International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 04 | Apr 2026 www.irjet.net p-ISSN: 2395-0072

Electricity Theft Detection Using Machine Learning: A Comparative Analysis of SVM and KNN

Chinta Naga Pavithra[1], Macharla Naga Mahesh[2], Kagita Adarsh[3], Ambati Pavan Kalyan[4], Kakarla Sashi Venkat[5], Polisetti Karimulla[6]

[1,2,3,4,5]*Undergraduate Students, Department of Electrical & Electronics Engineering, Bapatla Engineering College, Bapatla, Andhra Pradesh, India*

[6]*Assistant Professor, Department of Electrical and Electronics Engineering, Bapatla Engineering College, Bapatla, Andhra Pradesh, India*

Abstract - *Electricity theft is a major issue faced by power distribution companies, leading to significant financial losses and reduced operational efficiency. Conventional detection methods, such as manual inspections and rule-based approaches, are often time-consuming, costly, and less accurate. To overcome these challenges, this paper presents a machine learning-based electricity theft detection system using both Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) classifiers. The proposed system utilizes smart meter consumption data to identify abnormal electricity usage patterns effectively. Initially, the collected data is preprocessed to eliminate noise, handle missing values, and ensure data consistency. Subsequently, feature extraction is performed to identify key parameters such as energy consumption, time-based usage patterns, and voltage variations. These features are then used to train and evaluate both SVM and KNN models for classification. The system categorizes consumers into two classes: normal users and electricity theft cases. The performance of the models is evaluated using standard metrics such as accuracy and confusion matrix to ensure reliable results. Experimental results demonstrate that both models are effective in detecting electricity theft; however, the SVM model achieves higher accuracy of 99% compared to the KNN model. This indicates that SVM provides better classification performance for the given dataset. The proposed approach improves detection accuracy and helps reduce non-technical losses in power distribution networks, offering an efficient and reliable solution for modern smart grid systems.*

Key Words: **Electricity Theft Detection, Machine Learning, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Smart Meter Data, Classification.**

1. INTRODUCTION

Electricity theft is a serious issue in many countries, particularly in developing regions, where it contributes significantly to non-technical losses in power systems. It can occur through various methods such as illegal connections, meter tampering, bypassing meters, and manipulation of billing information. These activities not only result in major financial losses for power distribution companies but also affect power quality, system reliability, and the stability of the electrical network.

Conventional electricity theft detection methods mainly rely on manual inspections and physical meter readings. These approaches require significant manpower, involve high operational costs, and are often inefficient in detecting complex and hidden theft patterns. Therefore, there is a need for more advanced, automated, and accurate detection techniques.

With the advancement of smart grid technologies and the deployment of smart meters, a large amount of electricity consumption data is now available. Machine learning techniques can analyze this data to identify unusual patterns and detect electricity theft effectively. These methods improve detection accuracy while reducing human effort and operational costs.

In this project, two machine learning algorithms, Support Vector Machine (SVM) and K-Nearest Neighbors (KNN), are used for electricity theft detection. KNN is a simple and easy-to-implement algorithm that classifies data based on similarity with neighboring data points. However, it requires high computational time and may be less effective for large datasets. On the other hand, SVM is a powerful supervised learning algorithm that can handle high-dimensional data and provides better classification accuracy by finding an optimal decision boundary.

The experimental results show that both algorithms are capable of detecting electricity theft, but the SVM model achieves higher accuracy compared to the KNN model. Therefore, SVM is considered more suitable for this application, as it provides efficient and reliable classification of electricity consumption data into normal and theft categories.

2. LITERATURE REVIEW

Electricity theft detection has gained significant attention in recent years due to its impact on power distribution systems and the increasing demand for reliable energy management. Traditional methods such as manual inspections and rule-based techniques have been widely used; however, these

approaches are time-consuming, labor-intensive, and often fail to detect complex and hidden theft patterns. As a result, there is a growing need for automated and intelligent detection systems

With the development of smart grid technologies and the deployment of smart meters, large volumes of electricity consumption data are now available. Researchers have explored the use of machine learning techniques to analyze this data and identify abnormal consumption patterns. Various algorithms such as Decision Trees, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) have been applied for electricity theft detection. These techniques provide improved accuracy and efficiency compared to conventional methods.

