



Exacta

ISSN: 1678-5428

exacta@uninove.br

Universidade Nove de Julho

Brasil

da Silva Carvalho, Norma Alice; Castro Souza, Reinaldo; Kahn Epprecht, Eugenio
Segmentação do mercado consumidor cativo alta e média tensão de uma distribuidora de
energia elétrica: aplicação do Mapa Auto-Organizável de Kohonen para descoberta de
padrões de inadimplência no setor
Exacta, vol. 14, núm. 3, 2016, pp. 495-509
Universidade Nove de Julho
São Paulo, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=81047687012>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe , Espanha e Portugal
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

Segmentação do mercado consumidor cativo alta e média tensão de uma distribuidora de energia elétrica: aplicação do Mapa Auto-Organizável de Kohonen para descoberta de padrões de inadimplência no setor

Targeting of the captive consumer market for high and medium voltage of an electric power distributor company: application of the Kohonen Self-Organization Map for discovery of industry delinquency patterns

Norma Alice da Silva Carvalho

Doutoranda em Engenharia de Produção e Mestre em Metrologia pela Pontifícia Universidade Católica - PUC-Rio, Graduada em Matemática pela Universidade Federal Fluminense - UFF.
Rio de Janeiro, RJ [Brasil]
nasc86@gmail.com

Reinaldo Castro Souza

Doutor em Estatística pela University of Warwick, Coventry, Inglaterra, Professor Titular da Pontifícia Universidade Católica do Rio de Janeiro – PUC-RIO.
Rio de Janeiro, RJ [Brasil]

Eugenio Kahn Epprecht

Doutor em Informática pela Facultés Universitaires Notre-Dame de la Paix, Bélgica, Professor Assistente da Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio.
Rio de Janeiro, RJ [Brasil]

Resumo

O modelo de regulação incentivada, no qual operam os agentes distribuidores de energia elétrica, é constantemente ameaçado pela inadimplência, que afeta o equilíbrio econômico financeiro do mercado de distribuição. Em decorrência disso, é fundamental que os agentes distribuidores disponham de uma ferramenta metodológica que permita identificar antecipadamente o perfil de pagamento de unidades consumidoras, visando à formulação de estratégias capazes de reduzir o grau de inadimplência desse mercado. O mercado de distribuição é caracterizado pelo monopólio natural e pela predominância de consumidores cativos. Neste estudo, propõe-se o uso de uma rede neural não supervisionada para agrupar unidades consumidoras, conforme informações armazenadas em base de dados e, então, obter um padrão da inadimplência, por meio da análise exploratória de dados, considerando as características de similaridades dos grupos gerados. Os resultados mostram que esse estudo é eficaz e recomendado para descoberta de padrões em outros mercados.

Palavras-chave: Análise exploratória de dados. Inadimplência. Redes neurais artificiais. Setor elétrico.

Abstract

The incentivized regulation regime, in which electricity distribution agents operate, is frequently threatened by delinquency that affects the financial-economic balance of the distribution market. In light of this, it is essential that distribution agents have a methodological tool that allows identifying preventively customer payment profiles and thus formulating strategies for tackling delinquency in this market. The distribution market is characterized as a natural monopoly and by the predominance of captive consumers. This paper proposes the use of an unsupervised neural network for grouping consumer units according to information stored in a database and then obtaining a pattern of delinquency by exploratory data analysis, considering the similarities of the created groups. The results show that this study is effective and suitable for discovering patterns in others markets.

Key words: Artificial neural networks. Electrical sector. Delinquency. Exploratory data analysis.

1 Introdução

Estudos recentes (ARAUJO, 2007; CARVALHO; SOUZA; EPPRECHT, 2014; CARVALHO; SOUZA; FROTA, 2013a, b; SOUZA; PESSANHA; OLIVEIRA, 2013; SOUZA et al., 2011) revelam que o conhecimento inerente às operações diárias de uma distribuidora de energia elétrica tem proporcionado a compreensão de fenômenos que envolvem a atividade de distribuição – em particular, a inadimplência (montante da receita faturada e não recebida pela distribuidora).

Araújo (2007), Depuru et al. (2011), Instituto Acende Brasil (2007a) e Medeiros (2013) consideram que a inadimplência nas distribuidoras está relacionada a fatores setoriais, sociais e econômicos. Para Smith (2004), países com alto nível de corrupção, baixa estabilidade de política e eficácia do governo tendem a ter níveis elevados de perdas não técnicas¹. Ainda conforme esse autor, além dos motivos supracitados, outros podem estar associados aos casos de inadimplência, como, por exemplo: mudança de endereço, falência, falta de instrumentos que punem os infratores (principalmente, se estes forem consumidores governamentais ou que possuam alto padrão social/político).

A gestão da inadimplência é essencial para a capacidade de caixa das distribuidoras (FONSECA; REIS, 2012), o que implica no equilíbrio econômico-financeiro dessas instituições. Tal equilíbrio está diretamente relacionado ao alcance das metas de produtividade impostas pelo regime de regulação incentivada², no qual a tarifa de fornecimento de energia é definida de forma a compartilhar com as unidades consumidoras uma parcela dos ganhos de produtividade das distribuidoras (ARAUJO, 2007; INSTITUTO ACENDE BRASIL, 2007b).

A tarifa de fornecimento de energia, mensalmente arrecadada pelas distribuidoras de energia elétrica, representa o valor agregado desse setor, pois é composta por custos inerentes à compra de energia elétrica, ao uso do sistema de transmissão, aos impostos e encargos setoriais, bem como, aos custos operacionais, de remuneração do capital e de reposição dos ativos dos sistemas de distribuição.

Estima-se que, no Brasil, a inadimplência gere uma perda em torno de R\$ 1 bilhão ao ano (INSTITUTO ACENDE BRASIL, 2007b). Atualmente, a política adotada pela Agência Nacional de Energia Elétrica (ANEEL) para combatê-la, baseia-se na determinação do recolhimento de um percentual dessa perda na tarifa e na autorização da realização do corte de fornecimento pelas distribuidoras sob algumas condições. No entanto, esses instrumentos não têm sido suficientes na recuperação dessa perda visto que muitas distribuidoras possuem nível de inadimplência acima do reconhecido na composição da tarifa e/ou apresentam dificuldades para realizar a execução sistemática das atividades do corte do fornecimento dos inadimplentes (FONSECA; REIS, 2012; SOUZA et al., 2011). Em decorrência disso, estudos que abordem ações, iniciativas e novas tecnologias no combate a inadimplência são necessários (ARAUJO, 2007; FONSECA; REIS, 2012).

