

FUNDAÇÃO ESCOLA DE COMÉRCIO ÁLVARES PENTEADO

FECAP

MESTRADO PROFISSIONAL EM ADMINISTRAÇÃO

MARCOS SANTOS OLIVEIRA

PROPOSTA DE UM MODELO DE CREDIT SCORING PARA

UMA CARTEIRA DE CRÉDITO CONSIGNADO VISANDO

AÇÕES DE CROSS-SELL

São Paulo

2016

MARCOS SANTOS OLIVEIRA

**PROPOSTA DE UM MODELO DE CREDIT SCORING PARA UMA
CARTEIRA DE CRÉDITO CONSIGNADO VISANDO AÇÕES DE
CROSS-SELL**

Dissertação apresentada à Fundação Escola de
Comércio Álvares – FECAP, como requisito para
obtenção do título de Mestre em Administração.

Orientador: Prof. Dr. Eduardo Pozzi Lucchesi

São Paulo

2016

FUNDAÇÃO ESCOLA DE COMÉRCIO ÁLVARES PENTEADO – FECAP

Reitor: Prof. Dr. Edison Simoni da Silva

Pró-reitor de Graduação: Prof. Dr. Ronaldo Frois de Carvalho

Pró-reitor de Pós-graduação: Prof. Dr. Edison Simoni da Silva

Diretor da Pós-Graduação Lato Sensu: Prof. Alexandre Garcia

Coordenador de Mestrado em Ciências Contábeis: Prof. Dr. Cláudio Parisi

Coordenador do Mestrado Profissional em Administração: Prof. Dr. Heber Pessoa da Silveira

FICHA CATALOGRÁFICA

O48p

Oliveira, Marcos Santos

Proposta de um modelo de Credit Scoring para uma carteira de crédito consignado visando ações de Cross-Sell / Marcos Santos Oliveira. - - São Paulo, 2016.

73 f.

Orientador: Prof. Dr. Eduardo Pozzi Lucchesi.

Dissertação (mestrado) – Fundação Escola de Comércio Álvares Penteado - FECAP - Mestrado Profissional em Administração.

1. Administração de crédito. 2. Sistemas de avaliação de risco de crédito (Finanças). 3. SPSS (Programa de computador).

CDD 658.88

MARCOS SANTOS OLIVEIRA

**PROPOSTA DE UM MODELO DE *CREDIT SCORING* PARA UMA CARTEIRA DE
CRÉDITO CONSIGNADO VISANDO AÇÕES DE *CROSS-SELL***

Dissertação apresentada à Fundação Escola de Comércio Álvares Penteado - FECAP, como requisito para a obtenção do título de Mestre Profissional em Administração.

COMISSÃO JULGADORA

**Professor Dr. João Carlos Douat
Fundação Getúlio Vargas – (FGV)**

**Professora Dr. Joelson Oliveira Sampaio
Fundação Escola de Comércio Álvares Penteado – FECAP**

**Professor Dr. Eduardo Pozzi Lucchesi
Fundação Escola de Comércio Álvares Penteado – FECAP
Professor Orientador – Presidente da Banca Examinadora**

São Paulo, 28 de setembro de 2016.

Agradecimentos

Agradeço a minha família por todo apoio. À minha mãe Luiza e aos meus irmãos Lenice, Liene, Luciano e Diógenes.

Aos meus amigos Job, Claudemir, Clarckson, São Rodrigo, Roberto Reis e Cristiane.

Aos professores da FECAP e em especial ao meu orientador Eduardo Pozzi.

A todos os meus colegas do curso André, Arthur, Estevão, José Vasques, Julian, Luiz, Marcelo, Marcos Silva, Maschio e Junior.

RESUMO

Este trabalho tem o objetivo de analisar a eficiência do modelo de *credit scoring* na ação de *cross-selling* para proporcionar uma maior rentabilidade alinhada ao risco do novo produto. A realização deste estudo se diferencia dos demais por utilizar uma base de dados com clientes que realizaram empréstimo Consignado, a partir da modelagem convencional de um *Credit Scoring* ofertar outro produto, o Cartão de Crédito que exige um melhor perfil para cumprimento dos pagamentos. O estudo resultou em 3 cenários de rentabilidade e desempenho. No Cenário 1 sem uso do escoreamento apresentou rentabilidade de R\$ 0,5 milhões e inadimplência de 16,1%. Nos demais cenários com uso de escores as rentabilidades ultrapassaram R\$ 2,3 milhões e inadimplências abaixo de 9%. Os Cenários 2 e 3 apenas com escore de empresas Bureau. O Cenário 4 inclui o modelo *Crédit Scoring* desenvolvido neste trabalho, apresentou a melhor discriminação entre clientes bons e maus e a maior taxa de aprovação, sendo 75% contra 64% do melhor Bureau. Para isso, utilizou-se de dados fornecido por uma instituição financeira. Utilizando o SPSS e técnicas estatísticas, a análise de Risco Relativo, construção de *dummies* e a análise de correlação de Spearman, foi gerado o modelo de Regressão Logística Binária, validado com o teste Kolmogorov-Smirnov, a Curva ROC e outros. O modelo de *Credit Scoring* desenvolvido apresentou resultados satisfatórios quanto a seu poder de classificação dos clientes. A eficácia da Regressão Logística, como ferramenta de predição de performance de crédito, habilita a aplicação da utilização do modelo *Credit Scoring* pela instituição financeira provedora dos dados para melhorar a rentabilidade e a inadimplência da carteira de clientes com Cartão de Crédito oriundo da carteira de clientes do empréstimo Consignado.

Palavras-Chave: *Credit Scoring*. *Cross-Selling*. Risco de Crédito. Rentabilidade.

ABSTRACT

This work has the objective to analyze the efficiency of the credit scoring model in cross-selling action to provide greater profitability aligned with the risk of new product. This study differs from others by using a database of clients who Payroll-linked loan from the conventional modeling of a Credit Scoring offer another product, the credit card that requires a better profile for meeting payments. The study resulted in 3 of profitability and performance scenarios. In Scenario 1 without use of shoring showed profitability of R\$ 0.5 million and delinquencies of 16.1%. In the others scenarios with the use of the yields scores exceeded R\$ 2.3 million and delinquencies below 9%. Scenarios 2 and 3 with just score Bureau companies. Scenario 4 includes Credit scoring model developed in this work, we showed the best discrimination between good and bad customers and the highest rate of approval, 75% against 64% of the best Bureau. For this, we used data provided by a financial institution. Using SPSS and statistical techniques, the risk analysis Relative, construction of dummies and Spearman correlation analysis, generated the model Logistic Regression Binary, validated with the Kolmogorov-Smirnov test, the ROC curve and others. The model developed credit scoring showed good results as to their power of customer classification. The effectiveness of Logistic Regression as credit performance prediction tool enables the application of the use of credit scoring model by the financial institution provider of data to improve profitability and default of the customer portfolio by credit card coming from the customer base of payroll loan.

Keywords: Credit Scoring. Cross-Selling. Credit Risk. Profitability.

LISTA DE GRÁFICOS

GRÁFICO 1 - Relação Crédito por PIB Brasil	12
GRÁFICO 2 - Inadimplência por tipo de crédito no Brasil	16
GRÁFICO 3 - Evolução Mensal da Relação Crédito por PIB Brasil	21
GRÁFICO 4 - Distribuição em Modelo de <i>Credit Scoring</i>	28
GRÁFICO 5 - Saldo de Crédito Pessoa Física no Brasil	36
GRÁFICO 6 - Taxa média de juros mensal do Crédito Pessoa Física no Brasil	38
GRÁFICO 7 - Inadimplência Cartão de Crédito	54
GRÁFICO 8 - Representação da Curva ROC	58
GRÁFICO 9 - Representação do teste KS2	58
GRÁFICO 10 - Representação do teste KS1	60

LISTA DE QUADROS

QUADRO 1 - Segmentação da Base de dados do estudo	52
QUADRO 2 - Variáveis da Amostra	53
QUADRO 3 - Precificação de consultoria para construção do modelo	55
QUADRO 4 - Variáveis independentes aproveitadas	56

LISTA DE FIGURAS

FIGURA 1 - Etapas para desenvolvimento do <i>Credit Scoring</i>	28
FIGURA 2 - Tipos de crédito ao consumidor	39
FIGURA 3 - Exemplo de Análise do Teste Wald	48
FIGURA 4 - Curva ROC	50
FIGURA 5 - Passos para desenvolvimento do estudo	52
FIGURA 6 - Metodologia de definição do valor de cada cliente	54

LISTA DE TABELAS

TABELA 1 - Interpretação do RR	46
TABELA 2 - Exemplo de Resultado do Teste de Hosmer Lemeshow	48
TABELA 3 - Discriminação da área da Curva ROC	49
TABELA 4 - Discriminação do teste KS2 para Credit Scoring	51
TABELA 5 - Equação Gerada	57
TABELA 6 - Tabela de Classificação	57
TABELA 7 - Resultado do Teste de Hosmer-Lemeshow	59
TABELA 8 - Testes Bases para Validação	59
TABELA 9 - Inadimplência e Rentabilidade por Faixa	61
TABELA 10 - Decisão por Custo, Retorno e Performance	62

LISTA DE ABREVIATURAS E SIGLAS

AUROC: *Area Under Receiver Operating Characteristic*

BCB: Banco Central do Brasil

CART: Classification and Regression Tree

DIEESE: Departamento Intersindical de Estatística e Estudos Socioeconômicos

IF: Instituição Financeira

KS: Teste Kolmogorov Smirnov

PIB: Produto Interno Bruto

ROC: *Receiver Operating Characteristic*

RR: Risco Relativo

SPSS: *Statistical Package for the Social Sciences*

SUMÁRIO

1 INTRODUÇÃO	11
1.1 SITUAÇÃO PROBLEMA	12
1.2 OBJETIVOS	15
1.3 JUSTIFICATIVA	15
2 FUNDAMENTAÇÃO TEÓRICA	18
2.1 RISCO DE CRÉDITO	20
2.2 CREDIT SCORING	25
2.3 PRODUÇÕES INTERNACIONAIS	28
2.4 PUBLICAÇÕES NACIONAIS	31
2.5 OS PRODUTOS: CONSIGNADO E CARTÃO DE CRÉDITO	34
2.5.1 O CRÉDITO CONSIGNADO	35
2.5.2 O CARTÃO DE CRÉDITO	37
2.6 <i>CROSS-SELLING</i> E VIABILIDADE FINANCEIRA	39
3 METODOLOGIA	41
3.1 REGRESSÃO LOGÍSTICA BINÁRIA	42
3.2 VARIÁVEL <i>DUMMY</i>	43
3.3 RISCO RELATIVO (RR)	44
3.4 CORRELAÇÃO DE SPEARMAN	45
3.5 TESTE WALD E TESTE DE HOSMER LEMESHOW	46
3.6 CURVA <i>RECEIVER OPERATING CHARACTERISTIC</i> (ROC)	47
3.7 TESTE KOLMOGOROV-SMIRNOV (KS)	49
4 ANÁLISE DOS RESULTADOS	50
4.1 A AMOSTRA E DESCRIÇÃO DOS DADOS	50
4.2 ANÁLISE DOS DADOS E PROCEDIMENTOS	53
4.3 RESULTADO DO MODELO	58
5 CONCLUSÕES	62
REFERÊNCIAS	64
APÊNDICE A – Análise RR e Geração das <i>Dummies</i>	69
APÊNDICE B – Matriz de Correlação de Spearman	72
APÊNDICE C – Dados da Modelagem	73

1 INTRODUÇÃO

O mercado do crédito exerce papel fundamental no desenvolvimento e crescimento financeiro dos países no mundo. Proporcionar o avanço da concessão de crédito de forma responsável, sustentável e rentável é o que as instituições financeiras buscam para manter-se dentro desse mercado competitivo e altamente regulamentado, sendo esse o grande foco dos órgãos reguladores nacionais para evitar que ingerências e concessões indevidas prejudiquem toda a engrenagem que compõe o sistema financeiro e, por consequência, toda a sociedade.

No contexto econômico e social, muitos tipos ou linhas de crédito surgiram e foram disponibilizadas para consumidores em geral, além das empresas. No Brasil podemos destacar o crédito consignado em salário ou em folha de pagamento; os cartões magnéticos ou, agora, com *chip*, que são popularmente conhecidos como cartões de crédito; a concessão para aquisição de imóveis; para financiamento de veículos, de máquinas e equipamentos; os empréstimos rurais e tantos outros.

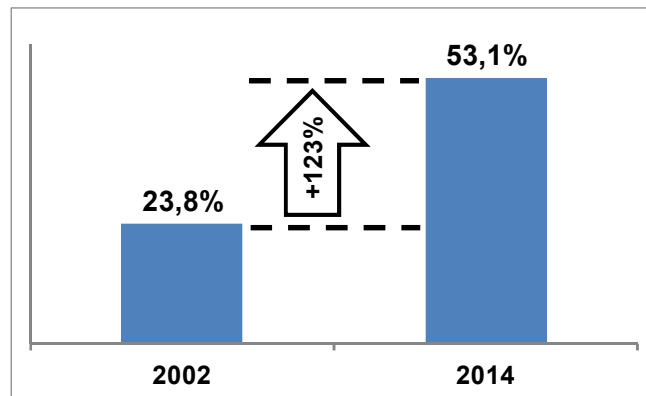
De acordo com Ribeiro et al. (2012, p. 27) tem-se que

A consolidação do ambiente macroeconômico mais favorável no Brasil, oportunizado também pelo crescimento da economia mundial, possibilitou uma diminuição importante nas taxas de juros no país e consolidou fatores condicionantes para a sustentabilidade do nosso crescimento econômico nos últimos cinco anos.

Segundo DIEESE (2014) a concessão do crédito é muito importante no fomento da economia, porque contribui para o investimento nos setores produtivos e para financiar o consumo das famílias. Isso ajuda a entender que existe uma contribuição do aumento do crédito e o crescimento econômico do Brasil no início do século 21.

O Gráfico 1 mostra a relação do crédito por PIB no Brasil nos anos de 2012 e 2014, elaborado com dados de séries temporais disponíveis no Banco Central do Brasil (BCB). Observa-se um crescimento muito acentuado entre esses dois momentos, que corresponde ao aumento de aproximadamente 123% no range de 12 anos.

GRÁFICO 1 - RELAÇÃO CRÉDITO POR PIB BRASIL



Fonte: Adaptado de Banco Central do Brasil (2016).

Mas para acompanhar essa demanda se faz necessária à procura por ferramentas que auxiliem na liberação dos créditos pelas instituições financeiras, que utilizem de aplicativos computacionais baseados em métodos estatísticos, para altos ou baixos valores, alteração de limites, recusas ou cancelamentos de disponibilidades e até ofertas de outros produtos. Tudo no intuito de promover uma maximização dos resultados do negócio.

Qualquer instituição ou credor vai querer coletar informações sobre o mutuário, independentemente de ser uma empresa, um indivíduo, ou de uma unidade governamental. Quando é considerada a possibilidade de emprestar há o interesse em informações sobre a posição financeira atual do mutuário; perspectivas financeiras; e "histórico", no que diz respeito às obrigações financeiras passadas (WHITE, 2013).

Em um estudo realizado por Chang et al. (2014) em bancos na China indicou que os mesmos possuem informações capazes de desempenhar um papel importante para liberar crédito e que o efeito da informação varia em relação a força do relacionamento bancário e da qualidade da informação armazenada.

1.1 SITUAÇÃO PROBLEMA

Utilizar os dados de clientes para auxiliar em decisões gerenciais, classificando-os através de métodos estatísticos e definindo o risco financeiro já é uma prática bastante utilizada nas instituições que concedem crédito e em portfólios ou produtos que possuem uma probabilidade alta de inadimplência. A possibilidade

de discriminar dados relevantes e com utilidade para auxiliar na tomada de Decisão é uma realidade (HORTA; ALVES; CARVALHO, 2014).

A boa utilização do crédito por parte dos clientes favorece a um satisfatório retorno financeiro e a utilização de instrumentos que visem resguardar a empresa contra o não recebimento dos valores disponibilizados na carteira de crédito, identificando e a estimando perdas potenciais com clientes, pode contribuir para uma mensuração mais real dos ativos das instituições.

Com o crescimento de concessões de crédito requer que a atenção para a manutenção de uma boa carteira de clientes seja ponto crucial para estruturação sistemática de grandes volumes de dados com o interesse de contribuir para fortalecimento do relacionamento individual (BOTELHO; TOSTES, 2010).

O *Credit Scoring* é uma ferramenta muito utilizada nas grandes instituições de crédito (bancos, financeiras, cartões de crédito, etc.) do mundo para dar suporte à captação de novos, manutenção dos existentes e recuperação de clientes, principalmente os melhores. Com essa ferramenta, podem ser definidas diversas ações de acordo com o risco de inadimplência.

É crucial ter conhecimento do perfil de risco do cliente para ofertar outros produtos de crédito. Segundo Berry e Linoff (2004), o *cross-selling* ou venda-cruzada de produtos consiste no levantamento de informações suficiente para definir o perfil do cliente com condições a aceitar outros itens que possa satisfazê-lo. O *Credit Scoring* também é aplicado na ação de seleção de clientes com bom perfil de risco para ofertar outros produtos do portfólio.

De acordo com Gouvêa, Gonçalves e Mantovani (2013, p. 97) tem-se que:

Os modelos de análise para concessão de crédito conhecidos como modelos de *credit scoring* baseiam-se em dados históricos da base de clientes existentes para avaliar se um futuro cliente terá mais chances de ser bom ou mau pagador. Os modelos de *credit scoring* são implantados nos sistemas das instituições, permitindo que a avaliação de crédito seja on-line.

Trata-se de um modelo desenvolvido para identificar a probabilidade dos clientes quitarem as suas faturas (ou parcelas) a vencer ou de determinar sua conduta de comportamento, se é boa ou ruim. Assim, sabendo a probabilidade do cliente entrar ou não em inadimplência, pode-se decidir com objetividade para quais clientes deve ser dada ou renovada a concessão do crédito.

São modelos estatísticos de previsão baseados em informações dos clientes, no seu histórico na instituição e no mercado. Para construir o modelo, é imprescindível que exista histórico de comportamento dos clientes disponível, e que estes tenham apresentado uma conduta de crédito (boa ou má) a partir de um conjunto de características disponíveis, sejam elas pessoais, profissionais, de pagamento ou da própria operação de crédito.

Um *Credit Scoring* tem por objetivo estimar a chance que cada cliente avaliado tem de corresponder a um melhor resultado, em termos de retorno positivo à instituição que concede crédito. Conhecendo esta previsão, é possível se implantar políticas de concessão diferenciadas. É importante lembrar que quanto maior o score, menor o risco, pois, por convenção, quanto melhor o atributo mais ganha-se pontos e, do contrário, perde-se (MANFIO, 2007).

Convém ressaltar que modelos não possuem uma precisão absoluta, isto é, ter um poder de previsão com assertividade total. Porém, a promoção do interesse em melhorar a acurácia do potencial preditivo acarretará na geração de rentabilidade (GOUVÊA; GONÇALVES; MANTOVANI, 2013).

