International Journal of Research in Science advanced Technology and management studies (IJRSTMS)

ISSN: 2459-425X ● Website: www.ijrstms.com

Enhancing the Security of Smart Grid using Neural Networks

Dr. ARVIND PRASAD

[1] Assistant Professor, Department of Computer Science and Engineering, University College of Engineering JNTU Kakinada

[2] Student, Department of Computer Science and Engineering, University College of Engineering JNTU Kakinada

ABSTRACT

The primary goal of this initiative is to contend with the global issue theft of electricity concern, which has adverse effects on both electricity consumers and utility companies. This illegal activity interrupts the economic growth of utility companies, poses electric hazards, and leads to higher energy costs for users. To combat this problem, we aim to develop an effective approach to predict and detect theft of electricity from smart grids, using Artificial Neural Networks (ANN) as the core technology. The project will utilize a dataset of electricity usage, sourced from the popular web repository Kaggle, as the basis for training the ANN. The collected data will undergo preprocessing to ensure it is suitable for the ANN. By feeding this pre-processed data into the ANN model It will be able to gain knowledge from the trends and outliers found in the data on electricity consumption. The accuracy, precision, recall and F1-score for the suggested model are 98%, 98%, 97% and 98%, respectively.

Keywords: Electricity Theft, Smart Grid, Artificial Neural Network, Deep Learning, Electricity Consumption, Agglomerative Clustering, Smart Meter Data

1. INTRODUCTION

The rapid advancement of technology has led to the development of smart grids, which leverage digital communication and intelligent control systems to enhance the efficiency, reliability, and sustainability of power distribution. However, alongside these benefits, smart grids also face challenges such as electricity theft, which can significantly impact the revenue of utility companies and compromise the stability of the grid.

Electricity theft refers to the illegal and unauthorized consumption or tampering of electrical power without proper metering or payment. Traditional methods of theft detection, relying solely on manual inspections and periodic meter readings, are often insufficient in identifying sophisticated theft techniques and patterns. As a result, utility companies are increasingly turning to advanced technologies, such as data analytics and machine learning, to enhance their theft detection capabilities in smart grids.

This project's objective is to create a system for predicting energy theft in smart grids that is based on Artificial Neural Networks (ANN). The suggested method would make use of a dataset on electricity use that was taken from the well-known web repository Kaggle. The pre-processed data will be sent into the artificial neural network (ANN), which will train itself to spot patterns and abnormalities in the consumption data. The ANN

model will be tested using data that includes cases of electricity theft after being trained using a dataset of acceptable usage patterns.

The decision to validate the ANN model using a diverse dataset, encompassing both instances of theft and instances of acceptable usage, underscores its versatility and adaptability. Rigorous testing underpins the model's reliability in distinguishing between genuine deviations and actual theft occurrences. The iterative nature of machine learning also allows the ANN to evolve alongside the shifting landscape of energy theft, ensuring that it remains a stalwart defense against emerging tactics.

Furthermore, the ANN-based method is not a static solution but a dynamic framework that evolves with the changing landscape of smart grids. Regular updates to the model's training data and algorithmic parameters enable it to adapt to emerging theft techniques. This adaptability positions utility companies to stay ahead of innovative theft methods and provides a sustained defense against unauthorized consumption.

In the pursuit of a sustainable energy future, the proposed ANN-based method offers a significant step forward in securing the efficiency and integrity of smart grids. By amalgamating cutting-edge technology, data-driven insights, and industry collaboration, this approach not only detects electricity theft but also fosters a resilient energy ecosystem that can fully capitalize on the transformative potential of smart grids. As energy systems continue to evolve, embracing proactive and intelligent solutions will be instrumental in ensuring a reliable and sustainable energy supply for generations to come.

2. RELATED WORK

Liu et al.,(2015) introduce an analysis of cyberattacks involving energy theft through metering manipulation for bill reduction. They propose a diagnostic approach utilizing Bollinger bands and partially observable Markov decision processes (POMDP). Simulation results indicate an average detection rate of 92.55% for energy theft, effectively mitigating neighborhood impacts.

Gao et al., (2019) develop a data-driven model with physical insights to identify electricity theft through smart meter data. This model solely relies on voltage and electricity usage data from smart meters, avoiding the need for imprecise parameters and secondary network information. The model's efficacy is verified using real smart meter data, showcasing its effectiveness in detecting cases of electricity theft.

