

Detecção de Irregularidades no Consumo de Energia Elétrica Usando Árvores de Decisão

Y. C. S. Correa, R. L. S. Carvalho, F. P. Monteiro,
A. S. Tobias, J. P. A. Vieira, M. V. A. Nunes,
U. H. Bezerra, M. E. L. Tostes
CEAMAZON/UFPA
Belém, Brasil

R. C. de
Oliveira
CELPA
Belém, Brasil

Abstract— Este artigo visa apresentar uma metodologia de detecção de irregularidades no consumo de energia elétrica em redes de distribuição usando a técnica de mineração de dados árvore de decisão (AD). A técnica de AD foi utilizada para classificar as unidades consumidoras (UCs) com padrões de irregularidades. Através da metodologia Descoberta de Conhecimento em Base de Dados ou (KDD, do inglês “*Knowledge Discovery in Databases*”), o trabalho apresenta duas abordagens para aplicação de técnicas de pré-processamento visando à otimização do desempenho do classificador. Testes preliminares foram realizados utilizando dados reais do histórico de consumo de energia elétrica e de fiscalização de irregularidades em UCs localizadas na cidade de Belém, PA. Os resultados obtidos mostraram que o método desenvolvido possui elevado desempenho na tarefa de detectar irregularidades em UCs.

Index Terms—Irregularidades no consumo de energia, fraudes, perdas comerciais, mineração de dados, árvore de decisão.

I. INTRODUÇÃO

O mercado de distribuição de energia elétrica vem sofrendo com sérios prejuízos ao longo do tempo devido às perdas comerciais, as quais são o maior fator de diminuição de receita das empresas. Segundo levantamento da Agência Nacional de Energia Elétrica (ANEEL), o prejuízo com as perdas comerciais de energia elétrica no Brasil é de R\$ 7 bilhões de reais ao ano, o que representa em torno de 13% da energia consumida [1]. Diante de valores estratosféricos, é fundamental combater às perdas comerciais de maneira eficiente, segura e rápida. Atualmente, o combate é realizado por meio de inspeções nas unidades consumidoras (UCs), entretanto, a quantidade de UCs, por alimentador, é muito grande. Além disso, a análise sobre o perfil de consumo dos clientes não é realizada com eficiência, principalmente em áreas cuja irregularidade no consumo de energia não é evidente.

A redução de perdas comerciais é um fator bastante relevante para as distribuidoras de energia elétrica; porém, a identificação das UCs com comportamento fraudulento ou com problemas na medição é uma tarefa complexa, uma vez que envolve inspeções aleatórias *in loco*.

Dessa forma, é ideal encontrar um perfil que indique um comportamento suspeito, em que o especialista possa recomendar que a UC seja inspecionada a partir da

descoberta de padrões realizada por meios de técnicas de mineração de dados.

A mineração de dados ou *data mining* é o processo realizado através de estratégias automatizadas que tem como objetivo a descoberta de conhecimento valioso em grandes bases de dados. A mineração de dados baseia-se na utilização de técnicas capazes de “vasculhar” grandes bases de dados de modo eficiente e revelar padrões interessantes. Vários trabalhos mostraram aplicações de mineração de dados ou de técnicas de inteligência computacional na detecção de irregularidades ou fraudes no consumo de energia [2]-[13].

A árvore de decisão é uma técnica de mineração de dados que realiza o processo de aprendizagem ao utilizar algoritmos para a tarefa de classificação. É uma representação simples do conhecimento que constrói classificadores que predizem classes baseadas nos valores de atributos de um conjunto de dados.

O artigo aborda as perdas comerciais e seus impactos econômicos nas empresas distribuidoras, e a necessidade da aplicação de técnicas de mineração de dados, como ferramenta para detecção de irregularidades no consumo de energia elétrica. No presente artigo são utilizados dados do histórico de 4 anos de consumo de energia elétrica e dados de fiscalização de irregularidades em UC's, os quais foram fornecidos pela CELPA. Os resultados preliminares obtidos mostraram que a abordagem proposta baseada em árvores de decisão apresenta elevada taxa de acertos para classificar padrões de irregularidades em UC's.

