

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

MARCOS SANTOS OLIVEIRA
CENTRO UNIVERSITÁRIO ÁLVARES PENTEADO (FECAP)
marcos@exogenia.com.br

EDUARDO POZZI LUCCHESI
CENTRO UNIVERSITÁRIO ÁLVARES PENTEADO (FECAP)
edupozzi@uol.com.br

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

1. INTRODUÇÃO

A utilização de dados de clientes para auxiliar em decisões gerenciais, classificando-os através de métodos estatísticos e definindo o risco financeiro é uma prática bastante utilizada nas instituições que concedem crédito e na gestão de portfólios ou produtos sujeitos à inadimplência.

Com o crescimento do volume de concessões de crédito tornou-se crucial para a manutenção de uma carteira de clientes saudável a estruturação sistemática de grandes volumes de dados por parte das instituições (BOTELHO e TOSTES, 2010). Esses dados alimentam modelos de análise para concessão de crédito conhecidos como modelos de *credit scoring* os quais se baseiam em dados históricos da base de clientes existentes para avaliar se um futuro cliente terá mais chances de ser bom ou mau pagador (Gouvêa et al., 2013).

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

O objetivo de um modelo de *credit scoring* é estimar a chance que cada cliente avaliado tem de corresponder a um melhor resultado em termos de retorno positivo à instituição que concede o crédito. Conhecendo esta previsão, a instituição pode implantar políticas de concessão diferenciadas.

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. Assim, sabendo a probabilidade de o cliente entrar ou não em inadimplência, pode-se decidir com objetividade para quais clientes deve ser dada ou renovada a concessão do crédito.

Para construir um modelo de *credit scoring*, 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.

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 suficientes para definir o perfil do cliente com condições a aceitar outros itens que possa satisfazê-lo. Portanto, o *credit scoring* também é aplicado na ação de seleção de clientes com bom perfil de risco para ofertar outros produtos do portfólio. Tal aplicação do *credit scoring* é particularmente pertinente quando o grau de risco do produto adicional ofertado é maior. Assim, torna-se interessante o desenvolvimento de um método de seleção de clientes que possuem crédito consignado (um produto de baixíssimo risco e com uma garantia de recebimento maior devido ao desconto automático em folha de pagamento) para a realização de oferta de cartões de crédito (produto com maior rentabilidade e alto grau de risco creditício).

Diante do exposto, neste trabalho o objetivo é analisar se o desenvolvimento de um modelo de *credit scoring* visando ações de *cross-selling* teria condições de promover um incremento na rentabilidade. De forma específica é feita a proposição de uma modelagem convencional de *credit scoring* em uma carteira de crédito consignado de uma Instituição Financeira visando a oferta de cartões de crédito para essa base de clientes. A hipótese central é que os investimentos

realizados na construção do modelo de *credit scoring* em uma carteira de crédito consignado para fins de oferta de cartões de crédito podem ser justificados pelo incremento na rentabilidade da Instituição Financeira.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. *Credit Scoring*

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 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 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 *et al.* (1999) pode ser compreendida em duas categorias, os modelos de aprovação de crédito, bastante conhecidos por *credit* ou *aplicattion 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.

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.

