*International Journal of Engineering Applied Sciences and Technology, 2022*
**Vol. 6, Issue 10, ISSN No. 2455-2143, Pages 282-287**
Published Online February 2022 in IJEAST (http://www.ijeast.com)

# AN EFFICIENT ELECTRICITY THEFT DETECTION USING XG BOOST

Disha Mhaske, Riya Satam, Snehal Londhe, Tanushree Kohad, Sonali Kadam
Department of Computer Engineering,
Bharati Vidyapeeth's College of Engineering for Women, Pune-411043

***Abstract*—Electricity theft is a significant occurrence that is happening about all nations from one side of the planet to the other. This adversely affects the nation's economy. Theft identification in the power sector is a difficult task for power distribution organizations overall, since this power theft prompts to monetary losses just as loss of electric energy. There are numerous ways by which power is stolen like controlling energy meters or tapping cables at the consumer's end, and so on. Since this robbery is going on in enormous amount, manual examination of such burglary is a hectic task. So, automatic detection of power theft is a vital need of the hour. This paper presents an electricity theft detection model using smart grid meter data dependent on Extreme Gradient Boosting (XGBoost) and OCR. Feature Selection is made used of in the model in order to pick the most relevant features from the dataset to that of our electricity theft detection model. XGBoost is remarkable as it utilizes a more regularized model formalization to have command over-fitting which results in better execution significantly quicker. OCR is employed in order to know what object led to electricity theft detection.**

***Keywords*—XGBoost, Data Pre-processing, Smart grid meter, Advanced metering infrastructure, OCR, Feature selection, Machine learning, Power theft, Electricity theft detection.**

## I. INTRODUCTION

India, the biggest democracy with an approximate population of around 1.04 billion, is on a path to quick-paced development in economy. Energy primarily electricity, is a vital contribution for speeding up economic growth.

Electricity is a key aspect of improvement and financial development. As probably the biggest consumer of power, India has been reliant upon its power sector generally. The interest for power in India has projected to develop at 5% per annum through 2030. The theft of electricity is a criminal offence and power utilities are losing billions of rupees in this account.

With the execution of the advanced metering infrastructure (AMI) in smart meters, power utilities acquired enormous amount of power utilization information at a high frequency from smart meters, which is useful for us to identify power burglary. However, each coin has different sides; the smart grid meter opens the entryway for some new power robbery attacks. These attacks in the smart grid can be sent off by different means like advanced tools and digital attacks.

The wastage of energy in power transmission and conveyance is a significant issue faced by power organizations everywhere. The energy losses are normally categorized into technical losses (TLs) and nontechnical losses (NTLs).

The technical loss is inborn to the transportation of power, which is brought about by internal activities in the power framework parts, for example, the transmission liner and transformers; the Non-Technical Loss is characterized as the contrast between total loss and TL.

With the aid of machine learning algorithms like Extreme Gradient Boosting (XGBoost) and Optical Character Recognition (OCR), we determine non-technical loss (NTL) detection in electricity theft.

Power Theft Methods

- Wires/Cables: Illegal connection to uncovered wires or underground cables, Connecting wires to other household's supply or to other power utilities
- Transformers: Illicit terminal taps of overhead lines on the low side of the transformer
- Meters: Damaging or detaching the meters, Meddling with meters and seals, Detouring/Diverging the meters
- Billing Irregularities: Done by meter readers from electric organisations
- Unpaid bills: By people, government organizations and untouchable VIPs.

