International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

ELECTRICITY THEFT DETECTION USING MACHINE LEARNING

[1] Dr. SHANKARAGOWDA BB, [2] MOHAMMED TAHIR B

[1]Associate Professor and HOD, Department of Master of Computer Application, BIET, Davangere
[2]Student, Department of MCA, BIET, Davangere

Abstract: Utility companies are quite concerned about electricity theft. Individual user power usage is one of the vast amounts of data that the smart grid (SG) system generates. Machine learning and deep learning methods can precisely identify electricity theft customers using this data. We develop a convolutional neural network (CNN) model for automatically detecting electricity theft. This study takes into account experimentation to determine the sequential model's (SM) ideal configuration for categorising and identifying electricity theft. The two layers with the best performance each have 64 nodes in the second layer and 128 nodes in the first layer. The precision was up to 0.92. This makes it possible to create high-performance electricity signal classifiers for a variety of applications.

Key Words:Electricity Theft, SmartGrids ,KNN, CNN,Decision Tree, Machine Learning.

I.INTRODUCTION

Utility companies are quite concerned about electricity theft. Individual user power usage is one of the vast amounts of data that the smart grid (SG) system generates. Machine learning and deep learning methods can precisely identify electricity theft customers using this data. We develop a convolutional neural network (CNN) model for automatically detecting electricity theft. This study takes into account experimentation to determine the sequential model's (SM) ideal configuration for categorising and identifying electricity theft. The two layers with the best performance each have 64 nodes in the second layer and 128 nodes in the first layer. The precision was up to 0.92. This makes it possible to create high-performance electricity signal classifiers for a variety of applications.

The presence of SGs increases the likelihood that electricity theft will be resolved. SGs are made up of traditional electricity networks, communication grids that connect smart devices (such smart sensors and metres) in networks, and computing services to perceive and control networks. Smart networks that link service providers and employers transport information and energy. Smart sensors or metres can compile a variety of data in this way, including network status information, information about how much electricity is being used, information about finance, and electrical energy cost . The electrical network is regarded as an SG that can effectively combine the efforts of all employers associated with its producers, customers, and those who do both to supply sustainable, affordable, and safe sources of electricity.

II. LITERATURE SURVEY

The approaches of ETD, which employ consuming data from smart meters to find dishonest customers, were investigated in some literature. Academics have become concerned about the observation of consumer load profiles for signs of electricity theft in conventional power schemes up to this point. In order to create a typical form of power consumption for each user, Angelos et al used five parameters, including maximum consumption, mean consumption, inspection remarks summing, standard deviance, and the mean consumption of the neighbourhood. For the collection of consumers with the same profiles, fuzzy clustering based on K-means was accomplished. Customers who had easy access to the cluster centers were suspected of being dishonest. This ETD system's accuracy was limited by gathering consumers and relying on long-period data, which led to a significant detection delay.

Convolutional neural networks (CNNs) and deep learning have been the focus of prior research. The CNN innovation was explained by Abdel-Hamid et al. in the variety of fields related to developing appreciation, from image processing to speech recognition. The reduction in the number of factors in artificial neural networks (ANNs) is the most favourable aspect of CNNs. This accomplishment has prompted designers and scholars to approximate larger models to address challenging problems, which was not likely with traditional ANNs [20]. In fact, Mallat developed a mathematical framework to analyse the characteristics of broad CNN architectures using methods that were previously provided. At a major level, the extension was made possible by switching from the necessity of contractions and invariants to adaptable collections of local symmetries and contractions.

Additionally, a focus is placed on reviewing various studies that have been done on the use of CNNs for ETD. In this regard, Krizhevsky et al. investigated the use of CNNs to the issue of detecting theft. The periodicity of consecutive data is significant for the classifier and is driven by the numerical model method. Therefore, improving the accuracy of the detection of electricity theft can benefit from a sufficient description of the periodicity. The multiscale DenseNet, which can automatically detect the short- and long-term periodic properties of the successive data, was specifically advised to be adjusted.

Three deep learning techniques—specifically, CNNs, (LSTM) recurrent neural networks (RNNs), and loaded autoencoders—were examined by Bhat et al. for the detection of electricity theft. However, the operation of the detectors was investigated using synthetic data, which prevented a trustworthy evaluation of the detector performance related to shallow structures. The wide and deep CNN model ETD in SGs was the last suggestion made by Zheng et al. Due to their reliance on one-dimensional (1-D) data on electricity use and inability to stop the periodicity of electricity consumption, they discovered that the majority of current systems had poor accuracy in detecting electricity theft.

