Journal of Optimization in Industrial Engineering
Vol.15, Issue 2, Summer & Autumn 2022, 67- 73
Doi: 10.22094/joie.2022.1941123.1894

Research Article

Electricity Theft Detection in Power Utilities Using Bagged CHAID-Based Classification Trees

Muhammad S. Saeed [a,*] Mohd. Wazir Bin Mustafa [a], Usman Ullah Sheikh [a], Attaullah Khidrani [a], Mohd Norzali Haji Mohd [a]

[a] *School of Electrical Engineering, University Technology Malaysia, Skudai, Johor Bahru 81310, Malaysia*

Received 23 September 2021; Revised 09 November 2021; Accepted 07 March 2022

Abstract

Electricity theft and fraud in billing are the primary concerns for Distribution System Operators (DSO). It is estimated that billions of dollars are lost each year due to these illegal activities. DSOs around the world, especially in underdeveloped countries, are still utilizing conventional time consuming and inefficient methods for Non-Technical Loss (NTL) detection. This research work attempts to solve the mentioned problems by developing an efficient energy theft detection model to identify the fraudster customers in a power distribution system. The key motivation for the current study is to assist the DSOs for their campaign against energy theft. The proposed method initially utilizes the monthly consumption data of energy customers, obtained from Multan Electric Power Company (MEPCO) Pakistan, to segregate the honest and the fraudulent customers. The Bagged Chi-square Automatic Interaction Detection (CHAID) Decision Tree (DT) algorithm is used to classify the honest and fraudster consumers. Furthermore, based on the mentioned metrics, the performance superiority of the Bagged CHAID-based NTL detection method is validated by comparing its efficacy with that of few well-known state-of-the-art machine learning algorithms such as Support Vector Machine (SVM), Artificial Neural Network (ANN) Logistic Regression (LR), Bayesian Network (BN) and Discriminant Analysis. The proposed NTL detection method provides an Accuracy of 86.35% and Area Under Curve (AUC) of 0.927, respectively, which are significantly higher than that of the same for the mentioned algorithms.

Keywords:Electricity theft; fraud billing; Non-Technical Loss; Chi-square Automatic Interaction Detection

1. Introduction

Modern civilizations are greatly relying on electrical energy for normal living. Energy consumption is steadily increasing for the emerging markets due to the increase in economic development and population (Zhang & Cheng, 2009). However, large parts of the produced and distributed electricity are not paid and therefore, do not add to the profit margin of the Power Distribution Companies (PDCs)(Navani et al., 2012).Operational losses occur in all three major power system aspects (generation, transmission and distribution) (Saeed, 2020). Unlike the generation side, where the losses are theoretically specified, the power system's losses on the Transmission and Distribution (T&D) side can not be precisely calculated (Zheng & Otuoze, 2019). This is because, in addition to the technical ones, the T&D means some non-technical energy losses. Technical losses occur as a result of the dissipation of power in various elements of the power system such as transmission lines, transformers, electrical appliances, switches and many other components of power systems. These losses can be calculated by measuring just two parameters; total grid load and total energy billed (Liu & Hu, 2015) Non-technical losses, on the other side, account for billing errors, poor quality of infrastructure, malfunctioning of facilities, supply without the metering device and illicit business practices such as fraud, corruption and organized crime (Jamil & Ahmad, 2018). The non - technical losses are very difficult to predict compared to the technical losses and therefore, their minimization is one of the major concerns for any Power Distribution Company (PDC). The issue is so severe that nearly half of the total electrical power generated is turned into NTLs in most extreme cases, resulting in annual losses of billions of dollars (Yip, 2018). On average, the power companies suffer a loss of $25 billion per annum, owing to energy theft (Micheli, 2019). The problem, as mentioned above, is present in all countries; however, in underdeveloped countries, the effects are more harmful than in developed ones. For instance, PDCs in Pakistan, the T&D losses for 2017-2018 were reported as 17.5 per cent, which is doing tremendous harm to the country's weak economy and is far higher than other Asian countries, as T&D losses were recorded as 8 per cent and 3.6 per cent respectively in China and Korea (Saeed, 2019). It is estimated that 33 per cent of the country's electrical system's T&D losses account for NTL losses (Kessides, 2013). Besides, the PDCs used the findings obtained from the random assessment of consumer billing profiles to perform a study of metering systems for the detection of the NTLs. Since the power usage pattern is not taken into consideration during the identification process, the performance rate of these random inspections is exceedingly low. The arbitrary existence of the above process makes it ineffective and inefficient form of theft detection as only random bills are chosen for checking, while many are left unobserved. Another major drawback of this method is that it is very costly and time-consuming (Guerrero, 2018). This exists in developing countries, since, the