Esta pesquisa se insere nesse contexto e, nesta, objetivou-se propor o uso de uma rede neural não supervisionada para segmentar o mercado consumidor de uma distribuidora de energia elétrica, conforme informações armazenadas em base de dados e, então, obter um padrão da inadimplência, por meio da análise exploratória de dados, considerando as características de similaridades que segregaram esse mercado em grupos. Esse artigo encontra-se organizado como segue: na primeira seção, tem-se a introdução, que

caracteriza e descreve o problema a ser estudado; na segunda seção, descrevem-se a metodologia e a base conceitual usada para segmentar e obter um padrão da inadimplência de unidades consumidoras dos sistemas de distribuição de energia elétrica; na terceira, apresentam-se e discutem-se os resultados da aplicação da metodologia proposta; e na quarta são apresentadas as principais conclusões.

2 Formulação metodológica

Nesta pesquisa, propõe-se a obtenção de um padrão da inadimplência de unidades consumidoras dos sistemas de distribuição de energia, por meio da análise de dados da segmentação desse mercado consumidor, dado um conjunto de informações armazenadas em banco de dados de um sistema de distribuição de energia. Assim, conforme definições de Vergara (2013), esta pesquisa classifica-se como: bibliográfica, documental, descritiva, aplicada, metodológica e de investigação *ex post fato*. Essencialmente, constitui-se em um estudo de caso com características, predominantemente, quantitativas. Nesta seção, a base conceitual do estudo é apresentada.

2.1 Sistemas de distribuição de energia elétrica

Os sistemas de distribuição de energia elétrica são constituídos por empresas distribuidoras de energia elétrica – agentes titulares de concessão ou permissão de energia que integram uma das atividades de valor³ – que conectam e atendem o mercado consumidor ao setor elétrico (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA – ANEEL, 2008).

O mercado consumidor de energia elétrica é composto por pessoas físicas ou jurí-

dicas que solicitam o fornecimento de energia e assumem a responsabilidade pelo pagamento das faturas e outras obrigações fixadas pela ANEEL (ASSOCIAÇÃO BRASILEIRA DE DISTRIBUIDORES DE ENERGIA ELÉTRICA, s.d.). Essencialmente, esse mercado é composto por consumidores cativos que, de modo obrigatório, se conectam ao agente distribuidor, concessionário ou permissionário do serviço de energia, da localidade na qual se encontram instalados. O consumidor cativo pode ser classificado como consumidor grupo A (alta e média tensão) ou consumidor grupo B (baixa tensão).

A tarifa de fornecimento de energia do consumidor cativo é cobrada considerando a TUSD (tarifa associada aos custos operacionais para distribuição de energia) e TE (tarifa associada aos custos relacionados à compra de energia), sendo essa última definida com base na demanda de potência (kW) e consumo de energia (kWh). No caso do consumidor cativo baixa tensão, a TE é estabelecida somente em função do consumo de energia (kWh). E, para o consumidor cativo média e alta tensão, considera-se tanto o consumo de energia (kWh) quanto a demanda de potência (kW). A Tabela 1 apresenta as possíveis modalidades tarifárias de fornecimento de energia aplicáveis aos consumidores média e alta tensão e, na Tabela 2, tem-se a composição da TE, conforme modalidade tarifária adotada para esses consumidores.

Tabela 1: Modalidades tarifárias consumidor cativo grupo A

Tensão (kV)	Demanda contratada (kW)	Modalidades tarifárias		
		Horo-sazonal azul	Horo-sazonal verde	Convencional
< 69	< 300	Horo-sazonal azul	Horo-sazonal verde	
< 69	> 300	Horo-sazonal azul	Horo-sazonal verde	
= 69		Horo-sazonal azul		

Fonte: Fugimoto (2010).

Tabela 2: Estrutura tarifária, por modalidade tarifária, consumidor cativo grupo A

Modalidade tarifária	Período	Tarifa demanda de potência (R\$/kW)		Tarifa consumo de energia (R\$/kWh)	
		Ponta (P)	Fora ponta (Fp)	Ponta (P)	Fora ponta (Fp)
Horo-sazonal azul	Úmido (u)	TD(P)	TD(Fp)	TC(Pu)	TC(Fpu)
	Seco (s)			TC(Ps)	TC(Fps)
Horo-sazonal verde	Úmido (u)	TD	TD	TC(Pu)	TC(Fpu)
	Seco (s)			TC(PS)	TC(Fps)
Convencional	Úmido (u)	TD	TD	TC	
	Seco (s)			TC	

Fonte: Dávila (2011).

2.2 Segmentação do mercado consumidor

Entende-se por segmentação a tarefa de agrupar e/ou classificar determinados consumidores conforme características de demografia, comportamentos, valores, etc. Para Hiziroglu (2013) e Tkaczynski e Rundle-Thiele (2011), a segmentação do mercado foi introduzida na literatura de *marketing* por Smith (1956), o qual a considera uma estratégia de gestão e *marketing*.

Segundo Kotler (2003), deve-se considerar os seguintes critérios para uma segmentação eficaz: (i) mensurabilidade, capacidade de medir as variáveis dos segmentos; (ii) acessibilidade, capacidade de verificar a existência de determinado segmento como representativo de uma categoria de clientes; (iii) capacidade de diferenciação, os segmentos devem ser distinguíveis; (iv) substancialidade, a dimensionalidade de segmentos deve ser proporcional ao estudo realizado; (v) capacidade de ação, possibilidade de implementação de ações para alocar ou realocar clientes nos segmentos.

Conforme Jain et al. (1999), a segmentação por agrupamento envolve quatro etapas, a saber: 1^a etapa – seleção de dados; 2^a etapa – definição de padrão de proximidade; 3^a etapa – escolha do método, técnica e algoritmo de agrupamento e 4^a etapa – interpretação e validação dos resultados. A segmentação pode ser realizada por meio de técnicas estatísticas multivariadas (técnicas tradicio-

nais) ou por técnicas *soft computing*⁴. Hiziroglu (2013) menciona que as técnicas *soft computing* têm crescido como área potencial em pesquisas de segmentação.

Os Mapas Auto-Organizáveis de Kohonen (isto é, as redes neurais Self-Organizing Maps – Redes SOM) foram introduzidos na literatura nos anos 1980, seguem o paradigma de aprendizado não supervisionado e têm sido amplamente aplicados na segmentação por agrupamento em diferentes áreas, como na área de indústria, finanças, ciências naturais e linguísticas (KOHONEN, 2013; VESANTO, 1999; YANG; HUNG; CHEN, 2012).

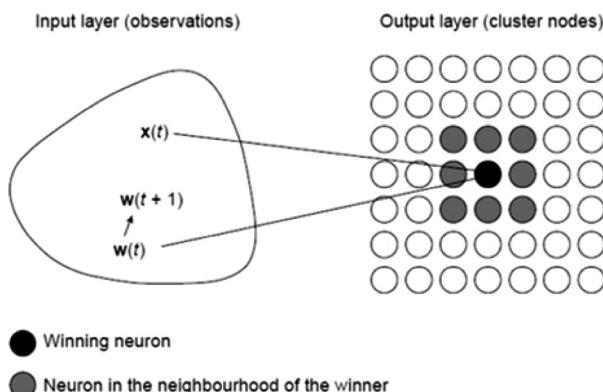
2.3 Materiais e métodos

Nesta subseção, apresentam-se os métodos e a base de dados que orientam e alimentam, respectivamente, o desenvolvimento desta pesquisa.