II. PERDAS NO SISTEMA ELÉTRICO

Para facilitar a investigação na caracterização de perdas é importante dividir as mesmas em diferentes categorias, sendo habitual fazer a distinção entre perdas técnicas e não técnicas. As perdas técnicas ocorrem nos equipamentos elétricos, especialmente em linhas e transformadores. As perdas não técnicas estão associadas a fenômenos não esperados no funcionamento normal do SEE.

A. Perdas Técnicas

Constituem na quantidade de energia elétrica dissipada entre os suprimentos de energia da distribuidora e os pontos de entrega nas instalações das unidades consumidoras ou distribuidoras supridas. São derivadas de fenômeno físico

associado à transferência de energia elétrica das centrais de geração até os consumidores. Neste processo, uma parte da energia é dissipada, essencialmente em condutores e transformadores. Assim, as perdas ocorrem nos mais diversos equipamentos elétricos e podem ser divididos em dois grupos:

- Perdas Ativas: são resultantes do efeito Joule, constituindo normalmente a maior parte das perdas técnicas e estão presentes nos condutores elétricos. Essas perdas são proporcionais ao quadrado da corrente elétrica que percorre nos condutores, pela resistência do mesmo. No caso da transmissão de energia em corrente alternada, o efeito pelicular [6] pode contribuir para o aumento das perdas, através do aumento da resistência do condutor. Pois, com este fenômeno, a corrente tende a ser maior na superfície do condutor e menor no seu eixo, assim, alterando a secção útil do condutor.
- Perdas Reativas: ocorrem nas reatâncias dos equipamentos. Nos sistemas de transmissão em corrente alternada as reatâncias armazenam energia em cada meio ciclo, para em seguida devolver essa mesma energia à fonte. Pelo fato referido anteriormente, a energia reativa em questão não será efetivamente perdida, sendo, contudo necessária para o funcionamento dos componentes. Embora não cause perdas de forma direta, também as influencia, uma vez que contribui para o aumento da corrente nas linhas. Os motores elétricos são um exemplo de cargas que consomem uma energia reativa significativa, contribuindo para o aumento da componente reativa na rede.

B. Perdas Não-Técnicas

As perdas não técnicas, também denominadas por perdas comerciais, contribuem para o acréscimo da energia fornecida pelo distribuidor que posteriormente não é faturada. Essa perdas são apuradas pela diferença entre as perdas totais e as perdas técnicas, considerando, portanto, todas as demais perdas associadas à distribuição de energia elétrica. As principais causas das perdas não técnicas são as seguintes:

- Ausência ou erros de medição;
- Erros de leitura, processamento,
- Consumo estimado e falta de medição;
- Desvio antes da medição;
- Desvios ou ligações clandestinas (furtos de energia);
- Fraude ou irregularidades nos equipamentos de medição.

Essas perdas estão diretamente associadas à gestão comercial da distribuidora.

C. Fatores de influência em fraudes e furtos

Ao contrário das perdas técnicas, as perdas não técnicas não deveriam ocorrer, sendo, portanto, “inaceitáveis nos sistemas elétricos”. De acordo com [7], dentre outros fatores, as perdas comerciais têm expressiva participação de diversos fatores exógenos ao setor elétrico, como: a questão social, nível de emprego e renda; composição da balança comercial regional e nacional. Em [8], o autor busca entender como o problema de perdas comerciais é tratado em diversos países e propõe que as diferenças de perdas de energia elétrica estejam associadas a problemas de governança, tais como: aspectos de processos políticos, liberdade civil, burocracia, independência dos serviços públicos às pressões políticas, instabilidade política e violência, além da efetividade do judiciário na aplicação da lei, mostrando a correlação entre estes aspectos e o nível de perdas de energia elétrica.

