IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

Impact of Machine Learning on Electricity Theft

Prof. Pankaj Phadtare[1], Priti Chavan[2], Jay Kadam[3], Rugved Shinde[4], Velankanni Yadavar[5]
Department of Computer Engineering
Trinity College of Engineering, Pune, India[1,2,3,4,5]

Abstract: *The principal source of electrical power loss that has a substantial impact on both the quantity and quality of electrical power is electricity theft. However, the approaches currently in use for detecting this theft-related criminal activity are varied and complex since it is difficult to extract useful information from time-series data due to the uneven nature of the dataset. This research develops a novel approach for detecting electricity theft by combining three algorithms into a pipeline. The suggested approach first balances the dataset using the synthetic minority oversampling technique (SMOTE), then integrates kernel function and principal component analysis (KPCA) to extract features from highly dimensional time-series data, and uses support vector machines (SVM) to classify the data. Additionally, the effectiveness of the system.*

Keywords: Machine Learning

I. INTRODUCTION

Modern existence requires electricity as a basic necessity. It powers electric machinery and appliances and is used for lighting, cooling, and heating. Electricity has changed modern medical and surgical procedures, entertainment, communication, and transportation to comfort people. A variety of steps are being done to make power capable of meeting the demands as demand and usage grow daily [1]. However, the largest danger to the electrical management system continues to be power outages.

An intelligent energy system and smart grid have been created in response to the requirement to minimise power losses and maximise the usage of electricity (SG). Advanced metering infrastructure (AMI) is the foundation of SG [2]. The energy system now uses smart metres (SM), which have been replaced by AMI.

Transmission, distribution, and consumption are possible points of energy loss in the power system [1], [8]. Technical losses (TLs) and commercial losses, commonly referred to as non-technical losses (NTLs) [1, [5,] [9], are the two groups into which these losses are split. The energy loss in the conductors, distribution lines, and transmission lines is what causes TLs. NTLs can occur for a variety of reasons, including poor installation, broken metres, billing problems, tampering with the metres, hacking into smart metres, manipulating the data, direct hooking on other homes, etc. [5], [10]. The power used by customers but not billed by the utility is known as NTLs, according to utilities [11].

Energy system damage from NTLs is considerable. Customers' dishonest behaviour is to blame for the economy's troubles. The primary reason for the power utility's revenue loss is NTL [8]. It also has an impact on supply quality. NTLs are also to blame for the rise in energy prices that impacts all consumers. Because when tariffs are calculated, such losses are distributed among all consumers [12]. Additionally, it makes the electricity grid less stable and reliable [11]. The globe loses $89.3 billion annually to electricity theft, claims Northeast Group LLC [13]. NTLs are a serious problem for rich countries as well as underdeveloped countries.

II. LITERATURE SURVEY

Paper Name - Wide & Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids.

Abstract - Electricity theft can be harmful to power grid suppliers and cause economic losses. Integrating information flows with energy flows, smart grids can help to solve the problem of electricity theft owning to the availability of massive data generated from smart grids. The data analysis on the data of smart grids is helpful in detecting electricity theft because of the abnormal electricity consumption pattern of energy thieves. However, the existing methods have poor detection accuracy of electricity-theft since most of them were conducted on one dimensional (1-D) electricity

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

consumption data and failed to capture the periodicity of electricity consumption. In this paper, we originally propose a novel +electricity-theft detection method based on the Wide & Deep Convolutional Neural Networks (CNN)model to address the above concerns. In particular, the Wide & Deep CNN model consists of two components: the Wide component and the Deep CNN component. The Deep CNN component can accurately identify the non-periodicity of electricity-theft and the periodicity of normal electricity usage based on two dimensional (2-D) electricity consumption data. Meanwhile, the Wide component can capture the global features of 1-D electricity consumption data. As a result, the Wide & Deep CNN model can achieve excellent performance in electricity-theft detection.Extensive experiments based on realistic dataset show that the Wide & Deep CNN model outperforms other existing methods.

Proposed Methodology

Wide component: As seen in Fig. 5, the Wide component (shown enclosed in the red dashed box) is a fully connected layer of neural networks that learns the global knowledge from the 1-D data on electricity consumption. According to the preliminary analysis in Section III, whereas regular power usage indicates periodicity, energy thieves' consumption is less periodic or non-periodic, and customers' electricity consumption varies from time to time. A single customer's electricity usage is effectively a one-dimensional (1-D) time series of data.

Effect of : is a parameter that regulates the amount of neurons in the Wide component's fully-connected layer.

To investigate the impact of αon the prediction results, we vary the values of αfrom 10 to 120 with the step value of 1. At the same time, we fix β= 64 and γ= 10. We conduct two groups of experiments with the training ratio with 60% or 80%, respectively.