2.3.1 Rede SOM

A Rede SOM, ilustrada na Figura 1, é formada por uma camada de entrada com observações -dimensionais e uma camada de saída representada por uma grade que consiste em neurônios associados a pesos -dimensional (EVERITT et al., 2011).

Na Rede SOM, cada vetor de entrada seleciona a unidade neural que melhor o representa (neurônio vencedor), considerando uma relação

**Figura 1: Arquitetura da Rede SOM**

Fonte: Everitt et al. (2011).

de proximidade entre vetor peso sináptico e o vetor de entrada, que é estabelecida pela medida de distância euclidiana, como demonstra a equação 1 (CARVALHO; BRAGA; LUDERMIR, 1998; KOHONEN, 1990, 2013). Em seguida, cada neurônio ativado (isto é, neurônio vencedor) determina a localização espacial de uma vizinhança topológica de neurônios excitados que com ele interagem, como mostrado na equação 2 (HAYKIN, 2001). E, então, a cada instante de tempo, até que um melhor ajuste aos dados de entrada seja alcançado pela rede, uma atualização dos pesos de todos os neurônios da grade que se encontram dentro da vizinhança topológica do neurônio vencedor é realizada, conforme equação 3 (HAYKIN, 2001; KOHONEN, 2013).

$$i(x) = \operatorname{argmin}_j x - w_j \quad (1)$$

$$h_j, i_x(t) = \exp - d_{j,i}^2 / 2\sigma^2(t) \quad (2)$$

$$w_{it+1} = w_{it} + \eta(t) h_j, i_x(t) x - w_i(t) \quad (3)$$

Nas equações (1) a (3), tem-se:

$i(x)$: neurônio que melhor casa com o vetor de entrada x – isto é, neurônio vencedor;

$j=1, 2, \dots, l$. Sendo: l , número total de números na grade;

x : vetor de entrada selecionado aleatoriamente do espaço de entrada com m -dimensão – representado, matematicamente, por $x=x_1, x_2, \dots, x_m$; w_j : vetor peso sináptico do neurônio j na grade com mesma dimensão do espaço de entrada – representado, matematicamente, por $w_j=w_{j1}, w_{j2}, \dots, w_{jm}$;

t : instante de tempo. Assim: $t=0, 1, 2, \dots$;

$h_j, i_x(t)$: vizinhança topológica do neurônio vencedor em um instante de tempo;

$d_{j,i}^2$: distância entre neurônio excitado j e neurônio vencedor i ao quadrado;

$\sigma(t)$: representa a “largura efetiva” da vizinhança topológica no instante t . Esse parâmetro deve diminuir com o tempo a fim de satisfazer uma das condições da função de vizinhança topológica (HAYKIN, 2001). Assim: $\sigma(t) = \sigma_0 \exp - t\tau_1$, sendo: σ_0 , valor de σ na inicialização do algoritmo SOM e τ_1 , uma constante de tempo do algoritmo de aprendizagem do SOM;

$\eta(t)$: taxa de aprendizagem do algoritmo. Esse parâmetro deve decrescer gradualmente com o tempo. Em geral, adota-se: $\eta(t) = \eta_0 \exp - t\tau_2$, sendo: η_0 , valor de η na inicialização do algoritmo SOM e τ_2 , uma constante de tempo do algoritmo de aprendizagem do SOM.

A atualização dos pesos sinápticos consiste em um processo iterativo de aprendizagem

(treinamento) que é realizado em duas etapas, a saber: (i) fase de auto-organização ou de ordenação, que consiste na ordenação topológica dos vetores de peso; (ii) fase de convergência, que consiste no aprimoramento do agrupamento realizado. Para detalhamentos sobre o número de iterações do algoritmo SOM em cada fase, veja Carvalho et al. (1998) Haykin (2001) e Kohonen (1997). E, para atribuição de valores iniciais aos parâmetros neurais , e , consulte Kohonen (2013) e Haykin (2001).

É válido ressaltar que a vizinhança pode assumir vários formatos diferentes, tais como hexagonal e quadrado (FARIA et al., 2010; VESANTO et al., 1999, 2000). Conforme Carvalho et al. (1998, p. 114), “[...] a definição do formato mais adequado depende do problema atacado e da distribuição de seus dados. Geralmente, o formato da vizinhança é definido por tentativa e erro [...]”.

A interpretação gráfica da Rede SOM pode ser realizada por meio de diversos métodos existentes, os quais podem ser classificados nas três seguintes categorias (VESANTO, 1999):

- 1^a categoria – métodos que fornecem uma ideia global e possível da estrutura de grupo. Nesse caso, tem-se: a projeção realizada pelo próprio SOM, que fornece a identificação do vetor de entrada por neurônio e a quantidade de vetor de entrada por unidade do mapa por meio da proporção de preenchimento desta unidade. Desta forma, verifica-se que regiões mais densas representam grupos de unidades neurais cujo vetor de atributos é similar aos seus vizinhos; e áreas menos densas, regiões de separação entre grupos.
- 2^a categoria – métodos de detecção de grupos. Nesse caso, tem-se: (i) U-matriz, matriz de distância unificada entre unidades

vizinhas. Esse método permite visualizar relações topológicas do mapa por meio de tonalidades de cor, atribuídas a cada unidade neural, conforme distância existente entre unidades vizinhas, de modo que, tons mais claros representem menor distância entre vetores de entrada e respectivo vetor de pesos e, tons mais escuros, maior distância entre esses vetores. Assim, grupos são detectados pelas delimitações dessas regiões; (ii) Similaridade por cor, matriz de distância unificada com atribuição de cor semelhante às unidades do mapa que possuem proximidade no espaço de entrada. Para isso, verifica-se a distância entre a projeção das observações -dimensionais de entrada sobre cada vetor de peso do mapa gerado por meio da técnica de análise de componentes principais⁵.

- 3^a categoria – métodos de análise das características dos grupos. Nesse caso, tem-se a U-matriz por atributos, matriz de distância por atributos entre unidades vizinhas. Esse método permite verificar a existência de correlação entre os -dimensionais atributos dos vetores de entrada , de modo que, U-matrizes com padrões de coloração semelhantes indicam correlação positiva enquanto, padrões de coloração invertidos indicam correlação negativa.

Em relação à validação, a literatura aborda duas métricas para avaliação dos resultados do algoritmo SOM, a saber: (i) erro de quantização, verifica o quanto bem ajustado estão os vetores de entrada na rede; (ii) erro topográfico, avalia a capacidade do mapa em representar a topologia dos dados de entrada. Tais métricas são descritas em Faria et al. (2010), Sassi (2006) e Vesanto et al. (2000).

