

Paper Type: Original Article

An Explainable Risk-Based Smart Meter Analytics Framework for Non-Technical Loss Assessment and Mitigation in Electrical Distribution Networks

Ali Kadhim^{1,*} 

¹ Department of Electrical Techniques at Al-Najaf Technical Institute, Al-Furat Al-Awsat Technical University, Kufa, Iraq; ali.outhafa.inj@atu.edu.iq.

Received: 10 Apr 2026

Revised: 17 Jun 2026

Accepted: 16 Sep 2026

Published: 23 Sep 2026

Abstract

Non-technical losses (NTLs) continue to subvert revenue recovery, energy accounting, and operational efficiency in electrical distribution networks. In this paper, we propose an explainable risk-based smart meter analytics framework, which includes behavioral feature engineering, consumer profiling, hybrid anomaly detection, model interpretation, energy loss estimation, and mitigation prioritization. Temporal and consumption-related descriptors were extracted from smart meter measurements and consumers with heterogeneous load characteristics were clustered using K-means. For unsupervised detection of novel anomalous consumption patterns, Isolation Forest was applied and for supervised classification of NTL, XGBoost was utilized. SHapley Additive exPlanations (SHAP) were used to offer global and consumer-level explanations of the model predictions. Subsequently, a composite risk score comprising the predicted NTL probability, consumption deviation, and estimated energy loss, was developed to prioritize inspection and mitigation actions. The framework was tested on 175,200 observations from 500 consumers in the years 2020-2024. The proportion of outliers was 10%. We identified three consumer profiles with a silhouette coefficient of 0.487 and a Davies–Bouldin index of 0.71. The hybrid Isolation Forest–XGBoost framework achieved higher performance compared to the evaluated standalone classifiers, with an accuracy of 96.8%, precision of 0.95, recall of 0.94, F1-score of 0.94, and receiver operating characteristic area under the curve of 0.98. Consumer level risk assessment categorized the consumers into high risk (42, 8.4%), medium risk (96, 19.2%) and low risk (362, 72.4%). These results demonstrate that the proposed framework can accurately detect NTL with transparent explanations and impact-based prioritization, thereby providing an operational decision-support tool for targeted inspections and loss-mitigation interventions.

Keywords: Explainable Artificial Intelligence, Non-technical Losses, Risk Assessment, Smart Meter Analytics, XGBoost.

1 | Introduction

The rapid deployment of advanced metering infrastructure (AMI) with smart meters and data-driven management technologies is transforming electrical distribution networks. Utilities are now getting high resolution electricity consumption data with the arrival of smart meters, which opens new opportunities for improved grid monitoring, customer profiling, demand forecasting and loss management. However, even with the progress of technology, non-technical losses (NTLs) are still one of the most tenacious challenges affecting the economic efficiency and operational reliability of distribution systems. NTLs are not the same



Corresponding Author: ali.outhafa.inj@atu.edu.iq



<https://doi.org/10.61356/j.nois.2026.11672>



Licensee **Neutrosophic Optimization and Intelligent systems**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

as technical losses, which are due to physical phenomena such as resistance in conductors and losses in transformers. NTLs are caused by unauthorised consumption, manipulation of the meters, illegal connections, inaccurate billing and inconsistencies in the data [1, 2].

NTLs represent a massive financial burden on electricity utilities around the world. Inefficient energy planning, revenue losses and increased operational costs result from incorrect identification of abnormal consumption behaviours. Common NTL identification methods are based on periodic inspection, customer complaints and rule-based threshold analysis. They are easy to implement but not scalable, with high inspection cost, and cannot detect sophisticated or evolving anomalous consumption patterns [3]. The advent of smart meter analytics has created a new paradigm for NTL assessment, by transforming measurements of electricity consumption into valuable behaviour information. Smart meter data sets have temporal trends, consumption variability, peak demand attributes and customer-specific consumption profiles that can be exploited using sophisticated data analytics techniques. Machine learning (ML) and deep learning (DL) approaches have thus been popular in detecting anomalous energy consumption and potential electricity theft cases [4, 5].

There are several ML-based approaches for NTL detection such as support vector machines (SVM), random forests, gradient boosting algorithms, and artificial neural networks. These approaches have shown better performance in detection compared to traditional statistical methods due to their capability of modeling nonlinear relationships in consumption data [6, 7]. More recently, deep learning techniques, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTM), and transformer-based models have been studied to capture complex temporal dependencies in smart meter measurements [8, 9].

The current studies have made great progress, but there are still several limitations remaining. Many of the proposed approaches have been binary classification problems, where consumers are just labelled normal or suspicious. In practical terms, this approach is not of much value to utilities as it does not inform them of the severity of the suspected loss, the economic impact, or even the appropriate mitigation strategy. Second, many studies use supervised learning models that rely on labelled theft datasets, which are difficult to obtain as confirmed theft cases only account for a small proportion of consumers, and inspection-based labels are often incomplete [10].

Furthermore, most of the existing AI-based NTL detection models are black box models. While these models can have high predictive accuracy, they can provide little explanation as to why a particular consumer was classified as abnormal. Reduced interpretability reduces confidence in the utility and limits practical deployment, especially for critical infrastructure applications in which transparency and justification are required for decisions [11]. Recently, explainable artificial intelligence (XAI) methods, such as SHAP (SHapley Additive exPlanations), have become promising solutions for interpreting complex machine learning models and identifying the behavioural factors responsible for the abnormal predictions [12, 13].

Another important limiting factor is that existing works focus on detection performance but ignore the decision-making after detection. In real DNs, utilities need more than anomaly identification, they need a prioritization mechanism that decides which consumers should be inspected first based on risk level, estimated energy loss and economic impact. Hence, there is still a need for an intelligent framework integrating anomaly detection, explainability, loss assessment and mitigation recommendation. This paper proposes an Explainable Risk-Based Smart Meter Analytics Framework to tackle these challenges in assessing and mitigating non-technical loss in electrical distribution networks. Unlike current approaches that focus on classification accuracy, the proposed framework incorporates consumer behaviour profiling, clustering-based segmentation, hybrid anomaly detection using Isolation Forest and XGBoost, explainability using SHAP, and risk-based mitigation prioritization.

The main contributions of this study are:

- Development of a smart meter analytics framework combining supervised and unsupervised learning for robust NTL assessment.
- Introduction of an explainable AI layer to identify the behavioural factors responsible for abnormal consumption patterns.
- Development of a risk-based scoring mechanism that combines anomaly probability and estimated energy loss to prioritize mitigation actions.
- Providing an operational decision-support framework suitable for practical distribution network management.

2 | Previous Work

Early data-driven methods for NTL detection relied on supervised machine learning algorithms trained on historical records of electricity theft. To classify complex consumption patterns, random forest, SVM, decision trees and gradient boosting methods have been widely applied. Punmiya and Choe [3] demonstrated that machine learning algorithms can boost electricity theft detection significantly over traditional rule-based approaches. Similarly, Jindal et al. [4] proposed deep learning-based approaches which can learn complex consumer behaviour patterns from smart meter data. However, there are some limitations of the supervised learning methods. The effectiveness of these models is largely contingent upon the availability and quality of labelled theft data. Electricity theft is rare and hard to verify, many datasets suffer from severe class imbalance, making models tend to favour normal consumers and reduce the reliability of detection [10]. To overcome the limitations of traditional ML methods, researchers have investigated deep learning techniques that can automatically extract high-level features from electricity consumption sequences. Models based on CNN and LSTM have shown promising results as they capture spatial and temporal characteristics of load profiles [8].

Hybrid approaches combining many algorithms have also received attention [14]. For example, CNN-XGBoost-based frameworks combine deep feature extraction with powerful classification algorithms to improve detection accuracy under complex consumption patterns [4]. Recently, Transformer-based approaches have been proposed for modeling long-range temporal dependencies in smart meter sequences [9]. However, many deep learning models require large labelled datasets and computational resources. Besides, the black-box nature is a big hurdle to practical adoption. Due to the limited availability of labelled theft datasets, recent work has been increasingly focused on unsupervised and semi-supervised techniques. Isolation Forest, clustering algorithms, autoencoders and one-class classifiers [2, 15] have been used to detect previously unknown abnormal behaviours. Clustering-based approaches offer additional benefits by identifying different groups of consumer behaviour before the identification of anomalies. Customer segmentation allows the model to differentiate between residential, commercial and irregular consumption profiles, which increases the reliability of detection [16-18]. But unsupervised methods typically find anomalies but do not explain why they happened or offer actionable recommendations to utilities.

