

Modelo Predictivo para la Detección del hurto de la energía Eléctrica

Predictive Model for the Detection of Electrical Energy Theft

 Pedro Martín Lezama Gonzales
plezama@unfv.edu.pe
Universidad Nacional Federico Villarreal, Perú

 Orestes Cachay Boza
ocachayb@unmsm.edu.pe
Universidad Nacional Mayor de San Marcos, Perú

 Máximo Zevallos León
mzevallos@unfv.edu.pe
Universidad Nacional Federico Villarreal, Perú

Resumen

El acceso y uso ilegal de la electricidad, conocido como robo de electricidad, representa una amenaza significativa para la industria energética y la sociedad en su conjunto. Este fenómeno fraudulento socava la integridad del sistema eléctrico, afecta negativamente a los proveedores de servicios y tiene graves consecuencias económicas, sociales y ambientales. Además, los consumos anómalos de electricidad también plantean desafíos importantes en términos de detección temprana de irregularidades y optimización del consumo energético. En esta investigación, se analizaron las causas y motivaciones que impulsan el robo de electricidad. Asimismo, se examinó los diversos métodos utilizados por los infractores para manipular los medidores de energía y ocultar sus actividades fraudulentas, poniendo de relieve la necesidad de soluciones innovadoras para combatir esta problemática. De todas las soluciones existentes nos focalizamos en el análisis y predicción de consumos mediante técnicas de Machine Learning. El empleo de modelos algorítmicos de aprendizaje automático se explora como una herramienta clave para detectar y prevenir el robo de electricidad y consumos anómalos. Presentando una mejora importante en la detección del hurto de la energía eléctrica.

Palabras claves: Machine Learning, algorítmicos de aprendizaje automático, robo de electricidad.

Abstract

Illegal access and use of electricity, known as electricity theft, represents a significant threat to the energy industry and society as a whole. This fraudulent phenomenon undermines the integrity of the electrical system, negatively affects service providers and has serious economic, social and environmental consequences. Furthermore, abnormal electricity consumption also poses significant challenges in terms of early detection of irregularities and optimization of energy consumption. In this research, the causes and motivations that drive electricity theft were analyzed. Likewise, the various methods used by offenders to manipulate energy meters and hide their fraudulent activities were examined, highlighting the need for innovative solutions to combat this problem. Of all the existing solutions, we focus on the analysis and prediction of consumption using Machine Learning techniques. The use of algorithmic machine learning models is explored as a key tool to detect and prevent electricity theft and anomalous consumption. Presenting a significant improvement in the detection of electrical energy theft.

Keywords: Machine Learning, machine learning algorithms, electricity theft.



Publicado: 14/03/2024
Aceptado: 12/03/2024
Recibido: 09/01/2023

Open Access
Article scientific

 <https://doi.org/10.47422/jstri.v5i1.44>

Este artículo es publicado por la Journal of Scientific and Technological Research Industrial, Professionals On Line SAC. Este es un artículo de acceso abierto, distribuido bajo los términos de la Licencia Creative Commons Atribución 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/>) que permite compartir (copiar y redistribuir el material en cualquier medio o formato) y adaptar (remezclar, transformar y construir a partir del material) para cualquier propósito, incluso comercialmente.





INTRODUCCIÓN

La presente investigación se centra en la predicción y detección del robo de electricidad y consumos anómalos mediante técnicas avanzadas de análisis de datos y aprendizaje automático. El objetivo principal es desarrollar un modelo de detección proactiva que permita a las compañías eléctricas identificar patrones sospechosos y comportamientos inusuales en tiempo real, con el fin de prevenir pérdidas financieras y salvaguardar la integridad del sistema.

En este documento, se analizarán las causas y motivaciones que impulsan el robo de electricidad, incluyendo la falta de acceso a servicios legales de energía, la creciente demanda no satisfecha y la búsqueda de beneficios económicos ilícitos.

Asimismo, se examinarán los diversos métodos utilizados por los infractores para manipular los medidores de energía y ocultar sus actividades fraudulentas, poniendo de relieve la necesidad de soluciones innovadoras para combatir esta problemática.

El análisis y predicción de consumos anómalos también serán abordados en este estudio. La detección temprana de patrones inusuales de consumo energético puede ayudar a los usuarios y proveedores de servicios a identificar posibles fallas en equipos, fugas de energía y malos hábitos de consumo, contribuyendo a una gestión más eficiente y sostenible de la electricidad.

