Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Non-Technical Loss Detection in Electric Meter Systems: A Predictive Approach with Business Intelligence in North Lima

Dayanna Pérez, Miguel Flores, Pedro Castañeda

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

Non-technical losses (NTL) in power distribution systems, including electricity theft, device failures, and maintenance issues, represent a significant challenge for electric utility companies, especially in emerging economies. This study presents a predictive model that uses advanced business intelligence (BI) and machine learning techniques, specifically ARIMA and XGBoost models, to detect non-technical losses in electric metering systems in northern Lima. The methodology employed includes data extraction, transformation, and loading (ETL) from various sources, such as the National Open Data Platform and electric utility registries. After a series of preprocessing steps involving anomaly detection, feature engineering, and cross-validation, the model optimizes its accuracy in predicting irregular consumption patterns, energy theft indicators, and other NTL. The results obtained demonstrate that the combination of ARIMA and XGBoost is effective in identifying atypical consumption patterns, contributing to improving both the reliability of the electric system and its economic efficiency. Furthermore, the model respects data governance policies under Legislative Decree 1412, ensuring quality and security of information. The solution is scalable and can be adapted to other similar contexts, offering a useful tool for energy distribution companies in the early detection of consumption irregularities. Thus, it presents an effective alternative to reduce non-technical losses through advanced analysis, improving energy management in areas with distribution challenges.

1. Introduction

Energy theft refers to the deliberate or illegal use of electrical energy by various means, the main cause is the theft of electricity that represents approximately 80% of the total loss. In addition, non-technical (NTL) losses include electricity theft, device installation errors, lack of maintenance, and counting errors. In addition, they can be caused by a wide variety of causes such as altered meter readings through unauthorized access to the database, incorrect calculations of non-technical losses, container fraud, faulty metering, theft of electricity through distribution lines, non-payment by customers, billing errors, and so on [1, 2].

It is estimated that the world loses about $89.3 billion to electricity theft annually, with emerging nations accounting for $58.7 billion." [3]. For example, in Latin America and the Caribbean, NTLs accounted for approximately 15% of the total energy generated in 2018. However, this percentage varied between 5% and 30% depending on the country, due to the strong correlation with social, economic, political and technical variables [4]. In the particular case of Metropolitan Lima, statistics reveal an alarming incidence of NTLs, especially in the residential sector, where they reach a worrying 69%. This data not only reflects a vulnerability in the electricity distribution system, but also has a significant economic impact. According to data from Enel Peru, it is estimated that these losses have represented a loss of close to 490 million soles in the last five years due to energy theft [5].

In this context, an approach based on the use of data from smart meters and auxiliary databases as raw material for a supervised machine (XGBoost) is the best learning algorithm to perform the predictive model [6]. Likewise, [7] indicates that the prediction of electrical energy NTL using Business Intelligence (BI) is useful when proposing a detector based on deep learning and together to detect false readings in real time in advanced metering smart grids. Through the analysis of energy consumption data in real time and the use of deep learning models, it seeks to improve the accuracy in the detection of false readings, which contributes to predicting and preventing electrical energy NTLs more effectively through technologies such as BI.

However, the arguments mentioned in the articles on how to predict with a higher rate of effectiveness depend on how the solution is proposed, the aspects mentioned give a vision of what tools can be used in the future and how the implementation of Business Intelligence helps it to be successful.

In view of this, the present work suggests proposing a predictive model, starting with a data processing phase that includes ETL (Extract, Transform and Load) using Python to extract data from multiple sources, followed by an exploratory data analysis (EDA). This EDA, partly automated by AWS Glue, takes care of data preparation and cleansing, removing outliers and normalizing data if necessary. In addition, the pre-processed data is stored in Amazon S3 and used to train machine learning algorithms, optimized using hyperparameterization. The results generated can be presented in binary format or with detailed information, depending on the richness of the available data, and are continuously evaluated and adjusted to improve the accuracy of the model.