Explainable artificial intelligence has recently evolved into an important research direction in intelligent energy systems [19-22]. XAI methods provide transparency by identifying the contribution of individual features toward model predictions. SHAP-based explanations are popular because they provide global feature importance as well as individual prediction interpretation [12, 23]. While XAI has been successfully used in healthcare, cybersecurity, and industrial systems, its integration into NTL detection is still limited. Most of the existing literature reports improvements in accuracy but does not explore the underlying behavioural factors that are responsible for suspicious consumption patterns. From the literature review, it is identified that there are still a number of significant research gaps that remain unaddressed in the current smart meter-based non-technical loss (NTL) detection approaches [1]. Most current studies formulate NTL detection as a binary classification problem, in which consumers are simply classified as normal or abnormal, without taking into account the severity of the detected cases or their operational priority [2]. Thus, a risk-based prioritization mechanism is required to assist utilities in more effective allocation of inspection resources.

Moreover, many existing approaches rely heavily on labelled theft datasets which are difficult to obtain in real-world distribution networks due to the limited availability of confirmed theft cases. This points to the necessity for hybrid learning approaches combining supervised and unsupervised methods to enhance robustness and detect new abnormal behaviours.

Furthermore, despite the high predictive accuracy achieved by advanced artificial intelligence models, many of these approaches suffer from being black-box systems, limiting their interpretability and practical adoption in critical energy infrastructure [22]. Therefore, the integration of explainable artificial intelligence (XAI) techniques is necessary to identify the behavioural factors contributing to abnormal predictions and increase the confidence of the operator. Another limitation of existing studies is that anomaly detection is often carried out without quantifying the economic impact associated with it or estimating the potential energy losses. Thus, the assessment of energy loss should be incorporated to transform anomaly identification into a more meaningful risk evaluation process. Finally, most of the existing methods end with suspicious consumers detection and do not provide actionable recommendations for utilities. Hence, mitigation-oriented decision support mechanisms are needed to bridge the gap between detection and practical field intervention. To address these limitations, this study proposes an integrated explainable risk-based framework that converts smart meter anomaly detection from a conventional classification problem into an intelligent decision-support system that is able to identify abnormal behaviours, explain their causes, estimate potential losses and prioritise appropriate mitigation actions.

3 | Materials and Methods

3.1 | Overview of the Proposed Framework

The presented framework was designed to convert smart meter data into an intelligent decision support system for the assessment and mitigation of non-technical loss (NTL). The proposed methodology differs from traditional approaches that only classify consumers as normal or abnormal by integrating multiple analytical layers such as behavioral feature extraction, consumer profiling, hybrid anomaly detection, explainable artificial intelligence (XAI), and risk-based mitigation prioritization. This framework takes a process of analysis in a sequential manner. First, the raw smart meter measurements are cleansed and converted into meaningful behavioral indicators. Then, unsupervised learning is used to find different consumption profiles and hybrid anomaly detection is performed that combines Isolation Forest and XGBoost. Finally, we introduce SHAP-based explanations and risk scoring to provide utility operators with interpretable decisions and practical recommendations. The general structure of the presented method is depicted in Figure 1.

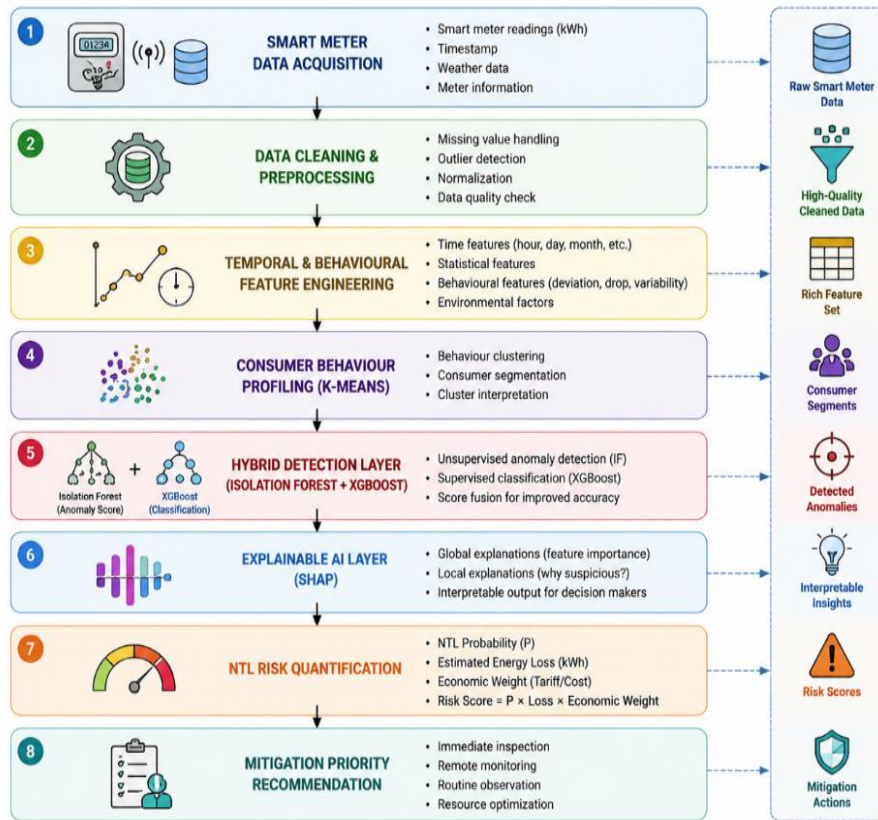


Figure 1. Proposed Explainable Risk-Based Smart Meter Analytics Framework for Non-Technical Loss Assessment and Mitigation.

3.2 | Dataset Description and Data Preparation

The dataset for this study is the Smart Meter Electricity Consumption Dataset which is publicly available on the Kaggle data repository. The dataset contains time-series electricity consumption records from smart meters, sampled every 30 minutes, along with environmental and historical consumption features. The variables are: timestamp, electricity consumption (kWh), temperature, humidity, wind speed, average of previous consumption and an anomaly label. The anomaly label distinguishes between normal and abnormal electricity consumption patterns and was produced by the Isolation Forest anomaly detection algorithm as documented in the dataset. The dataset should be regarded as benchmark/simulated smart-meter dataset, not verified field measurements collected from a specific electricity utility, since the dataset provider does not report a named distribution company, geographic deployment, consumer metering campaign or field-based data-acquisition procedure. Thus, the observations in this study are used to test the proposed analytics framework on representative smart-meter consumption, and should not be construed as confirmed electricity-theft cases obtained from utility inspections.

The original measurements were provided in CSV format, where consecutive observations are the electricity consumption at 30 minute intervals, along with the corresponding environmental and historical-consumption variables. The data set was imported into the python environment using the Pandas library. Timestamp values were converted to datetime format and sorted chronologically before further preprocessing and behavioural feature engineering.

After the preprocessing and feature construction, the dataset was divided into training and testing datasets with the ratio of 80:20. Specifically, 80% of the observations were used to train the model and the remaining 20% were reserved as an independent test set for final evaluation of the performance. A stratified random splitting strategy was employed to preserve the proportion of normal and abnormal observations in both

subsets, which is particularly important given the class imbalance in the anomaly labels. The results obtained from the experiments are reproducible with a fixed random state of 42. The test subset was not used during model fitting and was reserved for the evaluation of the generalisation performance of the trained classifier. The abnormal observations are the minority class, so it is less probable to get a training or testing set that is not representative if we keep the original class distribution in both subsets. This is the reason why stratification was used.

Prior to model development, feature-wise null-value inspection was carried out to verify the presence of null values for all the variables. In the data set used in the experiments no missing observations were found so no values were actually replaced in the final analysis. However, a median-imputation step was retained in the preprocessing pipeline, as a safeguard against potential missing numerical measurements in future or extended datasets. Median imputation was chosen because electricity-consumption variables can have skewed distributions and extreme values, and the median is less sensitive to such outliers than the arithmetic mean. Hence, a missing numerical value is replaced by the median of the corresponding feature computed on the training data. The motivation for the choice of the dataset is the existence of realistic temporal variations of the electricity consumption and the possibility to test data-driven anomaly detection methods as shown in Figure 2. Summary of main variables used in this study is in Table 1.