El desarrollo de modelos algorítmicos de aprendizaje automático se explorará como una herramienta clave para detectar y prevenir el robo de electricidad y consumos anómalos. La utilización de datos históricos, técnicas de analítica de datos como series de tiempo permitirá identificar patrones sutiles y anomalías que puedan indicar actividades fraudulentas o ineficiencias en el uso de la energía.

Formulación del Problema

Problema General

¿En qué medida el uso de un modelo empírico y analítico predictiva mejorará la detección del hurto de energía eléctrica?

Problemas Específicos

- ¿En qué medida la definición de un modelo empírico mejorará la detección del hurto de energía eléctrica
- ¿En qué medida la definición de los modelos de analítica predictiva mejorará la detección del hurto de energía eléctrica?

Justificación e importancia

Conveniencia:

El desarrollo de la presente investigación servirá para que, a partir de definir un modelo empírico y analítico predictivo para la detección del hurto de la energía eléctrica, se evitarán cortes totales o parciales de energía eléctrica, el uso de recursos naturales escasos, permitiendo una menor contaminación ambiental.

Relevancia Social:

Mediante la aplicación del modelo empírico y analítico predictivo para el hurto de la energía eléctrica, se tendrá un menor incremento en las tarifas eléctricas, e incendios originados por malas conexiones.

Implicaciones Prácticas:

A través de la presente investigación se pretende resolver problemas prácticos que se derivan de la detección de la pérdida no técnica de la energía eléctrica, permitiendo a las empresas prestadoras del servicio eléctrico, ser proactivos en la ejecución de las actividades de visita y focalización de los usuarios que están involucrados en el robo de energía eléctrica.

Valor Teórico:

El modelo planteado, combinará enfoques orientados a mejora de procesos (el uso de criterios estadístico de 1-sigma y 3-sigma), y el empleo de modelo de Machine Learning y Deep Learning, proponiendo un nuevo modelo de detección de las pérdidas no técnicas de energía eléctrica.

Aportes:

Con la presente investigación se podrá mejorar la detección de las personas que hurtan energía eléctrica, permitiendo una mejor usabilidad de recursos naturales escasos, reducir la contaminación, aumento de tarifas eléctricas, incendios entre otros.

Objetivos

Objetivo General

Determinar el grado de mejora en la detección del hurto de energía eléctrica, al implementar un modelo empírico y analítico predictiva.

Objetivos Específicos

- Definir el modelo empírico, para mejorar la detección en el hurto de la energía eléctrica.
- Definir los modelos de analítica predictiva, para mejorar la detección del hurto de la energía eléctrica.



MATERIALES Y METODOS

Se toma en consideración los siguientes puntos:

a) **Ámbito espacial y temporal del estudio**

- **Ámbito Espacial:** La investigación se realizará para los usuarios de la empresa Distriluz-Perú.
- **Ámbito Temporal:** El período de tiempo del cual se tienen los datos es desde el 2010 hasta el 2023, teniendo en cuenta que las mediciones de los medidores eléctricos son realizadas en forma mensual.

b) **Universo o la muestra del estudio**

Universo: Empresa Distriluz, que cuenta con una cobertura de 3 millones de suministros.

n = 150,000 (muestra inicial de las unidades de negocio de Chimbote, Conchucos, Huaraz y La Libertad)

c) **Unidad de análisis**

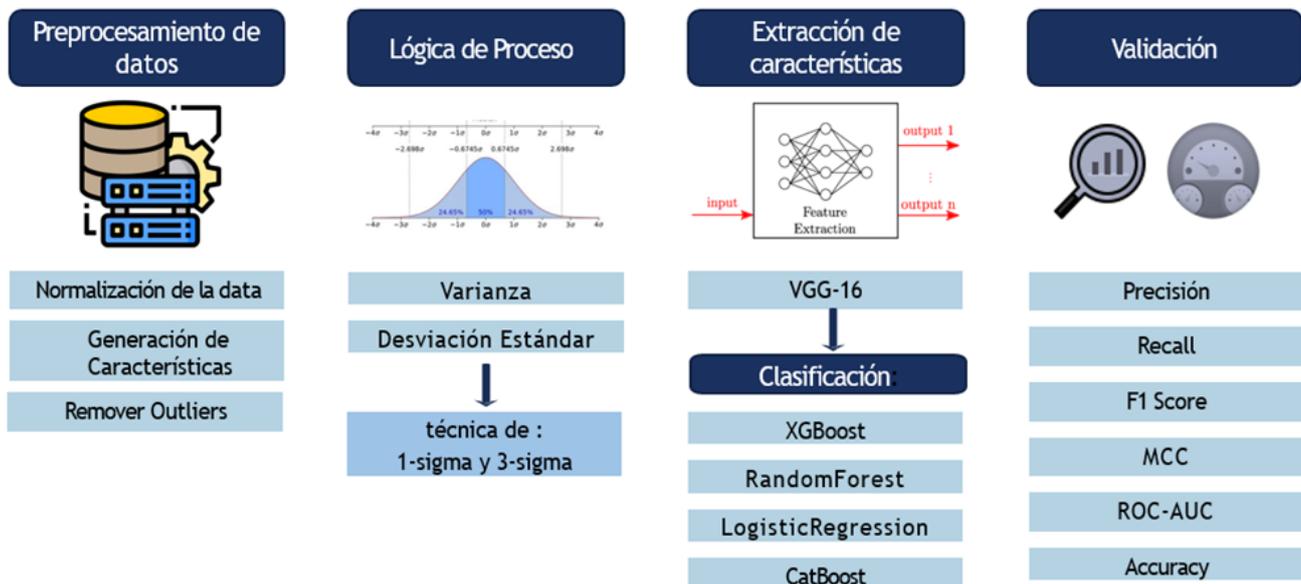
- Detección de las pérdidas no técnicas: Eficacia en la detección del modelo

