

Electricity Theft Detection in Power Consumption using Superiority of Machine Learning Algorithm

Prof. Smita Khot¹, Sudha Dhore², Sejal Thorat³, Priyanka Musmade⁴, Varsharani Sargar⁵

^{1,2,3,4,5} Department of Computer Engineering

^{1,2,3,4,5} Dr. D.Y. Patil Institute of Technology, Pune, Maharashtra, India

Abstract - Electricity theft is the act of stealing electricity by deceptive techniques. Electricity theft represents a large chunk of NTL (Non- Technical Loss). These are the losses caused by misidentified, mis-allocated, or incorrect energy flows. Electricity theft is a major issue in India, as it is in most of the developing countries. Although the theft can be detected using machine learning techniques. In this study, the system is proposed that reads smart meter data (input) using OCR (Optimized Character Recognition). OCR is a technique for recognizing text characters in digital images that have been printed or handwritten. The smart meter data image which consists of electricity usage units is converted into machine-readable text by OCR. Furthermore, the SARIMAX (Seasonal Auto Regressive Integrated Moving Average with exogenous factors) algorithm is utilized to monitor customers electricity consumption and detect electricity theft. If theft is identified, an alert message and details of the theft are sent to an electrical board worker. The worker then manually verifies/checks and updates the status. If no theft is discovered, a bill is generated.

Key Words: Electricity Theft, Machine Learning, Optical Character Recognition, SARIMAX

1. INTRODUCTION

Electricity theft costs crores of rupees in India. As per the Section 135 of the Electricity Act 2003, electricity theft happens when a person taps electricity supply lines, tampers with electricity meters or transformers or uses a device that interferes with reading or damages equipment such as electric meters or uses electricity in an unauthorized manner. If electricity theft is discovered in this circumstance, the main power supply is immediately disconnected. The individual is then sentenced to pay three times the amount of money stolen. If the same person is caught again, he is restricted from receiving energy for three months, or even a year. Electricity theft is a big problem in India, costing billions of dollars each year, and is consequently the most pressing economic issue that people must address. As a result, it is critical for the government to highlight this issue openly and guide residents through it. There are several sorts of energy power theft, each linked to a different component of electrical equipment, such as meters, cables, and overhead lines. The focus of this discussion is on meters specifically. In the proposed system, a machine learning algorithm is used to identify whether theft occurred. This is done if there is some uncertainty in the power consumption of electricity at users end. If there is significant discrepancy then theft is detected.

2. LITERATURE REVIEW

Electricity theft is the illegal act of stealing electricity from the power utility and paying less for electricity than the real consumption. Electricity can be used in a variety of ways. Electricity suffers losses throughout generation, transmission, and distribution. Technical Losses (TLs) and Non-technical Losses (NTLs) are the two types of losses. Non-Technical Losses (NTLs) are caused by transmission line energy dissipation and magnetic losses. Transformers Electricity theft is the leading cause of NTLs [1]. The crime can occur in following ways faulty or broken electricity meters, tampering with meters in order to record lesser consumptions, bypassing meters by connecting directly to power sources, bribing meter readers to arrange phony meter readings. An unmetered supply, errors in meter readings, billing, and data processing caused by human or technical error. Formerly, consumption data is unlabeled; however, there should be a good mechanism for labelling the data. Besides the fair customers outnumber the fraudulent consumers, which has a negative impact on the classification algorithm's performance. To solve this problem an updated ETD model was proposed, in which an optimized classifier Differential Evaluation Random Under Sampling Boosting (DE-RUSBoost) is utilized. The proposed classifier DE-RUSBoost is optimized utilizing the Differential Evaluation metaheuristic optimization approach (DE) [1]. A lack of integrated infrastructure for coordinating electrical load data analysis operations is addressed by using a TCN (temporal convolutional network) and handling imbalanced data [2]. Another proposed model is to employ a deep neural network-based classification strategy to detect theft using comprehensive information in the temporal and frequency domains. Further improves the detection performance of electricity theft by optimizing hyperparameters with a Bayesian optimizer and utilizing an adaptive moment estimation optimizer to run trials with different values of critical parameters to identify the ideal settings [3].

