

SELEÇÃO DE CLIENTES PARA AÇÕES DE COBRANÇA ATRAVÉS DE REGRESSÃO LOGÍSTICA NA INADIMPLÊNCIA DO CONSUMO DE ENERGIA ELÉTRICA

RENATA V. ARAÚJO, LUCIANA MARTINEZ, FERNANDO AUGUSTO MOREIRA.

Universidade Federal da Bahia

Rua Aristides Novis, nº 2, Federação, Salvador-BA, Brasil, 40210-630

E-mails: valente.araujo@gmail.com, lucianam@ufba.br, moreiraf@ufba.br

Abstract— This paper presents a methodology for selection of customers with electricity consumption bills late for the proposition of collection actions. Variables determined from business information such as debt amount, overdue invoices and time debt of residential customers of the Company of the State of Bahia Electricity (COELBA) are used in this study. By logistic regression, it was possible to build a model to support intelligent decision in favor of combating delinquency phenomenon.

Keywords— Delinquency, Logistic Regression, Intelligent Decision Making, Collections

Resumo— Este trabalho propõe uma metodologia de seleção de clientes com faturas de consumo de energia elétrica em atraso para proposição da ação de cobrança de suspensão do fornecimento. Variáveis determinadas a partir de informações comerciais, como montante da dívida, quantidade de faturas vencidas e tempo da dívida de consumidores residenciais da Companhia de Eletricidade do Estado da Bahia (COELBA) são utilizadas como estudo de caso. Através da ferramenta de regressão logística foi possível a construção de um modelo de apoio à tomada de decisão inteligente em favor do combate ao fenômeno da inadimplência.

Palavras-chave— Cobrança, Inadimplência, Regressão Logística, Tomada de Decisão Inteligente.

1 Introdução

De maneira geral, a inadimplência é um risco existente em qualquer negócio que ofereça produtos ou serviços ao consumidor, para pagamento em um período posterior ao seu recebimento. Por este motivo, empresas de diferentes setores demonstram grande interesse no desenvolvimento de ferramentas que auxiliem a previsão do comportamento de seus clientes, visando mitigar o impacto deste tipo de perda.

O modelo de negócio das concessionárias distribuidoras de energia elétrica brasileiras faz com que o risco de inadimplência de todos os seus consumidores seja invariavelmente aceito. Ao contrário de outros setores que enfrentam a mesma questão, a concessionária não pode negar o fornecimento para clientes que apresentem histórico de inadimplência. Além disso, não existe um mecanismo que permita limitar o consumo do usuário.

Por outro lado, estas empresas possuem abrangentes bancos de dados dotados de informações associadas às suas rotinas operacionais, que apresentam o comportamento histórico dos seus clientes e podem auxiliar no levantamento de variáveis físicas, econômicas e sociais que auxiliem na construção de sistemas suporte à decisão.

Desta forma, este trabalho apresenta a aplicação da regressão logística aos dados de clientes inadimplentes, de maneira a selecionar as melhores oportunidades de cobrança e uma melhor distribuição dos recursos disponíveis para estas ações. A regressão logística é comumente encontrada em pesquisas e estudos

acadêmicos relacionados à inadimplência, sendo também aplicada nos modelos de *credit scoring*¹ adotados por instituições financeiras.

A identificação de perfis de comportamento de pagamento de consumidores de energia através das técnicas estatísticas de regressão logística binária e análise de agrupamentos proposta por Silva (Silva, 2008), obteve um modelo com um nível de assertividade em torno de 84% em uma série histórica de seis meses. No estudo desenvolvido, foram agrupados clientes adimplentes, clientes que quitam as faturas com atraso, mas sem a necessidade de qualquer ação de cobrança, e clientes que só realizam o pagamento após a suspensão do fornecimento.