Among these methods, the K-Nearest Neighbors (KNN) algorithm is widely used due to its simplicity and ease of implementation. It classifies data based on similarity with neighboring data points. However, KNN requires higher computational time, and its performance is highly dependent on the selection of the 'k' value. In contrast, the Support Vector Machine (SVM) algorithm is a powerful supervised learning method capable of handling high-dimensional data and providing better classification accuracy by determining an optimal decision boundary.

Several studies have reported that SVM-based models outperform other machine learning techniques in detecting electricity theft with higher accuracy and reliability. These models are effective in identifying complex consumption patterns and minimizing misclassification errors.

Based on the analysis of existing literature, it is observed that machine learning-based approaches provide an efficient solution for electricity theft detection. Therefore, this project focuses on implementing and comparing SVM and KNN algorithms, where SVM demonstrates superior performance in terms of accuracy and efficiency.

3. METHODOLOGY OF SVM ALGORITHM

The proposed electricity theft detection system follows a structured process consisting of multiple stages, as illustrated in Fig. 1. Each stage plays a significant role in accurately identifying abnormal electricity consumption patterns.

3.1 Electricity Consumption Data

The system begins with the collection of electricity consumption data from smart meters. This data includes parameters such as energy usage (kWh), time-based consumption patterns, and voltage levels. The collected data serves as the primary input for further processing and analysis.

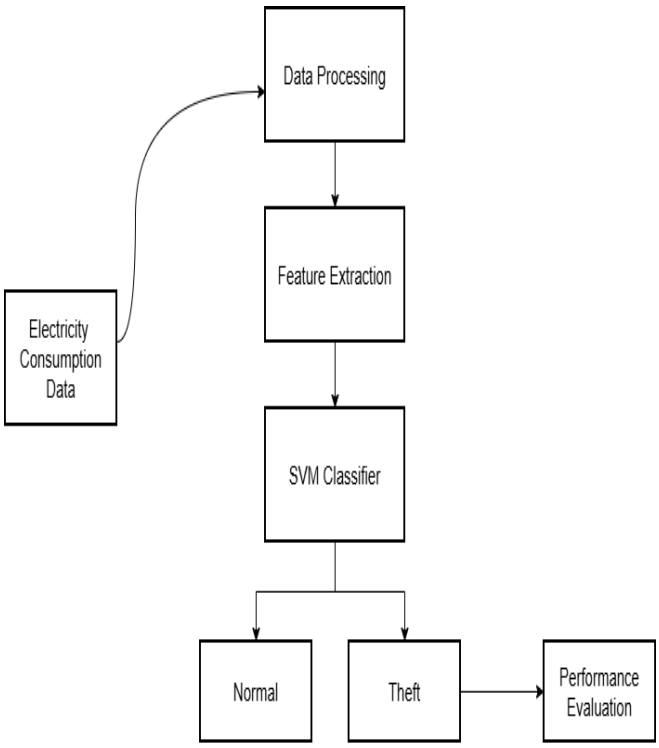

Fig -1: Block Diagram of Electricity Theft Detection Using SVM Classifier

3.1.2 Data Processing

The raw data may contain noise, missing values, and inconsistencies. Therefore, it is processed in this stage using preprocessing techniques such as data cleaning, normalization, and handling of missing values. This step ensures that the dataset is accurate, consistent, and suitable for further analysis.

3.1.3 Feature Extraction

In this stage, important features are extracted from the processed data. These features include daily energy consumption, peak usage time, load variations, and voltage fluctuations. Feature extraction helps reduce data dimensionality and improves the efficiency and accuracy of the classification model.

