Detection of Non-Technical Losses in Electrical Metering Systems in Northern Lima Using Predictive Modeling and Business Intelligence

Dayanna Perez
Faculty of Information Systems Engineering, Universidad Peruana de Ciencias Aplicadas, Lima, Peru
u202019969@upc.edu.pe

Miguel Flores
Faculty of Information Systems Engineering, Universidad Peruana de Ciencias Aplicadas, Lima, Peru
u20201b548@upc.edu.pe

Pedro Castaneda
Faculty of Systems Engineering and Electrical Mechanics, Universidad Nacional Toribio Rodriguez de Mendoza, Amazonas, Peru
pedro.castaneda@untrm.edu.pe

Jose Santisteban
Information Systems Engineering, Universidad Peruana de Ciencias Aplicadas, Lima, Peru
pcsilsan@upc.edu.pe (corresponding author)

Alejandra Onate-Andino
Escuela Superior Politecnica de Chimborazo, Riobamba, Ecuador
monate@espoch.edu.ec

Received: 19 September 2025 | Revised: 21 October 2025 and 30 October 2025 | Accepted: 1 November 2025

ABSTRACT

Non-Technical Losses (NTLs) of electric energy compromise the operational efficiency and sustainability of the electrical system, particularly in the residential sector. This study addresses this problem by developing a predictive model that can estimate energy consumption and detect anomalous patterns. For this purpose, data were collected from the Plataforma Nacional de Datos Abiertos and the Osinergmin website. The study integrates two approaches: ARIMA, which is used to represent time series with well-defined seasonal patterns, and an approach based on XGBoost to represent non-linear behavior in more heterogeneous consumption intervals. The results suggest that ARIMA demonstrated optimal performance in stable cases, with errors close to zero in several cases, where the most representative systems are SR0148 with Mean Absolute Error (MAE) = 0.000124 and Root Mean Square Error (RMSE) = 0.003549, and SE1095 with MAE = 0.000287 and RMSE = 0.004481. XGBoost, on the other hand, reached its maximum performance in the interval "From 1 to 30 kWh", with MAE = 2.81, RMSE = 5.80, and a Coefficient of Determination (R^2) of 0.13. This validates the effectiveness of the proposed approach based on the integration of more than one algorithm to identify electric consumption anomalies.

Keywords-Non-Technical Losses (NTLs); business intelligence; data analysis; XGboost; regression; ARIMA; time series; energy anomalies

I. INTRODUCTION

NTLs are electricity consumed but not billed or recorded due to non-technical factors such as energy theft, billing errors, or meter manipulation [1, 2]. Globally, electricity theft causes losses of about \$89.3 billion (USD) per year, mostly in emerging economies [3]. In Latin America and the Caribbean, NTLs represent around 15% of the total energy generated, reaching up to 30% in some countries [4]. In Metropolitan Lima, residential NTLs reach 69%, generating losses of around 490 million soles over five years [5].

To address this issue, predictive models using supervised learning have been proposed, mainly XGBoost combined with Business Intelligence (BI), to analyze smart meter data and detect anomalies in real time [6, 7]. The proposed solution, AmpTrack, integrates public data from Plataforma Nacional de Datos Abiertos [8] and Osinergmin [9] through ETL processes stored in PostgreSQL, training ARIMA and XGBoost models to forecast consumption and identify irregularities. This allows scalable and adaptive monitoring to prevent NTLs.

The model choice directly affects detection accuracy. Studies using XGBoost achieved good results [6, 10], while others utilized deep learning models such as EMB-YOLO [11] and FL-ConvGRU [12]. Statistical tests [13], IoT-based neural networks [14], explainable AI [15], synthetic data generation [16], and data protection methods [17] have been also applied to improve reliability. It has been indicated that automated metering can reduce losses by 30% [18].

Authors in [19] developed an RNN-BiLSTM-CRF model that reached 93.05% accuracy in theft detection, while authors in [20] improved detection deploying oversampling techniques such as SMOTE. Moreover, combining diverse data sources or PCA-XGBoost increased detection accuracy up to 97% [21, 22].

