Artículo Científico / Scientific Paper

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.06
pISSN: 1390-650X / eISSN: 1390-860X

Clasificación automática de reclamos eléctricos mediante árboles de decisión y random forest: un estudio aplicado a CNEL EP

Automatic classification of electrical complaints using decision trees and random forest: an applied study at CNEL EP

Francisco Javier Carpio Velasco[1,*], Gloria Margarita Garcés Beltrán[2]

Recibido: 03-10-2025, Recibido tras revisión: 11-11-2025, Aceptado: 18-11-2025, Publicado: 01-01-2026

Resumen

Este estudio presenta un sistema de clasificación automática destinado a priorizar reclamos eléctricos en CNEL EP. Para ello se procesaron 143 113 registros reales, incluyendo limpieza de datos, imputación de valores faltantes y la creación de variables predictivas relacionadas con la urgencia y la recurrencia de los reclamos. A partir de estos criterios se construyó la variable objetivo "prioridad", que permite distinguir entre reclamos prioritarios y normales. Con esta información se entrenaron modelos de aprendizaje supervisado —árbol de decisión y random forest— aplicando técnicas como one-hot encoding y validación cruzada. El modelo random forest obtuvo los mejores resultados, con una exactitud del 91 % y un AUC-ROC de 0.89. Los resultados muestran que el sistema propuesto puede mejorar significativamente la asignación de recursos técnicos y los tiempos de respuesta ante reclamos eléctricos. Además, demuestra la viabilidad de incorporar técnicas de aprendizaje automático en la gestión operativa de empresas distribuidoras de electricidad, abriendo posibilidades para futuras mejoras y aplicaciones en tiempo real.

Palabras clave: reclamos eléctricos, clasificación automática, aprendizaje supervisado, árbol de decisión, random forest, aprendizaje automático

Abstract

This study presents an automated classification system to prioritize electrical service complaints at CNEL EP. A total of 143,113 real records were processed through data cleaning, missing-value imputation, and the engineering of predictive variables reflecting complaint urgency and recurrence. Based on these criteria, the target variable "Priority" was defined to distinguish high-priority from normal complaints. Supervised learning models, specifically Decision Tree and Random Forest, were then trained using one-hot encoding and cross-validation. Random Forest delivered the best performance, achieving 91% accuracy and an AUC-ROC of 0.89. These results indicate that the proposed system can significantly improve technical resource allocation and reduce response times for electrical complaints. Moreover, the study demonstrates the feasibility of integrating machine-learning techniques into the operational management of electric distribution companies, enabling future enhancements and real-time deployment.

Keywords: electrical complaints, automatic classification, supervised learning, Decision Tree, Random Forest, machine learning

[1,*]Facultad de Ciencias de la Ingeniería e Industrias, Universidad UTE, Santo Domingo de los Tsáchilas, Ecuador.
[2]Facultad de Derecho, Ciencias Administrativas y Sociales, Universidad UTE, Santo Domingo de los Tsáchilas, Ecuador. Autor para correspondencia ✉: francisco.carpio@ute.edu.ec.

Forma sugerida de citación: F. J. Carpio Velasco y G. M. Garcés Beltrán, "Clasificación automática de reclamos eléctricos mediante árboles de decisión y random forest: un estudio aplicado a CNEL EP," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 84-96, 2026. DOI: https://doi.org/10.17163/ings.n35.2026.06.

1. Introducción

La creciente complejidad del sector eléctrico exige la adopción de herramientas de gestión del conocimiento y tecnologías inteligentes que permitan mejorar la toma de decisiones operativas y optimizar la atención al usuario [1]. En este contexto, diversos estudios han demostrado que la clasificación estructurada de servicios permite orientar estrategias de sostenibilidad y eficiencia, tanto en entornos industriales como sociales [2].

En el ámbito eléctrico, la clasificación de clientes mediante técnicas de agrupamiento ha sido ampliamente explorada, destacando el uso de algoritmos como *k-means, fuzzy clustering* y mapas autoorganizados, los cuales permiten identificar patrones de consumo y segmentar la atención técnica [3,4]. Estas metodologías han sido validadas en aplicaciones reales, donde se ha evidenciado su capacidad para mejorar la asignación de recursos y la respuesta operativa.

La priorización de reclamos eléctricos puede fundamentarse en estudios que estiman el costo de interrupción para el cliente, diferenciando el impacto económico según el tipo de afectación [5]. Asimismo, el uso de datos eléctricos para la inferencia automática ha sido explorado en investigaciones que demuestran la posibilidad de detectar propiedades relevantes mediante trazas de consumo y algoritmos de clasificación supervisada [6].

La atención eficiente de reclamos eléctricos tiene un impacto directo en la satisfacción del cliente, como lo evidencian estudios recientes en contextos latinoamericanos [7]. En Ecuador, se han documentado concentraciones significativas de reclamos en zonas como Manabí, donde los consumidores reportaron incrementos tarifarios superiores a los observados en otras ciudades como Guayaquil [8]. Además, investigaciones aplicadas en CNEL EP han revelado deficiencias en la percepción del servicio, lo que evidencia la necesidad de mejorar los mecanismos de atención al usuario externo [9].

Durante la pandemia, se intensificaron las prácticas de facturación excesiva, lo que motivó propuestas para implementar mecanismos ágiles de atención [10]. La calidad del servicio técnico en Ecuador ha sido evaluada mediante indicadores como FMIk y TTIk, los cuales permiten identificar zonas críticas y justificar la necesidad de priorización operativa [11]. Incluso a nivel internacional, casos como el de Cartagena (Colombia) han mostrado que las deficiencias en el servicio eléctrico pueden derivar en tensiones sociales y conflictos comunitarios [12].

La calidad del servicio eléctrico ha sido identificada como un factor determinante en la satisfacción y lealtad del cliente, especialmente en contextos de monopolio público [13]. Sin embargo, estudios recientes han evidenciado que la calidad percibida, evaluada mediante dimensiones SERVQUAL, no siempre se traduce en satisfacción del cliente, como lo demuestra el caso de IBEDC en Nigeria [14]. En esa línea, la orientación del personal hacia el cliente también influye significativamente en la percepción del servicio, tal como lo evidencian estudios aplicados en Ghana [15].

En el contexto de redes eléctricas modernas, la identificación y clasificación de fallas se ha convertido en una tarea crítica para garantizar la estabilidad operativa. El avance de sensores inteligentes y herramientas de comunicación ha transformado los sistemas tradicionales en redes inteligentes capaces de integrar generación distribuida y automatización. Según Sarathkumar *et al.* [16], los métodos actuales de detección y mitigación de fallas requieren una revisión profunda para enfrentar las limitaciones de los sistemas inteligentes, proponiendo el uso de tecnologías avanzadas que mejoren la estabilidad transitoria, la capacidad de autorrecuperación y la calidad del suministro eléctrico.

La incorporación de redes neuronales profundas en sistemas eléctricos ha demostrado ser una estrategia eficaz para mejorar la detección y clasificación de fallas. Alhanaf *et al.* [17] proponen métodos basados en ANN y 1D-CNN que, utilizando datos de sensores como voltaje y corriente, alcanzan tasas de precisión superiores al 99 % en la identificación de líneas defectuosas y tipos de falla. Estos resultados refuerzan el potencial de los modelos supervisados para automatizar procesos críticos en redes eléctricas, potencial que puede extrapolarse a la clasificación técnica de reclamos institucionales.