*Corresponding author Email address: rainaveed77@gmail.com

distribution feeders serve a large number of consumers and are usually quite large. The smart meters with dedicated communication links have recently emerged as one of the most viable solutions for efficient detection of NTLs. Their deployment and operating costs, however, need billions of dollars, rendering it an unfeasible option for many developing countries, including Pakistan (Depuru, 2011). Despite being one of the country's major challenges to the economy, there is no significant research on detection of NTL in Pakistan's PDCs, thus becoming the core justification of the current research work.

2. Literature Review

There are several methods used in literature for NTL detection. These methods can be broadly classified into three major categories: Artificial Intelligence (AI) based, state-based and game theory-based (Singh , 2018) Jiang *et al.*, 2014). The AI-based methods make use of machine learning techniques to assess the consumers' load profiles patterns through classification and clustering techniques in order to detect fraudulent consumers. The operation is based on the fact that the consumption patterns for the fraudulent consumers is irregular as compared to that of the genuine consumers and hence can be easily identified. In classification methods, the pre-labelled dataset is used to train the classifier while the clustering method can work with an unlabeled dataset (Zheng et al., 2019).. The second category of NTL detection methods is a state-based method. The State-based methods require measurement of a few additional parameters like voltage, current and power in order to accurately detect the electricity thefts in the distribution network (Singh, 2017) Although these methods can provide a high detection accuracy, yet the need for precise network topology and additional energy meters makes it an unfeasible option in many cases (Singh, 2018)). In-game theory, a game between the power utilities and the fraudulent consumers is considered based on the game equilibrium, which can be directly obtained from the genuine and fraudulent users' energy consumption patterns (Cardenas, 2012) The game theory-based methods rely on strong theoretical analysis and assumptions and are very complex in their basic operational mechanism; thus; such approaches are beyond the scope of this research. On the other hand, the State based methods like Advance Metering Intrusion Detection System (AMIDS) are generally used to identify the abnormality in the consumption pattern by using several sensors (McLaughlin, 2013). The major drawback associated with this method is that it requires a higher sampling rate for proper functioning, which results in revealing all types of appliances that are in use of customers along with their period of usage. This results in destroying the costumer's confidentiality and hence can't be deployed for NTL detection in a number of scenarios. The authors in (Yip, 2018) used a linear regression model to identify NTL activity by evaluating the resistance and active power in the distribution system. However, the process was time-consuming as the resistance of every single line needs to be calculated accurately along with other parameters to identify fraudster customers. An Artificial Neural Networks (ANN) method for NTL detection was explored in (Costa et al, 2013) The authors utilized the customer's consumption usage to form a database and then utilized the ANN-based scheme to classify the customers as honest or fraudsters. However, the technique was found to be ineffective due to unbalanced datasets and hence achieved less precision which ultimately resulted in a huge number of false positives. An ANN-based fuzzy logic method for NTL detection purpose was studied in reference (Muniz et al. 2009). The technique used simple rules to identify the fraudster consumers; however, the model suffered from low accuracy. In another study (Dos Angelos et al. 2011), the author used fuzzy clustering and fuzzy classification approaches for classifying the energy usage pattern of energy costumers. The main drawback of the proposed scheme was that it required a considerable amount of data like maximum consumption, average consumption data, sum of the remarks from inspection and neighborhood consumption average for every single consumer. Collecting an enormous volume of data for every customer is a very tough job, hence resulted in a huge detection delay.

Recently, smart prepaid energy meters were proposed for NTL detection in (Mohammad et al. 2013) These meters are very efficient and tempered proof; however, during direct hooking or illegal tapings, the installed sensors in meter generally used to show zero units; thus results in unmeasured supplied electricity. The author in (Jokar, et al. 2016) proposed a consumption pattern-based energy theft detector (CPDETD) Anomaly detection algorithm along with the SVM algorithm, but the main problem with the proposed scheme was that the power utility had to invest extra money for installing the transformer metering device in addition to the smart meters. An SVM based model for NTL detection was proposed in reference (Nagi *et al*. 2009). The proposed method utilized the monthly energy usage data for 25 months along with a creditworthiness rating for classifying fraudsters customers. To increase the detection rate in previous work (Nagi *et al*. 2009)., a fuzzy inference system (FIS) was utilized along with SVM by authors in (Nagi *et al*. 2010). However, as compared to the previous research work the mentioned parameter was hardly increased by 12% (from 60% to 72%), which is still very low as compared to the current research work.