The goal of this work is to address all of the aforementioned concerns by offering an effective method of electricity theft detection (ETD). To identify electricity thieves, we initially propose a convolutional neural networks (CNN) with the blue monkey (BM) algorithm model, a recently suggested nature-inspired metaheuristic optimization algorithm. A pooling layer, a totally connected layer, and multiple convolutional layers make up the CNN component. The CNN component may primarily record the periodicity of data related to electricity usage. This approach combines the advantages of the BM algorithm and the CNN component to provide effective ETD. To the best of our knowledge, this is the first study to propose and implement a deep algorithm model to examine electricity theft in smart networks (mixing CNN with the BM algorithm). Additionally, we have conducted extensive tests on a sizable accurate dataset of electricity consumption.

III.EXISTING SYSTEM

Since the nation's economy, public security, and safety heavily rely on the energy grids, privacy and security concerns have recently been the subject of in-depth investigation. Unfortunately, concerns about privacy are not often fully appreciated in SG metering, thus more needs to be done to address the dangers of electricity theft. The topics of privacy and security in the realm of SGs have been covered in some excellent survey studies. They outlined the general difficulties in cybersecurity, including trust models, connectivity, security management, customer privacy, software vulnerabilities, and human aspects.These problems also have potential remedies, according to. In particular for the transmission subsystem inside the SG, Deng and Shukla examined the vulnerabilities and mitigation strategies . They concentrated on the areas where the phasor measurement units (PMUs) and wide area measuring system (WAMS) technologies were poor. Wang and Lu looked at security issues in the SG grid, which includes household area networks (HANs), advanced metering infrastructures (AMIs), and distribution and transmission subsystems. With the help of matter studies, they demonstrated the need for security and calculated network worries. A study on smart homes and safety was presented by Komninos et al. These authors typically considered that communication between the SG and smart home environments is classified as posing safety risks.

IV. PROPOSED SYSTEM

Here, we provide a system for detecting electricity theft. A statistical analysis of the electrical energy consumption statistics of both energy thieves and regular users reveals that energy thieves often utilize less electricity than regular consumers, either rarely or frequently. This monitoring makes it easier to categorize irregular electricity use and irregularity in electricity usage.

However, it is difficult to analyse the periodicity of the data on power consumption for a variety of reasons, including:

1) The data on electricity consumption is a 1-D time series and is very large, making it difficult to research the periodicity.
2) The data on electricity consumption is frequently inaccurate and noisy.
3) Due to the complexity of the calculations and the limited capacity for simplification, several traditional methods of data analysis, such as ANN and support vector machine (SVM), cannot be applied directly to the data on electricity consumption. Thus, the CNN approach has been used in this work to address the aforementioned difficulties.

V. METHODOLOGY

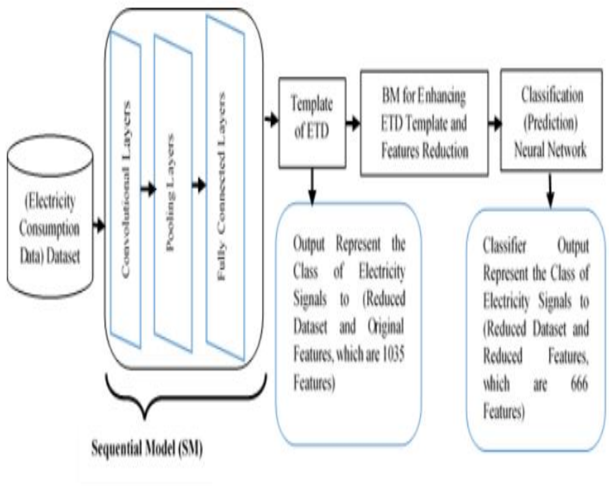


Fig.1: System Architecture

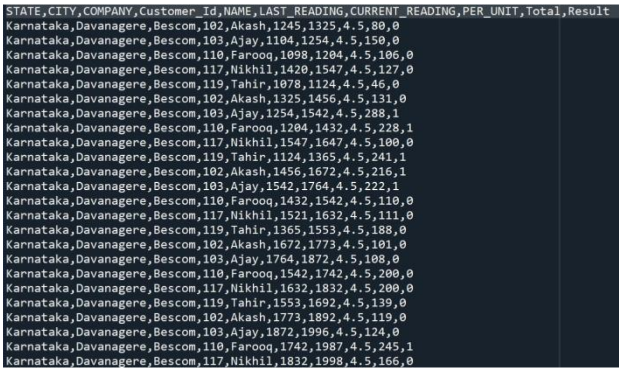
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