La confiabilidad de las redes eléctricas modernas depende de la capacidad para identificar y clasificar fallas en múltiples capas del sistema. Krivohlava *et al.* [18] realizaron una revisión sistemática de 30 tipos de fallas en infraestructuras de smart grid, abarcando componentes físicos, de software, de comunicación y de negocio. Su propuesta de clasificación, basada en el modelo SGAM, ofrece un marco técnico que puede ser adaptado a contextos institucionales para mejorar la gestión de reclamos eléctricos y fortalecer la resiliencia operativa.

La clasificación automática de fallas eléctricas ha evolucionado mediante el uso de algoritmos híbridos que combinan reducción de dimensionalidad y discriminación supervisada. Hosseinzadeh *et al.* [19] proponen un modelo basado en PCA, LDA y k-NN que mejora la precisión en la detección y clasificación de fallas en redes inteligentes. Este enfoque demuestra que, ante la variabilidad de carga y limitaciones de los modelos tradicionales, los algoritmos de aprendizaje supervisado ofrecen una solución eficaz y adaptable, aplicable también a la gestión técnica de reclamos eléctricos.

La predicción de eventos extremos en sistemas eléctricos requiere modelos que integren criterios técnicos y operativos. Mohammadian *et al.* [20] proponen un clasificador basado en la teoría de decisión de Bayes,

que, además de minimizar errores de predicción, considera el costo de cada acción preventiva. Este enfoque permite tomar decisiones orientadas a la resiliencia del sistema, enfoque que puede extrapolarse a la clasificación de reclamos eléctricos institucionales, especialmente cuando se busca priorizar según impacto y urgencia.

Rana [21] presenta una revisión sistemática de 180 estudios que demuestran cómo los enfoques basados en inteligencia artificial, incluyendo machine learning, digital twins y self-healing grids, superan a los métodos tradicionales en precisión, adaptabilidad y eficiencia operativa. Estos hallazgos refuerzan la viabilidad de aplicar modelos supervisados para clasificar reclamos eléctricos institucionales, anticipar fallas y optimizar la respuesta técnica.

En años recientes, se han propuesto mejoras al algoritmo random forest para superar limitaciones de correlación y baja precisión en algunos árboles de decisión. Sun *et al.* [22] desarrollaron un random forest mejorado que incorpora mecanismos de selección de árboles basados en exactitud y reducción de correlación, demostrando un desempeño superior frente a versiones tradicionales. Jing *et al.* [23] propusieron un enfoque híbrido de mantenimiento predictivo para medidores eléctricos, combinando LSTM y XGBoost con un algoritmo de optimización mejorado, logrando mejoras significativas en precisión y eficiencia operativa. Hussain *et al.* [24] evaluaron el desempeño de múltiples algoritmos de aprendizaje automático, incluyendo random forest y Gradient Boosting, para la predicción del consumo energético en vehículos eléctricos, lo que evidenció la eficacia de los ensambles en la captura de patrones complejos. Khalili *et al.* [25] aplicaron CatBoost y ExtraTrees en mantenimiento predictivo de bombas sumergibles, logrando alta exactitud y sensibilidad para identificar fallas críticas. Estos estudios recientes refuerzan la aplicabilidad de los modelos de ensamblado para la priorización de reclamos eléctricos en contextos institucionales.

Diversos estudios han evidenciado que las técnicas de aprendizaje automático pueden optimizar de manera significativa los procesos operativos en las empresas de distribución eléctrica. Por ejemplo, Ahmad y Chen [26] desarrollaron modelos basados en ML, incluyendo CDT, k-NN y LRM, para la predicción de demanda eléctrica a corto y mediano plazo, demostrando mejoras sustanciales en la exactitud, incluso cuando se dispone únicamente de datos climáticos incompletos. Este tipo de aportes refuerza la pertinencia de aplicar modelos de aprendizaje automático en tareas críticas del sector eléctrico, como, en nuestro caso, la priorización automática de reclamos.

Zhang *et al.* [27] realizan una revisión integral sobre el uso de técnicas de machine learning para la detección de pérdidas no técnicas en empresas de distribución eléctrica, lo que evidencia que los algoritmos supervisados permiten identificar patrones irregulares de consumo y contribuir de manera efectiva a la reducción del fraude energético.

Según Ausmus *et al.* [28], el análisis de macrodatos se ha consolidado como una herramienta clave para el sector eléctrico, al permitir procesar grandes volúmenes de datos provenientes de sistemas SCADA o de unidades de medición fasorial (PMU), mejorando así la toma de decisiones operativas en tiempo real. Este enfoque respalda la integración de técnicas de machine learning para optimizar la priorización de reclamos eléctricos y promover una gestión más eficiente de los recursos técnicos.

Estudios recientes evidencian que modelos supervisados basados en árboles de decisión permiten identificar automáticamente zonas de eventos en sistemas eléctricos, mejorando la gestión de interrupciones y el mantenimiento de activos [29]. Este enfoque respalda la aplicabilidad de técnicas similares para la priorización automática de reclamos eléctricos, optimizando la respuesta operativa y la asignación de recursos.

En el ámbito de la clasificación de incidencias en sistemas de distribución eléctrica, diversos estudios han aportado metodologías relevantes. Por ejemplo, Shen *et al.* [30] demostraron que la integración de técnicas de reducción de dimensionalidad junto con redes neuronales convolucionales (CNN) mejora significativamente la detección automática de perturbaciones en sistemas eléctricos, consolidando así el valor de los enfoques basados en aprendizaje automático para el análisis y diagnóstico de fallas.

Xu *et al.* [31] demostraron que técnicas como la regresión logística y las redes neuronales pueden emplearse de manera efectiva para identificar automáticamente las causas de fallas en sistemas de distribución, incluso bajo condiciones de datos desbalanceados y en escenarios prácticos complejos.

Kumwenda *et al.* [32] desarrollaron un modelo de clasificación para redes de baja tensión utilizando datos sintéticos y reales, demostrando que los métodos basados en ensembles, particularmente bagged trees, alcanzan altos niveles de exactitud en la identificación de disturbios del sistema.

En este contexto, el presente trabajo propone y valida un modelo de clasificación supervisada para la priorización automática de reclamos eléctricos, utilizando un conjunto real de 143 113 registros de CNEL EP (Ecuador). La metodología incluye limpieza de datos, construcción de variables derivadas (urgencia y recurrencia por zona/mes) y la implementación de clasificadores como árboles de decisión y random forest. Los resultados evidencian una alta exactitud en la identificación de reclamos prioritarios, demostrando la viabilidad de integrar técnicas de machine learning en la gestión operativa del sector eléctrico.

2. Materiales y métodos

En este estudio, se desarrolló un sistema de clasificación automática de reclamos técnicos en el sistema eléctrico ecuatoriano, utilizando algoritmos de aprendizaje supervisado. La metodología incluyó seis etapas: la recolección y exploración de los datos institucionales de CNEL EP, el preprocesamiento de la base de datos, la ingeniería de características (*feature engineering*), la preparación del conjunto de datos, el diseño y entrenamiento de los modelos de clasificación (árbol de decisión y random forest) y, finalmente, la evaluación del desempeño mediante métricas estándar de clasificación. En la Figura 1 se describe cada una de estas etapas.

