International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Real-Time Detection of Energy Theft in Smart Grids Using Machine Learning

Sourav Pandey[1], Satyam Kandpal[2], Prof Deepika Rawat[3]

Department of Computer Science and Engineering (Internet of Things)
Raj Kumar Goel Institute of Technology, Ghaziabad, India

ABSTRACT—

the rising integration of smart meters within modern power grids has significantly improved the efficiency of electricity distribution. However, energy theft is still a big challenge, which leads to significant economic losses and operational inefficiencies. Conventional detection methods are often reactive and inefficient for real-time detection. This paper proposes a machine learning based Support Vector Machine (SVM) and XGBoost, which are calculated using real-world synthetic smart meter datasets. The system is designed to work on streaming data to identify defective consumption patterns. Comparative analysis based on precision, accuracy, recall, and F1-score points out the effectiveness of the proposed approach. In addition, the paper also discusses practical challenges in deployment, such as false positives, data privacy, and computational overhead. The results show that machine learning offers a promising pathway to securing smart grids against non-technical losses and enhancing overall system performance.

Keywords—Smart Grid, Traditional Electricity Grid, Software Systems, Renewable Integration, PSO, SVM, Machine Learning.

INTRODUCTION

Smart meter data analysis consists of all the strategies that involve information processing collected using smart meters. The advantages that the integration of smart meters brings into a network are diverse, including, the collection of energy data, management of electricity consumption, safer billing, more reliable service, a reduction of technical losses, a reduction of energy theft, an improvement of security profile for consumers, and the optimization of data management. Smart meters will therefore allow daily reading and collection of consumption data in real time. Electricity theft generally consists of any attempt to extract energy from the electricity network without consumption being officially recorded. Many electrical energy consumers make installations to avoid the recording of their energy consumption from electrical meters. Theft detection in this context is defined as monitoring customer behavior in order to estimate, detect, or avoid unwanted behavior. Works in use data in the smart grid for anomaly detection. Using smart meters, the operator can know customer consumption data and thus identify any anomaly that may occur. Conventionally, energy theft detection methods have been based upon manual inspections, static data analysis, and periodic audits. However, these methods require a lot of resources and are often inefficient in detecting fraudulent consumption patterns in real-time. Given the increasing volume of data generated by smart meters and the complexity of modern grids, traditional techniques fail to provide the scalability and efficiency needed to detect energy theft. That's why a transition to automated, real-time theft detection systems based on Machine Learning algorithms offers the required solution. This paper puts forward a machine learning-based framework for the real-time detection of energy theft in smart grids. The proposed framework uses the power of machine learning models to analyze smart meter data and identify abnormal consumption patterns indicative of energy theft. Several algorithms, such as RF, SVM, and XGBoost, are evaluated to assess their performance in terms of detection accuracy, precision, recall, and F1 score. The results show that machine learning-based solutions can significantly improve the speed and accuracy of energy theft detection, offering a more scalable and cost-effective approach in comparison to traditional methods.

DIFFERENT METHODS OF THEFT DETECTION

It has become very important to stop energy theft in smart grids, particularly after the deployment of advanced metering infrastructure, which gives real-time monitoring of electricity consumption. Many researchers have proposed various techniques to detect energy theft, ranging from traditional to modern machine learning-based approaches. This particular section will provide information about existing techniques of theft detection in smart grids.

Traditional Methods for Energy Theft Detection

Traditional methods of energy theft detection were based on manual inspections and static monitoring systems. These techniques often involved checking of electric meters and detecting discrepancies between reported consumption and expected usage based on various consumer profiles. While effective to some extent, these methods are labor-intensive, time-consuming, and can only detect theft after significant loss has occurred, and they are unable to provide real-time theft detection, which is essential for minimizing losses in a smart grid environment.

Data-Driven Approaches for Energy Theft Detection

Since the inception of smart grid technologies, researchers have moved towards data-driven techniques for energy theft detection, especially those using data analytics and machine learning. In different technical papers, the authors have written about the use of clustering algorithms to group consumers based on similar consumption patterns and identified anomalous groups that could potentially identify theft. These approaches have struggled with scalability and were not able to detect more sophisticated methods of theft.