d) **Técnicas de recolección de datos o Instrumento**

- Análisis documental
- Análisis de Pareto
- Histogramas
- Muestreo
- Recolección de datos
- Tabulación de datos
- Mejora de procesos
- Modelos de Machine Learning
- Modelos de Deep Learning

Figura 1

Procedimiento Propuesto para la Investigación



1. Pre-Procesamiento de datos

a. Datos del consumo de electricidad

Tabla 1

Variables del Dataset

Característica	Descripción
FLAG	Indicador binario que señala si el servicio eléctrico ha sido objeto de hurto (1) o no (0).
Nombre Unidad Negocio	Nombre de la unidad de negocio responsable del suministro eléctrico.
IdNro Servicio	Identificador único del servicio eléctrico.
SED	Identificador de la Subestación de Distribución.
alimentador	Identificador del alimentador eléctrico.



abreviatarifa	Abreviatura de la tarifa eléctrica del servicio.
Tipo Conexión	Tipo de conexión eléctrica (por ejemplo, monofásica o trifásica).
Sector Típico	Categorización del sector según su consumo eléctrico típico.
PeriodoInicio	Fecha de inicio del período de consumo.
consumo	Consumo eléctrico registrado en cada uno de los períodos mensuales desde enero de 2010 hasta julio de 2022.

a) Procedimiento:

El procedimiento propuesto para la siguiente investigación es el siguiente:

Cada característica contiene información relevante sobre los servicios eléctricos y su consumo en diferentes períodos de tiempo. La característica "FLAG" es el objetivo del modelo, ya que indica si un servicio eléctrico ha sido víctima de hurto o no. Las características como "NombreUnidadNegocio", "IdNroServicio", "SED" y "alimentador" proporcionan información sobre la ubicación y el servicio específico. La "abreviatarifa" indica la tarifa eléctrica aplicada al servicio, mientras que "Tipo Conexión" y "Sector Típico" ayudan a categorizar el tipo de conexión eléctrica y el sector del cliente, respectivamente. La característica "PeriodoInicio" marca la fecha de inicio del período de consumo. Por último, las características "consumo" representan el consumo de electricidad registrado en cada período mensual.

b. Data Cleaning

Se llevará a cabo la limpieza y preparación de los datos de consumo de electricidad antes de proceder con el análisis y modelado. El objetivo principal es asegurar que el conjunto de datos esté libre de errores, valores faltantes y registros inválidos. A continuación, se describen los puntos relevantes a realizar en esta fase:

- **Identificación de Datos Faltantes:** Se identificarán las celdas del conjunto de datos que contengan valores nulos o faltantes. Estos valores pueden ser causados por diversos factores, como fallos en la medición o problemas en la recolección de datos.
- **Manejo de Datos Faltantes:** Se seleccionará una estrategia para manejar los datos faltantes, que puede incluir la eliminación de registros con valores faltantes, la imputación de valores utilizando técnicas estadísticas o el uso de algoritmos de imputación más avanzados.
- **Eliminación de Registros Inválidos:** Se eliminarán registros inválidos que no aporten información útil o que puedan afectar negativamente el análisis.

c. Generación de características

Los datos se prepararán para el análisis y modelado específico. Se realizarán transformaciones y manipulaciones en las características para mejorar la

calidad de los datos y garantizar la eficacia del modelo de pronóstico de hurto de electricidad. A continuación, se describen los puntos relevantes a realizar en esta fase:

- **Selección de Características Relevantes:** Se analizarán todas las características disponibles para identificar aquellas que tienen mayor relevancia para el pronóstico de hurto de electricidad. La selección de características ayuda a mejorar la eficiencia computacional y reducir el riesgo de overfitting.