Below is the breakdown of the rest of the text in the article. Section one will cover the introduction to this project. The contributions that helped the research work will be presented in section two. In section three, the design of the architecture of the model will be evidenced. Likewise, in section four the experiments and results of this work will be reflected, section five presents the discussion. The conclusions are evidenced in section six and will end with acknowledgements.

2. RELATED WORKS

An easy Studies on learning models for NTL detection in electric meters reveal a common focus on optimizing deep learning techniques and advanced classification models. The studies reviewed agree on using deep architectures and optimization algorithms to improve both accuracy and computational efficiency in anomaly detection. Among them, the article by [8] that develops the EMB-YOLO model stands out, which focuses on lightness and speed when reducing parameters, while [9] uses XGBoost together with anomaly detectors in smart homes, achieving an accuracy of 98.91% in the detection of electricity theft. Both articles highlight the relevance of adjusting models to specific consumption contexts and data environments, focusing on computational efficiency and accuracy in different scenarios.

Likewise, the articles highlight the importance of integrating preprocessing methods and feature selection techniques to improve the performance of learning models in detecting non-technical losses. [10] incorporates Principal Component Analysis (PCA) to power the XGBoost model, which strengthens the accuracy and efficiency of the detection system. Similarly, [11] they employ pattern-matching techniques in conjunction with near-neighbor methods, which allow anomaly scores to be obtained based on specific consumption behaviors. These approaches highlight the relevance of advanced preprocessing methods to manage large volumes of data and improve the ability of models to identify irregular consumption patterns.

That is why the use of advanced classification assemblies and models stands out as a trend to address the detection of losses in electrical networks. [12] employ an ADASYN and assembly-based reinforcement classification model, achieving outstanding ROC-AUC and PR-AUC values compared to other models. This approach is particularly useful in scenarios where there is a significant imbalance of data, a common challenge in detecting non-technical losses. This group of studies underscores the effectiveness of assembly models and advanced classification techniques in overcoming accuracy and data balancing challenges, providing a comprehensive perspective on the most appropriate approaches to loss detection in electrical metering systems.

On the other hand, other studies show a common focus on optimizing and using energy data to improve the effectiveness of these models. The articles reviewed agree on the importance of analyzing consumption patterns and using advanced techniques such as reservoir computing, neural networks and regression models. For example, [13] it features the SpeCluRC-NTL model, which uses energy data from smart meters and time series to detect fraudulent behavior in smart grids, enabling simplification and efficiency in classification. Similarly, [14] it leverages convolutional neural networks and LSTM to process large volumes of consumer data and achieve high accuracy in fraud detection, confirming the effectiveness of deep learning in this type of analysis.

Another common aspect is the use of explanatory techniques and feature selection methods, which improve the interpretability and accuracy of predictive models. [15] implements methods such as SHAP and LIME, which provide detailed explanations of the variables that affect non-technical loss detection, making the models more transparent and easier to interpret. [16], on the other hand, uses a model selection technique to establish regulatory goals in distribution companies in Brazil, identifying external variables such as population density and poverty, which reinforces the focus on the adaptability of the models to contexts. Specific energy data. These studies agree on the usefulness of explainability and statistical analysis methods to strengthen models and improve their applicability in the energy sector.

Finally, the articles coincide in highlighting the need to process energy data in time and frequency domains to capture specific consumption patterns and optimize predictive models.

[17] It employs features in both domains along with feature selection and hyperparameter optimization techniques, resulting in high accuracy in detecting electricity theft. The inclusion of consumption variables in time series is evidenced as an effective practice to capture relevant variations in consumer behavior, improving the robustness of the models in the identification of fraudulent patterns and non-technical losses. Together, these studies underscore the relevance of leveraging energy data holistically to improve the detection of irregularities in consumption and optimize predictive models in smart grid environments.