Real case studies provide valuable insights into effective solutions for detecting non-technical energy losses. For instance, authors in [23] developed a digital smart meter with an error below 0.5% for improved real-time monitoring, while authors in [24] applied ResNet34 and deep learning and obtained an AUC of 80.12% in anomaly detection. In [25-27], it was demonstrated that AI and adaptive algorithms can increase detection by up to 35%, reduce false positives by 25%, and recover significant annual revenue. Likewise, authors in [28] proposed a CNN-BiLSTM model for real-time load prediction, enabling precise detection of the anomalies linked to fraud or measurement errors.

II. SYSTEM DESIGN

A. Architecture

The architecture of the proposed predictive model is based on a modular structure that facilitates the integration, transformation, and analysis of data coming from open sources. Figure 1 illustrates the former's workflow that starts with the extraction stage, in which two main data sets are collected: the first one, in Excel format, contains the energy consumption history and is obtained from the Plataforma Nacional de Datos Abiertos [8]; the second one comes from an SQL backup of the Osinergmin public database [9], which provides detailed information on electricity supplies and systems.

Similarly, in the transformation stage, the data are unified and standardized. For this purpose, the Pentaho Data Integration tool is used to convert the source formats to a PostgreSQL-compatible structure. In addition, a Python script is implemented to adapt the SQL queries to the appropriate syntax for subsequent storage.

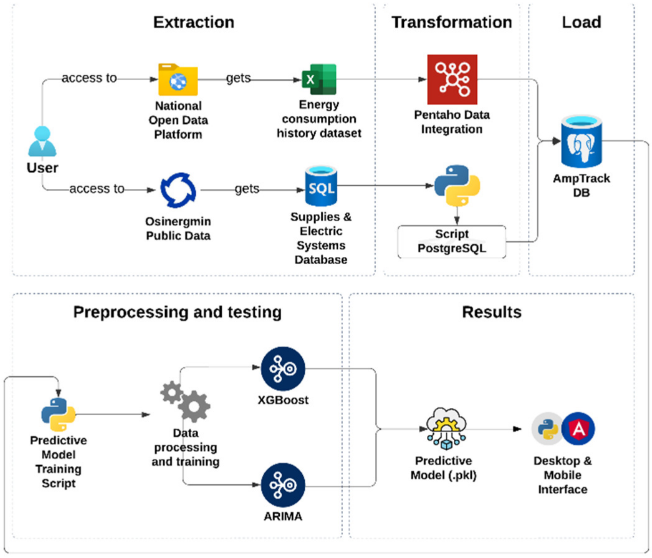

Fig. 1. Workflow of the proposed model.

Finally, in the loading phase, the processed data are integrated into a centralized database called AmpTrack DB, implemented in PostgreSQL. From this database, data are trained with ARIMA, with a focus on predicting time trends through monthly series, and XGBoost, which analyzes several variables to predict consumption in kilowatt-hours (kWh). Both predictive models are trained and evaluated periodically, and their outputs are stored in .pkl format files to be retrieved from the backend of the system. These predictions are visualized on the interface, allowing platform users to examine energy behavior and identify possible NTLs.

B. Methodology

1) Dataset

The developed predictive model is based on information obtained from two official public sources of the Peruvian government, which guarantee free access, traceability, and information quality standards. The first source is Plataforma Nacional de Datos Abiertos [8], from which historical energy consumption records were downloaded in Excel format. These files contain monthly consumption data expressed in kWh for the period between January 2022 and January 2025, as reported by electricity distribution companies nationwide.

The second source of information comes from an SQL backup of Osinergmin's public database [9], which provides technical details on electricity systems, active supplies, tariff categories, and other relevant characteristics of the distribution infrastructure. This database is crucial for understanding consumption patterns, and aligns with the structure of the national energy system.

Both datasets comply with Legislative Decree No. 1412 [29], which regulates governance, interoperability, and digital security for state entities. This regulation ensures that the data are up-to-date, consistent, reliable, and of high quality, which is vital for robust predictive analytics.

The collected data were transformed and cleansed before integration into the system. For this purpose, Pentaho Data Integration was used to convert the source files to a format compatible with the central AmpTrack DB database, implemented in PostgreSQL. Additionally, Python scripts were developed to automate the adaptation of SQL queries to a standardized format, ensuring data compatibility during loading and processing.