Correlación, La correlación puede medirse utilizando el coeficiente de correlación de Pearson.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Donde:

* x_i y y_i son los valores de las características del hurto para el i -ésimo registro.

* \bar{x} y \bar{y} son las medias de las características y del hurto de electricidad.

LDA (Reducción de Dimensionalidad)

$$Z = X \cdot W$$

Donde

* Z es la matriz de características proyectadas.

* X es la matriz de características centrada.

* W es la matriz de auto-vectores correspondientes a los autovalores más grandes.

ANOVA, permite realizar la comparación de las medias de dos o más grupos para determinar si hay diferencias significativas entre ellos.

$$F = \frac{MS_{Between}}{MS_{within}}$$

Donde:

* F es el estadístico de prueba ANOVA.



* MSBetween es la media de cuadrados de las diferencias entre los grupos (variabilidad explicada por el factor categórico).

* MSWithin es la medida de cuadrados de las diferencias dentro del dato grupos (variabilidad no explicada por el factor categórico)

- Transformación de Datos: Se aplicarán transformaciones a las características si es necesario, como logaritmos, exponentes u otras funciones que ayuden a linealizar relaciones no lineales en los datos.

Escalado, se utiliza para escalar las características a un rango común, generalmente entre 0 y 1, o a una distribución específica.

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Donde:

* x es el valor original de la característica.

*xnorm es el valor normalizado

Logarítmica, se utiliza para reducir la variabilidad de datos que presentan una distribución sesgada o exponencial

$$x_{log} = \log(x)$$

Donde:

* x es el valor original de la característica.

*xlog es el valor transformado

- Codificación de Variables Categóricas: Si existen variables categóricas, se convertirán en variables numéricas para su inclusión en el modelo.
- One-Hot Encoder: técnica de codificación utilizada para convertir variables de tipo categóricas en una representación numérica. Para cada valor único en la variable definida como categórica, se crea una nueva columna binaria, donde el 1 representa la presencia de la categoría y un 0 representa la ausencia.

$$\begin{cases} 1, \text{ si } X_{ij} = \text{Categoría}_k \\ 0, \text{ en otro caso} \end{cases}$$

Donde:

* X_{ij} es el valor de la variable caregorica para la i-ésima fila y j-ésima categoría.

* Caegoría_k es la k-ésima categoría única en la variable categorica.

- División del Conjunto de Datos: Se dividirá el conjunto de datos, en conjuntos de datos de entrenamiento y conjunto de datos de prueba para evaluar la eficacia del modelo en datos no vistos.
- Manejo de Desequilibrio de Clases: Si existe un desequilibrio significativo entre las clases (hurto y no hurto), se aplicarán técnicas de balanceo de clases para mejorar el performance del modelo.

SMOTE (Synthetic Minority Over-sampling Technique), técnica de sobremuestreo que genera muestras sintéticas de las clases minoritarias para igualar la cantidad de muestras en cada clase

$$A_{sintetica} = x_i + rand(\text{vecino} - x_i)$$

Donde:

* x_i es una muestra de la clase minoritaria A.

* Vecino es uno de los vecinos cercanos de x_i.

* rand número aleatorio entra 0 y 1 que controla el equilibrio entra x_i y Vecino.

2. Lógica de Proceso

a. Análisis de los datos empleando la técnica de 1-sigma y 3-sigma

Se realizará un análisis exhaustivo de los datos de consumo de electricidad utilizando la técnica de 1-sigma y 3-sigma. Esta técnica es útil para identificar y manejar valores atípicos (outliers) en el conjunto de datos. Los outliers pueden afectar negativamente el rendimiento y la precisión del modelo de pronóstico de hurto de electricidad, por lo que es esencial identificarlos y tratarlos adecuadamente.

- Cálculo de Media y Desviación Estándar: Se calculará la media (μ) y la desviación estándar (σ) para cada característica del conjunto de datos. Estos valores son fundamentales para determinar el rango dentro del cual la mayoría de los datos se encuentra.
- Definición de Límites: Utilizando la técnica de 1-sigma y 3-sigma, se definirán los límites inferior y superior para cada característica. Los valores que caigan fuera de estos límites se considerarán outliers.
- Identificación de Outliers: Se identificarán los valores atípicos en cada característica del conjunto de datos. Estos outliers pueden ser causados por errores en la medición, registros incorrectos o situaciones excepcionales.
- Tratamiento de Outliers: Los outliers identificados pueden tratarse de diferentes maneras, como eliminación, imputación o ajuste a los límites. La elección del método dependerá del análisis y del conjunto de datos específico.