Furthermore, to construct relevant strategies for the power company's practical energy theft detection requirements, an electricity theft detection approach based on stacked autoencoder (SAE) and the under sampling and re-sampling based random forest (UaRe-RF) algorithm was proposed. Reasonable tactics were developed for specific detection goals based on the distribution of suspicion levels, and key electricity theft users are given a greater detection priority, reducing workload [4]. Owing to the theft behaviors among customers, collaborative energy theft, has become particularly common. A group of people who steal electricity in constant proportions was considered. When the theft exists in the same area conventional correlation-sorting-based methods have trouble handling these electricity thieves. To overcome such limitation, the mathematical model of non-technical loss (NTL) and the load

data of fixed ratio electricity thieves (FRETs) is established. Trends to locate FRETs are observed and analyzed. A correlation analysis-based detection method is proposed based on this trend. It measures the correlation between the NTL and user data. A series of tests via comparisons with other state-of-the-art methods regarding accuracy, stability, and scalability is conducted. The results indicate that the proposed method has superior and stable performance in detecting FRETs, particularly when it comes to multiple FRETs in one area [5].

Moreover, the performance of Random Under Sampling (RUSBoost) is enhanced by DE and Jaya optimization algorithms. The proposed classifiers DE-RUSBoost and Jaya-RUSBoost achieve better AUC values of 0.83 and 0.90 as compared to 0.82 and 0.78 for RUSBoost and WADCNN benchmark scheme. Both Classification accuracy and execution time required for classification can be increased by Parameter tuning [6]. Advanced meters, such as AMI Smart meters, are able to measure consumer energy consumption at precise intervals of time, such as every five, fifteen, or twenty minutes. Through the incorporation of Information and Communication Technologies (ICTs), an intelligent electricity grid optimizes the generation, distribution, and consumption of electricity. In order to study time series of power consumption and on the development of a straightforward forecast model using a comparable day method, this research presents a thorough examination of 5-minutes 100 anonymized commercial building meter data sets. Energy companies can use big data analytics to make accurate forecasts, detect anomalies, support intelligent grid control, and improve energy and service management. In this more predictive analysis and enhanced the forecast model with the comparison of ARIMA, exponential smoothing, etc. is done which resulted in a lower prediction error. ARIMA produced more accurate forecast on aggregated consumption than exponential smoothing [7]. However, smart meters also introduce numerous new methods of electricity theft. Malicious users can hack into smart meters by using advanced instruments or cyberattack techniques. These attacks cause huge amount of financial loss every year. In another paper an electricity theft detection scheme based on the Extreme Gradient Boosting (XGBoost) for advanced metering infrastructure (AMI) is proposed. In this XGBoost based electricity theft detector, the gradient boosting builds an ensemble of Decision Trees, continuously adding new trees to correct the errors made by existing model [8].

Adding to the existing solutions, AMI is a major component of smart grid that offers a massive amount of data, making technologies like data mining more suitable for detecting electricity theft. Many models in the realm of electricity theft are prone to under-fitting due to the unbalanced dataset. An outlier detection strategy based on clustering and local outlier factor (LOF) was presented to avoid this problem. Outlier candidates are customers with load characteristics that are far from the cluster centers. The LOF was then used to calculate the anomalous degrees of outlier candidates [9]. However, there are certain drawbacks to the proposed strategy. The proposed method solely examines electricity use data, which may be incomplete. Other information, such as meteorological elements (temperature), geographical factors, and some electric factors (current and voltage), should be researched in addition to meter reading data [8], [9].

3. PROPOSED SYSTEM

The proposed system is provided with smart meter image as an input. Using optical character recognition, the smart meter readings are retrieved from the image. Machine predicts the outcome and classifies if there is theft or not, this is aided by SARIMAX (Seasonal Auto Regressive Integrated Moving Average with exogenous factors) algorithm. If theft is discovered, a message of theft detection is issued. Therefore, in this study, the OCR (Optical Character Recognition) and SARIMAX algorithms are used to identify electricity theft.

Web Interface:

Admin can log in with their user credentials. The primary graphical user interface (GUI) for electricity theft detection will be displayed to the admin after a successful login.

Primary Interface:

The smart meter image can be entered into the system by the admin via this interface. The meter readings from the image are transformed into a format that is machine readable. Machine Learning model calculates the average units consumed and predicted units. If theft is detected then theft detection message is displayed on the screen and notification alerting the theft is sent accordingly. Otherwise, bill is generated for that user.

OCR:

OCR is employed to extract essential data from the image. In this model, optical character recognition is carried out using Google Application Programming Interface (API). The readings from the smart meter image are transformed into a format that can be interpreted by machines.

