Anirudh Srivastav, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 11, Issue 6, (Series-IV) June 2021, pp. 60-65

RESEARCH ARTICLE OPEN ACCESS

Power Theft Detection Using Machine Learning

Anirudh Srivastav*, Atul**, Subham Kumar Ojha***, Ujjwal Lal****, Dr. Dhanalakshmi R*****

**(Department of Electrical & Electronics, Dayananda Sagar College Of Engineering, Bangalore, India)*
***(Department of Electrical & Electronics, Dayananda Sagar College Of Engineering, Bangalore, India)*
****(Department of Electrical & Electronics, Dayananda Sagar College Of Engineering, Bangalore, India)*
*****(Department of Electrical & Electronics, Dayananda Sagar College Of Engineering, Bangalore, India)*
******(Department of Electrical & Electronics, Dayananda Sagar College Of Engineering, Bangalore, India)*

ABSTRACT
Power theft is one of the serious issues of electric utilities. The untrustworthy clients produce monetary misfortune to the service organizations. It's anything but conceivable to review them physically. The power utilization energy information acquired from the Smart Meter introduced at client premise have the data that is utilized for distinguishing the irregularity of clients. This paper proposes a way to deal with distinguishing the presume clients utilizing the client power use design. Machine learning algorithm is utilized for this reason. The dependability of clients is checked and is chosen for the theft detection. This investigation is done by tweaking the real Smart Meter information to make false information. The ANN arrangement model is created utilizing managed learning calculation that assists with separating the clients profile dependent on their certifiable movement and deceitful action in power utilization. Reproduction result shows that the proposed framework is productive in distinguishing the suspects with high exactness.

Keywords – AMI, K Means Clustering, Machine Learning, Power theft, Smart Meter

---

Date of Submission: 15-06-2021 Date of Acceptance: 30-06-2021

I. INTRODUCTION

A network of transmission lines, substations, transformers termed as electric grid which delivers electricity from the power plant where electricity is produced to consumers. Electric grid is made keen by permitting two-path correspondence between the usefulness and its clients, and the detecting along the channelling lines. Advanced Metering Infrastructure (AMI) networks are used by modern smart grids for billing and monitoring purposes. However such an approach also results in power theft. High financial losses are the result of electricity bills for many countries such as the US which is $6 billion/year and India which is $17 billion/year. Almost 50% of their electric revenue is lost by developing countries due to theft. To deal with this problem we have come up with a solution to identify suspicious clients by using their power usage pattern. Automated Metering Infrastructure (AMI), Phasor Measurement Unit and Communication network are the components of the smart grid. The AMI depicts the entire framework from a keen meter to a two way communication system to control focus hardware and every application that is related and collecting energy use data in close to continuous. Smart meter, meter data acquisition system, communication network, meter data management system are the main components of AMI. Following works are done by AMI : framework dependability, energy cost, and power burglary. Service switching, time-based rates, remote programming to control smart devices, power quality measure, and a user interface for real-time monitoring are the functionalities included in AMI . It is a computerized gadget having the highlights to gather the utilization information generally on an hourly premise (may vary). Our approach uses Machine Learning Algorithm For Efficient Power Theft Detection using Smart Meter Data. Regular reports of customer power consumption are used by the AMI which is sent by the smart meter installed at their homes. Physical electricity theft such as meter tampering and linehooking can be stopped by this approach. The stakeholders here are the Electricity Distribution Companies through local substations to detect fraudulent clients and the Electricity Distribution Companies through local substations to detect fraudulent clients.

II. LITERATURE REVIEW

The most common way of doing theft is bypassing the meter with the help of a piece of wire which is detected by the electricity theft detection