## II. LITERATURESURVEY

### 1. Electricity theft detection in AMI using customers consumption patterns [1]

As one in all the predominant elements of the nontechnical losses (NTLs) in distribution networks, the power robbery causes massive damage to energy grids, which impacts energy supply quality and decreases working profits. In order to assist utility groups to solve the issues of inefficient power inspection and abnormal energy intake, a novel hybrid convolutional neural network-random forest (CNN-RF) version for automated power robbery detection is provided in this paper. In this version, a convolutional neural network (CNN) first of all is designed to examine the functions among specific hours of the day and specific days from large and ranging smart meter records via the operations of convolution

and downsampling. In addition, a dropout layer is introduced to retard the chance of overfitting, and the back propagation algorithm is carried out to update network parameters in the training phase. And then, the random forest (RF) is trained primarily based on the received features to discover whether the consumer steals power. To construct the RF in the hybrid version, the grid search algorithm is followed to decide best parameters. Finally, experiments are carried out primarily based on actual electricity intake records, and the outcomes display that the proposed detection model outperforms other techniques in terms of accuracy and efficiency.

2. **Large-scale detection of non-technical losses in unbalanced data sets [2]**

Non-technical losses (NTL) which includes power robbery cause widespread damage to our economies, as in a few nations they'll vary as much as 40% of the overall power distributed. Detecting NTLs needs expensive onsite inspections. Accurate prediction of NTLs for clients by the usage of machine learning is consequently crucial. To date, associated studies largely ignore that the 2 classes of regular & non-regular customers are exceptionally unbalanced, that NTL proportions can also additionally change and in most cases consider small data sets, often no longer permitting to deploy the outcomes in production. In this paper, we present a complete technique to evaluate 3 NTL detection models for distinct NTL proportions in huge real world data sets of 100Ks of clients: Boolean rules, fuzzy logic and Support Vector Machine. This work has led to appreciable outcomes which are about to be deployed in a leading industry solution. We trust that the considerations and observations made in this contribution are essential for future smart meter studies so as to record their effectiveness on imbalanced and huge real world data sets.

3. **XGBoost Tree boosting is a notably powerful and extensively used machine learning method. [3]**

In this paper, we describe a scalable end-to-end tree boosting system called XGBoost, which is used broadly through data scientists to obtain modern outcomes on many machine learning challenges. We suggest a unique sparsity aware algorithm for sparse records and weighted quantile sketch for approximate tree learning. More importantly, we offer insights on cache access patterns, data compression and sharding to construct a scalable tree boosting system. By combining these insights, XGBoost scales beyond billions of examples by the usage of far fewer sources than existing systems.

4. **Electricity theft: overview, issues, prevention and a smart meter based approach to control theft [4]**

Non-technical loss (NTL) within side the transmission of electrical energy is a high hassle in developing countries and it is been very difficult for the utility companies to find out and fight the people responsible for theft. Electricity theft forms a prime chunk of NTL. These losses have an impact on quality of supply, boom load on the generating station, and feature an impact on tariff imposed on genuine customers. This paper tells information about the aspects that influence the customers to steal electricity. In view of these ill effects, numerous techniques for detection and estimation of the theft are discussed. This paper proposes an architectural format of smart meter, external control station, harmonic generator, and filter circuit. Motivation of this model is to detect illegal customers, and maintain and effectively use energy. Also, the smart meters are designed to provide records of several parameters related to instantaneous electricity consumption. NTL within side the distribution feeder is computed with the useful resource of the usage of external control station from the sending end information of the distribution feeder. If a considerable amount of NTL is detected, harmonic generator is operated at that feeder for introducing more harmonic issue for destroying appliances of the illegal customers. For illustration, cost-gain assessment for implementation of the proposed system in India is presented.

5. **The Challenge of Non-Technical Loss Detection Using Artificial Intelligence: A Survey [5]**

Non-technical losses (NTL) detection includes electricity theft, faulty meters or billing faults which is point of attention for researchers in electrical engineering and computer science. NTLs cause major harm to the economy, as in some of the countries it may range up to 40 percentage of the total electricity distributed. The principal research direction is employing artificial intelligence to predict electricity theft. This paper first provides information of how non-technical losses are defined and their influence on economies, which include loss of revenue and profit of electricity providers and decline the stability and reliability of electrical power grids. It then surveys the state-of-the-art research efforts in an up- to-date and analysis of algorithms, features and data sets used. It finally identifies and suggests how this could be addressed in the future.