Zheng et al., (2018) suggest a novel strategy for combating electricity theft using a Wide & Deep Convolutional Neural Network (CNN) model. A Wide feature and a Deep CNN feature are integrated in the model. The Deep CNN function accurately detects between routine usage trends and theft abnormalities using two-dimensional electrical consumption data. AUC (0.993), precision (0.90), recall (0.91), f1-score (0.89), and accuracy (0.89) are some of the model's remarkable performance measures.

Peng et al., (2021) address under-fitting challenges caused by imbalanced datasets in energy theft studies. Their solution involves an outlier detection approach combining clustering and local outlier factor (LOF). Employing k-means, they examine load profiles, select outlier candidates with profiles distant from cluster centers. The LOF method is, so used to quantify the abnormal degrees of these candidates, enhancing the detection of various theft attacks.

Ismail et al., (2020) examine consumer manipulation of renewable-based DG unit monitors in smart meters, causing inaccurate energy supply reporting and excessive utility charges. Their proposed solution employs a deep convolutional-recurrent neural network trained on DG smart meter, weather, and SCADA data, yielding a notable detection rate (99.3%) and minimal false alarms (0.22%). In a related vein, Liu et al., (2020) address hidden electricity theft (HET) attacks utilizing emerging MP systems. Their study highlights the potential for hackers to manipulate smart meters, leading to reduced-rate electricity billing. To counter HET attacks, they

suggest strategies such as meter-based protection, attack cycle limitation, and billing process enhancements, fortifying smart grid security.

Khan et al., (2023) underscore the importance of synchronized procedures for classifying electricity theft, proposing an integrated ETD structure with three modules. Their innovative approach tackles challenges related to unstandardized usage data, outliers, and imbalanced datasets. In a related context, Ullah et al., (2022) apply in an integrated DL model optimized for effective electricity theft detection in smart grids. By combining AlexNet to handle dimensionality issues and AdaBoost for final categorization, their model addresses issues like class imbalance and hyper-parameter tuning.

Khan et al., (2022) tackle challenges associated with missing values, data volatility, and non-linear relationships in significant power consumption datasets. They introduce a comprehensive ETD model with three integrated components, utilizing diverse machine learning techniques to manage these complexities. In a related context Eddin et al., (2022) explore power theft attacks on smart meters, emphasizing adversaries manipulating energy generation claims. Their proposed methods, tailored to single generator and fuel types, address the effects of subtle perturbations caused by attackers. They present an advanced multitask deep-learning-based detector, featuring heightened detection rates, adaptability to various fuel types, and efficiency with a single data source.

3. PROPOSED WORK

Problem Statement

Electricity theft in smart grids is a growing concern that poses significant challenges for utility companies. The existing methods for electricity theft detection often rely on traditional statistical approaches, which may lack the ability to capture complex patterns and adapt to evolving theft techniques. As a result, there is a need for an efficient and accurate approach that can effectively detect employing artificial neural networks (ANNs) to detect theft of electricity in smart grids.