2) Model

The model used in the present study combines the ARIMA model and the XGBoost algorithm to calculate non-technical energy losses in electric meters. ARIMA, a time series model, helps to identify regular and seasonal patterns in energy use; this is very useful in detecting unusual changes in continuous rhythm. XGBoost, on the other hand, is a learning algorithm that uses decision trees with boosting techniques to model complex relationships between variables, making it possible to identify unusual patterns in energy consumption. Putting these two models together ensures a firm forecast, which facilitates the detection of energy theft.

Moreover, data cleaning is an important modeling process for identifying NTLs. This involves cleaning historical consumption data, which include removing repeated records, identifying and correcting outliers that would harm the analysis, and filling in missing values using techniques such as linear and mean interpolation.

The data are normalized to ensure consistency in the scales of the variables used. In the time series, smoothing techniques are applied to eliminate noise and ensure the quality of the data before being processed by the ARIMA model. No data augmentation is performed, as the data are divided equally into training, validation, and test sets, without mixing examples between them. This ensures the integrity of the analysis in the validation and testing phases.

In XGBoost, extracting features is key to improving forecasting accuracy. Important variables, such as current tariff or supply usage history, are analyzed to identify possible NTLs. The tariff type or usage quantity is encoded into numerical values so the learning algorithm can process them.

At this stage, the relevance of each characteristic is evaluated using selection techniques, making it possible to eliminate irrelevant or redundant variables that do not contribute to the predictive power. This procedure ensures that the model is trained with only the most significant data, optimizing its performance and accuracy.

3) Training

A strategy to train the predictive models is implemented using previous electricity consumption records. The ARIMA model is adjusted utilizing time series to identify seasonal behaviors in consumption, while the XGBoost algorithm is trained using explanatory variables, such as geographic location, type of supply, and electricity tariff, among other factors that affect consumption patterns.

Therefore, the predictive models employed in the development of the solution are:

- ARIMA: This model was trained by adjusting the differencing (d), autoregression (p), and moving average (q) parameters to adequately model the energy consumption time series. The optimal configuration was selected using the Akaike Information Criterion (AIC), which helps identify the model that achieves a balance between the accuracy of the fit and its complexity.
- XGBoost: For this algorithm, different configurations were trained by adjusting key hyperparameters such as maximum tree depth, learning rate, and number of iterations. Cross-validation was applied during training to reduce the risk of overfitting, and L2 regularization was incorporated as a mechanism to control complexity and improve the generalizability of the model.

To ensure efficient training and accurate evaluation of the predictive model, the dataset was divided into three main subsets. As the data are chronological, the split was performed sequentially to prevent data leakage from future periods into the training set. The oldest 80% of the data was used for training, the subsequent 10% for validation, and the most recent 10% was reserved for the final test. This segmentation enables optimizing learning, adjusting the hyperparameters, and evaluating the model's generalization capability [30]. Table I details the proportion and purpose of each subset used during the process.

TABLE I. PERCENTAGE OF DATA USED FOR MODEL TRAINING AND EVALUATION

Subset	Percentage	Description
Training	80 %	Used to adjust the model's internal parameters during the learning phase.
Validation	10 %	Used to fine-tune the model's hyperparameters, leading to improved performance without overfitting.
Test	10 %	Reserved for evaluating the final performance of the model, providing an objective measure of its generalization capability.

4) Interface

The Amptrack platform was designed to facilitate the use of the predictive model and promote its application in real operating environments. It offers a functional interface that improves efficiency in NTL detection. Figure 2 shows a summary module of Amptrack, providing a general overview with/using indicators such as the number of supervised companies, electrical systems, registered supplies, and residential tariffs. In addition, the record of the observed and reported cases is displayed, using annual filters.

Figure 3 displays the analytics module, which is accessed by entering a supply number. After that, the related data are loaded, including the prediction panels. This is one of the most important sections, since the prediction is visualized by ARIMA (shown in yellow in the graph) and also by XGBoost, which provides its prediction based on the consumption value for the date selected by the user. Also, if any anomalous consumption is observed, it allows it to be recorded by clicking on the consumption point within the graph in the consumption section (blue colored graph).

Figure 4 illustrates the observation history module, where detected anomalies are managed. It contains records that highlight consumption outside the expected pattern, which can be marked as reported. In addition, filters by company, date, and record status can also be included.