Machine Learning Techniques for Energy Theft Detection

Due to continuous advancements in machine learning, researchers have inclined towards these modern methods for energy theft detection in smart grids. Many studies have applied supervised machine learning algorithms to identify fraudulent consumption patterns based on labelled training data. For instance, support vector machines (SVMs) have been widely used due to their ability to classify anomalies in consumption data. An SVM-based approach was proposed to detect energy theft by analyzing consumption patterns from smart meters. However, this technique required a large amount of labelled data and had limited performance in dealing with unseen data. More recently, Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) have gained popularity due to their ability to handle large datasets and complex patterns.

MATERIALS AND METHODS

Description of the Dataset

The dataset was obtained from the Cameroon Electricity Company. This dataset contains data from nearly 1000 consumers, of which 85.2% are normal consumers and 14.8% have abnormal consumption, which could be considered as electricity theft. The dataset was collected over an interval from January 1, 2020, to December 31, 2020. Table 1 presents a description of this dataset.

Table 2. Description of the Dataset

Explanations	Values
Electricity Consumption Period	Jan 1, 2020 to Dec 31, 2020
Total No of Consumers	1000
No of Normal Consumers	852
No of Fraudulent Consumers	148

OpenDSS-G

Some functionalities, such as line geo-localization, can only be implemented in OpenDSS-G, which is the evolution of simulation tools based on OpenDSS. Its interface has adopted OpenDSS functionalities to make advanced components of the platform easier for the user. This version includes parallel processing of OpenDSS elements. OpenDSS-G also allows modification of network elements during the simulation process.

OMNet++

Objective modular network in C++ (OMNet++) is a tool for simulating components intended for communication networks. OMNet++ is also an Eclipse platform that expands on other features, such as the editor and design wizard. OMNet++ also has properties for creating and configuring models (NED and ini file), ensuring batch execution and analysis of simulation results, while Eclipse provides the C++ editor and other optional elements (UML modelling, access to the database). It allows you to simulate a real existing communication network without any risk. Simulation results can be analysed to determine how the real network may be affected. In our simulation, OMNet++ makes it possible to create a NAN network to recover data at the transformer stations on the electricity network. It is used to build and simulate the communication network and enable simulation with OpenDSS through the COM interface. OMNet++ is capable of creating our communication network, simulating communication in the smart grid with OpenDSS, and ensuring better retransmission of information in the various local networks.

MATLAB

MATLAB is software that was initially developed by Cleve Moler in the 1970s. In addition, the MathWorks Company ensures its continuous development to this day. It is used for matrix calculations in order to analyse data and classify these. MATLAB also allows programming for the intelligent resolution of data mining problems. In our work, MATLAB allows us to implement all AI algorithms, such as SVM and PSO. All simulations of this work were carried out using the MATLAB R2020b 64-bit version.

Computer

All simulations were carried out on a computer with the following characteristics: Intel Core i5, 3.5 GHz, 8 GB RAM, 500 GB hard disk, and Windows 7/64 bit system.

METHODOLOGY OF ELECTRICITY THEFT DETECTION USING THE SVM-PSO TECHNIQUE

The analysis in this study focuses on detecting irregularities in customers' consumption behaviours by analysing their load profiles, which reflect their general electricity usage patterns. These profiles vary significantly, allowing for the comparison of typical customers against potential fraudsters. The detection process is powered by a Support Vector Classification (SVC) model, an advanced artificial intelligence approach used to identify fraudulent activities. The system employs a hybrid SVM-PSO (Particle Swarm Optimization) algorithm, where PSO optimizes the initial parameters of the SVM, such as velocity and position.