6. **A multi-sensor energy theft detection framework for advanced metering infrastructures [6]**

The advanced metering infrastructure (AMI) is a key component of the smart grid, substituting traditional analog devices with computerized smart meters. Smart meters have not only permitted for efficient management of many end-users, but also have made AMI an attractive target for remote exploits. Smart meters have multiple sensors and data sources that can detect energy theft. In this paper, we present an AMI intrusion detection system that uses information fusion that combines the sensors and consumption data from a smart meter to detect energy theft accurately. In AMIDS, meter audit logs of physical and cyber events are combined with consumption data to more accurately model and detect theft-related behaviour. Our experimental results on normal and

abnormal load profiles display that AMIDS can recognize energy theft efforts with high correctness. Also, AMIDS correctly identified legitimate load profile changes that more elementary analyses classified as harmful.

7. **Improving knowledge-based systems with statistical techniques, text mining, and neural networks for nontechnical loss detection [7]**

The purpose of this project is the detection of NonTechnical Losses (NTLs) in power utilities. Nontechnical losses (NTL) detection include electricity theft, faulty meters or billing faults on client side. Initially, research was made to study the application of techniques of data mining and neural networks. After several researches, the studies are extended to other research fields: expert systems, text mining, pattern recognition, statistical techniques, etc. These techniques have provided an computerized system for detection of NTLs on company databases. This project is gone through testing phase and it is applied in real time cases in company databases.

8. **Fraud Identification in Electricity Company Customers Using Decision Trees [8]**

Power consumer fraudulence is a difficulty faced by all power services universal. Finding well-organized measurements for spotting fake electricity consumption has been an active investigate area in current years. This idea presents a new tactic on the way to Non-Technical Loss (NTL) detection in power utilities using a combination of data mining and artificial intelligence (AI) based techniques, precisely: Support Vector Machine (SVM) and Fuzzy Inference System (FIS). The main incentive of this research is to support Tenaga National Berhad (TNB) in peninsular Malaysia to decrease its NTLs in the distribution area. The intellectual system established in this research study preselects chary clients to be examined on the spot by the TNBD SEAL (Strike Enforcement Against Losses) panels for recognition of the fake activities. This methodology also provides a technique of data mining, which includes feature selection and extraction from ancient customer utilisation data. The Support Vector Classification (SVC) technique applied in this thesis practices consumer load profile data in order to reveal irregular behaviour that is acknowledged to be extremely allied with the NTL activities. The FIS is working as an information post processing system, which uses intelligence of human proficiency joint with the outcomes from the SVC, in order to pick out potential fraud disbelieves for on the spot review. The recommended SVC and FIS model is accomplished by using TNB Distribution's historical kwh (kilo watt hour) utilization data for the Kuala Lumpur (KL) Barat station, which is chronicled with one of the peak rates of fraud happenings in the state of Selangor in Malaysia. Model testing and confirmation is executed by using customer info from three towns in the state of Kelantan in Malaysia. The response from TNBD for onsite inspection tells that the fraud detection organization developed is more effective as paralleled to the present actions taken by them. With the execution of this new fraud detection system, TNBD's usual hit ratio for onsite customer review will become 40%, which increases their current inspection hit ratio 35-37% from a mere 3-5%.

9. **High performance computing used for recognition of power theft**

Spread and circulation of electricity include technical as well as Non-Technical Losses (NTLs). Illegal utilization of electricity establishes a main quantity of the NTL at distribution feeder level. Seeing the seriousness and shocking results of the problem, illegal use of electricity has to be noticed promptly in real-time. Lastly, this paper examines the option and the part of High Performance Computing (HPC) algorithms in finding of illegitimate consumers. Also, this paper strategies and gears an encrypting technique to shorten and change customer energy utilization data for faster examination without negotiating the value or individuality of the data. This paper parallelizes general customer grouping process. The parallelized algorithms have led to in noticeable outcomes as shown in the results sector of the paper.