3.1.4 SVM Classifier

The extracted features are fed into the Support Vector Machine (SVM) classifier. SVM is a supervised machine learning algorithm that determines an optimal decision boundary (hyperplane) to separate different classes. The model is trained using labeled data to classify electricity consumption patterns as either normal or theft.

3.1.5 Normal and Theft Classification

Based on the trained SVM model, the system classifies the input data into two categories: normal consumption and electricity theft. This classification helps in identifying users involved in fraudulent electricity usage.

3.1.6 Performance Evaluation

The performance of the system is evaluated using standard metrics such as accuracy and confusion matrix. These metrics help in analyzing the effectiveness of the model in detecting electricity theft. The proposed system achieves high accuracy, indicating reliable and efficient performance.

3.2 METHODOLOGY OF KNN ALGORITHM

The proposed electricity theft detection system also utilizes the K-Nearest Neighbors (KNN) algorithm for classification. The overall system architecture is illustrated in Fig. 2. The system processes electricity consumption data through multiple stages to classify users as either normal or theft.

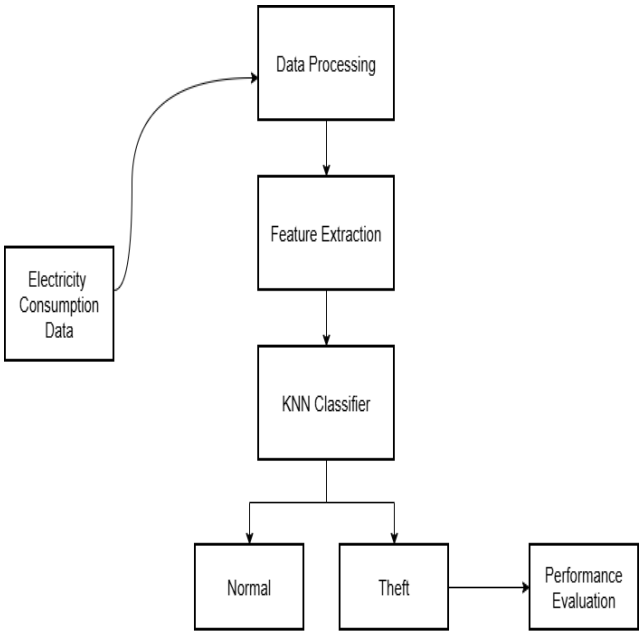

Fig -2: Block Diagram of Electricity Theft Detection Using KNN Classifier

3.2.1 Electricity Consumption Data

The system begins with the collection of electricity consumption data from smart meters. This data includes parameters such as energy usage (kWh), time-based consumption patterns, and voltage levels. The collected data serves as the primary input for further analysis.

3.2.2 Data Processing

The collected raw data may contain noise, missing values, and inconsistencies. Therefore, preprocessing techniques such as data cleaning, normalization, and handling of missing values are applied. This step ensures improved data quality and reliability for subsequent processing.

3.2.3 Feature Extraction

In this stage, relevant features are extracted from the processed data. These features include daily consumption patterns, peak usage time, load variations, and voltage fluctuations. Feature extraction reduces data complexity and enhances the overall performance of the model.

3.2.4 KNN Classifier

The extracted features are provided to the K-Nearest Neighbors (KNN) classifier. KNN is a supervised machine learning algorithm that classifies data based on similarity. It identifies the 'k' nearest data points (neighbors) from the training dataset and assigns the class based on majority voting. The selection of 'k' plays a crucial role in determining the accuracy of the model.

3.2.5 Normal and Theft Classification

Based on the KNN algorithm, the system classifies the input data into two categories: normal consumption and electricity theft. This classification is performed by comparing the input data with its nearest neighbors.

3.2.6 Performance Evaluation

The performance of the KNN model is evaluated using standard metrics such as accuracy and confusion matrix. Although KNN is simple and easy to implement, it requires higher computational time and may provide lower accuracy compared to SVM, especially for large datasets.