System Methodology

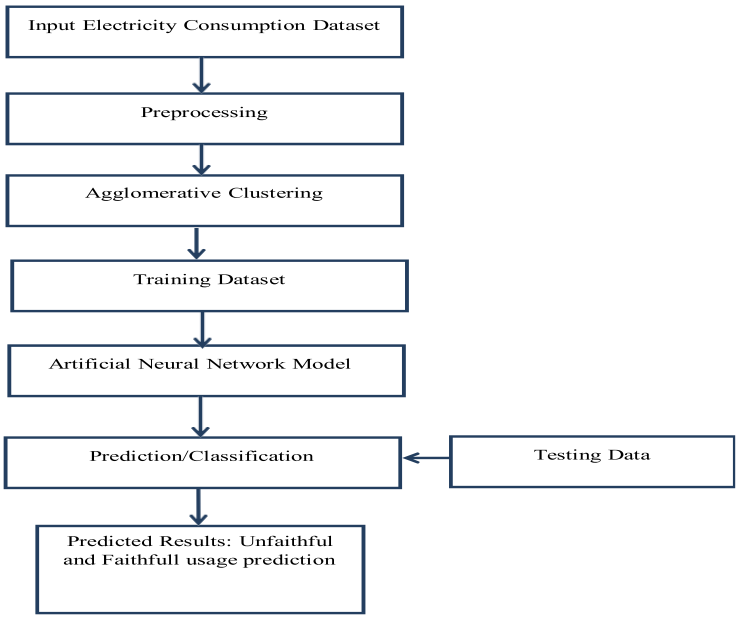

Figure 1 :Proposed System Architecture

Methodology

3.1 Input Electricity Consumption Dataset

The first module is where we gather the data. The actual process for developing a machine learning model and accumulating data establishes now. This is an essential stage that will determine how effective the model is; the more and better data we collect, the more effectively our model will function. Web scraping and other manual interventions are instances of such data collection strategies. The widely used online repository Kaggle is where the Electricity Consumption dataset is found.

3.2 Data Preprocessing

3.2.1 Handling Missing Values

The isnull().sum() function uses to check in the code for missing entries in the daily dataset. For each column, the number of missing values is printed.

3.2.2 Exploratory Data Analysis (EDA)

The code performs exploratory data analysis on the daily dataset using various visualization techniques such as box plots and line plots. It analyzes the distribution of the "energy_count" variable and identifies outliers.

3.2.3 Data Cleaning

Rows in the daily dataset where the "energy_count" is less than or equal to 47 are dropped using the drop() function. This step removes the rows that do not have a complete set of 48 half-hourly energy counts.

3.2.4 Data Manipulation

The code divides the sampled customer IDs into three groups (theft1, theft2, theft3) and applies different transformations to each group. The transformations involve modifying the values of the columns using random numbers drawn from specific distributions (standard normal, beta, and gamma distributions).

3.2.5 Labeling

The transformed data subsets (theft1, theft2, theft3) are labeled with a value of 1 to indicate potential fraudulent activity, while the original data (final) is labeled with a value of 0 to indicate faithful behavior.

3.3 Training Data

The training dataset is created by splitting all of the data into training as well as testing sets. The test_size option is set to 0.2, which means that 20% of the collected information will be used for testing and the remaining 80% for training.

3.4 Artificial Neural Network (ANN) Model

In this module, an Artificial Neural Network (ANN) model will be created and trained using the extracted features and corresponding labels (faithful or unfaithful). The complexity of the neural network that optimizes generalization performance is found through order selection. That many neurons are needed to reduce the inaccuracy in the selection cases.

Biological neural networks that exist in the human brain are the basis for a class of machine learning models known as artificial neural networks (ANNs), which are based on their structure and operation. The interlinked nodes that make up an ANN, known as artificial neurons or units, function together to process and relay information.

The perceptron is the fundamental component of a synthetic neuron. It applies an activation function to a set of input values multiplied by corresponding weights to create an output. The Architecture of ANN is depicted in Figure 2.

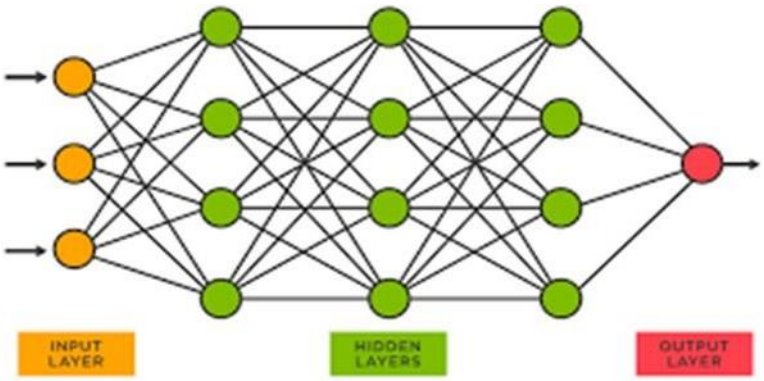

Figure 2: Architecture for Artificial Neural Networks (ANN)

3.5 Testing Data

In the testing data process, a trained Artificial Neural Network (ANN) model's performance is assessed using a different set of data that it was not exposed to during training. This procedure gives information about the model's correctness and dependability and aids in evaluating how well it generalizes to new data. The testing data method starts by segregating a portion of the available data specifically for testing purposes once the ANN model has been trained using a training dataset. By performing this, it is ensured that unbiased, random data will be utilized to evaluate the model's performance.

3.6 Prediction/Classification

The classification process involves training the model on labelled training data and evaluating its performance on unseen testing data. The classification of an ANN model involves training the model on labelled training data to learn patterns, and then evaluating its performance on separate testing data to assess its generalization and classification accuracy.