Existem várias técnicas estatísticas e computacionais que podem contribuir no desenvolvimento de modelos discriminantes de comportamento. Os mais conhecidos são Árvores de Decisão, Análise Discriminante, Redes Neurais, Algoritmos Genéticos, Análise de Clusters e, o foco de nosso estudo, a Regressão Logística Binária.

A Regressão Logística é muito utilizada e consiste na aplicação de uma única variável dependente, sendo a dicotômica ou binária a mais usual no mercado financeiro, de duas ou mais variáveis independentes (explanatórias) e retorna uma equação que atribui pesos (pontos) às características dos clientes. O somatório dos pesos chama-se score total.

Cada score (ou classes de scores) corresponde a uma probabilidade de o cliente ter uma determinada conduta para o período de previsão. De acordo com Gouvêa, Gonçalves e Mantovani (2013) afirmam que os resultados gerados no estudo por eles desenvolvido com o uso da técnica Regressão Logística corrobora com outros pesquisas anteriores, o que valida a seguridade na aplicação desses modelos na mitigação do risco de crédito.

Diante do exposto, neste trabalho, procura-se responder a seguinte pergunta: Uma ação de *cross-selling* a partir de uma carteira de clientes consignados, com suporte de um modelo de *credit scoring*, contribui para mitigar de forma mais eficiente o risco de um novo produto e promover maior rentabilidade?

1.1. OBJETIVOS

O objetivo deste trabalho é o de analisar a eficiência do modelo de *credit scoring* na ação de *cross-selling* para proporcionar uma maior rentabilidade alinhada ao risco do novo produto.

Objetivos Específicos:

- a) Propor um modelo estatístico de previsão e de cálculo de probabilidade da inadimplência;
- b) Verificar se um *Credit Scoring* discrimina uma carteira de empréstimo pessoal consignado;
- c) Analisar a viabilidade financeira da oferta de um novo produto a partir da carteira de consignado.

1.2 JUSTIFICATIVA

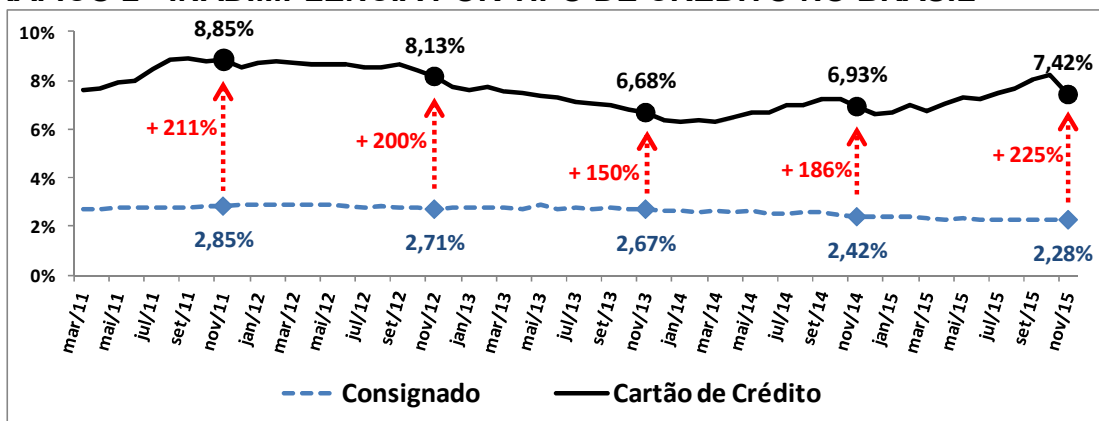
O mercado do crédito no Brasil vem crescendo em um aspecto acelerado, tendo uma participação maior no Produto Interno Bruto e voltando bastante a atenção à qualidade da disponibilidade dos financiamentos para população e dos serviços prestados pelos bancos e instituições, seja através dos meios de pagamento como cartão de crédito, limite de cheque especial, empréstimo pessoal ou consignado e tantos outros; ou, seja através das regulamentações definidas por entidades responsáveis pelo bom andamento desse mercado.

Abordar temas que auxiliam na tomada de decisão na dinâmica financeira favorece a busca por rentabilidade através de técnicas apoiadas por metodologias científica e estatística, para aplicar em uma carteira de clientes de forma massiva, público caracterizado como Pessoa Física pela Receita Federal do Brasil. Uma instituição atuante nesse mercado tem que a mitigação do risco em uma abordagem

massiva é essencial para que a utilização de uma agilidade e perfeição sejam elementos do sucesso da organização.

Com isso, torna-se interessante o desenvolvimento de um método de seleção de clientes que possuem o produto de crédito consignado para ofertar o Cartão de Crédito, produto com maior rentabilidade, mas que possuem maior risco creditício. A iniciativa da aplicação dos modelos de *credit scoring* é baseada no histórico de pagamentos das pessoas para estimar a probabilidade de ser adimplente ou inadimplente (GOUVÊA; GONÇALVES; MANTOVANI, 2013).

GRÁFICO 2 - INADIMPLÊNCIA POR TIPO DE CRÉDITO NO BRASIL



Fonte: Adaptado de Banco Central do Brasil (2016).

O Gráfico 2 mostra uma evolução temporal da inadimplência registrada no Brasil para as linhas de financiamento de crédito por Consignado e por Cartão. Tem uma evidência que essas linhas de crédito representam perfis de risco muito diferentes, correspondendo à suas características distintas. Em destaque temos os meses de novembro, dos anos de 2011 a 2015, exibindo o quanto alta é a inadimplência do Cartão de Crédito.

A realização deste estudo se diferencia dos demais por utilizar uma base de dados com clientes que realizaram empréstimo Consignado, um produto de baixíssimo risco e com uma maior garantia de recebimento do pagamento devido descontar automaticamente do salário dos clientes. A partir da modelagem convencional de um *Credit Scoring* ofertar outro produto, o Cartão de Crédito, com maior risco e sem garantia, mas que exige um melhor perfil para cumprimento dos pagamentos.

Em continuidade, o estudo alinhará ao desenvolvimento da modelagem da equação do score a potencial viabilidade da sua utilização, quais os ganhos no

desempenho dos pagamentos dos clientes selecionados a partir da decisão de escore e o potencial ganho financeiro e redução de perdas. Destacando que a escassez do uso de técnicas seletivas em uma carteira de clientes com ótimo desempenho em um determinado produto pode comprometer a rentabilidade de outro produto que exige um perfil diferenciado.

O estudo está dividido em cinco capítulos, incluindo esta introdução como o primeiro. No segundo capítulo é apresentada a fundamentação teórica dividida em seis partes: (i) abordagem sobre o risco de crédito, procedimentos e sua evolução normativa; (ii) o que é um *Credit Scoring*, conceitos e aplicações; (iii) as produções internacionais sobre modelagem de escore e mitigação de risco de crédito; (iv) destaque das produções nacionais sobre o tema; (v) os produtos Consignado e Cartão de Crédito; e (vi) discussão sobre a ação de *Cross-selling* e a necessidade da viabilidade financeira. No terceiro capítulo apresenta-se a metodologia, descrição de seus aspectos operacionais e as ferramentas estatísticas utilizadas. No quarto capítulo são apresentadas as estatísticas descritivas, a base de dados utilizada e informações adicionais exploradas, faz-se a discussão das variáveis e os resultados obtidos das análises. No quinto e último capítulo, indica-se o alcance dos objetivos proposto desse estudo e a aplicação prática dos resultados da pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

Vive-se uma era de análise de dados baseada nos avanços tecnológicos alinhados a engenharias, cálculos e informática que remete uma realidade palpável, mas que é denotada como virtual. Trata-se de uma era de mudanças constantes, com aplicações tecnológicas quase que imediatas, mostrando que o dia a dia está envolvido em uma revolução científica, com uso de estatísticas, matemáticas, físicas, dentre outras ciências.

Para análise de crédito existe uma grande preocupação de possuir o levantamento de dados no cadastro do tomador. É através do conjunto de informações disponíveis que se obtém o parecer de liberar ou negar os recursos solicitados. E as fontes adicionais podem estar desde indicadores financeiros existentes em balanços de uma empresa às declarações de imposto de renda ou contracheque, como também informações disponíveis no mercado. (SECURATO, 2012).

É notório o aproveitamento do conteúdo de informações armazenadas nos bancos de dados. Nos mais diversos tipos de entidades é uma linha de pesquisa e área de trabalho que está se desenvolvendo com uma velocidade surpreendente. Melhorias tecnológicas resultaram em acesso mais imediato e mais barato para os dados em todo o mundo em tempo real por parte de investidores nacionais e internacionais (SAUNDERS; CORNETT, 2012).

O uso de dados está presente nas instituições financeiras, responsáveis por concessão de crédito, que estão buscando alinhamento com a utilização cada vez mais frequente de técnicas que possibilitem a boa cessão de fundos e redução de perdas (SAUNDERS, 2000). Que proporcione uma mitigação de risco eficiente para disponibilizar uma rentabilidade que agrade seus acionistas, mantendo o negócio e empregados, além de satisfazer os clientes.

A disponibilidade de recursos financeiras através de instituições ou bancos possui o importante objetivo de proporcionar o impulso da atividade econômica, e a oferta de recursos financeiros para a sociedade, seja para pessoas comuns ou grandes corporações, contribui para realizar os anseios eventuais, permanentes ou, até mesmo, estratégicos (SAUNDERS; CORNETT, 2012).

Sendo assim, volta-se a atenção à finalidade de saber dar bem o crédito e que proporcione ganhos para uma organização, reduzindo potenciais perdas. Para isso, conhecer bem esse cliente que receberá um valor monetário para utilizar do modo que gere sua auto satisfação pessoal é fator primordial e, como também, obrigatório.

Nota-se grande atuação das instituições financeiras na aplicação de modelos quantitativos no apoio, liberação do crédito e no suporte para gestão das carteiras de clientes. A modelagem de risco creditício baseia-se em técnicas para assessorar os gestores com informações, com a intenção de auxiliar na tomada de decisão e no cumprimento das diretrizes estipuladas nas políticas institucionais para concessão (BRITO; ASSAF NETO, 2008).

Se conhecer o cliente passa ser de fundamental importância para aplicação efetiva do conhecimento científico com o intuito de mitigar riscos financeiros. Então cada dado, contendo particularidades de um indivíduo, pode contribuir em uma gama de possibilidade diretiva para tomada de decisão na continuidade rentável de um negócio, através do desenvolvimento de modelos para classificação do risco.

As técnicas de modelagem de risco tem o objetivo de avaliar o perfil de um tomador ou de uma operação de investimento, através da atribuição de pesos que indiquem a expectativa de cumprimento de contrato, seja através da classificação de risco, o *rating*, ou por pontuação ou score, o *scoring*. (BRITO; ASSAF NETO, 2008).

Com o uso apropriado de técnicas pode-se desenvolver esse trabalho de diagnosticar massivamente o perfil de risco para cada proponente ao crédito e para manutenção do mesmo em uma carteira de financiamentos, possibilitando a uma continuidade saudável do engenho creditício. Essas técnicas vão desde estatísticas até artifícios computacionais.

A previsão de inadimplência com o uso da modelagem massiva é baseada em aplicações estatísticas de análise multivariada, como regressão linear, análise discriminante e regressão logística. As redes neurais e os algoritmos genéticos são outras técnicas computacionais que também vem ganhando espaço em modelos de previsão do risco creditício (BRITO; ASSAF NETO, 2008).

Segundo Wang et al. (2012) um método de avaliação considerado bom para mitigar risco de crédito pode ajudar as instituições financeiras a concederem empréstimos a candidatos credíveis, assim, aumentar os lucros; ele também pode negar crédito para os requerentes não credíveis, assim diminuir as perdas. Além de serem utilizadas na melhoria de processos, na busca de eficiência em campanhas publicitárias ou de propaganda.

A utilização de ferramentas estatísticas associada aos avanços da área computacional vem contribuindo fortemente na utilização de modelos de predição. Isso, porque toda decisão do crédito carrega em sua composição o risco, e por consequência, estimar a probabilidade do não recebimento do crédito cedido (VICENTE, 2001).

2.1 RISCO DE CRÉDITO

O mercado de crédito é utilizado pelos agentes econômicos para suprir a necessidade de recursos financeiros de curto e médio prazo. Seja através da viabilização de bens para consumo das pessoas comuns quanto para a possibilidade da oferta de capital de giro das empresas. Esses recursos financeiros ou créditos são supridos pelas instituições financeiras ou intermediários bancários (SILVA, 2008).

Securato (2012) indica que o crédito está no dia a dia dos indivíduos, com o objetivo de proporcionar a venda ou aquisição de produtos ou serviços. A palavra crédito vem do latim *creditum*, que remete a existência de confiança na interação de duas ou mais partes em uma operação.

Para Sicsú (2010) o ato de conceder crédito compõe a decisão diante de incertezas. Em qualquer modalidade de financiamento sempre existe a possibilidade de perda e se possuir as ferramentas necessárias para que se possa estimar essa probabilidade de ocorrência, tem-se uma maior confiabilidade nas decisões tomadas. Por consequência, o crédito se torna mais ajustado ao perfil de risco que se deseja assumir.

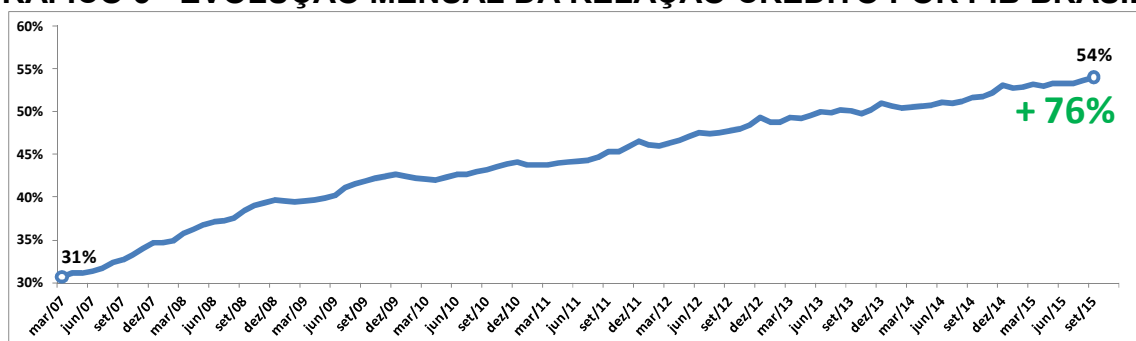
A operação de crédito é um empréstimo de dinheiro para pagamento após um período estabelecido, de acordo com esse período é aplicada uma renumeração,

os juros. Assim, quando uma loja oferta seus produtos à venda, com o pagamento a prazo, também está financiando ou emprestando (SECURATO, 2012).

Segundo Caouette, Altman e Narayanan (1999) existe uma condição de risco quando um indivíduo faz a aquisição de um produto ou contrata uma prestação de serviço e não faz imediatamente o seu pagamento. O Cartão de Crédito é um bom exemplo dessa argumentação, quando a instituição emissora de cartão assume o risco de todos os portadores ao conceder prazo para pagamento dos bens ou serviços adquiridos.

O Gráfico 3 mostra a relação do crédito por PIB no Brasil do período de março de 2007 a setembro de 2015, elaborado com dados de séries temporais disponíveis no Banco Central do Brasil (BCB). Do mesmo modo que o Gráfico 1, é notório que o crescimento do crédito foi muito grande no Brasil nos últimos. Por essa razão que se faz necessário o interesse em melhorar a forma de concessão de empréstimos sempre alinhada ao melhor risco.

GRÁFICO 3 - EVOLUÇÃO MENSAL DA RELAÇÃO CRÉDITO POR PIB BRASIL



Fonte: Adaptado de Banco Central do Brasil (2016).

O Vicente (2001) dá o ponto de vista de que a função do crédito não é exclusividade do sistema financeiro, indicando que na indústria, ou no comércio, ou nos serviços, existe a necessidade de também assumir o papel de personagem que o utilizam como fomentador de vendas, isso, na visão de bancos e financeiras é o principal objetivo desse negócio.

Deve-se considerar que o risco de crédito é parte essencial do processo diário de toda empresa, não apenas das atuantes no sistema financeiro, mas também nas áreas de serviços, comercial e industrial. Claro que cada atividade possui uma atuação com caráter diferenciado, mas com o mesmo foco: aumentar a rentabilidade aproveitando oportunidades.

Existe risco no dia a dia, apesar de ser um conteúdo abstrato, porque há uma percepção na natureza humana das desarmonias entre risco e recompensa muito transparente, por exemplo, um custo já orçado que indica um risco com perda esperada versus um custo imprevisto que indica um risco muito além das perdas esperadas no decorrer da normalidade da vida cotidiana de uma pessoa. (CROUHY; GALAI; MARK, 2007).

Define-se em finanças que o risco é quando se tem dúvida nos resultados futuros ou quando há possibilidade de prejuízo financeiro e em crédito o risco se relaciona a três fatores: aos pagamentos em atraso, a existência de calote e ao aumento da taxa de captação (SECURATO, 2012).

Tudo inicia no ato de conceder crédito, nada mais do que disponibilizar um valor presente embasado na expectativa do cumprimento da promessa de quitação do mesmo em um período posterior, geralmente definido em meses ou ano. Essa esperança em honrar o compromisso em datas acordadas é quando se estabelece o risco de crédito, tratando-se da incapacidade do pagamento de uma operação de concessão e da mitigação do quanto esse risco pode ocorrer.

Empréstimos e financiamentos são atividades principais da maioria dos bancos e, representam volumes significativos nas empresas industriais, comerciais e de serviços. O processo de concessão requer que se façam julgamentos sobre as pessoas ou organizações às quais se pretende fornecer o crédito, porém, esses julgamentos nem sempre se fazem corretos, ou se tornam incorretos com o passar do tempo (VICENTE, 2001).

É sempre conveniente ressaltar, conforme Vicente (2001), que o risco gerado é quando há a ocorrência da inadimplência do contratante do crédito, isto é, quando não cumpre o que se assume em um contrato. Isso não apenas ao nos referirmos a empréstimos, mas também a todas as demais possíveis exposições ao risco, como garantias, aceites e investimento de títulos.

Apesar de ser lucrativo, diz Souza (2000), conceder crédito para o consumidor de forma massiva, em grande número e simultaneamente, é considerado pouco atraente por muitas instituições financeiras porque neste ramo existe o fato de que parte da parcela de empréstimo acarretará em prejuízo.

Definir e controlar o risco de modo que as perdas sejam previsíveis, e dentro de limites estabelecidos, são os objetivos de uma gestão estratégica dos riscos de créditos. Administrar esse risco é, portanto, uma das principais responsabilidades da gestão estratégica do crédito ao consumidor (SOUZA, 2000).

Podemos entender que o risco de crédito de um dado proponente, ou consumidor, pode ser determinado com maior ou menor exatidão quando é submetido a procedimentos que avaliam o risco. Estas avaliações podem ser mais ou menos detalhadas e cuidadas de acordo com a atividade em questão e o peso do crédito concedido no total do faturamento da empresa.

O maior nível de risco que uma organização creditícia pode assumir está alinhado com as diretrizes estipuladas por ela para seus produtos em oferta (GOUVÊA; GONÇALVES; MANTOVANI, 2013). Por características próprias o risco está alinhado ao tipo de negócio correspondente e aos seus produtos ofertados, que podem ter condições diferenciadas a cada tipo, denominando produtos de risco (MANFIO, 2007).