2.3.2 Análise exploratória de dados

A análise exploratória de dados consiste em um estudo descritivo de um conjunto de dados, por meio de medidas numéricas ou gráficas, cuja função é sintetizar informações neles contidas (BEHRENS, 1997; SELTMAN, 2013). Tal estudo pode seguir abordagem univariada ou multivariada, dependendo da quantidade de variáveis em questão. Em geral, utilizam-se as seguintes medidas para descrever um conjunto de dados:

Medidas de tendência central – referem-se a medidas que produzem um valor médio representativo de um conjunto de dados (SPIEGEL; STEPHENS, 2008), também denominadas por medidas de locação. A média aritmética, mediana e moda são as mais comumente usadas para esse fim.

Medidas de dispersão – referem-se a medidas que representam o grau de variabilidade entre dados (SPIEGEL; STEPHENS, 2008). Nesse caso, as mais utilizadas são: amplitude, desvio-padrão e variância.

Histograma de frequência – trata-se de um diagrama de barras no qual cada barra representa a frequência ou proporção de casos para um intervalo de valores (SELMAN, 2013). A escolha dos intervalos é definida pelo pesquisador arbitrariamente de modo que as classes tenham a mesma amplitude, e a frequência de observação dos dados sejam superiores a cinco.

2.3.3 Base de dados

A base de dados utilizada nesta pesquisa foi extraída de uma empresa de distribuição de energia elétrica brasileira que atende 2,8 milhões de consumidores residenciais, comerciais e industriais, em mais de 60 municípios, cobrindo uma área aproximada de 30 000 . Tal distribuidora faz parte de uma *holding* controlada por um grupo italiano que atua nos demais segmentos do setor elétrico brasileiro. A base de dados cedida

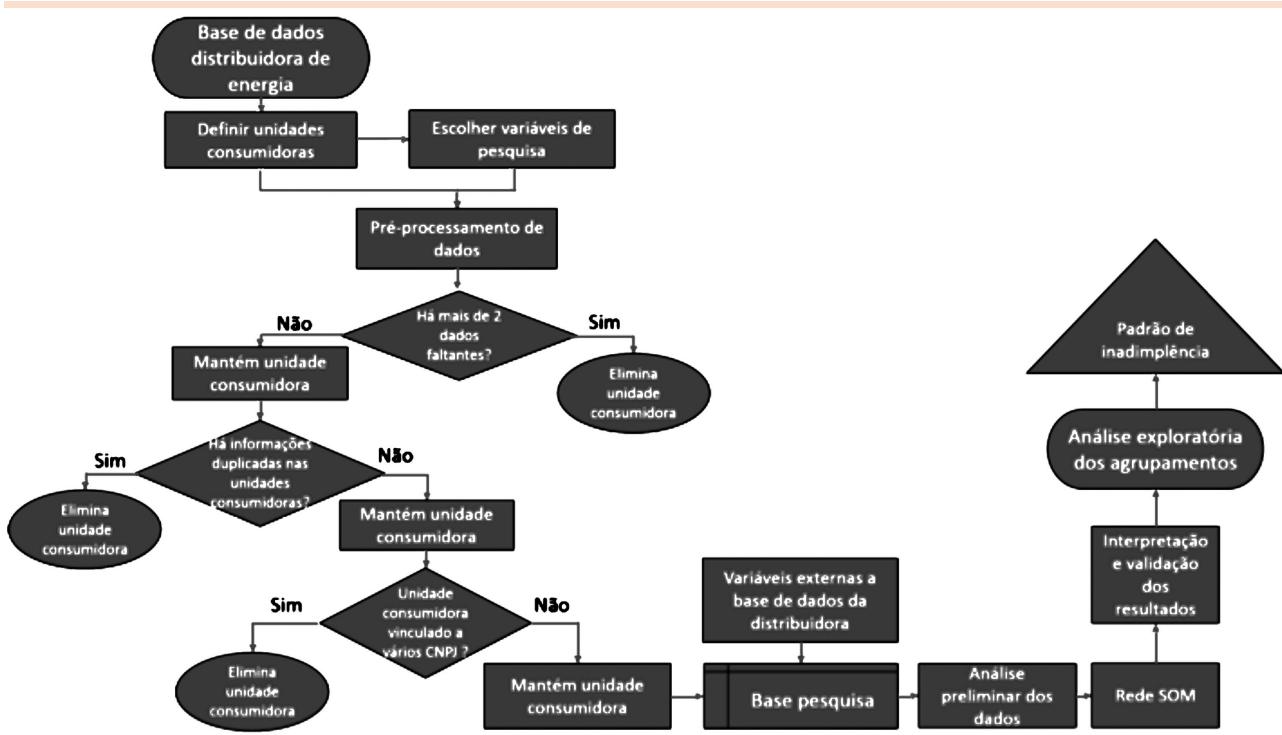
pela diretoria comercial da distribuidora é composta por: histórico de faturamento mensal, referente ao período de janeiro/2005 a abril/2011; histórico de arrecadação mensal, realizado de janeiro/2009 a abril/2011, e cadastro de unidades consumidoras cativas grupo A. Nesses históricos estão presentes as seguintes variáveis: número unidade consumidora, nome consumidor, município unidade consumidora, corte, pendência judicial, porte unidade consumidora, ano mês referência faturamento, data de pagamento, data de vencimento, valor fatura, tensão, nome setor, demanda contratada, demanda medida, consumo ativo.

3 Descoberta de padrões de inadimplência em uma distribuidora de energia elétrica mediante segmentação do mercado consumidor pela Rede SOM

A metodologia adotada para obtenção do estudo proposto na seção 2 será essencialmente constituída pela análise dos grupos gerados pela rede neural artificial não supervisionada Self-Organizing Maps. O estudo foi realizado em um *ultrabook* com sistema operacional de 64 bits, processador Intel(R) Core (TM) i5-3317U de 1,70 GHz e 4,00 GB de memória RAM. O *software* MatLab, especificamente o Toolbox SOM⁶, foi utilizado para realizar a tarefa de agrupamento dos dados selecionados. A Figura 2 ilustra os procedimentos adotados para realização deste estudo.

3.1 Pré-processamento de dados

O estudo foi realizado considerando comportamentos de consumo, valor fatura, demanda e atraso das unidades consumidoras alta e média

**Figura 2: Representação esquemática da pesquisa**

Fonte: Os autores.

tensão no período de jan./2009 a dez./2010. Antes da implementação da metodologia proposta, o procedimento limpeza, redução e transformação dos dados foi realizado na base disponibilizada pela distribuidora. No processo de limpeza, foram eliminados consumidores: (i) com três ou mais dados faltantes; (ii) com informações duplicadas ou conflitantes referentes a determinada variável no tempo, e (iii) cuja unidade consumidora, anteriormente, pertencia a outro. Os registros faltantes não foram tratados com metodologia estatística, pois adotou-se a mediana, medida de tendência central não sensível a valores extremos, dos registros por consumidores na entrada do modelo neural. Posteriormente, os dados foram normalizados, integrados e consolidados em uma base de dados única, denominada base pesquisa, que é constituída por variáveis categóricas, contínuas e binárias. Além das informações disponibilizadas pela distribuidora, a base pesquisa é composta por informações, como atividade econômica e

natureza administrativa, buscadas em *sites* de cadastro mantidos por órgãos públicos e privados. Na Tabela 3, constam as variáveis utilizadas na implementação da metodologia proposta.