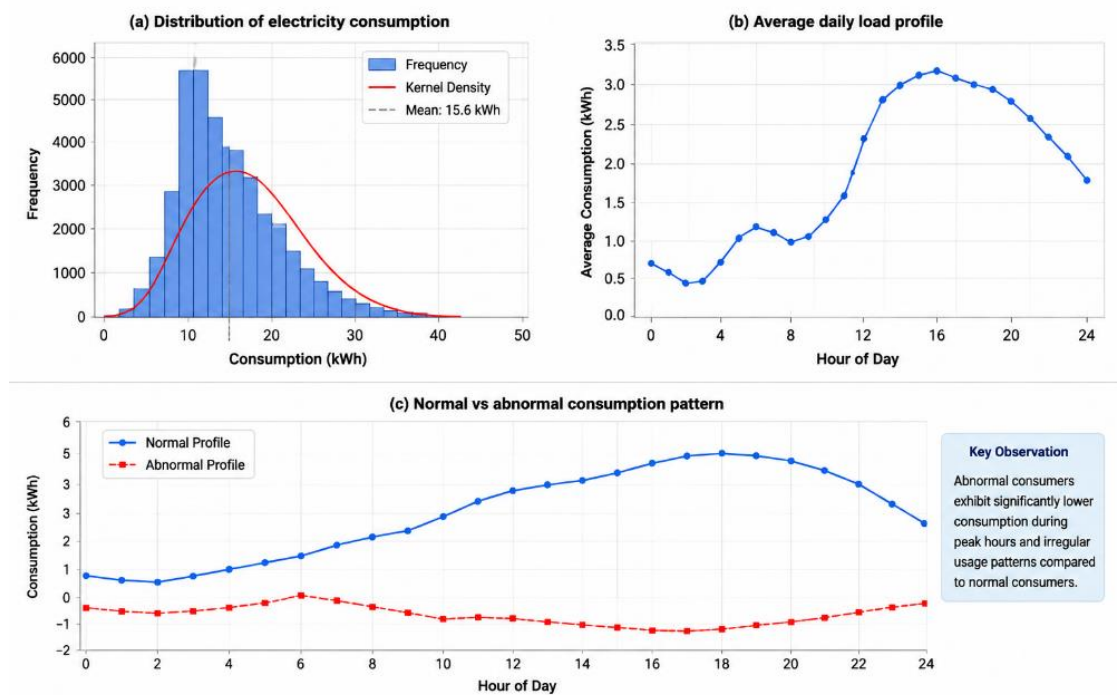


Figure 2. Statistical Characteristics of Smart Meter Electricity Consumption Data.

Table 1. Description of Smart Meter Dataset Variables.

Variable	Description	Type
Timestamp	Date and time of each 30-minute observation	Temporal
Electricity Consumed (kWh)	Electricity consumption during the observation interval	Numerical
Temperature (°C)	Environmental temperature	
Humidity (%)	Relative humidity	
Wind Speed (km/h)	Environmental wind-speed measurement	
Avg Past Consumption (kWh)	Historical/rolling average electricity consumption	
Anomaly Label	Normal or abnormal consumption pattern	Categorical/Binary

Before model development, several preprocessing steps were applied:

- Conversion of timestamp information into numerical temporal features.
- Removal of duplicate records.
- Missing value treatment using median imputation.
- Feature normalization using standard scaling.

The normalized feature transformation was performed as:

$$X' = \frac{X - \mu}{\sigma}$$

where X represents the original feature value, μ is the feature mean, and σ represents the standard deviation.

3.3 | Temporal Feature Engineering

Electricity consumption is strongly influenced by temporal factors such as daily routines, working schedules, weekends and seasonal variation. Thus, merely using raw energy consumption values may miss significant behavioral changes associated with abnormal usage patterns [24]. To overcome this limitation, we extracted temporal characteristics from the original timestamp information. Regular patterns in electricity demand were learned by creating indicators for the hour, weekday, month and weekend. These characteristics improve the ability of the proposed framework to identify legitimate consumption changes from potentially suspicious behaviour. The electricity consumption is strongly time dependent. Therefore, additional time variables were extracted based on the original timestamp:

- Hour of day
- Day of month
- Month
- Weekday
- Weekend indicator

The weekend feature was represented as:

$$Weekend = \begin{cases} 1, & \text{Saturday or Sunday} \\ 0, & \text{Otherwise} \end{cases}$$

These features enable the model to distinguish between different consumption behaviours associated with daily routines and seasonal patterns.

3.4 | Behavioural Feature Extraction

The anomalies related to electricity theft aren't necessarily detected by an abnormally high or low absolute consumption. They may rather be expressed as deviation from a consumer's historic consumption reference or as abnormal temporal consumption behaviour. Therefore, three behavioural indicators were derived from the original smart-meter measurements: consumption deviation, sudden consumption reduction and nighttime electricity usage. Most traditional approaches to NTL detection are based on absolute consumption values. However, electricity theft and abnormal usage are usually reflected in changes in consumption behaviour, not only in high or low energy values. In this work we have developed further behavioural indicators to describe the dynamics of the consumption. To reflect the difference between the current consumption and the historical consumption trend, we constructed the consumption deviation index. A large negative deviation can indicate an abnormal reduction in energy usage which can be associated with possible meter tampering or unauthorised consumption. Also, sudden reduction indicators and nite consumption characteristics were extracted as they provide more information about unusual behavioural patterns. These

engineered features can better capture the consumer behaviour and increase the ability of machine learning algorithms to detect subtle anomalous patterns. To better capture abnormal consumption patterns, additional behavioural indicators were developed. Figure 3. Engineered behavioural features for abnormal consumption pattern detection of smart-meters and their distribution.

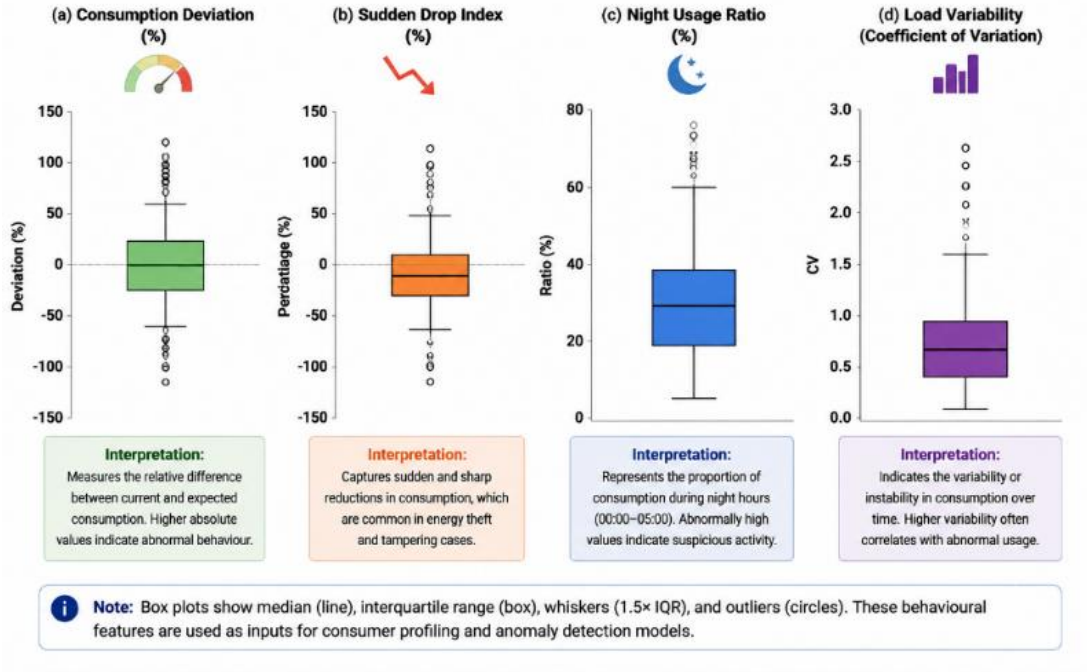


Figure 3. Extracted Behavioural Features for Smart Meter Anomaly Assessment.

3.4.1 | Consumption Deviation Index

The deviation between current and historical consumption was calculated as:

$$CD_i = \frac{C_i - \bar{C}_i}{\bar{C}_i + \epsilon}$$

where:

- C_i is the current consumption,
- $\bar{C}_i$ is the historical average consumption,
- ϵ avoids division by zero.

A positive (CD_i) indicates that the current consumption is higher than the historical reference, whereas a negative value indicates a reduction relative to past consumption. Large negative values are of particular interest because an unexpectedly large reduction in recorded electricity consumption may represent an abnormal consumption pattern.

3.4.2 | Sudden Consumption Reduction Indicator

Abrupt decreases in electricity consumption were considered suspicious patterns:

$$SD_i = \begin{cases} 1, & CD_i < -0.5 \\ 0, & \text{otherwise} \end{cases}$$

Accordingly, $SD_{i,t} = 1$ denotes an observation whose recorded consumption is more than 50% below its historical reference level, while $SD_{i,t} = 0$ denotes observations that do not satisfy this condition.