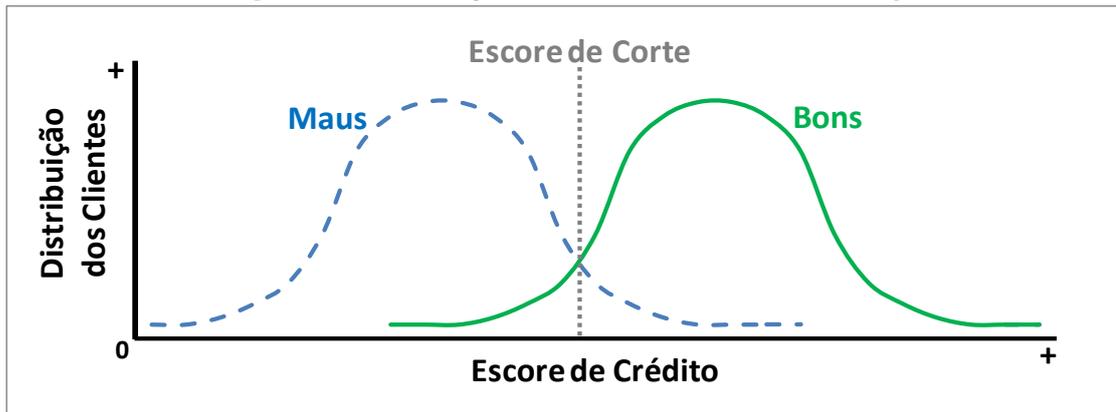
O *credit* e o *aplicattion scores* são a denominação utilizada para os modelos estatísticos que possuem como variáveis utilizadas as informações presentes no cadastro do proponente ao crédito ou para operação específica de produto. Esse modelo é direcionado em sua predominância para a tomada de decisão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

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 *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).

De acordo com Caouette *et al.* (1999) 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 mostra a Figura 1. Com base na distribuição de frequência dos escores dos clientes maus e dos clientes bons define-se um escore de rejeição ou corte.

Figura 1: Distribuição em Modelo de *Credit Scoring*

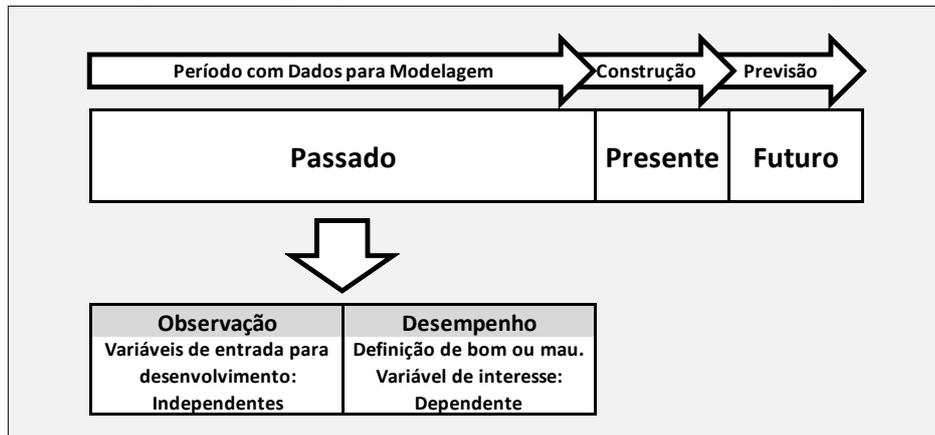


Fonte: Adaptado de Caouette *et al.* (1999, pág. 182).

Para desenvolvimento e classificação entre bons e maus é necessário um banco de dados, que pode corresponder a uma amostra dos proponentes aprovados em 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).

A Figura 2, adaptada de Diniz e Louzada (2012), apresenta a existência de três importantes etapas para o desenvolvimento do *credit scoring*. 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 escore. E a terceira é o futuro, momento onde o cálculo da equação retorna a previsão a ser aplicada.

Figura 2: Etapas para desenvolvimento do *credit scoring*



Fonte: Adaptado de DINIZ e LOUZADA (2012).

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.

2.2. Cross-selling

Ormeç e Aksin (2010) comentam que diversas empresas atuantes em serviços financeiros buscam o crescimento aprofundando o 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 os 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. A grande preocupação para oferta de produtos é a identificação da pessoa certa no momento conveniente para explorar o 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 e 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. Caso a organização não possua um sistema para *cross-selling* automatizado, a área responsável pelas vendas utilizará as informações perceptíveis para determinar se um cliente é apropriado para tentar uma venda cruzada.

Para Lau *et al.* (2004) o *cross-selling* deve ter uma atuação mais ampla e não se restringir a segmentação de ações puramente comerciais. Ele deve ser aplicado em uma visão abrangente, atuando em conjunto com os conceitos do marketing, como produto, preço, praça e promoção.

