J-40402082-9

Fundación Aula Virtual

# Aula Virtual



## Generando Conocimiento

http://www.aulavirtual.web.ve

ISSN: 2665-0398

Deposito Legal: LA2020000026

Vol. 7 Nº 15 Año 2026

Periodicidad Continua

AulaVirtual
Generando Conocimiento

**ISSN**: 2665-0398

# REVISTA CIENTÍFICA AULA VIRTUAL

**Director Editor:**

- Dra. Leidy Hernández PhD.
- Dr. Fernando Bárbara

**Consejo Asesor:**

- MSc. Manuel Mujica
- MSc. Wilman Briceño
- Dra. Harizmar Izquierdo
- Dr. José Gregorio Sánchez

**Revista Científica Arbitrada de Fundación Aula Virtual**

**Email:** revista@aulavirtual.web.ve
**URL:** http://aulavirtual.web.ve/revista




**ISSN:** 2665-0398
**Depósito Legal:** LA2020000026
**País:** Venezuela
**Año de Inicio:** 2020
**Periodicidad:** Continua
**Sistema de Arbitraje:** Revisión por pares. "Doble Ciego"
**Licencia:** **Creative Commons** CC BY NC ND
**Volumen:** 7
**Número:** 15
**Año:** 2026
**Período:** Julio 2026 – Diciembre 2026 (continua)
**Dirección Fiscal:** Av. Libertador, Arca del Norte, Nro. 52D, Barquisimeto estado Lara, Venezuela, C.P. 3001

La Revista seriada Científica Arbitrada e Indexada **Aula Virtual**, es de acceso abierto y en formato electrónico; la misma está orientada a la divulgación de las producciones científicas creadas por investigadores en diversas áreas del conocimiento. Su cobertura temática abarca Tecnología, Ciencias de la Salud, Ciencias Administrativas, Ciencias Sociales, Ciencias Jurídicas y Políticas, Ciencias Exactas y otras áreas afines. Su publicación es **CONTINUA**, indexada y arbitrada por especialistas en el área, bajo la modalidad de doble ciego. Se reciben las producciones tipo: *Artículo Científico* en las diferentes modalidades cualitativas y cuantitativas, *Avances Investigativos*, *Ensayos, Reseñas Bibliográficas, Ponencias o publicaciones derivada de eventos*, y cualquier otro tipo de investigación orientada al tratamiento y profundización de la información de los campos de estudios de las diferentes ciencias. La Revista **Aula Virtual,** busca fomentar la divulgación del conocimiento científico y el pensamiento crítico reflexivo en el ámbito investigativo.


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

**ISSN**: 2665-0398



# EVALUACIÓN COMPARATIVA DE MODELOS DE APRENDIZAJE AUTOMÁTICO PARA LA DETECCIÓN EN EL ROBO DE ELECTRICIDAD

# COMPARATIVE EVALUATION OF MACHINE LEARNING MODELS FOR DETECTING ELECTRICITY THEFT

**Tipo de Publicación:** Artículo Científico
**Área del Conocimiento**: Ciencias Sociales y Aplicadas
**Recibido:** 05/08/2026
**Aceptado:** 06/09/2026
**Publicado:** 07/09/2026
**Código Único AV:** e826
**Páginas:** 1(1365-1386)
**DOI:** https://doi.org/10.5281/zenodo.22646367

**Autores:**
**Ivan Carlo Petrlik Azabache**
Ingeniero de Computación y Sistemas
Maestría en Ingeniería de Sistemas
Doctorado en Ingeniería de Sistemas
https://orcid.org/0000-0002-1201-2143
**E-mail:** ipetrlik@unfv.edu.pe
**Afiliación:** Universidad Nacional Federico Villarreal
**País:** Republica del Perú

**Pablo Roberto Aparicio Montenegro**
Ingeniero Industrial
Maestría en Ingeniería de Sistemas
https://orcid.org/0000-0001-6034-9536
**E-mail:** paparicio@unfv.edu.pe
**Afiliación:** Universidad Nacional Federico Villarreal
**País:** Republica del Perú

**Resumen**

La sustracción de electricidad acarrea pérdidas no técnicas notables, así como también constituyen un reto para las empresas distribuidoras debido al desbalance y complejidad que implican los datos de consumo. El presente estudio tiene como objetivo evaluar comparativamente diferentes modelos de aprendizaje automático supervisado con el objetivo de detectar el fraude eléctrico. La metodología aplicada hace uso de datos reales obtenidos de medidores inteligentes, preprocesado, normalización y balanceo a través de SMOTE, introduciendo al mismo tiempo ocho modelos de clasificación. Los resultados de *accuracy* fueron los siguientes: Random Forest 86,19 %, SVM 83,91 %, Naive Bayes 82,38 %, Árbol de Decisión 81,18 %, Regresión Logística 80,92 %, XGBoost 78,74 %, KNN 77,65 %, y AdaBoost 73,94 %, donde el mejor modelo fue el correspondiente a Random Forest. Se concluye que la detección del robo de electricidad en contextos reales aún constituye un problema que se enfrenta como un gran reto, donde formulaciones más robustas son las que deben ser consideradas.

**Palabras Clave** Robo de electricidad, aprendizaje automático, medidores inteligentes, detección de fraude

**Abstract**

Electricity theft results in significant non-technical losses and also poses a challenge for distribution companies due to the imbalance and complexity inherent in consumption data. The objective of this study is to comparatively evaluate different supervised machine learning models for the purpose of detecting electricity fraud. The methodology applied uses real data obtained from smart meters, which was preprocessed, normalized, and balanced using SMOTE, while simultaneously testing eight classification models. The accuracy results were as follows: Random Forest 86.19%, SVM 83.91%, Naive Bayes 82.38%, Decision Tree 81.18%, Logistic Regression 80.92%, XGBoost 78.74%, KNN 77.65%, and AdaBoost 73.94%, with Random Forest emerging as the best model. It is concluded that detecting electricity theft in real-world contexts remains a significant challenge, and more robust approaches should be considered.

**Keywords** Electricity theft, machine learning, smart meters, fraud detection


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

## Introducción

El crecimiento poblacional y el progreso tecnológico han ido aumentando continuamente la demanda de electricidad. En este sentido, los proveedores de energía tratan de abocarse a cubrir ese crecimiento, aunque, a la larga, sufren problemas operativos que suponen enormes pérdidas económicas en los sistemas eléctricos (Saha, 2023). Estos problemas, que presentan pérdidas de energía en las redes de transmisión y distribución y la detección de robos de electricidad, son significativos para las empresas de electricidad en todo el mundo (Ghaedi et al., 2021).

El robo de electricidad ha crecido al igual que el aumento del consumo eléctrico, lo que justifica la falta de eficiencia en el funcionamiento de las redes eléctricas. Además, uno de los principales causantes de las pérdidas de electricidad no técnicas, se refiere a cualquier tipo de manipulación de medidores, empalmes ilegales y errores deliberados en el proceso de facturación. Es así que el robo de electricidad se ha normalizado y ha pasado a ser un inconveniente generalizado, el cual ocasiona importantes pérdidas económicas, tanto a los suministradores como a los consumidores (Khan et al., 2024). En definitiva, a pesar de los esfuerzos desplegados aún no se ha llegado a encontrar una solución pragmática y definitiva que anule las pérdidas existentes y que, conlleva porcentajes importantes en diferentes países (Savian et al., 2021; Ahmad, 2017).

Con respecto a sus efectos económicos, debe tenerse en cuenta que las actividades ilegales que el robo genera se traducen en unos serios riesgos para la seguridad eléctrica que pueden ser considerados como incendios, descargas eléctricas, o cortes de electricidad de dimensiones considerables. Asimismo, la importancia de una correcta identificación del robo de electricidad en el sentido de minimizar las pérdidas de los ingresos, contribuye a la justicia de las transacciones en la electricidad, o incluso salvar la seguridad de las personas consumidoras de electricidad (Chen et al., 2023; Emadaleslami et al., 2023).

