ISSN: 2633-4828 Vol. 5 No.3, September, 2023

International Journal of Applied Engineering & Technology

INTRUSION DETECTION IN SMART ENERGY METER

Sheeba R, Christine Mariam Mammen and Shyba Zaheer
Department of Electrical and Electronics Engineering, TKM College of Engineering, Kerala, India
sheebaeee@tkmce.ac.in, christinepadam@gmail.com and s.shyba@gmail.com

ABSTRACT
Smart meters and the Internet of Things (IoT) have been increasingly used to replace conventional analog meters in today's modern smart home. It converts collected data of meter readings into digital format. The data can be delivered wirelessly, which decreases the amount of human labour required. smart meters, on the other hand, bring a slew of new ways to steal electricity. Using advanced tools or cyberattack techniques, malicious users can break into smart meters. Every year, this illegal conduct results in a significant financial loss. Energy theft detection techniques face a difficult task as a result of this. The advanced metering infrastructure (AMI) has the capability of monitoring each consumer's consumption details, tracking their patterns of consumption, billing them, and detecting variations. With the help of the smart grid's communication capabilities, utilities have been able to save their customers' usage details. This database can be used to develop a theft detection model. Artificial intelligence-based technologies are widely used in AMI, which deploys machine learning algorithms to detect prospective electricity thieves, frequently. The most common classification approaches involve utilizing labels to identify unusual trends in customers' previous electricity usage data and then detecting possible electricity theft behaviours. In this work, supervised learning techniques were used to detect electricity theft. To assess classification accuracy, a comparison of several machine learning classifiers such as Support Vector Machine, Nave Bayes, Decision Tree, and Random Forest, is also presented. Unsupervised learning techniques such as K-means and DBSCAN are also used to quickly spot irregularities in the readings.

Keywords-Advanced Metering Infrastructure (AMI), Support Vector Machine (SVM), Naïve Bayes, Decision Tree, Random Forest, K-means.

1. INTRODUCTION
In today's modern world, the power grid has been a necessity. Many countries have been updating their existing power systems into smart grids, which feature two-way communication, high stability, real-time demand feedback, self-healing, and security, owing to the advancements in information and communication technology. Advanced Metering Infrastructure (AMI) is a critical component of the smart grid and is closely associated with people's daily lives [1]. AMI automates the electric metering system by replacing the traditional meters with smart meters that allow utility companies and energy users to communicate in real time. AMI integrates smart meters and Internet of Things (IoT) controlling devices that may collect enormous amounts of data in a short amount of time. AMI's rich information interchange and multilevel semi-open network structure, on the other hand, expand the security flaws for metering across entire public networks and offer numerous cyber-attack vulnerabilities. Data that vary from normal and predicted patterns are referred to as anomalies in IoT [2].

The electric utility experiences substantial income losses as a result of the electricity thieves. Technical Losses (TL) and Non-Technical Losses (NTL) are the two types of electricity losses in transmission and distribution (NTL). TL is caused by power losses in overhead power wires, transformers, and other substation devices. Electricity theft is the most common kind of NTL. Electricity theft is described as the use of energy without the permission of a power company. Direct hooking, bypassing the electricity meter, energy bribery of unauthorized connections, interfering with the meter reading, and bypassing the energy meter are all examples [3]. It is responsible for considerable revenue losses as well as a reduction in electricity quality. According to a recent report, global power utility firms lose more than $20 billion per year. Both industrialized and developing countries are affected by the NTL. In Pakistan, for example, losses of 17.5 percent in energy transmission and distribution were observed in the years 2017–2018. Each year, India loses roughly $4.5 billion due to electricity theft. According to a recent assessment, unlawful electricity use accounts for 20% of total electricity consumption

in India. Rich countries are also affected by this issue. Illegal electricity consumption costs the United States roughly $6 billion per year, while power losses in the United Kingdom can cost up to $175 million per year. Furthermore, electricity theft might have an impact on the power system's functioning and reliability. It reduces the quality of power by overloading transformers and voltage fluctuations. Electric utilities are having difficulty providing electricity to their customers in a timely manner due to a rise in the number of electricity thieves. As a result, more intensive study is required in order to accurately discover electricity thefts and restore a significant income loss for utility companies [4]. The advanced metering infrastructure has the capability of monitoring each consumer's usage details, recording their consumption patterns, charging them, and detecting any irregularities. Utilities have been able to store their customers' consumption patterns because of the smart grid's communication capabilities. By analysing the collected data from smart meters, this database can be used to create a theft detection model using a machine learning algorithm. The detection of electricity theft using multiple machine-learning classifiers is discussed in this work. Unsupervised learning methods are also employed to cope with unlabelled data.