2.3. Estudos e evidências empíricas sobre modelos de *credit scoring*

Diversos autores se dedicaram ao estudo de modelos de *credit scoring* tanto no Brasil quanto no exterior. Dentre os estudos realizados no exterior destacam-se os trabalhos de Berger *et al.* (2011), Einav *et al.* (2013), Smith (2013) e Agarwal e Sharma (2014) no mercado americano; os estudos de Kocenda e Vojtek (2011) na República Checa, Jagric e Jagric (2011) na Eslovênia e Medina e Selva (2013) na Espanha. Ainda no âmbito internacional merece destaque o estudo experimental de Wang *et al.* (2012), a proposição de uma nova estratégia para reduzir o tempo computacional em modelos de *credit scoring* de Hens e Tiwari (2012), as revisões da literatura desenvolvidas por Sabato (2010) e Abdou e Pointon (2011) além das abordagens inovadoras de Bruni *et al.* (2014) com a análise envoltória de dados (DEA) e Mendes e Fard (2014) com a proposição de um modelo de regressão logística modificado.

No Brasil destacam-se os estudos de Santos e Famá (2007), Araujo *et al.* (2007), Brito e Assaf Neto (2008), Santos e Santos (2009), Pinho (2009), Brito *et al.* (2009), Botelho e Tostes (2010), Bressan *et al.* (2011), Camargos *et al.* (2012), Ribeiro *et al.* (2012), Gonçalves *et al.* (2013), Vieira *et al.* (2013), Ferreira *et al.* (2013), Gouvêa *et al.* (2013), Horta *et al.* (2014) e Sousa e Figueiredo (2014).

3. METODOLOGIA

Para estimar o modelo de *credit scoring* foi utilizada a técnica da regressão logística (ou logit). Tal técnica consiste na investigação da relação entre variáveis explicativas, métricas e não métricas e uma variável dependente categórica binária ou múltipla e necessita de testes estatísticos para seleção de variáveis e de validação da equação, tais como o Risco Relativo, a Correlação de Sperman, o teste de Kolmogorov Smirnov (KS), a Curva ROC e outros.

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:

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

em que:

$\pi(x)$ é a variável dependente;

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 (1), obtém-se:

$$\pi(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} \quad (2)$$

Segundo Araújo *et al.* (2007) dentro do contexto das atividades e definições do risco de crédito determina-se que a variável dependente é definida pelo cliente adimplente ou bom e pelo cliente 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 *et al.* (2007), as informações referentes a dados demográficos, econômicos e financeiros, sejam de empresas ou de pessoas constituem variáveis independentes e representam os fatores que influenciam a ocorrência da inadimplência. Após os tratamentos necessários para viabilizar melhor o manuseio e capacidade de informação, essas variáveis são aplicadas à regressão logística.

3.1 Amostra e Procedimentos

As bases de dados utilizadas foram oriundas de uma seleção amostral proveniente da carteira de crédito de uma Instituição Financeira que solicitou confidencialidade. A amostra está segmentada temporalmente em três bases de dados, condição essencial para desenvolvimento do *credit scoring*. A Tabela 1, a seguir apresenta as características iniciais dos dados.

Tabela 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 na Tabela 1 a amostra possui um total de 20.018 clientes que solicitaram o crédito consignado entre o período de maio de 2012 a agosto de 2013, exceto no mês de

dezembro de 2012, 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 *Bureau* de Crédito e dados existentes no Brasil, tais como Serasa, Boa Vista Serviços, SPC Brasil e outros. Na Tabela 2 tem-se a indicação das variáveis fornecidas e as suas características.