La regla establece lo siguiente para un conjunto de datos que sigue una distribución normal:

- Aproximadamente el 68% de los datos se encuentran dentro de una desviación estándar (σ) de la media (promedio).
- Aproximadamente el 95% de los datos se encuentran dentro de dos desviaciones estándar (2σ) de la media.
- Aproximadamente el 99.7% de los datos se encuentran dentro de tres desviaciones estándar (3σ) de la media.

En términos matemáticos, si X representa la variable aleatoria en un conjunto de datos que sigue una distribución normal, con media (μ) y desviación estándar (σ), la regla se puede expresar como:

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &\approx 0.68 \\ P(\mu - 2\sigma < X < \mu + 2\sigma) &\approx 0.95 \\ P(\mu - 3\sigma < X < \mu + 3\sigma) &\approx 0.997 \end{aligned}$$

Varianza

$$\text{Varianza}_i = (\text{Consumo}_i - \text{Consumo}_{i-1})^2$$

Promedio

$$\text{Promedio} = \frac{1}{n} \sum_{i=1}^n \text{Consumo}_i$$

Desviación Estándar

$$\begin{aligned} &\text{Desviación Estándar} \\ &= \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Consumo}_i - \text{promedio})^2} \end{aligned}$$

1. Extracción de características

VGG16 es una red neuronal convolucional profunda diseñada para clasificar imágenes. Consiste en 16 capas, incluyendo capas de convolución y capas completamente conectadas. En lugar de trabajar directamente con las características extraídas en la fase anterior, VGG16 puede procesar directamente las imágenes originales como entradas. Durante el entrenamiento, la red aprende automáticamente las características relevantes a partir de los datos de consumo. La predicción final se obtiene a través de la capa del vector unidimensional, que asigna probabilidades a cada clase y elige la clase con la probabilidad más alta como la predicción final.

2. Clasificación

Random Forest:

Random Forest es un algoritmo de aprendizaje que combina varios árboles de decisión buscando mejorar la precisión y reducir el sobreajuste. Cada árbol en Random Forest se entrena con una muestra aleatoria de los datos de

entrenamiento, y durante la clasificación, cada árbol vota por la clase final. La predicción del Random Forest se obtiene promediando las predicciones de todos los árboles. La ventaja de este enfoque es que evita la sobreoptimización y tiene una alta precisión en la clasificación.

XGBoost (Extreme Gradient Boosting):

XGBoost es un algoritmo de aprendizaje supervisado basado en árboles de decisión que se enfoca en mejorar la velocidad y el rendimiento del modelo. A diferencia de Random Forest, XGBoost utiliza árboles de decisión débiles de forma secuencial, donde cada nuevo árbol se ajusta a los errores cometidos por los árboles anteriores. Utiliza una función de pérdida específica (como regresión logística para clasificación) para cuantificar los errores del modelo y ajustar los pesos de las instancias en cada iteración. XGBoost también incorpora regularización para controlar la complejidad del modelo y prevenir el sobreajuste. Su capacidad para manejar datos faltantes y su eficiencia lo convierten en una elección popular para problemas de clasificación y regresión.

Logistic Regression:

Es un modelo lineal para la clasificación, que utiliza la función logística para estimar la probabilidad de que una instancia pertenezca a una clase particular. Se ajusta a los datos de entrenamiento utilizando el método de máxima verosimilitud, minimizando la función de pérdida logarítmica. La salida de la regresión logística se transforma mediante la función logística, que produce valores en el rango $[0, 1]$, representando las probabilidades de pertenencia a la clase positiva. Para tomar decisiones de clasificación, se establece un umbral, y las instancias con probabilidades superiores al umbral se clasifican como positivas, mientras que las inferiores se clasifican como negativas. A pesar de su simplicidad, la regresión logística es eficaz en problemas lineales y es especialmente útil cuando se desea interpretar el impacto de cada característica en la predicción.

CatBoost:

CatBoost es un algoritmo de aprendizaje automático basado en árboles que está diseñado para manejar de manera eficiente variables categóricas sin requerir una codificación previa. Utiliza una estrategia de propagación de la gradiente que optimiza la función de pérdida a lo largo de las direcciones de los gradientes estimados, mejorando así la convergencia del modelo. CatBoost también incorpora técnicas de regularización para prevenir el sobreajuste. Su capacidad para manejar datos categóricos de manera nativa y su rendimiento robusto en una variedad



de conjuntos de datos lo hacen destacar en comparación con otros algoritmos.