To fix the deficiencies and weaknesses of preceding works, several redundant classifiers are combined for NTL identification to form Ensemble Learning Systems (ELS's). ELS's results in enhanced accuracy, better performance, robustness, and reduced uncertainties. The proposed model utilizes Bagging algorithm on Chi-square Automatic Interaction Detection (CHAID) Decision Tree (DT) for classifying the honest and fraudster customers. The Bagged CHAID-based scheme has achieved the maximum accuracy as compared to the conventional methods on Multan Electric Power Company (MEPCO) real-time dataset. The proposed classification scheme uses data from MEPCO (Pakistan) on household electricity consumption to identify the honest and fraudster customer.

3. Methodology and Detection Framework

This research work develops an AI-based efficient fraud detection model utilizing the data obtained from MEPCO. This research utilizes the Bagging algorithm to enhance the classification performance of the C5.0 DT model. The complete framework of the current study is shown in Figure 1.

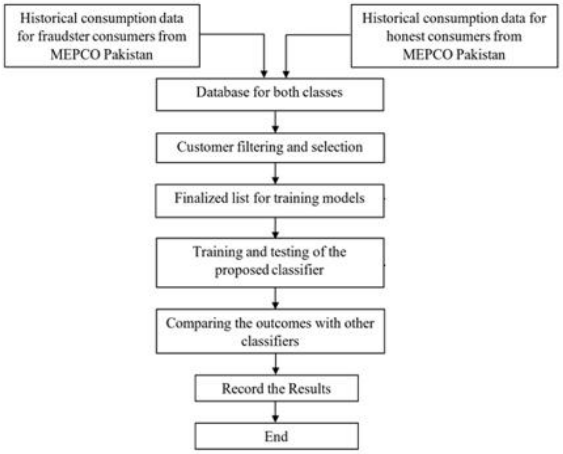

Fig. 1. Framework for the training and validation of the model.

3.1. Data collection

The collection of the data was one of the key tasks of this study. The historical energy consumption data required in this research study have been obtained from the Multan Electric Power Company (MEPCO). MEPCO is amongst the most prominent electric power utilities in Pakistan, with around 6 million customers. However, the 5.9 million customers out of 6 million belong to the domestic and commercial tariff. Therefore, this becomes one of the main reasons to focus on domestic and commercial customers in this study. The consumption data for honest consumers were obtained from the Shah Rukan-e-Alam feeder. The total number of registered customers in Shah Rukan-e- Alam feeders is 4391, with 3977 belonging to the domestic tariff, 312 belonging to the commercial tariff and remaining belongs to industrial and other categories. The Meter and Testing (M&T) laboratory officials are responsible for verifying the status of the meter. If the energy consumption of any consumer is suspicious, it is then the responsibility of the M&T laboratory to issue a thorough report after checking the site whether the consumer is involved in any illegal activity or not. The data for all the registered theft cases for the two years, i.e., 2016-2017 and 2017-2018, was gathered. The total number of registered theft cases in the Multan circle in the mentioned time was 1109. The data collected also include the meter reading, connected load, meter status, date of meter reading, date of inspection, type of theft, sanctioned load, and discrepancies etc.

3.2. Consumption pattern of fraudulent and honest consumers

The energy utilization pattern of the honest customer reveals a symmetrical pattern with a strong increase in summer as the temperature in the area under consideration approaches 50^{o}C while the same temperature in winter drop to around 20^{o}C. It is evident from figure 2 that energy consumption of dishonest customer undergoes abrupt changes compared to the honest customer, which is a clear indication of fraud.

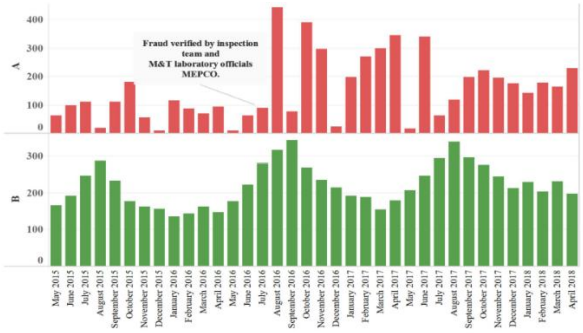

Fig. 2. Energy usage pattern of the fraudster and the honest customer.

