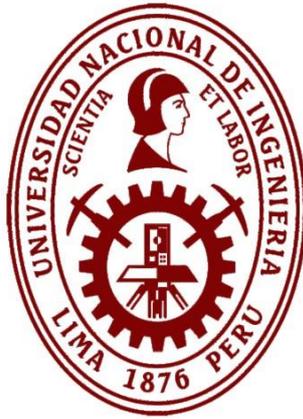


UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA MECÁNICA
UNIDAD DE POSGRADO



TESIS:

“MODELO DE APRENDIZAJE AUTOMÁTICO PARA DETECTAR ANOMALÍAS EN UN TRANSFORMADOR DE UNA EMPRESA DE DISTRIBUCIÓN ELÉCTRICA”

PARA OBTENER EL GRADO ACADÉMICO DE MAESTRO EN INGENIERÍA CON MENCIÓN EN GERENCIA E INGENIERÍA DE MANTENIMIENTO

ELABORADA POR:

MARTIN JONATHAN MORALES VILLANUEVA

ASESOR:

Dr. DHEYBI GROVER CERVAN PRADO

LIMA, PERÚ

2024

Dedicatoria

Para mi padre
que me apoyo desde siempre
y que desde el cielo sonr e
por cada triunfo en mi vida.

Agradecimiento

Quiero expresar mi más sincero agradecimiento:

A mi esposa y a mi hija, por su invaluable apoyo y por ceder parte de su tiempo, permitiéndome alcanzar este objetivo académico.

Al Dr. Dheybi Grover Cervan Prado, mi asesor, por su dedicación y orientación en la realización de este trabajo de investigación, asegurando que cumpla con los más altos estándares académicos.

A la Unidad de Posgrado de la Facultad de Ingeniería Mecánica, por brindarme la oportunidad de formar parte de esta maestría.

A mis compañeros de maestría, con quienes compartí valiosas experiencias que han enriquecido mi desarrollo profesional.

A la Universidad Nacional de Ingeniería, mi alma mater, que ha sido fundamental en mi formación académica y profesional.

Al programa CARELEC - MEM, por otorgar becas que permiten a profesionales del sector actualizarse en nuevos conocimientos y contribuir así al desarrollo del país.

A Pluz Energía Perú S.A.A., por permitirme ser parte de este programa de estudios.

A mis compañeros de trabajo, cuyas experiencias han sido esenciales para enriquecer mi conocimiento profesional y aportar significativamente en esta maestría.

Finalmente, a Dios, quien me ha dado un día más de vida para seguir creciendo tanto como profesional y como persona.

Índice de Contenidos

Dedicatoria	ii
Agradecimiento	iii
Índice de Contenidos	iv
Lista de tablas	viii
Lista de figuras.....	ix
Lista de ecuaciones	xi
Resumen.....	xii
Abstract	xiii
Introducción.....	xiv
CAPÍTULO I. PROTOCOLO DE LA INVESTIGACIÓN.....	1
1.1. Antecedentes de la investigación	1
1.2. Identificación y Descripción del Problema de Estudio	13
1.3. Formulación del problema	17
1.3.1. Problema principal	17
1.3.2. Problemas secundarios	17
1.4. Justificación e importancia.....	17
1.5. Objetivos	19
1.5.1. Objetivo General	19
1.5.2. Objetivos Específicos	19
1.6. Hipótesis	19
1.6.1. Hipótesis General	19

1.6.2.	Hipótesis Específicas.....	19
1.7.	Variables y Operacionalización de Variables	20
1.8.	Diseño Metodológico	21
1.8.1.	Unidades de Análisis	21
1.8.2.	Fuentes de Información	21
1.8.3.	Tipo de Investigación	21
1.8.4.	Selección de 01 transformador para la instalación del equipo DTM	22
1.8.5.	Técnicas e Instrumentos de recolección de Datos	22
1.8.6.	Técnicas e instrumentos de Análisis y Procesamiento de Datos	22
1.9.	Contribuciones	23
CAPÍTULO II. MARCO TEÓRICO.....		24
2.1.	Bases Teóricas.....	24
2.1.1.	Aprendizaje automático (Machine Learning).....	24
2.1.2.	Aprendizaje Automático Supervisado.....	25
2.1.3.	Tipo de Aprendizaje Automático Supervisado – Clasificación	25
2.1.4.	Árbol de Decisión (Decision Tree).....	26
2.1.5.	Métricas de aprobación del algoritmo Árbol de decisión	27
2.1.6.	Anomalías en redes de distribución	27
2.1.7.	Transformador de distribución.....	27
2.1.8.	Modos de Falla y Análisis de Efectos (AMFE)	28
2.1.9.	Criterios de aprobación del proyecto.....	28
2.2.	Marco Conceptual.....	29
2.2.1.	Aprendizaje Automático	29