The detection methodology follows these steps:

- Data Retrieval: Collect customer data, both normal and fraudulent profiles, from the field or energy distribution agencies.
- Pre-processing: Use data mining techniques to pre-process the customer and billing data, converting it into a format (e.g., Excel or CSV) that can be used by the modelling tool (e.g., MATLAB).
- Feature Extraction: Identify and extract key parameters from the data, such as customer consumption indices, that will serve as input for the SVM model.
- SVM Parameter Initialization: In this step, we need to initialize all the basic parameters of the SVM model for proper operation.
- SVM Parameter Optimization: Now we need to fine-tune the SVM's kernel and error penalty parameters of the SVM model for proper operation. Using PSO to maximize the model's accuracy.
- SVC Training: The SVC model needs to be trained by adjusting its parameters through optimization and refining the model to match input data with the expected output.
- Model Development: Create and test the SVC classifier to predict future consumption behaviour and improve the detection of anomalies.
- Final Data Processing: Develop an algorithm to classify and select potentially fraudulent customers based on the predicted outcomes from the SVC model and their actual consumption patterns.

This streamlined process aims to efficiently identify suspicious customer behaviour, ensuring accurate detection of energy theft.

EVALUATION OF PERFORMANCE AND SENSITIVITY ANALYSIS

In this study, we performed a sensitivity analysis on the process by calculating outcomes of electricity theft detection under alternative electrical assumptions to evaluate the impact of fraudulent consumption. Therefore, the sensitivity analysis allows us to test the robustness of the data analysis results and increase the understanding of the relationships between input and output variables on electrical consumption data. It also allows us to reduce the uncertainty through the identification of the hybrid deep learning model inputs, research the errors in the model, simplify the complexity of the model in the space of input factors, and identify the correlations between observations. These processes aim to study how the uncertainty in the output of the model can be divided and allocated to different sources of uncertainty in its inputs.

The sensitivity analysis is performed using the following operations:

- Regression analysis, which involves fitting a linear regression to the model response
- Analysis of variance, which quantifies the input and output uncertainties as a probability distribution, is as follows:

$$\mathrm{Var}\left(E_{X_{\approx i}}\left(Y \otimes X_i\right)\right).$$

Where Var and *E*, respectively, are the variance and expected value operators and $X_{\approx i}$ is the set of all input variables except for X_i

The variance formulation measures the contribution X_i alone to the uncertainty in *Y* and is known as the first-order sensitivity index.

Variogram analysis of response (VARS), which addresses the weakness of directional variograms and variograms through recognizing a spatially continuous correlation structure to the value *Y* and hence to the values ($\partial Y/\partial x_i$)

RESULTS AND DISCUSSIONS

Data collection enables the analysis of customers' electricity consumption patterns, considering factors such as demographic trends and financial and social affiliations. The data, gathered over one year, includes information from 1,000 participants who have installed smart meters in their homes. The typical consumption patterns of regular customers are shown in *Figure 1*, where increased consumption is observed during February, April, and November, while May sees a significant drop. To in usage analyze this data, we employed a hybrid artificial intelligence approach combining Support Vector Machines (SVM) and Particle Swarm Optimization (PSO) for detecting potential fraudulent consumers. In Figure 2, we can observe the comparison between the normal consumption patterns to those of suspected fraudsters for the year. The sudden fluctuations in the consumption of fraudulent users allow the SVM to identify possible instances of theft. Through this technique, 4 potential fraud users were traced within the sample of consumers. Thus giving the system an accuracy of 98.9%.

Fig. 1 Normal power consumption.