3.3. Customer filtering and selection

The obtained data from MEPCO was initially filtered to remove the consumers with incomplete information to have appropriate training of the model. The following consumers were screened and removed during this process.

1. Customers whose kWh consumption data for the complete duration were not available, i.e., the connection got disconnected due to non-payment.
2. All those customers who were registered after the month of May-2015.
3. All those costumers whose metering apparatus became faulty during the mentioned period.
4. All the customers who were charged average units due to the unavailability of the meter reading or any other valid reason.
5. All those customers who were charged nil units (0) throughout the studied period due to non-usage of electricity.

The finalized list includes 2774 consumers with 647 fraudsters and 2117 honest consumers, after removing all outliers and the filtration process. Although many customers were excluded after the filtering process, the remaining ones were sufficient for the training and testing of the proposed model.

4. Classification Using Bagged CHAID Decision Tree Algorithm

The CHAID belongs to the classification tree category (Tso & Yau, 2007). The CHAID DT has the capability of producing accurate rules and has better performance in terms of memory. On the other hand, Ensemble learning methods train many machine learning algorithms to reach a final judgement (Saeed et all.2020). Ensemble Learning Systems (ELS) are inspired by human behavior, which believes that seeking and applying the opinion of several experts can easily tackle any problem. Based on those diverse opinions, the decision is reached. ELSs provides better performance as compared to using a single

classifier. There are different ensemble algorithms with the most common being bagging and boosting.

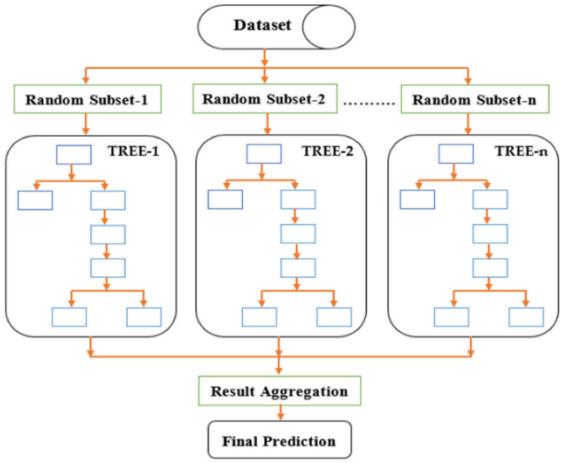

Fig. 3. Training procedure of Bagging algorithm

The bagging algorithm 's working mechanism begins by generating a random variation of the original data set into n number of data sets, as shown in Figure 1. It parallels the different classifier testing on those original dataset subsets. Finally, a model is developed based on the plurality of the individual models' votes. Then, a prediction is made based on the final model judgment. Bootstrapped subsamples are drawn within a given dataset. Each bootstrapped sample shall have a individual CHAID Decision Tree (DT). The CHAID DT outcomes are aggregated to produce the best and most accurate indicator.

5. Studied Classification Methods

The performance of the proposed classification approach was compared with a few state-of-the-art machine learning to validate the superiority of the proposed algorithm.

5.1. Support vector machines (SVM)

A Support Vector Machine (SVM) is a supervised machine learning algorithm that is used to solve classification and regression problems (Nagi et al. 2009) . Saeed *et al.* 2020) SVM usually constructs single or multiple hyperplanes for classifying non-separable groups in a higher-dimensional space. The use of a kernel trick is to distinguish non-separable groups by moving them from a lower-dimensional space into a higher-dimensional space. SVM has been used in the NTL detection problem many times (Nagi et al. 2009),(Nagi et al. 2009) , Messinis et al. 2008)

5.2. Artificial neural network

ANN is a computational model which is usually used in cases where the information provided is not sufficient to perform the classification task. ANN's has three linked layers (Ramos et all., 2011), (Saeed *et al.* 2020) The primary layer consists of input neurons, which send the data successively to the second layer and thus to the third layer. ANN has been used in the NTL detection problem many times, and some of the finest work can be found in references (Costa et al. 2013), (Muniz et al. 2016), (Ford et al. 2015).

5.3. Logistic regression

Logistic Regression (LR) has also been used for solving classification problems (Bonte & Vercauteren, 2018). Generally, LR classifiers utilize the linear combination of features value, or the response variables can be used as the argument for the sigmoid function. The LR also makes use of all data points like linear regression, but the points far from the threshold have far less impact due to logit transformation. The respective outcome of the sigmoid function lies between the value of 0 and 1. The class of the output is decided by comparing it with the middle value. If the output value is higher than one, then the object belongs to the class 0.5, similarly, if the output value is less than 0.5, then the objects to class 0.