Overall, the Amptrack platform modules enable a clear, interactive, and decision-oriented visualization of non-technical energy losses. This interface not only facilitates technical analysis but also improves traceability and control of the detected cases, serving as a bridge between the predictive model and its practical application in real environments.

Fig. 2. AmpTrack Summary module.

Fig. 3. Analytics module.

Fig. 4. Observation History module.

5) *Metrics*

To evaluate the performance of the predictive models used in AmpTrack to estimate residential energy consumption, standard regression metrics were used. The latter quantify the model's accuracy and explanatory power with respect to the observed values.

Table II presents the metrics used to evaluate the predictive models' performance, specifically MAE, RMSE, and R^2. These metrics were selected for their suitability for regression tasks, where the objective is to estimate the average energy consumption based on historical values.

TABLE II. METRICS FOR PREDICTIVE MODEL TRAINING

Metric	Description	Formula
MAE	Average absolute errors. Measures how far, on average, the predictions are from the actual values. Where *n* is the total number of observations, y_i is the actual observed value, and $\hat{y}_i$ is the predicted value.	$(1/n) * \Sigma \lvert y_i - \hat{y}_i \rvert$
Precision	Root of the average of squared errors. It penalizes larger errors more and shows the dispersion of the error.	$\sqrt{((1/n) * \Sigma (y_i - \hat{y}_i)^2)}$
R^2	Measures the proportion of the total variance of the true value that is explained by the model.	$1 - (\Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2)$

Here *n* is the total number of observations, y_i is the actual observed value, and $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the average of the true values.

III. RESULTS

The current study evaluates the performance of ARIMA and XGBoost using MAE and RMSE, and, in the case of XGBoost, it also utilizes R^2. The results reflect the behavior of the models in predicting energy consumption in different tariff segments.

A. *ARIMA Results*

The ARIMA model was applied to specific electricity systems and tariff combinations that had at least 12 monthly observations. Table III presents the ten results with the lowest error values, including cases where both MAE and RMSE were practically zero, exhibiting strong modeling capability in the case of low variability in the series. In particular, the SR0148 and SE1095 electrical systems had an extremely stable consumption profile throughout the study duration, which enabled the model to make highly accurate predictions. These were tested using an independent set of measurements, deploying a cross-validation method, ensuring the robustness of the fit and avoiding excessive under-fitting corrections.

TABLE III. COMBINATIONS IN ARIMA MODEL

Electrical system	Tariff	MAE	RMSE
SR0148	From 1 to 30 kWh	0.000124	0.003549
SE1095	From 1 to 30 kWh	0.000287	0.004481
SR0128	From 1 to 30 kWh	0.007149	0.010892
SR0129	From 1 to 30 kWh	0.009702	0.011817
SE0270	From 1 to 30 kWh	0.019784	0.020273

The ARIMA model was trained with a monthly series of energy consumption between January 2022 and January 2025, segmented by the electricity system and tariff. This series enabled identifying seasonal patterns and time trends, which are the main requirements for accurate projections. In the training phase, the model performed excellently in several combinations, with error values (MAE and RMSE) approaching zero, as presented in Table III.

Within the performed analysis, the month of July was chosen randomly to highlight a specific example of prediction. Figure 5 shows that the recorded consumption was 7.5 kWh in July 2022, 7.1 kWh in July 2023, and 7.5 kWh in July 2024. The prediction generated for July 2025 was 7.2 kWh, in line with the observed historical behavior. This result reinforces the model's ability to capture the dynamics of energy consumption and project values consistent with the observed historical behavior.

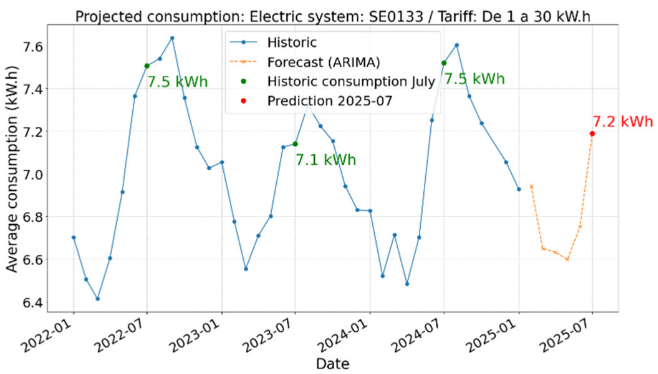

Fig. 5. Monthly consumption forecast for July 2025 using ARIMA.