La detección del robo eléctrico es una de las formas de detección más elementales de los fraudes eléctricos y que se basa en la actividad permanente de las personas que permiten determinar patrones anómalos que pueden corresponder a una determinada actividad ilícita, pero como este tipo de método del que se ha hablado no es suficiente para aclarar en última instancia el caso ( ya que se requiere la inspección in situ), se intenta centrar la atención en aquellos casos donde se puede suponer una mayor probabilidad de que haya un fraude, porque la verificación de todos los datos de medición es cara y laboriosa (Tsao et al., 2024; Emadaleslami et al., 2023).



Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 15, Año: 2026 (Julio 2026 - diciembre 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

En este escenario las redes y los medidores inteligentes, permite comportar grandes volúmenes de datos de consumo eléctrico que son relevantes para el seguimiento del comportamiento de las personas; Así, estos datos constituyen la base del aprendizaje automático y del aprendizaje profundo para el reconocimiento del robo de energía, ya que permiten encontrar patrones anómalos del consumo de la energía eléctrica (Elshennawy et al., 2025; Yang et al., 2023; Zidi et al., 2023). A pesar de esto, la escasez de muestras de robos y el desequilibrio entre clases limitan el rendimiento de los métodos, dificultando el reconocimiento de comportamientos fraudulentos (Emadaleslami et al., 2023; Ge et al., 2025).

A su vez, los métodos actuales de detección son aquellos que extraen patrones del consumo de la energía dado los datos específicos de cada cliente. De hecho, en general, estos métodos son válidos para el reconocimiento múltiples robos, pero las diferencias en los hábitos de consumo de diferentes usuarios se extraen de forma pasiva en el aprendizaje de los modelos iterativos en donde el margen de mejora de la detección es limitado (Jokar et al., 2016). Al mismo tiempo, se han encontrado seis categorías de robo de la electricidad que proceden de modificaciones en los patrones de consumo normalizados, desde reducciones sencillas de patrones a modificaciones más complejas que se producen en un intervalo de tiempo. Los métodos actuales de detección los cuales se fundamentan en la comparación de los patrones locales respecto a las características globales pueden hacer frente a casi todos los casos, sin embargo, poseen limitaciones para detectar el robo de la tarifa plana y para realizar comparaciones activas entre los usuarios (Liu et al., 2023).

La problemática se agrava aún más porque solamente existe un único medidor inteligente para medir la energía generada y consumida, dado que se trata de sistemas de medida neta. En los mencionados sistemas de medida, si el medidor devuelve lecturas negativas o positivas se debe en última instancia a las lecturas que efectúe a el balance energético del propio usuario, siendo pues relevantes los reembolsos y los pagos adicionales que se tienen que realizar. Dicha propiedad puede ser explotada por ciertos usuarios engañando al sistema, reportando menor consumo o una cantidad mayor de producción eléctrica fotovoltaica para así conseguir compensaciones por un uso indebido de la energía producida, lo que a su vez incrementa la probabilidad del fraude eléctrico (Blazakis et al., 2025).

Al final, el aumento del precio unitario del kWh ha propiciado y, a su vez, ha favorecido el aumento del uso ilícito de la corriente eléctrica, al tiempo que ha propiciado también en la reducción de las tasas de pago de clientes incautos por esta práctica. Y bajo estas consignas, para las empresas de distribución eléctrica, es ciertamente importante


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

no sólo la detección del robo de electricidad, sino también la estimación de la capacidad de pago de sus clientes potencialmente infractores con el que lograr aumentar la rentabilidad y el flujo de caja (Özay & Çakıt, 2025; Iqbal et al., 2025).

En esta misma dirección se ha comprobado que hay enfoques que deben ser explícitos y evaluados para la detección de la apropiación de la electricidad que sean más robustos y serviciales, donde pueda utilizarse de la mejor manera posible la generación que pueden realizar los medidores inteligentes, y que sean robustos para las limitaciones a las que nos encontramos hoy: desbalance de datos, escasez de muestras de fraude y patrones elementales de consumo. En esta línea interesa principalmente valorar comparativamente modelos de aprendizaje automático, para poder investigar el comportamiento, la buena generalización y la detección del comportamiento fraudulento en el trabajo con diferentes patrones de consumo eléctrico. De este modo, queremos poder llegar a afinar las técnicas que puedan robustecer las técnicas de detección de significados de la electricidad, para combatir las pérdidas no técnicas en las redes eléctricas.

## Revisión Bibliográfica

En consonancia con lo que argumentan Taruna et al., (2025), el robo de electricidad en usuarios residenciales con contadores convencionales es un problema que se ha tratado en el cual la detección depende de procedimientos manuales relativamente ineficientes. Como objetivo, propone la mejora de un modelo de aprendizaje automático, que permita optimizar el proceso de selección de clientes para realizar la inspección de las lecturas mensuales de los consumos eléctricos.

Para ello, la propuesta considera modificar los algoritmos de clasificación y también empezar a aplicar un esquema de una evaluación secuencial destacándose el Bosque Aleatorio + K-Vecinos Más cercanos. Los resultados numéricos demuestran un elevado rendimiento del modelo (exactitud 0,89 y F1 0,90), concluyendo al final que la investigación mejora la confiabilidad, reduce los errores humanos y optimiza la eficiencia operativa de la compañía.

De otro lado, Ullah et al., (2025), abordan el problema del robo de electricidad como responsable de pérdidas no técnicas en redes inteligentes mediante la observación de estos datos de consumo desbalanceados y la gran variabilidad del mismo. Su objetivo es mejorar la detección a partir de un análisis comparativo entre el aprendizaje automático y el aprendizaje profundo. En su metodología destaca el preprocesamiento, el remuestreo híbrido y el aprendizaje apilado, llevando a cabo la comparación entre diferentes clasificadores base como el Random Forest, SVM, AdaBoost, Árbol de Decisión y metaclasificadores como LSTM, GRU, Bi-LSTM, Bi-GRU y CNN.



Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 15, Año: 2026 (Julio 2026 - diciembre 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

La sección de resultados, mediante un análisis cuantitativo, muestra cómo el Bi-GRU con remuestreo híbrido trabaja para lograr la mayor precisión (99%), concluyendo en que el enfoque propuesto supera los modelos individuales y mejora significativamente la precisión y confiabilidad de la detección del robo de electricidad.

Del mismo modo, Pamir et al., (2023), abordan el aumento de las pérdidas no técnicas por el robo de electricidad, problema que ha aparecido entre diferentes impactos económicos y que distorsiona el suministro con baja calidad en redes inteligentes. En cuanto a su objetivo, es obtener un modelo de detección más eficaz que les permita superar las limitaciones de los métodos de aprendizaje automático convencionales. La metodología plantea el modelo SSA-GCAE-CSLSTM, donde se define una preparación y balanceo de datos, una reducción de dimensionalidad, la optimización de hiperparámetros y la clasificación, todo ello evaluado con datos reales y al que se compara con modelos clásicos y de referencia. Dichos resultados ofrecen unas respuestas cuantitativas, donde el modelo propuesto alcanza una precisión del 99,45 % y un f1-score del 95,93 %, concluyendo con un incremento en la eficacia de forma significativa respecto a la detección del robo de electricidad en las redes inteligentes.

A continuación, Saqib et al., (2026) presentan el robo de electricidad como un problema de naturaleza global que provoca pérdidas económicas y sobrecarga en la infraestructura y que a su vez, pone nuestra seguridad pública en riesgo ya que los métodos de detección actuales no son capaces de abordar esta problemática y es por esto que el objetivo es buscar un método más eficaz que permita subsanar el problema de la detección del fraude eléctrico en conjuntos de datos complejos y desequilibrados.