Tabela 2: Variáveis da Amostra

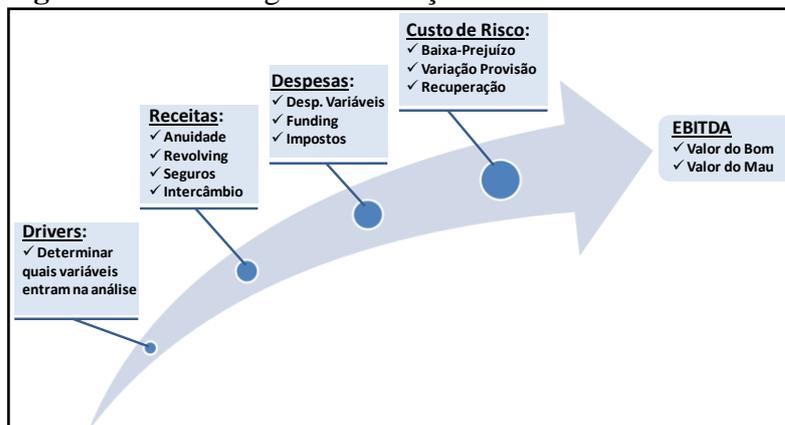
Variável	Descrição
Id	Numeração de controle e ordenamento dos casos, não corresponde a nenhum tipo de documento pessoal.
Safra	Mês de concessão do crédito, para a base de Desenvolvimento e Validação 1 a data da concessão do Cartão de Crédito, para a Validação 2 a data do Consignado.
Performance	Desempenho 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 <i>Bureau</i> de Crédito.
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_6 m	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 Instituição Financeira 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 3), no período dos 12 meses de observação das safras de desenvolvimento da equação.

Figura 3: 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 Instituição Financeira foi a inadimplência do produto cartão de crédito no período correspondente à base de desenvolvimento.

Dentre as informações de mercado utilizadas no trabalho, uma delas foi o preço médio unitário de R\$ 3,00 para consulta de score fornecida pelas empresas de Bureau de Crédito. Outra informação foram os honorários estimados de uma consultoria para construção de um modelo de score conforme demonstrado na Tabela 3. Nela é apresentado um detalhamento de horas estimadas, valores e as possíveis etapas.

Tabela 3: Precificação de consultoria para construção do modelo

	Local Atividade	Semanas de Atividades								Horas	
		1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	6 ^a	7 ^a	8 ^a		
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	Consultoria									16	
Desenvolvimento da Equação	Consultoria									72	
Documentação da equação e entrega	Consultoria									16	
										Total horas	140
										Valor/hora	R\$ 150
										Valor Atividade	R\$ 21.000

Fonte: Elaborado pelo autor, baseado em práticas do mercado

4. ANÁLISE DOS RESULTADOS

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 variáveis *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*.

A Tabela 4 apresenta as variáveis aproveitadas para modelagem e aplicadas à análise do RR na base de desenvolvimento. Para Gonçalves *et al.* (2013, pág. 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 *et al.* (2013, p. 108) em que os “grupos que apresentaram risco relativo (RR) semelhante foram reagrupados a fim de se diminuir o número de categorias por variável”.

Tabela 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*. 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.

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 inadimplência. 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 (Tabela 6), 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. Gouvêa *et al.* (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

A Curva ROC também validou a aplicabilidade da equação onde a área abaixo da curva correspondeu a 0,867 u.a. indicando uma discriminação excelente. Mendes e Fard (2014) obtiveram na melhor equação a área de 0,951 u.a. e Kocenda e Vojtek (2011) obtiveram 0,877 u.a. na melhor equação, ambos com a aplicação da regressão logística.

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-Sminorv (KS), especificamente o KS2. A equação de regressão logística binária resultou um KS2 de 55,6%. Gonçalves *et al.* (2013) obtiveram KS2 de 38% na regressão logística de desenvolvimento.

Verifica-se que os P-valores das variáveis na Regressão Logística Binária, o teste Wald, a Tabela de Classificação, a Curva ROC e o teste KS2 corroboram a adequação do modelo. Porém, o teste de Hosmer-Lemeshow foi o único que não atestou adequação com o modelo. 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, 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, mas, ainda assim os autores sugerem a aplicação da equação de regressão logística devido às consistências de outras validações do modelo.