system in which consumers simply bypass the electricity meter by placing a wire before and after the meter reading unit which is responsible for counting the current unit. In meters where the theft occurs, the proposed system is hidden. [1] This project will automatically collect the reading from houses or industries and compare with the overhead line values and vice versa. If any theft occurs it turns on the relay circuit and trips the main circuit. It also sends a message to the Electricity Board (EB) to inform about the theft and the area of its occurrence. Here a current sensor, voltage sensor and detection of short circuit is done in the lines. The output values of all the sensors are loaded into the PIC microcontroller and the comparison values are displayed on a LCD display. Also, if any mismatch occurs in the values a message will be sent to the EB. Immediately EB can find out the person who is responsible for the theft. Actions like switching off the power temporarily or turning on the relay can be done meanwhile. [2] For sending the data about power robbery to the EB, IOT and GSM technology is implemented in this proposed system which prevents the current scenario Raspberry pi and Arduino are interfaced with the system via serial communication and voltage and load current is sensed using sensors which are interfaced with raspberry pi. Thus, preventing electricity theft . The issues like force burglary, wastage of energy and transmission line shortcoming are solved by this system which are faced by power providers. Through this the system operators identify, electricity thefts, and generate bills for the electricity without violating customers privacy. [3] The secret sharing is used by PPETD to send masked data for the purpose of generating bills to the power providers. Arithmetic and binary circuits utilizing secure two-party conventions are executed to assess a summed up convolutional-neural organization model on the masked power utilization readings with the end goal of power theft recognition by the system administrator and every purchaser . To assess the security and the presentation of the PPETD , a broad investigation of genuine datasets is performed. In this manner affirming the exactness of the plan in identifying deceitful consumers with privacy preservation and acceptable communication and computation overhead. The majority of the past research work centers around client load profile data to uncover strange conduct that is known to be profoundly related with Non Technical Losses (NTL) exercises. [4] It gives internationally upgraded SVM hyper-boundaries utilizing a blend of arbitrary and prepopulated genomes. The decision tree strategy is carried out for tracking down the potential fraud action. The Artificial neural network is constructed to group the NTL – influence altering for brilliantly distinguishing the misfortunes by choosing the most required highlights from the client profile in. The extreme machine learning technique explains the activity of recognizing the client energy utilization design that characterizes veritable and unlawful profiled clients. Ordinary energy utilization information just as the encoded information is applied with the arrangement models through which relating grouping correctness and computational overhead is looked at.

The past work done in this field in theft detection focuses around the client power utilization profile information. The particular are the places where there is uniqueness in provided power and charging for that power. Every one of the clients having a place in that space are viewed as suspects. The disadvantage of the work talked about beforehand is that power theft distinguishing proof is dependent with the understanding that the clients are suspected to be misrepresentation. This case could recognize the trustworthy client as the fake client. This gives the inspiration to this exploration work to incorporate the false client power utilization information into the real force utilization information. The ML algorithms are utilized to dissect the information in a true sense and afterward arrange the client accordingly. The client's information is divided as fraud and genuine depending on their utilization design.

III. PROPOSED SYSTEM

Fig 1, presents the flowchart of the proposed system for power theft recognition. The clients profile dataset is given as contribution to k-means clustering algorithm. The k-means clustering algorithm will group these clients into k groups. The customers profile that is adequately close to the bunch head is the chosen clients for the robbery identification task. These clients are recognized as reliable clients. The chosen clients profile information are the contribution to the smart meter data. Three kinds of counterfeit information are created utilizing these chosen clients profile dataset. Then, at that point the classifier is prepared to classify the ordinary and bogus information that is created and included into the real authentic dataset. The exhibition of the classifier is estimated and assessed.

