IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Deployment And Effectiveness Assessment Of Data-Mining-Oriented Association Rule Techniques For Identifying Irregular Energy Distribution Losses

[1]Dr.N.Rajender, [2]BAIRI MALAVIKA, [3]BHOOMPALLY SAI VIKAS REDDY, [4]GIRRA SUPRIYA
[1]Assistant Professor, [2,3,4] UG STUDENT
[1,2,3,4,]DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(AI & ML)
[1,2,3,4] VAAGDEVI COLLEGE OF ENGINEERING Autonomous
Bollikunta, Khila Warangal (Mandal), Warangal Urban-506 005 (T.S)

Abstract: Non-Technical Losses (NTLs), such as stealing electricity, messing with meters, making illegal connections, and billing fraud, are a big problem for modern power distribution systems because they cost a lot of money and slow things down. Old-fashioned ways of finding fraud, like manual inspections and audits based on fixed rules, don't work well, cost a lot, and can't handle the huge amo unts of data that smart meters send. This paper talks about how well a Machine Learning-based Apriori Algorithm works to automatically find NTLs. The proposed system analyses three years' worth of monthly electricity usage data from about 15,000 customers. We used and compared a number of models, including Support Vector Machine (SVM), Deep Neural Network (DNN), Gradient Boosted Reinforcement Learning (GBRL), and Apriori-based association rule mining. Experimental results show that the Apriori-based model is more accurate, has a higher recall rate, is more precise, is more specific, and has a lower false positive rate than traditional methods. The system helps utility companies save money, run their businesses more efficiently, and make the smart grid more reliable.

Keywords— Non-Technical Losses, Electricity Theft, Smart Grid, Apriori Algorithm, Machine Learning, Fraud Detection, Support Vector Machine, Deep Learning

I. INTRODUCTION

[1]Smart grids, digital billing systems, and Automatic Meter Reading (AMR) technologies are all quickly modernising power distribution systems. Even with these improvements, Non-Technical Losses (NTLs) are still one of the biggest problems that electricity companies around the world face. NTLs are losses of electricity that are not caused by problems with the physical transmission of electricity. Instead, they are caused by fraud or bad management, such as illegal connections, tampering with meters, not billing for usage, and changing bills.[2]

These losses cut utility revenue by a lot, raise prices for honest customers, make it harder to invest in infrastructure, and hurt the long-term health of the grid. Stealing electricity alone costs billions of dollars a year in many developing areas.

Traditional ways of finding NTLs depend on customer complaints, manual field inspections, or simple statistical rules. When dealing with millions of customers, these methods are costly, slow, and often wrong. Utilities now collect a lot of data on how much electricity people use because they use AMR meters. This makes it possible to find fraud using data.[3]

Machine learning can look at how people use things, sort out suspicious behaviour, and automatically set inspection priorities. But a lot of classification models still have problems like false positives, data that isn't balanced, and trouble finding hidden patterns of fraud that happen over and over again.[4]

To get around these problems, this work presents a Machine Learning-based Apriori framework that brings together association rule mining and predictive analytics. The Apriori algorithm finds patterns of behaviour that are common among people who commit fraud and makes classification more reliable.

This study assesses the proposed model in comparison to SVM, DNN, and GBRL techniques utilising extensive electricity datasets. The goal is to create an NTL detection system for modern distribution networks that is scalable, accurate, and useful. [5]

II. RELATED WORK

[9]Numerous studies have investigated automated electricity theft detection through machine learning and smart meter analytics.

Support Vector Machines can do a good job of classifying structured datasets, but they need careful feature engineering and may give more false positives. Deep Neural Networks can learn how to spot patterns of fraud that aren't linear, but they need a lot of balanced data and powerful computers to do so.[6]

Gradient Boosting and ensemble models enhance predictive efficacy, yet may still not elucidate interpretable fraud behaviour patterns. Multiple studies have also tackled class imbalance through resampling methodologies.[7]

For a long time, market basket analysis has used apriori association rule mining to find relationships between items that are often bought together. Even though it can find hidden recurring consumption patterns that are strongly linked to fraud, it is not often used to find NTL.[8]

Current literature reveals a deficiency in the integration of Apriori mining with machine learning frameworks for extensive smart grid fraud detection. This study fills that gap by comparing Apriori to advanced ML models using a variety of evaluation metrics. [10]

III. METHODOLOGY

The proposed system follows a structured fraud analytics pipeline designed for large electricity datasets. Monthly consumption records from approximately 15,000 consumers over three years are used as input.