3.7 Predicted Results

The predictions are based on the trained model's classification decisions for each customer in the testing data. The model assigns a label of either 1 (indicating unfaithful) or 0 (indicating faithful) to each customer based on their energy usage patterns and other relevant features. The predicted results can be analysed and interpreted to identify potential instances of unfaithful energy usage.

4 RESULTS ANALYSIS

The project employed an ANN, a machine learning technique inspired by the human brain's neural networks, known for its ability to learn patterns from data. The system was trained on a comprehensive dataset comprising various parameters related to electricity consumption, grid behavior, and historical theft instances. By learning from this data, the ANN could identify subtle anomalies that might indicate unauthorized consumption or tampering. The ANN demonstrated promising results in detecting electricity theft. It showcased a high accuracy rate in identifying unusual patterns that deviated from normal consumption behaviors.

4.1 Confusion Matrix

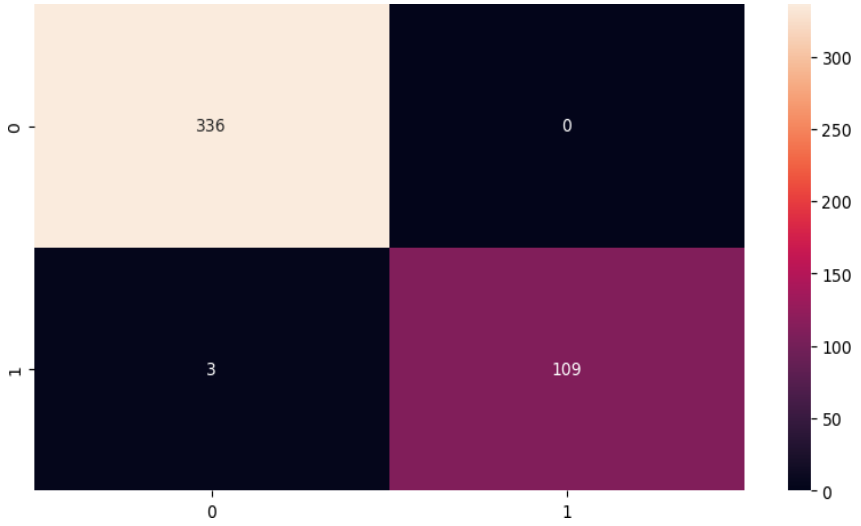

Figure 3 : Confusion Matrix

The confusion matrix is depicted in Figure 3 between the predicted class labels to the actual class labels of a dataset. In confusion matrix

4.1.1 True Positive

True Positive (TP): 336 – This represents how many instances were correctly anticipated as being positive.

4.1.2 False Negative

False Negative (FN): 3 – This indicates situations where the outcome was projected to be positive but turned out to be negative.

4.1.3 False Positive

False Positive (FP): 0 – This is the number of cases where a negative outcome was mistakenly expected to be a positive outcome.

4.1.4 True Negative

True Negative (TN): 109 – The number of accurately anticipated negative instances

4.2 Performance Metrics

Table 1: Performance Metrics

	Precisior	Recal	F1-scor	Suppor
0	0.97	1.00	0.99	336
1	1.00	0.94	0.97	112
Macro avg	0.99	0.96	0.97	448
Weighted avg	0.98	0.98	0.98	448

Performance measures are frequently used to rate the accuracy of predictions made by categorization models.

4.2.1 Precision

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (1)$$

For class 0, among all the occurrences the model predicted as class 0, 97% were actually class 0, while 3% were incorrectly predicted as class 0.For class 1, all instances predicted as class 1 were indeed class 1. This indicates that the model is very confident and accurate when predicting class 1.

4.2.2 Recall

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2)$$

The approach proved successful in accurately identifying all instances of class 0 in practice. This shows that the model didn't overlook any favorable examples for class 0.For class 1, the model correctly predicted 94% of the actual instances. The model was unable to predict certain positive instances for class 1.

4.2.3 F1- Score

$$F1\ score = \frac{2(recall * precision)}{recall + precision} \quad (3)$$

The F1-score for class 0 is high, indicating a strong balance between recall and precision. This shows that the model is successful in class 0 instance prediction and identification. The F1-score for class 1 is also high, indicating that class 1 forecasts have a good balance between precision and recall.