In the smart grid community, energy theft has become a big problem. Many countries have suffered significant losses in the billions of dollars. A smart meter is now installed at the end point of every distribution network to track energy consumption and create energy reports remotely. Hacking smart home appliances and, more commonly, straight hooking on other families' electrical supplies are two prominent means of energy theft [5]. Tampering with the smart meter's software and mechanism, as well as modifying data through cloud storage, are some of the other tactics used. As a result, attackers can lower their own electricity usage by manipulating other families' electricity usage through tampering and hacking, as long as the total cost for all consumers in the community remains the same. An example of an energy theft situation is a false data injection type attack. The higher-consumption home can minimize their own power usage by tapping into the power of another household through energy theft. It raises the electricity bills of the other family victim while lowering the expenses of the energy theft perpetrator. Researchers have recently concentrated their efforts on the development of improved theft detection systems based on artificial intelligence approaches. In [6], the new modern structure and security consideration in AMI network was discussed. Briefing basics of theft detection scheme using three categories, classification-based, state-estimation based and game theory-based. An architecture for theft detection in IoT data streaming is presented in [7][8] describes a deep learning-based system that utilizes Convolution Neural Network (CNN) Bad Data Detector (BDD) is employed to filter out low-quality data [9] proposes a two-stage theft-identifying system. At first, SVM is used to find out about the theft. Further stage attack is confirmed using a Temporal Failure Propagation Graph (TFPG). Long Short-Term Memory is employed in [10] to detect intrusion and also for predicting consumption patterns based on previous data. Whereas a combined application of CNN and LSTM architecture is suggested in [11]. In [12], theft detection using consumer usage patterns is presented and also addresses the data imbalance problem. The implementation of a Support Vector Machine (SVM) for identifying attacks or detecting theft is proposed in [13]. The Principal Component Analysis (PCA) is utilized in reducing the complexity of the process. To find out honest and dishonest consumers, deep recurrent vector embedding is employed in [14]. The Sequential grid search hyperparameter algorithm is used to improve the performance of the theft detection system. For the same purpose, the application of Extreme Gradient Boosting (XGBoost) in AMI is presented in [15]. By exploiting the relation between Non-Technical Loss and missing value patterns, it is possible to diagnose thefts. This methodology is implemented in [16], with the aid of CNN model to locate missing pattern [17] and present an approach that detects theft using the unsupervised learning method Firefly Algorithm based XGBoost. In this Visual Geometry Group (VGG-16) and normalization techniques were also applied to improve the performance of the proposed system. Comparison of SVM, Logistic, Regression (LR), and, CNN is also discussed. In [18], CatBoost Algorithm and SMOTETomek algorithm were employed in theft detection. The catBoost algorithm is a supervised method and utilizes feature engineering to choose the appropriate feature for the theft identification process. [19] proposes a hybrid CNN-RF (Random Forest) model for theft detection in power grids. All these methods are time-consuming and require, complex processing stages. This paper proposes a methodology utilizing supervised learning algorithms to perform theft detection. Various

models, such as SVM, Random Forest, Decision Tree, and Naïve Bayes, were analyzed. Unsupervised learning models, K-Means and DBSCAN, were also evaluated.

2. METHODOLOGY

In machine learning, computers impulsively understand and make smart decisions based on a given set of information. Supervised Learning, Unsupervised Learning, Semi-supervised learning and, Reinforced learning come under the machine learning process. Classification is a type of supervised learning in which supervision is accomplished by a set of labelled data. There are mainly two steps in the classification process. The first step is to generate a classifier model using a set of labelled data called the training set. In the second step, unlabelled data is given to the classifier for classification. In the case of unsupervised learning, data is not labelled. When a set of data is input into an unsupervised model, it will create clusters or groups based on similarities. In this work, the focus area is the utilization of supervised learning techniques for finding electricity theft by false data injection. Unsupervised learning methods were also analyzed. The flowchart of the proposed methodology is shown in Fig.1.

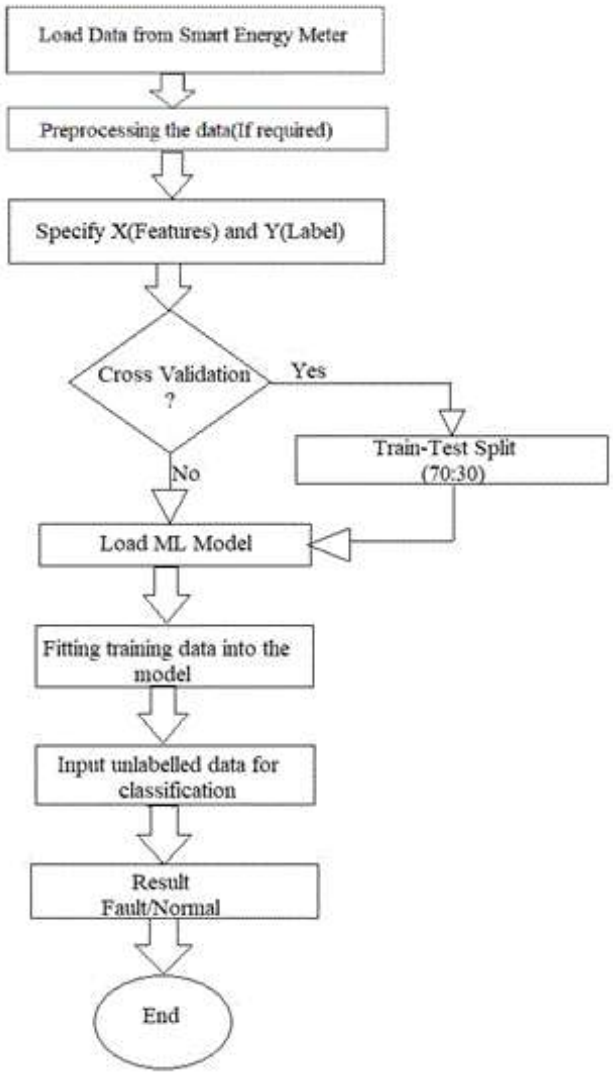

Fig. 1. Flowchart of proposed Methodology

A. DATA COLLECTION

Energy consumption data of domestic consumers for both normal and faulty conditions were collected from smart meters. A part of these datasets is given to the classifier model for training and the remaining is used to test the model for performance evaluation. The consumption details of 20000 consumers collected from a dataset of State Grid Corporation of China, were used. The dataset consists of time series data during 2016 -2017. Details include global active power which is the sum of power consumption other than submetering. Submetering is used to find out the consumption of warehouse-like areas. There were three submetering in the home area. The total consumption is calculated with these data. In the case of faulty consumers, there is a mismatch in the aggregate value and the sum of consumption from each area. whereas in normal cases, the aggregate value is equal to the sum of individual areas. Data from 10000 consumers is used for training the model. Among these 5000 is honest consumer and the remaining 5000 is faulty consumers. the remaining 10000 data are used for validating the model. In Fig.2, the consumption pattern when no theft is encountered is depicted. Whereas Figure 3 shows the consumption pattern during the theft. The theft is a fault data injection type attack. In which faulty consumption value is entered into the AMI system. Thereby consumption is made far away from the actual value and greedy attacker earns certain profit by reducing the electricity bill