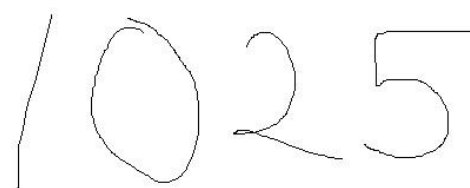


Fig -1: Meter Reading



Fig -2: OCR Extracted Digits

Average Units Consumed:

The actual readings retrieved from the image and the previous electricity use of a specific user are used to calculate the average units utilized.

Predictions and Classification:

The predicted value is determined and theft is identified using SARIMAX (Seasonal Autoregressive Integrated Moving Average Exogenous model) algorithm. Since the amount of electricity consumed varies depending on the season, SARIMAX is utilized to improve the model's efficiency. If there is a significant discrepancy between average value and anticipated value, theft has been detected.

Bill Generation:

A bill is generated if theft is not discovered. Bill is produced following a common formula:

Total kWh (kilowatt hour) (current month)/Units used = Current meter reading - Previous month's reading

Cost total = Units Used + cost

Total Bill equals total cost plus fixed fees.

Notification:

When a theft is detected, a message is sent and electrical board worker is advised to manually check for theft.

ALert ! Theft has been Detected.
[Click here](#) for details.

- Sent via FTWSMS

Fig -3: SMS Regarding Theft (Example Template)

Graph:

The Graph is generated for each user's electricity consumption, where x-axis represents months and y-axis represents units consumed.

4. ARCHITECTURE DIAGRAM

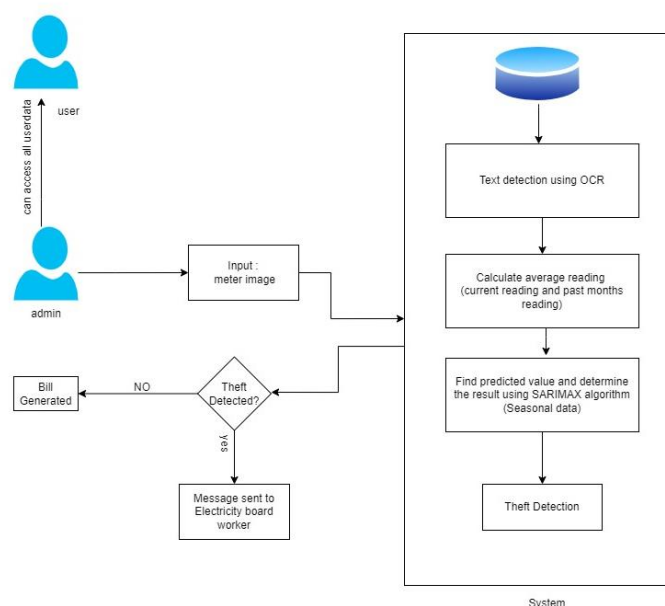


Fig -4: Architecture Diagram

5. ALGORITHM & RELATED MATHEMATICS

SARIMAX

SARIMAX stands for Seasonal Auto Regressive Integrated Moving Average with exogenous factors. It is updated version of Auto Regressive Integrated Moving Average (ARIMA) model. 'S' in SARIMAX denotes the seasonal aspect. ARIMA contains auto regressive integrated moving average while SARIMAX contains seasonal effects and exogenous factors with the auto regressive and moving average component.

In SARIMAX models parameter, there are two kinds of orders the first one is (p, d, q) and the other is the effect of seasonality or seasonal order. In seasonal order it is very essential to provide four numbers

- 1)Seasonal AR specification(p)
- 2)Seasonal Integration order(d)
- 3)Seasonal Moving Average(q)
- 4)Seasonal periodicity(s)

SARIMAX differs from Seasonal Autoregressive Integrated Moving Average (SARIMA) is that they do not use exogenous factors or variables. Exogenous variables are those variables which affect the model but are not affected by the model.

For example, in a farm's corn production crop eating pests and weather can be considered as exogenous variables. These pests and weather can reduce the crop growth but the crops cannot affect those pests.

What is Seasonality?

Seasonality occurs when certain patterns are inconsistent and which occur periodically. Certain products have a high demand in a particular period or season and very negligible demand during some days. For instance, check the monthly record of rainy products purchased over a year.