	Training ratio = 50%			Training ratio = 60%			Training ratio = 70%			Training ratio = 80%		
Methods	AUC	MAP@100	MAP@200	AUC	MAP@100	MAP@200	AUC	MAP@100	MAP@200	AUC	MAP@100	MAP@200
TSR (1)	0.5705	0.5056	0.5140	0.5698	0.5111	0.5140	0.5593	0.5365	0.5332	0.5676	0.5284	0.5363
TSR (2)	0.5903	0.5755	0.5577	0.5847	0.5955	0.5737	0.5720	0.5255	0.5277	0.5801	0.5764	0.5654
TSR (3)	0.5514	0.5362	0.5336	0.5526	0.4774	0.4989	0.5513	0.5281	0.5326	0.5498	0.5458	0.5266
TSR (4)	0.5069	0.4973	0.5039	0.5115	0.4988	0.4979	0.5135	0.5560	0.5383	0.5034	0.5676	0.5452
LR	0.6773	0.6442	0.5669	0.6944	0.6612	0.5746	0.6916	0.6666	0.5783	0.7060	0.6560	0.5781
SVM	0.7183	0.6862	0.5919	0.7317	0.7192	0.6071	0.7276	0.7244	0.6068	0.7413	0.7353	0.6195
RF	0.7317	0.9078	0.8670	0.7325	0.8869	0.8525	0.7372	0.9259	0.8864	0.7385	0.9054	0.8536
Wide	0.6751	0.8013	0.7675	0.6950	0.8096	0.7841	0.6866	0.8116	0.7768	0.6965	0.8096	0.7646
CNN	0.7636	0.9059	0.8835	0.7837	0.9394	0.9077	0.7779	0.9547	0.9154	0.7797	0.9229	0.8853
Wide & Deep CNN	**0.7760**	**0.9404**	**0.8961**	**0.7922**	**0.9555**	**0.9297**	**0.7860**	**0.9686**	**0.9327**	**0.7815**	**0.9503**	**0.9093**

Fig 1.Performance comparison with conventional schemes.

To examine the effect of on the prediction outcomes, we vary the values from 10 to 120 with the step value of 1. We correct = 64 and = 10 at the same time. Using training ratios of 60% or 80%, we run two groups of experiments.

Paper Name - Big data analytics: an aid to detection of non-technical losses in power utilities.

Abstract - The massive amount of data that the Advanced Metering Infrastructure has amassed can assist electric utilities in identifying energy theft, a problem that costs more than $25 billion per year globally. This study outlines a novel method for non-technical loss analysis in power utilities utilising a P2P computing variant that enables the detection of frauds in the lack of complete reachability of smart metres. The suggested method specifically applies a Multiple Linear Regression model to data collected by smart metres and collectors in the same neighbourhood to identify fraudulent clients. In order to validate the suggested method, the regression model has been contrasted with other data mining approaches, including SVM, neural networks, and logistic regression, using real utility data. The actual findings demonstrate that the Multiple Linear Regression Model may effectively identify energy thieves even in locations where it is difficult to obtain metres.

Proposed Methodology-

Empirical analysis- In this section, we present an empirical study carried out on a real utility database to assess the performances of the Multiple Linear Regression model with respect to the other data mining techniques. The numerical study has been carried out with reference to a database of a utility with 7000 customers divided into 48 Neighborhood Area Networks of different sizes: 38 NANs of small dimension (10 NANs with 20 users each, 11 with 50 users each, 9 with 100 users each and 8 with 150 users each) and 10 NANs with a high concentration of users (5 with 250 users each, 3 with 500 users each and 2 with 700 users each). The time horizon comprises 77 days, from 18 July 2016 to 2 October

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

2016, with 96 daily measurements (every 15 min), yielding 7392 measurements for every smart metre, as well as for the collector. All Neighborhood Area Networks presented a total reachability of smart metres (α100%).

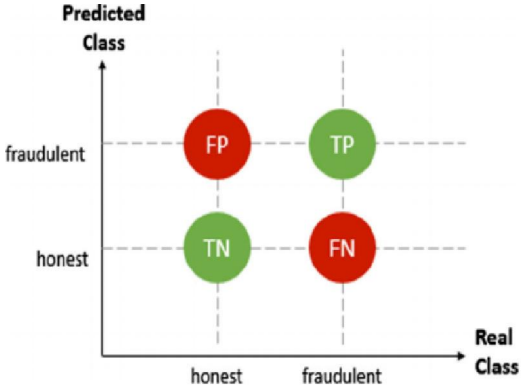

Fig: 2: Confusion matrix

Evaluation criteria -Accuracy, sensitivity, and specificity are three indices that can be computed to assess a classifier's performance. The confusion matrix, or the matrix that displays all the various findings in a classification problem, might be introduced in order to explain these indices. The confusion matrix specifically contains two rows and two columns and has the structure depicted in Fig. 2 in our two-class scenario. There are four possible outcomes:
"True Positives" (TP), "False Positives" (FP), "True Negatives" (TN), "False Negatives" (FN), "Fraudulent Users" (FU).

Paper Name -A Hybrid Model for Anomalies Detection in AMI System Combining K-means Clustering and Deep Neural Network.