10. **Fraud detection in electric power distribution networks using an ann based knowledge-discovery process**

Currently the power-driven services have to carry complications with the non-technical losses produced by fakes and thefts dedicated by a few of their consumers. In order to lessen this, some practices have been formed to execute the detection of customers that might be tricksters. In these circumstances, the usage of classification techniques can increase the hit rate of the fraud detection and rise the financial income. This paper suggests the usage of the knowledge- innovation in databases method built on artificial neural networks put on to the classifying procedure of consumers to be examined. An research made in a Brazilian electric power circulation company designated an development of over 50% of the planned approach if equated to the earlier methods consumed by that company.

11. **Power utility nontechnical misfortune examination with outrageous learning machine strategy.**

This paper presents another way to deal with nontechnical loss (NTL) examination for utilities utilizing the current computational method extreme learning machine (ELM). Nontechnical losses address a huge extent of power misfortunes in both developing and developed nations. The ELM-based approach introduced here utilizes client load-profile data to uncover unusual behaviour which is known to be profoundly connected with NTL activities. This approach gives a strategy for data mining for this reason, and it includes extricating patterns of client behaviour from historical kWh consumption information. The outcomes relents classification

*International Journal of Engineering Applied Sciences and Technology, 2022*
**Vol. 6, Issue 10, ISSN No. 2455-2143, Pages 282-287**
Published Online February 2022 in IJEAST (http://www.ijeast.com)

classes that are utilized to uncover whether any critical behaviour that arises is because of anomalies in consumption. In this paper, we have made use of both ELM and online sequentialELM (OS-ELM) algorithms to accomplish a superior grouping performance and to build accuracy of results. A correlation of this methodology with other classification techniques, like the support vector machine (SVM) calculation, is likewise embraced and the ELM execution and precision in NTL examination is demonstrated to be predominant.

**Inference of Literature Survey:**

The purpose of this literature surveys is to understand how a problem can be solved with the help of various methods. Lastly, we conclude that the literature survey helped us to know the various methods to detect electricity theft. Here, in this proposed model we will be using XGBoost and Optical Character Recognition (OCR) algorithm in order to detect electricity theft since it provides us with high accuracy.

## III. METHODOLOGY

*A.* Existing System Disadvantages:

1) Theft happens in large amount of data hence manual inspection is hectic or even not possible.
2) Possibility of theft not detected or bribing of the employees sent by electric utilities.
3) Ineffective and wasteful present techniques for detecting and preventing power robbery cause an income misfortune alongside harm to public and personal property.
4) Enormous measure of power shortage is caused because of power misuse and stealing.
5) The drawback of this system is that real time monitoring of the loads is not possible and location of theft is not determined.

*B.* Proposed System

India positions top in losing more money to power theft than any other country around the world. The province of Maharashtra loses $2.8 billion every year.

In this proposed framework, we use dataset which consists of power consumption of a smart grid (SG) meter. Making use of this dataset, we carry out preprocessing and feature selection on this particular dataset. At this point, we have enormous number of features in dataset; hence feature selection is a vital part in our Machine Learning model. As we use feature selection, it offers us with most significant features and this in turn makes our theft detection model more accurate and provides us with better results.

Henceforth, we take the help of Extreme Gradient Boosting (XGBoost) and Optical Character Recognition (OCR) algorithm to train our model for non-technical loss (NTL) detection.