A realistic electricity consumption dataset released by State Grid Corporation of China is used to train the models. This work is intended to identify electricity theft from the power consumption pattern of users, utilizing CNN-based deep learning and BM techniques. This classifier model is trained to utilize a dataset consisting of daily power consumption data of both normal and fraudulent users in a supervised manner. First, the data is prepared by a data-preprocessing algorithm to train the model. The preprocessing step also involves synthetic data generation for better performance. At the next stage, the proposed model is hyper-tuned and finally, the optimized model is evaluated via the test data. The overall system is depicted in Figure 1

KNN (K-Nearest Neighbour): KNN is a non-parametric method that may be applied to applications involving regression or classification. It is a lazy learning algorithm, therefore it memorises the training dataset rather than going through a particular training phase.

Decision Tree: A supervised learning method called a decision tree is employed for classification and regression problems. It builds a model of choices and potential outcomes that resembles a tree. The method creates a tree-like structure by breaking the data set up into smaller subgroups depending on the qualities.

VI. RESULTS



```
STATE,CITY,COMPANY,Customer_Id,NAME,LAST_READING,CURRENT_READING,PER_UNIT,Total,Result
Karnataka,Davanagere,Bescom,102,Akash,1245,1325,4.5,80,0
Karnataka,Davanagere,Bescom,103,Ajay,1104,1254,4.5,150,0
Karnataka,Davanagere,Bescom,110,Farooq,1098,1204,4.5,106,0
Karnataka,Davanagere,Bescom,117,Nikhil,1420,1547,4.5,127,0
Karnataka,Davanagere,Bescom,119,Tahir,1078,1124,4.5,46,0
Karnataka,Davanagere,Bescom,102,Akash,1325,1456,4.5,131,0
Karnataka,Davanagere,Bescom,103,Ajay,1254,1542,4.5,288,1
Karnataka,Davanagere,Bescom,110,Farooq,1204,1432,4.5,228,1
Karnataka,Davanagere,Bescom,117,Nikhil,1547,1647,4.5,100,0
Karnataka,Davanagere,Bescom,119,Tahir,1124,1365,4.5,241,1
Karnataka,Davanagere,Bescom,102,Akash,1456,1672,4.5,216,1
Karnataka,Davanagere,Bescom,103,Ajay,1542,1764,4.5,222,1
Karnataka,Davanagere,Bescom,110,Farooq,1432,1542,4.5,110,0
Karnataka,Davanagere,Bescom,117,Nikhil,1521,1632,4.5,111,0
Karnataka,Davanagere,Bescom,119,Tahir,1365,1553,4.5,188,0
Karnataka,Davanagere,Bescom,102,Akash,1672,1773,4.5,101,0
Karnataka,Davanagere,Bescom,103,Ajay,1764,1872,4.5,108,0
Karnataka,Davanagere,Bescom,110,Farooq,1542,1742,4.5,200,0
Karnataka,Davanagere,Bescom,117,Nikhil,1632,1832,4.5,200,0
Karnataka,Davanagere,Bescom,119,Tahir,1553,1692,4.5,139,0
Karnataka,Davanagere,Bescom,102,Akash,1773,1892,4.5,119,0
Karnataka,Davanagere,Bescom,103,Ajay,1872,1996,4.5,124,0
Karnataka,Davanagere,Bescom,110,Farooq,1742,1987,4.5,245,1
Karnataka,Davanagere,Bescom,117,Nikhil,1832,1998,4.5,166,0
```

Fig. 2: The Dataset of the project.