5.4. Bayesian network

A Bayesian Network (BN) is a widely used graphical model that computes the probabilistic relation between different variables (Hosseini & Barker, 2016), (Saeed, 2020). BN is used in scenarios where we have prior knowledge of probabilities and looking to forecast new probabilities in certain conditions. BN is used for performing both backward and forward analysis. The predictive analysis comes in the category of the forward propagation and diagnostic analysis is referred to as backward propagation. One of the primary advantages that can be found in Bayesian networks is that its diagrams can be easily understood by humans; thus, it could yield accurate results.

5.5. Discriminant analysis

The goal of Discriminant Analysis (DA) is to build discriminating functions, which are just the linear combination of independent variables with the ability to differentiate perfectly between the dependent variable groups (Stamenković et al. 2020). It allows the researchers to analyze whether there are significant differences between groups concerning the predictor variables. The DA can evaluate the validity of any classification. DA is characterized by the types of categories which the dependent variable possesses. If the explanatory variable has two types, then the form utilized is DA of two classes. There have been many types of DA which have been used for NTL detection problem (Antonelo & State, 2019), (Ghori et al. 2020)

6. Results and Discussions

The confusion matrix is commonly used to evaluate the classification performance of machine learning algorithms providing True for all correctly classified data and "Wrong" for all incorrectly classified datasets (Saeed *et al.* 2020) The word True Positive (TP) used in the confusion matrix refers to all fraudster costumers properly classified as fraudsters, while the term True Negative (TN) refers to all honest customers rightly classified as honest ones. Likewise, False Positive (FP) refers to all those costumers who are honest but mistakenly classified as fraudsters and

False Negative (FN) represent fraudster customers who are wrongly classified as honest. The accuracy of a classifier can be measured by utilizing equation 1:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

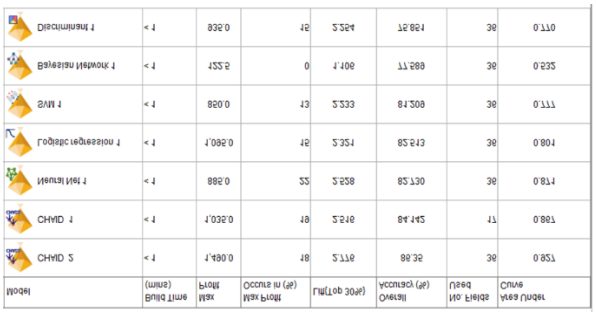

Fig. 4. Classification performance of all models.

It is clear from figure 4 that Bagging of CHAID algorithm considerably improves its classification performance. The accuracy of the bagged CHAID algorithm is 86.35 compared to simple CHAID, which is around 84.14. The Neural Network model achieves an accuracy of 82.73, which is significantly higher than LR, SVM, BN and DA methods. Figure 5 shows the simple CHAID DT for classifying honest and fraudster customers.

The Receiver Operating Characteristic (ROC) curve is another significant performance evaluator for a classifier(Andrew, 1997). The ROC curve evaluates the classifier's true performance by plotting the FPR against TPR and does not depend on the class distribution variation. The value for ROC covered area ranges from 0 to 1. A classifier with an AUC value greater than 0.5 performs better than the random forecast. The AUC value of 0.5, on the other hand, indicates that the model can not distinguish between the classes at all. The higher the AUC ranking, the greater will be the classifier 's performance. Figure 5 shows the CHAID DT until branch 3 for the classification of honest and fraudster customers. Node one is further sub-divided into three more nodes for further classification. The process continues until all the customers are classified. The nodes which do not play a significant role in the classification process are removed or pruned in the end to avoid the overfitting of the CHAID model to the training data.

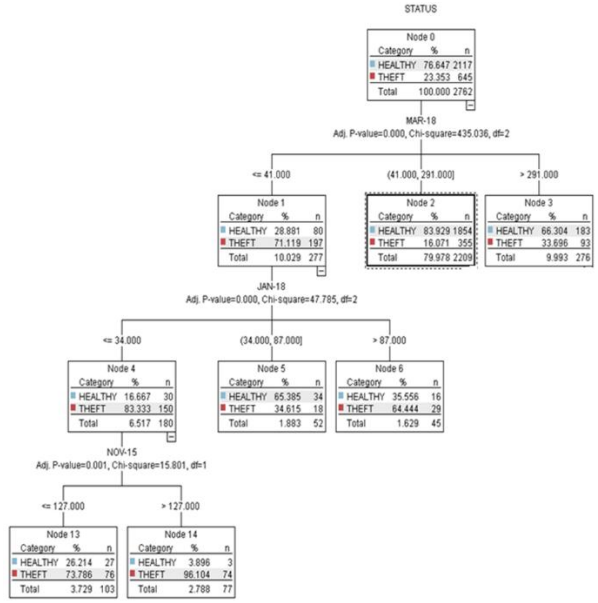

Fig. 5. CHAID DT for classification of honest and fraudster customers.