Avaliar o risco de um crédito é o pilar que proporcionará o efetivo funcionamento ou, até mesmo, a continuidade de um negócio. Entende-se, assim, que a avaliação equivocada ou mal elaborada pode acarretar em prejuízos, porque irá barrar o que poderia gerar lucros. Com isso, no setor do crédito, têm vantagens as empresas que sabem avaliar melhor por ficarem menos expostas a mitigações errôneas (GOUVÊA; GONÇALVES; MANTOVANI, 2013).

Ainda segundo o Gonçalves, Gouvêa e Mantovani (2013) a avaliação do risco pode ser realizada através do julgamento, caracterizado por atuar de modo mais subjetivo que se fundamenta em uma análise mais qualitativa, e através da classificação do proponente utilizando modelos de avaliação objetivos, fundamentando-se em uma análise mais quantitativa. Porém, na atualidade, os maiores agentes financeiros combinam essas duas formas.

Os principais fatores que compõem uma análise de crédito são representados pelos C's do crédito, os quais levam em consideração o risco de inadimplência para definir a decisão sobre a viabilidade e liberação de recurso, centrando-se na avaliação qualitativa de fatores (ARAÚJO; CARMONA; AMORIN NETO, 2007). Abaixo, tem-se um entendimento do que são os C's do crédito:

- a) Caráter: ter uma boa imagem no mercado de crédito, sendo idôneo e cumpridor dos pagamentos nos vencimento assumidos;
- b) Capacidade: aproveitamento do crédito absorvido, entendendo que se trata de investimento para obter ganhos;
- c) Capital: com está a situação financeira do proponente candidato ao crédito, visualizando se tem condições assumir os pagamentos;
- d) Colateral: Se possui a disponibilidade de bens, mesmo que sejam móveis, ou imóveis, ou até financeiros, para balizar a concessão do Crédito;
- e) Condições: quais impactos de fatores socioeconômicos que podem influenciar no não cumprimento dos pagamentos;
- f) Conglomerado: segundo Securato (2012) o quadro das empresas de um mesmo grupo pode influenciar uma única, sendo negativa ou positiva.

Para Saunders (2000) no processo qualitativo um analista especialista determina pesos a cada fator dos C's de forma subjetiva para chegar a uma decisão, para assim ter uma compreensão suficiente da situação do solicitante do crédito. O conhecimento especializado e o julgamento do analista são cruciais, determinantes e estão alinhados ao negócio. Esse método é bastante indicado para um volume baixo de propostas creditícias a serem analisadas.

Podem-se apresentar dois problemas nos processos que utilizam somente especialistas para definição da concessão de crédito. Um deles é a consistência do julgamento dos fatores comumente importantes na análise de diferentes tipos de proponentes. O outro é a subjetividade na definição dos pesos a serem atribuídos a cada fator e que podem variar de analista para outro e interpretações diferenciadas (SAUNDERS, 2000)

Para Sicsú (2010) o método quantitativo apresenta vantagens por causa da consistência nas decisões, ocorrendo de forma rápida, alinhada e adequada às expectativas de perdas e ganhos da operação, pode ser realizado à distância por causa do apoio das tecnologias, contribui para monitorar e administrar o risco de um portfólio de crédito. Não há muitas publicações no Brasil que mostrem o melhor resultado do *Credit Scoring* diante do tradicional.

Para grande volume de proponentes, ou análise massiva, aplica-se a análise quantitativa. Na análise quantitativa da concessão do crédito tem a avaliação

objetiva que busca a utilização de ferramentas, aplicando metodologias estatísticas com a finalidade de obter resultados baseados em dados matemáticos que capacitem a real potencialidade de um proponente honrar o seu pagamento (SANTOS; FAMÁ, 2007).

2.2 CREDIT SCORING

Conforme já comentado anteriormente, vimos em caráter de introdução ao tema, que o *Credit Scoring* é uma ferramenta proveniente de técnicas estatísticas, matemáticas ou computacionais que geram probabilidades ou pontuações e mitigam o risco, muito utilizada nas grandes instituições de crédito (bancos, financeiras, cooperativas, cartões de crédito, etc.) para dar suporte à concessão de crédito aos clientes.

Para Sicsú (2010) modelos de credit scoring é uma denominação dada pelo mercado para as equações que calculam as pontuações dadas aos clientes aos clientes que solicitam crédito. Essas pontuações ou escores tem o objetivo específico de dimensionar o risco do proponente ao crédito, estruturando a operação na definição de taxas, garantias, prazos e muitas outras características que auxiliam na gestão creditícia.

A modelagem de risco de crédito segundo Caouette, Altman e Narayanan (1999) pode ser compreendidos em duas categorias, os modelos de aprovação de crédito, bastante conhecidos por *Credit* ou *Application Score*, e os modelos de escoragem comportamental, o *Behavior* ou *Collection Score*. Esses modelos são utilizados para tomada de decisão, desde a iniciação ou na continuidade da concessão do crédito.

Para construção desses modelos técnicas de análise multivariada é aplicada para indicar as variáveis que melhor diferencia os proponentes, retornando a probabilidade de risco. As técnicas mais aplicadas são a análise discriminante, *probit*, *logit*, análise de sobrevivência e outros. Aproveitando de informações cadastrais, comportamentais e empresas externas para construção de equações capazes de separar bons e ruins (CAOUILLE; ALTMAN; NARAYANAN, 1999).

O risco de um proponente a crédito é segmentado a partir de componentes avaliados dentro da carteira, compreendendo além do risco de inadimplir, bem como risco de exposição do total de crédito disponibilizado e a incerteza do valor bem ajustado na concessão; e, o risco de recuperação do que foi concedido para se alinhar ao que pode ser recuperado pelo credor quando ocorre de fato a inadimplência (BRITO; ASSAF NETO, 2008).

Segundo Vasconcellos (2004) os modelos de pontuação surgiram no período da Segunda Mundial para auxiliar os jovens analistas, que sem experiência, necessitavam de alguma ferramenta para uma melhor tomada da decisão, pois os experientes estavam em combate. Iniciou-se a aplicação em instituições financeiras, mas sua aplicação ao ganhar credibilidade passou a fazer parte de outros ramos.

Existem denominações diferentes de modelos para pontuações de clientes, isso porque cada um terá finalidades diferenciadas; ou porque utilizará um conjunto de variáveis comum; ou porque o objetivo pretendido indicará os recursos essenciais para alcance do resultado. As denominações mais utilizadas no mercado são o *Credit Score*, *Application Score*, *Risk Score*, *Behavior Score* e o *Collection Score*.

O *Credit* e o *Application Scores* são a denominação utilizada para os modelos estatísticos que possuem como variáveis utilizadas as informações presentes no cadastro do proponente ao crédito ou para operação específica de produto. Esse modelo é direcionado em sua predominância para a tomada de decisão voltada à aceitação de novos clientes (MANFIO, 2007).

O Caprio e Honohan (2001) ressaltam que a discriminação entre os solventes e insolventes é o propósito de avaliação de crédito, mas os modelos estatísticos de pontuação para concessão do crédito que fazem uso de informação pessoal somente, e de outras que não sejam o histórico de crédito, pode ter o efeito indesejável de reforçar socialmente bolsões de desvantagem, quer geográfica, étnica ou outra.

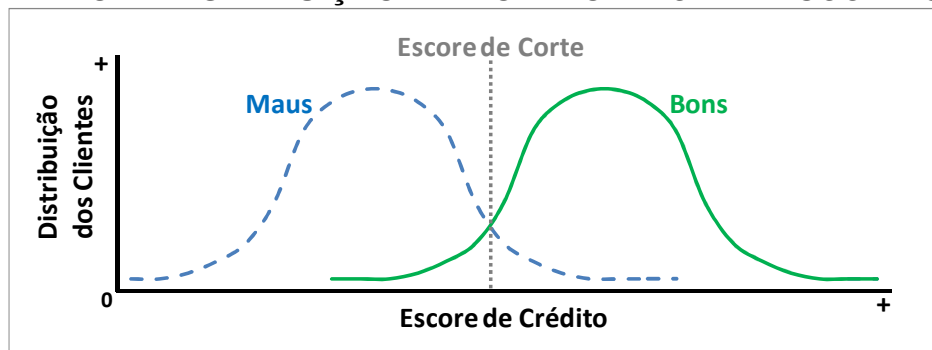
O *Risk Score* é utilizado para subscrição de contratos com risco considerado seguro e são mais voltados para empresas. Segundo Saunders (2000) mitiga o risco de *spread* relacionado em parte ao risco de mercado e baseia-se em calcular a perda esperada e inesperada em vez de mudanças de valor. Para o cálculo dessa perda também se baseia nos estados de adimplência e inadimplência.

O *Behavior Score* tem como grande aplicação à intenção de proporcionar uma maior rentabilidade da carteira de clientes, oferecendo direcionamento para renovação ou cancelamento, aumento ou diminuição de limites de crédito, contribuindo para utilização de encargos diferenciados, como também ao colaborar nas autorizações de compras (SOUZA, 2000).

E o *Collection Score*, segundo Manfio (2007) corresponde aos modelos estatísticos que se baseiam no perfil de clientes que já estão inadimplentes, com variáveis que indicam o comportamento de atraso, sendo bastante aplicado nas estratégias de recuperação de valores e, até mesmo, de potenciais clientes. O *Collection* estima a probabilidade que um cliente pode se dispor a pagar o seu débito.

De acordo com Caouette, Altman e Narayanan (1999) afirma que “a pressuposição nos modelos de *Credit Scoring* é de que exista uma métrica que separe os créditos bons dos maus, dividindo-os em dois grupos distintos”. Conforme apresenta o Gráfico 4, com a distribuição de frequência dos escores dos clientes maus com escores menores que a distribuição dos bons e de acordo com a definição de políticas define-se um escore de rejeição ou corte.

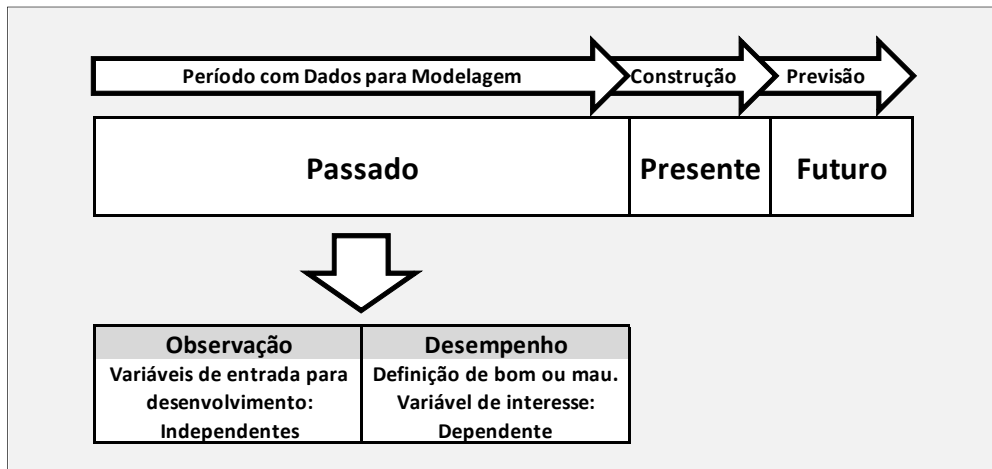
GRÁFICO 4 - DISTRIBUIÇÃO EM MODELO DE CREDIT SCORING



Fonte: Adaptado de Caouette, Altman e Narayanan (1999, p. 182).

Para desenvolvimento e classificação entre bons e maus é necessário um banco de dados, que pode corresponder a uma amostra dos proponentes aprovados é um determinado momento e esperar o seu desempenho em um período específico. Convém ressaltar que cada produto tem um período para desempenho específico e é determinado de acordo com as observação e determinação da instituição financeira (MANFIO, 2007).

FIGURA 1 - ETAPAS PARA DESENVOLVIMENTO DO CREDIT SCORING



Fonte: Adaptado de Diniz e Louzada (2012).

A Figura 1, adaptado de Diniz e Louzada (2012), onde esses autores que para desenvolvimento de modelos existe três importantes etapas. A primeira é o passado, etapa composta pelos dados observados e de desempenho dos créditos liberados. A segunda é o presente que corresponde ao momento atual de construção da equação de score. E a terceira é o futuro, momento onde o cálculo da equação retorna a previsão a ser aplicada.

Além da aplicação dos modelos de *Credit Scoring* nas instituições financeiras e companhias que atuam com o risco de crédito, o mundo acadêmico vem buscando contribuir, tanto no aprimoramento das metodologias existentes e nas indicações de novas técnicas.

2.3 PRODUÇÕES INTERNACIONAIS

O trabalho de Wang et al. (2012) construíram dois modelos de pontuação de crédito com base em árvore de decisão. Um modelo com uma única árvore de decisão que apresentou a menor eficiência entre as demais técnicas comparadas, tais como Regressão Logística e Análise Discriminante Linear. E outro modelo com a combinação de duas árvores de decisão com metodologias diferenciadas e apresentou a melhor eficiência.

Berger, Cowan e Frame (2011) abordam uma série de deficiências na literatura existente na construção de modelos de *Credit Scoring* e desenvolveram uma pesquisa para classificação de crédito em bancos comunitários para emprestar a pequenas empresas. Abrangendo dados sobre as empresas e os proprietários. Esse método para desenvolvimento de modelo não demonstrou eficiência na mudança de qualidade da carteira de crédito quando aplicado.

Os indianos Hens e Tiwari (2012) propuseram uma nova estratégia para reduzir o tempo computacional para análise de crédito utilizando uma técnica computacional denominada *Support vector machine* baseada em uma ferramenta estatística de comparação de dados, o teste F. Compararam o resultado do método por eles desenvolvido com o resultado obtido a partir de outro método bem conhecido e mostraram-se muito competitivo e com menos necessidade de tempo computacional.

Em um estudo desenvolvido por Sabato (2010) analisa os históricos e novos desenvolvimentos de modelos de *Credit Scoring*. Verificou-se que com o Acordo de Capital Basileia II os modelos para pontuação de crédito receberam significativa importância. Instituições Financeiras em todo o mundo estavam dispostas a modificar os seus modelos de risco de crédito interno para estar em conformidade com as regras e as melhores práticas do mercado.

Os ingleses Abdou e Pointon (2011) realizaram um levantamento bibliográfico global de 214 publicações entre livros e artigos que abordavam sobre o *Credit Scoring* em diversas áreas, enfatizando finanças e bancos. Também avaliaram as diferentes técnicas estatísticas e critérios de avaliação de desempenho dos modelos. O estudo revelou que não há em geral uma melhor técnica utilizada na construção e que não existe uma única para todas as circunstâncias.

Em Bruni, Beraldi e Iazzolino (2014) desenvolve um modelo estocástico de análise envoltória de dados para pontuação de crédito que, segundo os autores oferece uma melhora significativa em relação aos métodos estatísticos paramétricos atualmente utilizados pelos agentes de crédito, uma vez que não requer qualquer suposição funcional na relação funcional entre variáveis. O modelo é aplicável à atividade de concessão de empréstimos.

O estudo de Einav, Jenkins e Levin (2013) estudaram a aplicação de um *Credit Scoring* implantado em uma grande empresa de financiamento de veículos. Houve indícios de ter aumentado os lucros identificando dois benefícios distintos de classificação de risco: a capacidade de rastrear os mutuários de alto risco e a capacidade de direcionar mais empréstimos generosos aos mutuários de menor risco.

Kocenda e Vojtek (2011) estudaram a expansão do crédito no setor privado nos mercados emergentes da Europa. Usando dados de um banco da República Checa, construíram dois modelos de risco de crédito, com a regressão logística e outro com árvores de decisão. Ambos os métodos demonstraram eficiência semelhantes e variáveis financeiras e socioeconômicas determinantes para diferenciação do comportamento.

No artigo de Jagric e Jagric (2011), estuda a aplicação do *Credit Scoring* em um banco na Eslovênia, no momento da sua entrada na União Europeia. Também desenvolveram dois tipos de modelos, um baseado em regressão logística padrão e outro baseado no crescimento estruturas celulares, uma espécie de árvore de decisão, que demonstrou uma melhor capacidade de capturar relações não lineares entre as variáveis e melhor relação com as propriedades de variáveis categóricas.

Mendes e Fard (2014) propõem uma aplicação de uma técnica complementar para regressão logística denominada modelo de risco proporcional e direcionada para estudo de confiabilidade. Essa aplicação é adequada para modelagem com dados de previsão temporal que não tenha variáveis independentes categóricas. A aplicação do modelo de regressão logística modificada demonstrou ser inovadora e com potencial utilidade.

Em Agarwal e Sharma (2014) desenvolveram um modelo para prever o risco de fraude por partes dos bancos. O risco de fraude é muito diferente de risco de crédito porque não segue um padrão. O modelo de predição de risco de fraude utilizou a técnica de regressão logística, o score gerado adverte com antecedência a gestão de contas sobre prováveis perdas futuras e os resultados mostram que bom desempenho preditivo para fraudes.

No artigo de Smith et al. (2013) desenvolveram um modelo de escoramento com a pontuação de crédito comercializado pelas três principais agências de crédito

dos Estados Unidos. O estudo com representatividade nacional envolveu 1.000 indivíduos. O resultado do estudo indicou que os dados dos *bureaux* de crédito são suficientemente precisos para facilitar a concessão de empréstimos por instituições financeiras e proporcionar uma eficiência na gestão das contas por parte dos credores.

2.4 PUBLICAÇÕES NACIONAIS

Araujo, Carmona e Amorin Neto (2007) avaliaram a possibilidade de aplicar modelos de *credit scoring* em uma instituição de microcrédito em Recife. Foram desenvolvidos dois modelos com o apoio da técnica estatística regressão logística: um modelo para iniciação do crédito, *Aplicattion Score*, e outro comportamental para manutenção do crédito, o *Behavior Score*. Ambos demonstraram resultados satisfatórios, atingindo 80% de acertos entre dados observados e previstos.

Em Botelho e Tostes (2010) modelaram a probabilidade de abandono de clientes no relacionamento de uma organização. Também foi aplicada a regressão logística em uma amostra de clientes com cartão de crédito de uma grande rede varejista. O modelo gerado demonstrou uma boa adequação, oferecendo material para as empresas compreenderem a taxa de abandono e desenvolverem ações para manutenção do relacionamento com os clientes por períodos mais duradouros.

O Brito e Assaf Neto (2008) avaliaram o risco de crédito em empresas no mercado brasileiro. Através da classificação de uma amostra de empresas com capital aberto classificadas em solventes e insolventes, foi aplicada a regressão logística. Os resultados demonstraram que informações contábeis viabilizam a classificação de risco das empresas e mitigando a probabilidade de solvência dos financiamentos adquiridos por investidores ou fornecedores.

Na modelagem de Gonçalves Gouvêa e Mantovani (2013) exploraram a adequação da técnica de análise de regressão logística para a classificação de clientes de acordo com a adimplência. Com dados de uma instituição financeira compreendendo a variáveis cadastrais, de utilização e restrições de crédito desenvolveu-se um modelo bastante adequado e aplicado ao processo de decisão para concessão de financiamentos.