SAMPLE DATA

https://drive.google.com/drive/folders/1ZoxiMc9JTzfSBkMQjEbkj4yqy5cbGcO5?usp=sharing

4. RESULTS AND DISCUSSION

Electricity theft detection using supervised learning techniques is significantly influenced by the issue of class imbalance, where the number of normal (non-theft) consumers is considerably higher than that of fraudulent consumers. Due to this imbalance, relying solely on accuracy as a performance metric can lead to misleading results.

To overcome this limitation, multiple evaluation metrics derived from the confusion matrix are considered in this study. These metrics include accuracy, precision, recall, and false positive rate, which together provide a more

comprehensive and reliable assessment of the model's performance.

4.1 Confusion Matrix of SVM Algorithm

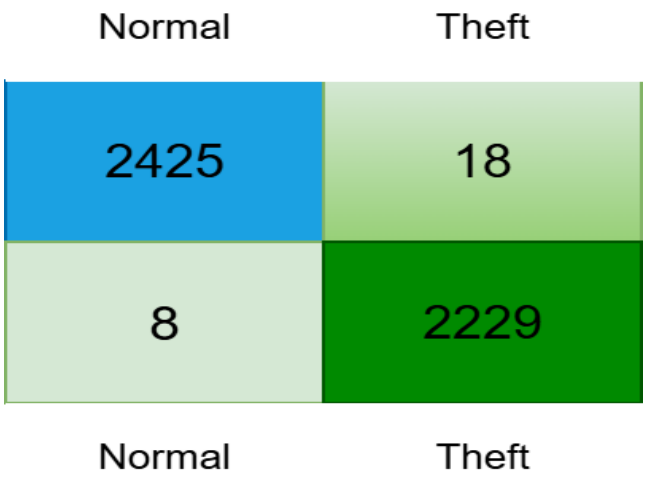

Fig -3: Confusion Matrix for SVM Approach

The confusion matrix summarizes the performance of the classification model in distinguishing between two classes: Normal and Theft. The rows represent the actual class labels, while the columns represent the predicted class labels.

True Positive [TP]: 2229

These are instances correctly classified as Theft. The model successfully identified 2229 theft cases.

True Negatives (TN): 2425

These correspond to Normal instances that were correctly classified. The model accurately labeled 2425 normal cases.

False Positives (FP): 18

These are Normal instances incorrectly classified as Theft. This indicates a small number of false alarms where normal behavior was mistaken for theft.

False Negatives (FN): 8

These represent Theft instances incorrectly classified as Normal. This is a critical error type, as actual theft cases were missed by the model.

4.1.2 Confusion Matrix Analysis

Fig. 3 shows the confusion matrix of the SVM model, where the rows represent the actual classes and the columns represent the predicted classes. The model achieves high classification performance, with 2229 true positives and 2425 true negatives. The number of misclassifications is very low, with only 18 false positives and 8 false negatives. This indicates that the model is effective in distinguishing between normal and theft instances, with a low false alarm rate and very few missed theft cases.

4.2 CONFUSION MATRIX FOR KNN ALGORITHM

Fig -4: Confusion Matrix for KNN Approach

The confusion matrix summarizes the performance of the classification model in distinguishing between two classes: Normal and Theft. The rows represent the actual class labels, while the columns represent the predicted class labels.

True Positive [TP]: 2366

These are instances correctly classified as Theft. The model successfully identified 2229 theft cases.

True Negatives (TN): 2220

These correspond to Normal instances that were correctly classified. The model accurately labeled 2425 normal cases.

False Positives (FP): 77

These are Normal instances incorrectly classified as Theft. This indicates a small number of false alarms where normal behavior was mistaken for theft

False Negatives (FN): 87

These represent Theft instances incorrectly classified as Normal. This is a critical error type, as actual theft cases were missed by the model.

4.2.1 Confusion Matrix Analysis

Fig. 4 shows the confusion matrix of the proposed model, where the rows represent the actual classes and the columns represent the predicted classes. The model correctly classifies 2220 theft instances and 2366 normal instances. However, some misclassifications are observed, with 77 false positives and 87 false negatives. This indicates that the model achieves good classification performance, although there are moderate errors in distinguishing between normal and theft instances.

