Electricity Theft Detection using Deep Learning and IoT

1st Mr. Abin M K
Jyothi Engineering College
Computer Science & Engineering
abinkurian47@gmail.com

2nd Ms. Aiswarya Suresh
Jyothi Engineering College
Computer Science & Engineering
aiswaryasuresh258@gmail.com

3rd Ms. Athira C
Jyothi Engineering College
Computer Science & Engineering
athira.c12345@gmail.com

4th Mr. Avin Joseph
Jyothi Engineering College
Computer Science & Engineering
avin.chelseafc@gmail.com

5th Mrs. Aswathy Wilson
Jyothi Engineering College
Computer Science & Engineering
aswathy@jecc.ac.in

***Abstract*—Electricity is an inevitable factor in our day-to-day life. Electricity plays a key role in expediting the socio-economic growth of the country. With the increase in the usage of electricity, we are witnessing a hike in electricity theft. Electricity theft results in wastage of huge amounts of energy and revenue due to which the customers have to pay excessive electricity bills, for the poor power supply they receive. Here we are introducing a novel hybrid CNN-XGBoost model for electricity theft detection, which outperforms the existing systems. In addition to this, our system is capable of locating the area of theft along with warning the concerned users. Here Convolutional Neural Network(CNN) is used to automatically extract the features from the input smart meter data, which consists of the electricity consumption pattern of residential and non-residential users on a half-hourly basis. XGBoost is used as a classifier that distinguishes between fraudulent and normal users. An IoT(Internet Of Things) system is implemented to locate and send warning signals to fraudulent users. Using Ubidots we can create real-time dashboards to analyze data, locate and control devices.**

***Index Terms*—CNN, XGBoost, IoT, Arduino, Ubidots**

I. Introduction

Electricity plays a vital role in our society. It acts as a pillar for economic development. Unethical usage of electricity is increasing in our daily lives. Electricity theft is the act of stealing electricity through meter tampering, meter hacking, illegal tapping, etc. Power losses due to electricity theft are mainly classified into technical and non-technical losses. Technical losses are the losses that occur within the distribution network during the transmission of electricity, due to the cables, overhead lines, and transformers. Non-technical losses occur due to unidentified, misallocated, or inaccurate energy flows, bypassing the electricity meter, or hacking the meter. Electricity theft comes under the category of non-technical losses. Therefore non-technical losses create a hindrance to power sectors all over the world. The world is losing around 89.3 billion US dollars due to electricity theft. The highest contributors to these losses are from India(16.2 billion USD) followed by Brazil(10.5 billion USD) and Russia(5.1 billion USD). Besides, electricity theft carries deadly risks. It is not just dangerous for those who steal but a menace to public safety. Awareness about the conduct of electricity theft in our society is generally low, so early detection of electricity theft assures public safety.

A better understanding of the existing system helps us to give a clear-cut idea of power theft detection. Latest machine learning algorithms along with k means clustering helps to cluster the data according to fraudulent and normal users. Here Artificial Neural Network (ANN) is used for classification of customer's profile [1], the clustered data is classified using machine learning algorithms like Random Forest [2]. Several Deep Learning methodologies are also used to detect electricity theft in smart meter data. CNN as a feature extractor used along with Lstm follows such an approach [3]. The idea of the Wide and Deep Convolutional Neural Networks (CNN) model is to identify the occurrences of electricity theft based on the consumption patterns of customers. Converting one-dimensional data to two dimensional helps to identify the periodicity and non-periodicity in the data as well as it helps to improve feature extraction. This model combines the power of memorization and generalization brought by both wide component and deep component respectively [4]. CNN-based deep learning method for identifying consumer socio-demographic information. CNN automatically extracts features from smart meter data. [5], CNN along with Random Forest approach is used, which is an ensemble technique used for classification [6]. Locating the non-technical loss using A - star algorithm is used in [7]. Unsupervised learning like Mean shift clustering along with convolutional neural network is used in [8]. Electricity theft can be detected and located with the help of IoT. Multiple transformers placed at specific distances help the authorities in identifying the exact location of the theft. [9] [10].

