aplicadas a realidade do mercado de microcrédito, por isso alguns dos valores máximos ou mínimos das variáveis numéricas são *outliers*, distantes da realidade do mercado de microcrédito, portanto foram excluídas das análises.

A consistência e confiabilidade da base de dados é um assunto importante a ser discutido. Os dados foram digitados a partir de informações preenchidas em formulário em papel pelo agente de crédito na visita ao microempreendedor. As divergências e

inconsistências das informações pode ser fruto de: (i) divergência entre a informação coletada na entrevista e a transcrita pelo agente no formulário em papel; (ii) divergência entre a informação do formulário em papel e a informação digitada na área operacional. A existência de base de dados com informações históricas consistentes é chave para o sucesso do desenvolvimento e utilização de modelos estatísticos.

Abaixo, a tabela 8 onde é possível observar a média e o desvio padrão de cada uma das variáveis da base e os respectivos valores para os bons e maus pagadores.

<i>Variável</i>			<i>Bom Pagador</i>		<i>Mau Pagador</i>	
	<i>Média</i>	<i>Desvio Padrão</i>	<i>Média</i>	<i>Desvio Padrão</i>	<i>Média</i>	<i>Desvio Padrão</i>
Idade	39,59	11,55	39,87	11,53	37,16	11,43
Quantidade de contrato	2,75	1,28	2,86	1,27	1,85	1,03
Quantidade de parcelas	5,83	1,10	5,81	1,10	5,97	1,17
Valor da parcela	996,59	638,94	1012,88	651,61	855,77	494,92
Valor do desembolso	1742,61	1079,97	1763,83	1079,97	1559,07	494,92
Valor da garantia	5168,68	3531,59	5231,47	3568,32	4625,83	3144,95
Vendas mensais	6958,12	6614,38	6000,81	6472,60	5589,15	7723,73
Custos vendas	3378,04	4479,05	3410,76	4357,32	3095,14	5411,09
Margem bruta	2580,09	3208,16	2590,05	3223,52	2494,01	3071,36
Margem líquida	38,45	15,13	38,27	15,13	39,99	15,75
Lucro líquido	2091,96	2775,39	2101,63	2793,20	2008,35	2615,43
Renda familiar	173,50	367,16	172,02	364,17	186,31	391,88
Despesa familiar	708,47	538,35	710,27	536,14	692,92	556,96
Despesa operacional	488,13	838,07	488,82	836,60	485,66	850,80
Capacidade pagamento	1556,98	2553,87	1563,38	2574,22	1501,74	2370,41
Capital do trabalho	7291,21	14168,41	7403,50	14416,56	6320,59	11765,61
Parcela sobre vendas	0,22	0,13	0,22	0,13	0,21	0,12
Despesa Oper. sobre vendas	0,07	0,06	0,07	0,06	0,08	0,06

Tabela 8: Resultados das médias e desvios padrão de cada variável. Elaborado pela autora.

Para o presente estudo foram escolhidos os métodos para desenvolvimento de modelos de *credit scoring* mais utilizados nos estudos acadêmicos de aplicação de modelos estatísticos de *credit scoring*: a análise discriminante e a regressão logística. No processo de seleção das variáveis chave para a classificação dos bons e maus pagadores, foram testadas todas as variáveis da base e excluídas as que não se mostraram relevantes.

A base completa com os 102.777 registros foi utilizada, mas com a presença de inúmeras informações inconsistentes ou inválidas inviabilizou a utilização da base completa, sendo que no final do processo a base foi reduzida para 28.620 registros. Essa é a base que será trabalhada na modelagem.

É importante ressaltar, que essa base com apenas registros válidos, mantém distribuição similar entre bons e maus pagadores da base original e completa. A base reduzida apresenta 89.7% de bons pagadores e 10.3% de maus pagadores.

Os primeiros resultados obtidos com amostras aleatórias da base apresentaram resultados muito positivos para a classificação dos bons pagadores, porém resultados muito negativos para a classificação dos maus pagadores. Desta forma com as modelagens iniciais não foi possível encontrar um modelo que pudesse classificar bons e maus pagadores de forma relevante.

Para reduzir o efeito da presença massiva de bons pagadores nas amostras, foram criadas amostras com proporções maiores de maus pagadores, na tentativa de melhorar a classificação dos maus pagadores. A primeira idéia foi montar o modelo como uma amostra que contivesse a mesma proporção de bons e maus pagadores: 50%-50%. Com esta base, os modelos passaram a classificar melhor os maus pagadores, mas a classificação dos bons pagadores foi muito prejudicada.

Não faz parte no escopo do estudo discutir e considerar na análise os custos que a classificação errada de um bom pagador em mau pagador ou vice-versa pode trazer ao negócio, porém encontrar um bom percentual de acertos tanto para o bom como para o mau pagador é fundamental.

A amostra que resultou numa melhor *performance* do modelo quanto ao percentual de acertos de bons e maus pagadores foi a amostra que contém 25% de maus pagadores e 75% de bons pagadores. Essa foi a proporção da base utilizada para o desenvolvimento do modelo.

Os primeiros resultados levavam a conclusões de que as variáveis financeiras não pareciam ser as determinantes para a classificação do bom e mau pagador. As variáveis que se

mostravam mais importantes eram em sua maioria as informações cadastrais. Durante as simulações dos modelos nas técnicas de regressão logística e análise discriminante, variáveis foram incluídas e excluídas dos modelos diversas vezes e com as mais variadas combinações, buscando realmente encontrar as variáveis chave e confirmar ou não os resultados iniciais.

O modelo de *credit scoring* escolhido considerando a metodologia de análise discriminante e de regressão logística classificam como relevantes as mesmas variáveis, exceto a variável: sexo, que por ser categórica, foi considerada apenas na regressão logística.

É importante ressaltar que na modelagem de análise discriminante as premissas relativas ao seu uso foram consideradas como verdadeiras, sendo elas: (i) amostra com dois grupos: bons e maus pagadores; (ii) mais que duas observações; (iii) quantidade de variáveis discriminantes menor que o número de observações menos dois; (iv) variável discriminante combinação de outras variáveis discriminantes; (v) matriz de covariância iguais entre cada grupo; (vi) população com distribuição normal.