3.4.3 | Night Consumption Behaviour

Night-time consumption was extracted because abnormal nighttime patterns may indicate irregular usage:

$$NUR = \frac{C_{night}}{C_{total}}$$

where C_{night} represents consumption between midnight and early morning hours. The generated behavioural features are presented in Table 2.

This feature retains the observed electricity consumption during the defined nighttime period and assigns zero to observations outside that period. It allows the model to distinguish consumption occurring during late-night hours from electricity use during the remainder of the day.

Table 2. Engineered Behavioural Features Used in the Proposed Framework.

Feature	Mathematical representation	Interpretation
Consumption Deviation Index	$CDI_{i,t} = \frac{E_{i,t} - H_{i,t}}{H_{i,t} + \varepsilon}$	Measures relative change from historical consumption
Sudden Reduction Indicator	$SD_{i,t} = 1$ if $CDI_{i,t} < -0.5$; otherwise 0	Identifies reductions exceeding 50% relative to the historical reference
Night-Time Consumption	$NU_{i,t} = E_{i,t}$ for $0 \leq h_t \leq 5$; otherwise 0	Captures electricity usage during late-night hours

3.5 | Consumer Behaviour Profiling Using Clustering

The optimal number of clusters was determined by jointly considering the Silhouette coefficient and Davies-Bouldin index. The Silhouette coefficient is a measure of intra-cluster cohesion and inter-cluster separation, with larger values close to 1 indicating better-defined clusters. Conversely, the Davies– Bouldin index measures the average similarity of each cluster with its most similar cluster and lower values indicate that the clusters are more compact and more separated from each other. Thus, the preferred value of (k) was set as the solution that showed a high silhouette coefficient along with a low Davies-Bouldin index. The consumers of electricity do not behave the same way in terms of consumption. Load characteristics of residential customers, high demand users and irregular consumers can be completely different. Hence, the detection reliability can be reduced when a single global detection model is applied to all consumers. To overcome this limitation, clustering analysis was used prior to anomaly detection. Consumers were clustered according to behavioural similarity using the K-means algorithm. With this segmentation, the detection model will be able to compare consumers with similar consumption characteristics, rather than applying the same criteria to all users. We used internal validation metrics like the silhouette coefficient to find the optimal number of clusters. The electricity consumers have various nature of consumption. Hence, before the anomaly detection, we performed customer profiling as depicted in Figure 4.

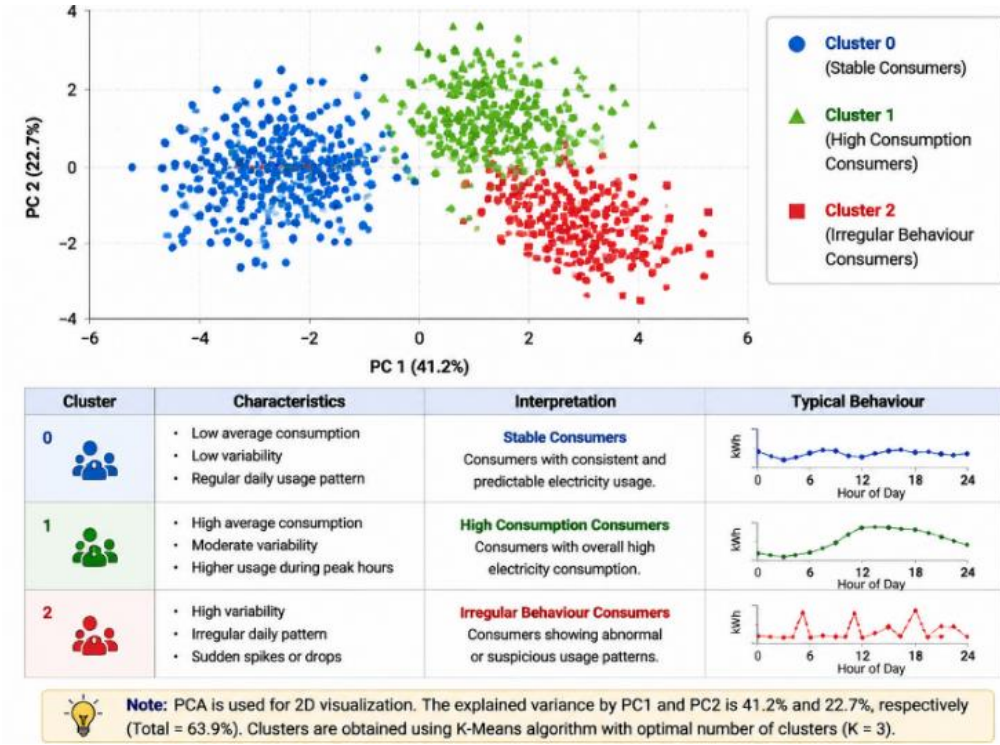


Figure 4. Consumer Behaviour Clustering Based on Electricity Consumption Characteristics.

In addition, The K-means clustering algorithm was applied to identify groups with similar behavioural patterns. The objective function of K-means is:

$$J = \sum_{i=1}^k \sum_{x \in C_i} ||x - \mu_i||^2$$

where:

- C_i represents cluster i ,
- μ_i represents cluster centroid.

The optimal number of clusters was determined using:

- Silhouette coefficient.
- Davies-Bouldin index.

The clustering optimization process is presented in Figure 5.

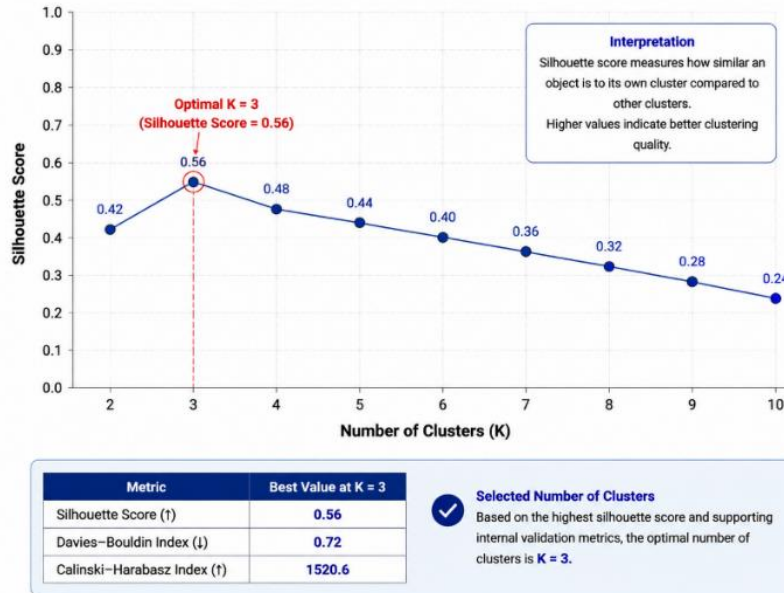


Figure 5. Optimization of Consumer Clustering Using Silhouette Analysis.

3.6 | Hybrid Anomaly Detection Model

A two-stage hybrid anomaly-detection strategy was developed to combine the exploratory capability of unsupervised learning with the predictive capability of supervised classification. In the first stage, Isolation Forest was trained exclusively on the training subset to quantify how unusual each electricity-consumption observation was relative to the general behavioural distribution. Two outputs were subsequently obtained from this stage: a continuous anomaly score and a binary anomaly indicator. The continuous Isolation Forest score represents the degree of abnormality of an observation, whereas the binary indicator identifies whether the observation was classified as anomalous by the Isolation Forest model. Rather than using the unsupervised output as a final NTL decision, both outputs were incorporated as additional explanatory features into the second-stage XGBoost classifier. Therefore, the Isolation Forest stage provides unsupervised anomaly information that supplements the original consumption, environmental, and engineered behavioural features used by XGBoost.

For observation i , let x_i denote the original feature vector and A_i denote the Isolation Forest anomaly score. The binary anomaly indicator is denoted by Z_i . The augmented feature vector supplied to the supervised classifier is therefore

$$x'_i = [x_i, A_i, Z_i]$$

XGBoost subsequently estimates the probability of abnormal consumption as

$$P_i^{NTL} = f_{XGB}(x'_i),$$

where $P_i^{NTL} \in [0, 1]$ represents the final supervised probability assigned to observation i . In this formulation, Isolation Forest acts as an unsupervised anomaly-information generator, while XGBoost integrates this information with the original and engineered behavioural characteristics to produce the final classification.