O trabalho desenvolvido por Selau (Selau, 2008) construiu modelos com as técnicas estatísticas de análise discriminante, regressão logística e redes neurais para a previsão de risco de concessão de crédito. O estudo propõe um método de seis etapas que contemplam desde a delimitação da amostra até os passos necessários para sua implantação.

Uma análise de dados de instituições financeiras foi proposta por Gonçalves (2005), com três modelos de classificação de clientes. Os modelos baseados em redes neurais e regressão logística apresentaram resultados satisfatórios, sendo o segundo ligeiramente superior. Outras aplicações de modelos de regressão logística podem ser vistas em Bruni (2007), Mendonça (2008) e Silva (2012). As referências apresentadas reforçam a aplicabilidade do modelo aqui adotado para a melhoria do processo de alocação de recursos de cobrança.

¹ Modelos quantitativos empregados comumente por instituições financeiras na mensuração e previsão do risco de crédito.

Portanto, diante do cenário apresentado, este trabalho tem como objetivo analisar os dados de consumidores de energia elétrica a fim de identificar fatores que contribuam para a tomada de decisão no processo de cobrança, como base em métodos de regressão logística. Dados de consumidores residenciais da Companhia de Eletricidade do Estado da Bahia (COELBA) foram utilizados como estudo de casos. Vale ressaltar que a identidade dos consumidores foi mantida em sigilo absoluto.

A COELBA detém o serviço de distribuição de energia em todo o território do estado da Bahia, atendendo, no ano de 2015, um número em torno de 5,5 milhões de consumidores ativos. Em 2014, a concessionária registrou uma inadimplência de 3,6% do seu faturamento.

2 O Problema

Para combater as perdas de faturamento decorrentes da inadimplência das faturas de energia elétrica, a COELBA possui uma política de cobrança que inclui uma série de ações que objetivam a recuperação financeira. Estas ações de cobrança são práticas consolidadas no mercado e estão alinhadas com a política regulatória vigente.

A ação de cobrança conhecida como "corte" consiste em interromper a conexão do cliente com o sistema de distribuição. A suspensão do fornecimento de energia elétrica, por se caracterizar como um serviço essencial, provoca grandes transtornos ao consumidor, limitando suas atividades comerciais, facilidades e conforto. Por este motivo, esta ação de cobrança é usualmente considerada como a mais efetiva e direta dentre todas as demais.

No entanto, a realização da desconexão física do consumidor atualmente requer atuação de uma turma composta de dois eletricitas *in loco*, requerendo assim um esforço logístico considerável.

Sendo assim, se faz necessária a seleção de consumidores que serão atingidos pela suspensão de fornecimento, de forma a adequar as operações ao orçamento e infraestrutura disponíveis.

2.1 Seleção de Variáveis

A despeito das diversas variáveis comerciais disponíveis no sistema da empresa, atualmente o modelo de seleção da Coelba realiza a seleção dos consumidores para "corte" apenas através do montante da dívida.

Em Araujo et al. (2014) foi realizado um estudo estatístico para a seleção de variáveis comerciais relacionadas com o fenômeno da inadimplência em clientes do serviço de distribuição da energia elétrica.

Inicialmente, foram selecionadas doze variáveis usualmente associadas ao atraso de pagamento e sua reincidência, dentre todo o universo de dados disponível nos sistemas comerciais: ocorrência de parcelamento na dívida atual; classificação baixa renda; situação do contrato (ativo ou suspenso);

situação da instalação (ativa ou encerrada); ocorrência de faturas de fraude na dívida atual; modalidade de pagamento da última fatura; padrão de fornecimento; número de faturas vencidas; montante vencido; tempo da dívida; histórico de ações de cobrança sofridas e número de dias decorridos desde a última ação de cobrança.

A identificação destas variáveis utilizou o conceito de Coeficiente de Correlação de *Pearson* e a metodologia bivariada da Análise de Componentes Principais.

De forma geral, os índices encontrados mostraram correlações lineares fracas entre as variáveis. No entanto, a análise de componentes principais demonstrou que um modelo com até quatro componentes principais se mostrou adequado para a representação dos dados.