In this paper, we design a novel CNN-XGBoost model for effective electricity theft detection using smart meter data along with an IoT system intended to locate areas of theft and warn the concerned users.

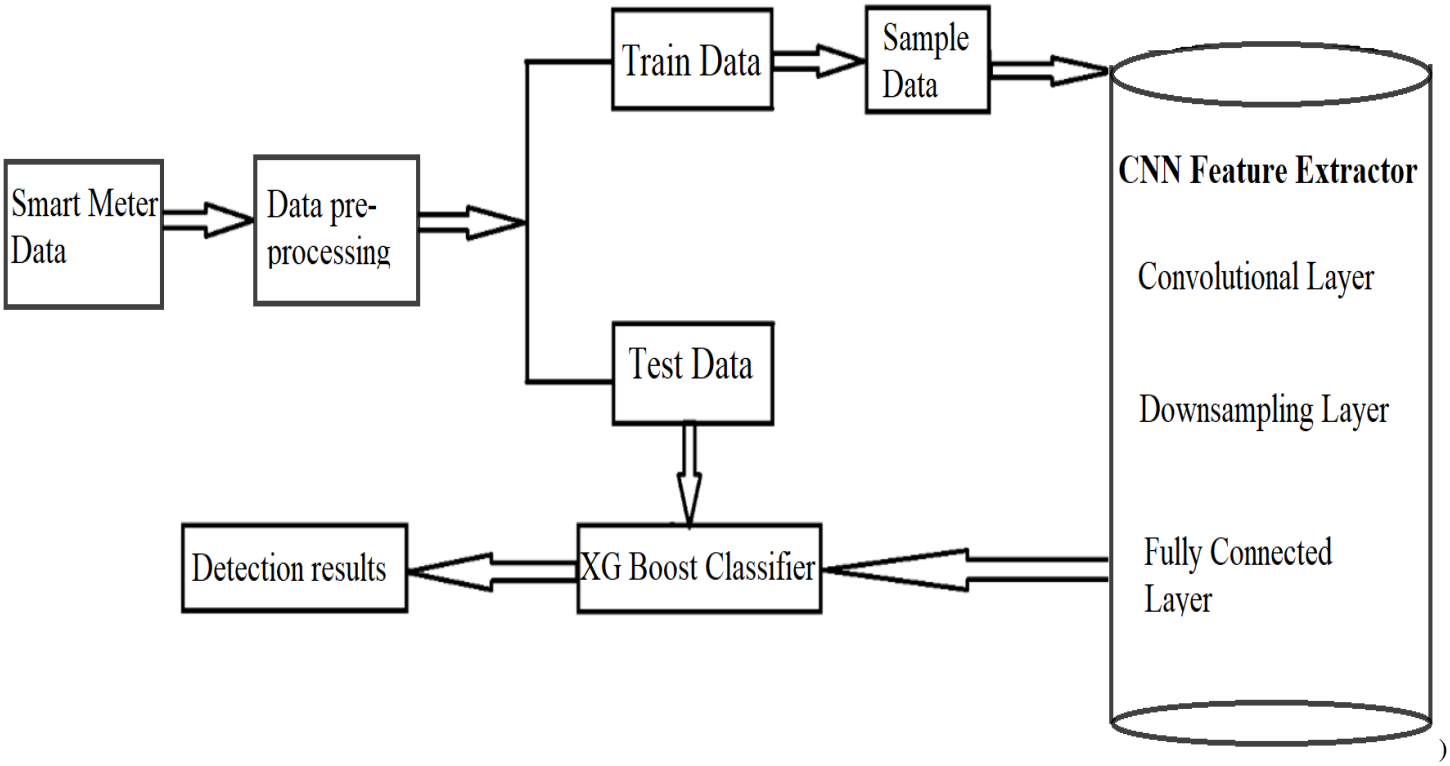

Fig. 1. ARCHITECTURE

II. Proposed System

Our proposed system is capable of detecting and locating the areas of electricity theft and alert the users of the same. Smart meter data is given as input to our system, which consists of the electricity usage data of the customers. The dataset was collected from the Sustainable Energy Authority of Ireland (SEAI) in January 2012. It consists of customer's electricity usage data from over 6000 residential and non-residential users. This dataset contains the half-hourly records of the electricity consumption of customers. Since the dataset was created for research purposes, it mostly contained honest users. Hence we added 1200 malicious customers as described in [11].