Este resultado não invalida a utilização da equação devido aos 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 mostra a Tabela 7, a seguir.

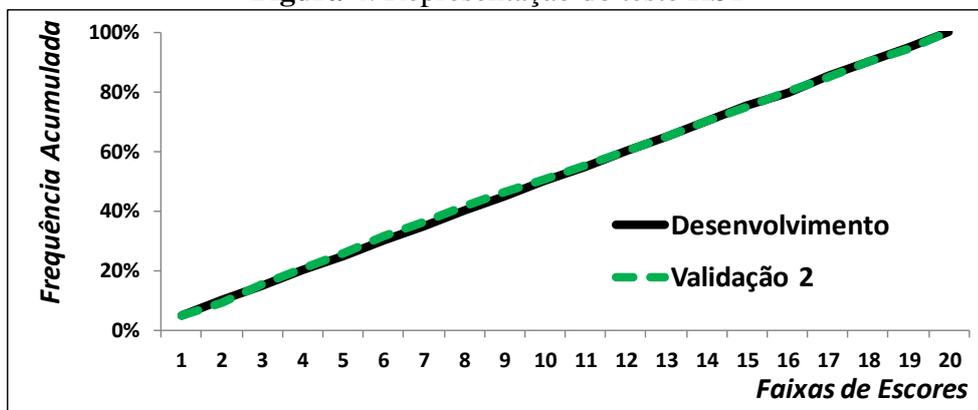
Tabela 7: Testes Bases para Validação

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 7 podem ser observados resultados amplamente satisfatórios, com destaque para o KS1 de 1,5% na base validação 2, conforme é visualizado na Figura 4. Para o KS1 quanto mais próximas forem as frequências, melhor será a aderência da equação.

Figura 4: 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.

Os dados da equação gerada e dos escores obtidos para todos os clientes fornecerá os resultados efetivos do modelo *credit scoring*. Esses resultados serão visualizados ao associar os números gerados pela equação aos dados de inadimplência e valor do cliente fornecidos pela Instituição Financeira, como também os custos dos *Bureaus* de Crédito e da consultoria para modelagem do *credit scoring*. A Tabela 8, abaixo, apresenta os dados da inadimplência mensal estimada e rentabilidade por faixa pelo volume de casos em cada categoria e faixas de escore, de acordo com as quantidades de clientes bons e maus. A inadimplência média de 16,09% da Instituição Financeira no período de desenvolvimento do modelo será utilizada para projetar e simular prováveis desempenhos.

Tabela 8: Inadimplência e Rentabilidade por Faixa

	Faixa Escore	Taxa de Maus*	Taxa de Maus**	Var. % Maus**	Inadimplência Mensal Estimada	Rentabilidade de 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.

A Tabela 8 indica que a taxa de Maus na carteira de clientes utilizada 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 (lembrando que um cliente bom possui valor de R\$ 793,76 e o cliente mau possui valor equivalente R\$ -1.235,48).

Ainda na Tabela 8, o modelo de *credit scoring* gerado indica nas cinco primeiras faixas de escore valores negativos na coluna “Rentabilidade por Faixa”. Isso mostra que a maioria dos clientes nestas faixas geram prejuízo. Os clientes posicionados a partir da faixa seis fornecem retorno positivo em sua maioria. Removendo as faixas de prejuízo tem-se que o modelo proporciona melhores indicadores para Inadimplência e Rentabilidade do produto Cartão de Crédito proveniente da oferta para clientes consignados.

Na tabela 9 tem-se um resumo para tomada de decisão, com números inferidos pela modelagem obtida, as informações fornecidas pela Instituição Financeira e os custos utilizados de itens de mercado bem como o valor da consulta aos *Bureaus* e consultoria para um período com 12 safras, gerando 4 cenários de observação.