Figura 2

Modelos y su configuración

XGBoost	Modelo	learning_rate	n_estimators	max_depth
	XGBoost	0.1	125	15
CatBoost	Modelo	learning_rate	iterations	depth
	CatBoost	0.1	100	20
Random Forest	Modelo	n_estimators	max_depth	
	Random Forest	200	18	
Regresión Logística	Modelo	C	max_iter	
	Regresión Logística	1	100	

3. Evaluación del desempeño en el campo

Una vez que hemos desarrollado y entrenado nuestros modelos utilizando técnicas de clasificación y agrupación para el pronóstico de hurto de electricidad, es crucial evaluar su desempeño en un entorno real o campo para asegurar que sean efectivos y puedan ser implementados con éxito en la detección de hurto de electricidad en la operación diaria de la compañía eléctrica.

Monitoreo del rendimiento:

Una vez implementados, es esencial monitorear constantemente el rendimiento de los modelos en un entorno de producción. Esto implica registrar y analizar las predicciones realizadas por los modelos en comparación con los casos reales de hurto de electricidad detectados y confirmados por la compañía. Esto nos permitirá evaluar la Acuracy, precisión y recall de los modelos en situaciones reales y realizar ajustes si es necesario.

DISCUSIÓN Y RESULTADOS

Marco de Trabajo y Preprocesamiento de Datos:

El enfoque empírico y analítico para la detección de pérdidas no técnicas de energía eléctrica se basó en un proceso integral de preprocesamiento de datos. El preprocesamiento incluyó la normalización de datos para estandarizar las características, la generación de nuevas características relevantes y la eliminación de outliers para mejorar la calidad de los datos. La aplicación de técnicas basadas en la varianza y la desviación estándar, especialmente la técnica de 1 y 3 sigma, permitió particionar eficazmente los datos, identificando posibles casos de hurto de electricidad.

Modelo de Detección:

Se implementaron varios modelos para la detección de pérdidas no técnicas, cada uno con su enfoque único en la extracción y clasificación de características. El modelo VGG16 fue utilizado para la extracción de características, seguido por algoritmos de clasificación como XGBoost, Random Forest, Logistic Regression y CatBoost.

Validación y Métricas de Evaluación:

La validación del rendimiento de los modelos se llevó a cabo mediante la evaluación de métricas clave, siendo esta precisión, recall, F1-Score, MCC (Coeficiente de Correlación Matthews), ROC-AUC y exactitud (accuracy). Estas métricas proporcionan una visión exhaustiva de la capacidad predictiva de los modelos en la identificación de casos de hurto de electricidad y la distinción entre las clases.

