AI-Based Detection of Electricity Theft in Smart Meter Networks

Mangesh Satish Kendre

Research Proposal

FEBRUARY 2025

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI……..…….. Artificial intelligence

ML…..…….Machine learning

IQR……...…..... Interquartile Range

RFE……..……... Recursive Feature Elimination

SMOTE……... Synthetic Minority Over-sampling Technique

Table of Contents

1. Problem Statement

Electricity theft is a widespread issue that leads to significant financial losses, operational inefficiencies, and unfair electricity pricing. It occurs when an entity manipulate their electricity consumption to avoid paying their actual usage. General methods of theft consists meter tamper, illegal connections, and bypass of smart meters. This problem is particularly severe in regions where regulatory enforcement is weak, leading to financial annual losses for power companies.

The consequences of electricity theft extend beyond financial losses. When unauthorized users consume electricity without paying, utility companies are forced to recover these losses by increasing tariffs, which affects honest consumers. Moreover, theft-related tampering can damage power infrastructure, leading to voltage instability, power outages, and increased maintenance costs. In extreme cases, unsafe connections can result in electrical fires, posing a serious safety hazard. Additionally, inaccurate billing due to theft undermines consumer trust in power providers, making revenue recovery even more challenging.

Traditional theft detection methods, such as manual inspections and consumer complaints, are slow, inefficient, and labor-intensive. Power companies often lack the resources to monitor large numbers of users effectively. Although smart meters generate detailed consumption data, there is no efficient way to analyze this information in real-time to detect fraud. This creates a gap in theft prevention and allows fraudulent activities to continue undetected.

This research proposes an AI-based classification model that utilizes smart meter data to detect electricity theft. By applying machine learning techniques, the model will analyze consumption patterns, identify anomalies, and classify users as either "stealer" or "non-stealer." Implementing such a system will help power companies minimize losses, enhance revenue protection, and ensure fair electricity distribution. A data-driven approach to theft detection is essential for improving grid efficiency and maintaining a sustainable energy supply.

2.1 Research Question

The Below research questions are suggested for the research objectives as follows.

- How AI and machine learning be used to detect electricity theft based on smart meter consumption data?
- How does class imbalance in theft detection datasets impact model performance, and what techniques can be applied to mitigate this issue?
- What machine learning algorithms are most effective for classifying users as "stealer" or "non-stealer"?

2.2 Aim and Objectives

The aim of this research is to develop an AI-based classification model for detecting electricity theft using smart metering data. The study seeks to leverage machine learning practices to analyze usage patterns and accurately classify users as either "stealer" or "non-stealer." By integrating this model into a real-time monitoring system, the research aims to provide an automated, efficient, and scalable solution for utility companies to minimize revenue losses, enhance grid security, and improve operational efficiency.

The research objectives are outlined as follows:

- To analyze electricity consumption patterns from smart meter data to identify key indicators of fraudulent behavior.
- To preprocess the dataset by applying suitable steps such as handling missing values, normalizing data, and addressing class imbalance to ensure model effectiveness.
- To investigate different ML modelling practices and their applicability in detecting electricity theft.
- To propose suitable classification models, including Classifier models for theft detection.
- To evaluate the performance of classification model based on metrics such as accuracy, F1 score, recall & precision.
- To assess practical impact of AI driven theft detection on power companies, focusing on revenue protection and operational improvements.

3.1 Research Methodology

The research methodology outlines the step-by-step approach to developing an AI-based classification model for detecting electricity theft. The methodology follows a structured process, from reviewing existing research to implementing, evaluating, and analyzing the proposed solution.