B. *XGBoost Results*

Table IV illustrates the results for each tariff range, indicating that XGBoost achieved its best performance in the most representative and frequent residential consumption segments, such as 1-30 kWh, with an MAE of 2.81 and an RMSE of 5.80. As higher consumption tariffs are analyzed, a greater dispersion in error metrics is observed. This variation responds to a technical characteristic of the dataset: an asymmetric distribution in the number of records per tariff segment. The predominance of information in the lower consumption range enabled a more robust training of the model in that group, while segments with lower representativeness present a greater complexity in the estimation, due to the limited volume of available data.

TABLE IV. COMBINATIONS IN XGBOOST ALGORITHM

Tariff	MAE	RMSE	R^2
From 1 to 30 kWh	2.810554	5.802939	0.126771
From 31 to 100 kWh	7.17998	50.661134	0.213494
From 101 to 140 kWh	4.630673	112.357523	0.008694
From 151 to 300 kWh	17.121668	193.987174	0.016267
From 301 to 500 kWh	25.162317	368.270339	0.010069

July was again chosen as the reference month so that the XGBoost prediction illustrated in Figure 6 could be compared with the ARIMA model. In this context, XGBoost generated a point forecast of 7.42 kWh for July 2025, a value close to that previously obtained with ARIMA (7.2 kWh). This agreement in the estimates reinforces the reliability of both approaches and supports the usefulness of hybrid models in projecting energy consumption in the residential sector.

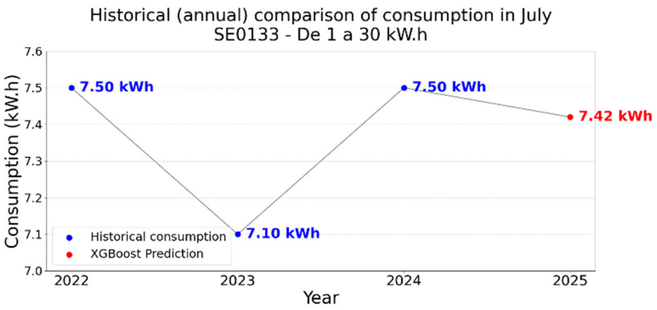

Fig. 6. Monthly consumption forecast for July 2025 using XBoost.

IV. DISCUSSION

AmpTrack provides a balanced and practical solution compared with the ones provided in existing literature. For example, while advanced models like the RNN-BiLSTM-CRF in [19] achieved a classification accuracy of 93.05% for theft identification, the regression-based approach of the present study provides a detailed consumption forecast, with the XGBoost model achieving an MAE as low as 2.81 in high-density data segments. This focus on prediction rather than binary classification offers a different utility for energy analysts. Furthermore, unlike [22, 26], where high F1-scores (97%) were obtained or significant improvements in detection (35%) were recorded by leveraging smart meter data or specific consumption profiles, the proposed model demonstrates its effectiveness using only publicly available administrative data. This makes AmpTrack immediately usable in contexts like Northern Lima, without requiring prior investment in advanced metering infrastructure.

This study highlights the potential and usefulness of ARIMA and XGBoost predictive models applied to residential energy consumption through AmpTrack. Although each model has areas where its performance varies according to the combination of electricity systems and tariffs, both provide significant and complementary utility in prediction.

The ARIMA model showed outstanding performance in time series analysis with stable patterns and sufficient historical data, achieving near-zero errors in several cases. This confirms that classical time series methods are still highly effective for certain energy consumption applications, particularly when demand varies seasonality or in well-defined trends. On the other hand, the XGBoost algorithm, based on automatic learning, demonstrated greater flexibility to handle heterogeneous tariff ranges and more variable consumption patterns, although with lower performance in ranges with limited data.