La metodología ofrece un uso conjunto de los modelos de Machine Learning, se utilizan técnicas de votación y apilamiento con regresión logística y random forest, además de utilizar ADASYN para equilibrar clases. Los resultados cuantitativos muestran que el modelo de apilamiento alcanza una puntuación en exactitud total de 0,94 y un ROC-AUC igualmente cercano a 0,94; concluyendo que la propuesta permite mejorar de manera significativa y robusta la detección del robo de electricidad, a través de la mejora en la detección de la clase minoritaria, sin perjudicar el rendimiento general.

Por otro lado, Iftikhar et al., (2024) establecen que el robo de electricidad representa una de las principales causas que subyacen al origen de las pérdidas no técnicas en las redes las eléctricas inteligentes. y esta situación es aún más grave si tenemos en cuenta que existe una serie de datos


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

desequilibrados de alta dimensión que afectan la eficacia de los métodos de detección más clásicos. En este sentido, el artículo propone un modelo de aprendizaje híbrido automático y profundo con tal de lograr una mejora en la detección del fraude eléctrico, a partir de datos reales e históricos de consumo.

Asimismo, la metodología combina el preprocesamiento de datos, el balanceo de datos a partir de la combinación k - means SMOTE y un modelo híbrido (Multilayer Perceptron- Gated Recurrent Unit, MLP-GRU), que se compara con AlexNet, GRU, BGRU y RNN (Accuracy de 87, 63 %, 74, 98 %, 83, 55 % y 83, 13 %, respectivamente) para diferentes configuraciones de entrenamiento. Los valores obtenidos para el modelo propuesto muestran que la precisión alcanza los un 93,33 % y el F1-score un 94 %. En consecuencia, la propuesta de trabajo ofrece una detección más robusta y precisa en términos de detección del robo de electricidad en redes las eléctricas automáticas.

Continuando con Mbey et al., (2024), indican la necesidad de un buen análisis de los grandes volúmenes de datos que se generan con medidores inteligentes para poder hacer frente al problema del fraude eléctrico, ya en redes inteligentes. La propuesta de los autores, es la de una técnica híbrida de inteligencia artificial que permite detectar consumidores fraudulentos a partir de cambios en patrones de consumo de forma abrupta. La metodología utilizada consiste en una máquina de vectores de soporte y una optimización de atrapamiento de partículas, con datos reales, utilizando indicadores de precisión, recuperación, f1 y AUC. Esta cantidad de resultados implica una tasa de detección del 98,9%, significando que el método propuesto es mejor que el anterior y efectivo para identificar el fraude energético.

Seguidamente, Saqib et al., (2024), presentan la problemática del suministro eléctrico en los países en desarrollo, donde no existe forma de tener Smartgrids y el robo de electricidad es algo a tener que solucionar y en este caso el objetivo era encontrar un modelo ligero de Deep Learning para detectar el fraude con sólo lecturas mensuales de consumo. La metodología utiliza ingeniería de características, así como métodos de reducción de dimensionalidad (PCA, t-SNE y UMAP) así como método de remuestreo como RUS, SMOTE y ROS, buscando aprender mejor la clase minoritaria.

La evaluación cuantitativa presenta una mejora significativa para la clase sobre el robo, con una precisión del 89 %, recuperación del 94 % y F1 del 91 % para el estudio de caso finalizando con la importancia de que el modelo propuesto es mejor que los existentes y que funciona bien en escenarios en los que no existen Smartgrids.

A partir de, Akintola et al., (2026), analizan las limitaciones que tienen los métodos tradicionales cualitativo para la detección del robo



Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 15, Año: 2026 (Julio 2026 - diciembre 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

de energía, que incluyen una baja consistencia en el resultado, así como un mayor riesgo de errores de clasificación. Se plantea un modelo SVM y este se evalúa frente a los métodos cualitativos. La metodología aplicada a cabo consiste en la comparación del modelo SVM y el análisis cualitativo a partir de las métricas estándar de desempeño establecido. El resultado cuantitativo demuestra que la SVM alcanza una precisión del 97,94 %, una precisión del 97,86 %, una recuperación del 99,93 % y un F1 del 98,88 % y concluye en el sentido de que proporciona una detección más precisa, confiable y escalable del robo de electricidad en las redes eléctricas del futuro.

En la misma línea, Khan et al., (2020) estudian el aumento del robo de electricidad, las principales limitaciones que les dificultan detectarlo (el tamaño del desbalance en los datos, el overfitting y las altas tasas de falsos positivos) y sugieren la posibilidad de desarrollar un modelo más exacto usando un aprendizaje automático supervisado que sea capaz de detectar el robo de electricidad a partir de datos reales del consumo. La propuesta metodológica consiste en la combinación del pre procesado de los datos, el oversampling según a ADASYN, la extracción de patrones mediante VGG-16 y la clasificación con FA-XGBoost, comparándolo con otros modelos como SVM, CNN o regresión logística. Los resultados cuantificables afianzan el alto rendimiento de la propuesta, que refleja que consigue F1 del 93,7 %, precisión del 92,6 % y recuperación del 97 %; concluyendo que la propuesta mejora la detección del robo del suministro eléctrico y puede resultar en una reducción, pero que a su vez puede ayudar considerablemente a disminuir las pérdidas económicas de las compañías eléctricas.

En la misma línea, Munawar et al., (2023) analizan el robo de la electricidad como un fenómeno menor significativo ya que origina grandes pérdidas económicas en las empresas de servicios públicos, debido a que la identificación de los consumidores que incurrieron en una actividad fraudulenta es complicada. La propuesta que se plantea consiste en diseñar un esquema robusto de detección que permita detectar eficazmente los comportamientos delictivos. La metodología apela al modelo AttenLSTMInception, a técnicas de inyección de datos o imitación de datos, balanceo con SMOTESVM bordelinde, y a la evaluación mediante Múltiples métricas y escenarios de validación.

Los resultados cuantitativos ponen de relieve la conclusión de que el modelo propuesto cuenta con una precisión y recuperación del 94 %, con una exactitud del 98 % y un AUC del 98 %. Las conclusiones derivadas hacen patente que el modelo propuesto es mejor que los modelos base como SVM, RF, DT y CNN - LSTM, y es la opción de detección de un robo de electricidad más confiable.


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

Por último, Bohani et al., (2021), conceptualizan el robo de electricidad como una de las causas de pérdidas no técnicas que perjudican la sostenibilidad económica de las empresas de electricidad, especialmente con conjuntos de datos muy desequilibrados. Su finalidad es la de comprobar el rendimiento de diferentes algoritmos de aprendizaje supervisados para determinar el modelo más conveniente para la detección de usuarios fraudulentos. A la metodología se le atribuyen Árbol de Decisión, ANN, DANN y AdaBoost utilizando datos reales de la State Grid Corporation of China y el cálculo de datos de exactitud, recuperación, F1-score y AUC. Los resultados indican que, si bien en términos de precisión el ANN es la red más precisa (92,54 %), el modelo DANN es el que presenta mejor rendimiento de forma global al ganar en retirada (40,94 %), F1-score (45,83 %) y AUC (0,69 %), concluyendo que los enfoques profundos poseen un mayor poder para detectar el robo de la electricidad en condiciones reales.

## Metodología

La metodología aplicada para llevar a cabo la investigación que aquí se describe ha estado determinada por un enfoque de aprendizaje automático supervisado para la detección de robo de la electricidad. Una vez en marcha el procedimiento metodológico, el mismo se articula en cinco etapas principales (A–E), que abarcan desde la etapa de adquisición y preparación de los datos, hasta la etapa de ajuste, evaluación y comparación de distintos modelos de clasificador. Todas las etapas fueron implementadas de forma secuencial, sistemática y reproducible, siguiendo según la representación de la Figura. 1.