4.3 ACCURACY OF SVM ALGORITHM

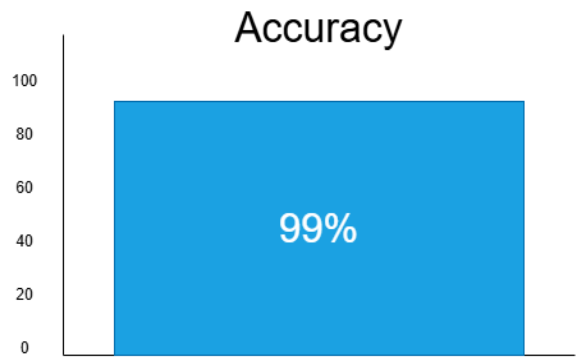

Fig -5: Accuracy of SVM

Fig. 5 shows the accuracy performance of the proposed model. The model achieves high accuracy, demonstrating its effectiveness in correctly classifying the data. The results indicate stable and reliable performance with minimal classification error.

Accuracy is defined as the ratio of correctly classified instances to the total number of instances.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

4.4 ACCURACY OF KNN ALGORITHM

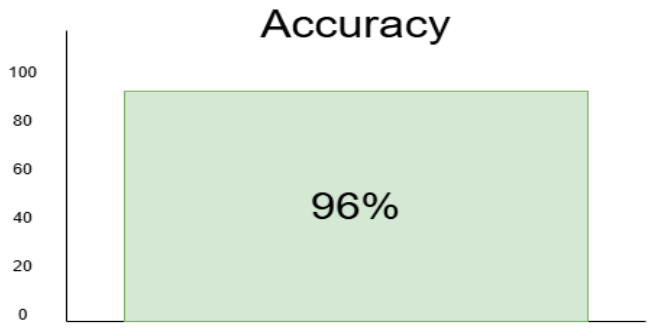

Fig -6: Accuracy of KNN

Fig. 6 shows the accuracy performance of the KNN model. The model achieves an accuracy of 96%, indicating good classification performance. It correctly classifies most instances, although some misclassifications are observed compared to higher accuracy models.

4.5 MODEL COMPARISON

The performance of the models is compared based on accuracy. The SVM model achieves an accuracy of 99%, while the KNN model achieves an accuracy of 96%. This shows that the SVM model performs better than the KNN model, with higher accuracy and fewer misclassifications. Hence, SVM is more effective for electricity theft detection in this study.

4.6 Performance Analysis of SVM

The performance of the Support Vector Machine (SVM) model is evaluated based on classification accuracy. The model is trained using the processed dataset and tested to determine its effectiveness in identifying normal and theft cases. The following graph illustrates the accuracy of the SVM model.

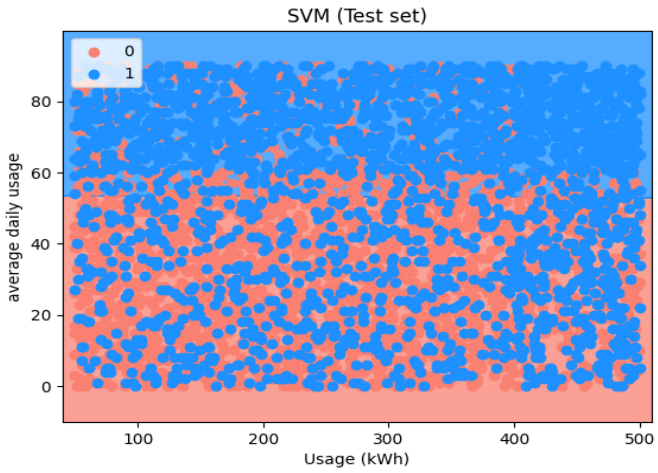

Fig -7: Performance of SVM

Fig. 7 shows the classification results of the SVM model on the test dataset, where the x-axis represents electricity usage (kWh) and the y-axis represents average daily usage. Orange indicates normal instances, while blue represents theft instances. The distribution shows that the model is able to distinguish between normal and theft patterns based on usage behavior. Theft instances are more prominent in higher usage regions, while normal instances are concentrated in lower usage regions. However, some overlap between the classes is observed, indicating minor misclassification. Overall, the model effectively separates the two classes and demonstrates reliable performance in detecting electricity theft.