Compared to other solutions, the proposed AmpTrack system offers a balanced combination of accuracy and regulatory applicability in real residential contexts. The results in [6, 10] show the high accuracy of XGBoost in heterogeneous environments, similar to the results obtained in the present study. However, methods such as EMB-YOLO [11] or FL-ConvGRU [12] are more oriented to image recognition or intelligent environments with specialized infrastructure, but not always suitable for NTL detection. Other works, such as [13, 14], incorporate deep neural networks and IoT, with promising results but dependent on advanced measurement devices. In contrast, AmpTrack relies on open and public data, accessible to institutions such as Osinergmin, which facilitates its replicability.

Authors in [18, 19, 22] confirm that the use of enriched data and over-sampling techniques can improve metrics, which is in line with the segment by tariffs and electricity systems adopted by the present study. Moreover, practical solutions such as those in [23, 26], focusing on IoT meters or consumption profiles, achieve significant economic impacts, but require infrastructure investment and are beyond the scope of the present study.

Overall, the reviewed studies confirm that using hybrid models with structured data, such as AmpTrack, represents an effective, scalable, and consistent alternative, suitable for developing countries, where predictive monitoring can be a key tool against NTLs.

V. CONCLUSIONS

This study developed and evaluated a hybrid predictive model, integrating ARIMA and XGBoost for the detection of Non-Technical Losses (NTLs) in residential energy consumption, through the analysis of official data provided by the Plataforma Nacional de Datos Abiertos and Osinergmin. The proposed model achieved a reliable performance with a Mean Absolute Error (MAE) of 4.57, a Root Mean Square Error (RMSE) of 5.83, and a Coefficient of Determination (R^2) of 0.82, validated on a set of structured data extracted from official sources such as Osinergmin and the Plataforma Nacional de Datos Abiertos. These results validate the effectiveness of the hybrid approach to capture both historical patterns and non-linear anomalies of energy behavior.

Segmentation by electricity system, tariff, and supply number proved to be significant in improving prediction accuracy, enabling identifying areas with a higher NTL incidence. The traceability functionality of the observed and reported cases was incorporated, strengthening institutional control and facilitating supervision by Osinergmin's analysts.

One of the main advantages of the AmpTrack lies in its approach to Business Intelligence (BI), which provides visualizing data in interactive dashboards, generating exportable reports (PDF, spreadsheet, CSV), and filters to facilitate analysis. This integration between data science and BI tools provides a user-friendly environment for decision-making, without requiring advanced programming knowledge from end users. Compared to previous studies, AmpTrack emerges as an effective tool, especially useful for developing countries where advanced infrastructure, such as IoT or smart sensors, is not available. Unlike models based on deep neural networks or computer vision approaches that require specialized hardware, the AmpTrack is implementable on existing data, reducing costs and adoption time.

Finally, the use of hybrid models, enriched with intelligent segmentation and dynamic visualization, represents an effective and affordable strategy to address NTLs. The developed model not only facilitates the proactive detection of anomalies but also contributes to institutional strengthening, regulatory efficiency, and sustainability of the electricity system in vulnerable urban sectors.

ACKNOWLEDGMENT

The authors express their gratitude to the Dirección de Investigación of the Universidad Peruana de Ciencias Aplicadas for supporting this research through the UPC-EXPOST-2025-2 incentive.

REFERENCES

[1] S. K. Gunturi and D. Sarkar, "Ensemble Machine Learning Models for the Detection of Energy Theft," *Electric Power Systems Research*, vol. 192, Mar. 2021, Art. no. 106904, https://doi.org/10.1016/j.epsr.2020.106904.

[2] I. M. R. Baldawi and T. İnan, "A Novel and Effective Method Based Deep Learning Model for Detecting Non-Technical Electricity Losses," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 14, no. 4, Dec. 2023, Art. no. 2464, https://doi.org/10.11591/ijpeds.v14.i4.pp2464-2473.

[3] R. K. Ahir and B. Chakraborty, "Pattern-Based and Context-Aware Electricity Theft Detection in Smart Grid," *Sustainable Energy, Grids and Networks*, vol. 32, Dec. 2022, Art. no. 100833, https://doi.org/10.1016/j.segan.2022.100833.

[4] M. Toledo-Orozco, C. Arias-Marin, C. Álvarez-Bel, D. Morales-Jadan, J. Rodríguez-García, and E. Bravo-Padilla, "Innovative Methodology to Identify Errors in Electric Energy Measurement Systems in Power Utilities," *Energies*, vol. 14, no. 4, Feb. 2021, Art. no.958, https://doi.org/10.3390/en14040958.