Figura 3

Evaluación Accuracy

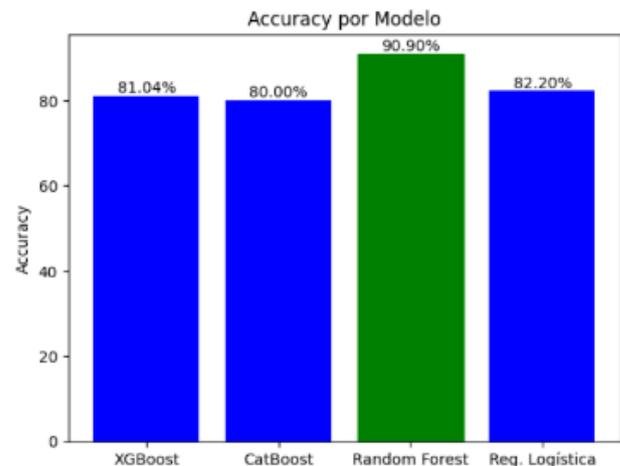


Figura 4

Evaluación MCC

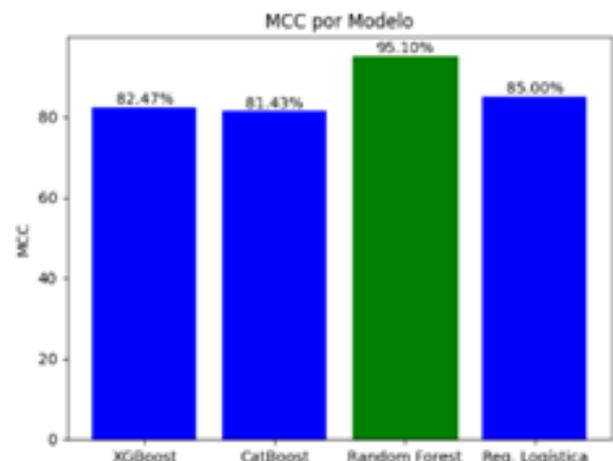




Figura 5

Evaluación AUC

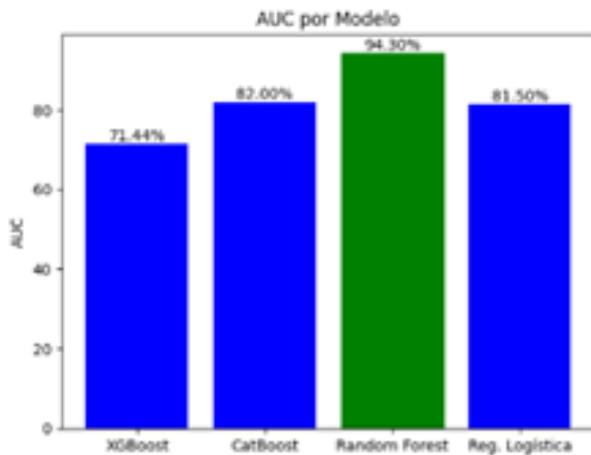
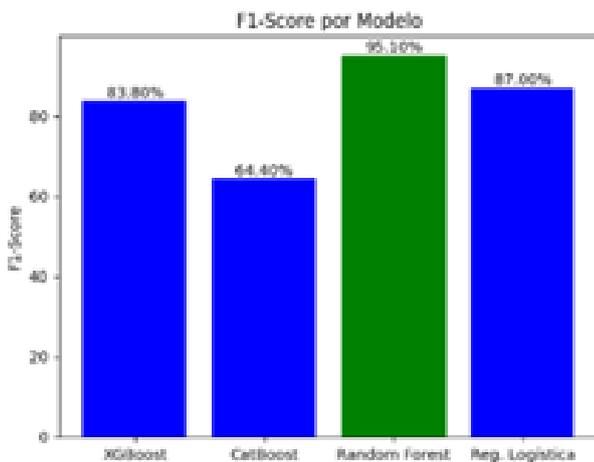


Figura 6

Evaluación F1-Score



En el marco de la investigación comparativa de modelos, el análisis revela que el Random Forest emerge como la opción más destacada debido a sus sólidas métricas de rendimiento. Con una impresionante exactitud del 95.0%, el modelo logra un alto porcentaje de predicciones correctas, indicando su eficacia general. El Coeficiente de Correlación Matthews (MCC) sobresale con un valor del 90.9%, demostrando un rendimiento excepcional y equilibrado en la clasificación.

El Área bajo la Curva ROC (AUC) del 95.1% subraya la excepcional capacidad del modelo para discriminar entre clases, un aspecto crítico en problemas de clasificación. El F1-Score, que alcanza un impresionante 94.3%, refleja un equilibrio notable entre precisión y recall, fortaleciendo la confianza en la validez general de las predicciones realizadas por el modelo.

En términos de sensibilidad (Recall), el modelo Random Forest identifica el 95.1% de los casos positivos, subrayando su capacidad para capturar eventos relevantes. La precisión del 94.2% destaca la exactitud en las predicciones positivas, consolidando la posición del Random Forest como el modelo más destacado en esta evaluación comparativa.

CONCLUSIONES

Inicialmente, los datos reales del medidor, que se recopilan de una empresa Distribuidora de luz (empresa peruana) en donde se tiene varios valores atípicos; con lo cual se procedió a realizar un preprocesamiento exhaustivo, de la regla de las tres sigmas y los métodos de normalización, aplicar una modelo variedad de modelos, dentro de los cuales Random Forest emerge como la opción más destacada debido a sus sólidas métricas de rendimiento. Con una impresionante exactitud del 95.0%, el modelo logra un alto porcentaje de predicciones correctas, indicando su eficacia general. El Coeficiente de Correlación Matthews (MCC) sobresale con un valor del 90.9%, demostrando un rendimiento excepcional y equilibrado en la clasificación. El Área bajo la Curva ROC (AUC) del 95.1% subraya la excepcional capacidad del modelo para discriminar entre clases, un aspecto crítico en problemas de clasificación. El F1-Score, que alcanza un impresionante 94.3%.