3.2 Seleção e análise preliminar dos dados

Conforme configuração descrita anteriormente, o estudo foi conduzido com 2.400 unidades consumidoras. Tais unidades são pessoas jurídicas públicas ou privadas estabelecidas na área de concessão da distribuidora em questão. Uma análise preliminar dos dados foi realizada antes da segmentação por agrupamento objetivando investigar o comportamento dos dados de cada variável descrita na Tabela 3. Por meio dessa análise, foi possível verificar o seguinte:

- a distribuição do consumo mediano e do valor fatura mediano apresentava características bastante assimétricas e com significativo

Tabela 3: Variáveis base pesquisa

Variável	Base de dados de origem
Número unidade consumidora	Cadastro/ Faturamento e arrecadação/ Histórico consumo e demanda
Nome consumidor	Cadastro
Região	Cadastro > Município unidade consumidora & Mesorregião a qual pertence
Corte	
Pendência judicial	Cadastro
Porte unidade consumidora	
Ano mês referência faturamento	Faturamento e arrecadação/ Histórico consumo e demanda
Atraso	Faturamento e arrecadação > Data de pagamento & Data de vencimento
Valor fatura	Histórico consumo e demanda
Modalidade tarifária	Histórico consumo e demanda > Tensão & Demanda contratada & Nome setor
Demandas contratadas	Histórico consumo e demanda > Maior valor (Demanda contratada fora ponta seco & Demanda contratada fora ponta úmida)
Demandas medidas	Histórico consumo e demanda > Demanda medida fora ponta
Consumo ativo	Histórico consumo e demanda > Consumo ativo fora ponta & Consumo ativo ponta
Atividade econômica	Cadastro > Nome consumidor & Busca nome consumidor em sites de cadastro mantidos por órgãos públicos e privados
Natureza administrativa	Cadastro > Nome consumidor & Busca Nome consumidor em sites de cadastro mantidos por órgãos públicos e privados

Fonte: Os autores.

número de valores discrepantes. Então, uma transformação logarítmica foi aplicada;

- cerca de 75% dos consumidores possuem comportamento mediano de atraso igual a zero, assim, optou-se por analisá-los em função dos perfis adimplente e inadimplente. Considerou-se inadimplente o consumidor que teve em algum momento, jan./2009 a dez./2010, registro de atraso;
- a maioria desses consumidores possuem transitoriedade no comportamento de pagamento. O perfil inadimplência é verificado em 59% deles;
- o comportamento de pagamento varia conforme porte e atividade econômica do consumidor.

Assim, na segmentação por agrupamento, foram consideradas as seguintes características comportamentais: *log* do consumo mediano, *log* do valor fatura mediano, mediana (demanda medida/demanda), corte e pendência judicial. As variáveis categóricas dessa pesquisa foram usadas na identificação de padrão dos grupos.

3.3 Resultados e discussões

As variáveis usadas na tarefa de agrupamento foram dispostas em uma matriz, conforme Figura 3.

1ª linha: Indicação do número de variáveis

2ª linha: Nome das variáveis precedidas por #n

1º vetor de dados

2º vetor de dados

n-ésimo vetor de dados

<i>m</i>					
#n	<i>X</i> ₁	<i>X</i> ₂	...	<i>X</i> _{<i>m</i>}	<i>Y</i>
	<i>x</i> ₁₁	<i>x</i> ₁₂	<i>x</i> ₁₂	<i>x</i> _{1<i>m</i>}	<i>y</i> ₁
	<i>x</i> ₂₁	<i>x</i> ₂₂	<i>x</i> ₂₂	<i>x</i> _{2<i>m</i>}	<i>y</i> ₂
	:	:	:	:	:
	<i>x</i> _{<i>n</i>1}	<i>x</i> _{<i>n</i>2}	...	<i>x</i> _{<i>n<i>m</i></i>}	<i>y</i> _{<i>n</i>}

Variável rótulo (atribui um nome a cada amostra)

Figura 3: Matriz de dados das unidades consumidoras

Fonte: Os autores.

Diferentes configurações para geração da rede SOM foram testadas, como mostrado na Tabela 4. Para definição do tamanho da rede, considerou-se uma fórmula heurística implementada pelo Toolbox SOM como sendo a rede de maior tamanho e, a partir dela, determinou-se uma rede pequena e média. Conforme Faria et al. (2010), a fórmula heurística para determinação das dimensões da rede é dada pela raiz quadrada da razão dos dois maiores autovalores do conjunto de dados, de modo que, o seu produto seja próximo ao número desejado de neurônios da rede, também definido por heurística, que corresponde a cinco vezes a raiz quadrada do tamanho da amostra de dados. Para definição de formato de vizinhança e inicialização do algoritmo, adotou-se o menor erro médio obtido para a rede de tamanho pequeno. Outros parâmetros neurais foram inicializados conforme literatura. O treinamento ocorreu de modo sequencial, isto é, a atualização dos pesos sinápticos era realizada toda vez que uma unidade consumidora fosse apresentada à rede.

Verifica-se, pela Tabela 4, que a rede de tamanho apresentou menor erro médio de quantização e topológico e, por isso, foi adotada como

configuração que melhor mapeou os dados do espaço de entrada nas unidades neurais. Então, procedeu-se uma análise mais detalhada da estrutura dessa rede.

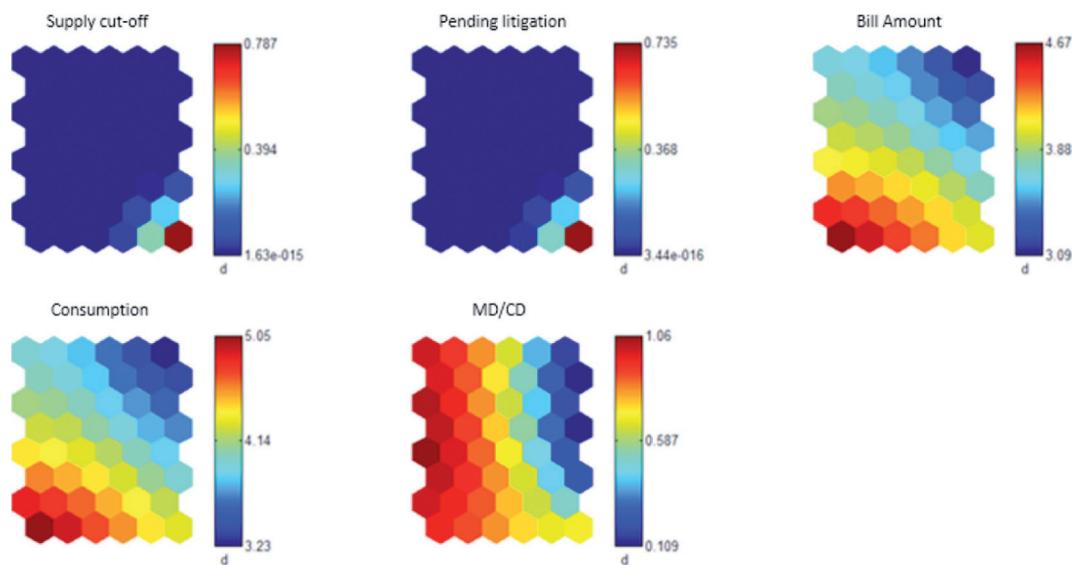
A existência de correlação entre os -dimensionais atributos dos vetores de entrada pode ser verificada na Figura 4. Observa-se, através do espetro de cores fornecido, que os atributos corte e pendência judicial são fortemente correlacionados, o mesmo pode ser visualizado entre valor fatura e consumo. Em relação à demanda, verifica-se que as unidades consumidoras cuja demanda contratada não é ajustada (isto é, $DM \neq DC$) apresentam consumo e valor fatura da demanda baixos (unidades neurais com tons azul marinho). Nota-se que unidades consumidoras que passaram por ocorrência de corte ou pendência judicial apresentam demanda, consumo e valor fatura moderado alto (unidades neurais com tons entre verde e amarelo).

A Figura 5 apresenta duas técnicas (a) e (b) que em conjunto torna evidente os grupos na rede SOM (c). Em (a), tem-se a rede SOM, propriamente dita, com a quantidade de vetor de entrada por unidade neural (*bits*) representada pela densidade de preenchimento da unidade. Por esse conceito,

Tabela 4: Configurações de rede testadas

Parâmetros	Pequena				Média	Grande
	Conf. 1	Conf. 2	Conf. 3	Conf. 4		
Função vizinhança	Gaussiana					
Raio final	1					
Algoritmo de treinamento	Sequencial					
Função de taxa aprendizagem	Exponencial					
Taxa de aprendizagem inicial	0,1					
Tamanho da rede						
Raio inicial	2	2	2	2	4	9
Formato da vizinhança	Retangular	Hexagonal	Retangular	Hexagonal	Hexagonal	Hexagonal
Algoritmo de inicialização	Linear	Linear	Aleatório	Aleatório	Aleatório	Aleatório
Nº total de iterações	8.500	8.500	8.500	8.500	25.000	118.000
Erro de quantização (QE)	0,0797	0,0798	0,0797	0,0807	0,0470	0,0237
Erro topológico (TE)	0,2100	0,0342	0,1317	0,0217	0,0250	0,0588

Fonte: Os autores.

**Figura 4: U-matriz por atributos**

Fonte: Os autores.

conclui-se que regiões mais densas apresentam similaridades para com seus vizinhos, constituindo, assim, possíveis grupos separados por regiões menos densas. Em (b), uma classificação por cores (similaridade por cor) sobre cada unidade neural é apresentada distinguindo os grupos. Em (c), os grupos são evidenciados considerando a visualização conjunta entre (a) e (b).

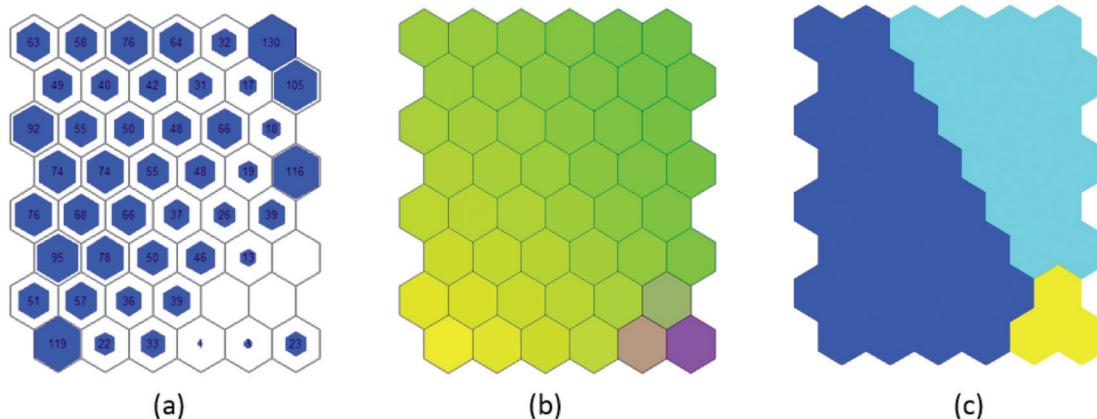
Assim, as unidades neurais que compõem cada grupo são descritas, a seguir:

Grupo 1={1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
19,20,21,22,23,24,29,30,31,32,39}

Grupo 2={17,18,25,26,27,28,33,34,35,36,37,38,
41,42,43,44,45,46}

Grupo 3={40,47,48}

Uma vez realizada a interpretação gráfica do SOM, uma análise exploratória dos grupos formados foi realizada para fins de identificação dos segmentos gerados (Tabela 5).

**Figura 5: Segmentação por agrupamento das unidades consumidoras**

Fonte: Os autores.

Grupos	Número de unidades consumidoras	Número de unidades com corte e/ou pendência judicial	Mediana DM/DC	Mediana valor fatura (R\$)	Mediana consumo (kWh)
Grupo1	1484	0	0,95	10981	22185
Grupo2	890	0	0,33	3068	4348
Grupo3	26	26	0,91	11153	24212

Tabela 5: Estatísticas dos segmentos gerados

Fonte: Os autores.

Posteriormente, um padrão da inadimplência no setor pode ser obtido considerando variáveis categóricas presentes na base pesquisa. As seguintes variáveis foram utilizadas para esse propósito: modalidade tarifária= {convencional, Horo-sazonal}; porte= {pequeno, médio, grande}; atividade econômica= {indústria, comércio, serviços}; região= {A, B, C, D}.

Verificou-se, assim, para cada grupo gerado, a proporção de unidades consumidoras que possuíam os atributos supracitados. E, então, obteve-se um padrão da inadimplência na distribuidora de energia elétrica em estudo. Desta forma, tem-se o que segue:

- o Grupo 1 é majoritariamente composto por unidades consumidoras que optam por uma estrutura tarifária horo-sazonal. Em geral, tais unidades conseguem um bom ajuste entre demanda medida e demanda contratada. Trata-se de empresas dos principais setores econômicos (indústria, comércio e serviços), sendo o setor de serviços responsável por 47% da inadimplência, seguido pela indústria, 28%. Cerca de 50% dessas empresas são de médio porte; e 20%, de grande porte.
- o Grupo 2 apresenta unidades consumidoras que possuem, essencialmente, estrutura tarifária convencional. As unidades consumidoras horo-sazonais não fazem um bom ajuste entre demanda medida e demanda contratada. Nesse grupo, o setor de serviços é responsável

por 70% da inadimplência, seguido pela indústria, 18%. Desses empresas, 46% são de médio porte; e 36%, de pequeno porte.

- o Grupo 3 tem unidades consumidoras que, em algum momento, passaram por corte e/ou tiveram cobrança judicial. Em geral, tais unidades apresentam valor fatura superior às unidades consumidoras dos demais grupos. As unidades consumidoras horo-sazonais que integram esse grupo conseguem um bom ajuste entre demanda medida e demanda contratada. Cerca de 50% das unidades consumidoras inadimplentes são do setor de serviços e de pequeno porte.

Notou-se, em todos os grupos, que as regiões com maiores índices de inadimplência são as de maiores níveis de desenvolvimento.

Desta forma, conforme critérios estabelecidos por Kotler (2003), veja seção 2.2, pode-se afirmar que a segmentação das unidades consumidoras se mostrou eficaz, tornando possível obter um padrão da inadimplência no setor.

Outras pesquisas, conforme relatado na introdução, têm apresentado estudos sobre o tema inadimplência em distribuidoras de energia elétrica. Araújo (2007) desenvolveu um estudo descriptivo e explicativo sobre perdas e inadimplência nos sistemas de distribuição de energia tendo como meios de investigação uma base documental. Nessa pesquisa, o autor levanta hipóteses para identificar quais variáveis se relacionam aos

fenômenos de perdas e inadimplência. Souza et al. (2011) e Souza et al. (2013) obtiveram bom resultado em uma modelagem baseada em lógica *fuzzy* e geoprocessamento para identificar a capacidade de pagamento de unidades consumidoras cativas baixa tensão. Carvalho et al. (2013a, b) propuseram o uso de metodologias estatísticas como ferramenta de auxílio para modelar e quantificar as incertezas presentes na arrecadação da fatura de energia emitida pelos agentes distribuidores aos consumidores cativos média e alta tensão. Enquanto que, em um estudo preliminar, Carvalho et al. (2014) verificaram a positiva capacidade da Rede SOM em agrupar unidades consumidoras conforme comportamento de atrasos. Em decorrência disso, no trabalho atual, propuseram-se verificar padrões de inadimplência nos agrupamentos gerados a partir das variáveis listadas na Tabela 3. Tais investigações apresentam bons resultados e peculiaridades na visualização do fenômeno inadimplência, no entanto, se complementam. Nota-se que, a principal característica que distinguem esta pesquisa das demais se encontra na unidade de análise do estudo e na metodologia usada para descoberta de padrões nos dados.

4 Conclusão

O conhecimento armazenado em base de dados tem mostrado ser um elemento fundamental à compreensão dos fenômenos que envolvem qualquer ambiente organizacional.

A gestão da inadimplência nas distribuidoras de energia elétrica é uma atividade primordial ao equilíbrio econômico financeiro delas e do setor elétrico. Os instrumentos usados para combate à inadimplência têm sido insuficientes para recuperação do faturamento perdido. Em recentes estudos, tem-se explorado o desenvol-

vimento de uma metodologia preventiva, mas, ainda são incipientes. Neste artigo, buscou-se contribuir com uma metodologia preventiva para combate a inadimplência em distribuidoras de energia elétrica. As principais limitações encontradas para a realização deste estudo foram: a obtenção de dados externos à base de dados cedidos pela distribuidora de energia e o pré-processamento de dados, pois ambos demandaram muito tempo de pesquisa.

Os resultados obtidos foram potenciais em explicar o comportamento de pagamento das unidades consumidoras, de maneira que os responsáveis na distribuidora de energia identifiquem possíveis inadimplentes nos grupos. Tal fator leva a atual investigação a contribuir como uma ferramenta de apoio à prevenção da inadimplência. Pretende-se, em um estudo futuro, usar os grupos gerados na formulação de regras para criar uma metodologia classificatória da capacidade de pagamento.

Notas

- 1 Fatores não técnicos que geram prejuízos ao sistema de distribuição e são causados devido à inadimplência de consumidores e ao uso irregular da energia elétrica, principalmente, de terceiros (furto de energia e fraude no consumo de eletricidade) ou por equipamentos defeituosos (erros de leitura, falha no faturamento).
- 2 Trata-se do modelo regulatório Price-Cap, que estabelece um valor teto para a tarifa, a qual se ajusta anualmente pela taxa de inflação descontada de um índice de ganho de produtividade pré-definido pelo regulador.
- 3 As atividades de valor são “[...] atividades física e tecnologicamente distintas, através das quais uma empresa cria um produto valioso para os seus compradores [...]” (PORTER, 1998, p. 34).
- 4 Conjunto de conceitos formado por uma parceria de métodos distintos que possuem como princípio orientador explorar a imprecisão e incerteza para alcançar robustez, baixo custo da solução e melhor proximidade com a realidade (ZADEH, 1994). Os métodos tradicionais relacionados ao termo *soft computing* são (MAGDALENA, 2010): lógica *fuzzy*, redes neurais artificiais, computação evolucionária e raciocínio probabilístico.
- 5 Técnica de análise exploratória de dados multidimensionais que transforma um conjunto $V = \{v_1, \dots, v_N\}$ de vetores $v_n = [v_{n1}, \dots, v_{nD}]^T \in \mathbb{R}^D$, $n=1, \dots, N$, em uma dada base ortonormal e encontra uma nova base ortonormal $\{u_1, \dots, u_D\}$

- capaz de gerar o espaço original, de modo que, ao optar por uma projeção de dados utilizando os P primeiros componentes principais, $P < D$, obtém-se uma representação do conjunto original em um espaço de menor dimensão.
- 6 Trata-se de uma biblioteca de rotinas para implementação das Redes SOM no MatLab. Foi desenvolvido pelo Laboratório de Computação e Ciências da Informação da Universidade de Tecnologia de Helsinki – Finlândia.

Referências

- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (ANEEL). *Atlas de energia elétrica do Brasil*. 3. ed. Brasília, DF: Agência Nacional de Energia Elétrica, 2008. 236 p. 978-85-87491-10-7.
- ARAUJO, A. C. M. De. *Perdas e inadimplência na atividade de distribuição de energia elétrica no Brasil*. 2007. 116 p. Tese (Doutorado em Ciências em Planejamento Energético)– Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2007.
- ASSOCIAÇÃO BRASILEIRA DE DISTRIBUIDORES DE ENERGIA ELÉTRICA. *Associação Brasileira de Distribuidores de Energia Elétrica*. [s.d.]. Disponível em: <www.abradee.org.br>. Acesso em: jan. 2016.
- BEHRENS, J. T. Principles and procedures of exploratory data analysis. *Psychological Methods*, v. 2, n. 2, p. 131-160 , 1997.
- CARVALHO, A. F.; BRAGA, A. P.; LUERMIR, T. B. *Fundamentos de redes neurais artificiais*. Rio de Janeiro: DCC/IM, COPPE/SISTEMAS, NCE/UFRJ, 1998.
- CARVALHO, N. A. S.; SOUZA, R. C.; EPPRECHT, E. K. *Topologia do perfil de pagamento dos consumidores de alta tensão de um distribuidor de energia elétrica*. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 46., 2014, [S.l: s.n.], Salvador. *Anais...* Salvador: SBPO, 2014. p. 1194-1205.
- CARVALHO, N. A. S.; SOUZA, R. C.; FROTA, M. N. Métodos estatísticos aplicados ao fluxo de caixa de uma distribuidora de energia elétrica. *Exacta-E P*, v. 11, n. 2, p. 137-148, 2013a.
- CARVALHO, N. A. S.; SOUZA, R. C.; FROTA, M. N. Statistical analysis of the collectability index: a case study of an electricity distribution utility. In: INTERNATIONAL SYMPOSIUM OPERATIONAL RESEARCH AND APPLICATIONS – ISORAP, 2013, Marrakesh. *Proceedings...* Marrakesh: [s.n.], 2013b, p. 565-570.
- DÁVILA, A. O. C. Medições de consumo e de demanda de energia elétrica: a base metrológica para simulação de tarifas de clientes do grupo A. 2011. 145 p. Dissertação (Mestrado em Metrologia para a Qualidade e Inovação)– Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2011.

DEPURU, S. S. S. R.; WANG, L.; DEVABHAKTUNI, V. Electricity theft: overview, issues, prevention and a smart meter based approach to control theft. *Energy Policy*, v. 39, p. 1007-1015, 2011.

EVERITT, B et al. *Cluster analysis*. 5. ed. [S.l.]: Wiley, Wiley Series in Probability and Statistics, 2011.

FARIA, E. L. et al. *Introdução ao toolbox de redes neurais de Kohonen*. Rio de Janeiro: [s.n.], 2010.

FONSECA, J. N.; REIS, L. B. *Empresas de distribuição de energia elétrica no Brasil: temas relevantes para gestão*. Rio de Janeiro: Synergia, 2012.

FUGIMOTO, S. K. *Estrutura de tarifas de energia elétrica – análise crítica e proposições metodológicas*. 2010. 195 p. Tese (Doutorado em Engenharia)– Universidade de São Paulo, São Paulo, 2010.]

HAYKIN, S. *Redes neurais*. 2. ed. ed. Porto Alegre: Bookman, 2001.

HIZIROGLU, A. Soft computing applications in customer segmentation: state-of-art review and critique. *Expert Systems with Application*, v. 40, p. 6491-6507 , 2013.

INSTITUTO ACENDE BRASIL. Perdas e inadimplência no setor elétrico. Cadernos de política tarifária p. 1-9 , 2007a. Disponível em: <<http://www.acendebrasil.com.br/br/estudos>>. Acesso em: jun.2016

INSTITUTO ACENDE BRASIL. Política tarifária e regulação por incentivos. Cadernos de política tarifária. p. 1-10 , 2007b. Disponível em: <<http://www.acendebrasil.com.br/br/estudos>>. Acesso em: jun. 2016

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. *ACM Computing Surveys*, v. 31, n. 3, p. 264-323, 1999.

KOHONEN, T. Essentials of the self-organizing map. *Neural Networks*, v. 37, p. 52-65, 2013.

KOHONEN, T. *Self organizing maps*. 2. ed. Heidelberg: Springer, Springer series in Information Sciences, 1997.

KOHONEN, T. The self-organizing map. [S.l]: IEEE, 1990. v. 78, n. 9. p.1464-1480.

KOTLER, P. *Marketing management*. 13. ed. New Jersey: Prentice-Hall, 2003.

MAGDALENA, L. What is soft computing? Revisiting possible answer. *International Journal of Computational Intelligence Systems*, v. 3, n. 2, p. 148-159 , 2010.

MEDEIROS, A. L. Alocação de equipes de campo para avaliação de perdas não-técnicas de energia elétrica: desenvolvimento de um sistema de apoio a decisão. 2013. 186 p. Tese (Doutorado em Administração, Gestão de Negócios, Economia e Mercado)– Universidade Federal de Lavras, Lavras, 2013.

PORTER, M. E. *Vantagem competitiva: criando e sustentando um desempenho superior*. 11. ed. Rio de Janeiro: Campos, 1998.

SASSI, R J. Uma arquitetura híbrida para descoberta de conhecimento em base de dados: teoria dos rough sets e redes neurais artificiais mapas auto organizáveis. 2006. 169 p. Tese (Doutorado em Sistemas Eletrônicos)-Universidade de São Paulo, São Paulo, 2006.

SELTMAN, H. J. *Experimental design and analysis*. 2015. Disponível em: <www.stat.cmu.edu/~hseltman/309/Book/Book.pdf>. Acesso em: jun.2016.

SMITH, T. B. Electricity theft: a comparative analysis. *Energy Policy*, v. 32, p. 2067-2076 , 2004.

SOUZA, R. C. et al. Índice de capacidade de pagamento dos consumidores residenciais BT: um método baseado em lógica fuzzy. 2011, Bento Gonçalves: [s.n.], 2011.

SOUZA, R. C.; PESSANHA, J. F. M.; OLIVEIRA, F. L. C. A residential consumer payment capability index based on fuzzy logic inference. *Journal of Intelligent & Fuzzy Systems*, v. 25, n. 3, p. 649-657, 2013.

SPIEGEL, M. R.; STEPHENS, L. J. *Schaum's outline of theory and problems of statistics*. 4. ed. [S.l.]: McGraw-Hill Inc., 2008. 577 p.

TKACZYNSKI, A.; RUNDLE-THIELE, S. R. Event segmentation: a review and research agenda. *Tourism Management*, v. 32, p. 426-434 , 2011.

VERGARA, S. C. *Projetos e relatório de pesquisa em administração*. 14. ed. São Paulo: Atlas, 2013.

VESANTO, J et al. *Self organizing map in Matlab: the SOM toolbox*. 1999, Espoo: [s.n.], 1999. p. 35-40.

VESANTO, J. *SOM-based data visualization methods. Intelligent Data Analysis*, v. 3, p. 111-126 , 1999.

VESANTO, J. et al. *SOM toolbox for Matlab 5*. Espoo: [s.n.], 2000.

YANG, M.S; HUNG, W L; CHEN, D H. Self-organizing map for symbolic data. *Fuzzy set and systems*, v. 203, p. 49-73, 2012.

ZADEH, L. A. Soft computing and fuzzy logic. *IEEE Software*, v. 11, n. 6, p. 48-56, 1994.

Recebido em 7 fev. 2016 / aprovado em 14 jun. 2016

Para referenciar este texto

CARVALHO, N. A. S.; SOUZA, R. C.; EPPRECHT, E. K. Segmentação do mercado consumidor cativo alta e média tensão de uma distribuidora de energia elétrica: aplicação do Mapa Auto-Organizável de Kohonen para descoberta de padrões de inadimplência no setor. *Exacta - EP*, São Paulo, v. 14, n. 3, p. 495-509, 2016.