The ROC curve for all the models is given in figure 6 and figure 7, respectively. The Bagged CHAID algorithm achieves a maximum AUC of 0.927, which validates the superiority of the proposed model. The CHAID algorithm performs second-best by achieving an AUC of 0.867.

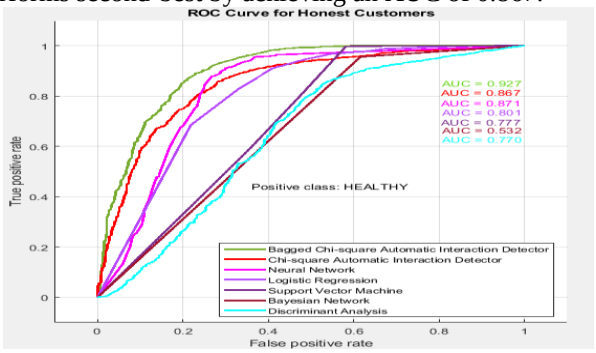

Fig. 6. ROC Curve for the negative class (Healthy consumption) customers.

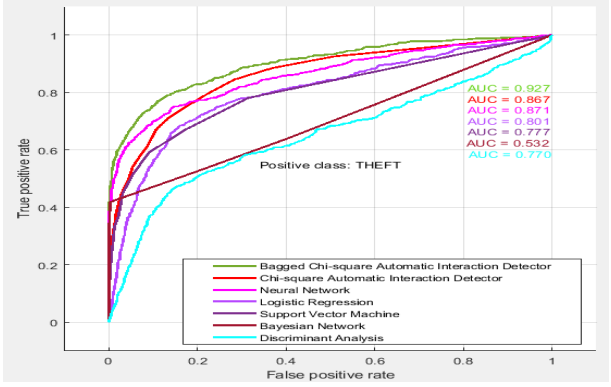

Fig. 7. ROC curve for the positive class (theft) fraudster customers

7. Conclusion

This work offered a new approach for the identification of NTL in PDCs using one of the most efficient classification algorithms called the Bagged CHAID DT algorithm. The suggested NTL detection framework has been used on

consumer historical consumption data obtained from MEPCO, which is one of Pakistan's largest power distribution firms. The performance of the proposed Bagged CHAID classifier was compared with few state-of-the-art machine learning algorithms to verify its effectiveness. The results of this study indicate that the Bagged CHAID algorithm performs much better than the above artificial intelligence techniques and achieves an accuracy of 86.35 per cent and an AUC of 0.927; which validates its superiority in performance. Furthermore, the AUC value of 0.927 indicates the Bagged CHAID algorithm 's tremendous classification capabilities. The results of this research work will help MEPCO and other power distribution firms to avoid trouble due to inefficient and expensive random inspections, which on the one side do not help to reduce NTL and on the other hand cause a huge loss of revenue on both NTL accounts and costly inspections to the PDC's.

References

Angelos, E. W. S., Saavedra, O. R., Cortés, O. A. C., & De Souza, A. N. (2011). Detection and identification of abnormalities in customer consumptions in power distribution systems. *IEEE Transactions on Power Delivery, 26(4)*, 2436-2442.

Antonelo, E. A., & State, R. (2019, October). On importance weighting for electric fraud detection with dataset shifts. *In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 3235-3242). *IEEE*.

Baskar, P., Joseph, M. A., Narayanan, N., & Loya, R. B. (2013, April). Experimental investigation of oxygen enrichment on performance of twin cylinder diesel engine with variation of injection pressure. *In 2013 International Conference on Energy Efficient Technologies for Sustainability* (pp. 682-687). *IEEE*.

Belloli, M., Melzi, S., Negrini, S., & Squicciarini, G. (2010). Numerical analysis of the dynamic response of a 5-conductor expanded bundle subjected to turbulent wind. *IEEE transactions on power delivery, 25(4)*, 3105-3112.