[5] "Energy Theft." *Pluz Energía Perú*, Oct. 2024. https://www.pluz.pe/es/ayuda/hurto-de-energia.html.

[6] A. Nawaz, T. Ali, G. Mustafa, S. U. Rehman, and M. R. Rashid, "A Novel Technique for Detecting Electricity Theft in Secure Smart Grids using CNN and XG-Boost," *Intelligent Systems with Applications*, vol. 17, Feb. 2023, Art. no. 200168, https://doi.org/10.1016/j.iswa.2022.200168.

[7] M. J. Abdulaal *et al.*, "Real-Time Detection of False Readings in Smart Grid AMI Using Deep and Ensemble Learning," *IEEE Access*, vol. 10, pp. 47541–47556, 2022, https://doi.org/10.1109/ACCESS.2022.3171262.

[8] Osinergmin, "Electrical energy consumption report (kWh-month) (Supervisory Body for Investment in Energy and Mining)." Datosabiertos, Nov. 2024, [Online]. Available: https://www.datosabiertos.gob.pe/dataset/reporte-del-consumo-de-energ%C3%ADa-el%C3%A9ctrica-kwh-mes-organismo-supervisor-de-la-inversi%C3%B3n-en-3.

[9] Osinergmin, "Aprobación de las Altas y Bajas 2023." Datosabiertos, 2023, [Online]. Available: https://www.osinergmin.gob.pe/seccion/institucional/regulacion-tarifaria/procesos-regulatorios/actividades-relacionadas/aprobacion-altas-bajas-2023.

[10] Pamir, N. Javaid, M. Akbar, A. Aldegheishem, N. Alrajeh, and E. A. Mohammed, "Employing a Machine Learning Boosting Classifiers Based Stacking Ensemble Model for Detecting Non-Technical Losses in Smart Grids," *IEEE Access*, vol. 10, pp. 121886–121899, 2022, https://doi.org/10.1109/ACCESS.2022.3222883.

[11] Z. Liu, Y. Li, F. Shuang, Z. Huang, and R. Wang, "EMB-YOLO: Dataset, Method and Benchmark for Electric Meter Box Defect Detection," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 2, Feb. 2024, Art. no. 101936, https://doi.org/10.1016/j.jksuci.2024.101936.

[12] M. H. Zafar *et al.*, "Step Towards Secure and Reliable Smart Grids in Industry 5.0: a Federated Learning Assisted Hybrid Deep Learning Model for Electricity Theft Detection using Smart Meters," *Energy Reports*, vol. 10, pp. 3001–3019, Nov. 2023, https://doi.org/10.1016/j.egyr.2023.09.100.

[13] B. Coma-Puig, A. Calvo, J. Carmona, and R. Gavaldà, "A Case Study of Improving a Non-technical Losses Detection System Through Explainability," *Data Mining and Knowledge Discovery*, vol. 38, no. 5, pp. 2704–2732, Sept. 2024, https://doi.org/10.1007/s10618-023-00927-7.

[14] P. K. Sebastian, K. Deepa, N. Neelima, R. Paul, and T. Özer, "A Comparative Analysis of Deep Neural Network Models in IoT-based Smart Systems for Energy Prediction and Theft Detection," *IET Renewable Power Generation*, vol. 18, no. 3, pp. 398–411, Feb. 2024, https://doi.org/10.1049/rpg2.12824.

[15] E. Correia, R. Calili, J. F. Pessanha, and M. F. Almeida, "Definition of Regulatory Targets for Electricity Non-Technical Losses: Proposition of an Automatic Model-Selection Technique for Panel Data Regressions," *Energies*, vol. 16, no. 6, Mar. 2023, Art. no. 2519, https://doi.org/10.3390/en16062519.

[16] A. Narayanan and T. Hardy, "Synthetic Data Generation for Machine Learning Model Training for Energy Theft Scenarios using Cosimulation," *IET Generation, Transmission & Distribution*, vol. 17, no. 5, pp. 1035–1046, Mar. 2023, https://doi.org/10.1049/gtd2.12619.