A administração das perdas comerciais é de alta complexidade, pois depende das características socioeconômicas da região atendida, no que diz respeito ao comportamento dos usuários, o qual está diretamente relacionado com o desenvolvimento social, econômico e educacional da população. Por isso, considerando os fatores exógenos, as perdas não técnicas podem ser influenciadas por: inadimplência, questão sociocultural, como educação, renda, desigualdade, área de concessão, tarifa, dentre outras

III. ÁRVORE DE DECISÃO

As árvores de decisão são algoritmos de classificação de dados no ambiente da chamada Mineração de Dados (*Data Mining*). E tem como uma importante característica a função de particionar de modo recursivo um conjunto complexo de dados, até que cada subconjunto assim obtido contenha apenas os casos de uma única classe possibilitando uma melhor análise. Após a construção da árvore de decisão, os resultados obtidos, são formados por dados organizados de maneira simples e de fácil entendimento e podem servir como importante ferramenta para tomada de decisão.

Os primeiros classificadores baseados em árvore de decisão surgiram no final da década de 50, a partir do trabalho de Hunt, foram apresentados vários experimentos para indução de regras. Posteriormente, foram desenvolvidos os seguintes algoritmos: CART (*Classification And Regression Trees*) [14], ID3 (*Iterative Dichotomiser 3*) e o C4.5.

As árvores de decisão são construídas baseadas no modelo hierárquico *top-down*, isto é, do nó raiz em direção às folhas. Embora haja diferenças na forma de realizar os passos, os algoritmos utilizam a técnica dividir para conquistar. Esta técnica está baseada nas sucessivas divisões do problema estudado em vários subproblemas de menores dimensões, até que seja encontrada uma solução mais simples para cada um dos problemas.

Uma árvore de decisão é essencialmente uma série de declarações *if-then*, que quando aplicados a um registro de uma base de dados, resultam na classificação daquele registro. A Fig 1, representa um exemplo de uma árvore hipotética, e é constituída dos seguintes elementos:

- Raiz: é o nó do topo da árvore.
- Nós: são todos os elementos que estão conectados por ramos;
- Ramos: Ligações entre nós;
- Folhas: são os últimos nós da árvore e representam as classes do conjunto de dados.

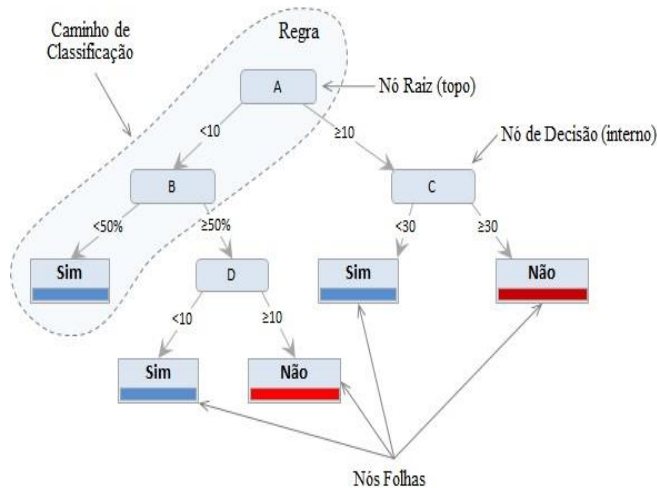


Figura 1. Exemplo de uma árvore de decisão

Os algoritmos J48, Random Forest e SimpleCART estão entre os mais conhecidos, e foram selecionados para a aplicação da metodologia proposta neste artigo.

IV. METODOLOGIAS

A Fig.2 ilustra o fluxograma das metodologias adotadas no presente trabalho.

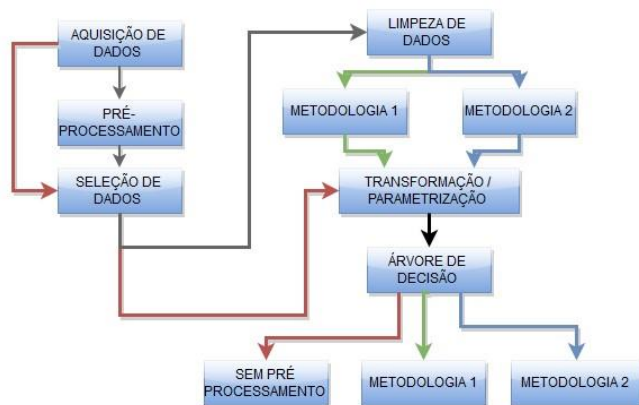


Figura 2. Fluxograma da Metodologia

A. Aquisição de dados

Para início da análise a rede Celpa disponibilizou tabelas com informações gerais das UC's, histórico de consumo do

alimentador PD05(2010 até 2013) com total de 4.728 UC's com 31 atributos sendo no total dos 4 anos 22.180 amostras no formato Excel e "csv". Os dados das fiscalizações (2013 e 2014) da região metropolitana de Belém.