3. SYSTEM DESING

A. Architecture

The proposed predictive model begins with comprehensive data processing, including Extract, Transform, and Load (ETL) for the generation of predictions. The data is extracted from various sources such as the Open Data Portal and electricity company registries, using Python scripts. After extraction, an exploratory data analysis (EDA) is performed to understand its structure and detect inconsistencies.

The EDA process is divided into two stages: one automated by AWS Glue, which prepares and integrates data, and another manual, where information is cleaned by removing outliers and normalizing data. The hyperparameters of the machine learning algorithms are then adjusted to optimize their performance.

The preprocessed data is stored in Amazon S3 and is used to train predictive models that can perform regressions, classifications, or clusters. The results are presented in binary form or with details on non-technical energy losses (NTLs), depending on the information available.

The model's iterative approach, which includes cross-validation and continuous parameter adjustment, ensures accurate and effective results for the detection of non-technical losses in the affected area's power grid.

FIG. 1 PROPOSED SOLUTION

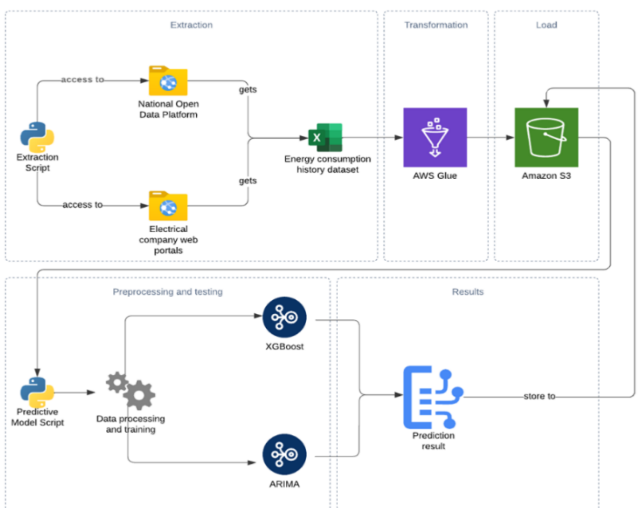

Fig. 1 shows the processing of the proposed solution

B. Methodology

1) Dataset: The data that will be used in the predictive model comes from two main sources: the Open Data Portal and the websites of electricity distribution companies [5]. The Open Data Portal, which is a state-run website, publishes historical consumption data for all electricity companies in Peru [18]. Since state entities are governed by Legislative Decree 1412 [19], which promotes data governance and digital security, it is ensured that the data used is up-to-date, reliable, and with an adequate level of governance, which strengthens the quality of predictive analytics.

The dataset is divided into test, training, and validation sets. For validation, production data provided by OSINERGMIN and/or electricity distribution companies is used. This validation process ensures that the model predicts accurately in real-world environments. As for the training and testing set, historical data from the Open Data Portal and records of electricity distribution companies are used. This process includes the extraction, transformation and loading (ETL) of the data, followed by the preparation and cleaning of the data, ensuring the integrity and quality of the information. As for the validation set, once the model has been trained and tested with the historical data, the validation is performed in a real production environment. The validation data comes from the stakeholders (OSINERGMIN or the electricity distribution companies), where the predictions generated by the model are applied to real situations of electricity consumption. This approach ensures that the model is robust and that predictions accurately reflect anomalies or potential energy theft.

The fact of using data regulated under Legislative Decree 1412 [19], guarantees that the project handles data with a high standard of security and quality. This regulatory framework promotes interoperability, data governance, and digital security, ensuring that the project aligns with data governance principles and that the data used is reliable and valid for predictive analytics.

In this way, the data provided by OSINERGMIN are strategically divided between the training and test sets, while the validation is carried out in real environments, maximizing the accuracy and reliability of the predictions of the non-technical energy loss model.

2) Model: The predictive model developed in this project uses a combination of ARIMA and XGBoost to detect non-technical energy losses in electric meters. ARIMA is a time series model that allows the identification of seasonal patterns and trends in energy consumption, being ideal for detecting anomalies related to regular changes over time. On the other hand, XGBoost is a boosting-based model that handles large volumes of data and complex relationships between variables, allowing it to detect atypical patterns in energy consumption, related to possible fraud. The combination of these two models guarantees a robust and accurate prediction, which facilitates the identification of energy thefts.