The research methodology consists steps as below:

- **Literature Review:** Practice a brief review of available work to understand electricity theft detection techniques through AI-driven approaches.
- **Experimental Design:** Design experiments which outlines the structured approach for developing, training, and evaluating the AI-based electricity theft detection model.
- **Data Collection:** Gather the data of the smart meter consumption records for experiments.
    1. **Data set:** The data required is: Smart_meters_dataset.
    2. **Data processing:** Data processing steps done for electricity theft detection using machine learning are as below:
        - **Handling Missing Data:** Identify and address missing values in electricity consumption records to improve data integrity. Techniques such as mean, median, mode imputation, forward or backward filling for time-series data, or predictive imputation using regression models can be used to fill gaps.
        - **Detecting and Handling Outliers:** Extreme values in power consumption may indicate faulty readings or electricity theft. Methods like Z-score, Interquartile Range (IQR), and visualization tools (e.g., box plots, histograms) can be used to detect anomalies. Identified outliers can be removed or adjusted using Winsorization or log transformations.
        - **Noise Reduction:** Smart meter data may contain noise due to sensor errors or environmental factors. Smoothing techniques such as moving averages, median filtering, or removing duplicate records help clean noisy data and improve consistency.
        - **Data Type and Format Verification:** Ensure that numerical values, timestamps, and categorical variables are correctly formatted and stored in appropriate data types to prevent errors during model training.

- **Normalization and Standardization:** Scale numerical features to maintain uniformity in model training.
    - **Normalization (Min-Max Scaling):** Rescales values between 0 and 1 to ensure all features have equal influence.
    - **Standardization (Z-Score Scaling):** Converts data to a standard normal distribution mean 0 and a standard deviation 1.
- **Feature Engineering:** Extract and create meaningful attributes from raw electricity usage data. Features like peak usage times, daily consumption trends, and unusual spikes can help improve theft detection accuracy.
- **Feature Selection:** Remove redundant or irrelevant features to enhance model performance and reduce complexity. Correlation analysis, recursive feature elimination (RFE), and statistical tests such as Chi-Square test used to select the most significant attributes.
- **Handling Class Imbalance:** Electricity theft cases are usually much fewer than normal cases, leading to an imbalanced dataset. Practices like Synthetic Minority Over-sampling Technique (SMOTE), undersampling of majority class can help balance data.
- **Encoding Categorical Data:** Conversion of categorical attributes (e.g., consumer type, region) into numerical values using encoding methods:
    - **One-Hot Encoding:** Converts categorical variables into binary columns.
    - **Label Encoding:** Assigns numerical values to categorical variables.
- **Data Splitt:** Divide dataset into train and test sets, commonly in an 80-20 or 70-30 ratio. Additionally, K-Fold Cross-Validation can be applied to improve generalization and prevent overfitting.

By applying these data pre-processing techniques, the dataset becomes clean, balanced, and ready for effective electricity theft detection using AI models.

Implementation: Implement machine learning algorithms and methodologies for detecting electricity theft using smart meter data. The plan includes training and evaluating multiple classification models to determine the most effective approach.

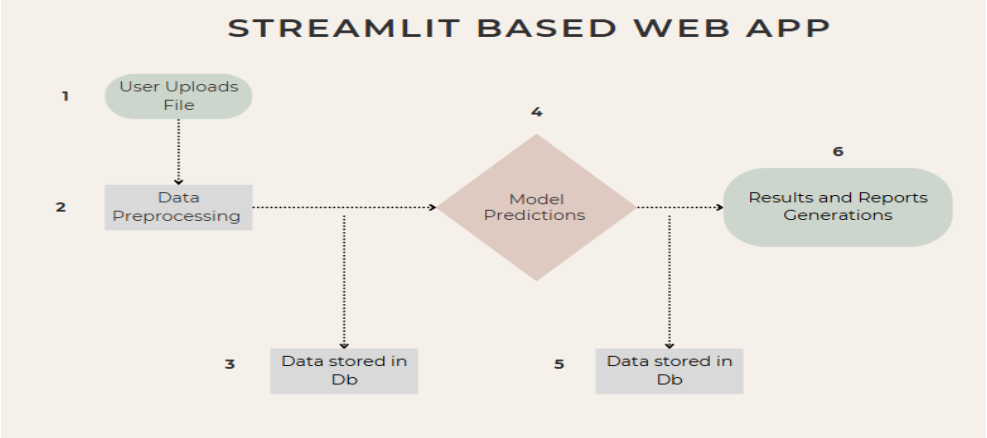

Figure 1.1: Basic Architecture of Electricity theft Detection web app.