2.2.2.	Aprendizaje Automático Supervisado.....	29
2.2.3.	Tipo de Aprendizaje Automático Supervisado – Clasificación	29
2.2.4.	Árbol de decisiones.....	29
2.2.5.	Transformador de distribución.....	29
2.2.6.	Anomalías en un transformador de distribución.....	30
2.2.7.	Detectar anomalías.....	30
2.3.	Marco Legal	30
2.3.1.	Calidad de Producto – Tensión.....	30
2.3.2.	Calidad de Suministro – Interrupciones	30
CAPÍTULO III. METODOLOGÍA.....		31
3.1.	Introducción.....	31
3.2.	Análisis de datos históricos	35
3.2.1.	Parametrización de anomalías	35
3.2.2.	Etiquetado de datos	36
3.3.	Entrenamiento del modelo basado en el árbol de Decisiones.....	38
3.3.1.	Preparación de datos.....	38
3.3.2.	Entrenamiento del algoritmo.....	39
3.3.3.	Métricas de desempeño	41
3.3.4.	Verificar el modelo basado en el árbol de decisiones.....	43
3.4.	Evaluación Económica	43
CAPÍTULO IV. RESULTADOS.....		46
4.1.	Recolección de datos	46
4.2.	Resultados de la parametrización de anomalías	52

4.3.	Resultado del etiquetado de datos	54
4.4.	Resultados del entrenamiento del algoritmo	56
4.4.1.	Preparación de datos	56
4.4.2.	Entrenamiento del algoritmo	56
4.4.3.	Métricas de desempeño	62
4.4.4.	Resultados de verificar el modelo basado en el árbol de decisiones	73
4.5.	Resultados de la evaluación económica	74
4.6.	Discusión	77
	CONCLUSIONES	79
	RECOMENDACIONES	81
	BIBLIOGRAFÍA	83
	ANEXO A. MATRIZ DE CONSISTENCIA	87
	ANEXO B. ALGORITMOS	88

Lista de tablas

Tabla 1. Documentos referidos al uso de algoritmos de inteligencia artificial	9
Tabla 2. Variables usadas para el entrenamiento de algoritmos de inteligencia artificial .	10
Tabla 3. Etapas del monitoreo y control de condiciones subestándares.....	11
Tabla 4. Aplicación de algoritmos de inteligencia artificial en instalaciones eléctricas	12
Tabla 5. Operacionalización de las variables	20
Tabla 6. Plantilla para el AMFE.....	28
Tabla 7. Criterios para etiquetado de datos con anomalía	37
Tabla 8. Flujo de caja para el análisis financiero del proyecto	44
Tabla 9. Modos de falla y análisis de efectos de un transformador.....	52
Tabla 10. Modos de controles actuales	53
Tabla 11. Anomalías del presente trabajo de investigación	54
Tabla 12. Filas de datos con etiqueta de anomalía.....	55
Tabla 13. Métricas de desempeño en las iteraciones	72
Tabla 14. Datos etiquetados manualmente del DataSet 02.....	73
Tabla 15. Datos detectados como anomalías en el DataSet 02.....	74
Tabla 16. Costos de inversión del proyecto de instalación de 20 DTM.....	75
Tabla 17. Ahorro promedio anual por transformador	75
Tabla 18. Flujo de caja durante 05 años	76