REFERENCIAS BIBLIOGRÁFICAS

1. Amin, S., Schwartz, G., Cardenas, A., & Sastry, S. (2015). Gametheoretic models of electricity theft detection in smart utility networks: Providing new capabilities with advanced metering infrastructure. *IEEE Control. Syst. Mag.*
2. Avila, N., Figueroa, G., & Chu, C. (2018). NTL detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random under sampling boosting. *IEEE Trans. Power Syst.*
3. Biswas, P., Cai, H., Zhou, B., Chen, B., Mashima, D., & Zheng, V. (2019). Electricity Theft Pinpointing through Correlation Analysis of Master and Individual Meter Readings. *IEEE Trans.*
4. Buzau, M., & Tejedor, J. (2018). Detection of non-technical losses using smart meter data and supervised learning. *EEE Trans. Smart Grid.*
5. Buzau, M., Tejedor, J., Cruz, P., & Gomez, A. (2019). Hybrid deep neural networks for detection of non-technical losses in electricity smart meters. *IEEE Trans. Power Syst.*
6. Ding, N., Ma, H., Gao, H., & Tan, G. (2019). Real-time anomaly detection based on long short-Term



- memory and Gaussian Mixture Model. *Comput. Electr. Eng.*
7. ENEL. (2021). Hurto de Energía - enel.pe. Obtenido de <https://www.enel.pe/es/ayuda/hurto-de-energia.html>
 8. Glauner, P., Valtchev, P., Glaeser, C., Dahringer, N., State, R., & Duarte, D. (2018). Non-Technical Losses in the 21st Century: Causes, Economic Effects, Detection and Perspectives. Obtenido de <https://www.researchgate.net/publication/325297875>
 9. Hammerschmitt, B. (2020). Non-Technical Losses Review and Possible Methodology Solutions. *Proceedings - 2020 6th International Conference on Electric Power and Energy Conversion Systems, EPECS*, 64–68. doi:10.1109/EPECS48981.2020.9304525
 10. Hasan, M., Toma, R., Nahid, A., Islam, M., & Kim, J. (2019). Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach.
 11. Jamil, A., Alghamdi, T., Khan, Z., Javaid, S., Haseeb, A., Wadud, Z., & Javaid, N. (2019). An Innovative Home Energy Management Model with Coordination among Appliances using Game Theory. *Sustainability*.
 12. Jiménez, R., Serebrisky, T., & Mercado, J. (2014). Power Lost: Sizing Electricity Losses in Transmission and Distribution Systems in Latin America and the Caribbean. *Inter-American Development Bank*. doi:10.18235/0001046.
 13. Leite, J., & Mantovani, J. (2016). Detecting and locating non-technical losses in modern distribution networks. *IEEE Trans. Smart Grid*.
 14. Li, S., Han, Y., Yao, X., Yingchen, S., Wang, J., & Zhao, Q. (2019). Electricity Theft Detection in Power Grids with Deep Learning and Random Forests. *Electr. Comput. Eng.*
 15. Lydia, M., Kumar, G., & Levron, Y. (2019). Detection of Electricity Theft based on Compressed Sensing. In *Proceedings of the 2019 5th International Conference on Advanced Computing and Communication Systems (ICACCS) IEEE*. Coimbatore, India.
 16. McDaniel, P., & McLaughlin, S. (2009). Security and privacy challenges in the smart grid. *IEEE Secur. Priv.*
 17. Ramos, C., Rodrigues, D., de Souza, A., & Papa, J. (2016). On the study of commercial losses in Brazil: a binary black hole algorithm for theft characterization. *IEEE Trans. Smart Grid*.
 18. Razavi, r., & Fleury, m. (2019). Socio-economic predictors of electricity theft in developing countries: An Indian case study. *Energy Sustain. Dev.*
 19. Razavi, R., Gharipour, A., Fleury, M., & Akpan, I. (2019). A practical feature-engineering framework for electricity theft detection in smart grids. *Appl. Energy*.
 20. Saeed, M., Mustafa, M., Sheikh, U., Jumani, T., & Mirjat, N. (2019). Ensemble Bagged Tree Based Classification for Reducing Non-Technical Losses in Multan Electric Power Company of Pakistan.
 21. Savian, F., Siluk, J., Garlet, T., Nascimento, F., Pinheiro, J., & Vale, Z. (2021). Non-technical losses: A systematic contemporary article review. *Renewable and Sustainable Energy Reviews*. doi:10.1016/J.RSER.2021.111205
 22. Wang, S., & Chen, H. (2019). A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network. *Appl. Energy*.
 23. Zahoor, A., Muhammad, A., Nadeem, J., Malik, S., Muhammad, S., & Jin-Ghoo, C. (2020). Electricity Theft Detection Using Supervised Learning Techniques on Smart Meter Data. *MDPI*.
 24. Zheng, K., Chen, Q., Wang, Y., Kang, C., & Xia, Q. (2019). A novel combined data-driven approach for electricity theft detection. *IEEE Trans. Ind. Inform*