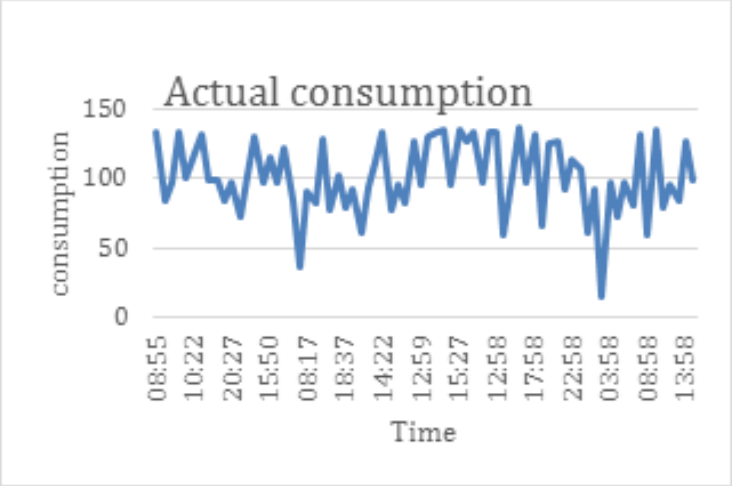

Fig.2. Actual Consumption

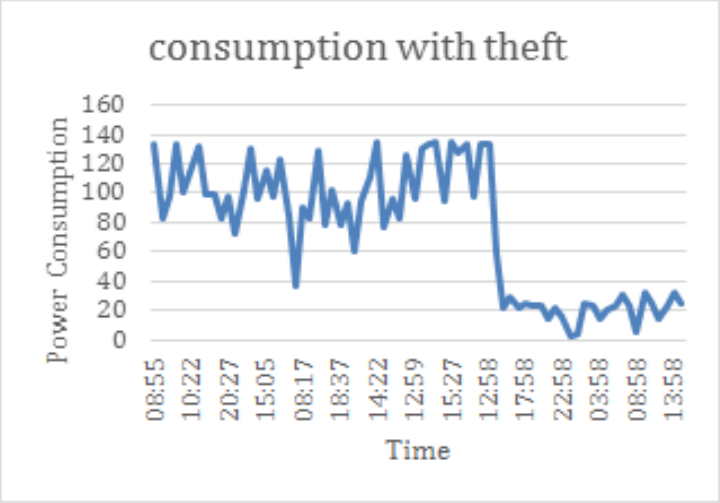

Fig. 3. Consumption with theft

Table- 1

Sl No	CLASSIFIER MODELS
1	Support Vector Machine (SVM)
2	Naïve Bayesian Classifier
3	Decision Tree Classifier
4	Random Forest
5	K-Means
6	DBSCAN

The list of supervised and unsupervised classifier models used in the proposed work is included in Table 1

B. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis or Karhunen-Loeve or K-L method is a dimensionality reduction technique. Assume, the input data vector consisting of 'n' feature is required to reduce to a data vector with 'k' feature such that $k \leq n$.Dimensionality reduction is achieved by projecting the actual data vector onto a smaller space. The PCA identifies the important aspects regarding the original data and creates an equivalent smaller set of features. The general procedure is given below,

- Normalize the input data. This avoids the dominance of the large domain features over the small domain feature. So that every attribute comes in the same range.
- PCA calculates k orthonormal vectors, unit vectors that are perpendicularly projected, called Principal Components.
- These principal components are arranged in descending order of strength.
- Data size is reduced by removing weaker components, ie, a component with lower variance. In Figure 4, Y1 and Y2 are the principal components of actual data mapped on X1 and X2 axes.

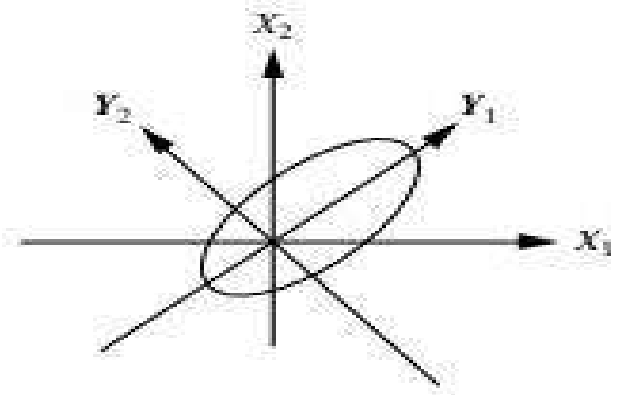

Fig. 4. Principal components

C. CROSS VALIDATION

In a model, overfitting or underfitting problems may arise as a result of using the same data set for training and testing. In the case of underfitting, it is difficult for the model to identify the differences in each class. It occurs when the model is too simple. Whereas in the case of overfitting the model is complex and can't generalize, when testing a model with the same data used for training, the model performs satisfiable. But, if a set of unseen data is given to the model, the accuracy is poor and fails to generalize the pattern which is called overfitting. Cross-validation techniques are employed to evaluate the performance of a model by using a part of the data for training and holding the remaining part, unseen data, for validation. The train-Test split method can be used if the dataset is large enough and the data are equally distributed. In this technique, the data is randomly selected for training

and testing. Common splitting ratios are 70:30 and 80:20. The accuracy is the same in both cases. Suppose the splitting ratio is 70:30,70% of the data is used for training the model and 30% is kept as unseen data for evaluating the performance of the model with new data.