A modelagem pela técnica de regressão logística apresenta menor número de premissas e por isso é a mais utilizada nos modelos de crédito. A premissa básica apontada por Pampel (2000), para a regressão logística, é que a amostra deve apresentar não mais que dois grupos. Esta é a condição do modelo objeto deste estudo.

As variáveis identificadas como significativas no modelo foram: (i) variáveis cadastrais: idade, sexo, estado civil, tempo de residência, situação do imóvel e região geográfica de atuação; (ii) informações da operação e histórico com a instituição financeira: prazo da operação, quantidade de operações de microcrédito com a instituição financeira e indicação de contrato novo ou não; (iii) informações financeiras do microempreendedor: margem líquida, percentual de comprometimento do valor da parcela sobre as vendas mensais, percentual de comprometimento do valor da parcela sobre a margem líquida do negócio e capital do trabalho que é o excedente de ativos de curto prazo sobre passivos de curto prazo.

Avaliando mais detalhadamente as variáveis selecionadas como chave pelos modelos estatísticos, optou-se pela exclusão da variável capital do trabalho. Esta variável é calculada

com base nas informações que são colhidas nas entrevistas do agente de crédito: ativo circulante e passivo circulante. O cálculo de ativo e passivo circulante envolve um grande número de informações, como: contas a receber, caixa com bancos, adiantamentos, estoques, fornecedores, adiantamento à clientes, dívidas com terceiros e dívidas com a instituição. Como os microempreendedores não têm registros contábeis, estas informações são colhidas de forma subjetiva, com grande probabilidade de haver erro nas informações fruto principalmente da falta de gestão dos negócios desta natureza ou por risco moral, onde as informações fornecidas pelo tomador do empréstimo pode não refletir a realidade do negócio.

Neste cenário, as informações chave do modelo e consideradas mais confiáveis e com menor risco de manipulação pelos microempreendedores são as informações de vendas e custos mensais. Desta forma, as variáveis financeiras relevantes para a classificação de bons e maus pagadores são: margem líquida e percentual de comprometimento do valor da parcela sobre as vendas mensais. Sendo a margem líquida uma medida em percentual e não em valor, calculada pela diferença entre vendas mensais, custos mensais e despesa operacional dividido pelas vendas mensais

A tabela 9 a seguir, resume os resultados encontrados no software SPSS versão 17, aplicando-se os modelos de análise discriminante e regressão logística a uma amostra de calibração que contém 10.560 onde 25% da amostra são de maus pagadores e 75% de bons pagadores. As variáveis significativas ($\text{sig} < 0,05$) estão marcadas com **.

ANÁLISE DISCRIMINANTE - VARIÁVEIS ESCOLHIDAS		REGRESSÃO LOGÍSTICA - VARIÁVEIS ESCOLHIDAS	
	Sig.		Sig.
Idade	.000 **	IDADE	.000 **
QUANTIDADE DE CONTRATOS	.000 **	SEXONUM	.039 **
QUANTIDADE DE PARCELAS	.000 **	SEXONUM(1)	.013 **
TIPO DE ABERTURA DE CONTRATO	.000 **	SEXONUM(2)	.501
MARGEMLÍQUIDA	.013 **	QTDECONTRATO	.000 **
TEMPORRESIDENCIA	.000 **	QTDEPARCELAS	.000 **
PARCELASOBRE VENDAS	.025 **	ABERTURA CONTRATO	.000 **
SUCURSAL	.003 **	MARGEMLÍQUIDA	.013 **
SITIMOVEL	.000 **	TEMPORRESIDENCIA	.000 **
ESTADOCIVIL	.025 **	PARCELA SOBRE VENDAS	.025 **
		SUCURSAL	.003 **
		SITIMOVEL	.000 **
		ESTADOCIVIL	.025 **
			.000 **

Tabela 9: Resultados SPSS Análise Discriminante e Regressão. Elaborado pela autora.

Na análise discriminante foi elaborado modelo com 69.2% de acerto sendo o acerto de bons pagadores de 67.9% e de maus pagadores de 73%. Como variáveis chave foram mantidas as já escolhidas anteriormente. As variáveis positivamente relacionadas são: idade, quantidade de contratos, quantidade de parcelas, índice de parcelas sobre vendas e sucursal. As variáveis negativamente relacionadas são: tipo de abertura de contrato, situação do imóvel, margem líquida, tempo de residência e estado civil. Sendo sexo uma variável categórica, a mesma não foi utilizada no modelo de análise discriminante.

Na regressão logística foi elaborado modelo com 78.5% de acerto sendo o acerto de bons pagadores de 93.1% e de maus pagadores de 39%. As variáveis positivamente relacionadas são: quantidade de parcelas, tipo de abertura do contrato, margem líquida, tempo de residência, situação do imóvel e estado civil. As variáveis negativamente relacionadas são: idade, sexo, quantidade de contratos, e sucursal.

Os achados são similares aos de trabalhos acadêmicos sobre modelos de *credit scoring* para microcrédito já realizados. Viganó (1993) destaca as variáveis: sexo, idade e estado civil como significativas. Schreiner (2004) destaca o maior risco de atraso à medida que o cliente toma novos empréstimos. Dinh e Kleimeier (2007) destacam local de residência, empréstimos anteriores, idade, estado civil, tempo de residência e tipo de residência. Como no presente estudo, os estudos pesquisados encontram poucas informações financeiras capazes de serem significativas para a classificação de bons e maus pagadores.

Estritamente sob o ponto de vista de modelo de crédito os resultados encontrados sugerem que praticamente todas as informações sócio-econômicas levantadas pelo agente de crédito nas entrevistas iniciais não contribuem para a classificação do cliente em bom ou mau pagador. As únicas informações financeiras relevantes são o volume de venda mensal e a margem líquida do negócio, sendo esta calculada em função do volume de vendas, custos mensais e despesa operacional do negócio.

Como o percentual de acertos encontrado para os bons e maus pagadores são muito distintos, recomenda-se que em estudos futuros seja analisada a hipótese de desenvolvimento de modelos distintos para os bons e maus pagadores com base nas informações históricas disponíveis.