*C.* System Architecture

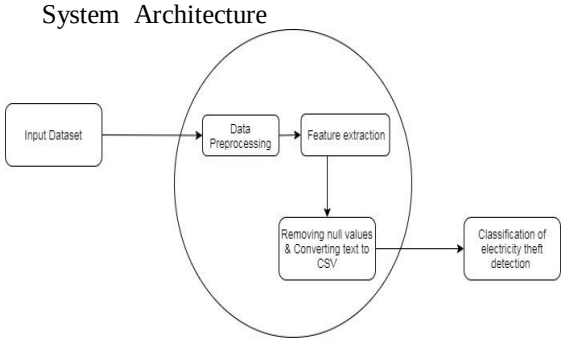


Fig. 1 System Architecture

Explanation:

1) Initially, we will take smart grid meter data as input for electricity theft detection model.
2) The data may have missing values and outliers. So, this can be removed by data pre-processing. Data pre-processing is the process of transforming raw data into an understandable format. It is also an important step in data mining as we cannot work with raw data. Pre-processing of data is mainly to check the data quality. The major tasks involved in data pre-processing:
   - Data cleaning
   - Data integration
   - Data reduction
   - Data transformation
3) The next step is feature selection. Feature selection is basically the part where in we reduce the number of input variables to those that seem most useful in a way of predicting our target Variable in our electricity theft detection model. Here we select best features such as current, voltage and electricity consumption of previous months by using feature selection method. Feature selection is useful in two ways: it helps in decreasing the computational cost and it additionally helps to improve the performance of the model.
4) Now that pre-processing is complete, we will train our model for which we use ensemble learning method (XGBoost) and OCR. Ensemble learning is a technique in which multiple models are re-created from same dataset and then it is combined to produce improved results. Extreme Gradient Boosting is an ensemble method based on decision tree alongside it employs gradient boosting framework for betterment of speed and improved performance. Optical character recognition empowers the computers to inspect printed or handwritten reports automatically and get ready text information into editable configurations for computers to productively deal with them. The basic use of OCR is to detect presence of object which led to electricity theft detection.

5) Now we have train our model. We can use this model after evaluation to detect the fraudulent or misuse of electricity.

## IV. CONCLUSION

The project of ours is pointed towards reducing the massive power and revenue misfortune that happens because of electricity theft or misuse by the consumers. By this design, it can be deduced that power burglary can be adequately kept under control by discovering where the power theft happens and enlightening the authorities. This will lead individuals to utilize the power productively and assist electrical utilities with limiting the theft.

To start with, smart grid meter data is given as input. 'State grid Corporation of China' dataset will be used as our dataset for the electricity theft detection model. The data may have missing values, punctuation marks, noise, etc. So, this can be removed by data pre-processing. Then, we have done feature selection to select important features from the dataset. Next, we have applied machine learning algorithms to train our model. This power theft is detected using XGBoost machine learning algorithm along with Optical Character Recognition. This algorithm is preferred as it gives quite accurate results in less time.

Consequently, by the above mentioned design we can effectively and viably address the issues connected with electricity theft. By this model, the wrongdoing of thieving power may be brought to an end and inevitably, a new bloom may be expected in the economy of our country and also furthermore there will be less undersupply for power usage.

## AUTHOR CONTRIBUTIONS

DishaMhaske and TanushreeKohad proposed and implemented the main idea and collected the data. SnehalLondhe did the literature survey and gained the knowledge of previous work done in this field. Riya Satam performed statistical calculation and prepared the mathematical model for our system. All authors have read and agreed to the final version of this paper.

## V. REFERENCES

[1] Muhammad Ismail1 , Mostafa Shahin ,Mostafa F. Shaaban , ErchinSerpedin, and Khalid qaraqe ”Efficient detection of electricity theft cyber attacks in AMI networks” IEEE Wireless Communications and Networking,2018

[2] Pandurang G. Kate,Jitendra R. Rana,” ZIGBEE based monitoring theft detection and automatic electricity meter reading”, IEEE International Conference on Energy Systems and Applications,2015

[3] Mahmoud Nabil , Muhammad Ismail , Mohamed Mahmoud , Mostafa Shahin , Khalid Qaraqe , and ErchinSerpedin,” Deep Recurrent Electricity Theft Detection in AMI Networks with Random Tuning of Hyper-parameters “,2018.