[17] S. Lavrijssen, B. Espinosa Apráez, and T. Ten Caten, "The Legal Complexities of Processing and Protecting Personal Data in the Electricity Sector," *Energies*, vol. 15, no. 3, p. 1088, Feb. 2022, https://doi.org/10.3390/en15031088.

[18] F. B. Abdullah, M. Arsalan Aqeeq, R. Iqbal, M. Abdullah, and F. S. Memon, "Enhancing Electricity Distribution Efficiency in Pakistan: a framework for Progress and Action," *Utilities Policy*, vol. 88, June 2024, Art. no. 101746, https://doi.org/10.1016/j.jup.2024.101746.

[19] A. Khalid, G. Mustafa, M. R. R. Rana, S. M. Alshahrani, and M. Alymani, "RNN-BiLSTM-CRF based Amalgamated Deep Learning Model for Electricity Theft Detection to Secure Smart Grids," *PeerJ Computer Science*, vol. 10, Feb. 2024, Art. no. e1872, https://doi.org/10.7717/peerj-cs.1872.

[20] Y. Maraden *et al.*, "Enhancing Electricity Theft Detection through K-Nearest Neighbors and Logistic Regression Algorithms with Synthetic Minority Oversampling Technique: A Case Study on State Electricity Company (PLN) Customer Data," *Energies*, vol. 16, no. 14, July 2023, Art. no. 5405, https://doi.org/10.3390/en16145405.

[21] M. B. Capeletti *et al.*, "Identification of Nontechnical Losses in Distribution Systems Adding Exogenous Data and Artificial Intelligence," *Energies*, vol. 15, no. 23, Nov. 2022, Art. no. 8794, https://doi.org/10.3390/en15238794.

[22] A. I. Kawoosa, D. Prashar, M. Faheem, N. Jha, and A. A. Khan, "Using Machine Learning Ensemble Method for Detection of Energy Theft in Smart Meters," *IET Generation, Transmission & Distribution*, vol. 17, no. 21, pp. 4794–4809, Nov. 2023, https://doi.org/10.1049/gtd2.12997.

[23] O. Munoz *et al.*, "Development of an IoT Smart Energy Meter with Power Quality Features for a Smart Grid Architecture," *Sustainable Computing: Informatics and Systems*, vol. 43, Sept. 2024, Art. no. 100990, https://doi.org/10.1016/j.suscom.2024.100990.

[24] S. Mitra, B. Chakraborty, and P. Mitra, "Smart Meter Data Analytics Applications for Secure, Reliable and Robust Grid System: Survey and Future Directions," *Energy*, vol. 289, Feb. 2024, Art. no. 129920, https://doi.org/10.1016/j.energy.2023.129920.

[25] M. A. Souza *et al.*, "Detection of Non-Technical Losses on a Smart Distribution Grid Based on Artificial Intelligence Models," *Energies*, vol. 17, no. 7, Apr. 2024, Art. no. 1729, https://doi.org/10.3390/en17071729.

[26] M. Žarković and G. Dobrić, "Artificial Intelligence for Energy Theft Detection in Distribution Networks," *Energies*, vol. 17, no. 7, Mar. 2024, Art. no. 1580, https://doi.org/10.3390/en17071580.

[27] A. B. Pengwah, R. Razzaghi, and L. L. H. Andrew, "Model-Less Non-Technical Loss Detection using Smart Meter Data," *IEEE Transactions on Power Delivery*, vol. 38, no. 5, pp. 3469–3479, Oct. 2023, https://doi.org/10.1109/TPWRD.2023.3280551.

[28] M. Bartouli, A. Msolli, A. Helali, and H. Fredj, "A Real-Time Charge Predictive Model for Intelligent Networks," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17091–17098, Oct. 2024, https://doi.org/10.48084/etasr.7845.

[29] Gobierno del Perú, *Decreto Legislativo No. 1412 – Ley de Gobierno Digital*, Lima, Perú, Sep. 2018. [Online]. Available: https://www.gob.pe/institucion/pcm/normas-legales/289706-1412.

[30] D. Pérez, M. Flores, and P. Castañeda, "Non-Technical Loss Detection in Electric Meter Systems: A Predictive Approach with Business Intelligence in North Lima," *Nanotechnology Perceptions*, vol. 20, pp. 3579–3590, 2024.