Table 2: Cross-validation using Train-Test Split

X (FEATURE)	Y (LABEL)
X TRAIN	Y TRAIN
X TEST	Y TEST

3. RESULT AND ANALYSIS

Model formulation and evaluation are carried out using Python programming. At first unsupervised models, K-Means, DBSCAN, were assessed. These models created clusters based on similarities of the input data. The clusters formed in both clustering techniques are shown in Fig 5.

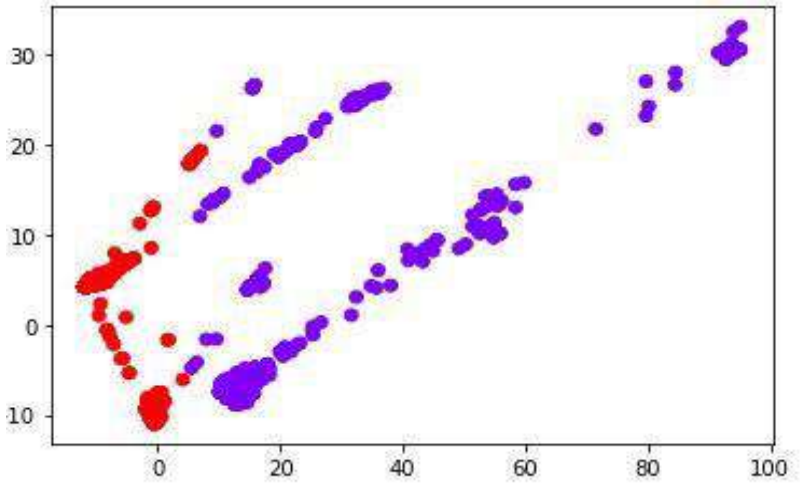

Fig. 5. Clusters in K-Means

Fig. 6. Clusters in DBSCAN

In Fig 6, the K-Means classifier creates two clusters from the given data, the red colour indicates Fault value while blue indicates normal data. Whereas in Fig 6, DBSCAN developed three clusters which will lead to misclassification. The original data has two class only, DBSCAN distribute data in three groups, which implies clustering using DBSCAN is not suitable for this scenario. In clustering methods, if new data is given to the model, it is difficult to identify the new data in the visual representation, i.e., labelling of unseen data is not possible. This can be used only for identifying the possibility of theft. In the case of the supervised learning technique, labelling of new unseen data is possible and beneficial for future training of the model. The performance evaluation of a classifier model is an important aspect. The matrix evaluation method or confusion matrix was most widely used to compute the accuracy of classification. Suppose a number of the class is m, such that,$m \geq 2$, then the confusion matrix will be $\times m$ matrix. In the ideal case, all the tuples are distributed in diagonal elements of the matrix. If the classifier predicts all the positive cases as positive, then it is a True Positive (TP) case. Otherwise, if the classifier predicts a negative value as positive, then it is False Positive (FP). The classification of actual negative cases as negative implies True Negative (TN). If not, i.e., the model groups positive data as negative, it is False Negative (FN). In the confusion matrix, a number of predictions belonging to each case (TP, TN, FP, FN) are given by the matrix elements. Accuracy can be computed using the equation(i),

$$Accuracy = \frac{TP+TN}{P+N} \quad \text{----- (1)}$$

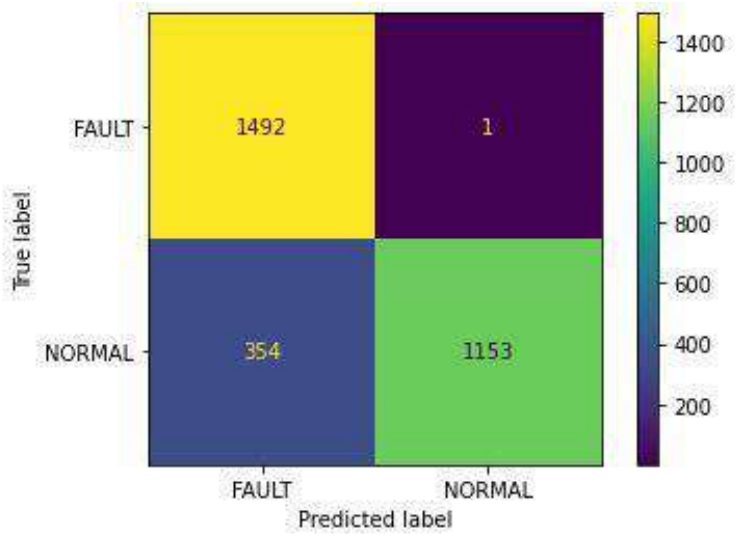

Fig. 7. Confusion Matrix of SVM (RBF kernel)

FP and FN are prediction errors since they refer to the number of misclassifications. The evaluation of each model with a confusion matrix is presented. For training the model 10,000 data were utilized. By employing the Train-Test Split method,7000 data were used for training, and 3000 is used for validation. Figure 7 shows the confusion matrix of SVM with RBF kernel.

Here C_{11}=1492, which is the number of true fault (TN) class members.C_{12}=1 represents False Normal (FP),C_{21}=354 refers to the False Fault (FN), and C_{22}=1153, is the True Normal (TP) value. In the false identification scenario is the number of actual normal data, and TN refers to the number of faulty values. Whereas FN and FP indicate the number of misclassifications of normal data as a fault and fault value as a normal value respectively. In the SVM classifier error from FP is higher than FN and testing accuracy is 88.167%. The decision boundary generated using different kernels, Linear, and Sigmoid is depicted in Fig8. To plot these boundaries, PCA is performed. As a result, the number of features is reduced to two equivalent features from the original five features. The SVM model with RBF kernel gives comparatively accurate classification.