4.6.2 Validação

Aplicando-se o modelo encontrado em uma amostra de validação sendo esta uma amostra aleatória de 3.550 registros selecionados de forma aleatória da base geral dos quais 354 registros de maus pagadores e 3.196 registros de bons pagadores obtemos os seguintes resultados: (i) na análise discriminante: o modelo obteve 67.6% de acerto, sendo 68.8% nos bons pagadores e 56.5% de acertos nos maus; (ii) na regressão logística: o modelo obteve 62.3% de acerto, sendo 63.3% nos bons pagadores e 53.4% de acertos nos maus.

Neste estudo não está sendo abordado ou considerado o custo que existe quando um modelo efetua a classificação errada de maus pagadores em bons pagadores ou vice-versa.

Frente à amostra de calibração, os resultados encontrados na amostra de validação tiveram piora na sua *performance* de maneira geral. Para calibração do modelo, foi necessário utilizar uma amostra que não representava a proporção de bons e maus pagadores da base total válida. Já a amostra de validação representa melhor a proporção da população da base, pois foi uma amostra aleatória.

A tabela abaixo resume os resultados:

Modelo / Técnica	Amostra Calibração			Amostra Validação		
	% Acerto	% Acerto Bons	% Acerto Maus	% Acerto	% Acerto Bons	% Acerto Maus
Análise Discriminante	69.2%	67.9%	73.0%	67.6%	68.8%	56.5%
Regressão Logística	78.5%	93.1%	39.0%	62.3%	63.3%	53.4%

Tabela 10: Resultado dos modelos nas amostras de calibração e validação. Elaborado pela autora.

4.7 BAGGING E BOOSTING

Bagging (abreviação de *Bootstrap Aggregating*) (Breiman, 1996) é um dos primeiros algoritmos de aprendizagem de máquina. É também um dos mais intuitivos e simples de implantar com uma melhora de desempenho surpreendente. A diversidade no *bagging* é obtida por meio de réplicas dos dados de treino, ou seja, subconjuntos diferentes da base de treinamento são sorteados aleatoriamente, com repetição e sempre com a mesma quantidade de registros. Cada subconjunto é usado para treinar o modelo. A escolha do resultado é obtida por votação, onde o resultado mais votado é o escolhido como verdadeiro. É um algoritmo que reduz de variância.

Boosting (Freund e Schapire, 1996; Schapire, 1990) é também uma técnica de aprendizagem de máquina onde a idéia básica é aplicar o modelo várias vezes a uma base de treino, produzindo assim uma sequência de treinamento para um número pré-definido de iterações, transformando um classificador fraco em um classificador forte.

Neste trabalho, utilizou-se das duas técnicas de aprendizagem de máquina, *bagging* e *boosting*, com o intuito de testar se estas técnicas trazem melhora de desempenho aos modelos de *credit scoring* construídos e comparar os resultados obtidos nas duas técnicas.

No caso do *bagging*, uma amostra de 1.000 registros foi selecionada e a partir desta foram criadas 80 sub-amostras aleatórias também de 1.000 registros cada uma com repetição de registros da amostra original. Os resultados dos 80 processamentos do modelo foram aplicados a uma amostra de validação com 3.550 registros, onde o resultado mais votado foi o escolhido como o correto.

Os resultados mostram melhora de *performance* com a aplicação do algoritmo de *bagging* tanto para o modelo de análise discriminante como para o modelo de regressão logística. Na análise discriminante houve uma melhora de 10.9% no percentual total de acertos, sendo importante ressaltar que quando separamos este percentual de acertos do modelo entre bons e maus pagadores, o *bagging* trouxe melhora apenas na *performance* dos bons pagadores. Para os bons pagadores, uma melhora de 13.9%, porém para os maus pagadores, uma piora de 15.3%.

Resultados similares são observados para a regressão logística, onde o *bagging* incrementou a *performance* de acertos do modelo em 18.6%, porém quando separamos os resultados entre bons e maus pagadores, o *bagging* só demonstrou ganhos para os bons pagadores. Para os bons pagadores, uma melhora de 23.0%, porém para os maus pagadores, uma piora de 21.2%.

Para o *boosting* foi utilizada uma amostra de 3.550 registros, a mesma amostra que foi utilizada para a validação do *bagging*. Foram processadas em 80 iterações, buscando transformar um estimador fraco em um estimador forte.

O modelo subdivide a amostra em duas partes, sendo uma a amostra de calibração e a outra, a amostra de treino. Os resultados encontrados mostram que o *boosting* melhorou a *performance* do modelo de maneira significativa. O algoritmo de *boosting* com as 80 iterações tem um acerto de 94.5%, sendo 99.5% nos bons pagadores e 55.6% nos maus pagadores. Neste caso, diferentemente dos resultados encontrados no *bagging*, foi obtida uma melhora no percentual de acertos tanto para os bons como para os maus pagadores, porém a melhora para os maus pagadores é praticamente nula.

Abaixo estão demonstrados os resultados obtidos nos testes dos dois algoritmos.

Modelo / Técnica	% Acerto	% Acerto Bons	% Acerto Maus
Análise Discriminante	67.6%	68.8%	56.5%
Regressão Logística	62.3%	63.3%	53.4%
Bagging Análise Discriminante	78.5%	82.7%	41.2%
Bagging Regressão Logística	80.9%	86.3%	32.2%
Boosting	95.6%	99.3%	62.1%

Tabela 11: Resultados *Bagging* e *Boosting*. Elaborado pela autora.

Segundo Wang et al. (2010), os algoritmos de *bagging* e *boosting* trazem melhoria de *performance* dos modelos de *credit scoring*. As conclusões do estudo de Wang et al. (2010) apontam uma melhor *performance* do *bagging* sobre o *boosting*. Os resultados do presente estudo apontam também um melhor resultado para o *boosting*.

Na maioria das vezes, espera-se que o *boosting* seja melhor, pois o algoritmo atribui um peso maior para as observações que são mais difíceis de serem classificadas. O *boosting* procura melhorar as estimativas, alterando as probabilidades nas amostras subseqüentes em função da dificuldade de classificação numa dada observação. No *bagging* repete-se várias vezes um mesmo procedimento em amostras de mesma quantidade de observações e com repetição de observações, sem a preocupação de ajuste de amostras existente no *boosting*. O *boosting* permite uma classificação melhor com menos re-amostragens.