Exploratory Data Analysis is performed to understand the varying load profiles. This dataset contains a certain amount of erroneous values, which leads to changes in consumption patterns. Missing values arise in the dataset due to various reasons such as smart meter failure, storage issues, etc. The interpolation method is used to recover the missing values. Normalization is done to change the values into a common scale. Min-max scaling is used to rescale the values to a range from 0 to 1. Data pre-processing is an essential step as it improves the efficiency of the proposed model.

Generation of train test data is needed to evaluate the performance of the model. Dataset is divided into two parts train and test in the ratio 70:30. Train data is used to train the model so that the model will learn the features that are required for predicting. Test data is used to evaluate the model by measuring its accuracy based on actual and predicted values. The data needs to be sampled to maintain the ratio between majority and minority carriers. SMOTE methodology is used for sampling the dataset.

CNN is used as an automatic feature extractor. Different layers of CNN are represented in Fig. 2. CNN consists of a convolutional layer, downsampling layer, and fully connected layer. The main purpose of the convolutional layer is to learn the feature map. It is performed by sliding the kernel over the entire input. Different filters are utilized to perform multiple convolutions to produce distinct feature maps. The pooling layer reduces the number of dimensions and spatial size of the activation map. It helps to control overfitting in networks. Max pooling is the pooling operation conducted here. We get a summarized version of the features detected in the input. Small changes are not addressed but retain the important contents. Fully connected layers are applied for flattening feature maps into one vector. An activation function is to add non-linearity to the model. Without activation function, our neural network will not be able to learn. ReLu activation function is commonly used for hidden layers of a neural network because it is differentiable and efficient for backpropagation.

The features extracted are used as an input to the XGBoost classifier. XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it