4.2.4 Support

It represents the number of instances belonging to a particular class. In classification tasks, support is frequently used to evaluate how the dataset's classes are distributed.

There are 336 instances of class 0 in the dataset, and the model's predictions for this class were highly accurate and complete. There are 112 instances of class 1 in the dataset, and while the model's predictions for this class were accurate, it missed some of the instances.

4.2.5 Macro average

The macro-average takes the average of metrics across all classes. It treats each class equally, regardless of its support. For the proposed model Macro avg Precision, Recall, F1-score and Support values are 0.99, 0.96,0.97 and 448 respectively

4.2.6 Weighted-average

The weighted-average takes into account the class distribution when calculating the average metrics. It gives more weight to classes with a larger number of instances. For the proposed model weighted avg Precision, Recall, F1-score and Support values are 0.98, 0.98,0.98 and 448 respectively

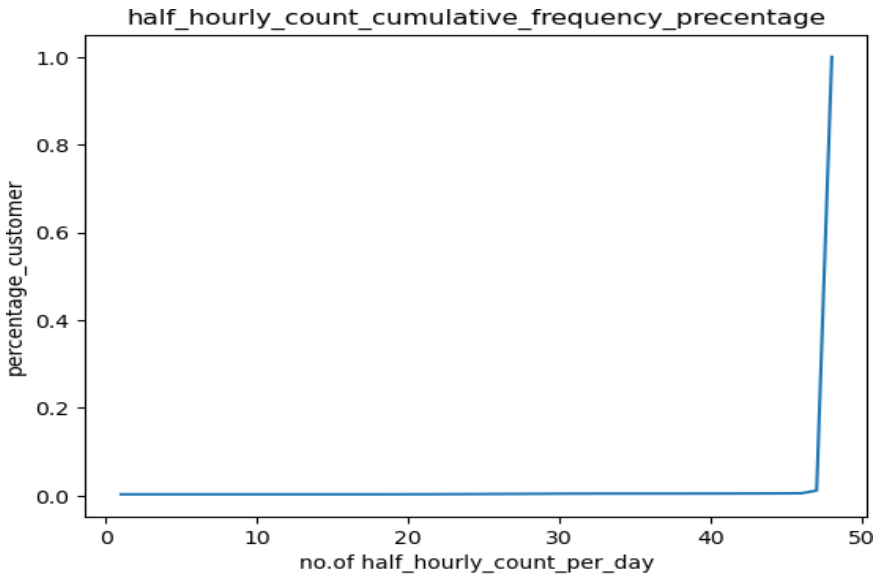

Figure 4: Plot for half hourly count cumulative frequency percentage

Figure 4 shows the plot for half hourly count cumulative frequency percentage, From the above plots and percentages we found that almost 99.99 precent people have all 48 half hourly counts per day so we can drop those 0.01 percentage rows as it is a nominal count.