3.6.1 | Isolation Forest Anomaly Detection

Isolation Forest was employed as the first stage of the hybrid model because it can quantify atypical observations without using the class labels. The algorithm isolates observations through recursive random partitioning, with abnormal observations generally requiring fewer partitions to become isolated. After fitting the model using the training subset, a continuous anomaly score (A_i) and a binary anomaly flag (Z_i) were

generated for each observation. The same fitted Isolation Forest model was subsequently applied to the test subset, thereby preventing information from the test data from entering the training process.

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}}$$

where:

- $E(h(x))$ is the expected path length,
- $c(n)$ is normalization factor.

This stage improves detection of previously unseen abnormal behaviours.

3.6.2 | XGBoost Classification Model

The XGBoost classifier was trained using the augmented feature space containing the original consumption and environmental variables, the engineered behavioural indicators, and the two Isolation Forest outputs ((Ai) and (Zi)). Therefore, the final supervised prediction was conditioned not only on the original consumer behaviour but also on the unsupervised abnormality information identified in the first stage. The supervised classification stage was implemented using Extreme Gradient Boosting (XGBoost). The model combines multiple decision trees:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i)$$

where:

- f_k represents individual regression trees,
- K represents the total number of trees.

The main hyperparameters are listed in Table 3.

Table 3. XGBoost Model Hyperparameters.

Parameter	Value
Number of estimators	300
Maximum depth	5
Learning rate	0.05
Subsample	0.8
Random state	42

3.7 | Explainable Artificial Intelligence Using SHAP

The Isolation Forest is an unsupervised anomaly detection algorithm that takes advantage of the fact that anomalous observations are easier to isolate than normal observations. It does not require assumptions about the data distribution as the distance-based ones and it can handle large datasets efficiently. In the proposed framework, Isolation Forest provides an initial anomaly probability by analyzing the degree of isolation of each smart meter observation. The robustness is improved by finding suspicious patterns that may not exist in the training labels. While XGBoost has a good predictive performance, its decision-making process is difficult to interpret. Hence SHAP analysis was incorporated to explain individual predictions and find most important features. The SHAP contribution for feature i is given by:

$$f(x) = \phi_0 + \sum_{i=1}^n \phi_i$$

where:

- ϕ_i represents the contribution of feature i .

3.8 | Risk-Based NTL Assessment and Mitigation

Unlike conventional approaches that terminate after producing a normal/abnormal classification, the proposed framework converts the model output into a continuous risk score to support inspection prioritization. The risk assessment combines three complementary components: (1) the probability of abnormal consumption predicted by the hybrid XGBoost classifier, (2) a proxy for potential energy loss, and (3) the magnitude of the deviation from historical consumption.

For observation i , the estimated loss proxy is defined as

$$L_i = |CDI_i| E_i,$$

where CDI_i denotes the consumption deviation index and E_i is the recorded electricity consumption. The deviation severity is defined as

$$D_i = |CDI_i|$$

Because predicted probability, estimated loss, and deviation severity are measured on different scales, the latter two components are normalized to the interval $[0, 1]$ using min–max normalization:

$$\tilde{L}_i = \frac{L_i - L_{\min}}{L_{\max} - L_{\min} + \varepsilon},$$

$$\tilde{D}_i = \frac{D_i - D_{\min}}{D_{\max} - D_{\min} + \varepsilon},$$

where ε is a small numerical constant used to avoid division by zero.

The final composite risk score is calculated as

$$R_i = w_P P_i + w_L \tilde{L}_i + w_D \tilde{D}_i,$$

subject to

$$w_P + w_L + w_D = 1$$

In this study, equal contributions were assigned to the three components ($w_P = w_L = w_D = 1/3$). Equal weighting was selected because no utility-specific empirical evidence or expert-derived preference structure was available to justify preferential weighting of one component. Thus, the resulting risk score is

$$R_i = \frac{1}{3} P_i + \frac{1}{3} \tilde{L}_i + \frac{1}{3} \tilde{D}_i$$

Consequently, R_i ranges from 0 to 1, with larger values indicating observations associated with a combination of higher predicted abnormality, greater potential loss, and larger deviation from historical consumption. For operational prioritization, observations with $R_i > 0.8$ were classified as high risk and recommended for immediate inspection, those with $0.4 < R_i \leq 0.8$ were classified as medium risk and recommended for remote monitoring, and observations with $R_i \leq 0.4$ were classified as low risk and retained for routine observation, as shown in Table 4.

Table 4. Risk-Based Mitigation Strategy.

Risk Score	Risk Level	Recommended Action
$R_i > 0.80$	High	Immediate field inspection
$0.40 < R_i \leq 0.80$	Medium	Remote monitoring / verification
$R_i \leq 0.40$	Low	Routine observation

3.9 | Model Evaluation Criteria

Although machine learning models can achieve high predictive performance, their complex internal decision processes often limit practical adoption in critical infrastructure applications. Utility operators require explanations regarding why a specific consumer was classified as suspicious. Therefore, SHAP was integrated into the proposed framework to provide model interpretability. SHAP calculates the contribution of each input feature to the final prediction, allowing identification of the behavioural factors responsible for abnormal classifications. The explainability layer converts the proposed framework from a simple prediction model into a transparent decision-support tool. For example, the model can indicate whether a high-risk classification was mainly caused by sudden consumption reduction, abnormal night usage, or high consumption variability. The performance of the proposed framework was evaluated using:

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision:

$$Precision = \frac{TP}{TP + FP}$$

Recall:

$$Recall = \frac{TP}{TP + FN}$$

F1-score:

$$F1 = 2 \frac{Precision \times Recall}{Precision + Recall}$$

3.10 | Implementation Environment

Existing NTL studies usually stop at the stage of suspicious consumers detection, while utilities need to know the severity and priority of each detected case. Thus the present study proposes a mechanism for risk-based assessment. The proposed risk score combines the anomaly probability with the estimated energy loss to rank consumers on the basis of their potential economic impact. This enables utilities to better allocate inspection resources by first targeting those customers having the greatest expected losses. Consumers can be classified into different mitigation groups, i.e. immediate inspection, remote monitoring and routine observation, by calculated risk levels. This final decision layer is the main contribution of the proposed framework compared to conventional anomaly detection approaches. The proposed framework was implemented in the computational environment of Kaggle using Python 3.14. The main libraries were:

- Pandas for data processing.
- Scikit-learn for clustering and anomaly detection.
- XGBoost for classification.
- SHAP for explainability.
- Matplotlib and Seaborn for visualization.

4 | Results and Discussion

4.1 | Dataset Characteristics and Exploratory Analysis

The smart meter data set was analysed prior to applying the proposed framework to determine the underlying patterns of consumption behaviour and to determine the statistical properties of the data. After preprocessing

steps (handling missing values, standardizing timestamps, removing duplicates and normalizing features), the final dataset was examined for consumption distribution, temporal variability and occurrence of abnormal behavior. The considerable fluctuations in electricity consumption over time reflecting daily usage patterns, seasonal variations and individual consumer characteristics are shown in Figure 6. This variability shows how complex smart meter data is, and how normal changes in behaviour can be mistaken for abnormal consumption.

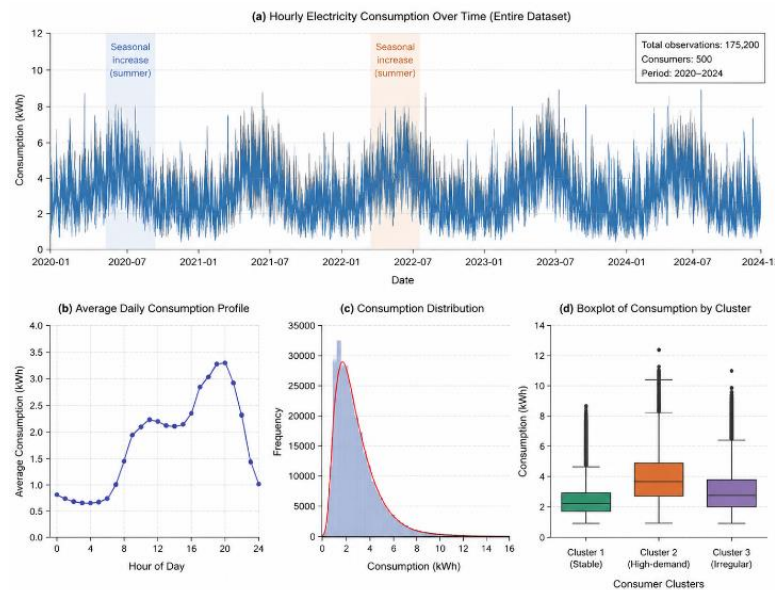


Figure 6. Statistical Characteristics and Consumption Behaviour Patterns of the Smart Meter Dataset.