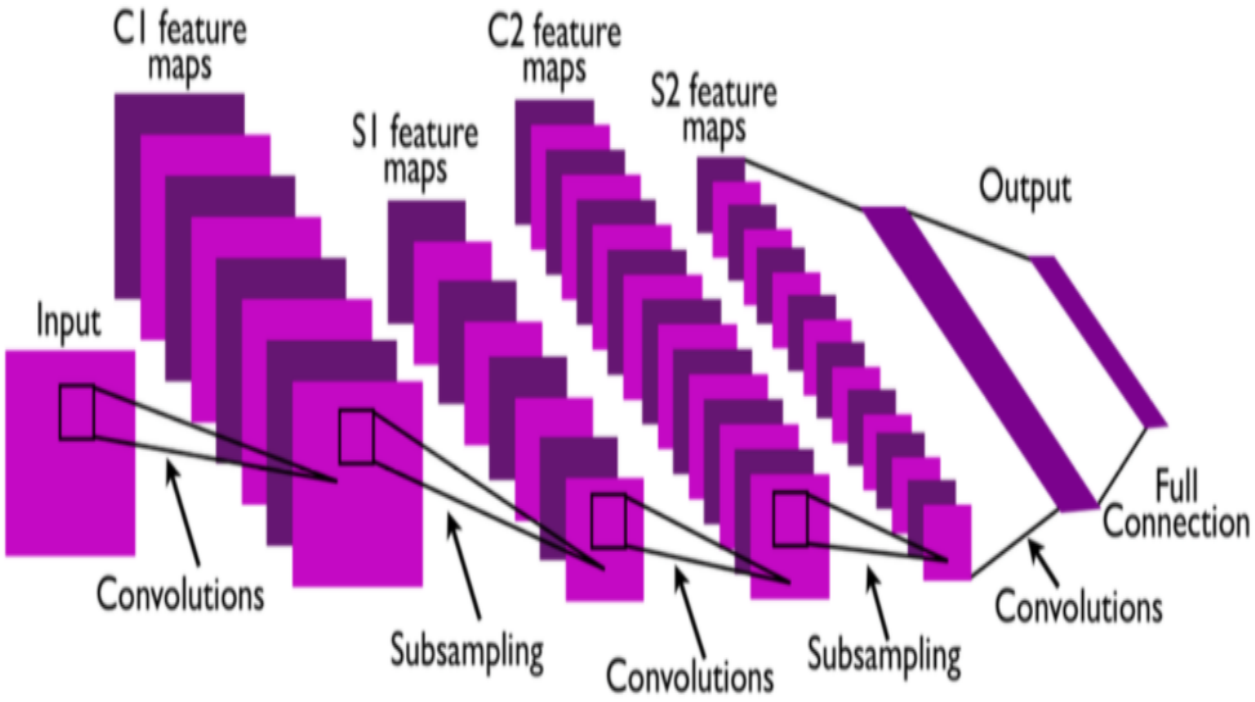

Fig. 2. Convolutional Neural Network (CNN)

comes to small-to-medium structured/tabular data, decision tree-based algorithms are considered best-in-class right now. XGBoost and Gradient Boosting Machines (GBMs) are both ensemble tree methods that apply the principle of boosting weak learners (CARTs generally) using the gradient descent architecture. However, XGBoost improves upon the base GBM framework through systems optimization and algorithmic enhancements.

The Internet Of Things refers to an embedded system with sensors, software, and internet-connected objects that are used to connect and exchange data over a wireless network without human intervention. In an ideal case, the outgoing amount of electricity should be equivalent to the sum of current consumed by users belonging to that particular area. The difference in the outgoing and consumed amount of electricity indicates the chances of occurrence of electricity theft. In practical cases, transmission losses may arise which results in a mismatch of the outgoing and consumed electricity [12]. A threshold value can be calculated by taking the average of differences in outgoing and consumed electricity. The location of power theft can be determined by placing multiple transformers in the load line. The serial number specified in the transformers is used to locate the areas prone to theft.

Electricity consumption data obtained from the smart meters are given to the Arduino UNO to monitor the consumption patterns. In our system, the threshold value will be sent via Arduino Ethernet Shield with an internet connection to the Ubidots IOT Cloud [13]. The Ubidots stores the collected electricity consumption data into the IoT database. Whenever the consumption value exceeds the threshold value, the Ubidots Event manager invokes a notification alert to the end-users.

Fig.3 describes how the Events Manager triggers alerts inside an active event window. Whenever the data (blue line) passes through the Threshold an event is triggered, i.e when the electricity consumption data exceeds the threshold value an alert message is sent to the corresponding users. For the Ubidots to trigger the next event, the data must fall below the threshold again. After an alert is triggered, subsequent values will not be triggered again, even if they comply with the trigger conditions. A second trigger cannot take place unless the data values return below the threshold value and exceed the threshold again.

III. Result

In this paper, electricity theft detection is considered as a binary classification problem, which distinguishes between normal and malicious users. By using CNN - XGBoost model the accuracy increases comparing to other existing models.

An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:

- True Positive Rate
- False Positive Rate

A ROC curve plots TPR vs. FPR at different classification thresholds. Lowering the classification threshold classifies