Bonte, C., & Vercauteren, F. (2018). Privacy-preserving logistic regression training. *BMC medical genomics, 11(4)*, 13-21.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition, 30(7)*, 1145-1159.

Cárdenas, A. A., Amin, S., Schwartz, G., Dong, R., & Sastry, S. (2012, October). A game theory model for electricity theft detection and privacy-aware control in AMI systems. *In 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (pp. 1830-1837). *IEEE*.

Costa, B. C., Alberto, B. L., Portela, A. M., Maduro, W., & Eler, E. O. (2013). Fraud detection in electric power distribution networks using an ann-based knowledge-discovery process. *International Journal of Artificial Intelligence & Applications, 4(6)*, 17.

Depuru, S. S. S. R., Wang, L., Devabhaktuni, V., & Gudi, N. (2011, March). Smart meters for power grid—Challenges, issues, advantages and status. *In 2011 IEEE/PES Power Systems Conference and Exposition* (pp. 1-7). *IEEE*.

Ford, V., Siraj, A., & Eberle, W. (2014, December). Smart grid energy fraud detection using artificial neural networks. *In 2014 IEEE symposium on computational intelligence applications in smart grid (CIASG)* (pp. 1-6). *IEEE*.

Ghori, K. M., Abbasi, R. A., Awais, M., Imran, M., Ullah, A., & Szathmary, L. (2019). Performance analysis of different types of machine learning classifiers for non-technical loss detection. *IEEE Access, 8*, 16033-16048.

Hosseini, S., & Barker, K. (2016). Modeling infrastructure resilience using Bayesian networks: A case study of inland waterway ports. *Computers & Industrial Engineering, 93*, 252-266.

Jain, M. B., Srinivas, M. B., & Jain, A. (2008, October). A novel web based expert system architecture for on-line and off-line fault diagnosis and control (FDC) of power system equipment. *In 2008 Joint International Conference on Power System Technology and IEEE Power India Conference* (pp. 1-5). *IEEE*.

Jamil, F., & Ahmad, E. (2019). Policy considerations for limiting electricity theft in the developing countries. *Energy policy, 129*, 452-458.

Jiang, R., Lu, R., Wang, Y., Luo, J., Shen, C., & Shen, X. (2014). Energy-theft detection issues for advanced metering infrastructure in smart grid. *Tsinghua Science and Technology, 19(2)*, 105-120.

Jokar, P. (2015). Detection of malicious activities against advanced metering infrastructure in smart grid (Doctoral dissertation, *University of British Columbia*).

Jokar, P., Arianpoo, N., & Leung, V. C. (2015). Electricity theft detection in AMI using customers' consumption patterns. *IEEE Transactions on Smart Grid, 7(1)*, 216-226.

Kessides, I. N. (2013). Chaos in power: Pakistan's electricity crisis. *Energy policy, 55*, 271-285.

Kim, J., Caire, G., & Molisch, A. F. (2015). Quality-aware streaming and scheduling for device-to-device video delivery. *IEEE/ACM Transactions on Networking, 24(4)*, 2319-2331.

Liu, Y., & Hu, S. (2015). Cyberthreat analysis and detection for energy theft in social networking of smart homes. *IEEE Transactions on Computational Social Systems, 2(4)*, 148-158.

McLaughlin, S., Holbert, B., Fawaz, A., Berthier, R., & Zonouz, S. (2013). A multi-sensor energy theft detection framework for advanced metering infrastructures. *IEEE Journal on Selected Areas in Communications, 31(7)*, 1319-1330.

Messinis, G. M., Rigas, A. E., & Hatziargyriou, N. D. (2019). A hybrid method for non-technical loss detection in smart distribution grids. *IEEE Transactions on Smart Grid, 10(6)*, 6080-6091.

Micheli, G., Soda, E., Vespucci, M. T., Gobbi, M., & Bertani, A. (2019). Big data analytics: an aid to detection of non-technical losses in power utilities. *Computational Management Science, 16(1)*, 329-343.

Mohammad, N., Barua, A., & Arafat, M. A. (2013, February). A smart prepaid energy metering system to control electricity theft. *In 2013 International Conference on Power, Energy and Control (ICPEC)* (pp. 562-565). *IEEE*.

Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., & Mohamad, M. (2009). Nontechnical loss detection for metered customers in power utility using support vector machines. *IEEE transactions on Power Delivery, 25(2)*, 1162-1171.

Navani, J. P., Sharma, N. K., & Sapra, S. (2012). Technical and non-technical losses in power system and its economic consequence in Indian economy. *International journal of electronics and computer science engineering, 1(2)*, 757-761.