Assim, baseado no estudo realizado em Araujo et al. (2014), foram selecionadas aqui as seguintes variáveis relacionadas à inadimplência:

1. Número de faturas vencidas (f);
2. Montante total da dívida (m);
3. Tempo da dívida (t).

2.2 Amostra de Estudo

A Resolução Normativa nº414/2010 (ANEEL, 2010) prevê a aplicação da suspensão de fornecimento de energia elétrica por inadimplemento. No presente trabalho, o banco de dados disponível para a aplicação do modelo consiste no universo de 5.000 consumidores que tiveram o fornecimento suspenso pela COELBA no mês de março de 2015.

A localização geográfica do consumidor é uma informação que pode contribuir para a identificação do seu perfil socioeconômico, associando-o a um padrão de pagamento e também de sensibilidade às ações de cobrança. Sendo assim, devido à abrangência da área de concessão da COELBA e sua consequente diversidade de comportamento, os dados selecionados referem-se à instalações residenciais situadas na capital do estado.

A amostra de teste do modelo implementado consiste em 7.000 consumidores que sofreram esta ação de cobrança no mesmo período, selecionados de forma aleatória, sem mecanismos que pudessem possibilitar a identificação das suas respectivas identidades.

Assim, foram excluídos contratos com dívidas que apresentavam algum tipo de bloqueio de cobrança, pois tais contratos estão formalmente resguardados de determinadas obrigações e são beneficiados por prerrogativas. Essas dívidas, por não estarem submetidas ao processo de cobrança vigente na concessionária, apresentam um comportamento destoante do padrão de inadimplência da empresa.

Analogamente, foram excluídos clientes com dívidas mais antigas que cinco anos, pois conforme o artigo 206 do Código Civil, após este prazo o direito à aplicação de ações de cobrança é prescrito.

Das amostras selecionadas ainda foram excluídos clientes com dívidas inferiores ao valor mínimo de faturamento, por não serem consideradas dívidas provenientes de consumo.

3 Metodologia

A regressão logística é aplicada quando se deseja identificar a influência de variáveis em um determinado processo. Sendo assim, sua modelagem tem como objetivo definir uma equação matemática que descreva este relacionamento a partir de um conjunto de observações (Bruni, 2007).

O modelo considera uma variável de saída e variáveis independentes, causais ou explicativas para descrever este comportamento. As variáveis deste estudo foram definidas de acordo com os dados disponíveis no sistema comercial da concessionária.

A aplicação da regressão logística ao problema que se deseja modelar consiste em estabelecer o evento dicotômico de interesse, como por exemplo a existência ou não de risco de inadimplência, como variável de saída. A partir daí, se sua probabilidade prevista for superior à 0,5, a hipótese de ocorrência é positiva; caso contrário, será negativa.

Na regressão logística, a variável de saída, também chamada de variável de resposta categórica, consiste na probabilidade de um evento ocorrer. Para tal, é determinada uma função com variáveis independentes associadas ao evento, que podem ser contínuas, binárias ou até qualitativas. Apesar dos valores assumidos pelas variáveis independentes, a solução dada pelo modelo será apresentada com uma curva de probabilidade da ocorrência da variável de saída. Ou seja, seu valor pode variar no intervalo entre 0 e 1, correspondendo portanto a uma distribuição binomial e não normal.

A sua representação gráfica em forma de “s”, ilustrada na Figura 1, consiste em uma distribuição sigmoide que pode assumir diversos comportamentos a depender da inclinação imposta pelos coeficientes da função (Braga, 2005).