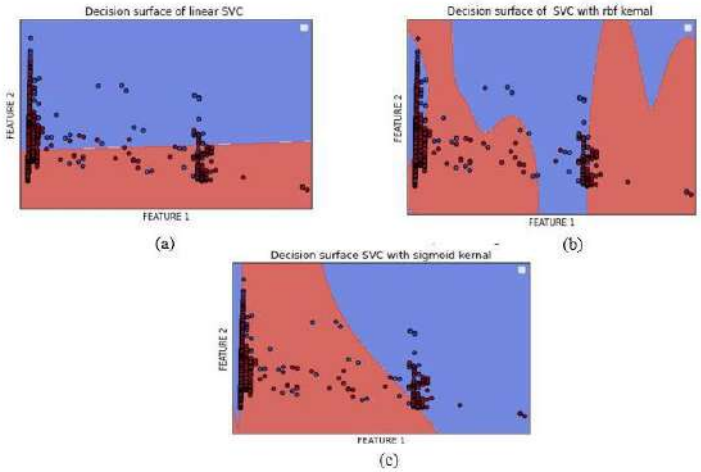

Fig. 8. Decision surface of SVM (a) Linear Kernel,(b) RBF Kernel,(c)Sigmoid Kernel

In Naïve Bayes Classifier Model (Fig9), uses three types, Gaussian, Multinomial, and Bernoulli models. For the Gaussian naïve Bayes classifier, True Normal (TP)=790 and True Fault (TN)=1407 and both errors are present in the model (Fales Normal (FP)=56, False Fault (FN)=747). The testing Accuracy of the model is 73.2%.

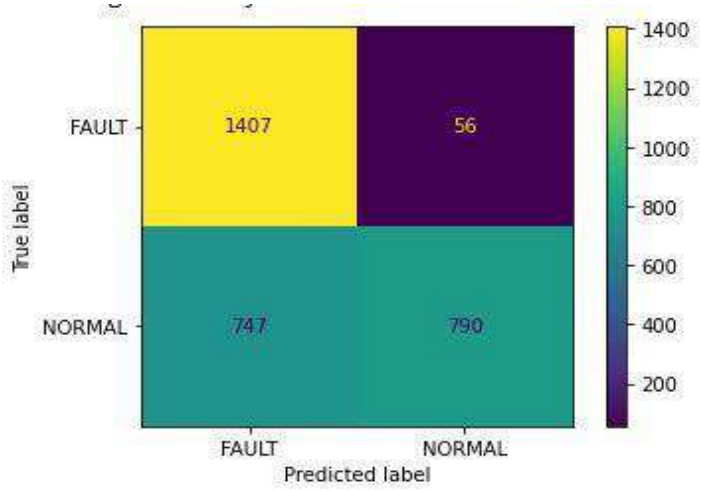

Fig. 9. Confusion Matrix of Gaussian NB

For Multinominal NB(Fig10), TP or True Normal=1446 and True Fault or TN=1370.Error due to wrong classification of normal value as faulty value is zero, i.e., False Fault (FN)=0, while the type of error due to False positive or False Normal is present, the number of such classification is equal to184. An accuracy of 93.86% is obtained for this model.

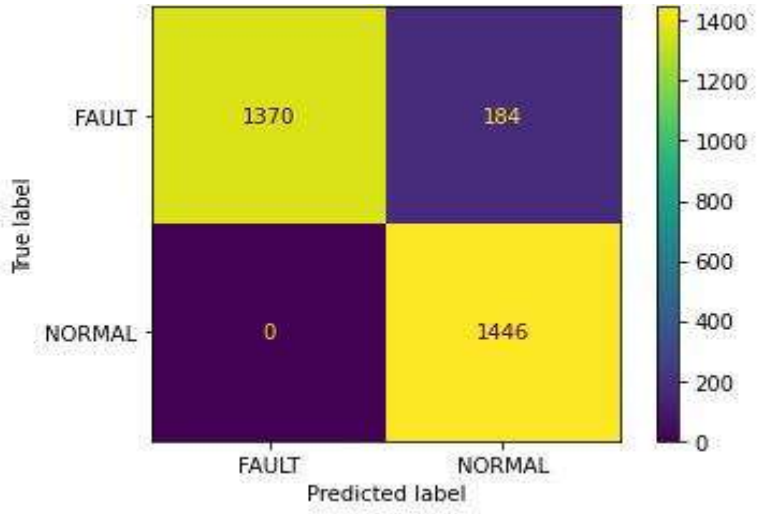

Fig. 10. Confusion Matrix of Multinominal NB

In the case of Bernoulli NB(Fig11), the accuracy of classification is lesser compared to other models. Both errors incurred in this model, FP=265 and FN=955. Testing accuracy is only 59.33%. True predictions TF and TP are 1224 and 556 respectively.

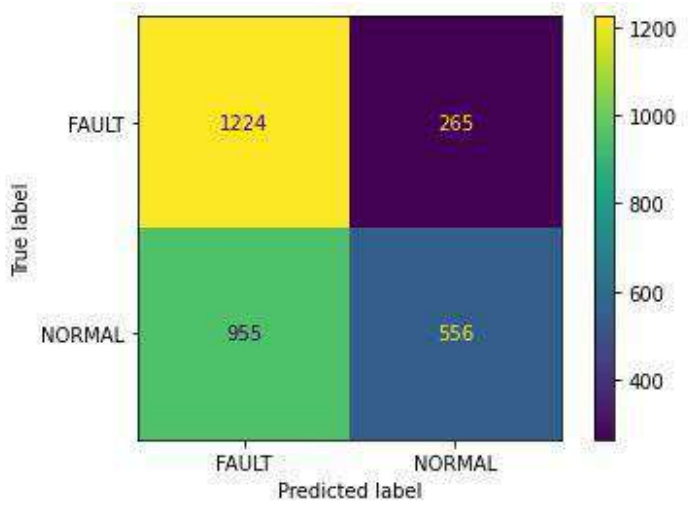

Figure. 11. Confusion Matrix of Bernoulli NB

The decision tree classifier (Fig12) is the best model for fault or theft detection. It gives an accuracy of 99.76%. Errors due to miss classification are much lower when compared with other models (FP=3 and FN =4). True fault and True Normal values are 1519 and 1474 respectively.