No primeiro cenário, denominado “Sem Escore”, a concessão do novo produto é mantida como está, não gerando custos adicionais. Tal cenário apresenta uma rentabilidade de R\$ 585 mil e uma inadimplência de 16,1%, com 100% dos créditos aprovados.

No segundo cenário, denominado *Bureau 1*, são utilizados apenas os escores pelo *Bureau 1*, com um custo adicional na produção em torno de R\$ 45 mil. Com a remoção das faixas de escore com perda observa-se uma rentabilidade de R\$ 2,3 milhões, uma inadimplência estimada de 8,0%, KS2 de 48,7%, AUROC de 0,809 e taxa de aprovação de 62% dos créditos.

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. Com a remoção das faixas de escore com perda observa-se uma rentabilidade de R\$ 2,55 milhões, uma inadimplência estimada de 7,3%, KS2 de 54,7%, AUROC de 0,848 e taxa de aprovação de 64% dos créditos.

No quarto cenário, denominado “Consultoria”, são utilizados os escores gerados pelo *credit scoring* estimado por meio da regressão logística. Nesse caso, considerou-se que o modelo 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 em função da utilização dos escores dos *Bureaus 1* e *2*.

Com a remoção das faixas de escore com perda observa-se uma rentabilidade de R\$ 2,56 milhões, uma inadimplência estimada de 8,5%, KS2 de 55,6%, AUROC de 0,867 e taxa de aprovação de 75% dos créditos.

Tabela 9: 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.

Analisando os quatro cenários sugere-se que a Instituição Financeira não deve conceder cartões de crédito a partir da carteira de clientes consignado por meio do seu processo sem utilização escore, isto é, deve abandonar o cenário “Sem Escore”.

Todos os demais escores fornecem um aumento da rentabilidade no mínimo quatro vezes maior. Dado o volume de dados da amostra e a forte aderência dos escores externos fornecidos pelos *Bureaus*, fica a critério da Instituição Financeira escolher apenas o melhor deles. No entanto, 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.

O modelo de *credit 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 escores 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 Instituição Financeira.

5. CONCLUSÕES

O objetivo neste trabalho foi analisar se o desenvolvimento de um modelo de *credit scoring* visando ações de *cross-selling* teria condições de promover um incremento na rentabilidade. De forma específica foi realizada uma modelagem convencional de *credit scoring* em uma carteira de crédito consignado de uma Instituição Financeira visando a oferta de cartões de crédito para essa base de clientes. Os dados da carteira de crédito consignado foram cedidos por uma Instituição Financeira em caráter de confidencialidade.

A hipótese central do trabalho é que os investimentos realizados na construção do modelo de *credit scoring* em uma carteira de crédito consignado para fins de ações de *cross-selling* (oferta de cartões de crédito) podem ser justificados pelo incremento na rentabilidade. Como o crédito consignado é um produto que possui uma garantia diferenciada, sobretudo quando se trata de funcionários públicos, aposentados e pensionistas, a oferta de um novo produto de crédito com risco distinto não deveria ser feita sem o suporte de um modelo de *credit scoring*.

Com base nos dados da carteira de crédito consignado foi estimado um modelo de *credit scoring* por meio da aplicação do método de regressão logística. O modelo foi validado pelos testes Wald, KS e Curva ROC e apresentou uma eficiência de 80,9% nos acertos gerais entre as observações e previsões.

Uma vez estimado o modelo de *credit scoring* foram gerados quatro cenários para verificar a rentabilidade e a inadimplência da estratégia de *cross-selling*. No primeiro cenário, a oferta do cartão de crédito é realizada sem o uso de escores; já no segundo e terceiro cenários são utilizados escores fornecidos por dois *Bureaus* de Crédito distintos. Por fim, no quarto cenário foram utilizados os escores produzidos pelo modelo de *credit scoring* estimado por meio de regressão logística.

