IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13582

Electricity Theft Detection in Smart Grids Using Sarimax & OCR

Mr. Abhale B.A[1], Ansari Aiman[2], Jamdar Omkar[3], Dange Satyam[4].

Asst. Professor, Department of IT Engineering, SND College of Engineering and Research Center,Yeola, India[1]

Students, Department of IT Engineering, SND College of Engineering and Research Center,Yeola, India[2-5]

Abstract: Electricity Theft has grown in tandem with the rise in electricity use. Electricity theft is causing widespread power outages and developing nations such as India are facing a major electricity crisis. When a consumer tampers with the units consumed it is called Electricity Theft. Because of illegal intervention with electric meters, power utilities are in deep financial trouble. In this paper, we are going to present a Desktop Application to Detect Electricity theft in Smart Distribution grids using machine learning. In this system, we utilize OCR (Optical Character Recognition) to recognize the meter reading (units used) from the meter image and convert it into machine-readable text. The SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with eXogenous components) algorithm is a new variant of the ARIMA algorithm that is used to anticipate consumer consumption and detect Electricity Theft. If theft is detected, a message will be delivered to the fraudulent customer. The suggested method would assist Electricity Boards in detecting and recovering losses from electricity theft.

Keywords: OCR, SARIMAX, Electricity Theft, Electricity Board, Desktop Application, Machine Learning.

I. INTRODUCTION

Electricity is lost in the generation, transmission, and distribution of electricity. There are two types of losses, Technical Losses (TL) and Non-Technical Losses (NTL). Energy dissipation in transmission lines and magnetic losses in transformers can contribute to technical losses. Electricity theft is the most common source of nontechnical losses. Electricity theft is a serious problem. Electricity theft is taking electricity from a power company and paying less for it than the actual use of electricity. Following are some examples of electricity fraud: 1) Meters that are broken or malfunctioning 2) Tampering with meters to inflate usage figures. 3) Connecting directly to electricity sources and bypassing meters 4) Supply is unmetered 5) Meter reading and data processing errors due to human or technological error globally, power utilities have lost money. The most obvious cause of the loss is energy theft. In wealthy countries, electrical utilities predict a 15% loss, but in India, the overall loss is 30% of total generation, which is roughly equal to 1.5 percent of the country’s GDP. Traditionally methods suggest using Intelligent Energy Meter or an IoT-based solution which is costly. In this paper, we have used the time series Estimation model which is a machine learning technique to analyze historical data of the customer and predict its estimated consumption. A time series is simply a series of data points ordered in time. In this model time is the independent variable and the goal is to make a forecast for the future. If Electricity theft is detected then the message is sent to the consumer. The message informs the consumer about the penalty to be paid for committing fraud. If the Unit Consumption is Genuine, no message will be sent. Now with the introduction of smart meters, the frequency with which household energy consumption data is collected has grown, allowing for extensive data analysis that was previously unavailable. The method of monitoring energy use is still largely human-dependent, and it must be done every month for each client owing to legal constraints. To meet these needs, power companies use various human resources, namely readers, who visit each customer (residential or commercial) once a month to read the electrical meter and issue a bill for that month’s usage. Each reader carries a gadget that allows them to enter the consumption figure, photograph the meter, and print the bill on the spot. Nonetheless, exhaustion, the difficulty of reading electrical meters owing to natural wear, the absence of meter uniformity, and other variables impact humans. As a result, having people interpret the numbers in the meter might result in incorrect invoicing. These billing mistakes raise power companies’ expenses, reduce customer trust in the firm, and cause difficulties with energy regulatory bodies. To avoid such issues, electricity providers attempt to estimate each customer’s projected usage and compare it to the value of the human reader. There are two parts to this: estimating the value and determining a tolerance around the expected value. When these two stages are combined, an upper limit for the values to be read is established. The bill is produced and provided to the consumer directly if the read value is near to the expected value (i.e. less than the projected value plus a tolerance). When the two numbers diverge (i.e., are more than the projected value plus a tolerance), a snapshot of the electrical meter is sent to an analyst, who will determine if the reading is correct later. Finally, each customer’s monthly usage figure is recorded in a log for future reference.

DOI: 10.17148/IJARCCE.2024.13582

II. LITERATURE SURVEY

Optical character recognition (OCR) is a technology that facilitates the translation of many forms of texts or images into analyzable, editable, and searchable data. During the last decade, researchers have employed artificial intelligence/machine learning techniques to automatically evaluate handwritten and printed materials in order to convert them to electronic format. The goal of this paper was to review the paper and outline previous research on character recognition in handwritten documents and offer research directions. An OCR system is primarily dependent on feature extraction and discrimination/classification of these characteristics. Handwritten OCR is gaining popularity as a subset of OCR. Based on this, it is further divided into offline systems and online systems. Jenny Mahoney was the assistant editor in charge of organizing the evaluation of this article and clearing it for publication.