Além dos dados de consumo, a Celpa forneceu os dados de relação das fiscalizações sobre irregularidade nos anos de 2013 e 2014, totalizando 335.429 amostras, com 176.203, no ano de 2013, e 159.226, em 2014. Contendo informação das UCs, tipo de irregularidade, data da fiscalização, código de retorno da fiscalização e informação sobre troca medidor. Os dados para a coluna irregularidade são divididas com valor "0" para caso de irregularidade não comprovada e valor "1" para unidades com alguma irregularidade comprovada.

Após cruzamento dos dados das tabelas de consumo e tabela do alimentador PD05 e de fiscalização, permaneceram apenas 887 amostras.

B. Pré-Processamento de Dados

O processo de preparação dos dados para a mineração pode ser denominado de pré-processamento e essa etapa compreende funções relacionadas à captação, organização, tratamento e à preparação da base de dados para a etapa de Data Mining.

Como a base de dados é pequena, a fase de pré processamento vem antes da fase de seleção dos dados.

Técnicas de pré-processamento são aplicadas para aumentar a qualidade e o poder de expressão dos dados a serem minerados. Esta fase busca eliminar os dados que estão "sujeitos", limpá-los, no caso com algum dos seguintes problemas Incompletos, valores ou atributos faltando, Ruidosos, contendo erros ou outliers, Inconsistentes, contendo discrepâncias em nome e na codificação.

C. Limpeza de Dados

As inconsistências são mineradas para a eliminação de linhas duplicadas, valores vazios, consumo zero, padronização de séries temporais e eliminação de outliers. Esse processo é denominado de Limpeza de Dados, em que para a execução, foram utilizadas "técnicas de limpeza" para o tratamento de valores ausentes e dados ruidosos. Com objetivo eliminar amostras que apresentem valores ausentes e valores mensais com consumo igual a zero. Para essa etapa, optou-se por fazer uso de duas metodologias e tipos diferentes de ferramentas de limpeza de dados: Manual e Automático.

D. Metodologia 1

Na primeira metodologia, o objetivo é eliminar amostras que apresentem valores ausentes e valores mensais com consumo igual a zero. Foi, então, utilizada a ferramenta OpenRefine (anteriormente Google Refine), que proporciona a limpeza e transformação dos dados para outro formato. É eficiente para trabalhar com dados desorganizados, pois nela é possível padronizar, filtrar e limpar bases de dados.

Após o término do processo de limpeza dos dados pelo primeiro método, as amostras, que estavam num total de 887, tiveram uma diminuição para o número de 790.

E. Metodologia 2

O objetivo da segunda metodologia adotada consiste em não descartar as instâncias que possuísem valores ausentes e ruídos, mas implementar formas de tratamento para esses tipos de dados. Assim, a base de dados não teria seu pequeno número reduzido ainda mais. Para essa tarefa, foi usado o software de mineração de dados *KEEL*, do inglês *Knowledge Extraction Based on Evolutionary Learning*.

A base de dados teve seus dados ausentes tratados pelo módulo *FKMeans*, que corresponde ao algoritmo de imputação *Fuzzy K-means*. Esse algoritmo trata valores ausentes por meio de computação de *k* fuzzy grupos de dados, atribuindo o valor à instância, dependendo do grau de pertinência da mesma para cada cluster. O algoritmo *Fuzzy K-means* é preparado para trabalhar com variáveis nominais, variáveis contínuas e discretizadas. Sendo assim, é possível utilizá-lo na base de dados.