Para a comparação dos resultados foi utilizada a mesma quantidade de iterações para os dois algoritmos de aprendizagem de máquina - *bagging* e *boosting*. Para testar a sensibilidade dos resultados dos algoritmos com a variação das quantidades de iterações, foi efetuada a aplicação das técnicas com quantidades distintas e crescentes de iterações. Os resultados obtidos são mostrados nas figuras e tabelas a seguir:

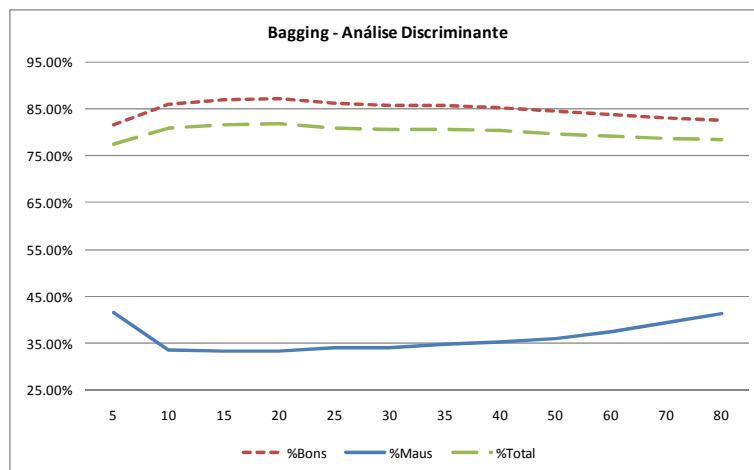


Figura 11: Percentual de acertos por Iterações. Elaborado pela autora.

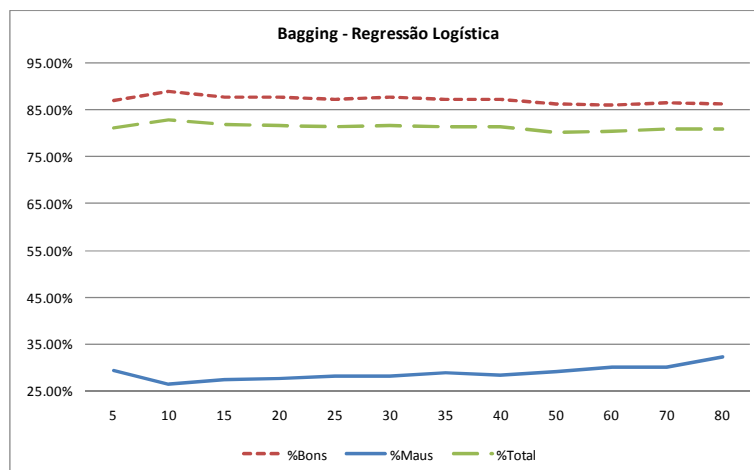


Figura 12: Percentual de acertos por Iterações. Elaborado pela autora.

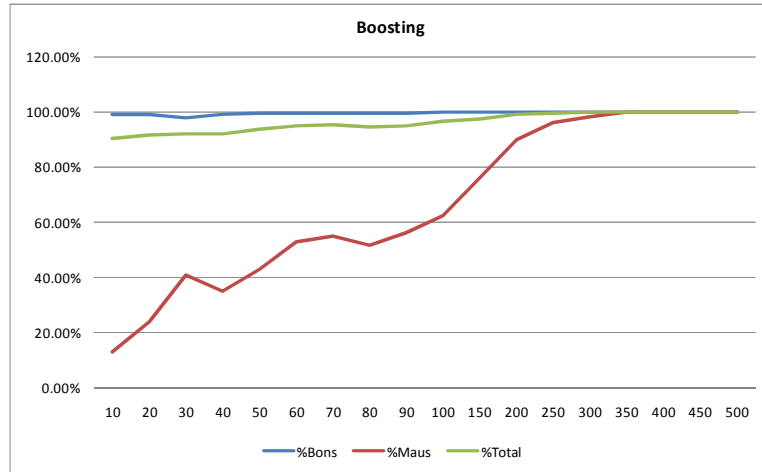


Figura 13: Percentual de acertos por Iterações. Elaborado pela autora.

Modelo / Técnica	% Acerto	80 iterações		% Acerto	50 iterações	
		% Acerto Bons	% Acerto Maus		% Acerto Bons	% Acerto Maus
Análise Discriminante	67.6%	68.8%	56.5%		68.8%	56.5%
Regressão Logística	62.3%	63.3%	53.4%	62.3%	63.3%	53.4%
Bagging Análise Discriminante	78.5%	82.7%	41.2%	79.7%	84.5%	35.9%
Bagging Regressão Logística	80.9%	86.3%	32.2%	80.6%	86.3%	29.1%
Boosting	95.6%	99.3%	62.1%	92.9%	99.3%	38.2%

Modelo / Técnica	% Acerto	30 iterações		% Acerto	10 iterações	
		% Acerto Bons	% Acerto Maus		% Acerto Bons	% Acerto Maus
Análise Discriminante	67.6%	68.8%	56.5%	67.6%	68.8%	56.5%
Regressão Logística	62.3%	63.3%	53.4%	62.3%	63.3%	53.4%
Bagging Análise Discriminante	80.7%	85.8%	34.8%	80.8%	86.0%	33.6%
Bagging Regressão Logística	81.7%	87.6%	28.2%	82.7%	88.8%	26.5%
Boosting	91.3%	97.7%	39.7%	91.3%	99.5%	10.4%

Tabela 12: Resultados *Bagging* e *Boosting* com quantidades diferentes de iterações. Elaborado pela autora.

Os testes foram efetuados processando-se os modelos variando a quantidade de iterações. Para o *bagging*, ao variar a quantidade de iterações, com o intuito de estudar a sensibilidade da *performance* em relação a esta variação, observa-se que de forma geral há apenas melhora na *performance* dos bons pagadores, sendo que não há incremento de *performance* à medida que há o aumento da quantidade de iterações. A melhor *performance* para os bons pagadores é obtida com 20 iterações para a análise discriminante e com 10 iterações para a regressão logística. O resultado esperado é que o aumento de iterações traga sempre melhora nos modelos originais.

No caso do *boosting*, à medida que há o aumento da quantidade de iterações há conseqüente melhora na *performance* do modelo. Esse comportamento é observado tanto para os bons como para os maus pagadores. É observada redução da *performance* dos maus pagadores quando há uma pequena quantidade de iterações.