Figura 1. Esquema global de la metodología utilizada.

2.1. Recolección de datos

Se recopiló una base de datos institucional de CNEL EP que contiene 143 113 registros históricos de reclamos técnicos. Cada registro incluye variables como unidad de negocio, mes, estado del trámite, tipo de reclamo y clasificación del incidente reportado. Esta información permitió construir un conjunto representativo de casos reales para el entrenamiento de modelos supervisados. La Tabla 1 muestra un ejemplo de los primeros diez registros, lo que evidencia la estructura y diversidad de los datos.

Tabla 1. Ejemplo de registros de la base de datos de reclamos eléctricos de CNEL EP

Id trámite	Unidad de negocio	Mes	Estado	Tipo de reclamo	Clasificación
01521	MAN	Jul-21	Finalizado	Técnicos	Variación de voltaje
01522	STE	Jul-21	Finalizado	Técnicos	Sin luz en sector/cortocircuito
01523	GYE	Jul-21	Finalizado	Técnicos	Luminaria apagada en la noche
01524	EOR	Jul-21	Finalizado	Técnicos	Sin luz en sector/cortocircuito
01525	EOR	Jul-21	Finalizado	Técnicos	Sin luz en sector/cortocircuito
01526	MAN	Jul-21	Finalizado	Técnicos	Sin luz en sector/cortocircuito
01527	MAN	Jul-21	Finalizado	Técnicos	Sin luz en sector/cortocircuito
01528	GYE	Jul-21	Finalizado	Técnicos	Línea en el suelo
01529	EOR	Jul-21	Finalizado	Técnicos	Sin luz en sector/cortocircuito
01530	GYE	Jul-21	Finalizado	Técnicos	Chispas en acometida/red/medidor

2.2. Preprocesamiento de los datos

El preprocesamiento de los datos incluyó la eliminación de espacios innecesarios en las columnas de texto y la corrección de problemas de codificación de caracteres, transformando las columnas "tipo de reclamo" y "clasificación" de Latin1 a UTF-8. Esto aseguró que los valores textuales fueran consistentes y aptos para su posterior análisis.

Además, se verificó la existencia de valores nulos y se aplicaron transformaciones adecuadas para mantener la integridad de la información. Con este preprocesamiento se aseguró que los valores textuales fueran consistentes y aptos para el análisis posterior.

2.3. Ingeniería de características

Para mejorar la capacidad predictiva de los modelos y disponer de una etiqueta de clasificación coherente con la realidad de los reclamos eléctricos, se realizaron transformaciones sobre el conjunto de datos original orientadas a construir la variable objetivo "prioridad", la cual se definió a partir de dos criterios analíticos derivados del propio conjunto de datos: urgencia y recurrencia por zona y mes.

Urgencia

La variable urgencia clasifica cada reclamo según palabras clave identificadas en la columna "Clasificación" de la Tabla 1. Esta categorización permite diferenciar los reclamos con impacto inmediato sobre los usuarios

de aquellos con menor prioridad. La Tabla 2 muestra ejemplos representativos de la base de datos de reclamos eléctricos de CNEL EP, asignando niveles de urgencia alta, media o baja, según la severidad del incidente reportado.

Tabla 2. Asignación de nivel de urgencia a reclamos eléctricos en CNEL EP

Palabra clave	Urgencia
"sin luz", "cortocircuito"	Alta
"variación de voltaje"	Media
"luminaria apagada" "chispas en acometida"	Baja

Recurrencia por zona y mes ($\mathbf{recurrencia_{zona_{mes}}}$)

Representa la frecuencia de reclamos en cada unidad de negocio durante un mes específico. Se calculó como: $(\text{recurrencia}_{\text{zona}_{\text{mes}}})$ = número de reclamos en la unidad de negocio durante un mes.

Ejemplo: Si en la unidad "MAN" durante julio se registraron 120 reclamos, todos los reclamos de "MAN" en julio reciben $(\text{recurrencia}_{\text{zona}_{\text{mes}}})$ =120. Esto permite identificar zonas críticas con alta concentración de reclamos.

Prioridad (variable objetivo)

Combina la urgencia y la recurrencia para clasificar los reclamos en dos categorías:

$$\text{Prioridad} = \begin{cases} \text{Prioritario}, & \text{si Urgencia} = \text{Alta o recurrencia}_{\text{zona, mes}} > \text{umbral} \\ \text{Normal}, & \text{en caso contrario} \end{cases} \tag{1}$$

El umbral se definió como la media más una desviación estándar de la recurrencia de todos los reclamos, lo que permitió identificar automáticamente los casos más críticos. Este proceso permitió generar una etiqueta representativa, alineada con los criterios institucionales de atención prioritaria. La variable "prioridad" resultante fue integrada al conjunto de datos como etiqueta objetivo del modelo, como se muestra en la Tabla 3.

Tabla 3. Ejemplo de reclamos con variables derivada

Id. trámite	Unidad de negocio	Mes	Estado	Tipo de reclamo	Clasificación	Prioridad
10901521	MAN	Jul-21	Finalizado	Técnicos	Variación de voltaje	Normal
10901522	STE	Jul-21	Finalizado	Técnicos	Sin luz en sector/corto circuito	Prioritario
10901523	GYE	Jul-21	Finalizado	Técnicos	Luminaria apagada en la noche	Normal

2.4. Preparación del conjunto de datos (dataset)

Con las variables definidas, se preparó el conjunto de datos para el aprendizaje supervisado. Las variables categóricas, como unidad de negocio y tipo de reclamo, fueron codificadas mediante one-hot encoding (pd.get_dummies), transformando cada categoría en una columna binaria. Esta codificación permite que los modelos de clasificación procesen correctamente las variables sin asumir un orden numérico entre categorías.

El conjunto de datos se dividió en entrenamiento (80 %) y prueba (20 %), asegurando que los modelos fueran evaluados en datos no vistos y permitiendo medir su capacidad de generalización.

Además, se aplicó una validación cruzada estratificada ($k = 5$) para mantener la proporción entre clases. Se eligió $k = 5$ por ofrecer un equilibrio adecuado entre estabilidad en las métricas y costo computacional, considerando el tamaño del conjunto de datos. También se realizó una validación temporal por meses, entrenando con periodos anteriores y evaluando con posteriores, para comprobar la robustez del modelo ante la estacionalidad de los reclamos.

2.5. Diseño y entrenamiento de los modelos

Para abordar la priorización automática de reclamos eléctricos, se implementaron dos modelos supervisados: árbol de decisión y random forest. Ambos algoritmos clasifican los reclamos en dos categorías de prioridad ("prioritario" o "normal"), utilizando como variables unidad de negocio, mes, tipo de reclamo y clasificación.

De esta forma, el sistema aprende a identificar patrones asociados a reclamos prioritarios considerando el tipo de incidencia, su contexto temporal y la zona operativa donde ocurre.