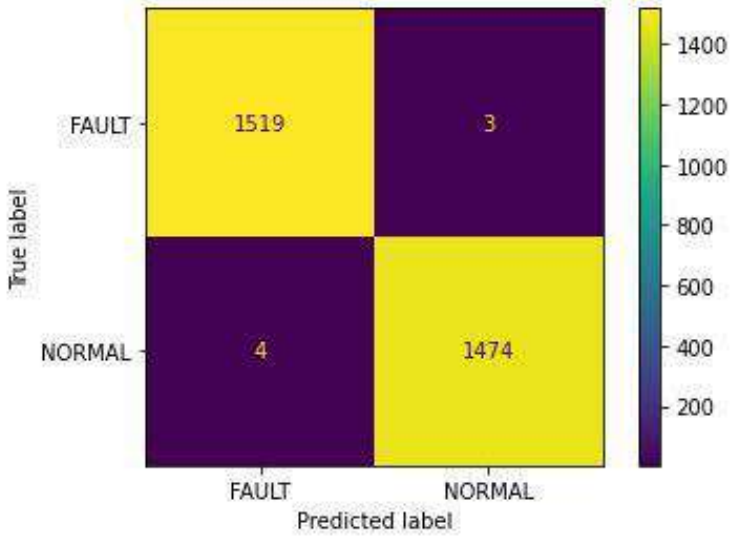

Fig.12. Confusion Matrix of Decision Tree

For Random Forest Classifier (Fig13), number of correct fault predictions is equal to 1363 and normal case classification is 961.Missclassification of the fault and normal cases are 459 and 157 respectively. The classification accuracy of the model is 78.13%.

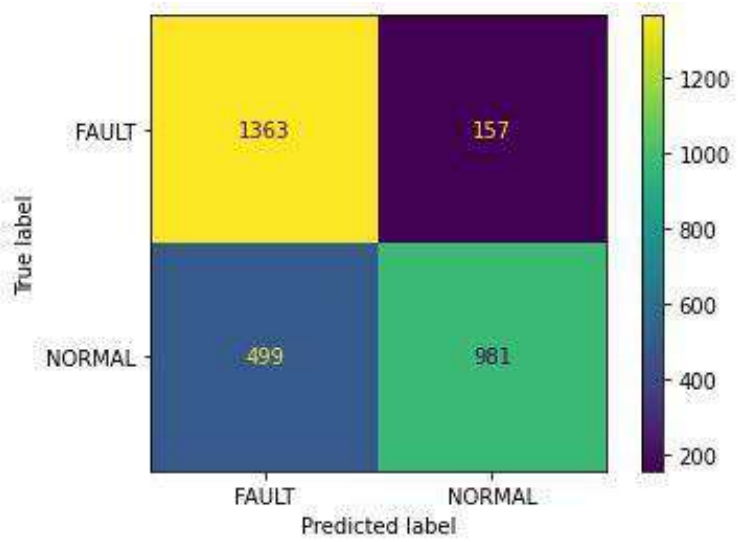

Fig.13. Confusion Matrix of Random Forest

The performance evaluation is carried out in different cases. Initially, accuracy finds out without performing cross-validation. In this case, trained classifiers using all 10,000 data. Validation is done using another set of 10,000 values, which will be unknown data for the model. An overfitting problem is detected in this stage, i.e., the training accuracy is higher while validation accuracy is reduced considerably. Accuracy comparison of various models in this situation is depicted in Fig4.

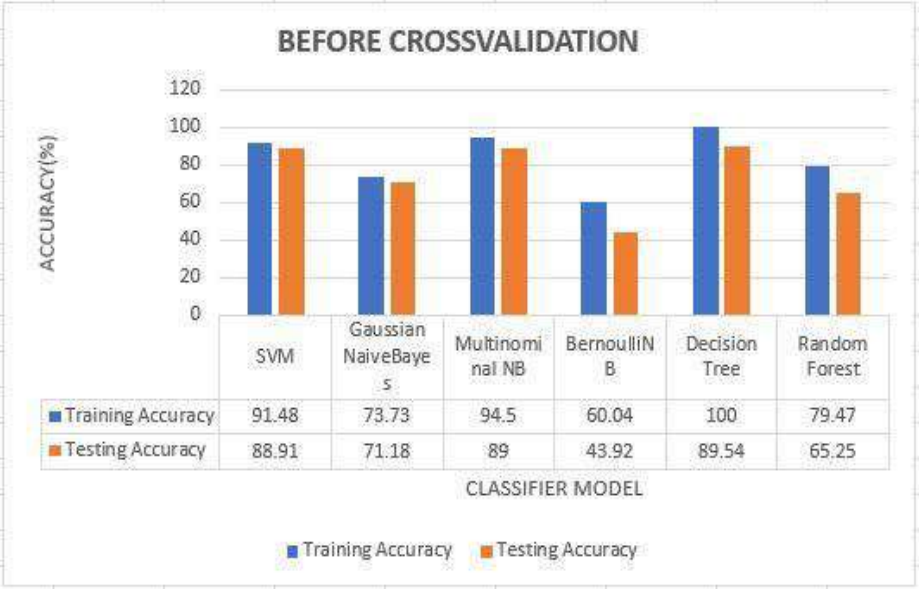

Fig.14. Accuracy Comparison of models before cross-validation

To overcome these overfitting problems, utilized the Train-Test Split method. The total training data is divided in the ratio 70:30., ie,70% is used for training, and 30%is used for testing. By this, both training and testing accuracy come in the same range. The accuracy of different models is shown in Fig15.