Abstract - The traditional energy grid has recently been greatly impacted by the drastic digital transition, becoming an intelligent network (smart grid). This mutation is based on the steady advancement of cutting-edge technology, including smart metres and advanced metering infrastructure (AMI), both of which are essential to the growth of the smart grid. AMI systems have the ability to significantly enhance energy efficiency, improve demand management, and lower the cost of electricity. However, one of the biggest problems facing electrical companies continues to be the potential for hacking smart metres and electricity theft. In this regard, we suggest a hybrid strategy based on a combination of two reliable machine learning algorithms, K-means and Deep Neural Network, to detect anomalies related with electricity theft in the AMI system.

Proposed Methodology-

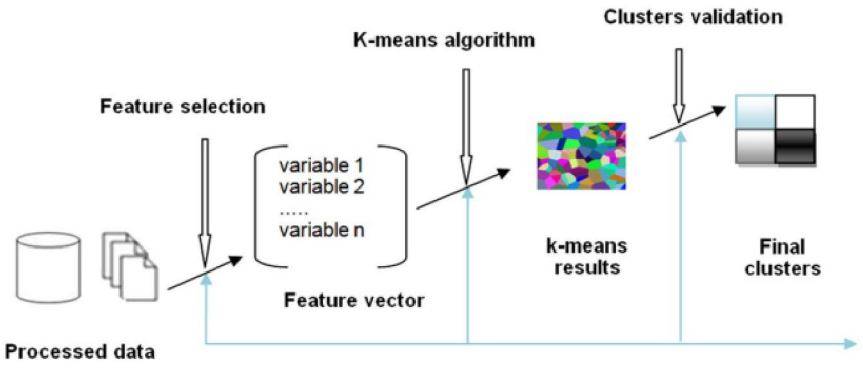

Fig 3: Three steps of the K-means process.

The k-means clustering method- A crucial stage in identifying the various typical categories of clients is the clustering technique. Based on the consumers' patterns of electricity consumption, it creates homogeneous clusters from a diverse population of users. Given its ease of implementation and ability to offer a decent approximation of the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

required segmentation, the k-means algorithm is perhaps the clustering technique that is employed the most frequently in this setting. Customer segments are divided up into clusters using the k-means algorithm. By analysing several client clusters, the K-means algorithm can thus be used to create an average pattern of electricity usage. Defining or extracting a collection of attributes that accurately captures each customer's consumption patterns is important before the clustering process can begin.

Paper Name - ZigBee Based Monitoring Theft Detection and Automatic Elec- tricity Meter Reading

Abstract - Now-a-days electricity metre reading and billing is conducted manually by door-to-door system. This system, as observed, requires a large amount of manpower and is also the time energy consumption. To overcome the limitations of this traditional system, propose a prototype module which includes advanced wireless technology called "ZIGBEE". The proposed module helps to reduce the time delay, errors and theft of electricity. The ZigBee is preferred over other wireless technolo- gies because it works in unlicensed frequency band, it does not require high speed data rate, also this device is low powered and low cost. The microcontroller based system continuously monitors the reading and theft detection that can be seen on the Liquid Crystal Display (LCD) display.

III. CONCLUSION

Each affiliation must include, at the very leastA pipeline is suggested in this paper to track down electricity theft in Singapore. SMOTE, KPCA, and SVM make up the suggested pipeline. SMOTE is used to address the issue of the imbalanced classes, KPCA is employed for feature extraction, and SVM is utilised for the classification of electricity theft. It is the most effective and straightforward method for effectively classifying fraudulent and non-fraudulent customers. In addition, a number of performance metrics, including ROC curve, precision, regression (LR), decision tree (DT), RF, CNN, and LSTM, have demonstrated the superiority of the suggested model in terms of prediction rate

REFERENCES

[1]Zheng, Z., Yang, Y., Niu, X., Dai, H.N. and Zhou, Y.,"Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids." 2017, IEEE Transactions on Industrial Informatics, 14(4), pp.1606-1615. Zheng,
[2] Micheli, G., Soda, E., Vespucci, M.T., Gobbi, M. and Bertani, A.,"Big data analytics: an aid to detection of non-technical losses in power utilities."2019, Computational Management Science, 16(1-2), pp.329- 343.
[3]Maamar, A. and Benahmed, K., A Hybrid Model for Anomalies Detec tion in AMI System Combining K-means Clustering and Deep Neural Network.",2019, vol.60, no.1, pp.15-39.
[4] Pandurang G.Kate , Jitendra R.Rana.ZigBee Based Monitoring Theft Detection and Automatic Electricity Meter Reading 10.1109/ICESA.2015.7503351
[5] Zheng, K., Chen, Q., Wang, Y., Kang, C. and Xia, Q., A novel combined data-driven approach for electricity theft detection. 2018, IEEE Transactions on Industrial Informatics, 15(3), pp.1809-1819.
[6] Razavi, R., Gharipour, A., Fleury, M. and Akpan, I.J.,A practical feature engineering framework for electricity theft detection in smart grids.2019, Applied energy, 238, pp.481-494.
[7] Fan, C., Xiao, F., Zhao, Y. and Wang, J.,Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in build ing energy data.2018, Applied energy, 211, pp.1123-1135.