Para Gouvêa, Gouvêa e Mantovani (2013) a comparação das técnicas regressão logística e redes neurais na promoção da distinção dos bons e maus pagadores em financiamentos bancários resultou em desempenho muito próximo e satisfatório par ambas aplicações técnicas, apresentando taxa de acertos acima de 65 por cento. O modelo de regressão logística demonstrou uma ligeira superioridade, mas não foi bem sustentada em testes estatísticos. As duas técnicas apresentaram ótima capacidade preditiva.

Horta, Alves e Carvalho (2014) a pesquisa foi baseada na amostra de empresas classificadas por empresas especializadas do mercado como solventes ou insolventes no período de 2004 a 2011. Compararam duas abordagens de avaliação de subconjuntos de atributos, o Filtro e o Wrapper. Comparam também modelos feitos pelas técnicas de regressão logística e de redes neurais. E o Filtro foi mais eficiente com as duas técnicas.

Em Ribeiro et al. (2012) calculou a probabilidade de inadimplência em uma instituição de ensino superior privada utilizando da regressão logística. As variáveis explicativas foram provenientes de um questionário socioeconômico, com 206 alunos respondentes aproveitados, 15,53% do total de matriculados. O modelo foi satisfatório, seus resultados e considerações estavam alinhados com a realidade da instituição privada de ensino.

Santos e Famá (2007) propuseram um modelo de *credit scoring* para créditos rotativos dos produtos: cheque especial e cartão de crédito. O banco de dados era composto por variáveis sistêmicas (internas à instituição) e não sistêmicas (externas). O objetivo alcançado era de proporcionar à redução do risco de inadimplência. Indicaram ser indispensável a realizações de avaliações contínuas dos modelos.

Santos e Santos (2009) estimaram a probabilidade de insolvência através do Modelo KMV ou *Credit Monitor*, baseado na hipótese que os preços de ações refletem as expectativas do mercado sobre a empresa, o risco de crédito está associado à variação dos ativos da empresa. O modelo demonstrou ser mais um recurso que pode ser aplicado na gestão de carteias de crédito e na mitigação da exposição ao risco em empresas.

Utilizando-se da análise discriminante e da tabela de *ratings* da agência Standard & Poor's, Ferreira et al. (2013) desenvolveu um estudo para avaliar a desenvoltura das empresas do segmento aeroviário. Aplicou-se o modelo Z Score de Altman. Foi possível determinar a capacidade financeira das empresas e indicaram uma possível descontinuidade das suas atividades para os anos subsequentes aos analisados.

Em Sousa e Figueiredo (2014) utilizando de um algoritmo baseado em árvore de decisão e outro algoritmo neural artificial para desenvolver modelos capazes de analisar a capacidade dos membros de uma cooperativa de crédito em liquidar seus compromissos. Os modelos demonstraram resultados estatisticamente semelhantes e com capacidade para ajudar no processo de tomada de decisão da cooperativa.

As espanholas Medina e Selva (2013) realizaram uma comparação entre as técnicas paramétricas análise discriminante e regressão logística e a não paramétrica árvore de decisão na modelagem de um *credit scoring*. Foi utilizado um banco de dados de várias instituições financeiras espanholas. Neste estudo o algoritmo da árvore de decisão, denominado CART, gerou o melhor resultado.

Bressan et al. (2011) com observações coletadas de uma amostra cooperativas Sicoob-Crediminas, entre 1995 e 2008, desenvolveram modelos com a regressão logística baseado no sistema de monitoramento de cooperativas denominado Pearls (proteção, estrutura financeira, ativos pela sua qualidade, retorno e custos, liquidez e sinais de crescimento). Demonstrou que os indicadores contábeis financeiros do sistema Pearls são relevantes para a avaliação de insolvência das cooperativas de crédito.

Na análise de conglomerados das empresas em Brito, Assaf Neto e Corrar (2009), na primeira etapa do estudo eles geraram uma pontuação baseada na regressão logística para em seguida aplicar a técnica estatística de *cluster* para classificar as empresas por nível de risco, gerando *ratings*. Com essa técnica foi possível monitorar com eficiência as mudanças dos níveis de risco das empresas em período temporais distintos antes da ocorrência de *default*.

No modelo de Camargos, Camargos e Araújo (2012) O modelo *credit scoring* desenvolvido por regressão logística identificou condições relevantes para

as empresas cliente de uma instituição financeira pública do Estado de Minas Gerais inadimplir. Apresentou 88,5% de acertos entre os dados observados e o previsto para as empresas, com boa precisão e qualificação semelhante em modelos das pesquisas consultadas.

Em Vieira, Pires e Galeano (2013) a pesquisa realizada desenvolve um modelo baseado em regressão estatística, utilizando variáveis da equipe de vendas que auxiliam no acompanhamento do desempenho comercial e a realizações de vendas cruzadas, o *cross-selling*. A pesquisa teve dados captados de segmentos comerciais e demonstrou previsão positiva e bom desempenho, indicando que a possibilidade de fazer combinações com a venda cruzada geram resultados.

Pinho (2009) desenvolveu um modelo para *cross-selling* utilizando-se de uma técnica de redes neurais com dados de uma revendedora de motos. O modelo teve como propósito identificar o interesse do cliente em comprar motos. Apresentou resultados satisfatórios e sua aplicação gerencial nos processos internos da empresa gerará uma lista de potenciais clientes interessados em comprar uma moto nova.

Entende-se com a revisão da literatura sobre *Credit Scoring* que dados cadastrais e comportamentais dos clientes favorecem como forte subsídio em uma tomada de decisão, porém a efetiva aplicação desses modelos necessita da utilização de indicadores estatísticos para construção, como também para medir a qualidade e eficiência da pontuação gerada que servirá para classificação dos clientes e auxílio na tomada de decisão.

2.5 OS PRODUTOS: CONSIGNADO E CARTÃO DE CRÉDITO

Para construção de um modelo de *credit scoring* é essencial que se defina o tipo de projeto e seus objetivos que se espera ser atingidos. Neste escopo deve ficar claro qual a operação de crédito que será analisada, em que mercado está atuando, como também público e outras características. Definindo assim o tipo de risco, modalidade, estágio, alcance e mercado (SICSÚ, 2010). O produto relaciona-se a essas condições.

Para Manfio (2007) o planejamento é uma condição para ter uma decisão menos arriscada do que ela pode de fato ser. O ato de planejar elimina a ação do “tentar adivinhar”, utilizando da observação e da valorização do conhecimento. O planejamento do crédito tem o objetivo de alinhar as características dos produtos com os direcionamentos estratégicos e financeiros que são determinados como foco de um negócio.

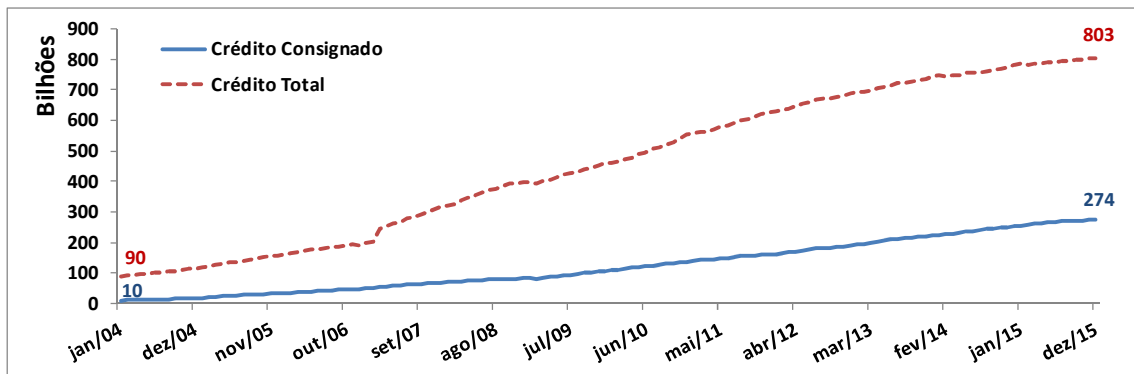
O objetivo deste trabalho é o de desenvolver um modelo de score para oferta o produto cartão de crédito para uma carteira de cliente que possuem o produto crédito consignado. Produtos caracterizados por níveis de risco totalmente distintos por causa dos mecanismos e critérios de concessão muito distintos.

2.5.1 O CRÉDITO CONSIGNADO

Para Coelho, Mello e Funchal (2012) evidências obtidas em seus estudos sugerem que o crédito consignado no Brasil foi um dos principais instrumentos que contribuiu para a criação de um mercado interno forte de consumidores. Eles documentaram que a utilização da renda futura como uma fonte de garantia contribuiu substancialmente em um grande aumento nos empréstimos pessoais no Brasil e fomentou a economia local.

A modalidade de crédito consignado foi introduzida no Brasil por lei a partir de 1990, regulamentando apenas para aposentados e servidores do setor público. A partir de dezembro de 2003, uma nova lei regulamentou a extensão ao uso do crédito consignado para os trabalhadores e aos aposentados ou beneficiários do setor privado. Desta forma os aposentados do setor privado representou a fatia maior de proponentes a ser explorada (COELHO; MELLO; FUNCHAL, 2012).

GRÁFICO 5 - SALDO DE CRÉDITO PESSOA FÍSICA NO BRASIL



Fonte: Adaptado de Banco Central do Brasil (2016).

O Gráfico 5 mostra uma evolução temporal do saldo de crédito no Brasil do período de janeiro de 2004 a dezembro de 2015, elaborado com dados de séries temporais disponíveis no Banco Central do Brasil (BCB). Visualizamos que o crescimento do crédito consignado parte de R\$ 10 bilhões no início da série e finaliza em R\$ 272 bilhões, apresentando um crescimento de 2.272%. O saldo do crédito total inicia em R\$ 90 bilhões e finaliza a série em R\$ 803 bilhões, com crescimento de 795%. O saldo do crédito consignado cresceu 3,4 vezes em relação ao saldo total de crédito no Brasil durante o período.

O crédito consignado refere-se a uma modalidade de empréstimo pessoal que apresenta como característica essencial que os pagamentos do valor principal e dos juros sejam deduzidos diretamente do contracheque do solicitante do crédito, abatendo diretamente do salário antes do acesso do mutuário recebê-lo. Podemos entender de forma sucinta que os empréstimos consignados utilizam a renda futura de um indivíduo em garantia de pagamento (COELHO; MELLO; FUNCHAL, 2012).

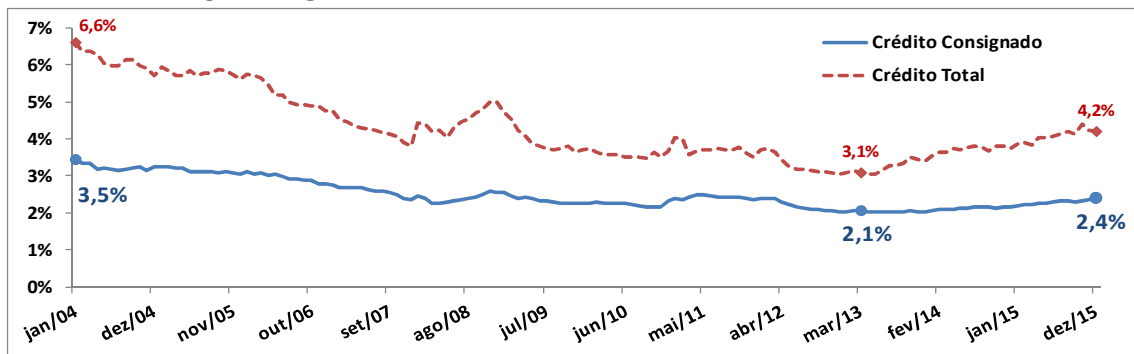
Ainda segundo Coelho, Mello e Funchal (2012) a utilização dessa fonte e forte garantia proporciona a mitigação do problema de risco moral, com relação ao não pagamento, com também reduz as consequências da omissão involuntária, levando a diminuição dos custos de empréstimos para as instituições. Trata-se de uma garantia valiosa para o não cumprimento involuntário devido aos pagamentos realizado automaticamente.

No Gráfico 2, apresentada na Introdução deste trabalho, a inadimplência do crédito consignado não ultrapassa os 3% nos últimos 5 anos. Por isso, para Coelho, Mello e Funchal (2012) um instrumento de garantia simples reflete ser muito valioso, demonstrando um impacto agregado muito positivo. E instrumento é relevante e que

contribuiu para aplicação de baixas taxas de juros. Que contribuiu também para redução das taxas de outras linhas de crédito com garantia, como por exemplo, o financiamento de veículos.

O Gráfico 6 mostra a evolução da taxa de juros mensal do crédito pessoa física no Brasil do período de janeiro de 2004 a dezembro de 2015, elaborado com dados de séries temporais disponíveis no Banco Central do Brasil (BCB). No início da série temos que a taxa média de juros para o crédito consignado em 3,5% a.m. e no fim da série reduz 30,4% apresentado taxa média de 2,4% a.m. de juros. A taxa de juros do crédito total para pessoa física tem um dos menores níveis de redução entre o início da série em janeiro/2004 e março/2013, que cai 53,3%, partindo da taxa média de 6,6% a.m. para 2,1% a.m. respectivamente. Em meados de 2013 a taxa de média de juros no Brasil volta a se elevar devido a desaceleração da economia.

GRÁFICO 6 - TAXA MÉDIA DE JUROS MENSAL DO CRÉDITO PESSOA FÍSICA NO BRASIL



Fonte: Adaptado de Banco Central do Brasil (2016).

Segundo Lopes (2010) as facilidades do crédito através da consignação de salários possibilita condições mais favoráveis de acesso ao crédito, com taxas de juros mais baixas, estimulando o consumo e a qualidade de vida. Por isso, a lei brasileira limitou a 30% o valor da parcela do empréstimo sobre o salário do tomador ligado ao setor privado, caso contrário poderia causar dificuldades financeiras ao comprometer a renda das pessoas.

2.5.2 O CARTÃO DE CRÉDITO

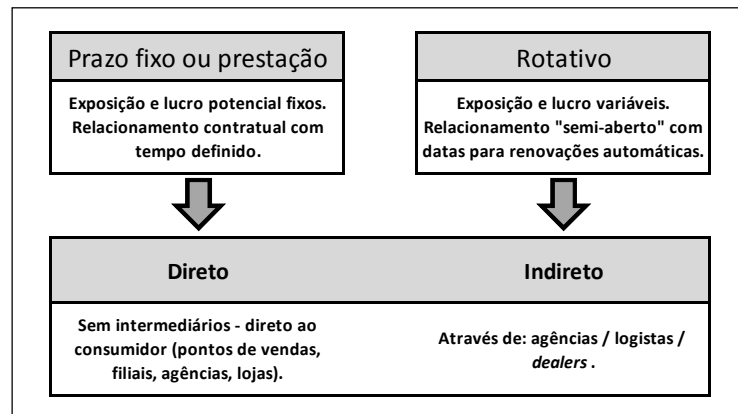
Para Bertaut e Haliassos (2005) os cartões de crédito são um meio de pagamento e de crédito que proporciona a facilidade de transações sem dinheiro,

oferecem a possibilidade de adiar o pagamento de um vencimento atual para o outro futuro e permitem que os consumidores gastem mais mesmo com deficiências temporárias de liquidez. É uma forma bastante dispendiosa de crédito quando o crédito rotativo é utilizado com frequência.

O cartão de crédito opera ligado a empresas denominados bandeiras, como Diners, Visa e MasterCard, como também há os cartões de crédito ligados a instituições financeiras, a lojas varejistas ou rede de supermercados. As compras obtidas através desta modalidade podem proporcionar o pagamento no próximo vencimento, ou parcelado sem juros, ou com juros ou podem ainda viabilizar que valores de um vencimento atual sejam refinanciados para o próximo com a aplicação dos encargos financeiros (SILVA, 2008).

Nos produtos de crédito ao consumidor existem duas opções para cumprir os pagamentos segundo Manfio (2007): Uma é caracterizada pelo pagamento parcelado, com prazo e prestações fixas; Outro é o rotativo que se caracteriza pela possibilidade de rolar a dívida ou parte dela, com a possibilidade de data ou não para a ocorrência de pagamentos. O cartão de crédito atua com as duas opções, conforme é demonstrado na Figura 2.

FIGURA 2 - TIPOS DE CRÉDITO AO CONSUMIDOR



Fonte: Adaptado de Manfio (2007).

Para Bertaut e Haliassos (2005) em sua pesquisa levantaram que no ano de 2001 nos Estados Unidos as ofertas de cartões de crédito eram muito mais comuns de serem provenientes de contas de já clientes da instituição através de suas carteiras de outros produtos. O potencial de novos negócios pode ser potencializado com os clientes ao oferecer novos produtos aproveitados de características individuais e ofertas alinhadas ao perfil (MANFIO, 2007).

A maioria dos portadores de cartão possuem a convicção que os cartões de crédito criam problemas ao auto-controle financeiro pessoal, decorrente pela disponibilidade do crédito fácil que possibilitam um forte probabilidade de gastos excessivos (BERTAUT; HALIASSOS, 2005). Os clientes se descontrolam com cartões de crédito quando não entendem que o limite concedido é um empréstimo. A maioria pensa ser uma renda complementar.

Segundo Silva (2008) para conceder crédito precisa atentar-se a possibilidade das várias formas de financiamento ou empréstimo para utilização dos clientes, a dimensão deste uso indica o nível de risco de crédito que está atribuído. O cartão de crédito é um tipo de produto de alto risco e a sua concessão não baseado em garantias, não é semelhante ao crédito consignado que possui uma garantia futura baseada no salário.

2.6 CROSS-SELLING E VIABILIDADE FINANCEIRA

Até aqui se observa que para viabilizar a gestão de um negócio de crédito massificado, como cartão de crédito, há de se render a ferramentas apropriadas, tecnológicas e seguras para obter o máximo de lucratividade de um negócio de risco. As decisões tomadas buscam a maximização da recompensa, custo e benefício, com a otimização do risco. Por isso é de fundamental importância o cálculo das contas (MANFIO, 2007).

Ormezi e Aksin (2010) comentam que diversas empresas atuantes em serviços financeiros recorrerem para o crescimento aprofundando do relacionamento com os clientes com o interesse de torná-los mais rentáveis. Essa rentabilidade provém de receitas geradas através de produtos e serviços vendidos para clientes existentes. A exploração da própria carteira de clientes é considerada a melhor perspectiva de vendas por ser mais barata que a busca por novos clientes.

Berry e Linoff (2005) citaram diversas empresas que se utilizam do *cross-selling* aproveitando dos dados armazenados com as informações sobre todos os seus clientes. E através das informações constroem modelos de mineração de dados para oferta de outros produtos. Evitando a fuga para concorrentes, o que banco entende por “atrito” e empresas de telefonia chamam de “*churn*”. Qualquer que seja a denominação é um grande problema.

A grande preocupação para oferta de produtos é a identificação da pessoa certa e no momento conveniente a ser explorado em processo de venda. Por isso acredita-se que o *cross-selling* é uma importante ação que tem o objetivo de viabilizar os clientes com maior propensão a adquirir vários produtos de uma mesma empresa, melhorando a retenção de clientes e promovendo mais receita com menor custo (ORMECI; AKSIN, 2010).