[4] R. E. Ogu and G. A. Chukwudebe,” Development of a Cost-Effective Electricity Theft Detection and Prevention System based on IoT Technology”, IEEE 3rd International Conference on Electro- Technology for National Development (NIGERCON),2017

[5] [5]Nikhil V. Patil,Rohan S. Kanase,Dnyaneshwar R. Bondar,P. D. Bamane,” Intelligent energy meter with advanced billing system and electricity theft detection“,2017

[6] Muhammad Saad,Muhammad Faraz Tariq,AmnaNawaz,Muhammad Yasir Jamal,” Theft detection based GSM prepaid electricity system”, IEEE International Conference on Control Science and Systems Engineerin,2017

[7] Daniel NikolaevNikovski, Zhenhua Wang, Alan Esenther, Hongbo Sun, Keisuke Sugiura, Toru Muso,aoru Tsuru, “Smart Meter Data Analysis for Power Theft Detection”,2013

[8] Sandeep Kumar Singh, Ranjan Bose, Anupam Joshi,” PCA based electricity theft detection in advanced metering infrastructure”, 7th International Conference on Power Systems,2017

[9] Muhammad Tariq,H. Vincent Poor” Real Time Electricity Theft Detection in Microgrids through Wireless Sensor Networks”, IEEE SENSORS,2016

[10] Jaime Yeckle,Bo Tang “Detection of Electricity Theft in Customer Consumption using Outlier Detection Algorithms”, IEEE 1st International Conference on Data Intelligence and Security,2018

[11] P. Jokar, N. Arianpoo, V.C.M. Leung, "Electricity theft detection in AMI using customers‟ consumption patterns", IEEE Trans. Smart Grid, vol. 7, no. 1, pp. 216-226,2016.

[12] M. Buzau, J. Aguilera, P. Romero, and A. Expósito, “Detection of Non-Technical Losses Using Smart Meter Data and Supervised Learning,” IEEE Trans. Smart Grid, Feb. 2018. [DOI: 10.1109/TSG.2018.2807925]

[13] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 2016, San Francisco, CA, USA [DOI 10.1145/2939672.2939785].

[14] Dorogush, V. Ershov, and A. Gulin, “CatBoost: gradient boosting with categorical features support,” Workshop on ML Systems at Neural Information Processing Systems (NIPS), 2017. 5. G. Ke, Qi Meng, T.Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." 31st Conference on Neural Information Processing Systems (NIPS), 2017

[15] Patrick Glauner et al. “Large-scale detection of nontechnical losses in imbalanced data sets”. In:


Innovative Smart Grid Technologies Conference (ISGT), 2016 IEEE Power & Energy Society. IEEE. 2016, pp. 1–5

[16] S.S.S.R. Depuru, L. Wang, and V. Devabhaktuni, "Electricity theft: overview,issues, prevention and a smart meter based approach to control theft," Energy Policy, vol. 39, pp. 1007–1015, Feb. 2011.

[17] Patrick Glauner et al. "The Challenge of Non-Technical Loss Detection Using Artificial Intelligence: A Survey". In: International Journal of Computational Intelligence Systems 10.1 (2017),pp. 760–775

[18] S. McLaughlin , B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz , "A multi-sensor energy theft detection framework for advanced metering infrastructures," IEEE J. Sel. Areas Commun., vol. 31, no. 7, pp. 1319–1330, Jul.2013

[19] Shuan Li, Yinghua Han, Xu Yao, Song Yingchen, Jinkuan Wang, Qiang Zhao, "Electricity Theft Detection in Power Grids with Deep Learning and Random Forests", Journal of Electrical and Computer Engineering, vol. 2019, Article ID 4136874, 2019.