Comparando-se os resultados, nota-se que o *boosting* tem uma significativa melhora de *performance*, principalmente nos maus pagadores, à medida que a quantidade de iterações aumenta, o que gera elevação do percentual de acertos dos maus pagadores de 10.4% para 62.1%. Vale destacar que para quantidades de iterações menores que 80, o *boosting* não trouxe melhora de *performance* no modelo para os maus pagadores, percentual este que já era próximo de 55%. Só há uma melhora do percentual de acertos dos maus pagadores quando se processa as 80 iterações, e, mesmo neste caso, quase nula.

Quando comparamos os resultados do *bagging*, os mesmos parecem contra intuitivos, pois o percentual de acertos do modelo com as 10 iterações é superior a *performance* para as 80 iterações. Vale salientar que mesmo com a menor quantidade de iterações já houve melhora do modelo, porém a melhora ocorreu apenas nos bons pagadores, não havendo melhora de percentual de acertos nos maus pagadores nem quando são processadas as 80 iterações.

Para um último teste, foi efetuado o processamento do *boosting* a partir de 100 até 500 iterações, com aumento de 50 em 50, com o objetivo de verificar a variação do percentual de acertos que este algoritmo traria o aumento gradual da quantidade de iterações, uma vez que o percentual de acerto do modelo para 80 iterações já é de 95.6%. Com 150 iterações foi produzido um acerto de 100% no modelo para os bons pagadores e com 300 iterações 100% de acerto no modelo para os maus pagadores. Segundo Schapire (1996) a aplicação do algoritmo de *boosting* pode transformar um estimador fraco em forte trazendo grande melhora no desempenho do modelo, isto é, 99% de acertos. Recomenda-se que o resultado de 100% de acertos, encontrado neste estudo, seja mais bem explorado e detalhado em estudos futuros.

5 CONCLUSÕES

O microcrédito, que surgiu com o Grameen Bank de Muhammad Yunus em 1976, foi a semente que trouxe ao mercado de microcrédito mais que uma inspiração sobre a forma de conceder crédito ao público de baixa renda. Trouxe inovação e metodologias que pudessem ser replicadas em uma série de países, beneficiando o pequeno tomador de crédito e também as IMFs.

A promessa de estar ao alcance dos pobres traz ao microcrédito o desafio de ser ofertado aos clientes de baixa renda de forma a atender às necessidades de desenvolvimento dos seus negócios para a geração de renda. Isso implica em oferecer aos clientes linhas de crédito com características mais flexíveis e que se adéquem melhor a este público: como parcelas com datas de vencimentos mais próximas (semanais, por exemplo), créditos menores no início e maiores à medida que o cliente se torna mais conhecido pela IMF, prazos mais curtos e etc. Não se trata de fornecer linhas de crédito para consumo, mas sim de possibilitar aos microempreendedores o crescimento do seu negócio e a geração de renda. É o chamado microcrédito produtivo orientado.

No Brasil, as concessões de operações de microcrédito têm crescido nos últimos anos, mas ainda são pequenas se comparadas com outros países. O microcrédito para consumo tem cedido espaço para o microcrédito produtivo orientado. Mesmo os incentivos governamentais e a entrada de algumas IMFs no segmento ainda não foram suficientes para que houvesse uma grande alavancagem e crescimento do microcrédito.

O crescimento depende, entre outras coisas, de as IMFs verem vantagens e quererem investir cada vez mais neste segmento. Apesar de frentes afirmarem que, para o público de baixa renda, ter crédito é mais importante do que o preço que se paga por ele, os preços das operações de microcrédito são regulados por lei e as IMFs precisam encontrar vantagens para investirem: o conhecido ganha-ganha. Dentre as barreiras, a redução dos custos envolvidos no processo é a maior e mais difícil delas.

Desde as primeiras operações de microcrédito, muitas inovações foram desenvolvidas e implantadas, sendo as mais conhecidas: o papel do agente de crédito e o grupo solidário. As

duas iniciativas reconhecidamente de sucesso. O grupo solidário trouxe ao negócio a mitigação de riscos; o agente de crédito, qualidade no conhecimento do cliente e adequação ao perfil deste cliente. Um caminho sem volta.

Crescer significa ter economia de escala e, neste contexto, o estudo propõe-se a discutir a inclusão de uma modelagem de crédito dentro do microcrédito como a próxima inovação neste mercado, como discutido em estudos acadêmicos de operações de microcrédito no Vietnã, Bolívia e África.

A partir de uma amostra de operações de microcrédito de uma instituição financeira foi desenvolvida uma modelagem de crédito, buscando encontrar as variáveis chave para distinguir os bons dos maus pagadores. As variáveis chave encontradas já foram discutidas em alguns estudos acadêmicos e bibliografias de microcrédito.

As mulheres são a maioria na amostra estudada e são melhores pagadoras do que os homens. Os grupos solidários trazem redução de riscos, mas a amostra não possuía operações de microcrédito individuais suficientes para afirmar empiricamente que os empréstimos com grupos solidários apresentam menor risco.

Cientes com maior quantidade de contratos com a instituição financeira são melhores pagadores. Operações com maior prazo apresentam maior risco de não pagamento. Os clientes casados são melhores pagadores e contra intuitivamente, os clientes de negócios com maiores margens líquidas se mostram como piores pagadores.

As variáveis selecionadas como chaves pelos modelos, utilizando as metodologias de análise discriminante e regressão logística, trazem uma evidência de que, sob o ponto de vista específico de risco de crédito, a maior parte das informações colhidas pelos agentes de crédito nas entrevistas para levantamento sócio-econômico não sejam relevantes. As únicas informações financeiras relevantes para o modelo de crédito foram volume de vendas mensais e margem líquida, sendo esta calculada em função do volume de vendas, custos mensais e despesa operacional do negócio.

As variáveis chave para o modelo são, em sua maioria, informações cadastrais dos clientes, sendo elas: idade, sexo, estado civil, região de atividade do microempreendedor,

tempo de residência e situação do imóvel. Além das informações financeiras e cadastrais já descritas, o modelo classificou a quantidade de contratos com a instituição, a quantidade de parcelas e o tipo de contrato (novo ou renovação) como significativas.