- **Evaluation:** Evaluate the models performance through key metrics like Accuracy, Precision, Recall, F 1 score and AUC ROC curve. Evaluation metrics for Classification models are based on different respective activities. The most commonly used evaluation metrics for classification models are as below:

    - **Accuracy:** Calculates the proportion of correctly predicted cases, offering an overall performance view. Example: If 100 users are tested and 90 are correctly classified as stealer or non-stealer, accuracy is 90%.
    - **Precision:** Measures the accuracy of theft predictions by minimizing false alarms. Example: Out of 50 flagged users, if 45 are actual thieves, precision is high.
    - **Recall (Sensitivity):** Evaluates how well the model detects actual theft cases. Example: If 100 theft cases exist but only 70 are detected, recall is 70%.
    - **F1-Score:** Gives a balanced measure of precision & recall. Example: Required when theft cases are rare, balancing false positive and false negative.
    - **AUC ROC Curve:** Shows the model's ability to differentiate between theft and non-theft. Example: A high AUC indicates strong detection capabilities.
    - **Model Comparison:** Test models such as Decision Trees, Random Forests, SVMs, and Neural Networks. Evaluate performance across metrics to identify the most effective classifier.

- **Cross Validation:** Use different validation practices to assess model reliability & reduce overfitting.

- **Analysis:** Analyse the results to develop the most efficient model for electricity theft identification.

3.2 Requirements Resources

The research will require hardware as well as software stacks for experimentation and implementation as below:

3.2.1 Hardware Requirements

The hardware requirements for the electricity theft detection typically include:

- **The Processing Unit:** A robust CPU (e.g., Intel i7 or higher) or a dedicated GPU (e.g., NVIDIA RTX 3060) to handle intensive model training tasks.
- **Memory:** A minimum of 16 GB RAM to ensure smooth data processing and analysis.
- **Storage:** At least 1TB SSD to accommodate large datasets and store model results.
- **Network:** High-speed internet for efficient data uploads and real-time dashboard functionality.

3.2.2 Software Requirements

The software stack for training models includes:

- **Programming Language:** Python 3.10 or later, suitable for model development and data analysis.
- **Libraries:** Pandas and NumPy for data manipulation, Scikit-learn and TensorFlow for model building, and Matplotlib for visualization.
- **Development Tools:** Jupyter Notebook or PyCharm for streamlined coding and experimentation.
- **Database System:** MySQL or MongoDB to securely store user data and prediction outcomes.

- **Dashboard Tools:** Streamlit or Flask for an interactive web interface to display results.
- **Version Control:** Git and GitHub for collaborative coding and version management.

3.2.3 Dataset Specifications

- **Large Primary Dataset:** Smart meter electricity consumption data from utility companies, including timestamps, consumption patterns, and theft indicators.
- **Data Attributes:** Key attributes include User ID, theft labels, daily consumption values on daywise timestamps spanning multiple years.
- **Data Quality Checks:** Handle missing values through imputation, remove noisy data, and verify data types for consistency.

3.2.4 Model Training Details

Training a model involves several steps and considerations:

- Pre-process: Dataset is cleaned by filling missing values and normalizing it for consistency. Class imbalance is treated using methods like oversampling or undersampling, for balanced model training.
- Feature Engineering: Important features such as time-based consumption patterns, statistical measures, and unusual usage behaviors are extracted to improve the model's understanding.Hyperparameters: Tuning hyperparameters such as learning rate, batch size, and dropout rate is crucial for effective training.
- Data Splitting: The dataset is distributed in train, and test sets to create & assess the algorithim effectively.
- Model Selection: Mostly all the possible classification models like Random forest classifier, KNN classifier, Bagging and stacking Classifier, etc were trained.
- Evaluation Metrics: Model performance is measured using accuracy, precision, recall, and F1-score to ensure reliable results.
- Experimentation & Deployment: Cross-validation is applied to avoid overfitting, and hyperparameters are tuned for the best model performance and deployed.