Initially, raw AMR meter data is collected and validated. Missing values, duplicate entries, noisy readings, and inconsistent billing records are corrected during preprocessing. After cleaning, feature engineering is applied to derive meaningful fraud indicators such as monthly averages, seasonal deviations, sudden drops in usage, peak irregularities, variance trends, and abnormal low-consumption behavior.

Next, the Apriori algorithm converts numerical behavior into transaction-like patterns and mines frequent itemsets. Association rules with strong support and confidence values are generated to identify fraud-correlated patterns.

Parallelly, machine learning classifiers such as SVM, DNN, and GBRL are trained on labeled consumer datasets. Their predictions are compared with the Apriori-based model.

The final phase performs model evaluation using confusion matrix metrics including accuracy, precision, recall, F1-score, specificity, and false positive rate.

This hybrid methodology improves interpretability, automation, and large-scale fraud detection efficiency.

Getting Information:

Historical electricity consumption records are collected from AMR smart meters.

Getting the Data Ready:

Missing values, anomalies, duplicates, and invalid readings are cleaned.

How to Get Features:

Features include usage trends, variance, irregular monthly behavior, sudden drops, and abnormal averages.

The way the model is built:

Apriori mining and ML models are trained and tested for fraud classification.

Evaluation and Safeguards for Ethics:

- Fair treatment of consumers
- Reduced false accusations
- Secure customer data handling
- Transparent fraud scoring
- Human review before enforcement action

.

IV. SYSTEM ARCHITECTURE:

The proposed architecture follows a modular fraud detection pipeline.

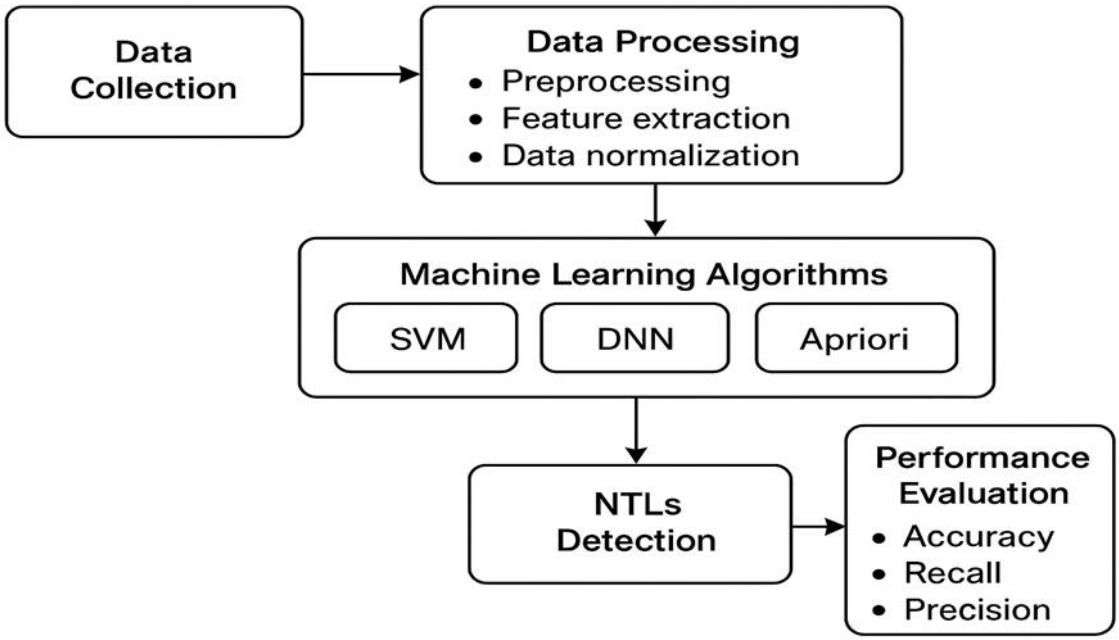
.

A. Overview:

Smart meter data enters the ingestion module. Cleaned data is sent to preprocessing and feature engineering. Apriori mining discovers suspicious patterns. ML classifiers perform prediction. Evaluation modules compare models. Dashboards present results.