Esta evidência sugere que há uma possibilidade de redução de custos restringindo o levantamento sócio-econômico apenas para as variáveis mais significativas. Esta alternativa não considera a utilidade que, porventura, as informações colhidas e registradas possam ter para as instituições financeiras sob o ponto de vista de relacionamento com o cliente, de serviço de consultoria financeira prestada aos clientes e de armazenamento de informações históricas. Trata-se de uma alternativa apenas sob a ótica da modelagem de crédito desenvolvida.

A importância do papel do agente de crédito e a contribuição que os grupos solidários trouxeram ao negócio não devem ser eliminadas com a implantação dos modelos de *credit scoring*.

Além do desenvolvimento do modelo de *credit scoring* para o microcrédito, o estudo discutiu a contribuição e melhora de desempenho que o modelo tem, utilizando-se dos algoritmos de aprendizagem de máquina: *bagging e boosting*.

Os resultados mostram que os algoritmos de *bagging e boosting* acrescentam melhora de desempenho nos modelos, sendo a *performance* do *boosting* superior a do *bagging* e os ganhos encontrados nos percentuais de acerto do modelo para os bons pagadores superiores ao dos maus pagadores. Recomenda-se que, em estudos futuros, as razões das diferenças de melhora no desempenho encontradas para os bons e maus pagadores sejam mais bem exploradas e aprofundadas. Recomenda-se também que o resultado encontrado no *boosting* de 100% de acertos para bons e maus pagadores com 150 e 300 iterações respectivamente, também seja mais bem explorado em estudos futuros.

A questão de geração de renda e melhoria dos negócios dos microempreendedores com as operações de microcrédito foi discutida de forma teórica, não sendo escopo deste trabalho a avaliação empírica das contribuições que o microcrédito trouxe aos negócios dos microempreendedores.

As questões básicas sobre a relevância das informações históricas para o desenvolvimento dos modelos abordadas por Schreiner (2000) foram observadas durante as análises da base de dados, pois diversas inconsistências reduziram a amostra inicial de 102.777 registros para 28.620 registros. Os dados disponíveis e válidos foram suficientes para o desenvolvimento do modelo; porém, não foi considerada, neste estudo, a imprecisão que pode haver no modelo devido às possíveis inconsistências da base ou à redução significativa da amostra por dados inválidos.

As instituições que pretendem aplicar modelos de *credit scoring* para a concessão de crédito das operações de microcrédito devem iniciar imediatamente o processo de obtenção e armazenamento de informações de qualidade, para que as mesmas possam ser utilizadas para os modelos de *credit scoring*: a próxima inovação do mercado de microcrédito.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ALFARO, Esteban et al. *Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks*. Decision Support Systems, 45(1): 110-122, 2008.
- ALTMAN, Edward. I. *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*. Journal of Finance, 23, 589-609, 1968.
- ARMENDÁRIZ, Beatriz.; MORDUCH, Jonathan. *The Economics of Microfinance*. First MIT Press Paperback edition, 2007. Massachusetts Institute of Technology, 2005.
- ANDERSON, Raymond. *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford University Press, 2007.
- ANTHONY, Denise. *Cooperation in Microcredit Borrowing Groups: Identity, Sanctions, and Reciprocity in the Productions of Collective Goods*. American Sociological Review, Vol.70, 2005.
- BALASUBRAMANIAN, Aditya. *Micromanagement: Fixing Microfinance in Argentina*. Harvard International Review – Fall, 2009.
- BANCO CENTRAL DO BRASIL. Acessado em 13 de abril de 2010, de: <http://www.bcb.gov.br>
- BARTLETT, Peter.; SHAWE-TAYLOR, John. *Generalization Performance of Support Vector Machines and Other Pattern Classifiers*. In "Advances in Kernel Methods, Support Vector Learning", Bernhard Scholkopf, Christopher J.
- BAUER, Eric.; KOHAVI, Ron. *An empirical comparison of voting classification algorithms: Bagging, boosting, and variants*. Machine Learning, 36(1/2), 105–139, 1999.
- BREIMAN, Leo. *Bagging Predictors*. Machine Learning, 24, 123-140. Kluwer Academic Publishers, Boston, 1996.
- BREIMAN, Leo. *Bias, variance, and arcing classifiers*. Technical Report 460, Department of Statistics, University of California, Berkeley, CA, 1996b.
- BREIMAN, Leo. *Arcing classifiers*. The Annals of Statistics, 26(3): 801-849, 1998.
- BRUZZONE Lorenzo.; COSSU Roberto.; VERNAZZA Gianni. *Detection of land-cover transitions by combining multivariate classifiers*. Pattern Recognition Letters, 25(13): 1491-1500, 2004.
- BUHLMANN, Peter.; YU, Bin. *Boosting with L2 loss: Regression and classification*. Journal of the American Statistical Association, 98, 324-338, 2003.

CULL, Robert; DEMIRGUÇ-KUNT, Asli; MORDUCH, Jonathan. *Microfinance Meets the Market*. Journal of Economic Perspectives-Volume 23, Number 1, Pages 167-192, 2009.

DENISON, David D. et al. *Nonlinear estimation and classification*. New York, 2003.

DICIONÁRIO AURÉLIO ON LINE DE LÍNGUA PORTUGUESA. Acessado em 18 de maio de 2010, de: <http://www.dicionariodoaurelio.com>

DINH, Thi H. T.; KLEIMEIER, Stefanie. *A credit scoring for Vietnam's retail banking market*. International Review of Financial Analysis.

DROLET, Juliet. *Women and Microcredit: Implications for Social and Economic development*. International Consortium for Social Development, 2009

FREUND, Young.; SCHAPIRE, R. E. *Experiments with a New Boosting Algorithm*. Machine Learning: Procediments of the Thirteenth International Conference, 1996.

FREUND, Young.; SCHAPIRE, R. E. *Discussion of the paper "Arcing Classifiers"* by Leo Breiman. The Annals of Statistics, 26(3)Ç 824-832, 1998.

GOMEZ, Rafael; SANTIR, Eric. *Membership has its privileges: the effect of social capital and neighborhood characteristics on the earnings of microfinance borrowers*. Canadian Journal of Economics. Revue canadien d'Économique. Vol. 34, No.4, 2001.

HAYKIN, Simon. *Redes Neurais – Princípios e Práticas*. Bookman – 2a. Edição, 1999.