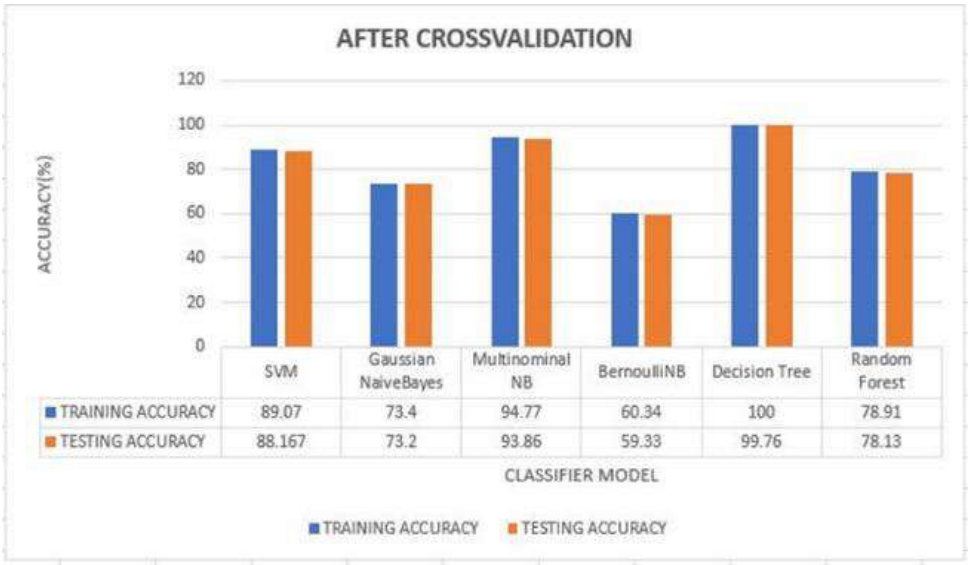

Fig.15. Accuracy Comparison of models after cross-validation

The performance of various classifiers with cross-validation was also analyzed with a set of new 10,000 unknown data. In all the cases Decision Tree classifier has higher accuracy. The Multinominal Naive Bayes classifier is better than SVM classifier with an accuracy of 94.77%. Among all these classifiers Bernoulli NB offers the lowest accuracy (59.33%).In Fi16, testing the accuracy of models with a new set of unlabelled data is plotted. The validation accuracy of the decision tree model is 90.7% and that of Bernoulli is only 49.46%.

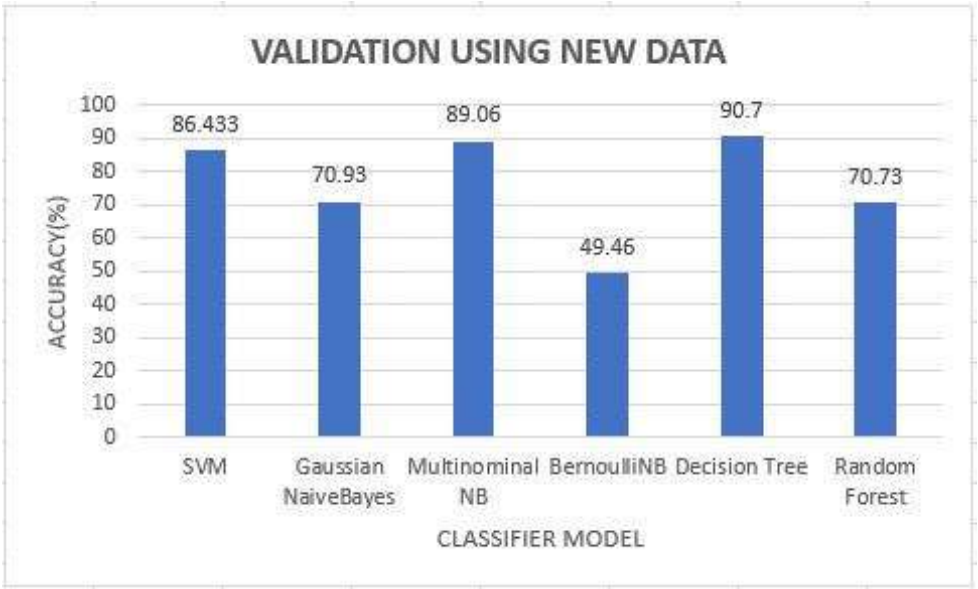

Fig.16. Accuracy Comparison of models with the unknown dataset

4. CONCLUSION AND FUTURE SCOPE

One of the major challenges faced by the energy sector is the development of an efficient energy theft detection system. Nowadays more research is going on in this area. Currently, available methods are becoming inapt in the era of advanced technologies. With the aid of changes in the AMI structure and the development of adequate tools intruders can easily attack smart metering systems. The emerging Artificial Intelligent techniques provide effective solutions to this problem. In this paper, a Machine learning-based theft detection strategy is proposed,

and evaluated the efficiency of these models is. Unsupervised learning-Means and DBSCAN is used. The performance of K-Means clustering is comparatively higher than DBSCAN. But using unsupervised models, unable to identify the state of the new unseen data through visual realization. With the help of supervised models, labeling of new unseen data as faulty or normal is possible. SVM, Gaussian Naïve Bayes, Multinominal Naïve Bayes, Bernoulli Naïve Bayes, Decision Tree, and Random Forest are the classifiers used in this work. Among these classifiers, the Decision tree classifier achieved higher validation accuracy of 99.67% with cross-validation and Bernoulli Naïve Bayes has the least accuracy (59.33%). The train-Test split method is used to overcome the overfitting problem. This work deals with only false data injection-type attacks. In which intruders perform by manipulating actual consumption data y various means. In the future, we can improve the model by identifying various types of attacks. For this, more data under different types of theft should be collected from the real smart metering system both in normal and theft cases. However, the effectiveness of the use of supervised learning techniques in theft detection systems is verified with available data from smart energy meters.