Os resultados mostraram que, no primeiro cenário, sem uso de escores, a Instituição Financeira apresenta uma rentabilidade de R\$ 0,5 milhões e índice de inadimplência de 16,1%. Nos demais cenários as rentabilidades ultrapassam R\$ 2,3 milhões, quatro vezes maior que no primeiro cenário e o índice de inadimplência permanece abaixo de 9%, aproximadamente a metade do primeiro cenário. O segundo e terceiro cenários utilizam somente o escore de Bureaus de Crédito e apresentam resultados satisfatórios. No entanto, o quarto cenário foi aquele que apresentou a maior rentabilidade para a Instituição Financeira e a melhor discriminação entre bons e maus clientes, com 75% de aprovação de clientes para o novo produto, contra 64% do melhor Bureau. Tais resultados justificam a realização de investimentos por parte de Instituições Financeiras para o desenvolvimento de modelos de *credit scoring* visando ações de *cross-selling*.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABDOU, H. A.; POINTON, J. Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature. John Wiley & Sons: Intelligent Systems in Accounting, Finance and Management, v. 18, pp 59-88, june/2011.
- AGARWAL, N.; SHARMA M. Fraud Risk Prediction in Merchant-Bank Relationship using Regression Modeling. Vikalpa: Journal for Decision Makers, v. 39, n. 3, July - September 2014.
- ARAÚJO, E. A.; CARMONA, C. U. M.; AMORIN NETO, A. Aplicação de modelos credit scoring na análise da inadimplência de uma instituição de microcrédito. Rev. Ciênc. Admin., Fortaleza. Vol.13: pág. 110-121, 2007.
- 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, v. 39, issue 1-2, pp 1-17, abril/2011.
- BERRY, M. J.; LINOFF, G. S. Data Mining Techniques for Marketing, Sales, and Customer Relationship Management. 2 ed. New York: John Wiley & Sons, 2004.
- BOTELHO, D.; TOSTES, F. D. Modelagem de Probabilidade de Churn. São Paulo: Revista de Administração de Empresas da FGV, v. 50, n. 4, p. 396-410, outubro-dezembro, 2010.
- BRESSAN, V. G. F.; BRAGA, M. J.; BRESSAN, A. A.; RESENDE FILHO, M. A. Avaliação de insolvência em cooperativas de crédito: uma aplicação do sistema pearls. RAM – Revista de Administração Mackenzie, v. 12, n. 2, mar/abr 2011.
- BRITO, G. A. S.; ASSAF NETO, A. Modelo de Classificação de Risco de Crédito de Empresas. R. Cont. Fin., USP, São Paulo, v. 19, n. 46, pág. 18 a 29, janeiro/abril 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, setembro/dezembro 2009.
- BRUNI, M. E.; BERALDI, P.; IAZZOLINO, G. Lending decisions under uncertainty: a DEA approach. International Journal of Production Research, vol. 52, No. 3, pp 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, v. 19, n. 3, pp. 467-486, jul./set. 2012.
- CAOUCETTE, J. B.; ALTMAN, E. I.; NARAYANAN, P. Gestão do Risco de Crédito: o próximo grande desafio financeiro. Rio de Janeiro: Qualitymark Editora, 1999.
- CAPRIO, G.; HONOHAN, P. Finance for Growth: Policy Choices in a Volatile World. World Bank and Oxford University Press. Washington and New York, 2001.
- DINIZ, C.; LOUZADA, F. Modelagem Estatística para Risco de Crédito. 20º SINAPE, João Pessoa – PB: ABE, 2012.

EINAV, L.; JENKINS, M.; LEVIN, J. The impact of credit scoring on consumer lending. *The RAND Journal of Economics*, Vol. 44, No. 2, pp. 249–274, Summer 2013.

FERREIRA, R. S.; CARMO, C. R. S.; MARTINS, V. F.; SOARES, A. B. 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*, Vol. 20, n. 03, p. 325-344, jul/set 2013.