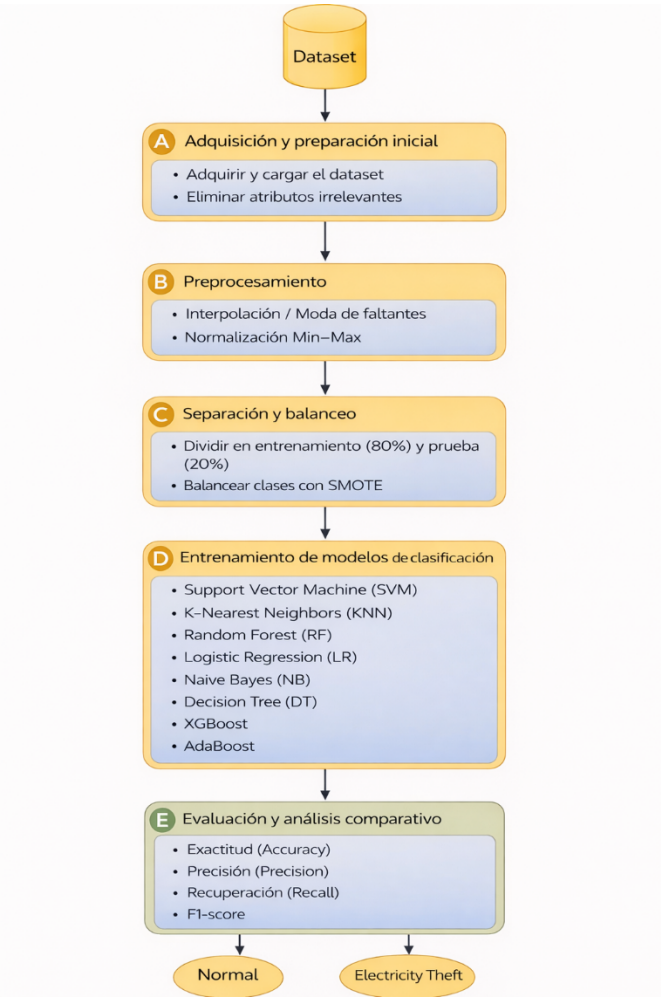



**Figura 1.** Esquema metodológico para la detección del robo de electricidad basado en el aprendizaje supervisado.

### *Adquisición y preparación inicial del conjunto de datos*

La primera fase fue aquella que incluyó tanto la adquisición como la carga de los datos a partir del



Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 15, Año: 2026 (Julio 2026 - diciembre 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

archivo en el que se realiza la identificación como conjunto de datos públicos, se encuentra disponible públicamente de Kaggle (2023) en formato CSV, recogidos en registros de consumo eléctrico obtenidos mediante contadores los eléctricos inteligentes, donde se originan principalmente de la State Grid Corporation of China (SGCC). El conjunto de datos contiene información de 42.372 clientes, de la cual 38.757 (91,5%) corresponden a consumidores con comportamiento no anómalo (no fraudulentos) y 3.615 (8,55%) a clientes asociados con el fraude el eléctrico (robar electricidad). La información que configura el conjunto de datos corresponde al consumo diario de energía eléctrica durante el período comprendido entre el 1 de enero del 2014 hasta el 31 de octubre del 2016, lo que conlleva unas 147 semanas de observaciones continuadas.

El conjunto de datos tiene una distribución desbalanceada comprendida por clases, lo que indica el comportamiento que se espera del análisis de una clase no técnica. Durante la preparación previa al análisis exploratorio del conjunto de datos, se eliminaron los atributos no interesantes del conjunto de datos (Consumer ID y CONS_NO). Su eliminación tiene como objetivo reducir el ruido del conjunto de datos y a la vez evitar el proceso del entrenamiento. A partir de la preparación previa, se obtuvieron estadísticas descriptivas y se analizó la distribución de clases. Esta descripción fue la necesaria para poder analizar la estructura del conjunto de datos.

### *Preprocesamiento y tratamiento de valores faltantes*

El conjunto de datos contiene registros con datos faltantes o inválidos a partir de los datos que se tienen de la información registrada por los medidores inteligentes, algo habitual en este tipo de datos por las causas habitualmente provocadas por los dispositivos medidores, el almacenamiento o cualquier otro problema de transmisión de datos. Esta circunstancia nos llevó sin duda a la elección de implementar una fase de preprocesamientos de los mismos para así tener unos datos de mejor calidad y más confiables para el entrenamiento de los modelos.

La forma de tratar los valores perdidos en la serie temporal haciendo referencia al consumo fue a partir de la interpolación, concretamente de la interpolación que sigue de forma natural la tendencia del consumo de electricidad. La serie temporal presente caso de valores atípicos en los cuales no se podía aplicar la técnica de interpolación, ya que eran los límites del total de la serie temporal (inicio o final de toda la serie) donde la acumulación no puede tener lugar o bien no disponemos de datos suficientes para la interpolación de forma que decidimos imputar de forma convencional los valores perdidos en cuanto a la media de la variable.


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

**ISSN**: 2665-0398

Finalmente, se aplicó el protocolo de normalización del tipo Min-Max a todas las variables predictoras escalando todos los valores al rango [0,1]. Con esta normalización, se homogeneiza el rango de los atributos y se mejora la estabilidad numérica y el rendimiento de los algoritmos de aprendizaje automático. Ello excluye a la variable objetivo (FLAG) de esta normalización. Una vez realizada, esta normalización fue ajustada en el conjunto de entrenamiento y posteriormente aplicamos dicha normalización en el conjunto de prueba para evitar la fuga de información (data fuga).

### *Separación de datos y balanceo de clases*

Cuando finalizó la etapa del preproceso, se hizo evidente que el conjunto de datos original se podía separar en uno que incluía un conjunto de variables predictivas (X) y el resto se convertía en uno que incluía un conjunto de variables de objetivo (Y). Hecho lo anterior, se procedió a dividir el conjunto de datos en entrenamiento, en un 80% y el de la prueba, en un 20%. Lo que se perseguía a era mantener una semilla de aleatoriedad fija; de este modo se buscaba asegurar la reproducibilidad de los resultados. Antes que este último, las separaciones de los conjuntos ya se habían hecho sin que se introdujera el uso de ningún tipo de técnica de balanceo de clases, ya que también se quería evitar introducir ningún tipo de sesión o fuga para el conjunto de prueba.

Y dado el considerable desequilibrio existente entre las clases se aplicó la técnica SMOTE (Synthetic Minority Over-Sampling Technique) solamente sobre el conjunto de entrenamiento. Esta técnica permite crear instancias sintéticas de la clase minoritaria con el fin de equilibrar la distribución de las clases, además de mejorar la capacidad de los modelos para poder aprender los patrones asociados al conjunto de datos de robo de electricidad.

### *Entrenamiento de modelos de aprendizaje automático*

En esta sección del trabajo se entrenaron los distintos modelos de aprendizaje automático supervisados, a fin de comparar su funcionamiento y evaluar el rendimiento en la detección del robo de electricidad. Los diferentes modelos tradicionales de aprendizaje automático supervisados que se implementaron quedan descritos en la Tabla 1.



**Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua**
Volumen: 7, Número: 15, Año: 2026 (Julio 2026 - diciembre 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

| Modelo | Descripción |
|---|---|
| AdaBoost | De acuerdo al autor Pineda Pertuz (2021), la adaptación de boosting será una técnica alternativa a los árboles de decisión individuales y a los bosques aleatorios que, normalmente, suele mostrar mejoras sobre la exactitud de los resultados. |
| Decision Tree | Los dendogramas estructurales que ilustran los árboles de decisión son modelos que sirven como soporte en los procesos de resolución de la toma de decisiones; Utiliza un proceso de ramificación por medio de condiciones para mostrar resultados.<br>Son una técnica simple y muy utilizada en el aprendizaje automático supervisado, específicamente para la clasificación de datos (Kashyap, 2018). |
| KNN | El algoritmo KNN, el método KNN, representa una forma de cierto aprendizaje de tipo supervisado que se puede aplicar para tareas de aprendizaje de tipo clasificación y para tareas de aprendizaje de tipo regresión, que está limitada por los datos que tiene a su alrededor. Sin embargo, la versión estándar de KNN no es escalable a grandes conjuntos de datos. El método Fast KNN mejora dicho rendimiento, introduciendo técnicas sobre la búsqueda de vecinos y sobre el coste computacional (Thanigaivelan et al., 2024). |
| Logistic Regression | La regresión logística se usa cuando la variable dependiente es de tipo binaria o categórica.<br>Lo que hace es estimar la probabilidad de que un evento en concreto ocurra.<br>Se usa también en la predicción de ventas, de resultados electorales, de eventos naturales (Dulhare et al., 2020) |
| Naive Bayes | Naive Bayes es un método de Machine Learning que aplica el teorema de Bayes para realizar predicciones. Este método asume que cada variable de entrada es independiente de las demás variables de entrada (Mehra et al., 2022). |
| Random Forest | Random Forest es un procedimiento de clasificación que se basa en la técnica de ensemble, el cual busca mejorar el rendimiento y la fiabilidad de los modelos tratando de evitar problemas de sobreajuste, que son característicos de los árboles de decisiones individuales (Zhang et al., 2022). |
| SVM | La Máquinas de vectores de soporte (SVM) se considera un clasificador generalizado lineal, el cual se utiliza para la clasificación binaria. Esta es uno de los algoritmos más eficientes para la clasificación binaria dentro del aprendizaje supervisado (Cortes & Vapnik, 1995). Resalta ser eficaz en el análisis de conjuntos de datos reducidos, de datos con patrones no lineales y en la clasificación de un conjunto con alta dimensionalidad (Pan et al., 2022). |
| XGBoost | Como entrada el algoritmo XGBoost, utilizamos un conjunto de entrenamiento, además de una función de pérdida diferenciable, un cierto número de modelos débiles y una tasa de aprendizaje, su utilización realiza una aplicación estructurada a través de diferentes etapas, de forma que permiten una mejora de la eficiencia y del rendimiento del modelo en tareas de aprendizaje (Xia & Sun, 2026). |

**Tabla 1**. Modelos de aprendizaje automático


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

Todos los modelos que se han utilizado fueron entrenados con el conjunto de balanceado ya mencionado, y los parámetros de los algoritmos se configuraron en función de la mayoría de ellos con las configuraciones por defecto, modificándose únicamente cuando esos parámetros lo requerían, es decir, a partir de la propia naturaleza del modelo. De ese modo, fue posible comparar el comportamiento básico de los algoritmos en igualdad de condiciones, sin llegar a realizar ningún tipo de proceso de optimización de hiperparámetros.

***Evaluación del desempeño y análisis comparativo***

La evaluación del rendimiento de los modelos fue realizada sobre un conjunto de prueba no balanceado, con la finalidad de emular un conjunto de condiciones reales de operación. Para cada uno de los modelos fueron generadas las métricas estándares de clasificación (exactitud (exactitud), así como matriz de confusión, precisión, recuperación y puntuación F1). Asimismo, para los modelos que tienen la capacidad para estimar probabilidades se obtuvo la curva ROC y se estimó el Área Bajo la Curva (AUC), con lo cual se analizó la capacidad de discriminación de cada uno de los modelos.

Este conjunto de métricas permitió al realizar comparativas objetivas del rendimiento de los modelos y la identificación del enfoque más idóneo para la detección del robo de electricidad.

**Resultados**

La Tabla 2 que se presenta a continuación, resume los resultados obtenidos para cada uno de los modelos analizados y las métricas de desempeño más relevantes Accuracy, F1-score, Precisión y Recall.

| Modelo | Accuracy | F1-score | Precisión | Recall |
|---|---|---|---|---|
| SVM | 83.91% | 38.11% | 28.89% | 56.00% |
| KNN | 77.65% | 31.13% | 21.40% | 57.07% |
| Random Forest | 86.19% | 40.61% | 32.79% | 53.33% |
| Logistic Regression | 80.92% | 35.40% | 25.27% | 59.07% |
| Naive Bayes | 82.38% | 31.36% | 23.93% | 45.47% |
| Decision Tree | 81.18% | 33.18% | 24.19% | 52.80% |
| XGBoost | 78.74% | 35.32% | 24.17% | 65.60% |
| AdaBoost | 73.94% | 33.04% | 21.38% | 72.67% |

**Tabla 2**. Resultados de los modelos

La Tabla 2 permite observar la forma de comparación entre los modelos de clasificación a partir del uso de las métricas: Accuracy, Precisión, Recall y F1 - score macro, permitiendo observar la mayor exactitud global (exactitud) y al mismo tiempo observar el comportamiento, para cada uno de los clasificadores, al clasificar las clases Normal o Robo de Electricidad. En este caso, el mejor comportamiento global observado es el modelo Random Forest, el cual presenta la máxima precisión en cuanto a los valores de la métrica



**Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua**
Volumen: 7, Número: 15, Año: 2026 (Julio 2026 - diciembre 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

Accuracy (86.19%) y su mayor F1-score (40.61%), es decir, con lo que se puede asegurar que el mejor comportamiento equilibrado del modelo en cuanto a la capacidad de detección o de clasificación es una función del rendimiento de los otros modelos evaluados.

Aunque el modelo demostró un estado de una alta precisión, un valor de la medida del F1-score intermedio expresa la existencia de un tipo de dificultades que pueden llegar a tener lugar debido al desequilibrio de clases que existieron como resultado de las mediciones de los modelos que pusimos en la práctica. Por otra parte, en lo que se refiere a los modelos SVM, Logistic Regression, y finalmente XGBoost, también son capaces de obtener valores de precisión superiores al 78% y valores intermedios de recuperación pero a su vez los resultados de precisión y F1-score son relativamente bajos, en donde se concibe tal situación que trata de que tendencia a diagnosticar como instancias positivas que son falsos o bien no identifican adecuadamente la clase minoritaria, o bien ambas cosas.

En cuanto a Naive Bayes y Decision Tree, tienen un comportamiento muy similar ya que en ambos casos la precisión supera el 80%, mientras que en lo que se refiere a la métrica F1-score que resulta en valores por debajo del 35% lo que expresa también una capacidad intermedia para poder clasificar adecuadamente de forma equilibrada entre ambas clases del problema. En último lugar, K-Nearest Neighbors y AdaBoost obtienen los menores niveles de mediciones de la métrica F1-score, pero obtienen mediciones relativamente altas de Recall, lo cual sugiere que estos modelos son más sensibles de identificar correctamente la clase positiva, pero esto significa a su vez que podemos tener una baja precisión con el impacto muy negativo que puede resultar sobre el rendimiento total de los modelos.

En fin, los resultados obtenidos son reflejos de la complejidad del problema dado y ponen de manifiesto la necesidad de aplicar otras técnicas para equilibrar y optimizar con el fin de que se pueda llegar a mejorar la capacidad de generalización que se aprecia en los modelos. Así mostramos las matrices de confusión de todos los modelos.