REFERENCES

[1] S. Sahoo, D. Nikovski, T. Muso and K. Tsuru, "Electricity theft detection using smart meter data", *in Proceedings of IEEE Power & Energy Society Innovative Smart Grid Technologies Conference* , pp. 1-5,2015.

[2] H. Zubi and A. Alrmaih, "Smart Energy Meter System Design & Simulation Presenting Electricity Theft Methods, Detection and Protection*", in Proceedings of IEEE 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA*, pp. 533-538, 2011.

[3] C. Richardson, N. Race, and P. Smith, "A privacy-preserving approach to energy theft detection in smart grids",*in Proceedings of IEEE International Smart Cities Conference*, pp. 1-4, 2016.

[4] Jiang, Rong & Lu, Rongxing & Wang, Ye & Luo, Jun & Shen, Changxiang & Shen, Xuemin. "Energy-Theft Detection Issues for Advanced Metering Infrastructure in Smart Grid", *Tsinghua Science & Technology*, 19. Pp 105-120 ,2014.

[5] W. Li, T. Logenthiran, V. Phan and W. L. Woo, "A Novel Smart Energy Theft System (SETS) for IoT-Based Smart Home," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5531-5539.

[6] A. Yahyaoui, H. Lakhdhar, T. Abdellatif, and R. Attia, "Machine learning based network intrusion detection for data streaming IoT applications," " in Proceedings of IEEE 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), pp.51-56, 2021.

[7] R. Jiang, R. Lu, C. Lai, J. Luo, and X. Shen, Robust group key management with revocation and collusion resistance for scada in smart grid, in Proc. IEEE Globe Communication Conference (Globecom), 2013, pp. 824- 829.

[8] S. Wang, S. Bi and Y. -J. A. Zhang, "Locational Detection of the False Data Injection Attack in a Smart Grid: A Multilabel Classification Approach," in IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8218-8227, Sept. 2020, doi: 10.1109/JIOT.2020.2983911.

[9] C. -C. Sun, D. J. Sebastian Cardenas, A. Hahn and C. -C. Liu, "Intrusion Detection for Cybersecurity of Smart Meters," *in IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 612-622.

[10] X. Wang, T. Zhao, H. Liu, and R. He, "Power Consumption Predicting and Anomaly Detection Based on Long Short-Term Memory Neural Network," ", in Proceedings of IEEE *4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA),* pp. 487-491, 2019.

[11] R. U. Madhure, R. Raman and S. K. Singh, "CNN-LSTM based Electricity Theft Detector in Advanced Metering Infrastructure," in Proceedings of IEEE *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT),* pp. 1-6, 2020

[12] D. Syed, H. Abu-Rub, S. S. Refaat, and L. Xie, "Detection of Energy Theft in Smart Grids using Electricity Consumption Patterns," *2020 IEEE International Conference on Big Data (Big Data),* pp. 4059-4064, 2020

[13] R. N. Toma, M. N. Hasan, A. -A. Nahid and B. Li, "Electricity Theft Detection to Reduce Non-Technical Loss using Support Vector Machine in Smart Grid*," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pp. 1-6, 2019

[14] A. Takiddin, M. Ismail, M. Nabil, M. M. E. A. Mahmoud, and E. Serpedin, "Detecting Electricity Theft Cyber-Attacks in AMI Networks Using Deep Vector Embeddings," *in IEEE Systems Journal*, vol. 15, no. 3, pp. 4189-4198, Sept. 2021.

[15] Z. Yan and H. Wen, "Electricity Theft Detection Base on Extreme Gradient Boosting in AMI," in *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-9, 2021,

[16] J. Yang et al., "Non-technical Loss Detection using Missing Values' Pattern," 2020 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE), 2020, pp. 149-154, doi: 10.1109/ICSGCE49177.2020.9275601.

[17] Khan, Zahoor A., Muhammad Adil, Nadeem Javaid, Malik N. Saqib, Muhammad Shafiq, and Jin-Ghoo Choi. 2020. "Electricity Theft Detection Using Supervised Learning Techniques on Smart Meter Data" *Sustainability* 12, no 19: 8023.

[18] Saddam Hussain, Mohd. Wazir Mustafa, Touqeer A. Jumani, Shadi Khan Baloch, Hammad Alotaibi, Ilyas Khan, Afrasyab Khan, A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection, Energy Reports, Volume 7, Pages 4425-4436, 2021.

[19] Shuan Li, Yinghua Han, Xu Yao, Song Yingchen, Jinkuan Wang, Qiang Zhao, "Electricity Theft Detection in Power Grids with Deep Learning and Random Forests", *Journal of Electrical and Computer Engineering*, vol. 2019, Article ID 4136874, 12 pages, 2019.

[20] Han, J.W., Kamber, M. and Pei, J. (2012) *Data Mining Concepts and Techniques.* 3rd Edition, Morgan Kaufmann Publishers, Waltham.

[21] S. O. Tehrani, M. H. Y. Moghaddam, and M. Asadi, "Decision Tree based Electricity Theft Detection in Smart Grid," *in Proceedings of IEEE 2020 4th International Conference on Smart City, Internet of Things and Applications (SCIOT),* pp. 46-51, 2020.