4.1 Significance of the Study

This study is important because it helps detect electricity theft using AI, reducing financial losses for power companies. It ensures fair billing by identifying users who may be stealing electricity. The use of smart meter data improves accuracy and reduces the need for manual inspections. A real-time dashboard allows quick detection, helping companies take action faster. Overall, this research supports better energy management and a more reliable power supply.

4.2 Scope of the Study

This study focuses on detecting electricity theft using AI-based classification models applied to smart meter data. It analyzes electricity consumption patterns to identify fraudulent activities and classifies users as "stealer" or "non-stealer." The research covers data preprocessing, feature selection, and model evaluation to ensure accurate detection. The system is designed for integration into a web-based dashboard, allowing utility companies to upload consumption data and receive theft detection reports. The study does not cover hardware-based meter tampering detection but focuses solely on data-driven analysis for theft identification.

5. Background

Recently, Electricity theft is a serious problem affecting power distribution companies worldwide. It happens when people or businesses manipulate electricity meters or illegally connect to power lines to use electricity without paying. This leads to major financial losses for utility companies and can also disrupt the power supply, causing equipment failures and voltage fluctuations.

Traditionally, power companies have relied on physical inspections and customer complaints to detect electricity theft. However, these methods are slow, expensive, and not always

effective. With the introduction of smart meters, electricity usage data is now collected in real-time, creating opportunities for better monitoring and fraud detection.

Artificial Intelligence (AI) and Machine Learning (ML) offer advanced solutions to this challenge. By analyzing historical electricity consumption, AI models can detect unusual usage patterns that may indicate theft. These models can learn and improve over time, making them more accurate than traditional methods.

This research focuses on developing an AI-powered classification model that can identify electricity theft based on smart meter data. By implementing this model in a real-time monitoring system, power companies can detect theft faster, reduce financial losses, and ensure fair billing for all customers. The study also compares different ML techniques to search the most effective approach for theft detection. This research puts light on the role of AI ML in making electricity distribution more secure and efficient.

6. Abstract

Electricity theft is a critical issue in power distribution, leading to financial losses and grid inefficiencies. Traditional detection methods rely on manual inspections and outdated statistical techniques, making them ineffective and resource-intensive. This study presents an AI-based classification model that analyzes smart meter data to detect electricity theft accurately. Using supervised machine learning, the model classifies consumers as either "stealer" or "non-stealer" based on their usage patterns.

The system is trained on historical consumption data and deployed via a web-based dashboard for real-time monitoring. Users can upload electricity consumption records, and the model generates reports highlighting suspicious activity. This proactive approach enhances detection accuracy, enabling utility companies to take timely corrective actions. The study evaluates various machine learning models, including decision tree classifier, random forest classifier,

voting, bagging & boosting classifier models , comparing their effectiveness using, recall, precision , accuracy & F1-score.

By integrating AI into electricity theft detection, power distribution companies can reduce revenue losses, enhance grid stability, and ensure fair billing. This research demonstrates how AI-driven analytics can transform theft detection, optimize resource allocation, and contribute to the sustainability of the energy sector.

7. Research Plan

Below is detailed plan of research for 20 weeks from January 14th, 2025 i.e date of topic approval.

- **Research Proposal Initiate**
  - **Week 1:** Research and prepare objectives, its scope with methodology.
  - **Week 2:** Prepare the research proposal, and literature work.
- **Proposal Feedback**
  - **Week 3:** Feedback from the supervisor
  - **Week 4:** Update the correction suggested by Supervisor
- **Data Gathering**
  - **Week 5,6,7:** Gather the data through various online platforms like Kaggle.
  - **Week 8:** Start data analysis by appropriate statistical techniques.
- **Dataset evaluation**
  - **Week 9,10:** Work with various experiments for better results
  - **Week 11,12:** Prepare the results part of research proposal.
- **Working with documentation**
  - **Week 13,14:** Prepare the documentation as required.
  - **Week 15,16:** Crosscheck the documentation again.
- **Prepare for Presentation**
  - **Week 17:** work with the presentation part.
  - **Week 18:** Practice the presentation for conference.
- **Ppt finalization**

- **Week 19:** Open up with the findings from research with audience.
- **Week 20:** Finalize research paper with reviews from supervisor and submit for evaluation.