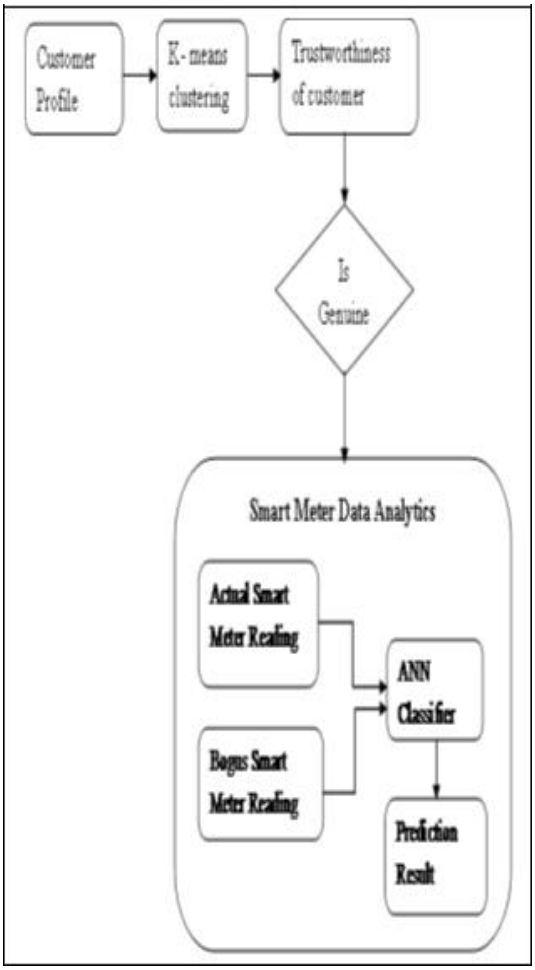

Fig 1 : Proposed Power Theft Detection Model

IV. METHODOLOGY

4.1 TRUSTWORTHINESS OF CUSTOMERS

The customer profile data of a specific area is chosen from Smart meter data which is available. It is taken into consideration that all the clients of the selected region are not fraudulent. Machine learning algorithm is used to identify the trustworthiness percentage of the clients. Unsupervised k-means clustering algorithm is used for this purpose. **K-means** algorithm is an iterative algorithm that tries to partition the dataset into *K* pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to **only one group**. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. The bunching is performed to choose the client's profile which is more certified in their force use design by gathering them in groups. The clients are grouped dependent on their force use readings acquired from their savvy meter. For this purpose, a sample customer's savvy meter reading for every half an hour in (KWH) for five days is considered.

4.2 K-Means Clustering

At first the client's profile is grouped dependent on the data which is obtained from the smart meter. The k-means clustering algorithm permits us to determine the quantity into the anticipated group. It bunches the clients depending on their five days power usage. K groups are shaped to oblige the client's profile in any one of the explicit bunch. 'K' is simply the quantity of groups decided. The quantity of groups is resolved in correspondence to the quantity of client profile contemplated. The groups are shaped with number of client's profile, where the bunch with less number of client profiles are discarded. The group with the greatest number of client profiles is chosen. The clients having a place with the group that fulfills these two conditions are observed for the Smart Meter Data Investigation. The unmistakable clients of the chose group are distinguished by ascertaining the Euclidean distance of every client profile to its comparing cluster head. The Euclidean distance is utilized for discovering the distance between the bunch part and group head. The client's profile which is near the group head is recognized as a real profile. The certifiable profiles are isolated and their energy utilization information are taken for additional examination. The dependability of the client is confirmed by utilizing the bunching calculation.

```
[ ] unique_elements, counts_elements = np.unique(y_kmeans, return_counts=True)
    print("Frequency of unique values of the said array:")
    print(np.asarray((unique_elements, counts_elements)))

Frequency of unique values of the said array:
[[  0   1   2   3]
 [327 310  26 316]]
```

Fig 2 : Divided In Optimal Clusters

4.3 Smart Meter Data Analytics

The real clients profile is acquired from the aftereffect of k-means clustering. For every client, the keen meter perusing is acquired for every half an hour. This inspecting rate is diminished to one perusing (normal perusing of forty eight interpretations per twenty four hours) out of every day per client. For every one of the considered examples in the dataset, three sorts of fake information tests are produced for consistently perusing. In type 1, an irregular worth is created between - 0.500000 and 0.50000. This irregular worth is increased with the normal perusing esteem determined for each day. It duplicates the adequate change in meter perusing. This is accomplished for each of the five days interpretations of all example

clients considered for the examination. In type 2, irregular days are taken where the genuine information esteems are supplanted with nothing. This suggests the situation where the meter doesn't work. In type 3, the mean worth of five days interpretations are duplicated with every one of the day readings.

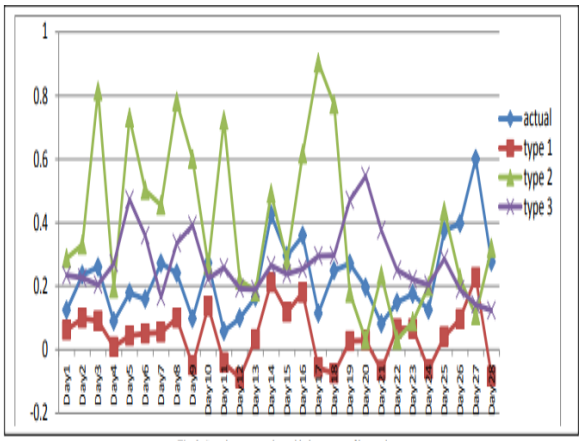

Fig 3 : Three types of Dataset

4.4 Classification

Subsequent to discovering the dependability of the clients, the certified profiled clients are considered for the characterization model by including the counterfeit information into real information. The Artificial Neural Network is worked to group the client's profile. The three kinds of fake information alongside the genuine information are considered to prepare the neural organization. Sixty or eighty percent of the dataset is used for preparing the neural organization. After a necessary number of emphasess, the neural organization is prepared to foresee any new client profile to veritable or extortion. The excess twenty or forty percent of the dataset is utilized for testing the dataset. The forecast is made by the ANN grouping model. Two boundaries exactness and mistake rate is utilized for the exhibition of the proposed framework. The distinction in real class esteem and the anticipated class esteem is considered for the exhibition assessment. The exhibition of the model relies upon the quantity of dataset thought about.