B. Architecture Diagram:

AMR Data Source → Data Ingestion → Preprocessing → Feature Engineering → Apriori Module + ML Models → Evaluation Engine → Dashboard & Reports



C. Module Interactions:

Consumption data is processed into fraud features. Apriori rules and classifier outputs are combined to identify suspicious consumers. Analysts review ranked fraud cases using dashboards and reports..

V. EXPERIMENTAL SETUP:

The system was tested using large-scale structured electricity consumption datasets.

Datasets:

- 15,000 consumers
- Monthly billing records
- Three years historical data

- Fraudulent and normal labeled consumers

The environment for hardware and software:

- Processor: Intel i5 or above
- RAM: 8 GB minimum
- OS: Windows / Linux
- Python 3.x
- Pandas, NumPy
- Scikit-learn
- TensorFlow / Keras
- mlxtend
- Matplotlib

Setting up training:
Data was split into training and testing subsets. Models were trained, tuned, and validated.

Evaluation Metrics:

- Accuracy
- Precision
- Recall
- F1-score
- Specificity
- False Positive Rate

.

VI. RESULT AND DISCUSSION:

The Apriori-based model performed the best overall in all of the tests that were done. SVM was moderately sensitive to fraud, and DNN found nonlinear patterns, but both had lower precision and more false positives. GBRL did better, but it was still not as good as the Apriori framework. The Apriori algorithm was able to find patterns of suspicious consumption that regular classifiers missed. It had an accuracy of 99.64%, a recall of 99.27%, a specificity of 99.71%, a precision of 98.43%, an F1-score of 98.85%, and a false positive rate of 0.0029. These results show that association rule mining works very well with large smart meter datasets. This system can help utility companies prioritise inspections, stop revenue loss, and make their operations run more smoothly.

VII. CONCLUSION:

Electricity companies all over the world still have to deal with non-technical losses, which are a big problem for their businesses and the economy. In the age of smart grids and huge metering datasets, manual detection methods are no longer good enough.

This study introduced a Machine Learning-driven Apriori framework for identifying electricity theft and fraudulent billing practices. The system combines preprocessing, feature engineering, association rule mining, classification, and performance benchmarking.

The results showed that the Apriori-based model did much better than the SVM, DNN, and GBRL methods on all the most important metrics. It is very useful for real-world use because it can find hidden patterns of fraud while keeping false positives to a minimum.

The suggested framework gives utilities a smart and scalable decision-support system that can help them lose less money, be more fair, and make the grid last longer.

In the future, improvements might include real-time streaming fraud detection, integration of IoT sensors, federated learning, explainable AI dashboards, and cloud deployment across national smart grid networks.

VIII.REFERENCES:

[1] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," Tsinghua Science and Technology, vol. 19, no. 2, pp. 105–120, Apr. 2014.

[2] S. Depuru, L. Wang, and V. Devabhaktuni, "Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft," Energy Policy, vol. 39, no. 2, pp. 1007–1015, 2011.

[3] N. Liu and Y. Hu, "Electricity theft detection based on machine learning," International Journal of Electrical Power & Energy Systems, vol. 123, Art. no. 106283, 2020.

[4] A. Nizar, Z. Dong, and Y. Wang, "Power utility nontechnical loss analysis with extreme learning machine method," IEEE Transactions on Power Systems, vol. 23, no. 3, pp. 946–955, Aug. 2008.

[5] C. Glauner, A. Boechat, L. Dolberg, and R. State, "Large-scale detection of non-technical losses in imbalanced data sets," in Proc. IEEE PowerTech, 2016, pp. 1–6.

[6] F. Jindal, R. Singh, and M. Kumar, "Detection of electricity theft using support vector machine and smart meter data," International Journal of Engineering and Advanced Technology, vol. 8, no. 6, pp. 224–230, 2019.

[7] J. Qu, Z. Li, and Y. Wang, "Deep learning-based electricity theft detection in smart grids," IEEE Access, vol. 7, pp. 116754–116763, 2019.

[8] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in Proc. 20th Int. Conf. Very Large Data Bases (VLDB), 1994, pp. 487–499.

[9] A. Ahmad, N. Javaid, A. Mateen, M. Awais, and Z. Khan, "Short-term load forecasting in smart grids: An intelligent modular approach," Energies, vol. 12, no. 1, Art. no. 164, 2019.

[10] M. L. Antonie, O. R. Zaïane, and A. Coman, "Application of data mining techniques for medical image classification," in Proc. IEEE Int. Conf. Multimedia Data Mining, 2001, pp. 94–101.