4.7 Performance Analysis of KNN

The performance of the K-Nearest Neighbors (KNN) model is evaluated based on classification accuracy. The model is trained using the processed dataset and tested to analyze its effectiveness in identifying normal and theft cases. The following graph illustrates the accuracy of the KNN model.

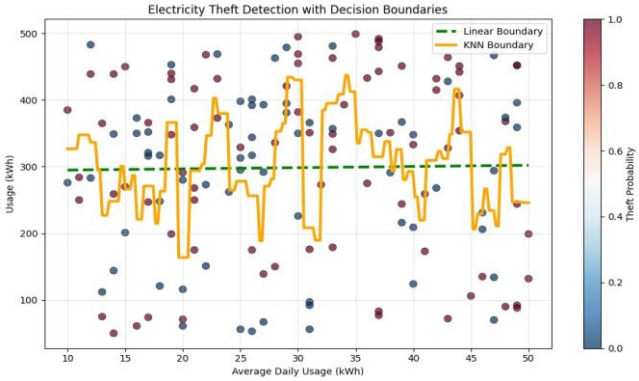

Fig -8: Performance of KNN

Fig. 8 illustrates the classification of electricity consumption patterns using machine learning models. The x-axis represents average daily usage (kWh), and the y-axis represents peak load (kW). Each data point corresponds to an individual consumer, while the color bar indicates theft probability, where higher values represent a greater likelihood of electricity theft.

The green dashed line denotes the linear decision boundary, which separates normal and suspicious consumption patterns using a linear approach. The orange curve represents the KNN (K-Nearest Neighbors) decision boundary, which adapts non-linearly to the data distribution.

From the figure, it can be observed that the KNN model captures complex variations in consumption behavior more effectively than the linear model. Regions exhibiting irregular usage patterns and abnormal peak loads are identified as potential theft cases. This indicates that non-linear models provide improved performance in detecting electricity theft.

5. CONCLUSION

In this paper, a machine learning-based approach for electricity theft detection has been proposed using Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) algorithms. The system utilizes smart meter data to analyze electricity consumption patterns and classify them as normal or theft.

Both SVM and KNN models were implemented and evaluated using standard performance metrics such as accuracy and confusion matrix. The experimental results demonstrate that both algorithms are capable of detecting electricity theft; however, the SVM model outperforms the KNN model in terms of accuracy and efficiency. The SVM model achieved an accuracy of 99%, indicating its effectiveness in identifying abnormal consumption patterns with minimal misclassification.

The proposed system helps reduce non-technical losses, improves the efficiency of power distribution systems, and minimizes the need for manual inspection. It provides an automated and reliable solution for electricity theft detection in modern smart grid environments.

In the future, the system can be enhanced by incorporating larger datasets, real-time monitoring, and advanced machine learning or deep learning techniques to further improve detection accuracy and scalability.

REFERENCES

[1] Safdar Ali Abro, Lyu Guang Hua, Javed Ahmed Laghari, Muhammad Akram Bhayo, and Abdul Aziz Memon, "Machine learning-based electricity theft detection using support vector machines," International Journal of Electrical and Computer Engineering (IJECE), vol. 14, no. 2, pp. 1240–1250, Apr. 2024, doi: 10.11591/ijece. v14i2.pp1240-1250.

[2] M. S. Saeed et al., "An efficient boosted C5.0 decision-tree-based classification approach for detecting non-technical losses in power utilities," Energies, vol. 13, no. 12, Jun. 2020, doi: 10.3390/en13123242.