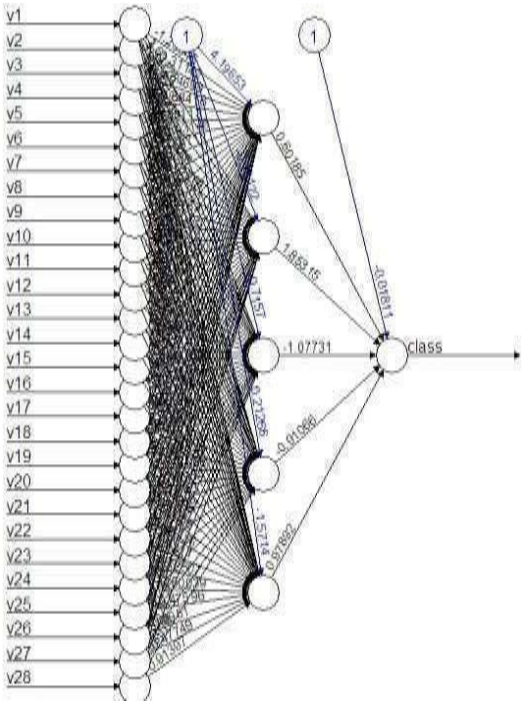

Fig 4 : Artifical Neural Network

V. TESTING RESULT & ANALYSIS

In order to test how our prediction model performs, we test it on a testing dataset. Performance Analysis function is called to predict the output for the testing dataset and the accuracy for the same is computed. Accuracy of the prediction model comes out to be 97 % and the reading at particular index is predicted and categorised as fraudulent or genuine. The model perfroms efficiently on the testing dataset. In order to validate the performance of the dataset further we build a Confusion Matrix of our predicted results.

```
Accuracy
0.9758064516129032
Reading at index 0 is fradulent.
Reading at index 1 is fradulent.
Reading at index 2 is fradulent.
Reading at index 3 is genuine.
Reading at index 4 is fradulent.
Reading at index 5 is fradulent.
Reading at index 6 is fradulent.
Reading at index 7 is fradulent.
Reading at index 8 is genuine.
Reading at index 9 is genuine.
Reading at index 10 is fradulent.
Reading at index 11 is fradulent.
Reading at index 12 is genuine.
Reading at index 13 is fradulent.
Reading at index 14 is fradulent.
Reading at index 15 is fradulent.
Reading at index 16 is fradulent.
Reading at index 17 is genuine.
```

Fig 5 : Predicted Result with Accuracy

5.1 Confusion Matrix

A confusion matrix is a two x two matrix that contains four possibilities given by a binary

classifier. Various metrics, such as error-rate, accuracy, specificity, sensitivity, and precision, are extracted from the confusion matrix.
A binary classifier predicts all information cases of a test dataset as one or the other positive or negative. This classification (or prediction) produces four outcomes – true positive, true negative, false positive and false negative.

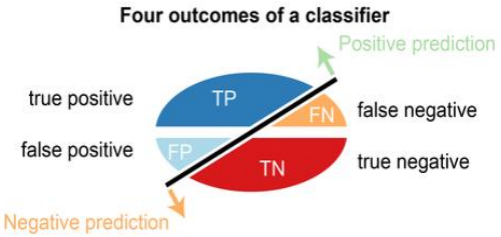

Fig 6 : Outcomes of Classifier

Other evaluation measures that can be derived from the confusion matrix are :
5.1.1 Sensitivity
Sensitivity (SN) is determined as the quantity of right certain expectations isolated by the complete number of positives. It is also called recall (REC) or true positive rate (TPR). The best affectability is 1.0, though the most noticeably awful is 0.0.

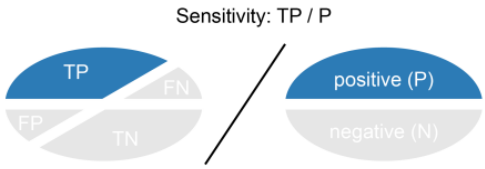

Fig 7 : Senisitivity

5.1.2 Specificity
Specificity (SP) is determined as the quantity of right adverse expectations isolated by the complete number of negatives. It is also called true negative rate (TNR). The best explicitness is 1.0, while the most exceedingly terrible is 0.0.