HSIEH, Nan-Chen; HUNG, Lun-Ping; HO, Chia-Ling. *A data driven ensemble classifier for credit scoring analysis*. Expert Systems with Applications, 37(1): 534-545, 2010.

HONGRUI, Xie et al. *Solving Credit Scoring Problem with Ensemble Learning: a Case Study*. Second International Symposium on Knowledge Acquisition and Modeling, 1:51-54, 2009.

KLECKA, William R. *Discriminant Analysis. Series: Quantitative Applications in the Social Sciences*. Sage Publications, 1980.

LAI, Kin K. et al. *Credit Risk Analysis Using a Reliability-Based Neural Network Ensemble Model*. Lecture Notes in Computer Science, Artificial Neural Networks – ICANN 2006.

LAI, Kin K.; HE, Kaijian; YEN, Jerome. *Modeling VaR in Crude Oil Market: A Multi Scale Nonlinear Ensemble Approach Incorporating Wavelet Analysis and ANN*. Lecture Notes in Computer Science, Computational Science – ICCS 2007.

LEDGERWOOD, Joanna. *Microfinance Handbook*. The International Bank for Reconstruction and Development - The World Bank, 1999.

LEIGH, William.; PURVIS, Russell.; RAGUSA, James M. *Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural networks, and genetic algorithm: a case study in romantic decision support*. Decision Support Systems 32(4): 361-377, 2002.

MACLIN, Richard.; OPITZ, David. *An empirical evaluation of bagging and boosting*. In Proceedings of the Fourteenth National Conference on Artificial Intelligence (pp. 546–551). Cambridge, MA: AAAI Press/MIT Press, 1997.

MAIMON, Oded.; ROKACH, Lior. *Ensemble of Decision Trees for Mining Manufacturing Data Sets, Machine Engineering*, vol. 4 N01-2, 2004.

MICROFINANCE INFORMATION EXCHANGE. Acessado em 05 de maio de 2010, de: <http://www.mixmarket.org>

MICROBANKING BULETTIN. Acessado em 05 de maio de 2010, de: <http://www.themix.org>

MCKERMAN, Mary. *The impact of microcredit programs on self-employment profits: do noncredit program aspects matter?* The Review of Economics and Statistics, 2002.

MOKEDDEM, Djamila.; BELBACHIR, Hafida. *A survey of distributed classification based ensemble data mining methods*. Journal of Applied Sciences, 9(20): 3739-3745, 2009.

MONZONI, Mario. *Impacto em Renda do Microcrédito*. Editora Petrópolis, 2007.

MORDUCH, Jonathan. *The Microfinance Promise*. Journal of Economic Literature. Volume XXXVII. December 1999. Pág. 1569-1614, 1999.

OPITZ, David.; MACLIN, Richard. *Popular ensemble methods: an empirical study*. Journal of Artificial Intelligence Research, 11: 169-198, 1999.

PALEOLOGO, Giuseppe; ELISSEEFF, André; ANTONINI, Gianluca. *Subagging for credit scoring models*. European Journal of Operational Research, 201(2): 490-499, 2010.

PAMPEL, Fred C. *Logistic Regression: A primer. Series: Quantitative Applications in the Social Sciences*. Sage Publications, 2000.

PRAHALAD, Coimbatore. K. *The Fortune at the Bottom of the Pyramid*. Wharton School Publishing, 2009.

PRIOR, Francesc.; ARGANDOÑA, Antonio. *Credit accessibility and corporate social responsibility in financial institutions: the case of microfinance*. Business Ethics: A European Review. Volume 18, Number 4, October 2009.

ROKACH, Lior. *Ensemble methods for classifiers. Data Mining and Knowledge. Discovery Handbook*, USÇ Springer 2005.

ROKACH, Lior. *Taxonomy for characterizing ensemble methods in classification tasks: a review and annotated bibliography*. Computational Statistics & Data Analysis, 53(12): 4046-4072, 2009.

SCHAPIRE, Robert E. *The boosting approach to machine learning: an overview*. In

SCHREINER, Mark. *Credit Scoring for Microfinance: Can it Work?* Center for Social Development. Washington University in St. Louis, 2000.

SCHREINER, Mark. *Scoring: The next breakthrough in microcredit?* CGAP Occasional Paper No. 7, <http://www.cgap.org>, 2003.

SCHREINER, Mark. *Scoring arrears at the microlender in Bolivia.* *Journal of Microfinance*, 6, 65-88, 2004.

SCHURMANN, Anna T.; JOHNSTONS, Heidi B. *The Group-lending Model and Social Closure: Microcredit, Exclusion, and Health in Bangladesh.* International Centre of Diarrhoeal Disease Research, Bangladesh, 2009.

SILVA, José P. da. *Gestão e Análise de Risco de Crédito.* Editora Atlas – 5ª.Edição, 2006.

SKURICHINA, Marina; DUIN, Robert P. W. *Boosting in Linear Discriminant Analysis.* Department of Applied Physics, 2002.

TAN, Aik C.; GILBERT, David; DEVILLE, Yves. *Multi-class Protein Fold Classification using a New Ensemble Machine Learning Approach.* *Genome Informatics*, 14:206- 217,2003.

THE CONSULTATIVE GROUP TO ASSIST THE POOR (CGAP). Acessado em 01 de maio de 2010, de: <http://www.cgap.org>

THE MICROCREDIT SUMMIT CAMPAIGN. Acessado em 20 de abril de 2010, de: <http://www.microcreditsummit.org>

TUKEY, John W. *Exploratory data analysis.* Addison-Wesley, Reading, Mass, 1977.

TUMER, Kagan; GHOSH, Joydeep. *Robust Order Statistics based Ensembles for Distributed Data Mining.* In KARGUPTA, H.; CHAN, P., eds, *Advances in Distributed and Parallel Knowledge Discovery*, pp. 185-210, AAAIMIT Press, 2000.

VIGANÓ, Laura. *A credit scoring model for development Banks: An African case study.* *Savings and Development*, 4, 441-482, 1993.

WANG, Gang et al. *A comparative assessment of ensemble learning for credit scoring.* Elsevier, 2010.

WITTEN, Ian H.; FRANK, Eibe. *Data Mining – Practical Machine Learning Tools and Techniques.* Elsevier, 2005.