GONÇALVES, E. B.; GOUVÊA, M. A.; MANTOVANI, D. M. N. Análise de risco de crédito com o uso de regressão logística. Belo Horizonte: *Revista Contemporânea em Contabilidade – UFSC*, v. 10, n. 20, p. 139-160, mai./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. Belo Horizonte: *Revista Contabilidade Vista & Revista da Universidade Federal de Minas Gerais*, 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. *Elsevier: Expert Systems with Applications*, v. 39, Pages 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. São Paulo: *Revista de Administração da Mackenzie*, v.15, n. 1, p. 125-151, Janeiro-Fevereiro, 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*, vol. 49, no. 6, pp. 74–96, November–December 2011.

KOCENDA, E.; VOJTEK, M. Default Predictors in Retail Credit Scoring: Evidence from Czech Banking Data. *Emerging Markets Finance & Trade*, Vol. 47, No. 6, pp. 80–98, November–December 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*, Vol. 11, n. 3, pp. 216–234, 2004.

MANFIO, Fernando. *O Risco Nosso de Cada Dia*. São Paulo: Estação das Letras e Cores Editora, 2007.

MEDINA, R. P.; SELVA, M. L. M. Análisis del Credit Scoring. *RAE – Revista de Administración de Empresas*, v. 53, n. 3, pp. 303-315, mai/jun 2013.

MENDES, A. C.; FARD, N. Binary Logistic Regression And PHM Analysis For Reliability Data. *International Journal of Reliability, Quality and Safety Engineering*, Vol. 21, No. 5, 2014.

ORMECI, E. L.; AKSIN, O. Z. Revenue Management Through Dynamic Cross Selling in Call Centers. *Journal Production and Operations Management*, Vol. 19, No. 6, pp. 742–756, November–December 2010.

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

RIBEIRO, C. F.; ZANI, J.; MACAGNAN, C. B.; ZANINI, F. A. M. Gestão de Universidade Privada: Risco de Inadimplência dos Alunos. São Paulo: *Revista Brasileira de Gestão de Negócios da FECAP*, v. 14, n. 42, p. 26-40, janeiro/março, 2012.

SABATO, G. Assessing the Quality of Retail Customers: Credit Risk Scoring Models. *The IUP Journal of Financial Risk Management*, Vol. VII, Nos. 1 & 2, pp 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 e Finanças, USP*, São Paulo, n. 44, p. 105-117, maio/agosto 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, abril-junho 2009.

- SICSÚ, A. L. Credit Scoring: desenvolvimento, implantação, acompanhamento. 1 ed. São Paulo: Blucher, 2010.
- SMITH, L. D.; STATEN, M.; EYSSELL, T.; KARIG, M.; FREEBORN, B. A.; GOLDEN, A. Accuracy of Information Maintained by US Credit Bureaus: Frequency of Errors and Effects on Consumers' Credit Scores. *The Journal of Consumer Affairs*, v. 47, n. 3, pp 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*, Vol. 11, No. 2, pp. 379-396, May/Aug., 2014.
- SOUZA, R. B. O Modelo de Collection Scoring como Ferramenta para a Gestão Estratégica do Risco de Crédito. Dissertação de Mestrado, FGV. São Paulo: 2000.
- VASCONCELLOS, R. S. Modelos de Escoragem de Crédito Aplicados a Empréstimo Pessoal com Cheque. Dissertação de Mestrado, FGV. Rio de Janeiro: 2004.
- VICENTE, E. F. R. A Estimativa do Risco na Constituição da PDD. Dissertação de Mestrado, USP. São Paulo: maio de 2001.
- VIEIRA, V.; PIRES, D.; GALEANO, R. Determinantes do desempenho empresarial e das vendas cruzadas no varejo. *RAE - Revista de Administração de Empresas*, Vol. 53, n. 6, p565-579, nov-dez 2013.
- WANG, G; MA, J.; HUANG, L.; XU, K. Two credit scoring models based on dual strategy ensemble trees. *Elsevier: Knowledge-Based Systems*, v. 26, Pages 61–68, February 2012.