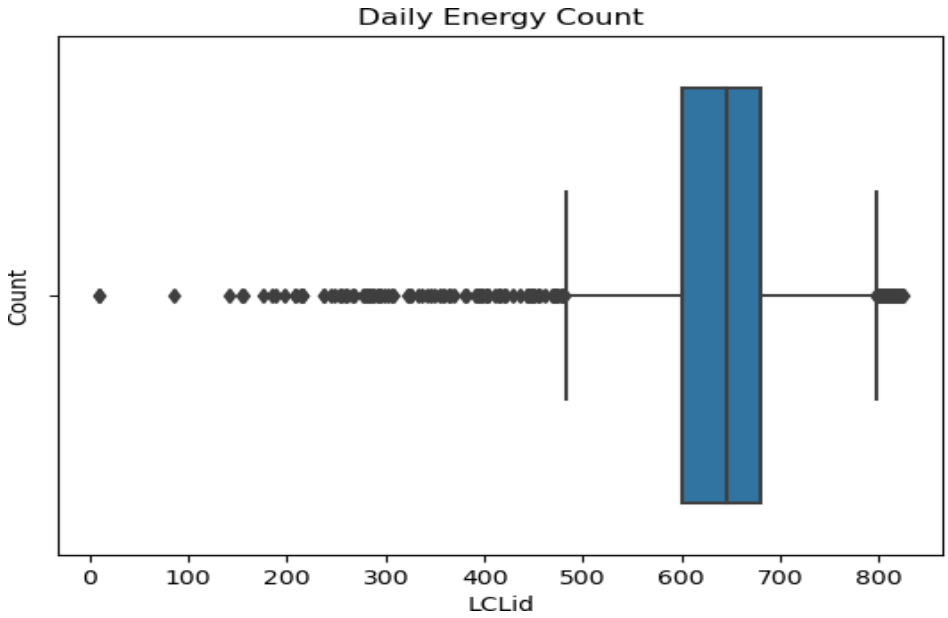

Figure 5 : Box Plot for daily_energy_count

Figure 5 plot would allow you to see the distribution of the counts of daily energy measurements for each "LCLid". Outliers, if present, can indicate data points that are significantly different from the rest. Overall, the boxplot would help you understand the variability in the number of daily energy measurements for different "LCLid" groups and identify potential outliers in the data.

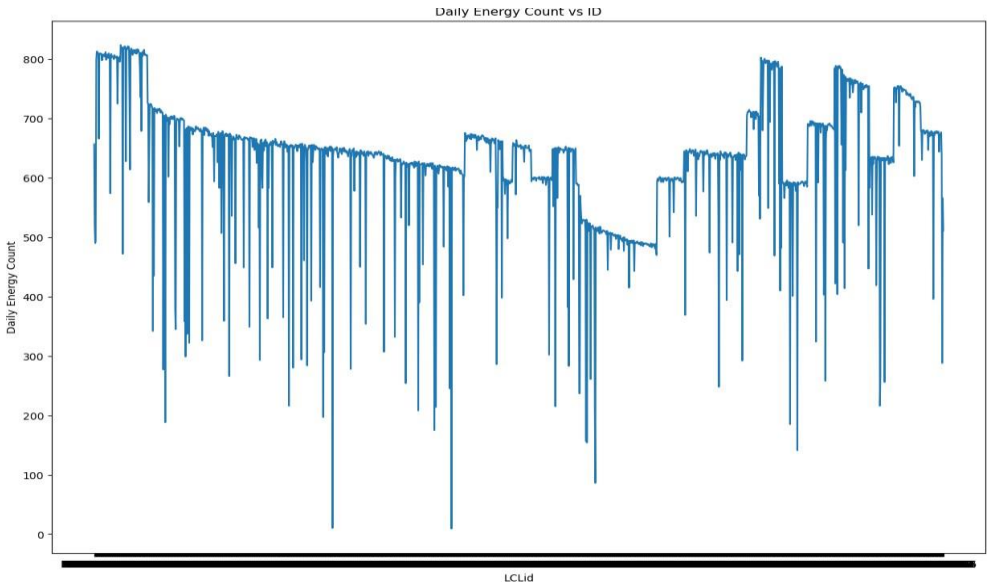

Figure 6: plot for id vs daily_energy_count

Figure 6 displays the plot for daily energy count of particular id's ,some of the ids have large duration which represents more instances, so we need to decide one period(Generally one year) where maximum number of ids will have entries, so that the analysis can be fruitful. You can identify "LCLid" values that have a high count of daily measurements (higher data density) and those with a low count (lower data density).Patterns, trends, or irregularities in the data can be observed through the line connecting the points.