Otuoze, A. O., Mustafa, M. W., Mohammed, O. O., Saeed, M. S., Surajudeen-Bakinde, N. T., & Salisu, S. (2019). Electricity theft detection by sources of threats for smart city planning. *IET Smart Cities, 1(2)*, 52-60.

Ramos, C. C. O., de Sousa, A. N., Papa, J. P., & Falcao, A. X. (2010). A new approach for nontechnical losses detection based on optimum-path forest. *IEEE Transactions on Power Systems, 26(1)*, 181-189.

Saeed, M. S., Mustafa, M. W., Hamadneh, N. N., Alshammari, N. A., Sheikh, U. U., Jumani, T. A., ... & Khan, I. (2020). Detection of non-technical losses in power utilities—A comprehensive systematic review. *Energies, 13(18)*, 4727.

Saeed, M. S., Mustafa, M. W., Sheikh, U. U., Jumani, T. A., & Mirjat, N. H. (2019). Ensemble bagged tree based classification for reducing non-technical losses in multan electric power company of Pakistan. *Electronics, 8(8)*, 860.

Saeed, M. S., Mustafa, M. W., Sheikh, U. U., Khidrani, A., & Mohd, M. N. H. (2020). Theft Detection In Power Utilities Using Ensemble Of Chaid Decision Tree Algorithm. *Science Proceedings Series, 2(2)*, 161-165.

Saeed, M. S., Mustafa, M., Bin, W., Sheikh, U. U., Salisu, S., & Mohammed, O. O. (2020). Fraud detection for metered costumers in power distribution companies using C5. 0 decision tree algorithm. *Journal of Computational and Theoretical Nanoscience, 17(2-3)*, 1318-1325.

Salman Saeed, M., Mustafa, M. W., Sheikh, U. U., Jumani, T. A., Khan, I., Atawneh, S., & Hamadneh, N. N. (2020). An efficient boosted C5. 0 decision-tree-based classification approach for detecting non-technical losses in power utilities. *Energies, 13(12)*, 3242.

Singh, S. K., Bose, R., & Joshi, A. (2018). Entropy-based electricity theft detection in AMI network. *IET Cyber-Physical Systems: Theory & Applications, 3(2)*, 99-105.

Singh, S. K., Bose, R., & Joshi, A. (2018, February). Energy theft detection in advanced metering infrastructure. *In 2018 IEEE 4th World Forum on Internet of Things (WF-IoT)* (pp. 529-534). *IEEE*.

Singh, S. K., Bose, R., & Joshi, A. (2018, February). Minimizing energy theft by statistical distance based theft detector in ami. *In 2018 Twenty Fourth National Conference on Communications (NCC)* (pp. 1-5). *IEEE*.

Stamenković, M., Steinwall, E., Nilsson, A. K., & Wulff, A. (2020). Fatty acids as chemotaxonomic and ecophysiological traits in green microalgae (desmids, Zygnematophyceae, Streptophyta): a discriminant analysis approach. *Phytochemistry, 170*, 112200.

Tso, G. K., & Yau, K. K. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy, 32(9)*, 1761-1768.

Yip, S. C., Tan, W. N., Tan, C., Gan, M. T., & Wong, K. (2018). An anomaly detection framework for identifying energy theft and defective meters in smart grids. *International Journal of Electrical Power & Energy Systems, 101*, 189-203.

Yip, S. C., Wong, K., Hew, W. P., Gan, M. T., Phan, R. C. W., & Tan, S. W. (2017). Detection of energy theft and defective smart meters in smart grids using linear regression. *International Journal of Electrical Power & Energy Systems, 91*, 230-240.

Zhang, X. P., & Cheng, X. M. (2009). Energy consumption, carbon emissions, and economic growth in China. *Ecological economics, 68(10)*, 2706-2712.

Zheng, K., Chen, Q., Wang, Y., Kang, C., & Xia, Q. (2018). A novel combined data-driven approach for electricity theft detection. *IEEE Transactions on Industrial Informatics, 15(3)*, 1809-1819.

This article can be cited: Saeed, M., Mustafa, M. W., Sheikh, U., Khidrani, A., & Mohd, M. N. H. (2022). Electricity theft detection in Power utilities using Bagged CHAID-Based Classification Trees. *Journal of Optimization in Industrial Engineering*, *15*(2), 67-73. Doi: 10.22094/joie.2022.1941123.1894