Data input the offline system is a static system with input data in the form of scanned photographs, but the type of input in the online system is more dynamic and is based on the movement of a pen tip with a certain velocity, projection angle, location, and locus point. As a result, an online system is deemed more complicated and advanced than an offline system since it overcomes the overlapping problem of input data that exists in the offline system. In addition to projecting power consumption, the SARIMAX models were employed as forecasting tools in a variety of sectors of application. To anticipate daily traffic counts, Cools et al. (2009) created the ARIMAX and SARIMAX models. In their study, they looked at seasonality in daily traffic statistics as well as the effects of holidays at various site locations. Incorporating weekly seasonality and vacation impacts at various site locations demonstrated that both the ARIMAX and SARIMAX models are superior frameworks.

III. AIM & OBJECTIVES

Aims:

- Investigate the effectiveness of integrating SARIMAX and OCR technologies for enhancing electricity theft detection in smart grids.
- Assess the potential benefits of employing advanced forecasting and meter reading analysis techniques in mitigating revenue losses due to electricity theft.
- Explore the feasibility of implementing SARIMAX and OCR-based theft detection systems in real-world smart grid environments.
- Contribute to the advancement of smart grid security by proposing innovative methodologies for detecting and preventing electricity theft.
- Provide insights and recommendations for utility companies and policymakers seeking to enhance the resilience and integrity of their power distribution networks.

Objectives:

- Review existing literature on electricity theft detection methods, smart grid technologies, SARIMAX modelling, and OCR technology.
- Develop a conceptual framework for integrating SARIMAX and OCR techniques for electricity theft detection in smart grids.
- Collect and analyze historical electricity consumption data and meter readings from a sample smart grid network.
- Design and implement SARIMAX and OCR-based algorithms for anomaly detection and theft identification.
- Evaluate the performance and accuracy of the proposed theft detection system through empirical testing and validation.
- Assess the scalability, cost-effectiveness, and practicality of deploying SARIMAX and OCR-based theft detection solutions in real-world scenarios.
- Identify potential challenges and limitations associated with the integration of SARIMAX and OCR technologies and propose strategies for overcoming them.
- Document the findings, insights, and recommendations derived from the research for dissemination to relevant stakeholders and academic communities.

IV. MOTIVATION

The motivation for this research is to tackle the persistent problem of electricity theft in smart grids. Traditional methods often fail to detect sophisticated theft, leading to financial losses and network instability.

By integrating SARIMAX and OCR technologies, the aim is to develop a proactive solution for accurate theft detection. This research seeks to safeguard revenue, ensure network integrity, and advance sustainable energy infrastructure.

APPLICATION:

The integration of SARIMAX and OCR technologies for electricity theft detection in smart grids has wide-ranging applications:

1. Utility Companies: Enhance revenue protection and network integrity.
2. Government Agencies: Enforce energy regulations and ensure fair competition.
3. Energy Consumers: Potentially stabilize electricity prices and improve affordability.
4. Technology Providers: Offer advanced solutions, driving innovation.
5. Research: Continuously improve detection accuracy and scalability for future systems.

V. SYSTEM ARCHITECTURE

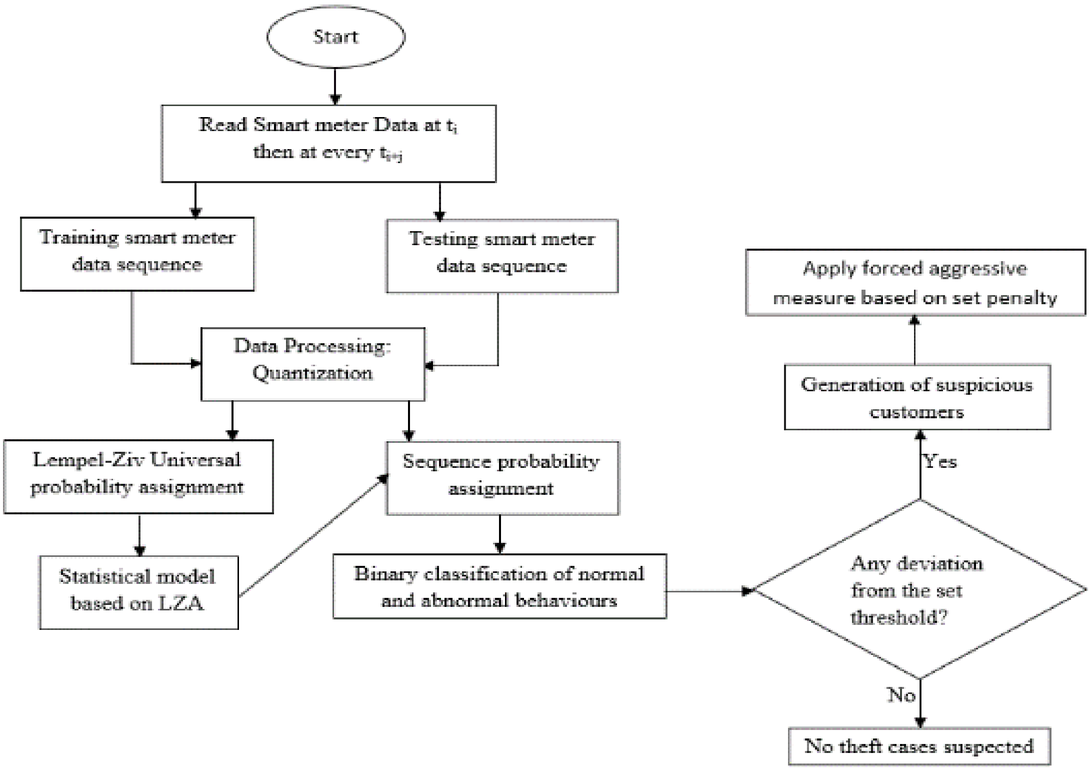
Start
Read Smart meter Data at t_i then at every t_{i+j}
Training smart meter data sequence
Testing smart meter data sequence
Data Processing: Quantization
Lempel-Ziv Universal probability assignment
Sequence probability assignment
Statistical model based on LZA
Binary classification of normal and abnormal behaviours
Any deviation from the set threshold?
Yes
No
Generation of suspicious customers
Apply forced aggressive measure based on set penalty
No theft cases suspected