Likewise, the preprocessing of the data is a key step in the predictive model for the detection of non-technical energy losses. A thorough cleaning of historical consumption data is performed. This cleanup includes the removal of duplicate values, the detection and correction of outliers that could distort the analysis, and the imputation of missing values using techniques such as mean or linear interpolation.

In addition, the data are normalized to ensure consistency in the scales of the variables used.

For time series, smoothing techniques are used to remove noise and ensure data quality before being processed by the ARIMA model. No data augmentation is performed, as the data is divided into training, validation, and testing sets uniformly, without mixing examples between these sets. This ensures the integrity of the analysis in the validation and testing stages.

On the other hand, in the XGBoost model, feature extraction is a crucial step in improving the accuracy of predictions. The most relevant characteristics are selected for the detection of possible energy theft, including variables such as the tariff applied and past consumption of the supply. Coding techniques are applied to transform categorical characteristics, such as tariff type or energy use, into numerical values, which can be interpreted by the machine learning model.

During this phase, the importance of each feature is evaluated using feature selection techniques, which allows irrelevant or redundant variables that do not provide predictive value to be eliminated. This process ensures that the model is trained on only the most relevant data, optimizing its performance and accuracy.

3) Training: To train the predictive models, an approach based on the use of historical energy consumption data is employed. The ARIMA model [10] is trained on time series data to capture seasonal patterns in energy consumption, while XGBoost is trained using a set of characteristics including geographic location, supply type, tariff, and other relevant factors that influence consumption.

That is why the algorithms to be presented are defined:

- ARIMA: It was trained by configuring the differentiation (d), autoregression (p) and moving average (q) parameters that best fit the time series of energy consumption. The selection of these parameters was made using the AIC (Akaike Information Criterion) methodology to select the best model.

- XGBoost: The model was trained by adjusting hyperparameters such as tree depth, learning rate, and number of iterations (n_estimators). The XGBoost model was trained by cross-validation to avoid overfitting, and L2 regularization was used to control the complexity of the model.

The data were also divided into three sets:

- Training (80%): Used to adjust model parameters.
- Validation (10%): Used to tune hyperparameters.
- Test (10%): Used to evaluate the final performance of the model.

4) Evaluation and Statistical Analysis: To evaluate the performance of the models, the following key metrics were used in the Energy Theft Detection Classification task.

TABLE I METRICS

Metric	Description	Formula
Accuracy	The proportion of correct predictions over the total predictions made. Evaluate the overall performance of the model.	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	The ratio of correct positive predictions (true positives) among all the positive predictions made. It measures	

	how many of the thefts detected are actually thefts.	$\frac{TP}{TP+FP}$
Recall	The proportion of true positives over the total number of actual positive cases. Evaluate the model's ability to detect theft.	$\frac{TP}{TP+FN}$
F1-Score	It is the harmonic mean between accuracy and sensitivity, useful when there is an imbalance in the data, as is the case with energy theft.	$2 \times \frac{Precision \times Recall}{Precision + Recall}$
AUC-ROC	It measures the performance of the model at different decision thresholds, assessing the ability to distinguish between theft and non-theft based on its rate of true positives and false positives.	$\mathrm{AUC} = \int 01\mathrm{TPR}(\mathrm{FPR})\mathrm{d}(\mathrm{FPR})$

Table I shows the metrics to evaluate the performance of the model

This table specifies the formulas and acronyms TP (True Positives), TN (True Negatives), FP (False Positives), FN (False Negatives), TPR (True Positive Rate) and FPR (False Positive Rate). When analyzing the different characteristics of the ARIMA model, we find functions that complement XGboost, which allows us to have a greater overview of the predictions to be generated. The parameters of the ARIMA model are also shown:

TABLE II ARIMA METRICS

Metric	Description
Student's t test	It was used when the prediction results followed a normal distribution, to compare the models in terms of accuracy.
Wilcoxon Test	It was used to assess the difference in performance between models when data were not normally distributed, considering paired outcomes in predictions.
DeLong Test	It was used to compare the ROC curves of the models and determine which one had a better performance in the classification of energy thefts.
P-valor (p-value)	In all cases, the results showed that XGBoost had a better overall performance in detecting power theft ($p < 0.05$), proving that it is a superior model in terms of accuracy and recall.

Table II shows the metrics of the ARIMA model, which are useful for forecasting future values in a time series based on past values.

4. RESULTS

FIG. 2 LINE CHART OF TRANSFORMED DATA

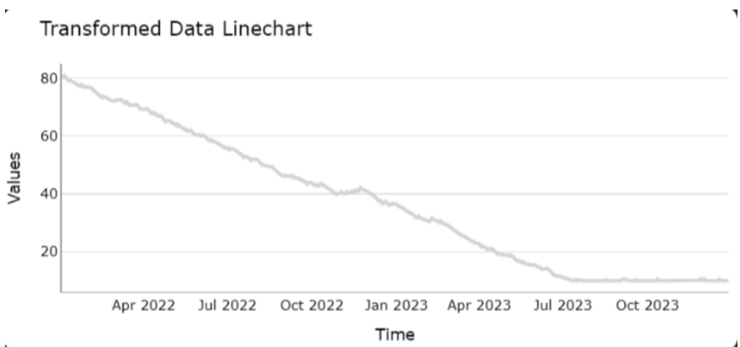

Fig. 2 Time series showing the downward trend in energy consumption after applying transformations, indicating possible non-technical losses

Figure 2 shows the time series of energy consumption after the transformations necessary to stabilize the mean and variance, which is essential for the applicability of the ARIMA model.

A clear downward trend is observed from values close to 80 to approximately 10 in the period analyzed. This marked decrease is atypical in a series of energy consumption, where a constant average would be expected under normal conditions. This downward trend suggests the possibility of non-technical losses (NTLs), such as energy theft, and reinforces the hypothesis that there are external factors that affect consumption behavior.

FIG. 3 LINE CHART OF TRANSFORMED DATA

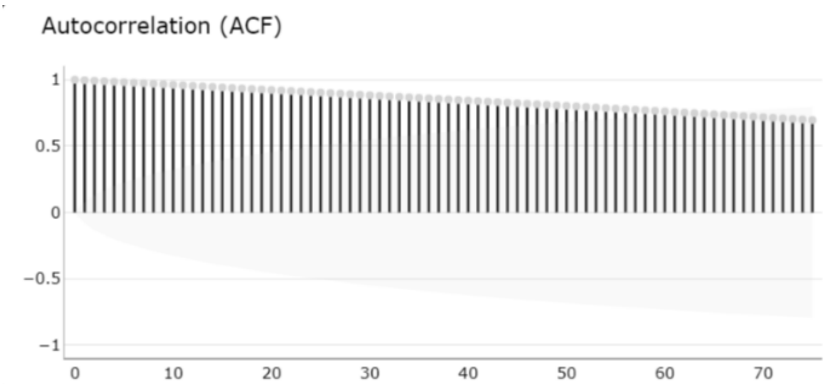

Fig. 3 The correlation gradually decreases with lags, indicating short-term dependence in energy consumption data

This graph shows that the autocorrelation function (ACF) shows a gradual decrease in correlation as lags increase, which is characteristic of a time series with short-term dependence.

This behavior indicates that consumption values in recent periods influence subsequent values, which validates the use of time series models such as ARIMA. In the context of non-technical losses, this dependence suggests that anomalies in a specific period can have persistent effects over time, allowing the model to capture patterns and deviations from expected consumption.