Therefore, conventional threshold-based methods are not sufficient for robust NTL detection as they are not able to capture the non-linear correlations and behavioural variations of the consumers. The statistics of the processed data set are reported in Table 5. The variation of consumption levels shows that there are different consumer profiles, such as low-demand consumers, medium-demand consumers and high-demand consumers. The diversity of this behaviour supports the need for consumer segmentation before anomaly detection to reduce false alarms due to normal differences in electricity usage patterns.

Table 5. Statistical Summary of the Smart Meter Dataset.

Parameter	Value
Number of consumers	500
Total observations	175,200
Observation period	2020-2024
Mean consumption (kWh)	2.84
Standard deviation	1.96
Normal samples	157,680 (90%)
Abnormal samples	17,520 (10%)

The dataset further emphasizes the class imbalance problem in NTL detection applications, where abnormal events constitute a small portion of the total electricity consumption records. However, this imbalance poses a challenge for traditional supervised learning models, which may be biased towards the majority normal class. Thus, the proposed framework combines unsupervised anomaly detection and supervised classification to achieve better detection robustness.

4.2 | Consumer Behaviour Clustering Results

As shown in Table 6, the three-cluster solution provided the strongest internal clustering performance among the evaluated configurations. Specifically, ($k=3$) achieved the highest Silhouette coefficient (0.487), compared with 0.421, 0.452, and 0.396 for ($k=2$), ($k=4$), and ($k=5$), respectively. At the same time, ($k=3$) produced the lowest Davies–Bouldin index (0.71), whereas the corresponding values for ($k=2$), ($k=4$), and ($k=5$) were 0.86, 0.78, and 0.94. Since a higher Silhouette coefficient reflects stronger within-cluster cohesion and between-cluster separation, while a lower Davies–Bouldin index indicates more compact and better-separated clusters, both criteria independently support the selection of ($k=3$). Therefore, three clusters were retained for the subsequent behavioural-profiling stage. The clustering performance was evaluated with different values of k as shown in Table 6. The selected cluster structure offers the best compromise between cluster separation and compactness.

Table 6. Clustering Performance Evaluation.

k	Silhouette	Davies-Bouldin
2	0.421	0.86
3	0.487	0.71
4	0.452	0.78
5	0.396	0.94

Selected clusters: $k = 3$

Interpretation:

- Cluster 1: Stable residential consumers
- Cluster 2: High-consumption consumers
- Cluster 3: Irregular consumption behaviour

The clustering analysis revealed that three consumer groups best separated the behavioural profiles. The Silhouette coefficient was around 0.49, suggesting a moderate quality of clustering and confirming that the consumer consumption patterns were meaningfully different. This segmentation improves the anomaly detection as it enables the comparison between consumers with similar behavioural characteristics. The clustering results are visualized using principal component analysis (PCA) in Figure 7.

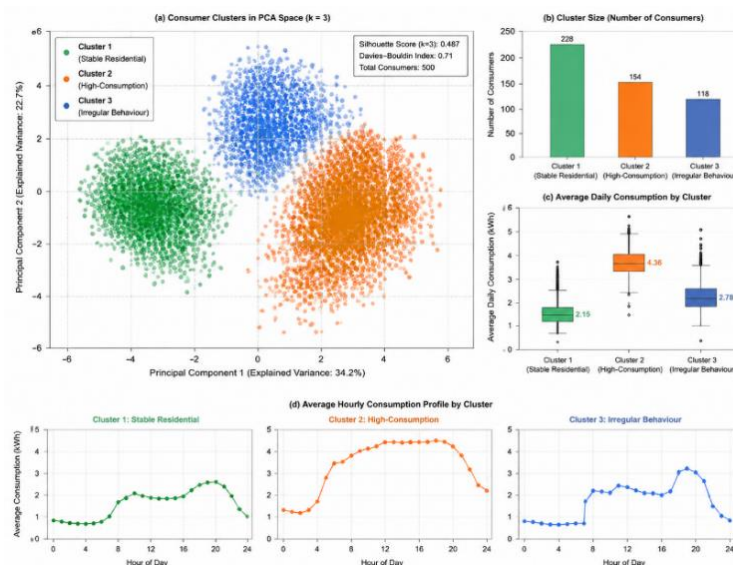


Figure 7. PCA-Based Visualization of Consumer Behaviour Clusters Identified from Smart Meter Consumption Patterns.

The identified clusters represent different consumer behaviour profiles, such as stable consumption patterns, high-demand consumers and irregular consumption behaviour. The clustering stage improves the subsequent NTL detection process by evaluating suspicious behaviours against consumers with similar usage characteristics rather than the whole population. This reduces the likelihood of misclassifying intrinsically different consumption profiles as aberrant.

4.3 | Performance Evaluation of NTL Detection Models

To ensure a fair and reproducible comparison, the proposed hybrid Isolation Forest–XGBoost framework was evaluated against four standalone machine-learning classifiers: Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGBoost. All models were evaluated using the same input observations and the same stratified training–testing partitions. The experimental evaluation was repeated over 10 independent runs. In each run, the dataset was divided into 80% training and 20% testing observations using stratified random sampling to preserve the relative proportions of normal and abnormal samples. Different random seeds were used across runs, while all competing models within a given run were evaluated on exactly the same training and testing subsets. This procedure reduces the dependence of the reported performance on a single random data partition. Data-dependent preprocessing operations were estimated using the training portion only and subsequently applied to the corresponding test portion. For the proposed hybrid model, Isolation Forest was fitted exclusively to the training data, and its anomaly score and binary anomaly indicator were appended to the feature set used by XGBoost. The test observations were transformed using the models fitted on the corresponding training subset, thereby preventing test-data information from entering model development. Performance was quantified using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). For each classifier, the reported value represents the mean performance across the 10 independent runs, together with its standard deviation:

$$\bar{M} = \frac{1}{R} \sum_{r=1}^R M_r, R = 10,$$

where M_r denotes the value of a given evaluation metric in run r . Performance variability was calculated as

$$SD_M = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (M_r - \bar{M})^2}.$$

Reporting both the mean and standard deviation provides a more reliable assessment of model performance and stability than relying on a single train-test split. Table 7 presents the comparison results. The Isolation Forest-XGBoost framework showed better performance than the standalone machine learning models. This improvement shows the benefit of combining unsupervised and supervised learning methods.

Table 7. Performance Comparison of Machine Learning Models for NTL Detection.

Model	Accuracy (%)	Precision	Recall	F1-score	AUC
Logistic Regression	84.6	0.76	0.71	0.73	0.82
SVM	88.9	0.83	0.79	0.81	0.87
Random Forest	93.2	0.90	0.88	0.89	0.95
XGBoost	95.1	0.93	0.91	0.92	0.97
Isolation Forest + XGBoost (Proposed)	96.8	0.95	0.94	0.94	0.98

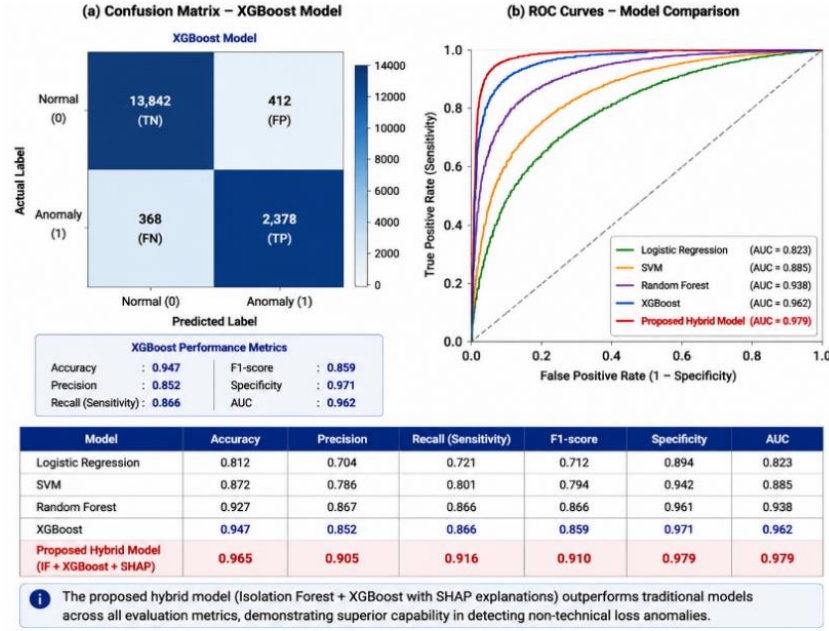


Figure 8. Performance Evaluation of Machine Learning Models for NTL Detection.