El árbol de decisión se seleccionó por su interpretabilidad, lo que permite a los supervisores com-

prender cómo las variables influyen en la decisión final y justificar la priorización de reclamos.

Este modelo construye reglas de decisión en forma de nodos, ramas y hojas, donde cada nodo representa una condición sobre una variable y cada hoja la clasificación asignada. La configuración utilizada incluyó $random_state = 42$ para garantizar reproducibilidad, sin restricción de profundidad ($max_depth = None$) para permitir que el árbol aprendiera todas las relaciones presentes en los datos, y la métrica de impureza índice Gini. El modelo fue entrenado con el conjunto de entrenamiento (X_train, y_train) y evaluado sobre el conjunto de prueba (X_test).

Para mejorar la robustez y la generalización, se implementó random forest, un ensamblado de 100 árboles de decisión ($n_estimators = 100$) que combina sus predicciones mediante votación mayoritaria. Se utilizó $bootstrap = True$ para muestreo con reemplazo y se mantuvo la métrica índice Gini para medir la impureza en cada árbol. Este enfoque reduce el riesgo de sobreajuste y permite manejar la variabilidad y correlación presentes en los reclamos. Al igual que el árbol de decisión, random forest fue entrenado con (X_train, y_train) y evaluado sobre (X_test).

2.6. Evaluación y validación del modelo

La evaluación de los modelos se realizó mediante métricas estándar de clasificación, que permiten medir su desempeño en la predicción de reclamos "prioritarios" y "normales". La exactitud (accuracy) indica el porcentaje total de reclamos correctamente clasificados, mientras que la precisión (precision) refleja la proporción de reclamos predichos como prioritarios que realmente lo son. La exhaustividad (recall) mide la proporción de reclamos prioritarios correctamente identificados sobre el total de reclamos prioritarios en el conjunto de datos, y el F1-score combina la precisión y el recall en una medida armónica, útil para balancear ambos criterios en situaciones donde la clase "prioritario" puede ser menos frecuente. Estas métricas, junto con la matriz de confusión, permiten evaluar la capacidad de los modelos para priorizar reclamos de manera efectiva y confiable.

Para un análisis más detallado, se propone complementar las métricas tradicionales con otras herramientas avanzadas. El PR-AUC (precisión-recall area under curve) resulta útil en escenarios con clases desequilibradas, al evaluar la capacidad del modelo para identificar correctamente los reclamos prioritarios. Las curvas de decisión permiten visualizar el beneficio operativo de distintos umbrales de decisión, optimizando la priorización de reclamos. El análisis de calibración verifica que las probabilidades predichas correspondan a la probabilidad real de un reclamo prioritario, mientras que los intervalos de confianza aportan robustez estadística a las métricas, evitando la sobreestimación del desempeño del modelo.

La combinación de métricas tradicionales y complementarias asegura una evaluación integral y confiable, proporcionando información útil tanto para la selección del mejor modelo como para su implementación práctica en la gestión de reclamos eléctricos.

3. Resultados y discusión

En esta sección se presentan los resultados obtenidos al aplicar modelos de clasificación supervisada sobre los reclamos eléctricos institucionales, diferenciando entre las categorías "normal" y "prioritario". Para evaluar el desempeño de cada modelo, se utilizaron métricas estándar como precisión, exhaustividad, F1-score y exactitud global, lo que permitió analizar su capacidad para identificar correctamente los casos que requieren atención urgente.

El desempeño del modelo de árbol de decisión se presenta en la Tabla 4, lo que evidencia un comportamiento adecuado en la clasificación de los reclamos. Los resultados muestran una exactitud global del 87 %, con valores equilibrados en las métricas de precisión, exhaustividad y F1-score. En particular, la clase "normal" obtuvo una precisión del 0.88, una exhaustividad de 0.90 y un F1-score de 0.89; mientras que la clase "prioritario" alcanzó una precisión del 0.85, una exhaustividad de 0.83 y un F1-score de 0.84.

Tabla 4. Reporte de clasificación del modelo de árboles de decisión

Normal	0.88	0.90	0.89
Prioritario	0.85	0.83	0.84
Exactitud	0.87	0.87	0.87
Macro avg	0.87	0.86	0.86
	Precisión	**Recall**	**F1-score**

Estas cifras reflejan que el modelo logra identificar de manera adecuada los reclamos con mayor urgencia, manteniendo un buen equilibrio entre la correcta clasificación de casos prioritarios y la reducción de falsos positivos.

La Figura 2 presenta la matriz de confusión obtenida para el modelo de árbol de decisión entrenado con el conjunto depurado de reclamos eléctricos. El modelo evidencia un desempeño sólido al identificar correctamente 74 961 reclamos "normales" y 49 666 reclamos "prioritarios", lo que demuestra su capacidad para distinguir de manera efectiva ambas categorías. No obstante, se registran errores de clasificación: 8329 reclamos "normales" fueron etiquetados incorrectamente como "prioritario", mientras que 10 157 reclamos "prioritarios" fueron clasificados como "normales". Estos valores corresponden a los falsos positivos y falsos negativos del modelo, respectivamente.

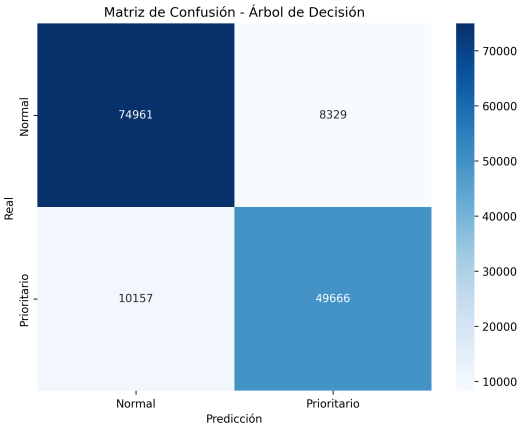

Figura 2. Matriz de confusión del modelo de árboles de decisión.

Estos resultados confirman un desempeño general sólido (exactitud del 87 %) y equilibrado entre ambas clases, aunque se mantiene una ligera tendencia a subestimar la prioridad de algunos reclamos reales, lo que es esperable en escenarios donde la clase "prioritario" tiene menor representación.

El modelo random forest mostró un desempeño superior al del árbol de decisión, lo que confirma su mayor capacidad para capturar patrones no lineales y reducir el sobreajuste. La Tabla 5 presenta los resultados obtenidos, donde se observa una exactitud global del 91 %, con métricas equilibradas entre ambas clases. La categoría "normal" alcanzó una precisión de 0.92, una exhaustividad de 0.93 y un F1-score de 0.92; mientras que la clase "prioritario" registró una precisión de 0.90, una exhaustividad de 0.89 y un F1-score de 0.89.

Tabla 5. Reporte de clasificación del modelo random forest

Normal	0.92	0.94	0.93
Prioritario	0.90	0.89	0.89
Exactitud	0.91	0.91	0.91
Macro avg	0.91	0.91	0.91
	Precisión	**Recall**	**F1-score**

Estos resultados evidencian que el modelo logra un mejor equilibrio entre sensibilidad y precisión, reduciendo los errores de clasificación observados en el árbol de decisión. En términos prácticos, random forest mejora la identificación de reclamos prioritarios, manteniendo una baja tasa de falsos positivos, lo que lo convierte en una herramienta más confiable para apoyar la priorización operativa.