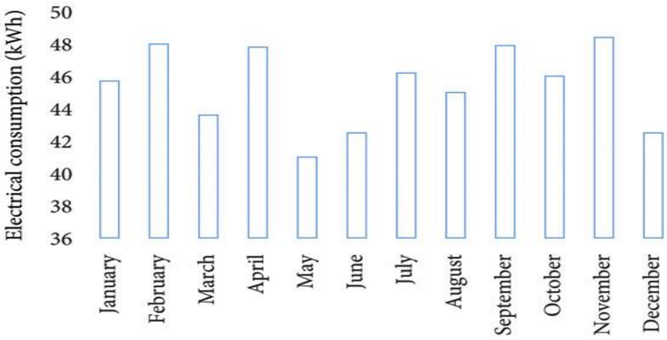
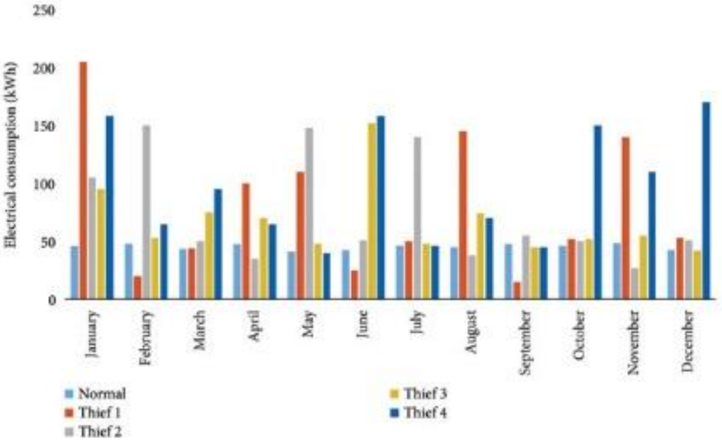


Fig. 2 Normal Consumers and Potential Thieves



CONCLUSION

This work presented a real-time model for electricity theft detection in a smart power grid. In which we used an OpenDSS-OMNet++ simulation platform to simulate the behaviour of a real smart grid to collect electricity consumption data in real time from smart meters installed for consumers. This model works based on support vector machines (SVMs) to detect energy theft in real time by observing different patterns of consumption of electricity between normal consumers and fraudulent consumers, and the Particle Swarm Optimization (PSO) algorithm was used to fine-tune the parameters of the SVM model. This system successfully identifies the fraudulent consumption patterns. Thus, the proposed intelligent model allows us to detect fraudulent values of electrical energy consumption. A consumption dataset of 1000 households was used to verify the effectiveness of the proposed method over one year. The simulation results give a performance of 98.9% for the detection of electricity fraud in a smart grid based on data obtained from smart meters. In addition, the detection time is relatively reduced and the AUC is increased, demonstrating the effectiveness of the proposed method compared to that in the literature. This method can be effective for implementation in a larger power network with thousands of customers, thereby enriching the learning database in the long term. The proposed model can also be optimized by using meta-heuristic prediction algorithms and wireless connection devices. The limitations of this work concern the nonlinearity of the smart meter data, which can affect the performance of the deep learning model. Moreover, further research can be conducted on the implementation of supervised deep learning techniques with different datasets to obtain better performance.

REFERENCES

[1] Yip S. C., Wong K., Hew W. P., Gan M. T., Phan C. W., and Tan Su W., Detection of energy theft and defective smart meters in smart grids using linear regression, International Journal of Electrical Power and Energy Systems. (2017) 91, 230–240, https://doi.org/10.1016/j.ijepes.2017.04.005, 2-s2.0-85018293402.

[2] Lydia M., Edwin Prem Kumar G., and Levron Y., Detection of electricity theft based on compressed sensing, Proceedings of the 5th International Conference on Advanced Computing and Communication Systems (ICACCS), March 2019, Coimbatore, India.

[3] Ahir R.K. and Chakra Borty B., Pattern-based and context-aware electricity theft detection in smart grid, *Sustainable Energy, Grids and Networks*. (2022) 32, https://doi.org/10.1016/j.segan.2022.100833.

[4] Punmiya R. and Choe S., Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing, *IEEE Transactions on Smart Grid*. (2019) 10, no. 2, 2326–2329, https://doi.org/10.1109/tsg.2019.2892595, 2-s2.0-85062320575.

[5] Gao Y., Foggo B., and Yu N., A physically inspired data-driven model for electricity theft detection with smart meter data, *IEEE Transactions on Industrial Informatics*. (2019) 15, no. 9, 5076–5088, https://doi.org/10.1109/tii.2019.2898171.

[6] Yem Souhe F. G., Boum A. T., Ele P., Mbey C. F., and Foba Kakeu V. J., A novel smart method for state estimation in a smart grid using smart meter data, *Applied Computational Intelligence and Soft Computing*. (2022) 2022, 1–14, https://doi.org/10.1155/2022/7978263.