Segundo Ormeci e Aksin (2010) é comum segmentar uma carteira de clientes em grupos com base na potencialidade de vendas, com foco exclusivo nas vendas para segmentos de alto potencial, quando não possui um sistema para *cross-selling* automatizado, a área responsáveis pelas vendas utilizará as informações perceptíveis para determinar se um cliente é apropriado para tentar uma venda cruzada.

A percepção em compreender que é possível montar um plano de retenção abordando as informações chave para atingir o público alvo os clientes certos. Em muitos setores, alguns clientes custam mais do que valem, com ações como adquirir um produto e não usar, ou podem ser pessoas que utilizam crédito, gerando o saldo devedor alto e se tornam inadimplentes sem perspectivas de pagamento (BERRY; LINOFF, 2005).

Para Lau, Chow e Liu (2004) o *cross-selling* deve ter uma atuação ilimitada e não se restringir a penas segmentação de ações puramente comerciais, mas deve ser aplicado em uma visão abrangente, atuando além dos conceitos do marketing, como produtos, preço, praça e a promoção. Como também no apoio para maximizar a rentabilidade do negócio, utilizando dos Cs do crédito e mitigando risco necessário para redução de perdas.

O objetivo deste trabalho é propor uma ação de *cross-selling* aplicando o modelo de *credit scoring* a uma carteira de cliente de um produto como o crédito consignado, que ao se utilizar de uma garantia futura, possui baixo risco, juros baixos e baixa inadimplência, e ofertar o cartão de crédito, linha de financiamento caracterizada pelos juros altos, risco elevado e alta inadimplência. Produtos totalmente opostos.

3 METODOLOGIA

Para realizar essa proposta em desenvolver um modelo de *credit score* fez-se uso de uma metodologia baseada em uma pesquisa experimental por se caracterizar em um objeto de estudo determinado, através de seleção de variáveis com capacidade de influenciar, definindo formas para controlar e para observar os resultados produzidos (GIL, 2002).

Para Creswell (2007) os experimentos adicionam dados verdadeiros, com escolha aleatória ou não, para tratamento e análise da informação. De acordo com Gil (2002) o experimento é um procedimento disponível para que pesquisadores possam responder a hipóteses e questionamentos. Oferece controle entre variáveis independentes que causam efeitos nas variáveis dependentes.

Trata-se, também, de uma pesquisa descritiva com a intenção de identificar a possibilidade de aproveitar as variáveis cadastrais de uma carteira de crédito consignado de uma instituição financeira, associadas a outras informações de mercado e internas para mitigar o risco em ofertar outra modalidade de crédito com nível de risco maior, maior rentabilidade financeira e maior alcance de clientes, o cartão de crédito. As pesquisas com caráter descritivo tem a intenção de desvendar as associações entre variáveis (GIL, 2002).

O estudo está classificado quanto a sua abordagem como quantitativo. Conforme Creswell (2007), por ter uma relação baseada na quantificação dos dados, com o interesse de proporcionar uma análise apurada para direcionar a interpretação das informações utilizadas na proposta de construção do *credit scoring*, com a aplicação de técnicas estatísticas.

A técnica estatística que será utilizada neste trabalho é a Regressão Logística (ou Logit) que consiste em investigar a relação entre variáveis explicativas, métricas e não métricas e uma variável dependente categórica binária ou múltipla, que necessita de testes estatísticos para seleção de variáveis e de validação da equação, tais como o Risco Relativo, a correlação de Spearman, o teste de Kolmogorov Smirnov (KS), a Curva ROC e outros.

O desenvolvimento do modelo de *credit scoring* proposto contará com a utilização de uma base de dados cedida por uma instituição financeira, que será

mantida em caráter de confidencialidade. Contará com o suporte do *software* estatístico SPSS versão gratuita para construção da equação, com as respectivas variáveis significativas que retornam os pesos para a geração da pontuação que classifica e diferencia o risco dos clientes.

3.1 REGRESSÃO LOGÍSTICA BINÁRIA

A técnica de Regressão Logística Binária foi desenvolvida em 1960 para investigar a relação entre variáveis explicativas, métricas e não métricas e uma variável dependente categórica binária (FÁVERO et al., 2009). É vista como uma alternativa para análise discriminante, pois é menos exigente no que diz respeito às pressuposições necessárias aos dados para que haja a submissão com êxito.

Sicsú (2010) ressalta que esse método de regressão estima a probabilidade de um indivíduo específico ser parte integrante de um dos dois grupos analisados no estudo. Ainda diz que um dos motivos do uso constante da regressão logística binária é a não linearidade do evento de interesse em relação ao score gerado e que ela prevê de forma direta a probabilidade.

Para Diniz e Louzada (2012) na modelagem de *credit scoring* a variável independente em sua maioria é caracterizada por ser binária, com o interesse de separar bons e maus clientes, por isso que a regressão logística é muito utilizada. Neste método de regressão a estimação dos coeficientes é obtida através da maximização das probabilidades dos conjuntos de dados observados partindo do modelo proposto.

Em particular, as premissas da regressão logística possuem como maior vantagem a possibilidade de não exigir algumas suposições, onde citamos a normalidade dos erros e a igualdade da matriz de covariância, como também possibilitar o entendimento direto dos coeficientes estimados por medida de associação (ALVES, 2008).

De acordo com Vicente (2001, p. 82)

[...] a Regressão Logística, ou *Logit*, é útil para situações nas quais se deseja prever a presença ou ausência de uma característica, ou resultado, baseado em valores de um conjunto de variáveis dependentes. No caso da mensuração do risco de crédito, a *Logit* é utilizada para a avaliação da

probabilidade de insolvência de determinado grupo de clientes, relativos à situação de empréstimo ou financiamento.

O modelo que se tem adotado para o estudo dessas relações é conhecido como função logística, e sua representação algébrica é:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} + \varepsilon \quad (I)$$

onde:

$\pi(x)$ é a variável dependente denotada por y ;

x é a variável explicativa;

e é a base do logaritmo natural;

β_0 e β_1 são os parâmetros a serem estimados; ε é o erro aleatório que se supõe com média zero.

Com transformações algébricas convenientes de (I), obtém-se:

$$\pi(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} \quad (II)$$

Segundo Araújo et al. (2007) dentro do contexto das atividades e definições do risco de crédito determina-se que a variável dependente é composta pelo cliente ser adimplente ou bom e ser inadimplente ou mau, assumindo valores 1 ou 0. Quando nos referimos a empresas também usamos os termos solventes para os bons e insolventes para os maus.

Ainda, conforme Araújo, Carmona e Amorin Neto (2007), as informações referentes a dados demográficos, econômicos e financeiros, sejam de empresas ou de pessoas constituem variáveis independentes e representam os fatores que influencia a ocorrência da inadimplência. Após os tratamentos necessários para viabilizar melhor o manuseio e capacidade de informação, essas variáveis são aplicadas à regressão logística.

3.2 VARIÁVEL DUMMY

De acordo com Fávero et al. (2009) a importância da variável *dummy* é de transformar variáveis qualitativas em numérica, onde uma categoria é evidenciada e posta em referência ao receber o atributo 1 (um) e determinando todas as demais

categorias que compõe a variável recebendo o atributo 0 (zero), representando a ausência de valor.

Com a formação de variáveis *dummies* tem-se a possibilidade de estudar separadamente o poder de influência que cada categoria de uma variável pode possuir em todo o contexto pesquisado. Pode-se denominá-las como variáveis binárias, dicotômicas ou indicadoras. Cada categoria que uma variável possui pode gerar uma *dummy*, tanto para variáveis qualitativas quanto para variáveis quantitativas categorizadas (SICSÚ, 2010).

3.3 RISCO RELATIVO (RR)

Conhecendo-se as variáveis independentes categóricas, ou categorizando-se as que não são, pode-se construir tabelas de frequências das variáveis explanatórias em relação à variável resposta observada (zero ou um). A partir disso é possível, por meio de uma análise bivariada, estudar a participação e influência dos casos em cada observação. Esta participação e influência foram denominadas Risco Relativo (RR).

Segundo Oliveira (2007) o cálculo do RR se dá pela razão entre os percentuais do estudo dos casos de ocorrência do evento na variável dependente dicotômica, que assume valor 1 e que é denominado como “bom cliente”, pela não-ocorrência do evento, representado pelo valor zero e denominado por “mau cliente”, em uma análise bivariada, com uma variável independente e suas categorias.

Em Gouvêa, Gonçalves e Mantovani (2013) afirma que o RR proporciona o entendimento do quanto uma categoria está ligada à ocorrência do evento de interesse na variável dependente ou não, além de colaborar no agrupamento das categorias para evitar a existência das que contém poucas observações e unificar as que possuem o mesmo risco.

Para Sicsú (2010) muitos critérios tem a possibilidade de decidir o agrupamento de categorias das variáveis dependentes e a vantagem de utilizar o RR está em o mesmo ter o entendimento simples e lógico para realizar a fusão de classes. Mas recomendam a utilização se a frequência de cada categoria seja

suficiente e grande. Em pequenas amostras devem ser conduzidas com cuidado porque podem levar a conclusões com baixo nível de confiabilidade.

Na tabela 1 tem-se uma categorização dos resultados do RR. É um resumo para auxiliar no entendimento e significa que o quanto mais próximo de zero muito ruim será a ocorrência do evento de interesse, têm-se mais clientes maus; quando aproximadamente 1 (um) indica que a categoria estudada não discrimina entre ocorrência e a não-ocorrência; mas quanto maior for que 1 (um) a ocorrência terá maior possibilidade de existir, têm-se mais clientes bons.

TABELA 1 - INTERPRETAÇÃO DO RR

Resultado RR	Interpretação
Menor ou igual a 0,5	Muito Ruim
Entre 0,5 e 0,9	Ruim
Entre 0,9 e 1,1	Neutro
Entre 1,1 e 2,0	Bom
Maior que 2,0	Muito Bom

Fonte: Elaborado pelo autor através de Gonçalves, Gouvêa e Mantovani (2013).

Pela análise das frequências relativas e do RR é possível aperfeiçoar a visualização da influência das categorias de cada variável e buscar, assim, uma redução destas categorias aproveitando-se melhor as variáveis qualitativas. Ao analisar a frequência das categorias pelas observações, tomam-se decisões de acordo com a representatividade, evitando-se estimativas pouco discriminantes.

Do mesmo modo, se a categoria tiver uma representatividade muito grande, têm-se efeitos indesejados. Após a verificação da participação da frequência, deve-se analisar o RR cuidadosamente. Categorias muito próximas podem ser agrupadas ou, a depender dos valores de seus índices obtidos, até mesmo serem descartadas.

3.4 CORRELAÇÃO DE SPEARMAN

Para Croux e Dehon (2010) comentam que o estimador de correlação não paramétrico como o de Spearman é amplamente utilizado em ciências aplicadas, considerado robusto no sentido de ser resistente às observações periféricas. A correlação de Pearson é um dos estimadores estatísticos mais frequentemente utilizados. Mas seu valor pode ser seriamente afetado por apenas um *outlier*.

A correlação de Spearman possui uma função de influência limitada, que corrobora com a crença de que esta medida não paramétrica de correlação é robusta a *outliers*. Além de serem robustas, é desejável que o estimador tenha uma alta eficiência estatística (CROUX; DEHON, 2010). Na modelagem de crédito no mercado utiliza-se a correlação de Spearman quando trabalha com a transformação das variáveis independentes em *dummies*.

Ainda Croux e Dehon (2010) indica que para a distribuição normal o estimador de correlação de Pearson é o mais eficaz, mas a eficiência estatística do estimador de correlação de Spearman permanece acima de 70% para todos os valores possíveis da correlação populacional. Portanto, eles fornecem um bom compromisso entre robustez e eficiência. Desta forma, na modelagem de *credit scoring* os índices abaixo de 70% considerado-se com correlação baixa.

3.5 TESTE WALD E TESTE DE HOSMER LEMESHOW

Para Brito e Assaf Neto (2008) o teste Wald é similar ao teste “t” comumente utilizado na modelagem linear e tem como hipótese nula a referência de ter o parâmetro estimado correspondendo a zero. Tem distribuição Qui-Quadrado, através do cálculo com o quadrado da razão do coeficiente pelo seu erro padrão.

Em Camargos, Camargos e Araújo (2012) ressaltam que o teste Wald pode apresentar falhas ao rejeitar coeficientes significativos e aconselham a utilização do teste da razão de verossimilhança. Na Figura 3 temos os valores Wald com p-valores significativos.

FIGURA 3 - EXEMPLO DE ANÁLISE DO TESTE WALD

Variável	Coefficiente	Erro-Padrão	Teste Wald	g.l.	Valor-p
Intercepto	-4,740	2,165	4,793	1	0,029
X12	-4,528	2,414	3,519	1	0,061
X16	18,433	8,297	4,936	1	0,026
X19	-14,080	6,998	4,048	1	0,044
X22	-11,028	6,103	3,265	1	0,071

Fonte: Brito e Assaf Neto (2008, p. 26).

O teste de Hosmer Lemeshow também é calculado automaticamente pelo SPSS. De acordo com Brito e Assaf Neto (2008) e consiste em avaliar a existência de diferenças entre o que foi observado para desenvolver o modelo e classificação

prevista gerada. O grau de boa qualidade do modelo ocorre quando as diferenças não são significativas.

A aplicação do teste Hosmer Lemeshow verifica se a classificação prevista, gerada pelo modelo, de cada grupo são iguais às observadas, por meio de estratificação das observações em faixas e da aplicação de um teste de Qui-quadrado. O interesse é que não rejeite a hipótese nula para que não haja essas diferenças significativas. O bom ajuste do modelo é indicado pela não significância do Qui-quadrado (FÁVERO et al., 2009).

TABELA 2 - EXEMPLO DE RESULTADO DO TESTE DE HOSMER LEMESHOW

Etapa	χ^2	GL	Sig
8	7,371	8	0,497

$H_0: \text{Clas. Obs.} = \text{Clas. Prev.}$

$H_1: \text{Clas. Obs.} \neq \text{Clas. Prev.}$

Fonte: Elaborado pelo Autor com auxílio do software SPSS.

Na tabela 2 representa um resultado onde a significância não apresentou rejeição da hipótese nula a um nível de significância de 0,05, ao apresentar o p-valor de 0,497 e concluindo que não há diferença significativa entre a previsão e a classificação observada.

3.6 CURVA RECEIVER OPERATING CHARACTERISTIC (ROC)

A Curva ROC é um gráfico elaborado para efetivar a análise preditiva do modelo e a abrangência da área correspondente à parte inferior da curva indica a discriminação de um modelo. É gerada a partir de uma Tabela de Classificação e do estabelecimento de um ponto de corte, cujos valores das probabilidades estimadas acima deste ponto indicam a ocorrência do evento de interesse, os valores abaixo indicam a não-ocorrência (FÁVERO et al., 2009).

Para Brito e Assaf Neto (2008, p. 27)

[...] a curva ROC (Receiver Operating Characteristic) constitui uma técnica bastante útil para validar modelos de risco de crédito e está baseada nos conceitos da sensibilidade e da especificidade. A sensibilidade é a proporção de acerto na previsão da ocorrência de um evento nos casos em

que ele de fato ocorreu. A especificidade é proporção de acerto na previsão da não ocorrência de um evento nos casos em que ele de fato não ocorreu

A área abaixo da Curva ROC é denominada por AUROC (Area Under ROC). Na Tabela 3 mostra a interpretação do indicador de área gerado através da curva em níveis que facilitam o entendimento ao discriminar. Tem-se que a discriminação é aceitável quando a área abaixo da Curva ROC for maior que 0,7 unidades de área (u.a.). Quando apresentar o tamanho da área abaixo do indicador de 0,5 está informando que o modelo não discrimina bem.

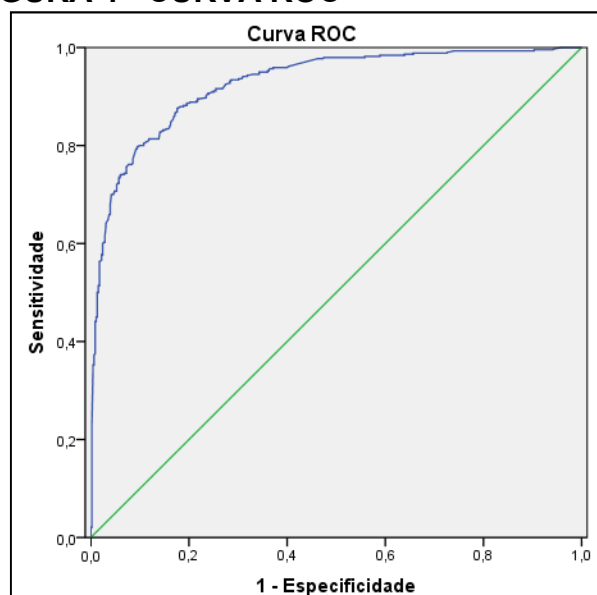
TABELA 3 - DISCRIMINAÇÃO DA ÁREA DA CURVA ROC

Área abaixo da Curva ROC (u.a.)	Interpretação
Menor ou igual a 0,5	Não há discriminação
Entre 0,7 e 0,8	Discriminação aceitável
Maior que 0,8	Discriminação excelente

Fonte: Fávero et al. (2009).

Na Figura 4 apresenta a representação da curva gerada. Através de Diniz e Louzada (2012) tem-se que a indicação a partir da Curva ROC destaca como deve ser o desempenho do modelo desenvolvido por toda a amplitude das pontuações de escores gerados. A linha diagonal reforça a medida da área 0,5 que descarta o uso de um modelo. Quando a curva fica abaixo da diagonal corresponde ao mesmo descarte.

FIGURA 4 - CURVA ROC



Fonte: Elaborado pelo autor com auxílio do SPSS.

3.7 TESTE KOLMOGOROV-SMIRNOV (KS)

Para Sicsú (2010) é a métrica para avaliação de desempenho dos modelos *credit scoring* mais popular entre os analistas de crédito e desenvolvedores. O KS consiste em medir a maior distância entre a distribuição da frequência acumulada dos bons clientes (ocorrência do evento) em relação à distribuição de frequência acumulada dos maus clientes (não ocorrência).

Segundo Manfio (2007) existem duas formas de aplicação do teste:

- a) Uma é na avaliação do perfil da população, é conhecido por KS1 e corresponde ao Indicador de Aderência ai medir a frequência modelada com a pontuada em outro período com a mesma equação. O ideal é ficar abaixo de 10%.
- b) A outra é a avaliação de discriminação final do modelo, é conhecida por KS2 e corresponde ao Indicador de Desempenho ao mostrar a separação entre os escores dos bons em relação aos maus. Quanto maior a diferença entre as duas curvas, melhor o ajuste do modelo. Considera-se uma discriminação aceitável quando a distância é maior que 20%.

Para Diniz e Louzada (2012) o KS1 e o KS2 também indicam se a equação gerada na modelagem do *credit scoring* ainda pode continuar a ser utilizada na segmentação de novos proponentes. A tabela 4 indica um resumo de fácil compreensão do indicador de desempenho KS2.