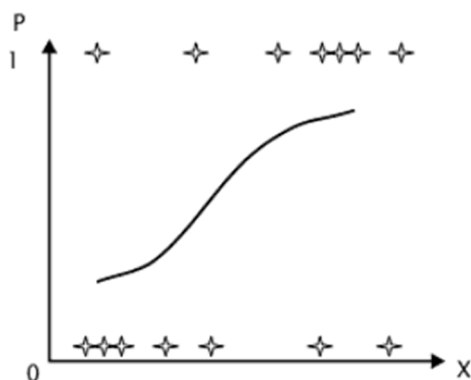


Figura 1. Forma da relação logística entre variáveis dependente e independentes - Fonte: (Braga, 2005).

4 Resultados

Para a aplicação do modelo de Regressão Logística computacionalmente, foi utilizado algoritmo *logit* existente na biblioteca do *software* livre de computação estatística "R Project", através do pacote "glm".

4.1 Aplicação do Modelo

No modelo de regressão logística, assume-se que o logaritmo das chances é linearmente relacionado com as variáveis explicativas, ou seja,

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_n \cdot X_n$$

Sendo, p = probabilidade de ocorrer a inadimplência;

$(1 - p)$ = probabilidade de ocorrer a não inadimplência;

X_i = variáveis preditoras;

b_i = coeficientes atribuídos a cada uma das variáveis preditoras.

Inicialmente, a amostra de dados obtida das variáveis de inadimplência: f (número de faturas vencidas), m (montante total da dívida) e t (tempo da dívida) foi aplicada à função logística, sendo obtidos os dados apresentados na Tabela 1.

Tabela 1. Resultados Encontrados $\text{logit}(f, t, m)$.

Variável	Coeficiente	Erro Padrão	Z	Pr
intercepto	$7,15 \cdot 10^{-1}$	$6,75 \cdot 10^{-2}$	10,59	$< 2 \cdot 10^{-16}$
f	$-1,97 \cdot 10^{-1}$	$1,47 \cdot 10^{-2}$	-13,39	$< 2 \cdot 10^{-16}$
t	$5,20 \cdot 10^{-4}$	$9,34 \cdot 10^{-5}$	5,57	$2,59 \cdot 10^{-8}$
m	$1,48 \cdot 10^{-5}$	$4,99 \cdot 10^{-5}$	0,30	0,766

Os erros padrão apresentados pelo algoritmo são uma estimativa de quanto, em média, esses coeficientes calculados podem variar quando novos dados forem aplicados ao modelo. Se dividirmos a estimativa do coeficiente por seu respectivo erro padrão, obtemos o *score* discriminante Z. Na última coluna, a variável Pr corresponde aos valores de Z normalizados em uma distribuição normal.

Dentro deste conjunto de parâmetros de saída, os resíduos médios são componentes que podem indicar erros no ajuste do modelo encontrado. Para este modelo foi encontrado um resíduo médio de -0,0101, sendo assim, encontramos um resíduo usual que indicando a qualidade de ajuste da metodologia escolhida.

Na Figura 2, é apresentada graficamente a função resposta, onde se pode verificar sua similaridade com a forma sigmoide da curva de regressão logística. Vale observar que a escala dos eixos se encontra normalizada de acordo com o padrão estabelecido pelo software, não apresentando os valores absolutos

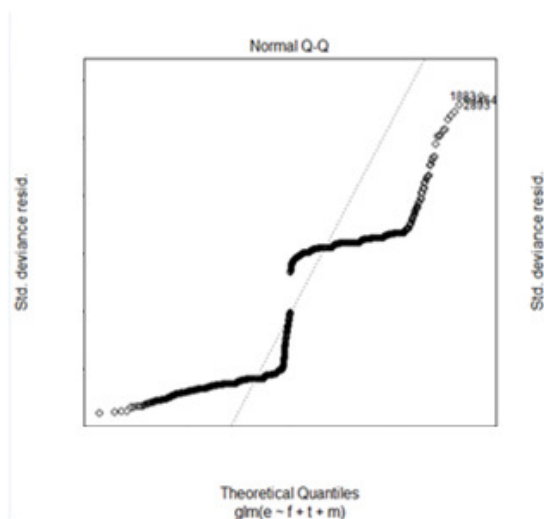


Figura 2. Gráfico da Função Resposta $\text{logit}(f,t,m)$

Sequencialmente, foi avaliada a relação logística entre cada par de variáveis independentes e a variável dependente. Sendo assim, o número de variáveis preditivas do modelo foi reduzido, de forma a verificar os resultados encontrados com a diminuição dos graus de liberdade da equação. A seguir são apresentados os resultados utilizando apenas as variáveis m e t .