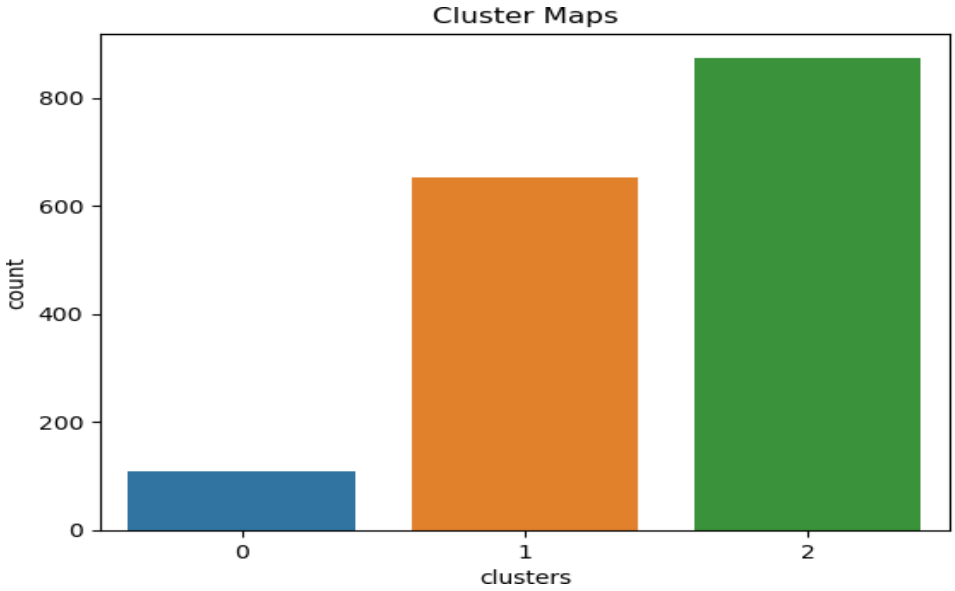

Figure 7:Cluster Maps for different Clusters

Figure 7 which will show the number of times each distinct value appears in the dataset. It clusters selected dataset into three clusters based on similarity of data points. Each bar on the plot corresponds to a category, and the height of the bar represents the number of occurrences of that category.

5 CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

In this study, the use of Artificial Neural Networks (ANN) to predict electricity theft in smart grids found that ANN showed promising results with a high Training and Validation Accuracy of 99%, it was still outperformed by the existing system for classification purposes.

The proposed system relies on analyzing consumption data patterns to identify instances of electricity theft. This approach not only has applications in power distribution networks but also has the potential to be used in anomaly detection across various fields. As a result, the impact of electricity theft can be minimized, leading to improved efficiency and security within the smart grid infrastructure.

5.2 Future Scope

As our method accurately predicts the electricity thefts in the current setting based on historic data, it may be expanded to Integrating real-time monitoring capabilities into the detection system can enable immediate response to potential theft incidents.

ANN can be combined with other machine learning techniques, such as clustering algorithms or expert systems, to create hybrid models for electricity theft detection.

6 REFERENCES

1. Y. Liu and S. Hu, “Cyberthreat analysis and detection for energy theft in social networking of smart homes,” IEEE Trans Comput Soc Syst, vol. 2, no. 4, pp. 148–158, Dec. 2015, doi: 10.1109/TCSS.2016.2519506.
2. Y. Gao, B. Foggo, and N. Yu, “A Physically Inspired Data-Driven Model for Electricity Theft Detection With Smart Meter Data,” IEEE Trans Industr Inform, vol. 15, no. 9, pp. 5076–5088, Feb. 2019, doi: 10.1109/tii.2019.2898171.
3. Z. Zheng, Y. Yang, X. Niu, H. N. Dai, and Y. Zhou, “Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids,” IEEE Trans Industr Inform, vol. 14, no. 4, pp. 1606–1615, Apr. 2018, doi: 10.1109/TII.2017.2785963.
4. 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.
5. M. Ismail, M. F. Shaaban, M. Naidu, and E. Serpedin, “Deep Learning Detection of Electricity Theft Cyber-Attacks in Renewable Distributed Generation,” IEEE Trans Smart Grid, vol. 11, no. 4, pp. 3428–3437, Jul. 2020, doi: 10.1109/TSG.2020.2973681.
6. Y. Liu, T. Liu, H. Sun, K. Zhang, and P. Liu, “Hidden Electricity Theft by Exploiting Multiple-Pricing Scheme in Smart Grids,” IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2453–2468, 2020, doi: 10.1109/TIFS.2020.2965276.
7. I. U. Khan, N. Javaid, C. J. Taylor, and X. Ma, “Robust Data Driven Analysis for Electricity Theft Attack-Resilient Power Grid,” IEEE Transactions on Power Systems, vol. 38, no. 1, pp. 537–548, Jan. 2023, doi: 10.1109/TPWRS.2022.3162391.
8. A. Ullah, N. Javaid, M. Asif, M. U. Javed, and A. S. Yahaya, “AlexNet, AdaBoost and Artificial Bee Colony Based Hybrid Model for Electricity Theft Detection in Smart Grids,” IEEE Access, vol. 10, pp. 18681–18694, 2022, doi: 10.1109/ACCESS.2022.3150016.

9. 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.
10. M. E. Eddin et al., "Fine-Tuned RNN-Based Detector for Electricity Theft Attacks in Smart Grid Generation Domain," IEEE Open Journal of the Industrial Electronics Society, vol. 3, pp. 733–750, 2022, doi: 10.1109/OJIES.2022.3224784.s