[3] H. Ghaedi, S. R. K. Tabbakh, and R. Ghaemi, "Improving electricity theft detection using combination of improved crow search algorithm and support vector machine," Majlesi Journal of Electrical Engineering, vol. 15, no. 4, pp. 63–75, Dec. 2021, doi: 10.52547/mjee.15.4.63.

[4] J. Nagi, A. M. Mohammad, K. S. Yap, S. K. Tiong, and S. K. Ahmed, "Non-technical loss analysis for detection of electricity theft using support vector machines," in Proc. IEEE 2nd Int. Power and Energy Conf. (PECon), Dec. 2008, pp. 907–912, doi: 10.1109/PECON.2008.4762604.

[5] R. Akram et al., "Towards big data electricity theft detection based on improved RUSBoost classifiers in smart grid," Energies, vol. 14, no. 23, Dec. 2021, doi: 10.3390/en14238029.

[6] W. Zhu, N. Zeng, and N. Wang, "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations," in Northeast SAS Users Group Conf.: Health Care and Life Sciences, 2010, pp. 19.

[7] 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. Ind. Informatics, vol. 14, no. 4, pp. 1606–1615, Apr. 2018, doi: 10.1109/TII.2017.2785963.Industrial Informatics, vol. 14,

no. 4, pp.1606–1615, Apr. 2018, doi: 10.1109/TII.2017.2785963.

[8] S. H. Hussain, A. Hussain, R. Shah, and S. A. Abro, "Mini rover-object detecting ground vehicle (UGV)," University of Sindh Journal of Information and Communication Technology, vol. 3, no. 2, pp. 104–108, 2019.

[9] L. K. Ramasamy, S. Kadry, Y. Nam, and M. N. Meqdad, "Performance analysis of sentiments in Twitter dataset using SVM models," Int. J. Electrical and Computer Engineering, vol. 11, no. 3, pp. 2275–2284, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2275-2284.

[10] M. Zulqarnain, R. Ghazali, Y. M. M. Hassim, and M. Rehan, "Text classification based on gated recurrent unit combined with support vector machine," Int. J. Electrical and Computer Engineering (IJECE), vol. 10, no. 4, pp. 3734–3742, Aug. 2020, doi: 10.11591/ijece.v10i4.pp3734-3742.

[11] I. Slimani et al., "Automated machine learning: the new data science challenge," Int. J. Electrical and Computer Engineering (IJECE), vol. 12, no. 4, pp. 4243–4252, Aug. 2022, doi: 10.11591/ijece.v12i4.pp4243-4252.

[12] T. B. Smith, "Electricity theft: a comparative analysis," Energy Policy, vol. 32, no. 18, pp. 2067–2076, Dec. 2004, doi: 10.1016/S0301-4215(03)00182-4.

[13] A. Foster and V. Pushak, "Ghana's infrastructure: a continental perspective," in The Oxford Handbook of Religious Diversity, 2010.

[14] D. Carr and M. Thomson, "Non-technical electricity losses," Energies, vol. 15, no. 6, Mar. 2022, doi: 10.3390/en15062218.

[15] M. N. Hasan, R. N. Toma, A. Al Nahid, M. M. M. Islam, and J. M. Kim, "Electricity theft detection in smart grid systems: A CNN-LSTM based approach," Energies, vol. 12, no. 17, Aug. 2019, doi: 10.3390/en12173310.

[16] K. Fatima, M. Rafique, A. M. Soomro, and M. Kumar, "Tailoring hydrogen adsorption and desorption properties of Li-doped SV (single vacancy) monolayer h-BN systems using ab initio calculations," Canadian Journal of Physics, vol. 101, no. 11, pp. 673–685, Nov. 2023, doi: 10.1139/cjp-2023-0072.

[17] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," SN Computer Science, vol. 2, no. 3, May 2021, doi: 10.1007/s42979-021-00592-x.

[18] H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," Applied Computing and Informatics, vol. 18, nos. 1–2, pp. 90–100, Jul. 2022, doi: 10.1016/j.aci.2018.12.004.