Tabela 2. Resultados Encontrados $\text{logit}(t,m)$.

Variável	Coeficiente	Erro Padrão	Z	Pr
intercepto	$8,91 \cdot 10^{-2}$	$3,97 \cdot 10^{-2}$	2,25	0,0248
t	$2,85 \cdot 10^{-5}$	$8,06 \cdot 10^{-5}$	0,35	0,7238
m	$-2,78 \cdot 10^{-4}$	$5,46 \cdot 10^{-5}$	-5,09	$3,51 \cdot 10^{-7}$

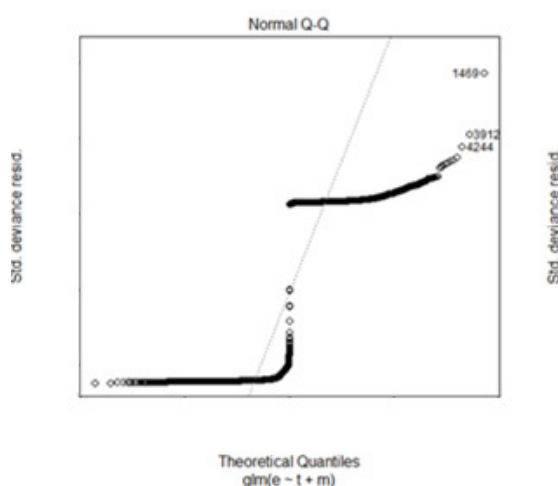


Figura 3. Gráfico da Função Resposta $\text{logit}(t,m)$

Seguem abaixo os resultados utilizando apenas as variáveis m e f .

Tabela 3. Resultados Encontrados $\text{logit}(f,m)$.

Variável	Coeficiente	Erro Padrão	Z	Pr
intercepto	$8,11 \cdot 10^{-1}$	$6,62 \cdot 10^{-2}$	12,27	$<2 \cdot 10^{-16}$
f	$-1,87 \cdot 10^{-1}$	$1,48 \cdot 10^{-2}$	-12,66	$<2 \cdot 10^{-16}$
m	$3,44 \cdot 10^{-5}$	$4,93 \cdot 10^{-5}$	0,70	0,486

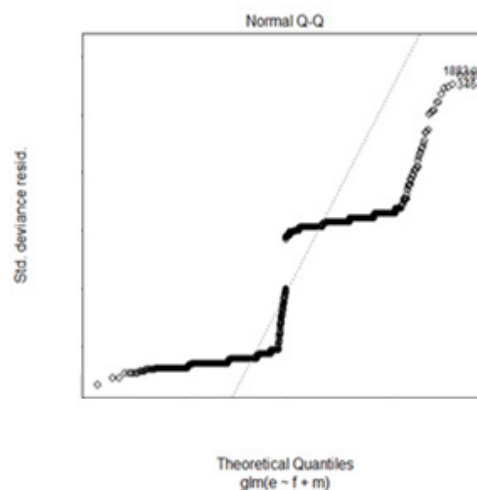


Figura 4. Gráfico da Função Resposta $\text{logit}(f,m)$

Por último, os resultados utilizando apenas as variáveis t e f .

Tabela 4. Resultados Encontrados $\text{logit}(t,f)$.