Nas duas fases seguintes do pré-processamento, foram utilizados os filtros *Normalize* e *Resample*. O filtro *Normalize* tem como propósito organizar os dados na mesma ordem de grandeza. Se não houvesse essa normalização, poderiam existir grandezas que teriam mais importância que outras — o que não será interessante para o classificador. O filtro *Resample* cria um subconjunto de valores aleatórios do conjunto de dados original, com reposição. Essa fase foi utilizada por haver desbalanceamento entre as classes de consumo, podendo comprometer a acurácia do classificador, em razão de um possível enviesamento do modelo.

F. Transformação dos Dados

Esta é a fase em que os dados são transformados e consolidados em formatos apropriados para a mineração. A organização deles é importante para que possam ser usados de forma eficiente, permitindo que haja singularidade, diminuição do uso da memória *RAM*, tornando o desenvolvimento do código do software mais transparente, simples e enxuto. A escolha da organização apropriada do banco de dados pode transformar um problema complicado num de solução relativamente simples.

O *PostgreSQL* foi o sistema de gerenciamento de banco de dados escolhido para o trabalho.

Para a fase de execução dos algoritmos de classificações foi necessário a criação de 12 visões para guardar valores calculados dos desvios padrões do histórico mensal de cada consumidor.

V. RESULTADOS

Nesse artigo, as simulações foram realizadas a partir da tarefa de classificação, com a *Árvore de Decisão*, utilizando

os algoritmos *J48*, *SimpleCART* e *Random Forest*, sendo a irregularidade tomada como atributo de saída para determinar se uma determinada unidade consumidora possui ou não alguma irregularidade.

A. Classificação com Dados não Processados

A primeira simulação foi feita com a base de dados não processados, estando presentes: os valores ausentes, as linhas duplicadas e os valores ruidosos. Essa base contém 887 amostras e 21 atributos. O resultado da classificação é mostrado na tabela I.

Algoritmos	Taxa de Acerto	Taxa de Erro	Índice Kappa
J48	69,7858%	30,2142%	0,3724
SimpleCART	71,3641%	28,6359%	0,4058
Random Forest	62,2322%	37,7678%	0,229

Tabela I. Acurácia dos algoritmos com dados não processados.

A classificação realizada com o algoritmo *SimpleCART* obteve os melhores resultados, em relação aos outros dois algoritmos, com taxa de acerto de 71,36%.

B. Classificação com Dados Processados

A classificação feita com os dados pré-processados, foi realizada seguindo o processo completo de mineração de dados e como dito na seção anterior, com duas metodologias, sendo a primeira utilizando no processo de limpeza dos dados, o software *OpenRefine*. Os resultados obtidos são mostrados na tabela II.

A primeira metodologia obteve acurácia 11,67% melhor que a metodologia anterior, cujos dados não pré-processados. O algoritmo que apresentou melhor desempenho de classificação foi o *Random Forest*, com 83,0380% de taxa de acerto e melhor índice *Kappa*. O índice *Kappa* tem como valor máximo 1, representando total concordância, até o valor 0, indicando nenhuma concordância.

Algoritmos	Taxa de Acerto	Taxa de Erro	Índice Kappa
J48	77,0886%	22,9114%	0,5319
SimpleCART	75,443%	24,557%	0,4964
Random Forest	83,0380%	16,962%	0,6504

Tabela II. Acurácia dos algoritmos com dados processados

A aplicação da segunda metodologia rendeu resultados melhores, todos acima de 80%. Sendo o melhor, dentre os três algoritmos, novamente, o *Random Forest*, com 847 instâncias classificadas corretamente, o que significa 85,47% de taxa de acerto. Nesta segunda metodologia o *Random Forrest* também apresentou o melhor índice *Kappa*. Esses resultados podem ser conferidos na tabela III.