Lista de figuras

Figura 1. Evolución del SAIDI del 2015 al 2022 de una empresa de Distribución.....	14
Figura 2. Evolución del SAIFI del 2015 al 2022 de una empresa de Distribución	15
Figura 3. Procesamiento de las mediciones puntuales de corriente y voltaje.	16
Figura 4. Interacción que se genera en un aprendizaje automático.....	24
Figura 5. Componentes de un árbol de decisiones	26
Figura 6. Nueva forma de trabajar aplicando aprendizaje automático.....	32
Figura 7. Diagrama de flujo de la metodología propuesta.....	33
Figura 8. Diagrama de flujo para identificar anomalías	36
Figura 9. DTM instalado en un transformador	36
Figura 10. Diagrama de flujo del código de etiquetado de datos	37
Figura 11. Subconjunto de datos que contienen datos normales y anómalos.....	38
Figura 12. Iteraciones del conjunto de datos para el entrenamiento y prueba	39
Figura 13. Procesos del algoritmo de árbol de decisiones	40
Figura 14. Diagrama de flujo para verificar el modelo basado en árbol de decisiones	43
Figura 15. Diagrama de carga de las tres fases de corriente durante el día 18/06/2024 .	47
Figura 16. Correlación entre Corriente y Voltaje en baja tensión	48
Figura 17. Correlación entre anomalías vs Corriente y Voltajes en baja tensión.....	49
Figura 18. Histograma de corriente en baja tensión	49
Figura 19. Densidad de datos de corriente en baja tensión	50
Figura 20. Histograma de voltaje en baja tensión.....	51
Figura 21. Densidad de datos de voltaje en baja tensión	51
Figura 22. Árbol de decisión – Iteración 01.....	57
Figura 23. Árbol de decisión – Iteración 02.....	57
Figura 24. Árbol de decisión – Iteración 03.....	58
Figura 25. Árbol de decisión – Iteración 04.....	58
Figura 26. Árbol de decisión – Iteración 05.....	59

Figura 27. Árbol de decisión – Iteración 06.....	59
Figura 28. Árbol de decisión – Iteración 07.....	60
Figura 29. Árbol de decisión – Iteración 08.....	60
Figura 30. Árbol de decisión – Iteración 09.....	61
Figura 31. Árbol de decisión – Iteración 10.....	61
Figura 32. Matriz de confusión para el entrenamiento – iteración 01	62
Figura 33. Matriz de confusión para el entrenamiento – iteración 02	63
Figura 34. Matriz de confusión para el entrenamiento – iteración 03	63
Figura 35. Matriz de confusión para el entrenamiento – iteración 04	64
Figura 36. Matriz de confusión para el entrenamiento – iteración 05	64
Figura 37. Matriz de confusión para el entrenamiento – iteración 06	65
Figura 38. Matriz de confusión para el entrenamiento – iteración 07	65
Figura 39. Matriz de confusión para el entrenamiento – iteración 08	66
Figura 40. Matriz de confusión para el entrenamiento – iteración 09	66
Figura 41. Matriz de confusión para el entrenamiento – iteración 10	67
Figura 42. Matriz de confusión para la prueba – iteración 01.....	67
Figura 43. Matriz de confusión para la prueba – iteración 02.....	68
Figura 44. Matriz de confusión para la prueba – iteración 03.....	68
Figura 45. Matriz de confusión para la prueba – iteración 04.....	69
Figura 46. Matriz de confusión para la prueba – iteración 05.....	69
Figura 47. Matriz de confusión para la prueba – iteración 06.....	70
Figura 48. Matriz de confusión para la prueba – iteración 07.....	70
Figura 49. Matriz de confusión para la prueba – iteración 08.....	71
Figura 50. Matriz de confusión para la prueba – iteración 09.....	71
Figura 51. Matriz de confusión para la prueba – iteración 10.....	72

Lista de ecuaciones

Ecuación 1. Cálculo de la incertidumbre en un árbol de decisiones clasificadorio	41
Ecuación 2. Cálculo de la impureza Gini.....	41
Ecuación 3. Cálculo de la precisión del modelo	42
Ecuación 4. Cálculo de la sensibilidad del modelo	42
Ecuación 5. Cálculo de Accuracy	42
Ecuación 6. Cálculo de F1–score.....	42
Ecuación 7. Cálculo del VAN del proyecto	44

Resumen

En el ámbito del mantenimiento eléctrico, es habitual realizar mediciones puntuales de parámetros eléctricos como corriente y voltaje de baja tensión. El tiempo que transcurre entre que el personal realiza la medición y se culmina el análisis de datos para detectar anomalías es de aproximadamente una semana. Las anomalías que no se identifican a tiempo pueden escalar rápidamente, afectando tanto el desempeño del sistema eléctrico y los indicadores clave del negocio. Las consecuencias de la no detección oportuna de una anomalía son la reducción en la venta de energía, reducción de la confiabilidad y el impacto en la satisfacción al cliente.