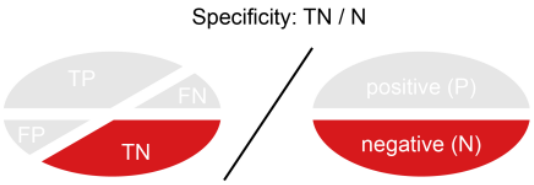

Fig 8 : Specificity

5.1.3 Precision
Precision (PREC) is determined as the quantity of right sure expectations isolated by the complete number of positive forecasts. It is also called positive predictive value (PPV). The best exactness is 1.0, though the most noticeably awful is 0.0.

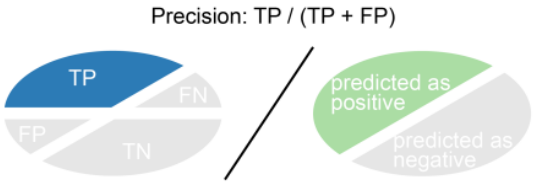

Fig 9 : Precision

5.2 ACCURACY AND CONFUSION MATRIX FOR THE ORIGINAL DATASET

```
Accuracy
0.9709677419354839
```

Fig 10 : Confusion Matrix & Accuracy (Original Dataset)

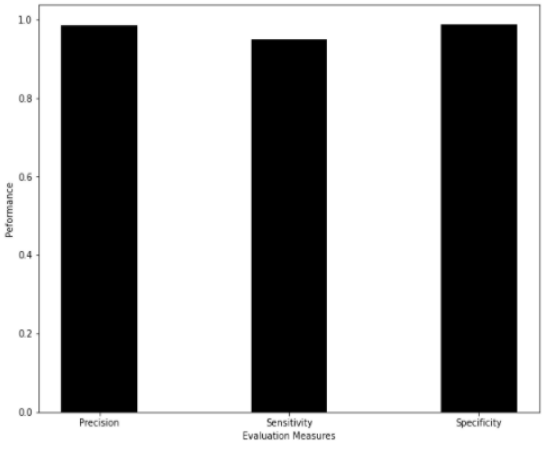

Fig 11 : Bar Graph for Evaluation Measures

VI. CONCLUSION

The model is presented with a novel K Means-ANN model which detects electricity theft. ANN is the output classifier in this model. A completely associated layer with a dropout rate is planned during the preparation stage to improve the danger of overfitting because of countless boundaries. The proposed model is a serious promising order technique in the power robbery

identification field as the framework demonstrates due to following features: The first is that highlights can be naturally extricated by the half and half model, while the accomplishment of most other customary classifiers depends to a great extent on the recovery of good hand planned highlights which is a relentless and tedious assignment. The second lies in that the crossover model consolidates the benefits of the K means and ANN, as both are the most famous and fruitful classifiers in the power robbery discovery field. As the security of the buyer is influenced by the discovery of power robbery, the future work will zero in on how the protection is influenced by researching the granularity and span of savvy meter information. Stretching out the proposed crossover model to different applications (e.g., load forecasting) is an assignment worth researching.

REFERENCES

[1]. Power Theft Detection Using GSM Technology, *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 2013
[2]. Power Monitoring and Theft Detection System using IoT, *International Conference on Physics and Photonics Processes in Nano Sciences*, 2017
[3]. PPETD : Privacy-Preserving Electricity Theft Detection Scheme With Load Monitoring and Billing for AMI Networks, IEEE.
[4]. Patrick Glauner et al. "Large-scale detection of non-technical losses in imbalanced data sets". In: *Innovative Smart Grid Technologies Conference (ISGT), 2016 IEEE Power & Energy Society*. IEEE. 2016, pp. 1–5
[5]. 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
[6]. Blue Yonder GmbH Maximilian Christ. How to add a custom feature. tsfresh. May 28, 2017.
[7]. Patrick Glauner et al. "The Challenge of Non-Technical Loss DetectionUsing Artificial Intelligence: A Survey". In:*International Journal of Computational Intelligence Systems 10.1* (2017), pp. 760–775

Anirudh Srivastav, et. al. "Power Theft Detection Using Machine Learning." *International Journal of Engineering Research and Applications (IJERA)*, vol.11 (6), 2021, pp 60-65.