Variável	Coeficiente	Erro Padrão	Z	Pr
intercepto	$7,17 \cdot 10^{-1}$	$6,73 \cdot 10^{-2}$	10,65	$<2 \cdot 10^{-16}$
t	$5,22 \cdot 10^{-4}$	$9,31 \cdot 10^{-5}$	5,60	$2,14 \cdot 10^{-8}$
f	$-1,97 \cdot 10^{-1}$	$1,46 \cdot 10^{-2}$	-13,48	$<2 \cdot 10^{-16}$

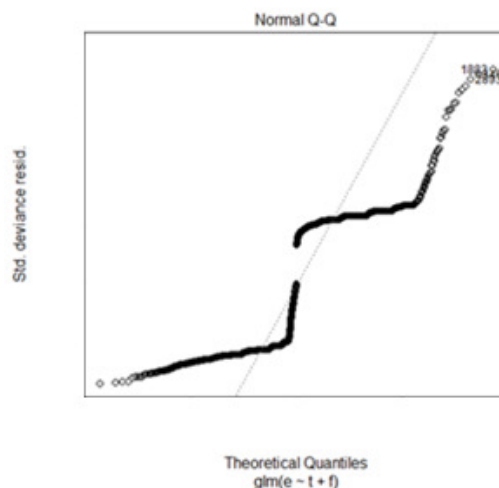


Figura 5. Gráfico da Função Resposta $\text{logit}(t,f)$

Inspecionando visualmente as funções apresentadas nas figuras 2 a 5, observamos que os gráficos possuem grande similaridade, sendo que o modelo $\text{logit}(t, m)$ apresenta maior distorção com relação à forma sigmoide.

Avaliando a qualidade do ajuste dos modelos com número de variáveis reduzidas, observamos na Tabela 5 que os resíduos médios encontrados se aproximam do resíduo médio da primeira função testada, $\text{logit}(f, t, m)$. O modelo $\text{logit}(t, m)$ apresenta resíduo médio superior aos demais, no entanto, ainda assim, podemos afirmar que a explicação dos modelos encontrados é equivalente, portanto todos os resultados podem ser considerados ajustados.

Tabela 5. Análise de Ajuste dos Modelos.

Modelo	Resíduo médio
$\text{logit}(f, t, m)$	-0,0101
$\text{logit}(t, m)$	-0,4057
$\text{logit}(f, m)$	-0,0096
$\text{logit}(f, t)$	-0,0089

4.2. Análise dos Resultados

Os quatro modelos testados apresentam resultados equivalentes em termos de ajuste. Sendo assim, seus resultados foram analisados comparativamente com relação às suas assertividades e eficiências financeiras.

Na Tabela 6 é apresentada a recuperação absoluta das ações realizadas por quantidade de consumidores e por montante da dívida. A coluna "Dívida Paga" corresponde aos valores efetivamente pagos após a ocorrência da ação de cobrança, independente da dívida total de cada consumidor. A primeira linha da Tabela 6 apresenta os dados que foram utilizados para testar os modelos construídos e consiste em um grupo de 7.000 clientes que tiveram o fornecimento suspenso. Nos modelos propostos, os consumidores cortados foram selecionados dentro deste conjunto de 7.000 consumidores, portanto, naturalmente eles apresentam resultados quantitativos reduzidos se comparados ao universo realizado.

Tabela 6. Resultados Absolutos.

Variável	Consumidores Acionados	Consumidores que pagaram	Dívida Acionada	Dívida Paga
Universe	7000	3217	R\$2.431.169,05	R\$ 633.609,02
$\text{logit}(f, m, t)$	997	399	R\$ 819.243,58	R\$ 218.659,15
$\text{logit}(m, t)$	91	28	R\$ 423.490,79	R\$ 139.501,91
$\text{logit}(f, m)$	750	253	R\$ 813.374,06	R\$ 197.273,98

$\text{logit}(f, t)$	1002	398	R\$ 862.452,78	R\$ 188.949,60
----------------------	------	-----	----------------	----------------

Vale observar que o modelo $\text{logit}(m, t)$ selecionou para corte apenas 1% dos consumidores, enquanto os outros modelos realizaram seleções bem superiores, variando entre 9 e 13%. No entanto, quando verificamos o montante da dívida selecionado, apenas 1% de consumidores selecionados pelo modelo $\text{logit}(m, t)$ representam 15% da dívida, enquanto as dívidas acionadas pelos demais modelos variaram entre 28 e 30%.