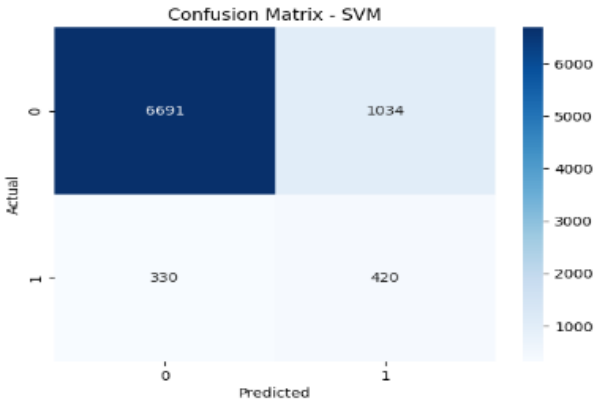



**Figura 2.** Matriz Confusión SVM


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

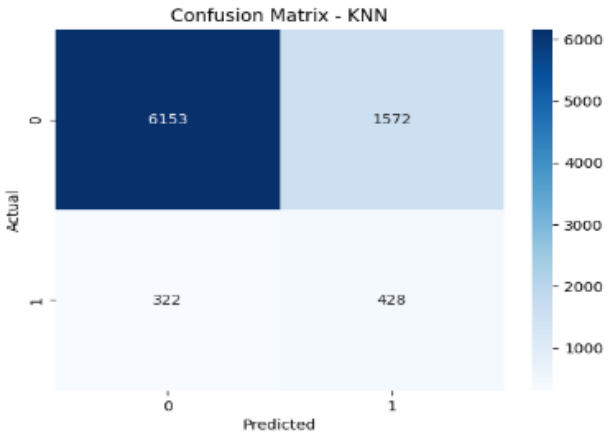



**Figura 3.** Matriz Confusión KNN

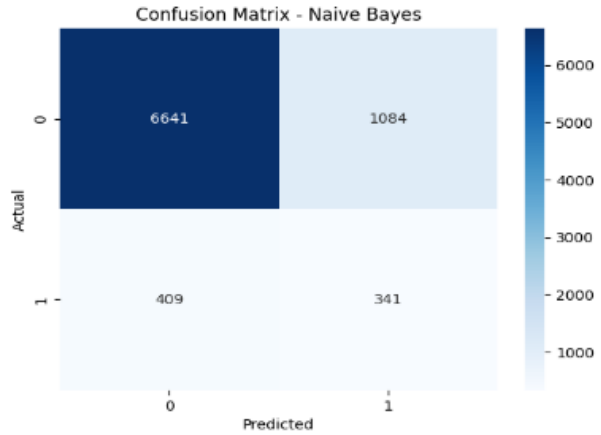



**Figura 6.** Matriz Confusión NB

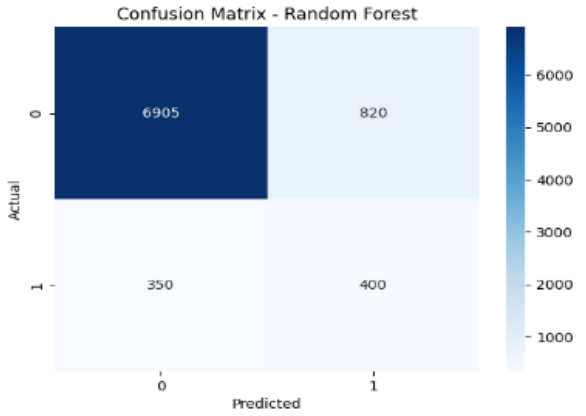



**Figura 4.** Matriz Confusión RF

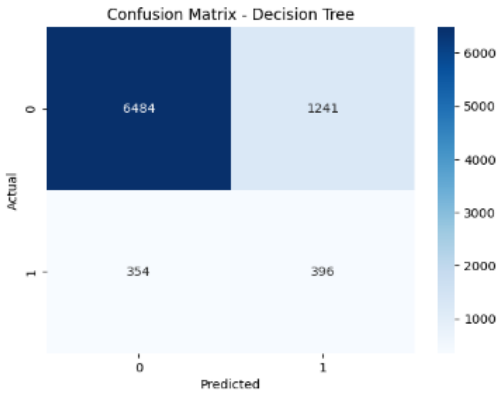



**Figura 7.** Matriz Confusión DT

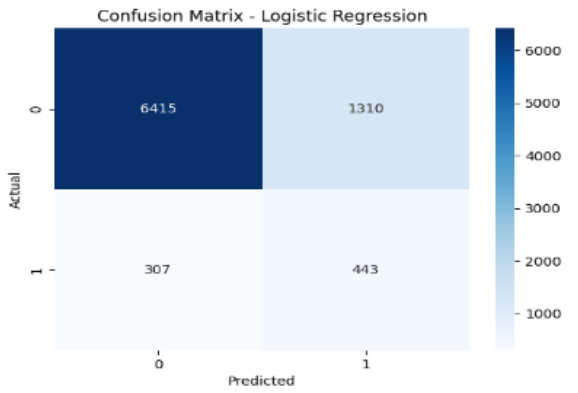



**Figura 5.** Matriz Confusión LR

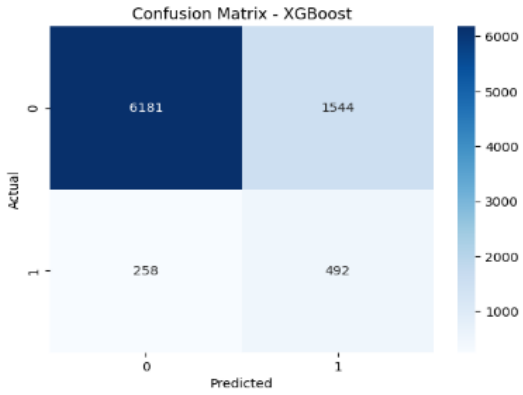



**Figura 8.** Matriz Confusión XGB



**Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua**
Volumen: 7, Número: 15, Año: 2026 (Julio 2026 - diciembre 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

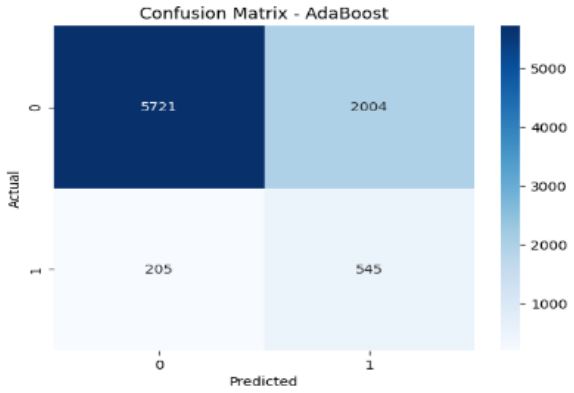



**Figura 9**. Matriz Confusión AB

Las matrices de confusión que pueden observarse brindan la posibilidad de conocer en profundidad la efectividad de los modelos de clasificación sin que las métricas globales mediaticen, además muestra cómo se distribuyen los verdaderos positivos (TP), los verdaderos negativos (TN), los falsos positivos (FP) y los falsos negativos (FN) de las clases Normal y Robo de Electricidad. En todos los modelos analizados se puede percibir un gran desbalanceo de las clases, el cual es altamente acentuado por la cantidad de instancias que pertenecen a la clase Normal.

Esta característica se deja ver en el gran número de verdaderos negativos que tuvieron las matrices de confusión de todos los modelos de clasificación analizados, razón por la cual los valores de precisión son relativamente altos, incluso en aquellos modelos que han evaluado con un bajo rendimiento global. El modelo Random Forest presenta un comportamiento equilibrado con una importante disminución de falsos positivos y falsos negativos y en comparación con el resto de los clasificadores. Su matriz muestra por un lado la buena capacidad para predecir de forma correcta los casos de Robo de Electricidad sin generar alarmas excesivas, por lo que no llama la atención que sea el modelo que presenta el mejor rendimiento global en las métricas globales.

El método de SVM y Logistic Regression poseen matrices con un mismo número de verdaderos positivos, aunque también muestran una cantidad alta de falsos positivos, o lo que es lo mismo: son modelos igualmente muy sensibles detectando el fraude, pero a costa de clasificar una cantidad alta de consumos normales como fraudulentos, tema que puede ser delicado para aplicaciones en el mundo real.

Naive Bayes y Decision Tree obtienen un comportamiento de doble filo, manteniendo un número alto de verdaderos negativos, aunque con elevadas cifras de falsos negativos, implicando una peor adecuación del modelo a la clase Electricity Theft, es decir, el modelo tiene una tendencia conservadora en la clasificación, a favor de la clase mayor.

En cambio, KNN obtiene un rendimiento negativo y un número constante de falsos positivos y falsos negativos; su matriz deja entrever otra cifra baja de capacidad de la separación entre clases y su


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

bajo F1 vuelve a darles un indicativo de problemas en la generalización del problema.