FIG. 4 BOX-COX PLOT

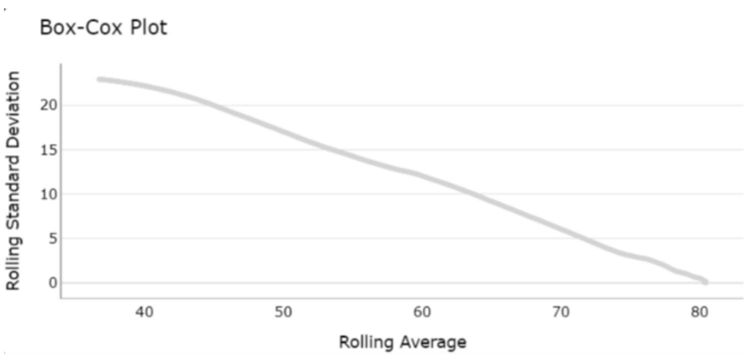

Fig. 4 Box-Cox transformation, stabilizing the variance of the energy consumption series by reducing heteroscedasticity, optimizing the detection of anomalous patterns

Figure 4 shows the Box-Cox transformation, which was applied to reduce heteroskedasticity in the consumption series, as can be seen in the graph.

The standard deviation decreases as the moving average increases, indicating that the transformation managed to stabilize the variance in the series. Variance stabilization is critical for linear time-series models, as it prevents extreme values or variations in magnitude from affecting model predictions. This optimizes the accuracy in the detection of anomalous patterns, making it easier to identify possible energy thefts by comparing atypical consumption patterns.

FIG. 5 LINE CHART OF TRANSFORMED DATA

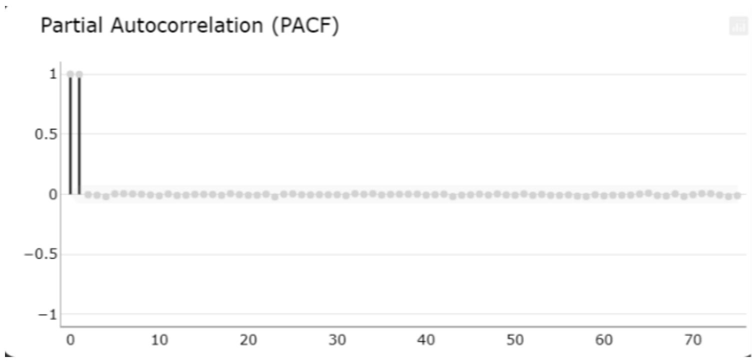

Fig. 5 Partial autocorrelation function, with a strong correlation at early lags and a sharp decline towards zero, indicating short-term dependence patterns and seasonal effects on energy consumption

This figure shows the PACF, which shows a strong correlation in the first lags, with an abrupt drop towards zero from the third lag onwards.

This behavior suggests that energy consumption in recent periods has a direct and significant influence on current consumption, but this influence decreases rapidly. This pattern is characteristic of time series where seasonal or recurrent short-term effects are observed. The presence of significant correlation in the first lags reinforces the selection of an autoregressive (AR) model within the ARIMA process to effectively capture these dependencies and detect possible consumption irregularities based on recent patterns.

FIG. 6 POST-PARAMETERIZATION RESULTS

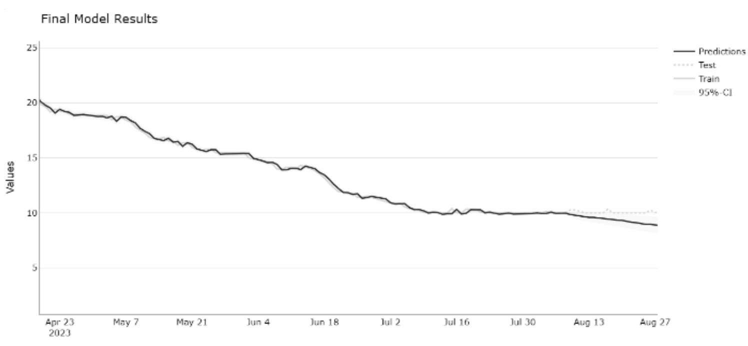

Fig. 6 Comparison of training and testing data and ARIMA-XGBoost model predictions, showing accurate fit and a decreasing trend in energy consumption, along with a 95% confidence interval.