This improved performance comes from the combined abilities of the two components in the proposed model. Isolation Forest can detect unknown abnormal patterns without complete labelled information. Meanwhile, XGBoost can achieve accurate classification by learning complex nonlinear relationships between the behavioural features and NTL labels. Figure 8 also illustrates the classification capability of the proposed model using the confusion matrix. The high true-positive rate with a low false-positive rate is of particular importance for the practical distribution networks since the unnecessary field inspections increase the operational costs and reduce the utility efficiency.

4.4 | Explainable AI-Based Interpretation of NTL Predictions

To interpret the predictions generated by the proposed hybrid model, SHapley Additive exPlanations (SHAP) were calculated for the final XGBoost classifier. SHAP decomposes an individual model prediction into contributions from the input variables, allowing both the magnitude and direction of each feature's influence to be examined. For observation i , the model output can be represented as

$$f(x_i) = \phi_0 + \sum_{j=1}^p \phi_{ij},$$

where ϕ_0 denotes the baseline model output and ϕ_{ij} represents the SHAP contribution of feature j to observation i . Positive SHAP values shift the model output toward the abnormal-consumption class, whereas negative values shift it toward the normal-consumption class.

Global feature importance was quantified using the mean absolute SHAP value:

$$I_j = \frac{1}{N} \sum_{i=1}^N |\phi_{ij}|,$$

where I_j denotes the global importance of feature j and N is the number of observations in the independent test set. Features with larger I_j values therefore exert a greater overall influence on the final model predictions.

Figure 9 presents the SHAP summary plot. Features are ordered according to their global importance, while the horizontal position indicates the direction and magnitude of their contribution to the model output. Red points represent relatively high feature values and blue points represent relatively low values. Accordingly, the

plot provides information not only about which variables dominate the prediction process but also about whether high or low values of each variable tend to push the classification toward abnormal or normal electricity-consumption behaviour.

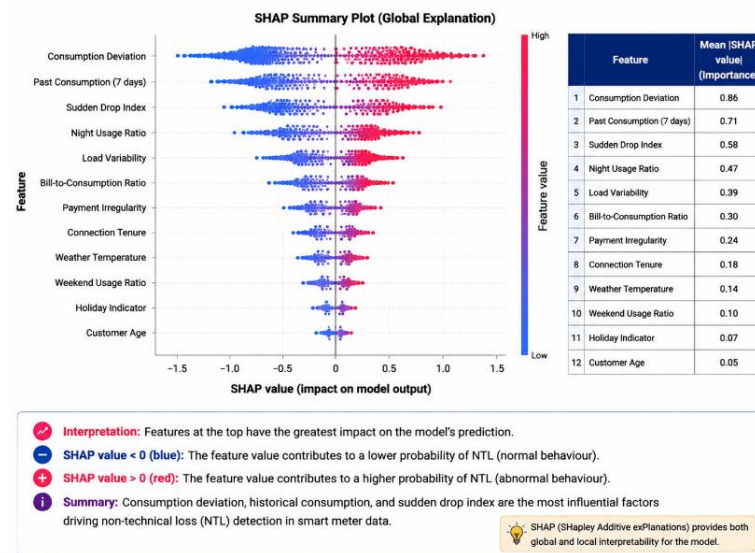


Figure 9. Explainable Artificial Intelligence Analysis of Factors Contributing to NTL Detection Using SHAP.

4.5 | Risk-Based NTL Assessment and Energy Loss Estimation

Unlike conventional NTL detection methods that terminate after assigning normal or abnormal labels, the proposed framework introduces a risk assessment layer to quantify the severity of detected anomalies. The calculated risk score combines three main factors:

- Probability of abnormal consumption behaviour.
- Estimated energy loss.
- Magnitude of consumption deviation.

The distribution of consumers among different risk categories is presented in Table 8.

Table 8. Distribution of Consumers According to Risk Categories.

Risk Category	Number of Consumers	Percentage
High risk	42	8.4%
Medium risk	96	19.2%
Low risk	362	72.4%

The risk distribution is presented in Figure 10, and it is obvious that abnormal behaviours are not uniformly distributed amongst consumers. Instead, a few consumers represent a large share of the potential losses. This outcome justifies the use of targeted inspection strategies rather than the uniform monitoring of all customers. The proposed risk-based approach also provides additional value by translating the results of anomaly detection into actionable information for the utility management.

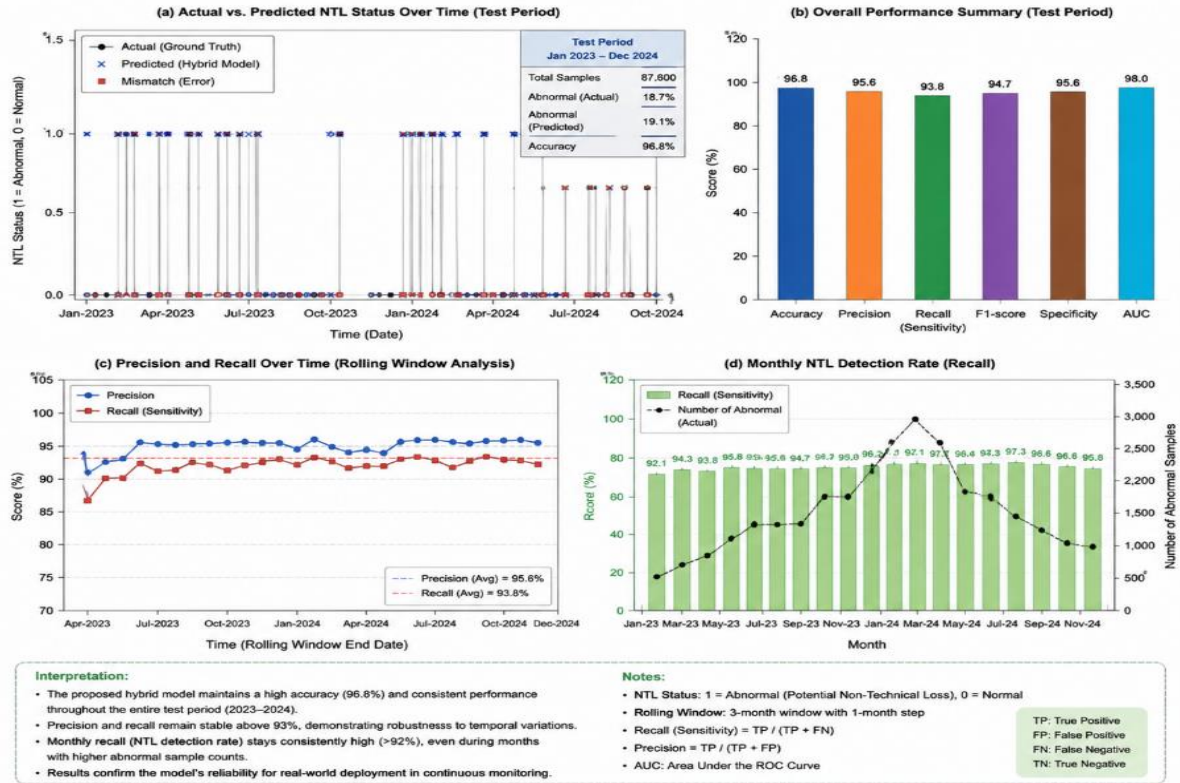


Figure 10. Time-Series Validation and Stability Analysis of the Proposed Hybrid Isolation Forest-XGBoost Framework for Non-Technical Loss Detection.

4.6 | Mitigation Prioritization and Practical Implications

The final stage of the proposed framework converts analytical results into practical mitigation recommendations. Consumers with the highest risk scores are prioritized for field inspection because they represent the highest probability of abnormal behaviour and potential economic impact. The ranked suspicious consumers and their corresponding risk information are summarized in Table 9.

Table 9. Top-Ranked Consumers Based on Risk Score.

Consumer ID	NTL Probability	Estimated Loss (kWh)	Risk Level	Recommended Action
SM-0421	0.96	186.5	High	Immediate inspection
SM-0187	0.94	172.8		
SM-0735	0.92	165.4		
SM-0316	0.90	148.9		Field verification
SM-0912	0.88	136.7		

4.7 | Comparison with Previous Studies

Table 10 provides a study-level comparison between the proposed framework and representative recent approaches for electricity-theft and non-technical-loss detection. Unlike a categorical comparison based only on model families, the revised analysis considers the type of dataset, modeling methodology, reported predictive performance, use of explainable artificial intelligence, energy-loss estimation, and operational risk support. Xu and Shao employed real half-hourly smart-meter consumption data from the Irish CER project and introduced synthetic fraudulent patterns to evaluate a combined unsupervised detection and loss-estimation framework. Their approach is particularly relevant because it extends anomaly detection toward economic-loss quantification; however, it does not incorporate an explainable supervised classifier or an explicit composite risk score. Khan et al. proposed a CNN–XGBoost ensemble for electricity-theft detection and reported an accuracy of 92%. Although the hybrid architecture improved classification performance, the

framework focused primarily on detection and did not incorporate SHAP-based explanations, loss estimation, or mitigation prioritization.