## Discusión

Los resultados de esto corroboran que, en escenarios reales, la detección del robo de electricidad sigue constituyendo un problema que es complicado principalmente por el elevado desequilibrio existente entre clases en el conjunto de datos de consumo. A pesar de que trabajos anteriores obtienen niveles muy elevados de exactitud con modelos profundos e híbridos (Ullah et al., 2025; Iftikhar et al., 2024), en general esos trabajos presentan niveles altos de exactitud en condiciones balanceadas y/o con altos costos computacionales de entrenamientos de modelos.

En contraposición, esta investigación presenta la evaluación de modelos clásicos de aprendizaje automático en un conjunto de datos no balanceado, lo cual se aleja de ser una evaluación teórica idealizada y se acerca, por lo contrario, a un contexto operativo real tal y como argumentan (Saqib et al., 2026; Bohani et al., 2021). Es necesario también en subrayar que, de los modelos analizados el Random Forest fue el que logró la mejor evaluación global, alcanzando la mayor exactitud y el F1-score macro mayor.

Esos resultados coinciden con antecedentes que dan cuenta de la robustez de este modelo frente a datos ruidosos y desequilibrados (Taruna et al., 2025; Bohani et al., 2021). A pesar de esto, los puntajes de F1-score y precisión son moderados y reflejantes que, a pesar de que la robustez de los modelos usados ayuda, llegan incluso a encontrar dificultades para clasificar de forma consistente la clase minoritaria, tal y como señalan (Jokar et al., 2016).

Igualmente, modelos como AdaBoost y KNN que tienen un alto recuerdo provocan el aumento de las falsas alarmas, lo cual limita su aplicabilidad, dado el incremento de los costos asociados a la inspección en caso de alertas erróneas. (Tsao et al., 2024; Özay & Çakıt, 2025). Esto da cuenta de la importancia de evaluar múltiples métricas, puesto que la exactitud total por sí sola no tiene ningún valor predictivo.

## Conclusiones

En el estudio se realizó un análisis comparativo de diferentes modelos de aprendizaje automático supervisado para detectar el robo de electricidad sobre datos reales de medidores inteligentes. Los resultados obtenidos muestran que, si bien los modelos analizados ofrecen resultados correctos en lo que respeta a la exactitud, su capacidad de la detección correcta de la clase minoritaria sigue siendo muy baja en sitios donde existe alta desproporción de clases.

Del estudio de los modelos, el modelo Random Forest es el que ofrece el mejor



Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 15, Año: 2026 (Julio 2026 - diciembre 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

compromiso entre el rendimiento global y la capacidad de detección, pero aún con un gran margen de mejora. De otro lado, se concluye que la métrica de precisión es muy simple para realizar una evaluación del sistema de detección de fraude eléctrico y teniendo en cuenta otras métricas como el recall y el F1 -score para obtener el rendimiento del modelo.

En resumidas cuentas, estos hallazgos evidencian la necesidad de seguir investigando de cara a lograr una detección de fraude eléctrico que no suponga un aumento demasiado grande de las falsas alarmas y que mantenga la implementación computacional y operativamente viable.

## Recomendaciones

Asimismo, la investigación debe de implementarse técnicas de ingeniería de características adicionales que permitan la captura de patrones más característicos de un comportamiento fraudulento: variabilidad, estacionalidad y comparación relativa entre usuarios, siguiendo lo que sugieren los estudios que abordan la detección de pérdidas no técnicas. Igualmente, deben evaluarse esquemas de aprendizaje híbridos que hagan una mezcla de modelos clásicos con otros más avanzados, y que consigan un equilibrio de desempeño y coste computacional.

Se recomienda, además explorar otras alternativas de tratamiento para el desequilibrio de clases, ya con técnicas de remuestreo más avanzadas como el SMOTE o bien con técnicas basadas en costos, para potenciar la detección de la clase minoritaria sin incrementar excesivamente los falsos positivos. En un plano operativo, se aconseja al priorizarlas que se recuerden el F1-score por encima de la exactitud global, sobre todo en escenarios bajo condiciones reales, y en último lugar, se aconseja validar los modelos en otras zonas geográficas o bien en otros períodos temporales para comprobar su capacidad de generalización y su robustez antes de ser utilizados en sistemas de detección de fraudes por robo de electricidad.

## Referencias

Ahmad, T. (2017). Non-technical loss analysis and prevention using smart meters. *Renewable and Sustainable Energy Reviews, 72*, 573–589. Documento en línea. Disponible https://doi.org/10.1016/j.rser.2017.01.100

Akintola, O., Adetokun, B., & Oghorada, O. (2026). Comparative analysis of energy theft detection in a power system using support vector machine and quantitative technique. *International Journal of Engineering Trends and Technology, 74*(1), 227–235. Documento en línea. Disponible https://doi.org/10.14445/22315381/IJETT-V74I1P117

Blazakis, K., Schetakis, N., Badr, M. M., Aghamalyan, D., Stavrakakis, K., & Stavrakakis, G. (2025). Power theft detection in smart grids using quantum machine learning. *IEEE Access, 13*, 61511–61525. Documento en línea. Disponible https://doi.org/10.1109/ACCESS.2025.3558143

Bohani, F. A., Suliman, A., Saripuddin, M., Sameon, S. S., Md Salleh, N. S., & Nazeri, S. (2021). A comprehensive analysis of supervised


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

**ISSN**: 2665-0398

learning techniques for electricity theft detection. *Journal of Electrical and Computer Engineering, 2021*, Article 9136206, 1–10. Documento en línea. Disponible https://doi.org/10.1155/2021/9136206

Chen, J., Nanehkaran, Y. A., Chen, W., Liu, Y., & Zhang, D. (2023). Data-driven intelligent method for detection of electricity theft. *International Journal of Electrical Power & Energy Systems, 148*, 108948. Documento en línea. Disponible https://doi.org/10.1016/j.ijepes.2023.108948

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297. Documento en línea. Disponible https://doi.org/10.1007/BF00994018

Dulhare, U. N., Ahmad, K., & Bin Ahmad, K. A. (Eds.). (2020). Machine learning and big data: Concepts, algorithms, tools and applications. *Wiley.* Documento en línea. Disponible https://doi.org/10.1002/9781119654834

Elshennawy, N. M., Ibrahim, D. M., & Gab Allah, A. M. (2025). An efficient electricity theft detection based on deep learning. *Scientific Reports, 15*, 12866. Documento en línea. Disponible https://doi.org/10.1038/s41598-025-93140-z

Emadaleslami, M., Haghifam, M.-R., & Zangiabadi, M. (2023). A two stage approach to electricity theft detection in AMI using deep learning. *International Journal of Electrical Power & Energy Systems, 150*, 109088. Documento en línea. Disponible https://doi.org/10.1016/j.ijepes.2023.109088

Ge, L., Li, J., Du, T., & Hou, L. (2025). Double-layer stacking optimization for electricity theft detection considering data incompleteness and intra-class imbalance. *International Journal of Electrical Power & Energy Systems, 165*, 110461. Documento en línea. Disponible https://doi.org/10.1016/j.ijepes.2025.110461

Ghaedi, H., Kamel Tabbakh, S. R., & Ghaemi, R. (2021). Improving electricity theft detection using combination of improved crow search algorithm and support vector machine. *Majlesi Journal of Electrical Engineering, 15*(4), 63–75. Documento en línea. Disponible https://doi.org/10.52547/mjee.15.4.63

Iftikhar, H., Khan, N., Raza, M. A., Abbas, G., Khan, M., Aoudia, M., Touti, E., & Emara, A. (2024). Electricity theft detection in smart grid using machine learning. *Frontiers in Energy Research, 12*, 1383090. Documento en línea. Disponible https://doi.org/10.3389/fenrg.2024.1383090