A figura 3 ilustra um dos ramos da árvore de decisão criada, a partir do uso da metodologia 2 (algoritmo *J48*), com a finalidade de apresentar os atributos mais críticos com relação a classificação das irregularidades das *UC's*. Como

pode ser observado na figura 6, o nó raiz é formado pelo atributo Tipo de fase de ligação, e derivam 3 ramos: monofásico, bifásico e trifásico. O ramo trifásico foi escolhido para análise, já que era o ramo que continha mais informações, como as classes de consumo. Em seguida, o nó subsequente é a classe de consumo: atributo importante na ligação trifásica, pois dela derivam ramos com os tipos de classe, como: residencial, comercial, serviços e outras atividades e, a partir desses são derivados o atributo desvio padrão de meses.

Algoritmos	Taxa de Acerto	Taxa de Erro	Índice Kappa
J48	82,2402%	17,7598%	0,6377
SimpleCART	80,2220%	19,778%	0,5964
Random Forest	85,4692%	14,5308%	0,7013

Tabela III. Acurácia dos algoritmos com dados não processados.

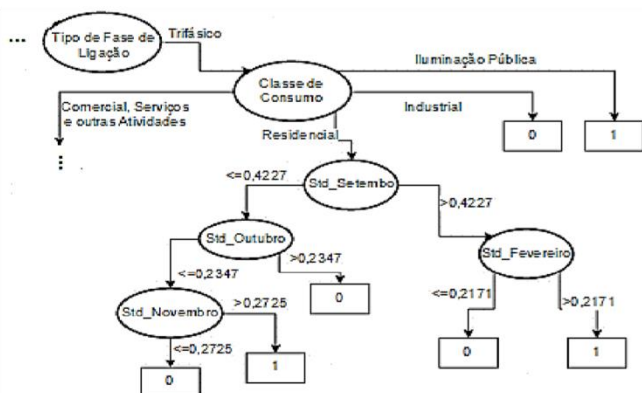


Figura 3. Ramo da árvore de decisão do algoritmo J48 com a aplicação da segunda metodologia e os principais atributos da ligação trifásica.

Analisando o ramo residencial mostrado na figura 5, pode ser destacar que o perfil de consumo pode ser irregular em duas situações; quando o desvio padrão dos meses de setembro e fevereiro são maiores que 0,4327 e 0,2171, respectivamente, ou quando o desvio padrão do mês de outubro é menor ou igual a 0,2347 e o desvio padrão de novembro é maior que 0,2725. Essas regras fornecidas pela árvore de decisão podem servir como diretrizes para tomada de ação no combate às fraudes e furtos no consumo de energia elétrica, que estão entre as principais irregularidades em UC's.

VI. CONCLUSÕES

Os resultados obtidos, a partir das classificações realizadas com os algoritmos da árvore de decisão foram considerados de satisfatórios a bons, uma vez que as metodologias utilizadas apresentaram taxas de acerto que variaram de 62,23% a 85,46%. A metodologia que não fez uso do procedimento de pré-processamento dos dados apresentou desempenho satisfatório como pode ser observado na tabela IV. Por outro lado, a aplicação da fase de pré-processamento

nas metodologias 1 e 2 proporcionou, como pode ser observado nas tabelas II e III, melhor desempenho quando comparado ao método que não fez uso dessa fase de pré-processamento. A metodologia 2 que consiste de processos mais específicos, como uso da ferramenta KEEL e dos filtros Normalize e Resample, apresentou melhoria em relação à metodologia 1, em pouco mais de 2%, quando em ambos os casos é utilizado o algoritmo Random Forrest. Enquanto que a melhoria obtida pela metodologia 2 em relação a metodologia 1 para os algoritmos J48 e SimpleCART, foram de 5,15% e 4,77%, respectivamente. A próxima etapa de desenvolvimento do projeto de P&D é considerar o histórico de consumo de energia elétrica de toda a região metropolitana de Belém, PA, no sentido de aumentar a precisão dos classificadores baseados em árvores de decisão.