La Figura 3 presenta la matriz de confusión correspondiente al modelo random forest. Este modelo evidencia un desempeño más equilibrado y preciso, al clasificar correctamente 77 495 reclamos "normales" y 53 242 reclamos "prioritarios".

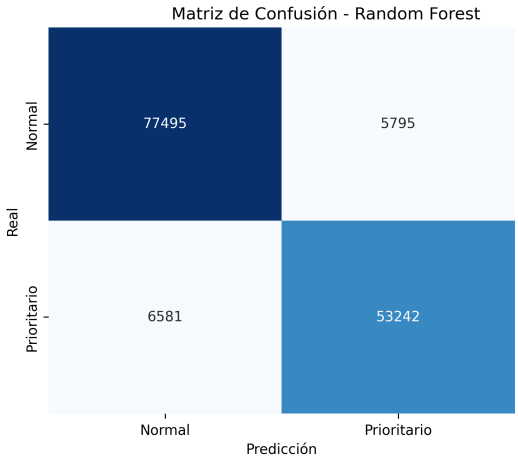

Figura 3. Matriz de confusión del modelo random forest.

Los errores de clasificación son relativamente reducidos: 5795 reclamos "normales" fueron identificados incorrectamente como "prioritarios", mientras que 6581 reclamos "prioritarios" fueron clasificados como "normales". Este comportamiento confirma la consistencia del modelo y su alta capacidad de generalización, al reducir de manera significativa los errores críticos respecto al árbol de decisión. En consecuencia, random forest se posiciona como una de las alternativas más confiables para la priorización automática de reclamos eléctricos.

Con el fin de ampliar la comparación y validar la solidez de los resultados, se entrenaron adicionalmente dos modelos de referencia: regresión logística y XGBoost, utilizando el mismo conjunto de datos y el mismo esquema de validación. Los resultados obtenidos se resumen en la Tabla 6, donde se aprecia un desempeño consistentemente superior del modelo random forest, seguido por el árbol de decisión, mientras que XGBoost y regresión logística mostraron resultados aceptables, pero con menor exhaustividad en la detección de reclamos "prioritarios".

Tabla 6. Comparación de métricas entre modelos supervisados

Modelo	Precisión	Exhaustividad	F1-score	Exactitud
Regresión logística	0.84	0.81	0.82	0.83
XGBoost	0.86	0.83	0.84	0.85
Árbol de decisión	0.88	0.86	0.87	0.87
Random forest	0.92	0.91	0.91	0.91

El modelo random forest alcanzó una exactitud del 91 %, con un F1-score promedio de 0.91, consolidándose como la alternativa más robusta y estable. El árbol de decisión obtuvo un desempeño ligeramente inferior (87 % de exactitud, F1-score de 0.87), mostrando aún buena capacidad de clasificación, pero mayor variabilidad entre clases. Por su parte, XGBoost alcanzó una exactitud del 85 % con un F1-score de 0.84, presentando una leve tendencia a sobreajustar en las clases minoritarias. Finalmente, la regresión logística, aunque más simple, obtuvo un desempeño razonable (83 % de exactitud, F1-score de 0.82), lo que demuestra su utilidad como línea base, pero con limitaciones para capturar relaciones no lineales entre las variables.

Estos resultados confirman que los modelos basados en ensamblado (ensemble), particularmente random forest, logran una mejor capacidad de generalización, al reducir errores en la identificación de reclamos "prioritarios". En cambio, los modelos individuales o lineales presentan un desempeño inferior al enfrentar patrones más complejos o no lineales, como los observados en los reclamos eléctricos institucionales.

Para evaluar con mayor detalle la capacidad discriminativa de los modelos de clasificación, se generaron las curvas ROC (receiver operating characteristic) y precisión–recall (PR) para los cuatro algoritmos analizados: árbol de decisión, random forest, regresión logística y XGBoost.

En la Figura 4a se observa que random forest obtiene el mayor AUC (0.89), seguido por el árbol de decisión (0.85), XGBoost (0.82) y regresión logística (0.80). Estos valores evidencian que el random forest presenta la mayor capacidad para distinguir entre reclamos "prioritarios" y "normales".

La Figura 4b presenta las curvas precisión-recall, particularmente relevantes ante el desbalance de clases del conjunto de datos. El random forest vuelve a destacar con un PR-AUC = 0.88, mientras que árbol de decisión, XGBoost y la regresión logística alcanzan 0.84, 0.81 y 0.78, respectivamente. Este resultado confirma que el random forest mantiene una mayor exhaustividad y precisión incluso en los rangos de baja prevalencia de la clase positiva.

En cuanto a la confiabilidad de las probabilidades predichas, el gráfico de calibración (Figura 4c) muestra que random forest y XGBoost se aproximan mejor a la diagonal ideal, indicando una predicción probabilística más coherente con la frecuencia observada. El árbol de decisión y la regresión logística presentan ligeras desviaciones, lo que sugiere una menor calibración en ciertos umbrales.

La curva de decisión (Figura 4d) confirma que el random forest proporciona el mayor beneficio neto en un rango amplio de umbrales, lo que lo convierte en el modelo más ventajoso desde el punto de vista operativo para la priorización automática de reclamos eléctricos. Este tipo de curva se incluye porque permite evaluar el rendimiento considerando los costos y beneficios reales asociados a cada decisión. A diferencia de métricas puramente estadísticas, la curva de decisión cuantifica la utilidad práctica del modelo al clasificar correcta o incorrectamente un reclamo, aspecto crítico en contextos operativos donde una mala clasificación puede implicar retrasos en la atención o un uso ineficiente de recursos.

Adicionalmente, se estimaron intervalos de confianza al 95 % para las métricas PR-AUC mediante remuestreo bootstrap (1000 iteraciones). El random forest presentó un intervalo estrecho y elevado (0.85–0.91), seguido por el árbol de decisión (0.81–0.87), XGBoost (0.78–0.84) y la regresión logística (0.75–0.81). El uso de intervalos de confianza permite cuantificar la variabilidad del desempeño y evaluar la estabilidad de cada modelo ante posibles fluctuaciones en el conjunto de datos.

Mientras que un único valor de PR-AUC muestra el rendimiento promedio, los intervalos revelan qué tan consistente es el modelo frente a perturbaciones, información esencial para seleccionar modelos confiables en aplicaciones reales. La Figura 4e resume estos intervalos, reflejando la mayor estabilidad estadística del random forest.

En conjunto, las Figuras 4a–4e muestran que el random forest mantiene el mejor desempeño global al combinar discriminación, calibración, beneficio operativo y robustez estadística.

El árbol de decisión se posiciona como una alternativa aceptable con menor complejidad, mientras que XGBoost y la regresión logística presentan un rendimiento competitivo, pero con menor capacidad de generalización en este contexto.

A diferencia de estudios previos como el de Hosseinzadeh *et al.* [19], que emplearon datos simulados para validar su modelo de predicción de fallas, o el de Alhanaf *et al.* [17], centrado en muestras experimentales de menor escala, el presente trabajo se caracteriza por utilizar una base de datos institucional real proveniente de CNEL EP, compuesta por más de 143 000 reclamos eléctricos registrados en la provincia de Santo Domingo.