Iqbal, M. S., Munawar, S., Adnan, M., Raza, A., Akbar, M. A., & Bermak, A. (2025). A critical review of technical case studies for electricity theft detection in smart grids: A new paradigm based transformative approach. *Energy Conversion and Management: X, 26*, 100965. Documento en línea. Disponible https://doi.org/10.1016/j.ecmx.2025.100965

Jokar, P., Arianpoo, N., & Leung, V. C. M. (2016). Electricity theft detection in AMI using customers' consumption patterns. *IEEE Transactions on Smart Grid, 7*(1), 216–226. Documento en línea. Disponible https://doi.org/10.1109/TSG.2015.2425222

Kaggle. (2023). Detección de robo de electricidad de SGCC [Conjunto de datos]. *Kaggle.* Documento en línea. Disponible https://www.kaggle.com/datasets/bensalem14/sgcc-dataset

Kashyap, P. (2018). Machine learning for decision makers: Cognitive computing fundamentals for better decision making. *Apress.* Documento en línea. Disponible https://doi.org/10.1007/978-1-4842-2988-0

Khan, N., Shahid, Z., Alam, M. M., Sajak, A. A. B., Nazar, M., & Mazliham, M. S. (2024). A novel deep learning technique to detect electricity theft in smart grids using AlexNet. *IET Renewable Power Generation, 18*(6), 941–958. Documento en línea. Disponible https://doi.org/10.1049/rpg2.12846



**Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua**
Volumen: 7, Número: 15, Año: 2026 (Julio 2026 - diciembre 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

Khan, Z. A., Adil, M., Javaid, N., Saqib, M. N., Shafiq, M., & Choi, J.-G. (2020). Electricity theft detection using supervised learning techniques on smart meter data. *Sustainability, 12*(19), 8023. Documento en línea. Disponible https://doi.org/10.3390/su12198023

Liu, Z., Ding, W., Chen, T., Sun, M., Cai, H., & Liu, C. (2023). An electricity theft detection method through contrastive learning in smart grid. *EURASIP Journal on Wireless Communications and Networking, 2023*, 54. Documento en línea. Disponible https://doi.org/10.1186/s13638-023-02258-z

Mbey, C. F., Bikai, J., Yem Souhe, F. G., Foba Kakeu, V. J., & Boum, A. T. (2024). Electricity theft detection in a smart grid using hybrid deep learning-based data analysis technique. *Journal of Electrical and Computer Engineering, 2024*, Article 6225510. Documento en línea. Disponible https://doi.org/10.1155/2024/6225510

Mehra, R., Meesad, P., Peddoju, S. K., & Rai, D. S. (Eds.). (2022). Computational intelligence and smart communication. *Springer*. Documento en línea. Disponible https://doi.org/10.1007/978-3-031-22915-2

Munawar, S., Khan, Z. A., Chaudhary, N. I., Javaid, N., & Raja, M. A. Z. (2023). Machine intelligence aware electricity theft detection for smart metering applications. *Waves in Random and Complex Media*, 1–21. Documento en línea. Disponible https://doi.org/10.1080/17455030.2023.2239368

Özay, E. C., & Çakıt, E. (2025). Assessing the relationship between electricity theft and customer payment habits: A machine-learning approach. *Utilities Policy, 95*, 101955. Documento en línea. Disponible https://doi.org/10.1016/j.jup.2025.101955

Pamir, Javaid, N., Javed, M. U., Houran, M. A., Almasoud, A. M., & Imran, M. (2023). Electricity theft detection for energy optimization using deep learning models. *Energy Science & Engineering, 11*(10), 3575–3596. Documento en línea. Disponible https://doi.org/10.1002/ese3.1541

Pan, J.-S., Balas, V. E., & Chen, C.-M. (Eds.). (2022). *Advances in intelligent data analysis and applications*. *Springer.* Documento en línea. Disponible https://doi.org/10.1007/978-981-16-5036-9

Pineda Pertuz, C. M. (2021). *Aprendizaje automático y profundo en Python: Una mirada hacia la inteligencia artificial.* Ediciones de la U. Documento en línea. Disponible https://api.pageplace.de/preview/DT0400.9789587923155_A43032537/preview-9789587923155_A43032537.pdf

Saha, O. (2023). *A comparative study on optimized machine learning and deep learning models for the detection of electricity theft* [Master's thesis, National College of Ireland]. NORMA@NCI Library. Documento en línea. Disponible https://norma.ncirl.ie/7312/

Saqib, S. M., Mazhar, T., Iqbal, M., Shahazad, T., Almogren, A., Ouahada, K., & Hamam, H. (2024). Deep learning-based electricity theft prediction in non-smart grid environments. *Heliyon, 10*(15), e35167. Documento en línea. Disponible https://doi.org/10.1016/j.heliyon.2024.e35167

Saqib, S. M., Mazhar, T., Iqbal, M., Almogren, A., Ghadi, Y. Y., Shahzad, T., & Hamam, H. (2026). Utilizing machine learning ensembles for effective electricity theft detection. *Energy Exploration & Exploitation, 44*(1), 526–553. Documento en línea. Disponible https://doi.org/10.1177/01445987251381989

Savian, F. de S., Siluk, J. C. M., Garlet, T. B., do Nascimento, F. M., Pinheiro, J. R., & Vale, Z. (2021). Non-technical losses: A systematic contemporary article review. *Renewable and Sustainable Energy Reviews, 147*, 111205. Documento en línea. Disponible https://doi.org/10.1016/j.rser.2021.111205

Taruna, A. P., Arisona, G., Irwanto, D., Bestari, A. B., & Juniawan, W. (2025). Electricity theft


Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

**ISSN**: 2665-0398

detection using machine learning in traditional meter postpaid residential customers: A case study on State Electricity Company (PLN) Indonesia. *IEEE Access, 13*, 7167–7191. Documento en línea. Disponible https://doi.org/10.1109/ACCESS.2025.3526764

Thanigaivelan, R., Suchithra, M., & Kaliappan, S. (Eds.). (2024). Metaheuristic and machine learning optimization strategies for complex systems. *IGI Global*. Documento en línea. Disponible https://doi.org/10.4018/979-8-3693-7842-7

Tsao, Y.-C., Rahmalia, D., & Lu, J.-C. (2024). Machine-learning techniques for enhancing electricity theft detection considering transformer reliability and supply interruptions. *Energy Reports, 12*, 3048–3064. Documento en línea. Disponible https://doi.org/10.1016/j.egyr.2024.08.068

Ullah, A., Khan, I. U., Younas, M. Z., Ahmad, M., & Kryvinska, N. (2025). Robust resampling and stacked learning models for electricity theft detection in smart grid. *Energy Reports, 13*, 770–779. Documento en línea. Disponible https://doi.org/10.1016/j.egyr.2024.12.041

Xia, Y., & Sun, J. (2026). Machine learning for microbiome statistics. *Chapman and Hall/CRC*. Documento en línea. Disponible https://doi.org/10.1201/9781003610281

Yang, K., Chen, W., Bi, J., Wang, M., & Luo, F. (2023). Multi-view broad learning system for electricity theft detection. *Applied Energy, 352*, 121914. https://doi.org/10.1016/j.apenergy.2023.121914

Zhang, Y., Jiang, Y., Tsai, S.-B., & Khosravi, M. R. (Eds.). (2022). Mild cognitive impairment recognition via gene expression mining and neuroimaging techniques. *Frontiers Media SA*. Documento en línea. Disponible https://doi.org/10.3389/978-2-8325-0840-4

Zidi, S., Mihoub, A., Qaisar, S. M., Krichen, M., & Abu Al-Haija, Q. (2023). Theft detection dataset for benchmarking and machine learning based classification in a smart grid environment. *Journal of King Saud University – Computer and Information Sciences, 35*(1), 13–25. Documento en línea. Disponible https://doi.org/10.1016/j.jksuci.2022.05.007



**Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua**
Volumen: 7, Número: 15, Año: 2026 (Julio 2026 - diciembre 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional