

aper:180193_1

Diferenciação de Perfis de Curva de Carga para Identificação de Perdas Não-Técnicas em Redes de Distribuição Utilizando Mineração de Dados e Aprendizado de Máquina

Jorge Gustavo Sandoval Simão, Raimundo Celeste Ghizoni Teive

Universidade do Vale do Itajaí

jorge.sandoval@univali.br, rteive@univali.br

Abstract. — *It is estimated that, in Brazil, approximately 15% of electricity is consumed fraudulently, a practice known as non-technical loss. This article is an applied study that analyzes the identification of these losses in the distribution networks in order to recognize distortions in relation to the respective consumption patterns of each consumer and, from the analysis of this information, to detect possible deviations that can be classified as frauds. For this purpose, data mining and machine learning techniques were applied to commercial consumers in the State of Santa Catarina, presenting evidence that these techniques can be used to identify patterns of irregular electricity consumption.*

Resumo. *É estimado que, no Brasil, aproximadamente 15% da energia elétrica é consumida de forma fraudulenta, prática essa denominada perda não-técnica. Este artigo é um estudo aplicado que analisa a identificação dessas perdas nas redes de distribuição para reconhecer distorções em relação aos respectivos padrões de consumo de cada consumidor e, a partir da análise desta informação, detectar possíveis desvios que possam ser classificados como fraudes. Para tal, foram aplicadas técnicas de mineração e dados e aprendizado de máquina nos consumidores comerciais do Estado de Santa Catarina, apresentando evidências de que essas técnicas podem ser usadas para identificar padrões de consumo de energia elétrica irregulares.*

1. Introdução

O crescimento das concessionárias de energia elétrica e o aumento dos prejuízos financeiros causados pelas perdas de energia elétrica consequentes das não-conformidades nas redes de distribuição, têm incrementado a busca por novas tecnologias de detecção de fraudes [Queiroz *et al.* 2016].

As perdas de energia elétrica são divididas em duas categorias: as técnicas e as não-técnicas. As perdas não-técnicas (também conhecidas como perdas comerciais) são causadas pela manipulação ilegal de medidores ou por falhas nas instalações de consumo, enquanto as técnicas são causadas através de efeitos físicos (o efeito Joule, por exemplo), devidos à distribuição de energia [Guerrero *et al.* 2014]. As perdas não-técnicas, dada sua natureza, geram diminuição no registro do consumo de energia, ocasionando assim modificações nas topologias das curvas de carga.

As fraudes podem ser divididas nas categorias de residenciais, industriais e comerciais, sendo estes últimos o foco deste estudo. Os consumidores comerciais representam uma parcela considerável da energia consumida. Em 2014, em Santa Catarina (Brasil), a energia consumida pela classe comercial foi de 3.946.188 MW/h, significando 16.9% do consumo total (representando um aumento em relação à 2013, que teve um consumo de 3.604.418 MW/h) [FIESC 2015]. Considerando que podem haver aumentos anuais do consumo de energia da classe comercial, o desvio passou a tornar-se cada vez mais significativo. Em 2016, também em Santa Catarina, os consumos de energia elétrica residencial e comercial subiram mais do que o PIB (Produto Interno Bruto) com uma taxa de aumento de 3,6% [EPE 2016]. Dados estes registros, a representatividade das perdas não-técnicas sobre a classe comercial torna-se significativa.

Muitos métodos existentes para identificação de perdas não-técnicas impõe um alto custo operacional, e requerem uso extensivo de recursos humanos [Nizar *et al.* 2006], porém há a necessidade de tentar mensurá-las, pois essas fraudes representam perdas econômicas, colocam em risco a segurança pública e criam impactos sociais. A Superintendência de Pesquisa e Desenvolvimento e Eficiência Energética (SPE) da ANEEL (Agência Nacional de Energia Elétrica) propôs, em 2013, a criação de um sistema de informações envolvendo os agentes do setor, que possibilitasse (dentre outros objetivos) a criação de uma base de dados consistente para a aplicação de técnicas de inteligência analítica e de mineração de dados [SPE 2013]. Considerando a implementação desta tecnologia para as concessionárias de energia, as técnicas de inteligência computacional também podem ser utilizadas sobre os dados dos clientes de uma distribuidora.

Sendo assim, este estudo propõe um método de detecção de fraudes de consumidores comerciais baseado na utilização de aprendizado de máquina, agrupando as curvas de carga através de sua tipologia utilizando o algoritmo *K-Means* e posteriormente aplicando uma rede neural não-supervisionada (*Autoencoder*) no resultado, visando identificar usuários que fogem ao perfil do *cluster* ao qual pertencem, durante o período analisado.

2. Identificação de Perdas Não-Técnicas

Uma das fraudes na rede de distribuição mais comuns é o caso de medidores com freios (furos e/ou agulhas) no disco de medição. Este tipo de fraude faz com que o disco medidor de energia fique travado em um determinado valor, e não meça toda a energia consumida. Este padrão de curva de consumo também acontece quando é realizada ligação direta antes do medidor de energia e o usuário utiliza a energia passada pelo medidor apenas para determinados fins, diminuindo assim, o consumo registrado pelo medidor [Queiroz *et al.* 2016].

As tipologias de curvas de carga dos consumidores comerciais seguem determinados padrões, baseados na sua atividade econômica e características funcionais das instalações, que permitem seu agrupamento pelas mesmas, pois muitas seguem comportamentos semelhantes. Nota-se uma distorção no perfil de consumo, quando um consumidor apresenta uma modificação de tipologia de curva de carga que o diferencia dos outros consumidores do mesmo *cluster*, indicando uma variação significativa do

padrão de consumo de energia, o que pode sinalizar perda não-técnica. Os tipos de padrões de consumo (anômalos ou não) variam de consumidor para consumidor, mas são mais comuns entre determinados grupos de consumidores e mantém-se padrão ao longo do ano, com pequenas diferenças causadas principalmente pela sazonalidade.

Considerando a existência de padrões de perfil de consumo e as distorções (não-conformidades) causadas no mesmo pelas perdas não-técnicas, é possível detectar padrões de consumo através das tipologias que podem representar possíveis fraudes nas redes de distribuição de energia elétrica, utilizando-se algoritmos inteligentes sobre os dados de curvas de cargas diários.

3. Metodologia Aplicada

3.1. Definição do Escopo

Para a identificação de perdas não-técnicas deste estudo foram analisados os consumos das divisões CNAE (Classificação Nacional de Atividades Econômicas) catarinenses mais representativas, focando exclusivamente na classe comercial (postos de gasolina, concessionárias de automóveis, mercados, entre outros) em um prazo pré-determinado (ano de 2011).

3.2. Banco de Dados de Curvas de Carga

Os valores de consumo de energia elétrica diário de cada usuário são armazenados pelas concessionárias após a leitura horária dos valores de consumo nos medidores em bases de dados. A base utilizada para este estudo pertence à uma distribuidora de energia brasileira e contém informações de consumo do ano de 2011, contendo o código da unidade consumidora, data e hora do consumo, bem como potência ativa e potência reativa consumida por aquele consumidor.

O número total de consumidores nesta base é de 503, sendo destes 233 consumidores comerciais. Para cada consumidor foram analisadas as tipologias das curvas de carga dos dias da semana (segunda à sexta), para todas as estações, sendo então 240 curvas de carga por consumidor (e 55.920 curvas de carga no total). Como complemento, esta mesma base possui informações fornecidas pelo consumidor para a distribuidora de energia, que apesar de não serem usadas diretamente na análise das tipologias, permitem uma melhor definição do perfil daquele consumidor.

3.3 Tipologia das Curvas de Carga

Os valores de consumo de energia elétrica diária de cada usuário são armazenados pelas concessionárias após a leitura dos valores de consumo nos medidores. Através da análise dos valores de energia consumida, é possível classificar os consumidores por padrões de consumo. Independentemente do tipo de consumidor, o consumo de energia elétrica possui um comportamento sazonal e cíclico que pode ser demonstrado através das tipologias de curvas de carga quando é feita a análise de consumo dos usuários. O comportamento regular desta curva é chamado de padrão ou perfil de consumo.

As curvas de carga podem ser definidas como residenciais, industriais, comerciais ou de serviços, para estações quentes e frias, em dias de semana ou fins de semana, e são obtidas agrupando os perfis de carga de acordo com sua similaridade

[Azad *et al.* 2014]. Nos dados utilizados para esta pesquisa, os perfis de consumo diário dos consumidores são registrados por hora, com dados coletados em um intervalo de cinco minutos. Estes dados também foram divididos sazonalmente, pois há mudança na tipologia das curvas de um mesmo consumidor em função da estação.

Através da identificação de características comuns entre os *clusters* e através da análise das curvas de consumo de energia elétrica, é possível detectar padrões de consumo que podem ser usados para classificar consumidores e detectar anomalias nas redes de distribuição de energia elétrica [Queiroz *et al.* 2016]. Além disso, esses padrões podem ser refinados com a aplicação da rede neural não-supervisionada.

3.4. Knowledge Discovery in Databases e Redução de Dimensionalidade

O KDD é o processo geral de conversão de dados brutos em informações úteis. Ele consiste em uma série de passos de transformação, do pré-processamento dos dados até o pós-processamento dos resultados da Mineração de Dados [Pang-Ning *et al.* 2005].

A maioria dos métodos de Mineração de Dados é baseada na aprendizagem de máquina, reconhecimento de padrões e estatísticas, e possui um impacto direto sobre o desempenho e a eficácia de todo o procedimento. Estes métodos podem ser categorizados em diversos grupos, tais como classificação, análise de regressão, clusterização e sumarização [Tan *et al.* 2015]. Para este estudo foi utilizada a clusterização (*K-Means*) e a classificação com uma rede neural não-supervisionada (*Autoencoder*).

O *K-Means* define o elemento como se fosse um centróide (elemento central), o que normalmente é a média de um grupo de pontos, e é tipicamente aplicado a objetos em um espaço contínuo n-dimensional [Pang-Ning *et al.* 2005]. Ele toma aleatoriamente k pontos de dados (dados numéricos) como sendo os centróides dos *clusters*. Em seguida cada ponto (ou registro da base de dados) é atribuído ao cluster cuja distância deste ponto em relação ao centróide de cada *cluster* é a menor de todas as distâncias calculadas. Um novo centróide é computado pela média dos pontos, caracterizando a configuração do *cluster* para a interação seguinte. O processo termina quando os centróides param de se modificar, ou após um número limitado de iterações especificado pelo usuário [Goldschmidt and Passos 2005]. Os métodos centróides calculam a proximidade entre dois grupos calculando a distância entre os seus centróides.

Para conectar um ponto ao seu centróide mais próximo, é necessária a utilização de uma medida de aproximação que quantifica a noção de “próximo” para os dados especificados. A distância euclidiana é muitas vezes utilizada para pontos em um espaço euclidiano conforme apresentado em (1):

$$D_{(p,q)} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

Após a aplicação do *K-Means*, as curvas de carga são divididas em *clusters*. Porém, mesmo sendo uma busca que é reduzida uma base diária, ainda não há como identificar determinadas perdas não-técnicas caso elas ocorram eventualmente (como por exemplo, um determinado equipamento que é ligado apenas em dias específicos e está conectado na linha de transmissão fora do medidor). Sendo assim, dados de alta dimensão podem ser convertidos em dados de baixa dimensão treinando uma rede neural multicamadas para reconstruir vetores de entrada de alta dimensão [Hinton and Salakhutdinov 2006], chamado de *Autoencoder*.

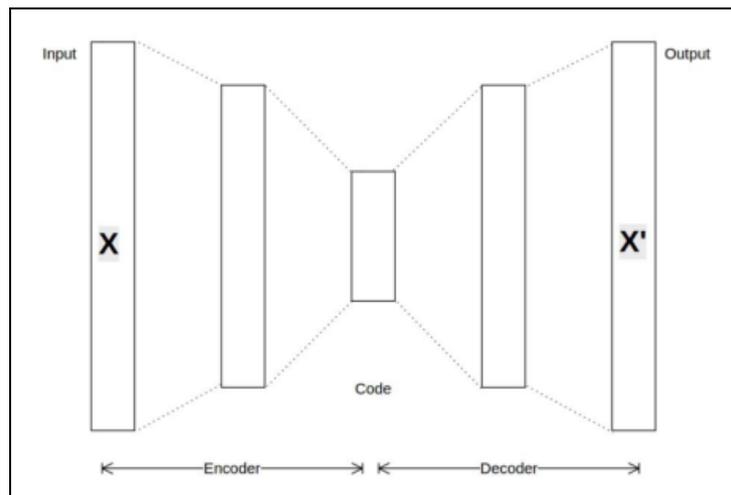


Figura 1. O leiaute de um Autoencoder

Conforme apresentado na Figura 1, um *Autoencoder* é um algoritmo de aprendizagem sem supervisão que tenta replicar a entrada depois de passá-la através de um gargalo de informações. É feito de várias camadas que primeiro comprimem o sinal, colocando-o através de um gargalo de informação. Este gargalo pode ser imposto por ter menos neurônios nas camadas sucessivas na primeira metade do *Autoencoder* (ou seja, o codificador). Também pode ser imposta por ter um critério de redução, pelo qual apenas uma certa fração de neurônios pode ser ativa. O modelo então tenta recriar o sinal o mais próximo possível do sinal original. Isso força o modelo a aprender recursos que podem ser usados para representar efetivamente os dados, apesar do gargalo da informação [Bhat *et al.* 2017]. O *Autoencoder* é treinado usando o algoritmo de *backpropagation*.

4. Resultados

Neste tópico são apresentados os resultados da clusterização *K-Means*, desde a seleção do número ideal de *clusters* até a clusterização dos dados, assim como o resultado da redução de dimensionalidade através da aplicação do *Autoencoder*.

Cada *cluster*, em cada uma das estações, apresenta um comportamento de consumo próprio e uma tipologia de curva de carga diferente para cada um dos consumidores. Porém, essas tipologias são similares entre si no *cluster* ao qual elas pertencem, pois os consumidores comerciais podem ter curvas de carga similares, independentemente da sua área de atuação, bastando apenas que o seu consumo em determinados horários siga padrões equivalentes aos de outro consumidor (o que caracteriza seu perfil de consumo).

Foram realizados testes para cada estação do ano com um número de *clusters* entre 2 (menor diluição possível) e 30 (maior diluição possível), visando encontrar o coeficiente de silhueta mais próximo de 0, mantendo-se um número de *clusters* que não diluísse demasiadamente as características entre as tipologias de curvas de carga. Para a demonstração deste processo, utilizou-se uma estação do ano como exemplo, a da Figura 1, que demonstra os coeficientes de silhueta para os dados durante os meses do inverno.



Figura 2. Análise dos coeficientes de silhueta para o inverno de 2 a 30 clusters

Coeficientes de silhueta próximos a 1 indicam que a amostra está longe da vizinhança. Um valor próximo a 0 indica que a amostra está dentro ou muito próxima do limite de decisão entre os *clusters* e valores negativos indicam que essas amostras foram associadas ao *cluster* errado. Na Figura 1, este coeficiente inicia-se em um valor mais próximo de 0 do que as outras estações analisadas nesta pesquisa (0.3528), o que significa que se não houvesse a necessidade de uma divisão de perfis por características mais específicas, muitas destas curvas de carga poderiam ser divididas em apenas dois *clusters*, considerando a similaridade de comportamento entre as tipologias. O valor de 15 *clusters* (mercado por um círculo vermelho), conforme apontado também pela clusterização hierárquica, foi selecionado para os dados do inverno pois o seu coeficiente de silhueta (0.3781) possui um valor médio mais próximo de 0 do que de 1, e não há uma diluição muito grande do número de características (como haveria dividindo os dados em mais de 20 *clusters*, por exemplo), o que permite um agrupamento por comportamento passível de estudo por esta pesquisa. A Tabela 1 apresenta os resultados dos números ideais de *clusters* apresentados para cada estação de acordo com a função de silhueta.

Tabela 1. Resultado do número ideal de clusters de acordo com a função de silhueta do K-Means

Verão	Outono	Inverno	Primavera
14	15	15	11

Na Figura 3 são apresentados exemplos de curvas de cargas dos *clusters* 8 e 10, para todos os dias da semana no período da primavera. Ela mostra que mesmo os consumidores possuindo perfis similares entre si, há diferenças nas tipologias das curvas

de cargas de cada *cluster*. Estas tipologias de curvas de carga representam, respectivamente e em sua maioria, as divisões CNAE 52 (mercados e lojas), 74 (assessorias contábeis e jurídicas), 70 (compra/locação de imóveis) e 85 (clínicas e hospitais). Os perfis agrupados em *clusters* indicam picos em horários de funcionamento de divisões CNAE que possuem um horário mais fixo (52, 74 e 70) e uma tipologia mais uníssona na 85, pois hospitais e clínicas possuem, em sua maioria, funcionamento 24 horas.

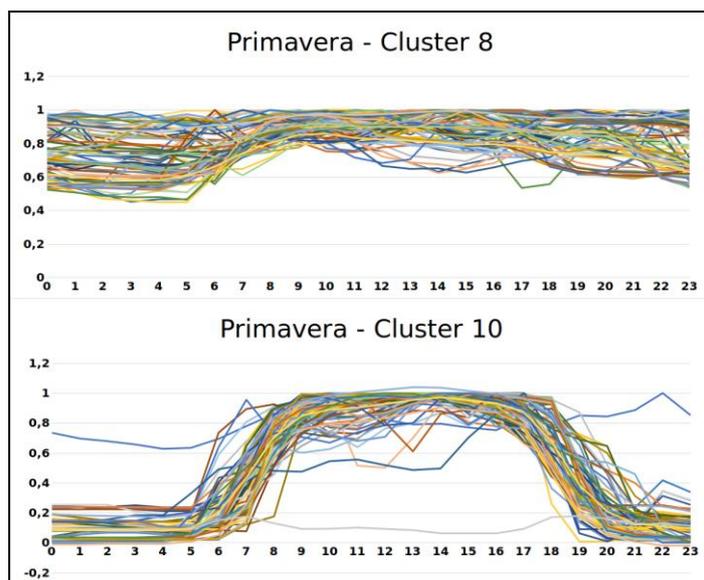


Figura 3. Perfis de consumo de dois clusters da primavera

Esta variação nas tipologias é dada pela diferença horária no consumo energético de cada consumidor comercial. O eixo *x* indica o horário, e o *y* o consumo normalizado pela fase de pré-processamento. Estes dados também foram divididos sazonalmente, pois há mudança no perfil de um mesmo consumidor em estações diferentes.

Tabela 2. Número de registros por cluster de acordo com cada estação.

Cluster/Estação	Verão	Outono	Inverno	Primavera
0	68	49	9	35
1	105	25	49	20
2	34	74	3	22
3	57	3	21	44
4	35	106	55	85
5	26	11	91	84
6	39	29	25	26
7	18	42	96	78
8	51	52	52	77
9	49	5	62	37
10	40	67	20	80
11	23	51	21	-
12	16	85	25	-
13	57	5	35	-
14	-	10	27	-

O perfil de consumo do usuário mantém um padrão similar durante todos os dias, pois as atividades comerciais dos consumidores utilizados neste estudo são regulares (ou seja, possuem intervalos definidos de início e fim, com interrupções sempre nos mesmos horários). As curvas de carga da Figura 3 apresentam o comportamento de cada consumidor que possui um perfil de consumo compatível com o do *cluster* no qual foi alocado. A Tabela 2 demonstra o número de registros por *cluster* utilizado para o estudo, de acordo com a estação.

Para estes resultados foram comparados todos os registros, visando identificar alguma dissociação entre os consumidores e os *clusters* aos quais os mesmos pertencem, ou seja, se durante a análise de todo o grupo de dados, algum consumidor alocava-se em um *cluster* diferente ao do grupo ao qual se associou originalmente.

Após a aplicação do *K-Means* sobre os dados da distribuidora de energia, detectou-se que muitos consumidores tendem a permanecer nos mesmos *clusters* durante todas as estações, ou seja, independentemente da sazonalidade, as variáveis climáticas aplicam sobre todos os membros daquele grupo o mesmo modificador, fazendo com que os mesmos mantenham seus perfis semelhantes durante todo o ano (e consequentemente, compartilhem o mesmo *cluster* em todas as estações).

Os perfis foram agrupados por *clusters*, de acordo com seu consumo em relação à todos os dias da semana. Com a aplicação do *Autoencoder*, utilizando 70% do volume de dados como treinamento e 30% como teste, percebeu-se que determinados dias, mesmo em um consumidor aparentemente constante durante a semana inteira, seu consumo poderia variar. A rede neural não-supervisionada, quando foi aplicada à dados já clusterizados através do *K-Means*, conseguiu reduzir a dimensionalidade, permitindo que distorções pudessem ser visualizadas em dias da semana específicos. A Figura 4 apresenta um perfil discrepante, especificamente às quintas feiras, para um consumidor específico cuja curva deveria seguir o mesmo padrão indicado nas linhas pontilhadas.

Em todos os agrupamentos, a maioria dos consumidores (70% ou mais) permanece junto aos outros consumidores do mesmo *cluster* por todo o ano, pois como a análise foi feita em clientes da distribuidora de uma mesma região, o impacto da variação climática é similar para todos. Quando a clusterização é aplicada, apenas perfis de consumo com uma discrepância evidente do *cluster* ao qual ele pertence são mostrados. Porém, com a aplicação do *Autoencoder* (feito com o uso de MATLAB), foi possível identificar curvas de cargas que possuíam padrões fraudulentos apenas em dias específicos, conforme apresentado na Figura 4. Assim, após o seu treinamento, a rede neural demonstrou curvas de cargas individuais em dias específicos que possuíam um comportamento susperito. A Figura 4 apresenta os desvios padrões superior e inferior de um consumidor no inverno, em relação à média do *cluster*, representada pela linha azul. Os círculos vermelhos indicam os momentos do dia em que houveram desvios por parte do usuário, aproximadamente às 06:00h, 14:00h, 16:00h e 19:00h. Essas diferenças no comportamento do perfil da curva de carga podem indicar que houve algum tipo de adulteração ou desvio no medidor, fazendo com que parte da energia consumida não fosse processada pelo mesmo.

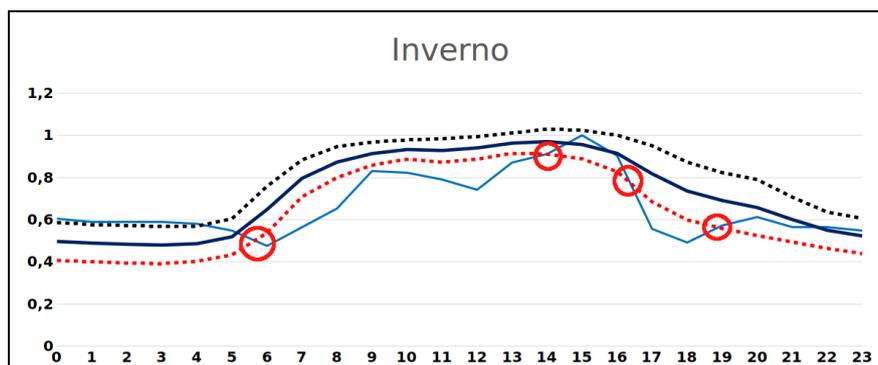


Figura 4. Média da tipologia das curvas de carga durante as quintas-feiras para o inverno no cluster 5.

Sendo assim, dentro do escopo desta pesquisa, qualquer usuário que apresentar desvios fora do padrão pode ser considerado suspeito de praticar fraude. De acordo com os resultados deste trabalho, então, as perdas não-técnicas podem ser inicialmente (pois maiores conjuntos de dados podem ser utilizados) identificadas pela sua diferença de comportamento em relação ao *cluster* ao qual esse usuário pertence, e se o mesmo possui desvios nesse padrão durante o horário comercial.

5. Conclusões

Considerando-se a quantidade de dados armazenados diariamente com os registros das leituras mensais de consumo de energia elétrica, é possível afirmar que a aplicação de mineração de dados e aprendizado de máquina para transformar tais dados em conhecimento pode auxiliar na identificação de distorções nos perfis de consumo dos usuários desta distribuidora.

Dentre os resultados obtidos através da clusterização permitiu-se, então, a divisão de consumidores pelo seu perfil de consumo (representado pela tipologia da curva de carga), possibilitando assim identificar quais elementos definem e diferenciam um consumidor que pertence à um determinado perfil. A classificação através do *Autoencoder* ampliou o nível de detalhamento, permitindo que um perfil pudesse ser analisado até dentro dos dias da semana.

Conclui-se, através da análise dos resultados obtidos, que determinados consumidores modificam o seu comportamento mesmo sofrendo influência da mesma variável, a sazonalidade. Como consumos referentes a dias típicos, tais como feriados, por exemplo, foram filtrados na fase de pré-processamento, infere-se que tais mudanças possam evidenciar possíveis distorções no consumo.

As variáveis que envolvem a modificação de uma tipologia estão associadas à uma variação frequente de consumo (tal como a prática de fraude), o que significa que os clientes da distribuidoras que estão apresentando uma diferença muito evidente no seu perfil, pode representar um comportamento suspeito que precisa ser submetido à uma avaliação mais detalhada. Apesar desta metodologia não garantir que um usuário possa ser responsável pelas perdas não-técnicas, os resultados obtidos com a aplicação da mesma pode direcionar as fiscalizações manuais que são realizadas pelas concessionárias de energia a fim de encontrar possíveis fraudes nas redes de distribuição de energia elétrica.

Referências

- Azad, S. A., Ali, A. B. M. S. and Wolfs, P. (nov 2014). Identification of Typical Load Profiles using K-means Clustering Algorithm. In *Asia-Pacific World Congress on Computer Science and Engineering*. IEEE. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5301726>.
- Bhat, R. R., Trevizan, R. D., Sengupta, R., Li, X. and Bretas, A. (2017). Identifying nontechnical power loss via spatial and temporal deep learning. *Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, n. November 2017, p. 272–279.
- EPE (2016). Consumo de Energia Elétrica por Classe (Regiões e Subregiões). [http://www.epe.gov.br/mercado/Paginas/Consumomensaldeenergiaelétricaporclasse\(regiõeses\),](http://www.epe.gov.br/mercado/Paginas/Consumomensaldeenergiaelétricaporclasse(regiõeses),) [accessed on May 1].
- FIESC (2015). Santa Catarina em Dados 2015. http://fiesc.com.br/sites/default/files/medias/sc_em_dados_site_correto.pdf, [accessed on Mar 21].
- Guerrero, J. I., León, C., Monedero, I., Biscarri, F. and Biscarri, J. (2014). Improving Knowledge-Based Systems with Statistical Techniques, Text Mining, and Neural Networks for Non-Technical Loss Detection. *Knowledge-Based Systems*, v. 71, p. 376–388.
- Hinton, G. E. and Salakhutdinov, R. R. (28 jul 2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, v. 313, n. 5786, p. 504–507.
- Nizar, A. H., Dong, Z. Y., Jalaluddin, M. and Raffles, M. J. (nov 2006). Load Profiling Method in Detecting non-Technical Loss Activities in a Power Utility. In *2006 IEEE International Power and Energy Conference*. IEEE. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4154468>.
- Pang-Ning, T., Vipin, K. and Steinbach, M. (2005). *Introduction to Data Mining*. Boston: Pearson Addison Wesley.
- Queiroz, A. de S., Franco, E. M. C. and López, G. P. (2016). Detecção de Fraudes nas Redes de Distribuição de Energia Elétrica Utilizando Técnicas de Inteligência Computacional. In *VI Simpósio Brasileiro de Sistemas Elétricos*.
- SPE (2013). SIASE – Sistema de Inteligência Analítica do Setor Elétrico. [http://www2.aneel.gov.br/arquivos/PDF/PD Estratégico 018-2013_SIASE.pdf](http://www2.aneel.gov.br/arquivos/PDF/PD%20Estrat%C3%A9gico%20018-2013_SIASE.pdf).
- Tan, Y., Zhang, C., Ma, Y. and Mao, Y. (jun 2015). Knowledge discovery in databases based on deep neural networks. In *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE. <http://ieeexplore.ieee.org/document/7334293/>.