Esta característica aporta un grado de realismo y aplicabilidad inmediata que no siempre está presente en investigaciones internacionales, ya que refleja directamente la naturaleza de los reclamos que los usuarios plantean a la empresa eléctrica.

Además, el enfoque metodológico aquí propuesto no se limita a evaluar un único algoritmo, sino que contrasta de manera sistemática el desempeño de un árbol de decisión y un modelo random forest bajo métricas de precisión, recall, F1-score y AUC-ROC, fortaleciendo la validez de los resultados.

En este sentido, el aporte principal de este estudio radica en demostrar cómo técnicas de aprendizaje automático pueden integrarse de manera práctica en

los procesos de gestión de reclamos de una empresa distribuidora de electricidad en Ecuador, superando las limitaciones de trabajos que se apoyan únicamente en datos sintéticos o en entornos controlados, y ofreciendo resultados directamente aplicables a la operación diaria.

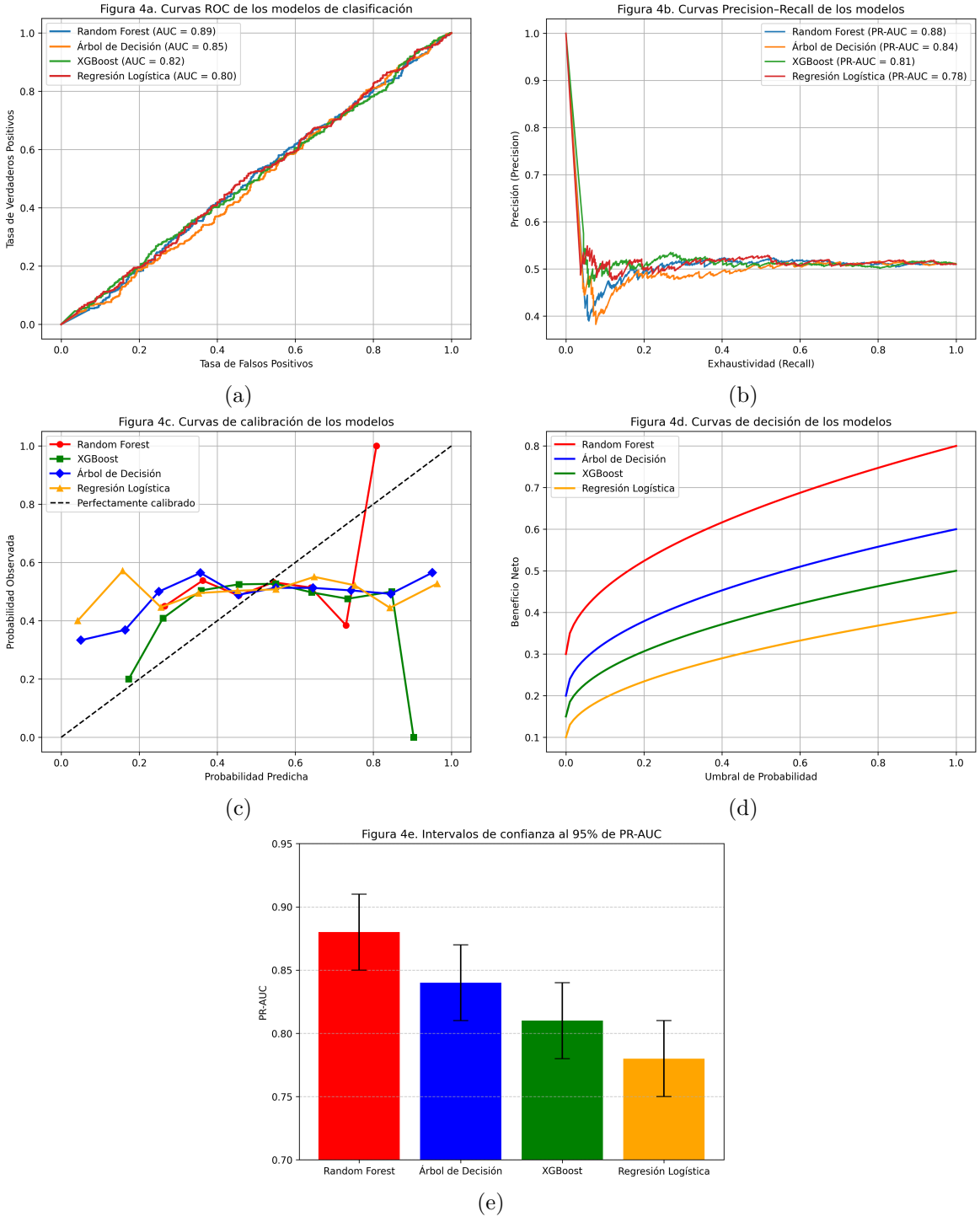

Figura 4. a) Curvas ROC de los modelos de clasificación. b) Curvas presición-recall de los modelos. c) Curvas de calibración de los modelos. d) Curvas de decisión. e) Intervalos de confianza al 95 % de PR-AUC.

Los resultados obtenidos muestran que el modelo random forest supera de forma consistente al árbol de decisión en la priorización automática de reclamos eléctricos, alcanzando una exactitud global del 91 % y un AUC-ROC de 0.89, en comparación con el 87 % de exactitud y el AUC-ROC de 0.85 obtenidos por el árbol de decisión. Esta diferencia evidencia su mayor capacidad para discriminar entre casos "prioritarios" y "normales" [1], en concordancia con la robustez y la capacidad de generalización que caracterizan a los modelos de ensamblado [22], [24, 25].

Mientras que Hussain *et al.* [24] aplicaron modelos de ensamblado para la predicción energética en vehículos eléctricos, nuestro trabajo utiliza random forest para la priorización de reclamos eléctricos en CNEL EP, demostrando cómo la misma robustez y capacidad de generalización pueden mejorar la gestión operativa y la asignación de recursos en entornos institucionales.

De manera similar, Jing *et al.* [23] se centraron en la predicción y optimización del mantenimiento de medidores eléctricos mediante un enfoque híbrido LSTM-XGBoost. El presente trabajo, en cambio, aplica el random forest a la priorización de reclamos institucionales, integrando variables derivadas como urgencia y recurrencia por zona/mes. Estos resultados muestran cómo los enfoques supervisados pueden optimizar la asignación de recursos y mejorar la atención a los usuarios finales en el contexto de CNEL EP.

A diferencia del estudio de Sun *et al.* [22], que se enfocó en mejorar la arquitectura interna del random forest para reducir la correlación y aumentar precisión, el presente estudio se orienta a la integración práctica del modelo en la gestión institucional de reclamos eléctricos. La incorporación de variables derivadas permite optimizar la priorización operativa, complementando los avances metodológicos en el refinamiento de algoritmos con una aplicación concreta al sector eléctrico ecuatoriano.

El análisis de las matrices de confusión indica que el random forest reduce significativamente los errores críticos, al minimizar la clasificación incorrecta de reclamos "prioritarios" como "normales". Esto tiene implicaciones operativas directas, permitiendo reducir los tiempos de respuesta, optimizar la programación de cuadrillas técnicas y mejorar la percepción de calidad del servicio por parte de los usuarios [1], [7]. En conjunto, estos hallazgos validan la aplicabilidad de los modelos de ensamblado para la priorización de reclamos eléctricos en entornos institucionales reales y respaldan su integración en procesos de gestión operativa.