Na Tabela 7 são apresentadas três diferentes formas de aferir a efetividade dos modelos propostos. A coluna "Acertos - Quantidade Consumidores" traz o percentual de consumidores selecionados que efetivamente realizaram o pagamento após a suspensão de fornecimento. A coluna "Acertos - Montante" apresenta o percentual da dívida dos consumidores selecionados que de fato foi recuperada após a ação de cobrança. A razão entre o montante recuperado e o investimento realizado nas ações de cobrança é apresentado na coluna "R\$ Recuperados/ R\$ Gastos".

Tabela 7. Resultados Percentuais.

Variável	Acertos - Quantidade Consumidores	Acertos - Montante	R\$ Recuperados/ R\$ Gastos
Universe	46%	26%	R\$ 1,62
$\text{logit}(f, m, t)$	40%	27%	R\$ 3,92
$\text{logit}(m, t)$	33%	31%	R\$ 27,43
$\text{logit}(f, m)$	34%	24%	R\$ 4,71
$\text{logit}(f, t)$	40%	22%	R\$ 3,37

Considerando que os valores de cada corte e sua respectiva religação, ambos executados pelas empreiteiras prestadoras de serviço e pagos pela COELBA, resultam em R\$ 55,89, na Tabela 8 é apresentado os custos das ações de cobrança do universo estudado e de cada um dos modelos.

Tabela 8. Avaliação dos Custos.

Modelo	Custos das Ações
$\text{logit}(f, t, m)$	R\$ 391.230,00
$\text{logit}(t, m)$	R\$ 55.722,33
$\text{logit}(f, m)$	R\$ 5.085,99
$\text{logit}(f, t)$	R\$ 41.917,50

A coluna "R\$ Recuperados/R\$ Gastos" apresentada na Tabela 7, corresponde aos valores apresentados na coluna "Dívida Paga" da Tabela 6 divididos pelos custos da Tabela 8. Este parâmetro relaciona os resultados obtidos com o investimento necessário para sua execução, se apresentando como uma interessante métrica de avaliação. Afinal, o método de seleção poderá ser extremamente eficiente em termos de recupe-

ração de inadimplência. Mas se o investimento necessário superar o retorno financeiro, sua utilização não será vantajosa.

No universo de 7.000 consumidores, 46% realizaram o pagamento após a suspensão do serviço o que equivale a 26% do montante da dívida total. Por outro lado, a métrica que evidencia a eficiência financeira do processo apresentou uma relação de R\$ 1,62 recuperados para cada real investido em ações de corte.

Com os dados da Tabela 7, observamos que nenhum dos modelos de regressão logística atingiu uma quantidade relativa maior de consumidores. Por outro lado, todos os modelos se apresentam vantajosos com relação à métrica financeira, uma vez que os montantes recuperados por cada R\$1,00 investidos em todos os modelos foram significativamente superiores ao universo real.

Realizando uma análise comparativa das simulações realizadas, com o objetivo de selecionar o modelo de regressão logística de maior eficiência, podemos eliminar o composto das variáveis m (montante total da dívida) e t (tempo da dívida) ($\text{logit}(m, t)$). Pois, apesar da precisão da seleção com R\$27,43 recuperados para cada R\$1,00 investido, sua seleção de consumidores foi de apenas 91, dentro de um universo de 7.000 inadimplentes. Portanto, a seleção foi muito restritiva, não apresentando uma abrangência satisfatória.