Here is Gantt chart of the project timeline:

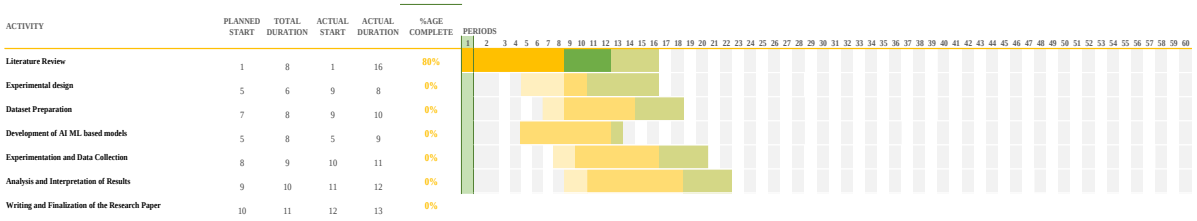

Figure 2. Project Plan

This plan provides a structured timeline for completing each phase of the research process within 20 weeks, ensuring adequate time for proposal development, data collection, analysis, writing, presentation, and finalization. Adjustments can be made as needed depending upon the particular requirement of the research project.

8. References

1. Dimf, G. P. ., Kumar , P. ., & Manju, V. N. . (2023). An Efficient Power Theft Detection Using Modified Deep Artificial Neural Network (MDANN). International Journal of Intelligent Systems and Applications in Engineering, 11(1), 01–11. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/2437

2. Ejaz Ul Haq, Can Pei, Ruihong Zhang, Huang Jianjun, Fiaz Ahmad, Electricity-theft detection for smart grid security using smart meter data: A deep-CNN based approach, Energy Reports, Volume 9, Supplement 1, 2023, Pages 634-643, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2022.11.072.

3. Joey Li, Munur Sacit Herdem, Jatin Nathwani, John Z. Wen, Methods and applications for Artificial Intelligence, Big Data, Internet of Things, and Blockchain in smart energy management, Energy and AI, Volume 11, 2023, 100208, ISSN 2666-5468, https://doi.org/10.1016/j.egyai.2022.100208

4. Jiang Z., Lin R., and Yang F., A hybrid machine learning model for electricity consumer categorization using smart meter data, Energies. (2018) 11, no. 9, 2235–2319, https://doi.org/10.3390/en11092235, 2-s2.0-85053804566.

5. Bhattarai B. P., Paudyal S., Luo Y., Mohanpurkar M., Cheung K., Tonkoski R., Hovsapian R., Myers K. S., Zhang R., Zhao P., Manic M., Zhang S., and Zhang X., Big data analytics in smart grids: state-of-theart, challenges, opportunities, and future directions, IET Smart grid. (2019) 2, no. 2, 141–154, https://doi.org/10.1049/iet-stg.2018.0261.

6.Yem Souhe F. G., Mbey C. F., Foba Kakeu V. J., Meyo A. E., and Boum A. T., Optimized forecasting of photovoltaic power generation using hybrid deep learning model based on GRU and SVM, Electrical Engineering. (2024) 1–20, https://doi.org/10.1007/s00202-024-02492-8.

7.Adil, M., Javaid, N., Qasim, U., Ullah, I., Shafiq, M., and Choi, J. G. (2020). LSTM and bat-based RUSBoost approach for electricity theft detection. Appl. Sci. 10 (12), 4378. doi:10.3390/app10124378

8.Ahir, R. K., and Chakraborty, B. (2022). Pattern-based and context-aware electricity theft detection in smart grid. Sustain. Energy, Grids Netw. 32, 100833. doi:10.1016/j.segan.2022.100833

9.Alameady, M. H., Albermany, S., and George, L. E. (2022). Energy theft detection and preventive measures for IoT using machine learning. Math. Statistician Eng. Appl., 7, 155–168.