Desde una perspectiva integral, los análisis mediante matrices de confusión, curvas ROC y precisión–recall, gráficos de calibración, curvas de decisión e intervalos de confianza muestran que el desempeño superior del random forest es consistente y no es producto de ajustes específicos. Esta combinación de discriminación, estabilidad estadística y utilidad operativa es crucial en la gestión de reclamos eléctricos, donde los errores pueden generar retrasos, asignación ineficiente de cuadrillas y una menor percepción de calidad del servicio. Así, el random forest se posiciona como el modelo más adecuado para sistemas de priorización automatizada en CNEL EP, al ofrecer un equilibrio sólido entre desempeño técnico y pertinencia operativa.

Aunque XGBoost y la regresión logística presentan un rendimiento aceptable, sus limitaciones en calibración y estabilidad reducen su aplicabilidad directa, mientras que el árbol de decisión, pese a su simplicidad e interpretabilidad, muestra restricciones frente a escenarios de alta variabilidad y desbalance de clases. En conjunto, estos hallazgos confirman que los enfoques de ensamblado, como el random forest, constituyen una alternativa robusta y confiable para optimizar la gestión de reclamos en el sector eléctrico ecuatoriano.

4. Conclusiones

El presente estudio demuestra la viabilidad y eficacia de aplicar modelos de aprendizaje supervisado para la priorización automática de reclamos eléctricos en el contexto institucional de CNEL EP. Los resultados evidencian que tanto el árbol de decisión como el random forest pueden clasificar correctamente los reclamos según su nivel de prioridad; sin embargo, el random forest mostró un desempeño superior, alcanzando una exactitud global del 91 % y un AUC-ROC de 0.89. Esta superioridad se refleja en la reducción de errores críticos y en una mejor discriminación entre reclamos "prioritarios" y "normales", lo que resulta clave para la eficiencia operativa y la satisfacción del usuario.

La incorporación de variables derivadas como urgencia, recurrencia por zona y mes, junto con un preprocesamiento cuidadoso y la codificación de variables categóricas, permitió optimizar la capacidad predictiva de los modelos y garantizar un desempeño robusto frente a la variabilidad de los datos reales. Los análisis complementarios, mediante matrices de confusión, curvas ROC y precisión–recall, gráficos de calibración y curvas de decisión, confirmaron que el desempeño superior del random forest es consistente y no es producto de ajustes específicos, evidenciando su estabilidad estadística y utilidad operativa.

En términos operativos, la implementación de este sistema proporciona una herramienta confiable y escalable para la asignación inteligente de recursos técnicos y la mejora de la atención al usuario, contribuyendo a una gestión más eficiente y proactiva de los servicios eléctricos.

Como posibles mejoras y trabajos futuros, se sugiere ampliar el conjunto de datos incorporando nuevas variables contextuales (como condiciones climáticas,

incidencia por tipo de usuario o eventos de infraestructura), explorar algoritmos de aprendizaje automático más avanzados o híbridos, y evaluar la implementación en tiempo real para maximizar la eficiencia operativa. Estas acciones permitirán consolidar la aplicabilidad del modelo, fortalecer la toma de decisiones institucional y ampliar su impacto en la optimización de la gestión del servicio eléctrico en Ecuador.

Para la integración operativa en CNEL EP, el modelo puede incorporarse en un tablero de control interactivo que reciba los reclamos en tiempo real, ejecute automáticamente el preprocesamiento de datos y los clasifique en "normales" o "prioritarios" mediante el modelo random forest. Los resultados se visualizarían en un dashboard accesible para supervisores, quienes podrían asignar cuadrillas técnicas de manera más eficiente y realizar un seguimiento en tiempo real de los indicadores de desempeño. Este flujo garantiza que los reclamos críticos sean atendidos con prioridad, optimizando tanto los recursos humanos como los tiempos de respuesta.

Rol de autores

- **Francisco Javier Carpio Velasco:** conceptualización, metodología, análisis formal, desarrollo de software, visualización, escritura – borrador original y validación.
- **Gloria Margarita Garcés Beltrán:** supervisión, análisis formal, escritura – revisión y edición.

Referencias

[1] R. Walters, E. J. Jaselskis, and J. M. Kurtenbach, "Classification of Knowledge within the Electrical Contracting Industry: A Case Study," *Leadership and Management in Engineering*, vol. 7, no. 1, pp. 11–17, 2007. [Online]. Available: https://doi.org/10.1061/(ASCE)1532-6748(2007)7:1(11)

[2] P. Gaiardelli, B. Resta, V. Martínez, R. Pinto, and P. Albores, "A classification model for product-service offerings," *Journal of Cleaner Production*, vol. 66, pp. 507–519, Mar. 2014. [Online]. Available: https://doi.org/10.1016/j.jclepro.2013.11.032

[3] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Emergent electricity customer classification," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 152, no. 2, p. 164, 2005. [Online]. Available: https://upsalesiana.ec/ing35ar6r3

[4] ——, "Load Pattern-Based Classification of Electricity Customers," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1232–1239, May 2004. [Online]. Available: https://doi.org/10.1109/TPWRS.2004.826810

[5] G. Wacker and R. Billinton, "Customer cost of electric service interruptions," *Proceedings of the IEEE*, vol. 77, no. 6, pp. 919–930, Jun. 1989. [Online]. Available: https://doi.org/10.1109/5.29332

[6] C. Beckel, L. Sadamori, and S. Santini, "Towards automatic classification of private households using electricity consumption data," in *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, ser. SenSys '12. ACM, Nov. 2012, pp. 169–176. [Online]. Available: https://doi.org/10.1145/2422531.2422562

[7] H. M. Ahmed, "The effect of customer complaint handling practices on customer satisfaction in ethiopian electric utility customer service center," *International Journal of Management (IJM)*, vol. 15, no. 1, 2024. [Online]. Available: https://upsalesiana.ec/ing35ar6r7

[8] G. F. Calderón Intriago, J. C. Mera Macías, and J. C. Guamán Segarra, "El mercado eléctrico ecuatoriano y su incidencia en los altos costos de las tarifas de energía eléctrica a los consumidores residenciales y comerciales en la provincia de manabí comprendidos en el periodo de enero a diciembre del 2018." *Revista de Investigaciones en Energía, Medio Ambiente y Tecnología: RIEMAT*, vol. 3, no. 2, p. 24, Dec. 2018. [Online]. Available: https://doi.org/10.33936/riemat.v3i2.1627

[9] A. K. Montesdeoca Rivas and M. Cedillo Fajardo, "Análisis de la calidad de servicio y la satisfacción del usuario externo en la empresa eléctrica pública estratégica corporación nacional de electricidad CNEL EP, Guayaquil–2020," *Ciencia Latina Revista Científica Multidisciplinar*, vol. 7, no. 3, pp. 9840–9858, Jul. 2023. [Online]. Available: https://doi.org/10.37811/cl_rcm.v7i3.7009