YU, Bin. (ed). *Non-linear estimation and classification.* Springer, 2003.

YUNUS, Muhammad. *O Banqueiro dos Pobres.* Editora Ática, 1997.

APÊNDICE A: Algoritmo de *Boosting*

```
#####
# Assunto: Algoritmo de boosting (usando loss logística) – R Project
# Banco: amostra.sav
#####

# Instruções:

## No meu micro:
setwd("C:/Users/XXXXX ")
# load(.RData)

# Pacotes necessários
require(foreign)
require(ada) # instalar na maquina...

# Carrega o banco de dados (amostra)
dados = as.data.frame(read.spss("amostra.sav"))

# Acerta os dados
dados$BOMMAU = as.factor(dados$BOMMAU)

# Olhando os dados
View(dados)

# Gera aleatoriamente os dados que serão treinados
# No help da função o exemplo faz algo parecido
# help(ada)
treino <- sample(1:dim(dados)[1], floor(.5*dim(dados)[1]), FALSE)
teste <- setdiff(1:dim(dados)[1], treino)

# Boosting
boostres <- ada(BOMMAU~., data = dados[treino,], iter = 80, loss = "logistic",
               type = "discrete", bag.shift=T)

# Primeiros Sumários
print(boostres)

# Adiciona teste
boostres <- addtest(boostres, test.x=dados[teste,-1], test.y=dados[teste,1])

# Mais sumários
print(boostres)

# O pairs não funciona se tiver variáveis categóricas...
# pairs(boostres, train.data=dados[,-1], gap=0)
summary(boostres)
varplot(boostres)
```



```
plot(boostres,T,T)  
  
# Sumário dos objetos do modelo  
boostres$confusion  
mosaicplot(boostres$confusion, col=c("blue", "red"))
```

APÊNDICE B: Algoritmo da Análise Discriminante

```

*****
* Assunto: Análise Discriminante – SPSS
*****
*
DISCRIMINANT
*
***** variável dependente *****
/GROUPS=BOMMAU(0 1)
*
*** ** variáveis independentes ****
/VARIABLES=IDADE QTDECONTRATO QTDEPARCELAS ABERTURACONTRATO
MARGEMLÍQUIDA TEMPORESIDENCIA PARCELASOBRE VENDAS
SUCURSAL SITIMOVEL ESTADOCIVIL
*
/ANALYSIS ALL
/SAVE=CLASS SCORES PROBS
/PRIORS EQUAL
/STATISTICS=MEAN STDDEV UNIVF BOXM COEFF RAW TABLE
/CLASSIFY=NONMISSING POOLED.

```

APÊNDICE C: Algoritmo da Regressão Logística

```

*****
* Assunto: Regressão Logística – SPSS
*****
*
LOGISTIC REGRESSION VARIABLES BOMMAU
/METHOD=ENTER IDADE SEXONUM QTDECONTRATO QTDEPARCELAS
ABERTURACONTRATO MARGEMLÍQUIDA TEMPORESIDENCIA
PARCELASOBRE VENDAS
SUCURSAL SITIMOVEL ESTADOCIVIL
*
***** variável categórica *****
/CONTRAST (SEXONUM)=Indicator(1)
*
/SAVE=PRED COOK LEVER RESID PGROUP
/CLASSPLOT
/CRITERIA=PIN(0) POUT(0) ITERATE(20) CUT(0.5).

```

APÊNDICE D: Algoritmo de Geração de Amostras com Repetição

```

*****
* Assunto: Gera amostras de 1000 em 1000 registro para o bagging – SPSS
* e processo a Análise Discriminante
*****
*
CSPLAN SAMPLE
  /PLAN FILE='C:\XXXX\sel_amostra.csplan'
  /PLANVARS SAMPLEWEIGHT=SampleWeight_Final_
  /PRINT PLAN
  /DESIGN
*
*** método de seleção dos registro para a amostra*****
  /METHOD TYPE=SAMPLE_WR ESTIMATION=DEFAULT
*
*** quantidade de registros da amostra *****
  /SIZE VALUE=1000
*
  /STAGEVARS INCLPROB(InclusionProbability_1_)
CUMWEIGHT(SampleWeightCumulative_1_)
  SAMPSIZE(SampleSize_1_) INDEX(Index_1_).
*
*** nome do arquivo a ser gerado *****
DATASET DECLARE AMOSTRAx.sav.
*
CSSELECT
  /PLAN FILE='C:\XXXX\sel_amostra.csplan'
  /CRITERIA STAGES=1 SEED=RANDOM
  /CLASSMISSING EXCLUDE
  /DATA RENAMEVARS
  /SAMPLEFILE OUTFILE='AMOSTRAx.sav'
  /PRINT SELECTION.
DATASET ACTIVATE AMOSTRAx.sav.
*
*** incluir programa de análise discriminante ou regressão logística *****
*
*** fecha arquivo*****
dataset close AMOSTRAx.sav.

```

APÊNDICE E: Algoritmo COMPLEX SAMPLE – SPSS

```

***** script sel_amostra.csplan *****
***** gerado pela opção de COMPLEX SAMPLE do SPSS *****
*
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<SPSSComplexSamples version="1.0">
<Header copyright="Copyright (c) SPSS Inc., 2006. All Rights Reserved."/>
<SampleDesign numberOfStages="1" wrEstimation="no">
<SampleStage selectionMethod="simple_wr" stageNumber="1">
<InputSampleSize>
<SingleValue value="1000"/>
</InputSampleSize>
<IncProb>
<Variable name="InclusionProbability_1_"/>
</IncProb>
<Index>
<Variable name="Index_1_"/>
</Index>
<OutputSampleSize>
<Variable name="SampleSize_1_"/>
</OutputSampleSize>
<CumulativeWeight>
<Variable name="SampleWeightCumulative_1_"/>
</CumulativeWeight>
</SampleStage>
<Weight>
<Variable name="SampleWeight_Final_"/>
</Weight>
</SampleDesign>
<AnalysisDesign SRSEstimator="wor" numberOfStages="1">
<AnalysisStage estimationMethod="wr" stageNumber="1">
<Weight>
<Variable name="SampleWeight_Final_"/>
</Weight>
</AnalysisDesign>
</SPSSComplexSamples>

```