REFERENCIAS

- [1] Disponível em: www.aneel.gov.br
- [2] A. H. Nizar, Z. Y. Dong, and Y. Wang, "Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Method", *IEEE Transactions on Power Systems*, Vol. 23, no.3, pp. 946-955, Aug, 2008.
- [3] J. Filho, "Fraud identification in electricity company costumers using decision tree," in *Proc. IEEE/PES Int. Conf. Systems, Man and Cybernetics*, The Hague, The Netherlands, 2004.
- [4] C. León, F. Biscarri, I. Monedero, J. I. Guerrero, J. Biscarri, and R. Millán, "Variability and trend-base generalized rule induction model to NTL detection in power companies," *IEEE Trans. Power Syst.*, vol.26, no. 4, pp. 1798-1807, Nov. 2011.
- [5] I. Monedero, F. Biscarri, C. León, J. I. Guerrero, J. Biscarri, and R. Millán, "Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees", *Electrical Power and Energy Systems*, Vol. 34, p-90-98, 2012.
- [6] COSTA. A. da P.O "Efeito pelicular ou efeito skinl. disponível em: <http://www.estv.ipv.pt/PaginasPessoais/paulomoses>.
- [7] VIEIRALVES, E. X. Proposta de uma metodologia para avaliação das perdas comerciais dos sistemas elétricos: o caso Manaus. 2005. 180 p. Dissertação (Mestrado em Planejamento de Sistemas Energéticos) – Universidade Estadual de Campinas, Campinas, 2005.
- [8] SMITH, T. B. (2004). "Electricity theft: a comparative analysis". *Energy Policy* 32: 2067-2076.
- [9] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines", *IEEE Transactions on Power Delivery*, Vol. 25, no.2, pp. 1162-1171, April, 2010.
- [10] C. C.O. Ramos, A. N. Souza, A. X. Falcão, and J. P. Papa, "A New Approach for Nontechnical Losses Detection Based on Optimum-Path Forest", *IEEE Transactions on Power Systems*, Vol. 26, no.1, pp. 181-189, Feb, 2011.
- [11] C. León, F. Biscarri, I. Monedero, J. I. Guerrero, J. Biscarri, and R. Millán, "Integrated expert system applied to the analysis of non-technical losses in power utilities", *Expert Systems with Applications*, Vol.38, no 3, pp.10274-10285, 2011.
- [12] C. C.O. Ramos, A. N. Souza, G. Chiachia, A. X. Falcão, and J. P. Papa, "A novel algorithm for feature selection using Harmony Search and its application for non-technical losses detection", *Computers and Electrical Engineering*, Vol. 37, pp.886-894, 2011.
- [13] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and F. Nagi, "Improving SVM-Based Nontechnical Loss Detection in Power Utility Using the Fuzzy Inference System", *IEEE Transactions on Power Delivery*, Vol. 26, no.2, pp. 1284-1285, April, 2011.
- [14] E. W. S. dos Angelos, O. R. Saavedra, O. A. C. Cortés, and A. N. de Souza, "Detection and Identification of Abnormalities in Customer Consumptions in Power Distribution Systems", *IEEE Transactions on Power Delivery*, Vol. 26, no.4, pp. 2436-2442, Oct, 2011.
- [15] C. León, F. Biscarri, I. Monedero, J. I. Guerrero, J. Biscarri, and R. Millán, "Variability and Trend-Based Generalized Rule Induction Model to NTL

- Detection in Power Companies”, *IEEE Transactions on Power Systems*, Vol. 26, no.4, pp. 1798-1807, Nov, 2011.
- [16] C. C.O. Ramos, A. N. Souza, A. X. Falcão, and J. P. Papa, “New Insights on Nontechnical Losses Characterization Through EvolutionaryBased Feature Selection”, *IEEE Transactions on Power Delivery*, Vol. 27, no.1, pp. 140-146, Jan, 2012.
- [17] A. M. S. Ferreira, C. A. M. T. Cavalcante, C. H. O. Fontes, J. E. S. Marambio, “A new method for pattern recognition in load profiles to support decision-making in the management of the electric sector ”, *Electrical Power and Energy Systems*, Vol. 53, p-824-831, 2013.
- [18] I. Benítez, A. Quijano, J. L. Díez, I. Delgado, “Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers ”, *Electrical Power and Energy Systems*, Vol. 55, p-437-448, 2014.
- [19] L. Breiman, J. Friedman, R. Olshen, C. Stone, *Classification and Regression Trees*. London, U.K.: Chapman and Hall/CRC, 1984.