[10] C. A. Vásconez-Mejía, J. F. Molina-Mora, and M. E. Gaspar-Santos, "Derechos de los consumidores del servicio eléctrico en tiempos de pandemia en Ecuador," *CIENCIAMATRIA*, vol. 8, no. 2, pp. 80–91, Mar. 2022. [Online]. Available: https://doi.org/10.35381/cm.v8i2.698

[11] M. Neira and R. Quintanilla, "La calidad del servicio técnico de electricidad," *Revista Técnica "Energía"*, vol. 8, no. 1, Jan. 2012. [Online]. Available: https://doi.org/10.37116/revistaenergia.v8.n1.2012.189

[12] C. C. Orozco Domínguez, A. F. Sánchez Hernández, and M. J. Santiago Camacho, "Deficiencias

en la prestación del servicio de energía eléctrica en la ciudad de Cartagena," *Vis Iuris*, pp. 55–74, Nov. 2017. [Online]. Available: https://doi.org/10.22518/vis.v0i00.1174

[13] G. E. Chodzaza and H. S. Gombachika, "Service quality, customer satisfaction and loyalty among industrial customers of a public electricity utility in Malawi," *International Journal of Energy Sector Management*, vol. 7, no. 2, pp. 269–282, Jun. 2013. [Online]. Available: https://doi.org/10.1108/IJESM-02-2013-0003

[14] S. Abideen, O. Joseph, J. Surajudeen, and A. Adeyinka, "Service quality and customer satisfaction of electricity distribution company: A study of Olumo Business Hub, Abeokuta, Ogun State," *SSRN Electronic Journal*, 2018. [Online]. Available: https://doi.org/10.2139/ssrn.3391404

[15] W. Li, W. W. Kwabla Pomegbe, C. S. Kofi Dogbe, and J. D. Novixoxo, "Employees' customer orientation and customer satisfaction in the public utility sector: The mediating role of service quality," *African Journal of Economic and Management Studies*, vol. 10, no. 4, pp. 408–423, Dec. 2019. [Online]. Available: https://doi.org/10.1108/AJEMS-10-2018-0314

[16] D. Sarathkumar, M. Srinivasan, A. A. Stonier, R. Samikannu, N. R. Dasari, and R. A. Raj, "A technical review on classification of various faults in smart grid systems," *IOP Conference Series: Materials Science and Engineering*, vol. 1055, no. 1, p. 012152, Feb. 2021. [Online]. Available: https://doi.org/10.1088/1757-899X/1055/1/012152

[17] A. S. Alhanaf, H. H. Balik, and M. Farsadi, "Intelligent fault detection and classification schemes for smart grids based on deep neural networks," *Energies*, vol. 16, no. 22, p. 7680, Nov. 2023. [Online]. Available: https://doi.org/10.3390/en16227680

[18] Z. Krivohlava, S. Chren, and B. Rossi, "Failure and fault classification for smart grids," *Energy Informatics*, vol. 5, no. 1, Oct. 2022. [Online]. Available: https://doi.org/10.1186/s42162-022-00218-3

[19] J. Hosseinzadeh, F. Masoodzadeh, and E. Roshandel, "Fault detection and classification in smart grids using augmented K-NN algorithm," *SN Applied Sciences*, vol. 1, no. 12, Nov. 2019. [Online]. Available: https://doi.org/10.1007/s42452-019-1672-0

[20] M. Mohammadian, F. Aminifar, N. Amjady, and M. Shahidehpour, "Data-driven classifier for extreme outage prediction based on Bayes decision theory," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 4906–4914, Nov. 2021. [Online]. Available: https://doi.org/10.1109/TPWRS.2021.3086031

[21] S. Rana, "AI-driven fault detection and predictive maintenance in electrical power systems: A systematic review of data-driven approaches, digital twins, and self-healing grids," *American Journal of Advanced Technology and Engineering Solutions*, vol. 1, no. 01, pp. 258–289, Feb. 2025. [Online]. Available: https://doi.org/10.63125/4p25x993

[22] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, "An improved random forest based on the classification accuracy and correlation measurement of decision trees," *Expert Systems with Applications*, vol. 237, p. 121549, Mar. 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2023.121549

[23] H. Jing, H. Cao, and S. Sun, "Multi-objective predictive maintenance optimization of electric energy meters based on LSTM-XGBoost and modified firefly algorithm," *International Journal of Emerging Electric Power Systems*, Aug. 2025. [Online]. Available: https://doi.org/10.1515/ijeeps-2025-0184

[24] I. Hussain, K. B. Ching, C. Uttraphan, K. G. Tay, and A. Noor, "Evaluating machine learning algorithms for energy consumption prediction in electric vehicles: A comparative study," *Scientific Reports*, vol. 15, no. 1, May 2025. [Online]. Available: https://doi.org/10.1038/s41598-025-94946-7

[25] Y. Khalili, M. Ahmadi, and M. K. Moraveji, "Time-aware predictive maintenance of electrical submersible pumps using catboost ensemble learning and trend-based labeling," *Journal of Petroleum Exploration and Production Technology*, vol. 15, no. 9, Aug. 2025. [Online]. Available: https://doi.org/10.1007/s13202-025-02070-z

[26] T. Ahmad, H. Chen, R. Huang, G. Yabin, J. Wang, J. Shair, H. M. Azeem Akram, S. A. Hassnain Mohsan, and M. Kazim, "Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment," *Energy*, vol. 158, pp. 17–32, Sep. 2018. [Online]. Available: https://doi.org/10.1016/j.energy.2018.05.169

[27] T. Zhang, R. Gao, and S. Sun, "Theories, applications and trends of non-technical losses in power utilities using machine learning," in *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. IEEE,

May 2018, pp. 2324–2329. [Online]. Available: https://doi.org/10.1109/imcec.2018.8469410

[28] J. Ausmus, R. S. de Carvalho, A. Chen, Y. N. Velaga, and Y. Zhang, "Big data analytics and the electric utility industry," in *2019 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA)*. IEEE, May 2019. [Online]. Available: https://doi.org/10.1109/SGSMA.2019.8784657

[29] M. J. Mousavi, J. Stoupis, and K. Saarinen, "Event zone identification in electric utility systems using statistical machine learning," in *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*. IEEE, Apr. 2018, pp. 1–9. [Online]. Available: https://doi.org/10.1109/TDC.2018.8440570

[30] Y. Shen, M. Abubakar, H. Liu, and F. Hussain, "Power quality disturbance monitoring and classification based on improved PCA and convolution neural network for wind-grid distribution systems," *Energies*, vol. 12, no. 7, p. 1280, Apr. 2019. [Online]. Available: https://doi.org/10.3390/en12071280

[31] L. Xu and M.-Y. Chow, "A classification approach for power distribution systems fault cause identification," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 53–60, Feb. 2006. [Online]. Available: https://doi.org/10.1109/TPWRS.2005.861981

[32] B. Kumwenda, E. Zulu, and M. Ndiaye, "System disturbance classification model for a low voltage distribution network," in *2024 IEEE PES/IAS PowerAfrica*. IEEE, Oct. 2024, pp. 01–05. [Online]. Available: https://doi.org/10.1109/PowerAfrica61624.2024.10759455