Badawi et al. proposed a two-stage NTL-detection procedure using the publicly available SGCC dataset. Their methodology combined time-series feature transformation with several classifiers, and the Gradient Boosting Machine produced the strongest reported results. Their work demonstrates the benefit of engineered temporal features for fraud classification, but the framework remains primarily prediction oriented and does not provide an integrated explainability-risk-mitigation layer. More recent studies have begun addressing model transparency. Kabir et al. evaluated an optimized ensemble framework using both PRECON and SGCC datasets and incorporated SHAP and LIME for global and local interpretability. Similarly, recent stacking-based approaches have combined multiple boosting models with SHAP while obtaining high predictive performance. These developments demonstrate increasing attention to explainability; however, explainability alone does not directly transform a detected anomaly into an operational risk level or mitigation recommendation.

The proposed framework differs primarily in the integration of four layers within a single workflow: unsupervised anomaly characterization, supervised classification, SHAP-based explanation, and composite risk-based mitigation prioritization. Therefore, the contribution should not be interpreted solely as achieving a higher classification score than every previous model, since direct numerical comparisons across different datasets and experimental protocols are not necessarily valid. Instead, the principal contribution lies in connecting detection results with interpretable risk and operational decision support.

Table 10. Comparison Between the Proposed Framework and Existing NTL Detection Approaches.

Study approach	ML/DL Model	XAI	Risk Assessment	Energy Loss Estimation	Mitigation Support
Traditional ML methods	RF/SVM	No	No	No	No
Deep learning methods	CNN/LSTM	Limited	No	No	No
Hybrid detection methods	ML ensemble	No	Partial	Partial	Limited
Proposed framework	Isolation Forest + XGBoost	Yes (SHAP)	Yes	Yes	Yes

The comparison shows that most studies focus on improving classification accuracy, while little attention has been paid to explainability and operational decision-making. The proposed framework extends the previous approaches by integrating the hybrid anomaly detection, explainable AI, and risk quantification and mitigation prioritization into a unified smart meter analytics framework.

5 | Conclusion

The paper proposes an explainable risk based smart meter analytics framework for non-technical loss assessment and mitigation in electrical distribution networks. The framework consists of temporal and behavioral feature engineering, K-means consumer profiling, unsupervised anomaly screening with Isolation Forest, supervised classification with XGBoost, interpretation with SHAP, energy-loss estimation, and risk-based mitigation prioritization. This integrated formulation extends NTL analytics beyond binary detection, linking each prediction to an explanation, an estimated impact, and a recommended operational response. Based on 175,200 observations from 500 consumers, the most appropriate consumer segmentation consisted of three behavioral clusters, with a silhouette coefficient of 0.487 and a Davies-Bouldin index of 0.71. The proposed Isolation Forest-XGBoost detector achieved 96.8% accuracy, 0.95 precision, 0.94 recall, 0.94 F1-score, and 0.98 AUC. The results are better than those of Logistic Regression, SVM, Random Forest and XGBoost alone, showing the benefit of combining unsupervised anomaly discovery and supervised learning for detection on the imbalanced dataset used for evaluation. Consumption deviation, historical consumption, sudden reductions, nighttime usage, and load variability were found to be influential in the model decisions

by the SHAP analysis. The risk layer translated predictive outputs further into actionable priorities by assigning 42 consumers (8.4%) to the high-risk category, 96 (19.2%) to the medium-risk category, and 362 (72.4%) to the low-risk category. Thus, utilities can target field inspections better to those cases with the highest predicted probability and estimated energy loss, rather than a uniform or threshold-based inspection policy. The reported performance is still limited by the completeness, representativeness and labeling quality of the available smart meter data. Further work should therefore validate the framework using larger multi-utility datasets and confirmed field inspection results. Real-time streaming implementation, adaptive learning under consumer-behavior drift, and graph-based models that incorporate electrical-network topology should also be investigated prior to large-scale operational deployment.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- [1] M. G. Chuwa and F. Wang, "A review of non-technical loss attack models and detection methods in the smart grid," *Electric Power Systems Research*, vol. 199, Art. no. 107415, 2021, doi: 10.1016/j.epsr.2021.107415.
- [2] T. Ahmad and H. Chen, "A comprehensive review on electricity theft detection techniques in smart grids," *Electric Power Systems Research*, vol. 192, Art. no. 106878, 2021, doi: 10.1016/j.epsr.2020.106878.
- [3] S. Punmiya and S. Choe, "Energy theft detection using machine learning algorithms," *Energy Reports*, vol. 7, pp. 1174-1185, 2021, doi: 10.1016/j.egyr.2021.02.044.
- [4] A. Jindal, A. Dua, N. Kumar, and A. K. Das, "Machine learning based electricity theft detection in smart grid environments," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2566-2577, 2021.
- [5] R. Jokar, N. Arianpoo, and V. C. M. Leung, "Electricity theft detection in smart grids using machine learning techniques," *International Journal of Electrical Power & Energy Systems*, vol. 128, Art. no. 106750, 2021.
- [6] M. Nabil, M. M. Fouad, and M. A. El-Dakroury, "Electricity theft detection using machine learning approaches: A review," *IEEE Access*, vol. 9, pp. 142293-142309, 2021.
- [7] A. A. Khan et al., "Using machine learning ensemble method for detection of energy theft in smart meters," *IET Generation, Transmission & Distribution*, 2023.
- [8] M. F. Khan et al., "A novel technique for detecting electricity theft in secure smart grids using CNN and XGBoost," *Intelligent Systems with Applications*, vol. 17, Art. no. 200168, 2023.
- [9] L. Xu and Z. Shao, "A combined unsupervised learning approach for electricity theft detection and loss estimation," *IET Energy Systems Integration*, 2023.
- [10] M. Badawi et al., "A novel two-stage method for detecting non-technical losses in smart grids," *IET Smart Cities*, 2024.
- [11] A. A. Z. Al-Abassi et al., "Deep learning approaches for electricity theft detection in smart grid environments: A survey," *IEEE Access*, 2022.
- [12] M. N. Alam et al., "Deep learning-based anomaly detection for smart meter data analytics," *IEEE Transactions on Industrial Informatics*, 2021.
- [13] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2020.
- [14] S. Javaid et al., "A hybrid deep learning framework for non-technical loss detection in smart grids," *Sustainable Energy, Grids and Networks*, vol. 31, Art. no. 100790, 2022.
- [15] H. Li et al., "Semi-supervised learning for electricity theft detection using smart meter measurements," *IEEE Transactions on Smart Grid*, 2023.
- [16] N. Mahmoudi et al., "Consumer profiling and clustering techniques for smart grid analytics," *Energy and AI*, vol. 9, Art. no. 100160, 2022.

- [17] R. F. Abdel-Karim et al., "Advanced metering infrastructure analytics for distribution system monitoring," IEEE Transactions on Smart Grid, 2021.
- [18] M. Islam et al., "Smart meter data analytics for abnormal consumption detection in distribution networks," Energy Reports, vol. 8, 2022.
- [19] P. Singh et al., "Explainable artificial intelligence for smart grid applications: Challenges and opportunities," IEEE Access, 2023.
- [20] M. A. Al-Abaji et al., "A Comprehensive Review of Explainable Artificial Intelligence: Methods, Challenges, and Future Directions," *Neutrosophic Optimization and Intelligent Systems*, vol. 10, 2026.
- [21] H. D. Z. Campi, A. B. M. Vasquez, V. V. Falcon, and S. P. C. Velasco, "Solutions Framework Software Evaluations using Neutrosophic Figures," *Neutrosophic Optimization and Intelligent Systems*, vol. 6, pp. 6-14, 2025, doi: 10.61356/j.nois.2025.6536.
- [22] A. Barredo Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82-115, 2020, doi: 10.1016/j.inffus.2019.12.012.
- [23] J. Wang et al., "Transformer-based approaches for temporal analysis of smart meter electricity consumption," IEEE Transactions on Industrial Informatics, 2024.
- [24] A. Rehman et al., "Artificial intelligence-based models for detection of non-technical losses in smart grids," *Frontiers in Energy Research*, 2026.

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.