Dentre os demais modelos, de acordo com os resultados da Tabela 7, os modelos apresentam uma relativa equivalência em termos de efetividade. No entanto, aquele que utilizou as variáveis f (número de faturas vencidas) e m (montante total da dívida) ($\text{logit}(f, m)$) demonstrou melhor eficiência financeira com R\$ 4,71 recuperados para cada R\$1,00 investido em ações de suspensão de fornecimento. Sendo assim, este modelo foi considerado o modelo de melhor desempenho dentre as simulações de regressão logística.

4 Conclusão

O trabalho desenvolvido se insere em uma área de pouca pesquisa no setor elétrico. Sua relevância é bastante recente e vem crescendo conforme a regulação setorial se aperfeiçoa, forçando um movimento de ganho de competitividade nas concessionárias de energia elétrica. É neste contexto de melhoria contínua de produtividade que surge e se desenvolve este estudo.

Desta forma, o trabalho apresenta contribuições para as áreas comerciais das distribuidoras de energia elétrica na medida em que introduz o conceito de eficiência e performance no tratamento da inadimplência da distribuição, como por exemplo com a aplicação do indicador de eficiência financeira “R\$ Recuperados/R\$ Gastos”.

Os métodos de regressão logística propostos, com a utilização de três ou duas variáveis, apresentaram melhoria da eficiência financeira da ação de cobrança, ressaltando que a adoção de quaisquer das soluções propostas traria bons resultados à distribuidora de energia elétrica. Por outro lado, não se pode afirmar

que os valores obtidos na formulação deste estudo devam ser encarados como definitivos, pois os modelos se alteram de acordo com os dados de entrada e as características da inadimplência são dinâmicas, variando ao longo do tempo sob efeito de diversos fatores internos e externos à organização.

Sendo assim, a implantação de um modelo deste tipo deve ser associada a um processo contínuo de novas simulações, modificando-se inclusive as variáveis utilizadas, que podem assumir maior ou menor relevância.

Agradecimentos

Os autores agradecem à Superintendência Comercial da Companhia de Eletricidade do Estado da Bahia.

Referências Bibliográficas

- ANEEL – Agência Nacional de Energia Elétrica. (2010) Resolução Normativa nº414/2010, Disponível em <http://www.aneel.gov.br/cedoc/ren2010414.pdf>. Acesso em Mar/2014.
- Araujo, R. V. de; Martinez, L. e Moreira, F. A. (2014). Statistical Analysis of the Relationship Between Payment Behavior Variables and Delinquency in Electricity Consumption. Transmission & Distribution Conference and Exposition - Latin America (PES T&D-LA), 2014 IEEE PES
- Braga, L. P. V. (2005). Introdução à Mineração de Dados. 2ª Edição. E-papers, Rio de Janeiro. Editora. pp. 47.
- Bruni, E. S. (2007). Uso de Regressão Logística para Precificação de Credit Default Swaps. Dissertação de mestrado. Universidade de São Paulo, São Paulo – SP.
- Gonçalves, E. B. (2005). Análise de Risco de Crédito com o Uso de Modelos de Regressão Logística, Redes Neurais e Algoritmos Genéticos. Universidade de São Paulo, São Paulo – SP.
- Selau, L. P. R. (2008). Construção de Modelos de Previsão de Risco de Crédito. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul, Porto Alegre – RS.
- Silva, R. K. (2008). Modelo de avaliação do risco de Inadimplência de Consumidores de Energia Elétrica. Universidade Salvador, Salvador – BA.
- Silva, A. M. M. (2012) Técnicas de Data Mining na Aquisição de Clientes para Financiamento de Crédito Direto ao Consumidor. Dissertação de Mestrado - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba - SP.
- Mendonça, T. S. (2008) Modelos de Regressão Logística Clássica, Bayesiana e Redes Neurais para Credit Scoring. Dissertação de Mestrado, Universidade Federal de São Carlos, São Carlos - SP.