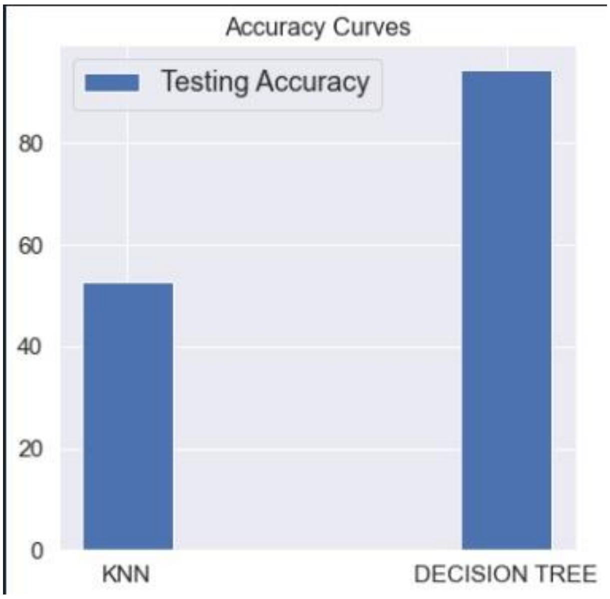
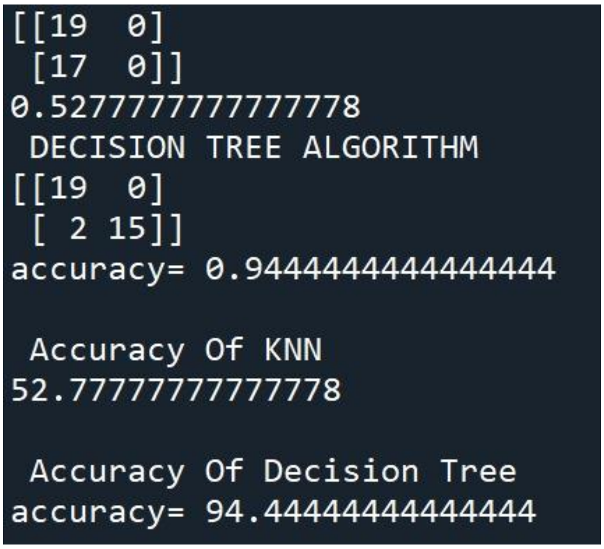


Fig.3: The Accuracy Curves



```
[[19  0]
 [17  0]]
0.5277777777777778
 DECISION TREE ALGORITHM
[[19  0]
 [ 2 15]]
accuracy= 0.9444444444444444

 Accuracy Of KNN
52.77777777777778

 Accuracy Of Decision Tree
accuracy= 94.44444444444444
```

Fig.4: The Accuracy in Decision tree and KNN Algorithm

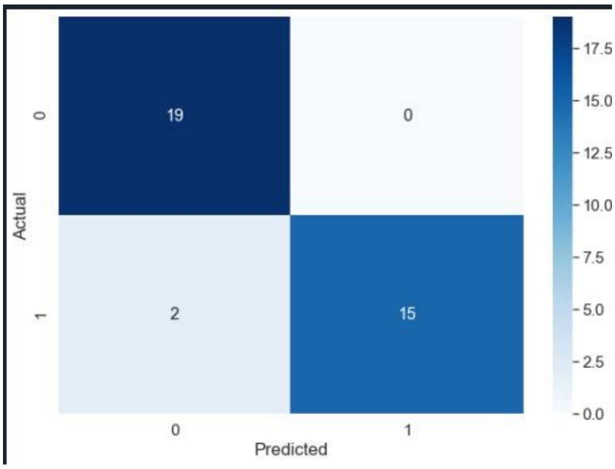


Fig.5: Conclusion Matrix

CONCLUSION

The most significant findings of this study are that supervised learning techniques outperform other techniques since they can train models with good performance using labelled data. Additionally, because pre-trained models are created using large datasets and sophisticated computers, they have a high power to address data on electricity consumption. When extracting data from a dataset using a standard CNN, accuracy may be lower than when utilising an SM to address data on electricity consumption. To improve the effectiveness of creating models and identifying fresh electricity signals, the dataset in this study was shrunk before the models were built. In fact, utilising an optimization technique (the BM algorithm) causes the extracted features to be reduced in order to improve the intended system's performance.

REFERENCE

[1] S. Tan, D. De, W. Song, J. Yang, S. K. Das, "Survey of Security Advances in Smart Grid: A Data Driven Approach," in IEEE Communications Surveys & Tutorials, vol. 19, no. 1, pp. 397-422, Firstquarter 2017, doi: 10.1109/COMST.2016.2616442.

[2] D. Alahakoon, X. Yu, "Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey," in IEEE Transactions on Industrial Informatics, vol. 12, no. 1, pp. 425-436, Feb. 2016, doi: 10.1109/TII.2015.2414355.

[3] M. Ehsani, Y. Gao, S. Longo, K. Ebrahimi, "Modern electric, hybrid electric, and fuel cell vehicles," CRC press,2018.

[4] M. H. Rehmani, M. Reisslein, A. Rachedi, M. Erol-Kantarci, M. Radenkovic, "Integrating Renewable Energy Resources Into the Smart Grid: Recent Developments in Information and Communication Technologies," in IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 2814-2825, July 2018, doi: 10.1109/TII.2018.2819169.

[5] J. Wu, S. Rangan, H. Zhang, "Green communications: theoretical fundamentals, algorithms, and applications," CRC press, 2016