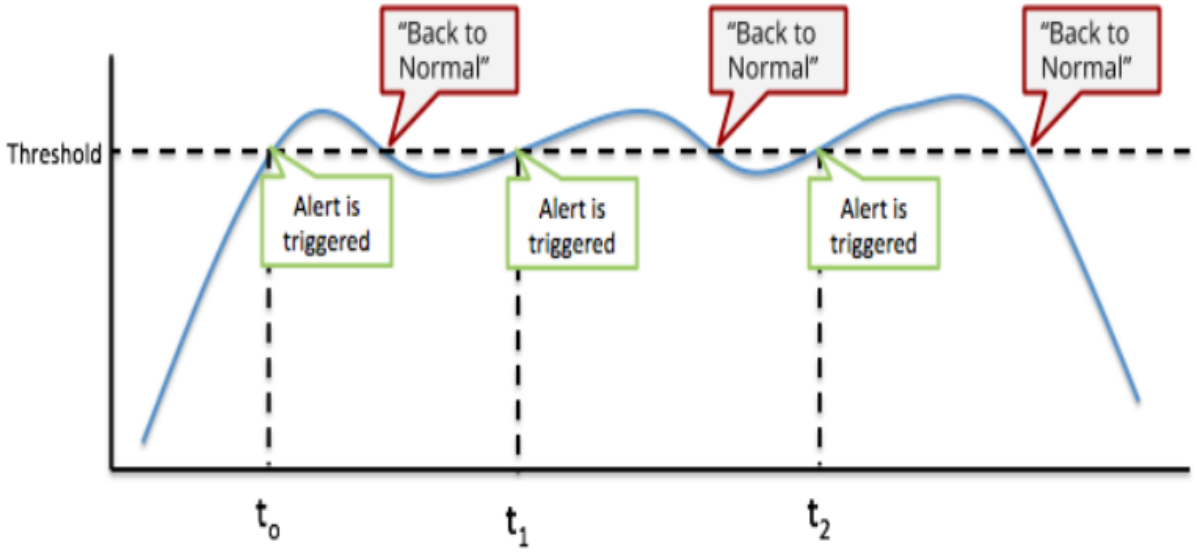

Fig. 3. Ubidot Event Manager

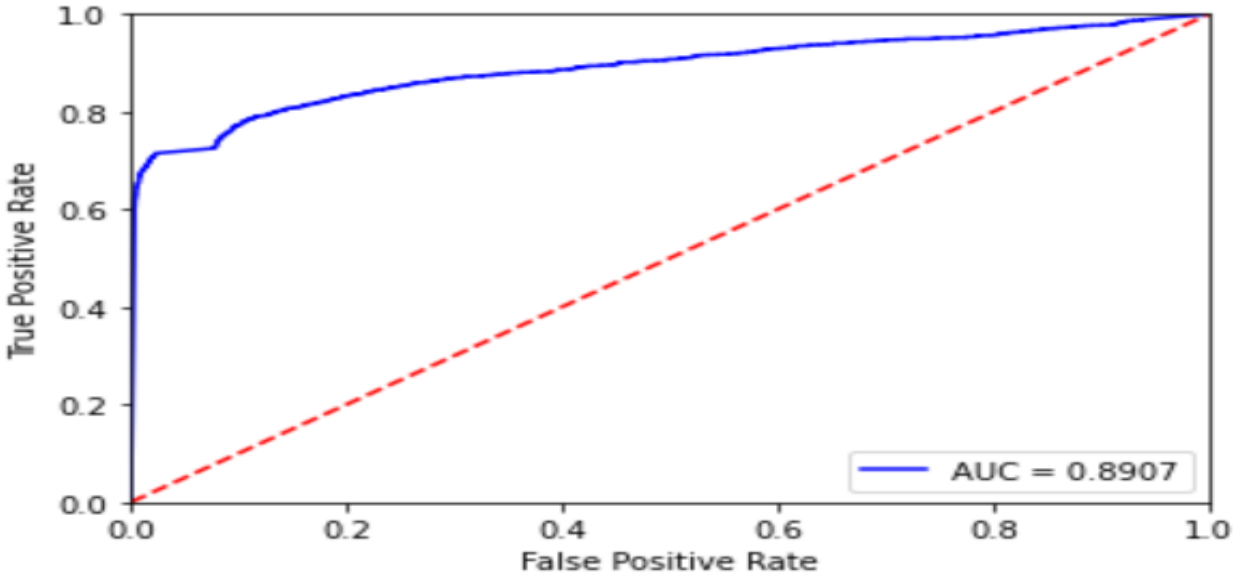

Fig. 4. ROC Curve and AUC

more items as positive, thus increasing both False Positives and True Positives. AUC stands for "Area under the ROC Curve." AUC provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is the probability that the model ranks a random positive example more highly than a random negative example.