TABELA 4 - DISCRIMINAÇÃO DO TESTE KS2 PARA CREDIT SCORING

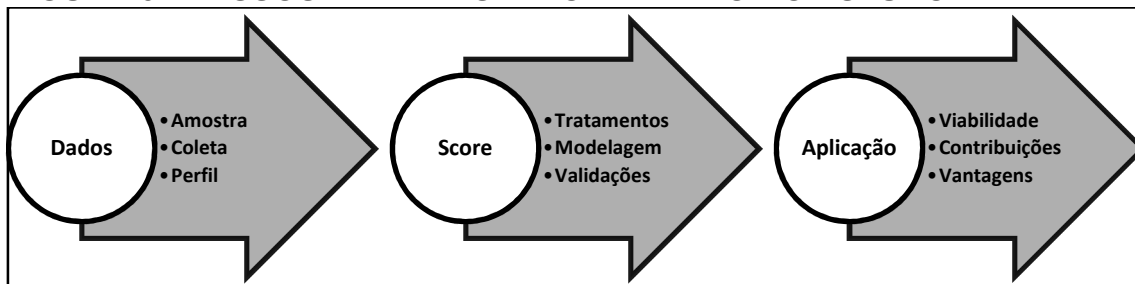
Indicador	Interpretação
Menor que 10%	Não há discriminação
Entre 10% e 20%	Discriminação baixa
Maior que 20%	Discrimina bem

Fonte: Elaborado pelo autor a partir de Diniz e Louzada (2012).

4 ANÁLISE DOS RESULTADOS

Esta seção do trabalho descreve a execução prática da pesquisa e a aplicação da metodologia. Partindo das características da amostra utilizada no estudo e da indicação das premissas para análises, mostrando o tratamento das variáveis, a aplicação da modelagem e técnicas estatísticas, a equação gerada e os resultados obtidos. Na Figura 5 demonstra os passos metodológicos realizados.

FIGURA 5 - PASSOS PARA DESENVOLVIMENTO DO ESTUDO



Fonte: Elaborado pelo autor.

4.1 A AMOSTRA E DESCRIÇÃO DOS DADOS

As bases de dados utilizadas foram oriundas de uma seleção amostral proveniente da carteira de crédito de uma Instituição Financeira (IF), que solicitou sua manutenção no anonimato. A amostra está segmentada temporalmente em três bases de dados, condição essencial para desenvolvimento do *Credit Scoring*. O Quadro 1 a seguir apresenta as características iniciais dos dados.

QUADRO 1 - SEGMENTAÇÃO DA BASE DE DADOS DO ESTUDO

Tipo da base	Desenvolvimento	Validação 1	Validação 2
Quantidade	8.721	2.489	8.808
Período	jan/13 até jul/13	ago/13	mai/12 até nov/12
Característica	Consignado com oferta de Cartão de Crédito e performance de comportamento em 12 meses	Consignado com oferta de Cartão de Crédito e performance de comportamento em 12 meses	Consignado que não houve oferta de Cartão de Crédito e sem performance
Finalidade	Construir equação	Validação da equação	Validação da equação

Fonte: Elaborado pelo autor.

Conforme indicado no Quadro 1 a amostra possui um total de 20.018 clientes que solicitaram o crédito consignado entre o período de maio/12 à agosto/13, exceto ao mês de dezembro/12, compreendendo 15 meses de observação. Verifica-se, também que apenas 11.210 clientes receberam o Cartão de Crédito convencional, com limite para utilização em compras e esse grupo foi

utilizado para desenvolver a equação do *Credit Scoring* e para a validação com desempenho de pagamento (Validação 1).

A amostra contém um total de 17 variáveis com informações da proposta de solicitação do crédito e com informações da situação do cliente no mercado financeiro. Essas informações são fornecidas por bureaux de crédito e dados existentes no Brasil, tais como Serasa, Boa Vista Serviços, SPC Brasil e outros. No Quadro 2 tem-se a indicação das variáveis fornecidas e as suas características.

QUADRO 2 - VARIÁVEIS DA AMOSTRA

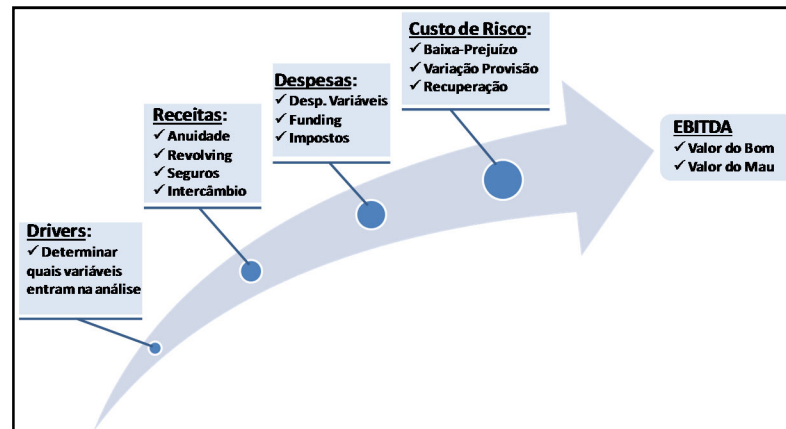
Variável	Descrição
Id	Numeração de controle e ordenamento dos casos, não corresponde a nenhum tipo de documento pessoal.
Safra	Mês de concessão do crédito, para a base de Desenvolvimento e Validação 1 a data da concessão do Cartão de Crédito, para a Validação 2 a data do Consignado.
Performance	Desempenho dos clientes com Cartão de Crédito no período de 12 meses após receber o limite de crédito. Atribui-se Mau para quem apresentou atraso superior a 90 dias no período e Bom para os demais. A base Validação 2 não possui essa variável.
Bureau_01	Pontuação do comportamento de crédito no mercado fornecida por um bureau de credito.
Bureau_02	Pontuação do comportamento de crédito no mercado fornecida por um segundo bureau de credito.
UF	Estado da federação de residência do Cliente.
Localidade	Cidade de residência do Cliente.
CEP	CEP da residência do Cliente.
Idade	Idade na data da concessão do crédito.
Tipo_Resid	Classificação da residência do Cliente em Própria ou Alugada-Outros.
Genero	Informe o gênero do cliente, se Masculino ou Feminino.
Estado_Civil	Classificação do estado civil do cliente em Casado, Solteiro, Divorciado ou Viúvo.
Renda	Renda na data da concessão do crédito.
Servidor	Classificação do cliente por categoria profissional como Aposentado, Estadual, Federal, For Armadas, Municipal, Pensão ou Policia Militar.
Instrucao_Escola	Nível de escolaridade do cliente, em Analfabeto, Fundamental, Medio ou Superior.
Possui_maior_6x_consultas_bureau	Informação externa de um bureau de crédito que indica Sim se o cliente possui mais de 6 consultas nos 90 dias anteriores a data do crédito ou Não em caso contrário.
Atraso_resolvido_6m	Informação externa de um bureau de crédito que indica Teve Atraso se o cliente quitou atrasos nos 6 meses anteriores a data do crédito ou Não Teve em caso contrário.

Fonte: Elaborado pelo autor.

A IF informou que o cliente bom apresenta o valor de R\$ 793,76 e o valor do cliente mau em R\$ -1.235,48. Essa informação é obtida através da Demonstração de Resultados dos clientes, considerando receitas, custo de risco e despesas

(Figura 6), no período dos 12 meses de observação das safras de desenvolvimento da equação.

FIGURA 6 - METODOLOGIA DE DEFINIÇÃO DO VALOR DE CADA CLIENTE

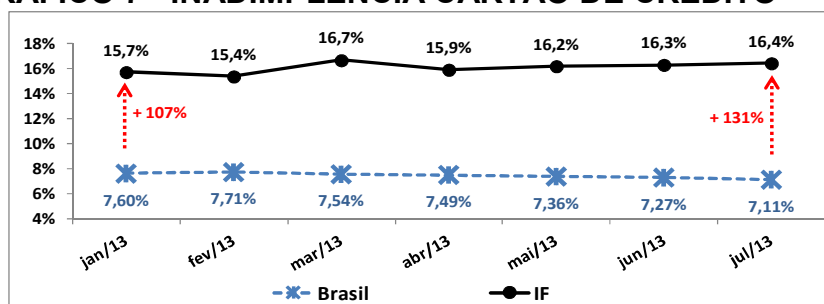


Fonte: Elaborado pelo autor.

Ter conhecimento de quanto vale o cliente Bom e o Mau contribui tanto para inferência da rentabilidade do negócio e estimar o ponto ótimo para retorno financeiro para a empresa como também colabora com a facilitação das dinâmicas para tomada de decisão baseada no ponto de corte que rejeita clientes através da utilização dos escores de classificação.

Outra informação fornecida pela IF foi a inadimplência do produto Cartão de Crédito no período correspondente à Base de Desenvolvimento. No gráfico 7 tem-se o evolutivo comparando com a Inadimplência de Cartão de Crédito no mercado brasileiro no mesmo período. A IF apresentou no período uma inadimplência média de 16,09%.

GRÁFICO 7 - INADIMPLÊNCIA CARTÃO DE CRÉDITO



Fonte: Elaborado pelo autor com dados da IF.

Dentre as informações de mercado utilizadas no trabalho, uma delas foi o preço médio unitário de R\$ 3,00 para consulta de escore fornecida pelas empresas de bureau de crédito. Outra informação foram os honorários estimados de uma

consultoria para construção de um modelo de escore conforme demonstrado no Quadro 3. Nele é apresentado um detalhamento de horas estimadas, valores e as possíveis etapas.

QUADRO 3 - PRECIFICAÇÃO DE CONSULTORIA PARA CONSTRUÇÃO DO MODELO

	Local Atividade	Semanas de Atividades								Horas
		1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	
Entendimento do Negócio	Empresa/Consultoria	■	■							36
Levantamento das bases	Empresa/Consultoria	■	■							
Entendimento das Variáveis	Empresa/Consultoria	■	■							
Análise de Indicadores de Desempenho	Empresa/Consultoria	■	■							
Sintetização das Informações colhidas	Consultoria			■						16
Desenvolvimento da Equação	Consultoria				■	■	■	■		72
Documentação da equação e entrega	Consultoria								■	16
Total horas										140
Valor por hora										R\$ 150
Valor Atividade										R\$ 21.000

Fonte: Elaborado pelo autor, baseado em práticas do mercado.

4.2 ANÁLISE DOS DADOS E PROCEDIMENTOS

Com base no banco de dados descrito anteriormente, o estudo teve como ponto de partida a categorização das variáveis contínuas e a análise do Risco Relativo (RR), definindo a construção das *dummies*. Em seguida, foi gerado um modelo de Regressão Logística Binária, validado com o teste Kolmogorov-Smirnov (KS), com a Curva ROC, além da Tabela de Classificação. O software utilizado para estimar o modelo de Regressão Logística Binária foi o SPSS, versão 20 demo, com base no método de análise Enter.

O Quadro 4 expõe as variáveis aproveitadas para modelagem e aplicadas à análise do RR na Base de Desenvolvimento. Para Gonçalves, Gouvêa e Mantovani (2013, p. 151) “o RR, além de auxiliar no agrupamento das categorias, ajuda a entender se a categoria em questão está mais ligada a clientes bons ou ruins”. Essa análise seguiu o mesmo modo executado por Gouvêa, Gonçalves e Mantovani

(2013, p. 108) os “grupos que apresentaram risco relativo (RR) semelhante foram reagrupados a fim de se diminuir o número de categorias por variável”.

QUADRO 4 - VARIÁVEIS INDEPENDENTES APROVEITADAS

Variável	Ação	Categorias	Nº <i>Dummies</i>
CEP	Categorizada	95	5
Renda	Categorizada	6	4
Gênero		2	1
Estado_Civil		4	2
Tipo_Resid		2	1
Idade	Categorizada	9	4
Servidor		7	2
Instrucao_escola		4	2
Bureau_01	Categorizada	11	5
Bureau_02	Categorizada	10	6
Possui_maior_6x_consultas_bureau		2	1
Atraso_resolvido_6m		2	1

Fonte: Elaborado pelo autor.

A partir da análise RR geram-se variáveis *dummies*, cujo detalhamento encontra-se no APÊNDICE A – Análise RR e Geração das *Dummies*. Das 12 variáveis independentes utilizadas na modelagem, com um total de 154 categorias, foram geradas 34 variáveis *dummies* para serem submetidas à análise de correlação de Spearman. Todas as *dummies* geradas ficaram com correlação de Spearman aceitável para aplicação do modelo, conforme apresentado no APÊNDICE B – Matriz de Correlação de Spearman.

Após preparação e pré-seleção das variáveis disponíveis para construção do modelo, utilizou-se o método Enter do software SPSS para a técnica da Regressão Logística Binária, a fim de classificar a probabilidade de inadimplir necessária para mitigação do risco. Foram selecionadas as *dummies* que satisfizeram o teste Wald e foram significantes com P-valor abaixo de 0,100, conforme é exibido na Tabela 5.

TABELA 5 - EQUAÇÃO GERADA

	B	Wald	P-valor
D_CEP_GRUPO_1	-0,356	19,632	0,000
D_CEP_GRUPO_6	0,252	10,177	0,001
D_Renda_1 (Renda <= R\$ 678)	0,143	4,965	0,026
D_EstCivil_Cas_1 (Casado)	0,165	7,405	0,007
D_Tipo_Resid_2 (Alugada ou outro)	-0,373	13,316	0,000
D_FX_idade_1 (Idade <=60)	-0,244	5,781	0,016
D_FX_idade_2a5 (Idade 61 – 67)	-0,286	13,205	0,000
D_FX_idade_67 (Idade 68 – 70)	-0,157	2,936	0,087
D_INSTRUCAO_ESCOLA_1 (Analfabeto)	-0,274	7,157	0,007
D_INSTRUCAO_ESCOLA_4 (Superior)	0,232	4,977	0,026
D_FX_BUREAU_01_12	-0,786	40,871	0,000
D_FX_BUREAU_01_3	-0,726	36,658	0,000
D_FX_BUREAU_01_4	-0,246	4,385	0,036
D_FX_BUREAU_01_6	0,223	3,546	0,060
D_FX_BUREAU_01_7a11	1,066	91,480	0,000
D_FX_BUREAU_02_1	-2,976	1031,083	0,000
D_FX_BUREAU_02_2	-1,139	80,035	0,000
D_FX_BUREAU_02_3	-1,033	110,791	0,000
D_FX_BUREAU_02_4	-0,779	63,644	0,000
D_FX_BUREAU_02_5	-0,484	21,238	0,000
D_Possui_maior_6x_consultas_bureau_1	-0,260	11,712	0,001
D_Atraso_resolvido_6m_1	0,207	3,807	0,051
Constante	1,727	178,751	0,000

Fonte: Elaborado pelo autor com auxílio do SPSS.

Na Tabela de Classificação apresentada a seguir, tem-se uma primeira visualização da eficiência da construção do modelo de regressão logística binária. Com 80,9% de acertos totais entre a relação dos casos de clientes bons e maus na amostra observada e a correspondente previsão gerada pela equação. Destaque para os 89,7% de acertos somente no grupo Bom. Em Gouvêa, Gonçalves e Mantovani (2013) obteve acertos totais de 70,2%.

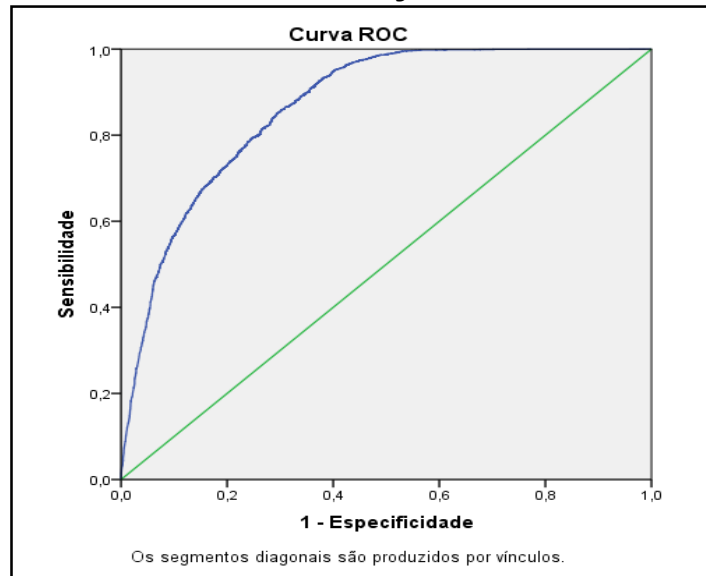
TABELA 6 - TABELA DE CLASSIFICAÇÃO

Observado		Previsto		
		Performance		Porcentagem de acertos
		MAU	BOM	
Performance	MAU	2036	1087	65,2
	BOM	577	5021	89,7
Porcentagem global				80,9

Fonte: Elaborado pelo autor com auxílio do SPSS.

Na Curva ROC também se validou a aplicabilidade da equação onde a área abaixo da curva correspondeu a 0,867 u.a. indicando uma discriminação excelente, conforme demonstrado no Gráfico 8 . Mendes e Fard (2014) obteve na melhor equação a área de 0,951 u.a. e Kocenda e Vojtek (2011) obteve 0,877 u.a. na melhor equação, ambos com a regressão logística.

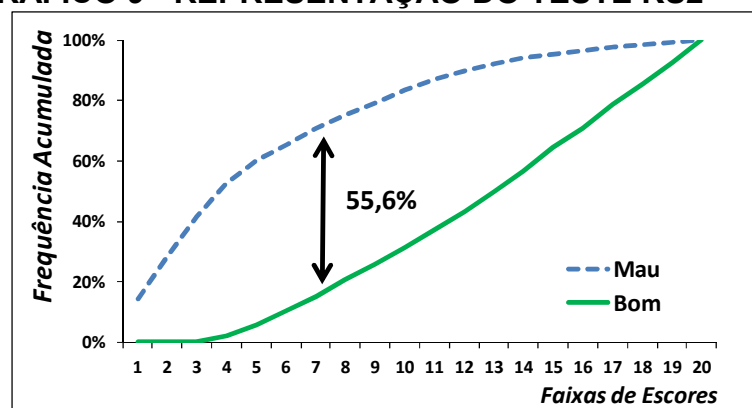
GRÁFICO 8 - REPRESENTAÇÃO DA CURVA ROC



Fonte: Elaborado pelo autor com auxílio do SPSS.

Outro teste que se utilizou para capacitar a efetiva utilização da equação gerada com a base de Desenvolvimento para mitigar o risco foi o Kolmogorov-Sminov (KS), especificamente o KS2. A equação de regressão logística binária resultou um KS2 de 55,6%. O Gráfico 9 demonstra excelente separação entre as frequências de clientes Bons e Maus. Gonçalves, Gouvêa e Mantovani (2013) obteve KS2 de 38% na regressão logística de Desenvolvimento.

GRÁFICO 9 - REPRESENTAÇÃO DO TESTE KS2



Fonte: Elaborado pelo autor.

Verificamos que os P-valores das variáveis na regressão logística, o teste Wald, a Tabela de Classificação, a Curva ROC e o teste KS2 corroboram com o quanto satisfatórios foram os parâmetros estatísticos das variáveis que compõem o modelo, fortalecendo o quanto adequado ficou ao nosso problema. Porém, o teste de Hosmer-Lemeshow foi o único que não atestou adequação com o modelo. Conforme podemos verificar na Tabela 7.

TABELA 7 - RESULTADO DO TESTE DE HOSMER-LEMESHOW

Etapa	χ^2	GL	P-valor
1	155,677	8	0,000

$$H_0: \text{Clas. Obs.} = \text{Clas. Prev.}$$

$$H_1: \text{Clas. Obs.} \neq \text{Clas. Prev.}$$

Fonte: Elaborado pelo Autor com auxílio do software SPSS.