Fig -1: System Architecture Diagram

VI. ADVANTAGES

The advantages of integrating SARIMAX and OCR technologies for electricity theft detection in smart grids include:

1. Enhanced Accuracy:
SARIMAX models provide precise consumption forecasts, improving the accuracy of theft detection.

2. Automated Analysis:
OCR technology automates meter reading analysis, reducing manual effort and human error.

3. Proactive Detection:
Early identification of anomalies enables utility companies to take prompt action, minimizing revenue losses.

4. Comprehensive Approach:
The integration of SARIMAX and OCR offers a holistic solution for detecting a wide range of theft scenarios.

DOI: 10.17148/IJARCCE.2024.13582

5. **Cost Savings:**
By reducing theft-related revenue losses, utility companies can potentially avoid the need for tariff increases, benefiting consumers.

6. **Improved Security:**
Enhanced theft detection strengthens the security and resilience of smart grid infrastructures, protecting against fraudulent activities.

7. **Scalability:**
SARIMAX and OCR-based systems can be scaled to accommodate varying network sizes and complexities.

8. **Adaptability:**
Advanced analytics and machine learning algorithms enable continuous refinement and optimization of theft detection capabilities.

9. **Regulatory Compliance:**
By detecting and preventing theft, utility companies can comply with energy regulations and industry standards.

10. **Sustainability:**
Effective theft detection contributes to the sustainability of energy distribution systems, promoting efficient resource utilization and environmental conservation.

VII. CONCLUSION

This proposed system detects the electricity theft using xgboost OCR, Sarimax machine learning method. This proposed system helps to electricity utilities to detect electricity theft and they will not have to bear loss. This system will detect electricity theft or not.

REFERENCES

[1]. J. Nagi, K. Yap, S. Tiong, S. Ahmed, and M. Mohamad, “Nontechnical loss detection for metered customers in power utility using support vector machines,” IEEE Transactions on Power Delivery, vol. 25, no. 2, pp. 1162–1171, 2010, cited by 104.

[2]. S. McLaughlin, D. Podkuiko, and P. McDaniel, “Energy theft in the advanced metering infrastructure,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 6027 LNCS, pp. 176–187, 2010, cited By 99.

[3]. G. Tsekouras, N. Hatziargyriou, and E. Dialynas, “Two stage pattern recognition of load curves for classification of electricity customers,” IEEE Transactions on Power Systems, vol. 22, no. 3, pp. 1120–1128, 2007, cited By 122. [Online].

[4]. Y. Zhang, W. Chen, and J. Black, “Anomaly detection in premise energy consumption data,” 2011, cited By 12. [Online].

[5]. S. Dua and X. Du, Data Mining and Machine Learning in Cybersecurity, 1st ed. Boston, MA, USA: Auerbach Publications, 2011.

[6]. V. Barnett and T. Lewis, Outliers in Statistical Data, ser. Wiley Series in Probability Statistics. Wiley, 1994. [Online]. Available: https://books.google.com.pr/books?id=B44QAQAAIAAJ

[7]. N. Billor, A. Hadi, and P. Velleman, “Bacon: Blocked adaptive computationally efficient outlier nominators,” Computational Statistics and Data Analysis, vol. 34, no. 3, pp. 279–298, 2000, cited By 154. College Short Form Name, Department of Computer Engineering 2021 40

[8]. E. Eskin, “Anomaly detection over noisy data using learned probability distributions,” in Proceedings of the Seventeenth International Conference on Machine Learning, ser. ICML ’00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 255–262.

[9]. E. M. Knorr and R. T. Ng, “Algorithms for mining distance-based outliers in large datasets,” in Proceedings of the 24rd International Conference on Very Large Data Bases, ser. VLDB ’98. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 392–403.

[10]. C. Aggarwal and P. Yu, “Outlier detection for high dimensional data,” 2001, pp. 37–46, cited by 433.