Classifiers that give curves closer to the top-left corner indicate better performance. The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes. In Fig.4 we observe that the positive (blue) line moves closer to the top left corner with an AUC value of 0.8907, which indicates that the model is having high accuracy.

IV. Conclusion

Here we are using deep learning techniques to detect and locate electricity thefts, overcoming the limitations of the existing theft detecting systems thus providing an efficient solution to tackle the issues caused. We have outlined the design of the proposed project, which aims to identify algorithms and features that can best detect and locate the electricity thefts occurring using the abnormal changes in the consumption pattern of customers from the smart meter. The proposed

model is based on the combination of CNN and XGBoost classifier, integrated with an IoT module. Here the CNN is similar to an automatic feature extractor in investigating smart meter data and the XGBoost is the output classifier. The mismatch in the outgoing electricity from the transformer and the consumed electricity by the users is identified and located. Whenever a mismatch is encountered a warning message is sent to the concerned users. The use of the technique proposed here will help the power utilities in detecting theft within a lesser amount of time and with comparatively high accuracy. The use of a hybrid neural network for electricity theft detection is presumed to improve the performance of detection. It can be concluded that this model is capable of effectively detecting power theft, thus reducing the rate of its occurrences.

REFERENCES

[1] J Jeyaranjani and D Devaraj. Machine learning algorithm for efficient power theft detection using smart meter data. International Journal of Engineering Technology, 7(3.34):900–904, 2018.

[2] Rohit Bamane, Mervyn Vinod, Jahnavi Shah, Shweta Ahuja, and Aditya Sahariya. Smart meter for power theft detection using machine learning.

[3] Md Hasan, Rafia Nishat Toma, Abdullah-Al Nahid, MM Islam, Jong-Myon Kim, et al. Electricity theft detection in smart grid systems: a cnn-lstm based approach. Energies, 12(17):3310, 2019.

[4] Zibin Zheng, Yatao Yang, Xiangdong Niu, Hong-Ning Dai, and Yuren Zhou. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. IEEE Transactions on Industrial Informatics, 14(4):1606–1615, 2017.

[5] Yi Wang, Qixin Chen, Dahua Gan, Jingwei Yang, Daniel S Kirschen, and Chongqing Kang. Deep learning-based socio-demographic information identification from smart meter data. IEEE Transactions on Smart Grid, 10(3):2593–2602, 2018

[6] Shuan Li, Yinghua Han, Xu Yao, Song Yingchen, Jinkuan Wang, and Qiang Zhao. Electricity theft detection in power grids with deep learning and random forests. Journal of Electrical and Computer Engineering, 2019, 2019.

[7] Jônatas Boás Leite and José Roberto Sanches Mantovani. Detecting and locating nontechnical losses in modern distribution networks. IEEE Transactions on Smart Grid, 9(2):1023–1032, 2016.

[8] G Johncy and A Anisha Felise. An efficient power theft detection using mean-shift clustering and deep learning in smart grid. In IOP Conference Series: Materials Science and Engineering, volume 983, page 012003. IOP Publishing, 2020.

[9] IOT based Power Theft Detection. R Giridhar Balakrishna, P Yogananda Reddy, M L N Vital.. International Journal of Innovations in Engineering and Technology (IJIET) ISSN: 2319-105, Volume 8, Issue 3, June 2017

[10] Ajay Mahato, Abhishek Nanda, Ajay Kumar Pal, Chandan Kumar Singh, "Electric Power theft detection and location Tracking using IOT", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 4, Issue 5, pp.35-39, March-April-2018.

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

[12] Z. H. Che Soh, I. H. Hamzah, S. A. Che Abdullah, M. A. Shafie, S. N. Sulaiman and K. Daud, "Energy Consumption Monitoring and Alert System via IoT," 2019 7th International Conference on Future Internet of Things and Cloud (FiCloud), Istanbul, Turkey, 2019.

[13] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot-BasedGarbageMonitoring and Clearance Alert System," 2018 IEEE 9thAnnual Information Technology, Electronics and MobileCommunication Conference (IEMCON), Vancouver, BC, 2018, pp.204-208.