O Teste de Hosmer-Lemeshow representa um resultado onde a significância apresentou a rejeição da hipótese nula a um nível de significância de 0,05, ao apresentar o P-valor de 0,000 e concluindo que há diferença significativa entre a previsão e a classificação observada. O estudo de Mendes e Fard (2014) também apresentou P-valores muito baixos nas 5 observações estudadas obteve 0,016513 como o maior e ainda assim sugere a equação da regressão logística para aplicação devido consistências de outras validações do modelo.

Este resultado não invalida a utilização da equação devido resultados satisfatórios nas 4 observações de validação anteriores e, também é interessante verificar o desempenho da aplicação da equação gerada nas bases de dados Validação 1 e 2, conforme verificará a seguir.

TABELA 8 - TESTES BASES PARA VALIDAÇÃO

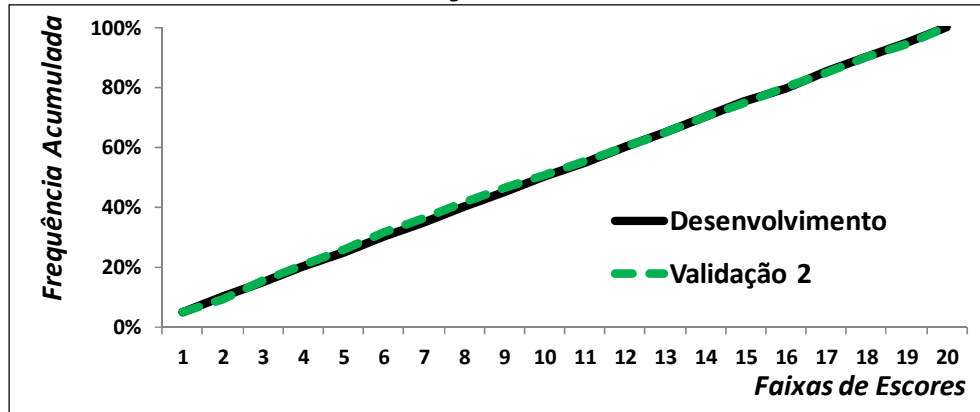
Testes	Validação	
	Base 1	Base 2
AUROC	0,840	
KS2	50,4%	
KS1	3,2%	1,5%

Fonte: Elaborado pelo Autor.

As bases para Validação utilizam casos fora do período da Base de Desenvolvimento e simulam a aplicação da equação como se fosse o ambiente de

produção. Na Tabela 8 temos excelentes resultados, com destaque para o KS1 de 1,5% na Base Validação 2, conforme é visualizado no Gráfico 10. Para o KS1 quanto mais próximas as frequências melhor é a aderência da equação.

GRÁFICO 10 - REPRESENTAÇÃO DO TESTE KS1



Fonte: Elaborado pelo autor.

Os testes aplicados nas bases de dados para validações apresentaram indicadores que contribuem para discriminação dos clientes em outro período que não o utilizado no desenvolvimento da equação. Isso corrobora com indicação da aplicação da equação para verificar os resultados específicos com as informações do negócio.

4.3 RESULTADO DO MODELO

A partir da equação gerada e os escores obtidos para todos os clientes propiciará os resultados efetivos do modelo *Credit Scoring* ao negócio de crédito pessoal. Esses resultados serão visualizados ao associar os números gerados pela equação aos dados de Inadimplência e valor do cliente fornecido pela IF, como também os custos por informações das empresas bureau de crédito e da consultoria para modelagem do *Credit Scoring*.

Gerou uma tabela para mitigar a inadimplência e a rentabilidade pelo volume de casos em cada categoria e faixas de escore, de acordo com as quantidades de clientes bons e maus, conforme é exposto na Tabela 9. A inadimplência média de 16,09% da IF no período de desenvolvimento do modelo será utilizada para projetar e simular prováveis desempenhos. Com as informações do cliente bom corresponder ao valor de R\$ 793,76 e o cliente mau ser equivalente a R\$ -1.235,48.

TABELA 9 - INADIMPLÊNCIA E RENTABILIDADE POR FAIXA

	Faixa Escore	Taxa de Maus*	Taxa de Maus**	Var. % Maus**	Inadimplência Mensal Estimada	Rentabilidade Por Faixa	Rentabilidade Acumulada
1	<= 0,06798	99,5%	35,8%		16,09%	-543.259,16	585.064,44
2	0,06799 - 0,08027	99,6%	32,4%	-9,5%	14,6%	-545.730,12	1.128.323,60
3	0,08028 - 0,10965	97,9%	28,6%	-20,2%	12,8%	-501.873,92	1.674.053,72
4	0,10966 - 0,31399	79,0%	24,6%	-31,2%	11,1%	-355.685,64	2.175.927,64
5	0,31400 - 0,42012	54,5%	21,2%	-40,7%	9,5%	-135.202,56	2.531.613,28
6	0,42013 - 0,50109	36,5%	19,0%	-46,9%	8,5%	23.430,20	2.666.815,84
7	0,50110 - 0,57560	39,0%	17,8%	-50,4%	8,0%	1.108,56	2.643.385,64
8	0,57561 - 0,64336	31,0%	16,1%	-54,9%	7,3%	72.131,96	2.642.277,08
9	0,64337 - 0,70855	29,7%	14,9%	-58,4%	6,7%	83.071,92	2.570.145,12
10	0,70856 - 0,76316	31,2%	13,6%	-62,1%	6,1%	70.102,72	2.487.073,20
11	0,76317 - 0,81273	24,8%	11,8%	-67,1%	5,3%	126.127,68	2.416.970,48
12	0,81274 - 0,85249	21,1%	10,3%	-71,1%	4,6%	159.389,28	2.290.842,80
13	0,85250 - 0,88355	17,4%	9,0%	-74,9%	4,0%	192.650,88	2.131.453,52
14	0,88356 - 0,90733	12,4%	7,8%	-78,2%	3,5%	237.294,16	1.938.802,64
15	0,90734 - 0,92460	7,7%	7,0%	-80,3%	3,2%	306.716,68	1.701.508,48
16	0,92461 - 0,93478	9,9%	6,9%	-80,8%	3,1%	226.898,96	1.394.791,80
17	0,93479 - 0,94229	6,9%	6,2%	-82,6%	2,8%	313.246,12	1.167.892,84
18	0,94230 - 0,94925	6,4%	6,0%	-83,3%	2,7%	270.300,08	854.646,72
19	0,94926 - 0,95493	7,5%	5,8%	-83,8%	2,6%	282.289,48	584.346,64
20	0,95494+	4,0%	4,0%	-88,8%	1,8%	302.057,16	302.057,16
Total		35,8%				585.064,44	

* Taxa de maus na faixa de escore; ** Taxa de maus acumulada decrescente por faixa de escore.

Fonte: Elaborado pelo autor.

Na Tabela 9 indica que a taxa de Maus na carteira de clientes utilizadas na modelagem foi de 35,8%, resultado da razão das quantidades de maus pela soma dos bons e maus, que corresponde a inadimplência média no período de 16,09%. Todas essas contas geram uma rentabilidade total de R\$ 585 mil, resultado do produto da quantidade e valor do Mau somado ao produto da quantidade e valor do Bom. O APÊNDICE C – Dados da Modelagem - apresenta o detalhamento dos dados.

Ainda na Tabela 9, o Modelo de *Credit Scoring* gerado indica nas cinco primeiras faixas de escore valores negativos na coluna “Rentabilidade por Faixa”. Mostra que a maioria dos clientes nestas faixas gera prejuízo. Os clientes posicionados a partir da faixa 6 dão retorno positivo em sua maioria. Removendo as faixas de prejuízo tem-se que o modelo proporciona melhores indicadores para a Inadimplência e na Rentabilidade do produto Cartão de Crédito proveniente da oferta para clientes consignados.

TABELA 10 - DECISÃO POR CUSTO, RETORNO E PERFORMANCE

Cálculos de Custos				
	Sem Escore	Bureau 1	Bureau 2	Consultoria
Bureaus		R\$ 44.850,86	R\$ 44.850,86	R\$ 89.701,71
Consultoria				R\$ 21.000,00
Custos	R\$ -	R\$ 44.850,86	R\$ 44.850,86	R\$ 110.701,71
Retorno Final				
	Sem Escore	Bureau 1	Bureau 2	Consultoria
Rentabilidade	R\$ 585.064,44	R\$ 2.344.224,96	R\$ 2.599.593,40	R\$ 2.666.815,84
Custos (-)	R\$ -	R\$ 44.850,86	R\$ 44.850,86	R\$ 110.701,71
Resultado	R\$ 585.064,44	R\$ 2.299.374,10	R\$ 2.554.742,54	R\$ 2.556.114,13
Performance				
	Sem Escore	Bureau 1	Bureau 2	Consultoria
KS2		48,7%	54,7%	55,6%
AUROC		0,809	0,848	0,867
Taxa Aprovação	100%	62%	64%	75%
Inadimplência	16,1%	8,0%	7,3%	8,5%

Fonte: Elaborado pelo autor.

Na tabela 10 tem-se um resumo para tomada de decisão, com números inferidos pela modelagem obtida, as informações fornecidas pela IF e os custos utilizados de itens de mercado como valor consulta bureau e consultoria, simulando para um período com 12 safras, gerando 4 cenários de observação:

- a) No primeiro cenário, denominado Sem Escore, mantém a concessão do novo produto como está, não gerando custos adicionais, mantendo a rentabilidade em R\$ 585 mil, a inadimplência em 16,1% e aprovando 100% dos créditos;
- b) No segundo cenário, denominado Bureau 1, utiliza apenas o escore fornecido pelo Bureau 1, com um custo adicional na produção em torno de R\$ 45 mil, removendo faixas de escore com perda gera a rentabilidade de R\$ 2,3 milhões. Além do monitoramento da performance do Score Externo com KS2 de 48,7% e AUROC de 0,809. A inadimplência estimada cai para 8,0% e aprova 62% da carteira Consignada;
- c) No terceiro cenário, denominado Bureau 2, apenas o escore fornecido pelo Bureau 2 é utilizado, com mesmo custo adicional na produção em torno de R\$ 45 mil, removendo faixas de escore com perda gera a rentabilidade de R\$ 2,55 milhões. Além do monitoramento da performance do Score Externo com KS2 de 54,7% e AUROC de 0,848. A inadimplência estimada cai para 7,3% e aprova 64% da carteira Consignada;

d) No quarto cenário, denominado Consultoria, utiliza o *Credit Scoring* gerado neste estudo, considerando que foi construído por uma consultoria contratada ao custo de R\$ 21 mil, com um custo adicional na produção em torno de R\$ 89,7 mil utilizando o Bureau 1 e o 2. Removendo as faixas de score com perda gera a rentabilidade de R\$ 2,56 milhões. Além do monitoramento da performance com KS2 de 55,6% e AUROC de 0,867. A inadimplência estimada cai para 8,5% e aprova 75% da carteira Consignada;

Analisando os quatro cenários sugere-se que a IF não deve se manter concedendo Cartão de Crédito a partir da carteira de clientes Consignado através do seu processo sem utilização score, isto é, deve abandonar o cenário “Sem Escore”.

Todos os demais scores fornecem um aumento da rentabilidade no mínimo 4 vezes maior. Dado ao volume de dados da amostra e a forte aderência dos scores externos fornecidos por empresas Bureau fica a critério da IF escolher apenas o melhor deles em rentabilidade e desempenho. Mas levando em conta a taxa de Aprovação tem-se que o cenário da Consultoria demonstra o melhor aproveitamento na mitigação de risco e com maior aproveitamento de clientes.

Convém ressaltar que na estimação da Rentabilidade outras variáveis são consideradas como Provisão, Receitas e custos. A Provisão é a linha de maior impacto porque quanto maior a faixa de atraso maior serão os valores provisionados, deixando o retorno do cliente menor e por isso que não há a proporcionalidade com a inadimplência.

A inadimplência considera-se apenas o saldo da carteira. Neste estudo a inadimplência estimada é uma relação linear com a Taxa de Maus Acumulada da modelagem, conforme se pode verificar na Tabela 9. A categorização dos scores contribui para uma maior concentração dos piores clientes nas piores faixas, que ao eliminá-las da oferta proporciona maior rentabilidade em relação não linear à queda da inadimplência.

O modelo de *Crédit Scoring* gerado neste trabalho demonstrou o melhor desempenho nos indicadores da equação e maior aprovação dos clientes Consignados. Esses dados indicam que a construção dos scores de acordo com as características do negócio possibilita encontrar o ponto ótimo para rentabilização através do profundo conhecimento das características da carteira de clientes da IF.

5 CONCLUSÕES

O objetivo neste trabalho foi analisar se o desenvolvimento de um modelo de *credit scoring* visando ações de *cross-selling* teria condições para promover uma maior rentabilidade na oferta de um novo produto com risco muito distinto. No caso deste estudo trata-se da oferta do Cartão de Crédito a partir de uma carteira de empréstimo pessoal consignado alinhada ao risco do novo produto. Os resultados expostos demonstraram que o escoramento contribui eficientemente para melhorar os resultados financeiros.

O Modelo de *Credit Scoring* obtido pela equação gerada com a regressão logística tem eficiência de 80,9% nos acertos gerais entre as observações e previsões. E validados pelos testes Wald, KS, Curva ROC entre outros, contribuiu para realização de dois dos objetivos específicos propostos neste trabalho. Um é a geração de um modelo estatístico de previsão e de cálculo de probabilidade da inadimplência e o outro é verificar se um modelo de *Credit Scoring* discrimina uma carteira de empréstimo pessoal consignado.

Para o terceiro objetivo específico, correspondente a análise da viabilidade financeira da oferta de um novo produto a partir da carteira de empréstimo pessoal consignado, destacam-se os resultados práticos do modelo gerado ao negócio de crédito, auxiliando no direcionamento de decisões e expondo os benefícios que a aplicação do modelo de *Credit Scoring* propicia, tais como rentabilidade, controle da inadimplência e aproveitamento da taxa de aprovação de clientes com perfil de risco sustentável.

Os cenários de rentabilidade e desempenho gerados para exposição dos resultados fortalecem a tomada de decisão. No Cenário 1, sem uso do escoramento, a IF possui uma rentabilidade de R\$ 0,5 milhões e desempenho ruim com inadimplência de 16,1%. Nos demais cenários as rentabilidades ultrapassam R\$ 2,3 milhões, quatro vezes maior que o Cenário 1, e inadimplência abaixo de 9%, aproximadamente a metade do Cenário 1.

Os Cenários 2 e 3 utilizam somente o score de empresas Bureau e apresentam resultados e desempenhos satisfatórios, indicando que podem ser aproveitados de forma único e exclusiva. E o Cenário 4 indica o quanto vale a pena investir em Modelos de *Crédit Scoring* por apresentar a melhor discriminação entre

bons e maus clientes e maior taxa de aprovação de clientes com o percentual de 75% para o novo produto, contra 64% do melhor Bureau.

Vale ressaltar que o estudo tem limitação com relação à aplicação prática da equação desenvolvida no cotidiano das atividades da instituição financeira que cedeu os dados. Outras técnicas de modelagem, outros testes para validação e monitoramento da efetiva aplicação podem ser validados e constatados.

Ao final do estudo concluiu-se que os resultados colaboram nas decisões sobre a oferta de crédito pelas instituições financeiras também nas ações de *cross-selling*. Não apenas nas práticas existentes no mercado e na maior parte das abordagens dos trabalhos acadêmicos existentes, tais como empréstimo pessoal, classificação de empresas, em renegociações de saldo devedor ou parcelamentos, na determinação de montante para renovação do crédito.

REFERÊNCIAS

- ABDOU, H. A.; POINTON, J. Credit scoring, statistical techniques and evaluation criteria: a review of the literature. **Intelligent Systems in Accounting, Finance and Management**, New York, v. 18, p. 59-88, June 2011.
- AGARWAL, N.; SHARMA M. Fraud risk prediction in merchant-bank relationship using regression modeling. **Journal for Decision Makers**, Vikalpa, v. 39, n. 3, July/Sept. 2014.
- ALVES, M. C. **Estratégias para o desenvolvimento de modelos de credit score com inferência de rejeitados**. 2008. 93 f. Tese (Dissertação de Mestrado) - Universidade de São Paulo, São Paulo – SP, 2008.
- ARAÚJO, E. A.; CARMONA, C. U. M.; AMORIN NETO, A. Aplicação de modelos credit scoring na análise da inadimplência de uma instituição de microcrédito. **Revista de Ciências Administrativas**, Fortaleza, v. 13, p. 110-121, 2007.
- BANCO CENTRAL DO BRASIL (BCB). **SGS: Sistema Gerenciador de Séries Temporais**. 2016. Disponível em: <www.bcb.gov.br>. Acesso em: 03 jan. 2016.
- BERGER, A. N.; COWAN, A. M.; FRAME, W. S. The surprising use of credit scoring in small business lending by community banks and the attendant effects on credit availability, risk, and profitability. **Journal of Financial Services Research**, [S.l.], v. 39, n. 1-2, p. 1-17, Apr. 2011.
- BERRY, M. J.; LINOFF, G. S. **Data mining techniques for marketing, sales, and customer relationship management**. 2nd ed. New York: John Wiley & Sons, 2004.
- BERTAUT, C. C.; HALIASSOS, M. Credit Cards: facts and theories. **Social Science Research Network**, Apr. 2005. Disponível em: <http://ssrn.com/abstract=931179>. Acesso em: 17 jan. 2016.
- BOTELHO, D.; TOSTES, F. D. Modelagem de probabilidade de Churn. **Revista de Administração de Empresas da FGV**, São Paulo, v. 50, n. 4, p. 396-410, out./dez. 2010.
- BRESSAN, V. G. F. et al. Avaliação de insolvência em cooperativas de crédito: uma aplicação do sistema pearls. **RAM – Revista de Administração Mackenzie**, São Paulo, v. 12, n. 2, mar./abr. 2011.
- BRITO, G. A. S.; ASSAF NETO, A. Modelo de classificação de risco de crédito de empresas. **Revista de Contabilidade Financeira da USP**, São Paulo, v. 19, n. 46, p. 18-29, jan./abr. 2008.
- BRITO, G. A. S.; ASSAF NETO, A; CORRAR, L. J. Sistema de classificação de risco de crédito: uma aplicação a companhias abertas no Brasil. **Revista Contabilidade & Finanças-USP**, São Paulo, v. 20, n. 51, p. 28-43, set./dez. 2009.

BRUNI, M. E.; BERARDI, P.; IAZZOLINO, G. Lending decisions under uncertainty: a DEA approach. **International Journal of Production Research**, Hudson, NY, v. 52, n. 3, p 766-775, 2014.

CAMARGOS, M. A.; CAMARGOS, M. C. S.; ARAÚJO, E. A. A inadimplência em um programa de crédito de uma instituição financeira pública de Minas Gerais: uma análise utilizando regressão logística. **REGE-USP**, São Paulo, v. 19, n. 3, p. 467-486, jul./set. 2012.

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

CAPRIO, G.; HONOHAN, P. **Finance for Growth: policy choices in a volatile world**. Washington; New York: World Bank; Oxford University Press, 2001.