Fig -5: Seasonality Trend

The purchase of umbrella occurs more during months of June, July, August, and September every year. However, it's purchase is a lot lower in other months like January and December. To account for such a pattern, they consider the values recorded during previous rainy season or previous two rainy seasons. SARIMAX will consider data for rainy season of previous ones and the current, it will not consider other seasons like summer or winter data.

Mathematically it is represented as:

$$Y = X \sum_{n=6}^9 (y_n + y_{n-12} + y_{n-24})$$

X - exogenous factor

Y is data for a particular 'n' value

n - months (June to September)

n -12 indicates last year month

n -24 indicates last to last year month

OCR

OCR is abbreviation for Optical Character Recognition. It converts handwritten text or images into machine readable data. It gives higher accuracy of text detection. It extracts information from image under different light conditions or from blurry images.



Fig -6: OCR Concept

In above figure it is seen that OCR extracts the text part "WAITING?" and "PLEASE TURN OFF YOUR ENGINE" from the image ignoring the symbols.

6. CONCLUSION

This proposed system detects the electricity theft using OCR, SARIMAX a seasonal machine learning algorithm. Whereas, in the existing system, XGBoost algorithm is used which calculates all the yearly readings of the customer. It is not a seasonal algorithm and hence gives less accurate results. The proposed method checks the data on seasonal basis giving more precise data.

With the use of this proposed system, power companies can prevent losses by detecting electricity theft and alerting the electricity board workers via text message. As a result, the system detects electricity theft by leveraging the supremacy of machine learning algorithms. The entire process can be automated from image entry to bill preparation and payment. Furthermore, the proposed method will take more types of electrical data as the input to verify the robustness and accuracy. In future establishment of a smart meter that can reduce man power by automating the entire process by accessing the meter like on and off can be implemented, prohibiting the users from manipulating readings or stealing electricity.

7. REFERENCES

1. S. Mujeeb, N. Javaid, R. Khalid, M. Imran, and N. Naseer, "DE-RUSBoost: An Efficient Electricity Theft Detection Scheme with Additive Communication Layer," in ICC 2020 - 2020 IEEE International Conference on Communications (ICC), Jun. 2020, pp. 1–6. doi: 10.1109/ICC40277.2020.9149315.
2. I. U. Khan, N. Javeid, C. J. Taylor, K. A. A. Gamage, and X. Ma, "A Stacked Machine and Deep Learning-Based Approach for Analysing Electricity Theft in Smart Grids," IEEE Trans Smart Grid, vol. 13, no. 2, pp. 1633–1644, Mar. 2022, doi: 10.1109/TSG.2021.3134018.
3. L. J. Lepolesa, S. Achari, and L. Cheng, "Electricity Theft Detection in Smart Grids Based on Deep Neural Network," IEEE Access, vol. 10, pp. 39638–39655, 2022, doi: 10.1109/ACCESS.2022.3166146.
4. G. Lin et al., "Electricity Theft Detection Based on Stacked Autoencoder and the Undersampling and Resampling Based Random Forest Algorithm," IEEE Access, vol. 9, pp. 124044–124058, 2021, doi: 10.1109/ACCESS.2021.3110510.

5. Y. Yang et al., "A Detection Method for Group Fixed Ratio Electricity Thieves Based on Correlation Analysis of Non-Technical Loss," IEEE Access, vol. 10, pp. 5608–5619, 2022, doi: 10.1109/ACCESS.2022.3141610.
6. S. Mujeeb et al., "Electricity Theft Detection With Automatic Labeling and Enhanced RUSBoost Classification Using Differential Evolution and Jaya Algorithm," IEEE Access, vol. 9, pp. 128521–128539, 2021, doi: 10.1109/ACCESS.2021.3102643.
7. M. H. Rashid, "AMI Smart Meter Big Data Analytics for Time Series of Electricity Consumption," in 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), Aug. 2018, pp. 1771–1776. doi: 10.1109/TrustCom/BigDataSE.2018.00267.
8. Z. Yan and H. Wen, "Electricity Theft Detection Base on Extreme Gradient Boosting in AMI," in 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), May 2020, pp. 1–6. doi: 10.1109/I2MTC43012.2020.9128712.
9. Y. Peng et al., "Electricity Theft Detection in AMI Based on Clustering and Local Outlier Factor," IEEE Access, vol. 9, pp. 107250–107259, 2021, doi: 10.1109/ACCESS.2021.3100980.