In this figure, the final results graph shows the comparison between the training data, the test data, and the predictions generated by the ARIMA-XGBoost model, along with the 95% confidence interval.

The solid line represents the model's predictions and closely follows the trend observed in the actual test data. The model predicts a sustained downward trend in consumption, which is consistent with the previously transformed series. The inclusion of the confidence interval suggests that the model maintains a controlled margin of error and fits appropriately with the test data.

This behavior of progressive reduction in consumption, although accurate according to the model's predictions, is unusual and constitutes a warning sign. Under normal conditions, energy consumption in residential and industrial sectors tends to vary around a stable average, especially in geographical regions with constant consumption patterns. The continued reduction observed in the data suggests the possibility of non-technical losses, such as energy theft, which can occur through meter manipulation practices or irregular connections.

5. DISCUSSION

The identification of downward patterns in energy consumption using the ARIMA model and its integration with the XGBoost model suggests that this approach is effective in detecting anomalies that could be associated with non-technical losses. The stabilization of the mean and variance through transformations such as Box-Cox, together with the analysis of autocorrelation and partial correlation (ACF and PACF), has made it possible to capture unusual consumption patterns that are aligned with cases of energy theft. This model has the potential to improve consumption monitoring by offering deeper insight into identifying tampering patterns in meters and irregular connections.

Compared to conventional monitoring and detection methods, this ARIMA-XGBoost-based approach can handle the non-stationary nature of energy consumption data, adjusting to the transformations needed to stabilize the time series. Many previous methods often rely on simple regression or clustering models without considering short-term dependency effects, limiting their ability to capture patterns of manipulation in time series. In this study, the use of the PACF function and correlations in early lags have allowed for an effective autoregressive approach, especially in situations where the data present seasonality or short-term cycles.

6. CONCLUSION

To effectively leverage the performance of predictive models, data management is crucial. He emphasized the importance of providing explanations for model predictions, highlighting that a lack of explanations can pose challenges in understanding model performance. This suggests that incorporating explainability into the data management process can help maximize the potential of predictive models.

The implemented predictive model, using ARIMA for time series analysis and XGBoost for the detection of anomalous patterns, demonstrated high effectiveness in identifying irregular trends in energy consumption. The results indicate an atypical and continuous decrease in consumption, which, under normal conditions, would suggest possible non-technical losses such as energy theft. The model's ability to closely track actual consumption patterns and accurately predict future values within a robust confidence interval validates its applicability in power grid environments where early detection of irregularities is essential for loss management.

Autocorrelation analysis and applied transformations confirm that the time series is stationary, allowing the model to capture seasonal patterns and anomalies with high accuracy. This approach is particularly useful in the context of the energy sector, where the identification of non-technical losses can result in significant economic savings and improved operational efficiency.

In conclusion, the proposed model is not only effective in predicting and detecting anomalous patterns, but also offers a scalable and adaptable solution for consumption monitoring in smart grids. This approach lays the foundation for implementing preventive strategies and improving monitoring mechanisms in the detection of energy theft.

ACKNOWLEDGEMENT

The authors are grateful to the Dirección de Investigación de la Universidad Peruana de Ciencias Aplicadas for the support provided for this research work through the economic incentive.