CHANG, C. et al. Information from relationship lending: evidence from loan defaults in China. **Journal of Money, Credit and Banking**, Ohio, v. 46, n. 6, p. 1225-1257, Sept. 2014.

COELHO, C. A.; MELLO, J. M. P.; FUNCHAL, B. The brazilian payroll lending experiment. **The Review of Economics and Statistics**, Harvard, v. 94, n. 4, p. 925-934, Nov. 2012.

CRESWELL, J. W. **Projeto de pesquisa: método qualitativo, quantitativo e misto**. 2. ed. Porto Alegre: Artmed, 2007.

CROUHY, M.; GALAI, D.; MARK, R. **Fundamentos da Gestão de Risco**. Rio de Janeiro: Qualitymark, 2007.

CROUX, C.; DEHON, C. Influence functions of the Spearman and Kendall correlation measures. **Statistical Methods & Applications**, Hudson, NY, v. 19, n. 4, p. 497-515, May 2010.

DIEESE. **Nota Técnica: a evolução do crédito na economia brasileira 2008-2013**, n. 135, maio 2014. Disponível em: <<http://www.dieese.org.br/notatecnica/2014/notaTec135Credito.pdf>>. Acesso em: 17 jan. 2016.

DINIZ, C.; LOUZADA, F. Modelagem estatística para risco de crédito. In: SINAPE. 20., 2012, João Pessoa – PB. **Anais...** João Pessoa: ABE, 2012.

EINAV, L.; JENKINS, M.; LEVIN, J. The impact of credit scoring on consumer lending. **The RAND Journal of Economics**, Califórnia, v. 44, n. 2, p. 249–274, 2013.

FÁVERO, L. P. et al. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009.

FERREIRA, R. S. et al. Análise discriminante e ratings: uma aplicação do Modelo Z'' score de Altman às empresas do setor aeroviário brasileiro, de 2005 a 2010. **Revista**

Alcance - Eletrônica, v. 20, n. 3, p. 325-344, jul./set. 2013. Disponível em: <<http://siaiap32.univali.br/seer/index.php/ra/article/viewFile/4472/2706>>. Acesso em: 17 jan. 2016.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GONÇALVES, E. B.; GOUVÊA, M. A.; MANTOVANI, D. M. N. Análise de risco de crédito com o uso de regressão logística. **Revista Contemporânea em Contabilidade – UFSC**, Belo Horizonte, v. 10, n. 20, p. 139-160, maio/ago. 2013.

GOUVÊA, M. A.; GONÇALVES, E. B.; MANTOVANI, D. M. N. Análise de risco de crédito com aplicação de regressão logística e redes neurais. **Revista Contabilidade Vista & Revista da Universidade Federal de Minas Gerais**, Belo Horizonte, v. 24, n. 4, p. 96-123, out./dez. 2013.

HENS, A. B.; TIWARI, M. K. Computational time reduction for credit scoring: an integrated approach based on support vector machine and stratified sampling method. **Expert Systems with Applications**, Amsterdam, v. 39, p. 6774–81, 2012.

HORTA, R. A. M.; ALVES, F. J. S.; CARVALHO, F. A. A. Seleção de atributos na previsão de insolvência: aplicação e avaliação usando dados brasileiros recentes. **Revista de Administração do Mackenzie**, São Paulo, v.15, n. 1, p. 125-151, jan./fev., 2014.

JAGRIC, T.; JAGRIC, V. A comparison of growing cell structures neural networks and linear scoring models in the retail credit environment. **Eastern European Economics**, Hudson, NY, v. 49, n. 6, p. 74-96, Nov./Dec. 2011.

KOCENDA, E.; VOJTEK, M. Default predictors in retail credit scoring: evidence from czech banking data. **Emerging Markets Finance & Trade**, Hudson, NY, v. 47, n. 6, p. 80-98, Nov./Dec. 2011.

LAU, K.; CHOW, H.; LIU, C. A database approach to cross selling in the banking industry: Practices, strategies and challenges. **Journal Database Marketing & Customer Strategy Management**, United Kingdom, v. 11, n. 3, p. 216-234, 2004.

LOPES, N. S. **Os impactos do empréstimo consignado no risco de crédito dos financiamentos imobiliários**. 2010. Dissertação (Mestrado) – Universidade Presbiteriana Mackenzie, São Paulo, SP, 2010.

MANFIO, F. **O risco nosso de cada dia**. São Paulo: Estação das Estrelas, 2007.

MEDINA, R. P.; SELVA, M. L. M. Análisis del Credit Scoring. **RAE – Revista de Administração de Empresas**, São Paulo, v. 53, n. 3, p. 303-315, maio/jun. 2013.

MENDES, A. C.; FARD, N. Binary logistic regression and PHM analysis for reliability data. **International Journal of Reliability, Quality and Safety Engineering**, San Francisco, USA, v. 21, n. 5, 2014.

OLIVEIRA, F. E. **SPSS básico para análise de dados**. Rio de Janeiro: Ciência Moderna, 2007.

ORMECI, E. L.; AKSIN, O. Z. Revenue management through dynamic cross selling in call centers. **Journal Production and Operations Management**, Baltimore, EUA, v. 19, n. 6, p. 742-756, Nov./Dec. 2010.

PINHO, A. G. Modelagem de cross-selling no marketing de relacionamento por máquinas de vetor de suporte. **Revista Pensamento Contemporâneo em Administração**, Rio de Janeiro, v. 3, n. 1, p. 66-79, jan./abr. 2009.

RIBEIRO, C. F. et al. Gestão de universidade privada: risco de inadimplência dos alunos. **Revista Brasileira de Gestão de Negócios da FECAP**, São Paulo, v. 14, n. 42, p. 26-40, jan./mar., 2012.

SABATO, G. Assessing the quality of retail customers: credit risk scoring models. **The IUP Journal of Financial Risk Management**, India, v. 7, n. 1/2, p. 35-43, 2010.

SANTOS, J. O.; FAMÁ, R. Avaliação da aplicabilidade de um modelo de Credit Scoring com variáveis sistêmicas e não-sistêmicas em carteiras de crédito bancário rotativo de pessoas físicas. **Revista Contabilidade Financeira da USP**, São Paulo, n. 44, p. 105-117, maio/ago. 2007.

SANTOS, J. O.; SANTOS, A. R. O Modelo KMV e sua utilidade no processo de análise do risco de crédito. **Revista de Gestão USP**, São Paulo, v. 16, n. 2, p. 73-82, abr./jun. 2009.

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

SAUNDERS, A.; CORNETT, M. M. **Financial markets and institutions**. 5th ed. New York: McGraw-Hill Irwin, 2012.

SECURATO, J. R. **Crédito: análise e avaliação de risco: pessoas físicas e jurídicas**. 2. ed. São Paulo: Saint Paul, 2012.

SICSÚ, A. L. **Credit Scoring: desenvolvimento, implantação, acompanhamento**. São Paulo: Blucher, 2010.

SILVA, P. J. **Gestão e análise de risco de crédito**. 6. ed. São Paulo: Atlas, 2008.

SMITH, L. D. et al. Accuracy of information maintained by US Credit Bureaus: frequency of errors and effects on consumers' credit scores. **The Journal of Consumer Affairs**, Hudson, NY, v. 47, n. 3, p. 588-601, 2013.

SOUSA, M. M.; FIGUEIREDO, R. S. Credit analysis using data mining: application in the case of a credit union. **JISTEM - Journal of Information Systems and Technology Management – Revista de Gestão da Tecnologia e Sistemas de Informação**, v. 11, n. 2, p. 379-396, May/Aug. 2014.

SOUZA, R. B. **O modelo de Collection Scoring como ferramenta para a gestão estratégica do risco de crédito**. 2000. 75 f. Dissertação (Mestrado) – Fundação Getúlio Vargas, São Paulo, SP, 2000.

VASCONCELLOS, R. S. **Modelos de escoragem de crédito aplicados a empréstimo pessoal com cheque**. 2004. 44 f. Dissertação (Mestrado) – Fundação Getúlio Vargas, Rio de Janeiro, RJ, 2004.

VICENTE, E. F. R. **A estimativa do risco na constituição da PDD**. 2001. 164 f. Dissertação (Mestrado) – Universidade de São Paulo, São Paulo, SP, 2001.

VIEIRA, V.; PIRES, D.; GALEANO, R. Determinantes do desempenho empresarial e das vendas cruzadas no varejo. **RAE - Revista de Administração de Empresas**, São Paulo, v. 53, n. 6, p. 565-579, nov./dez. 2013.

WANG, G. et al. Two credit scoring models based on dual strategy ensemble trees. **Knowledge-Based Systems**, Amsterdam, v. 26, p. 61-68, Feb. 2012.

WHITE, L. J. Credit rating agencies: an overview. **Annual Reviews of Financial Economics**, New York, v. 5, p. 93-122, 2013.

APÊNDICE A – Análise RR e Geração das *Dummies*

CEP

Grupos	Performance		Total	RR	Dummy
	MAU	BOM			
1	651	649	1300	0,556164	1
2	335	444	779	0,739396	2
3	465	680	1145	0,815821	3
4	780	1368	2148	0,978432	
5	371	838	1209	1,260112	5
6	521	1619	2140	1,733597	6
Total	3123	5598	8721	1	

Renda

Faixa Renda	Performance		Total	RR	Dummy
	MAU	BOM			
1 <= 678	1982	2790	4772	0,785307	1
2 679 - 950	273	476	749	0,97271	
3 951 - 1380	255	583	838	1,275462	3
4 1381 - 2000	243	622	865	1,427984	45
5 2001 - 3000	208	551	759	1,47784	45
6 3001+	162	576	738	1,983566	6
Total	3123	5598	8721	1	

Gênero

Classes	Performance		Total	RR	Dummy
	MAU	BOM			
1 F	1835	3362	5197	1,022117	
2 M	1288	2236	3524	0,96849	2
Total	3123	5598	8721	1	

Estado Civil

Classes	Performance		Total	RR	Dummy
	MAU	BOM			
1 Casado	1367	2904	4271	1,185133	1
2 Solteiro	134	191	325	0,795184	23
3 Divorciado	1156	1656	2812	0,799174	23
4 Viúvo	466	847	1313	1,013997	
Total	3123	5598	8721	1	

Tipo Residência

Classes	Performance		Total	RR	Dummy
	MAU	BOM			
1 Propria	1190	4461	5651	2,091339	
2 Alugada ou Outro	1933	1137	3070	0,328146	2
Total	3123	5598	8721	1	

Idade

	Faixa	Performance		Total	RR	Dummy
		MAU	BOM			
1	<= 60,00	498	606	1104	0,678863	1
2	61,00 - 62,00	378	522	900	0,770403	2a5
3	63,00 - 64,00	386	542	928	0,783341	2a5
4	65,00 - 66,00	404	603	1007	0,832674	2a5
5	67,00 - 67,00	210	323	533	0,858069	2a5
6	68,00 - 69,00	382	784	1166	1,144964	67
7	70,00 - 70,00	194	448	642	1,288295	67
8	71,00 - 72,00	335	841	1176	1,400523	89
9	73,00+	336	929	1265	1,542466	89
	Total	3123	5598	8721	1	

Servidor

	Classes	Performance		Total	RR	Dummy
		MAU	BOM			
1	Aposentado	2073	3663	5736	0,985773	
2	Estadual	262	578	840	1,230738	23
3	Federal	64	132	196	1,150623	23
4	For Armadas	23	36	59	0,8732	467
5	Municipal	589	1014	1603	0,960421	
6	Pensão	14	19	33	0,75712	467
7	Pol Militar	98	156	254	0,88805	467
	Total	3123	5598	8721	1	

Instrução Escolar

	Classes	Performance		Total	RR	Dummy
		MAU	BOM			
1	Analfabeto	334	473	807	0,790049	1
2	Fundamental	590	1016	1606	0,960685	
3	Medio	1943	3528	5471	1,012966	
4	Superior	256	581	837	1,266121	4
	Total	3123	5598	8721	1	

Bureau 1

	Faixas	Performance		Total	RR	Dummy
		MAU	BOM			
1	<= 37	647	156	803	0,134511	12
2	38 - 44	618	180	798	0,162489	12
3	45 - 63	514	323	837	0,350573	3
4	64 - 70	376	479	855	0,710701	4
5	71 - 74	231	450	681	1,086775	
6	75 - 78	270	803	1073	1,65917	6
7	79 - 80	111	412	523	2,070682	7a11
8	81 - 82	141	827	968	3,272092	7a11
9	83 - 84	103	684	787	3,704742	7a11
10	85 - 85	45	677	722	8,392962	7a11
11	86+	67	607	674	5,054206	7a11
	Total	3123	5598	8721	1	

Bureau 2

	Faixas	Performance		Total	RR	Dummy
		MAU	BOM			
1	<= 0	1669	297	1966	0,099275	1
2	1 - 38	210	175	385	0,464898	2
3	39 - 48	333	433	766	0,725409	3
4	49 - 54	262	543	805	1,156212	4
5	55 - 59	185	568	753	1,712836	5
6	60 - 64	183	722	905	2,201026	6a10
7	65 - 68	116	612	728	2,943286	6a10
8	69 - 73	83	890	973	5,982063	6a10
9	74 - 77	57	708	765	6,92943	6a10
10	78+	25	650	675	14,50482	6a10
Total		3123	5598	8721	1	

Possui mais de 6 consultas em 90d * Performance Tabulação cruzada

	Classes	Performance		Total	RR	Dummy
		MAU	BOM			
1	Sim	2243	2295	4538	0,570811	1
2	Nao	880	3303	4183	2,093944	
Total		3123	5598	8721	1	

Atraso_resolvido_6m * Performance Tabulação cruzada

	Classes	Performance		Total	RR	Dummy
		MAU	BOM			
1	Teve Atraso	1273	719	1992	0,315094	1
2	Nao teve	1850	4879	6729	1,47129	
Total		3123	5598	8721	1	

APÊNDICE C – Dados da Modelagem

Base Desenvolvimento

	Faixa Escore	Performance		Total	Taxa Aprovação
		MAU	BOM		
1	<= ,06798	441	2	443	100%
2	,06799 - ,08027	443	2	445	95%
3	,08028 - ,10965	412	9	421	90%
4	,10966 - ,31399	347	92	439	85%
5	,31400 - ,42012	236	197	433	80%
6	,42013 - ,50109	159	277	436	75%
7	,50110 - ,57560	170	266	436	70%
8	,57561 - ,64336	135	301	436	65%
9	,64337 - ,70855	130	307	437	60%
10	,70856 - ,76316	136	300	436	55%
11	,76317 - ,81273	108	327	435	50%
12	,81274 - ,85249	92	344	436	45%
13	,85250 - ,88355	76	361	437	40%
14	,88356 - ,90733	54	383	437	35%
15	,90734 - ,92460	37	444	481	30%
16	,92461 - ,93478	38	345	383	24%
17	,93479 - ,94229	33	446	479	20%
18	,94230 - ,94925	26	381	407	15%
19	,94926 - ,95493	33	407	440	10%
20	,95494+	17	407	424	5%
Total		3123	5598	8721	

Base Validação 1

	Faixa Escore	Performance 12m		Total
		MAU	BOM	
1	<= ,06798	83	1	84
2	,06799 - ,08027	97	0	97
3	,08028 - ,10965	149	2	151
4	,10966 - ,31399	80	28	108
5	,31400 - ,42012	72	51	123
6	,42013 - ,50109	69	119	188
7	,50110 - ,57560	50	79	129
8	,57561 - ,64336	67	103	170
9	,64337 - ,70855	45	105	150
10	,70856 - ,76316	27	81	108
11	,76317 - ,81273	25	91	116
12	,81274 - ,85249	33	94	127
13	,85250 - ,88355	21	93	114
14	,88356 - ,90733	19	92	111
15	,90734 - ,92460	11	71	82
16	,92461 - ,93478	7	107	114
17	,93479 - ,94229	9	112	121
18	,94230 - ,94925	9	135	144
19	,94926 - ,95493	14	112	126
20	,95494+	9	117	126
Total		896	1593	2489

Base Validação 2

	Faixas Escore	Frequência	Porcentual
1	<= ,06798	437	5,0
2	,06799 - ,08027	381	4,3
3	,08028 - ,10965	542	6,2
4	,10966 - ,31399	477	5,4
5	,31400 - ,42012	459	5,2
6	,42013 - ,50109	469	5,3
7	,50110 - ,57560	434	4,9
8	,57561 - ,64336	456	5,2
9	,64337 - ,70855	439	5,0
10	,70856 - ,76316	369	4,2
11	,76317 - ,81273	425	4,8
12	,81274 - ,85249	417	4,7
13	,85250 - ,88355	431	4,9
14	,88356 - ,90733	430	4,9
15	,90734 - ,92460	425	4,8
16	,92461 - ,93478	454	5,2
17	,93479 - ,94229	452	5,1
18	,94230 - ,94925	462	5,2
19	,94926 - ,95493	377	4,3
20	,95494+	472	5,4
	Total	8808	100,0

Bureau 1

Faixas Escore	Performance		Total	Taxa de Maus	Inadimplência	Rentabilidade por Faixa	Rentabilidade Acumulada	Taxa Aprovação
	MAU	BOM						
<= 37	647	156	803	35,81%	16,09%	-675.529	585.064	100%
38 - 44	618	180	798	31,27%	14,05%	-620.650	1.260.593	91%
45 - 63	514	323	837	26,10%	11,73%	-378.652	1.881.243	82%
64 - 70	376	479	855	21,39%	9,61%	-84.329	2.259.896	72%
71 - 74	231	450	681	17,83%	8,01%	71.796	2.344.225	62%
75 - 78	270	803	1073	15,53%	6,98%	303.810	2.272.429	54%
79 - 80	111	412	523	12,71%	5,71%	189.891	1.968.619	42%
81 - 82	141	827	968	11,30%	5,08%	482.237	1.778.728	36%
83 - 84	103	684	787	9,85%	4,43%	415.677	1.296.491	25%
85 - 85	45	677	722	8,02%	3,60%	481.779	880.814	16%
86+	67	607	674	9,94%	4,47%	399.035	399.035	8%
Total	3123	5598	8721			585.064		

Bureau 2

Faixa Escore	Performance		Total	Taxa de Maus	Inadimplência	Rentabilidade por Faixa	Rentabilidade Acumulada	Taxa Aprovação
	MAU	BOM						
<= 0	1669	297	1966	35,81%	16,09%	-1.826.269	585.064	100%
1 - 38	210	175	385	21,52%	9,67%	-120.543	2.411.334	77%
39 - 48	333	433	766	19,53%	8,77%	-67.717	2.531.877	73%
49 - 54	262	543	805	16,26%	7,30%	107.316	2.599.593	64%
55 - 59	185	568	753	13,52%	6,08%	222.292	2.492.277	55%
60 - 64	183	722	905	11,47%	5,15%	347.002	2.269.986	46%
65 - 68	116	612	728	8,95%	4,02%	342.465	1.922.984	36%
69 - 73	83	890	973	6,84%	3,07%	603.902	1.580.518	28%
74 - 77	57	708	765	5,69%	2,56%	491.560	976.617	17%
78+	25	650	675	3,70%	1,66%	485.057	485.057	8%
Total	3123	5598	8721			585.064		