References

1. Gunturi, S. K., & Sarkar, D., “Ensemble machine learning models for the detection of energy theft”, Electric Power Systems Research., vol. 192, March 2021.
2. Baldawi, I. M. R., & İnan, T., “A novel and effective method based deep learning model for detecting non-technical electricity losses,” International Journal of Power Electronics and Drive Systems., vol. 14, no.4, pp. 2464–2473, December 2023.
3. Ahir, R. K., & Chakraborty, B., “Pattern-based and context-aware electricity theft detection in smart grid,” Sustainable Energy Grids and Networks., vol. 32, December 2022.
4. Toledo-Orozco, M., Arias-Marin, C., Álvarez-Bel, C., Morales-Jadan, D., Rodríguez-García, J., & Bravo-Padilla, E., “Innovative methodology to identify errors in electric energy measurement systems in power utilities,” Energies., vol. 14, February 2021.
5. Ente Nazionale per l'Energia Elettrica., "Hurto de Energía," ENEL., August 2019.
6. Nawaz, A., Ali, T., Mustafa, G., Rehman, S. U., & Rashid, M. R., “A novel technique for detecting electricity theft in secure smart grids using CNN and XG-boost,” Intelligent Systems with Applications., vol. 17, pp. 200168, February 2023.
7. Abdulaal, M. J., Ibrahem, M. I., Mahmoud, M. M. E. A., Khalid, J., Aljohani, A. J., Milyani, A. H., & Abusorrah, A. M., “Real-Time Detection of False Readings in Smart Grid AMI Using Deep and Ensemble Learning,” IEEE Access., vol. 10, pp. 47541–47556, May 2022.
8. Liu, Z., Li, Y., Shuang, F., Huang, Z., & Wang, R., “EMB-YOLO: Dataset, method and benchmark for electric meter box defect detection,” Journal of King Saud University - Computer and Information Sciences., vol. 36(2), pp. 101936, February 2024.
9. Abraham, O. A., Ochiai, H., Hossain, M. D., Taenaka, Y., & Kadobayashi, Y., “Electricity Theft

Detection for Smart Homes: Harnessing the Power of Machine Learning With Real and Synthetic Attacks," IEEE Access., vol. 12, pp. 26023–26045, February 2024.
10. Kawoosa, A. I., Prashar, D., Faheem, M., Jha, N., & Khan, A. A., "Using machine learning ensemble method for detection of energy theft in smart meters," IET Generation, Transmission and Distribution., vol. 17(21), pp. 4794–4809, September 2023.
11. Ahir, R. K., & Chakraborty, B., "Pattern-based and context-aware electricity theft detection in smart grid," Sustainable Energy Grids and Networks., vol. 32, pp. 100833, December 2022.
12. Pamir, Javaid, N., Akbar, M., Aldegheishem, A., Alrajeh, N., & Mohammed, E. A., "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, November 2022.
13. Serra, A., Ortiz, A., Manjarrés, D., Fernández, M., Maqueda, E., Cortés, P. J., & Canals, V., "SpeCluRC-NTL: Spearman's distance-based clustering Reservoir Computing solution for NTL detection in smart grids,". International Journal of Electrical Power and Energy Systems, vol. 157, June 2024.
14. Sebastian, P. K., Deepa, K., Neelima, N., Paul, R., & Özer, T., "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(3), pp. 398–411, August 2023.
15. Coma-Puig, B., Calvo, A., Carmona, J., & Gavaldà, R., "A case study of improving a non-technical losses detection system through explainability," Data Mining and Knowledge Discovery., April 2023.
16. Correia, E., Calili, R., Pessanha, J. F., & Almeida, M. F., "Definition of Regulatory Targets for Electricity Non-Technical Losses: Proposition of an Automatic Model-Selection Technique for Panel Data Regressions," Energies., vol. 16(6), March 2023.
17. Lepolesa, L. J., Achari, S., & Cheng, L., "Electricity Theft Detection in Smart Grids Based on Deep Neural Network," IEEE Access., vol. 10, pp. 39638–39655, April 2022.
18. National Open Data Platform [Internet]. Datosabiertos.gob.pe. 2024 [cited 2024 Oct 30]. Available from: 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-0
19. Peru. Legislative Decree No. 1214, which strengthens Osinergmin's supervisory mechanisms in the energy and mining sector. 2015.