La presente tesis tiene como objetivo desarrollar un modelo de aprendizaje automático para la detección de anomalías en un transformador de una empresa de distribución eléctrica. Se centra en la identificación de tres tipos principales de anomalías: cortocircuito externo, sobrecarga en el transformador y niveles de voltaje. Para la detección automática de estas anomalías, se desarrolló un modelo de Árbol de Decisión que permitió clasificar con precisión los datos en categorías de anómalos y no anómalos. Los resultados obtenidos confirman la alta eficiencia y precisión del modelo propuesto, logrando identificar correctamente las anomalías dentro del extenso conjunto de datos analizados. Además de su precisión, la implementación del modelo demostró ser económicamente viable.

La aplicación de técnicas de aprendizaje automático para la detección de anomalías en transformadores permite reducir los tiempos en la toma de decisiones. Lo cual impacta de forma positiva en la gestión de mantenimiento, contribuye significativamente a mejorar la eficiencia operativa. La adopción de este tipo de tecnologías es un paso esencial para modernizar la gestión de mantenimiento y garantizar la continuidad del servicio eléctrico.

Palabras claves: Aprendizaje automático, detección de anomalía, transformador, árbol de decisión.

Abstract

In the field of electrical maintenance, it is common to conduct point measurements of electrical parameters such as low-voltage current and voltage. The time elapsed between conducting the measurement and completing the data analysis to detect anomalies is approximately one week. Anomalies that are not identified promptly can escalate quickly, affecting both the performance of the electrical system and key business indicators. The consequences of not detecting anomalies in time include reduced energy sales, decreased reliability, and negative impacts on customer satisfaction.

This thesis aims to develop a machine learning model for detecting anomalies in a transformer belonging to an electrical distribution company. It focuses on identifying three main types of anomalies: external short circuits, transformer overloads, and voltage levels. To enable the automatic detection of these anomalies, a Decision Tree model was developed, which accurately classified data into anomalous and non-anomalous categories. The results obtained confirm the high efficiency and accuracy of the proposed model, successfully identifying anomalies within the extensive dataset analyzed. In addition to its precision, the model's implementation proved to be economically viable.

The application of machine learning techniques for anomaly detection in transformers reduces decision-making times, positively impacting maintenance management and significantly improving operational efficiency. The adoption of such technologies is an essential step toward modernizing maintenance management and ensuring continuity in electrical service.

Keywords: Machine learning, anomaly detection, transformer, decision tree.

Introducción

La continuidad del suministro eléctrico enfrenta grandes desafíos, entre los cuales destaca la necesidad de mantener la confiabilidad del sistema eléctrico de distribución. Una manera efectiva de mejorar significativamente estos niveles de confiabilidad es implementar herramientas basadas en inteligencia artificial, en especial aquellas relacionadas con el aprendizaje automático, para la detección de anomalías en transformadores de distribución. El transformador de distribución es crucial para que los usuarios finales reciban la energía eléctrica en baja tensión; por ello, es vital que opere de manera ininterrumpida. Los trabajos de mantenimiento predictivo y preventivo deben garantizar que no se produzcan fallas que interrumpan la continuidad del servicio eléctrico. Pues actualmente la electricidad juega un rol importante en la sociedad.

El objetivo general de esta tesis es detectar anomalías en un transformador de una empresa de distribución eléctrica mediante técnicas de aprendizaje automático. Además, los objetivos específicos son: parametrizar las anomalías mediante el análisis de datos históricos; detectar anomalías usando el árbol de decisiones; y obtener los indicadores de rentabilidad mediante la evaluación económica.

La hipótesis general postula que el aprendizaje automático permite detectar anomalías en un transformador de una empresa de distribución eléctrica. Las hipótesis específicas son: a través del análisis de datos históricos, se logra la parametrización de anomalías; que, mediante el uso del árbol de decisiones, se logra detectar anomalías; y mediante la evaluación económica se logra obtener los indicadores de rentabilidad.

La tesis se organiza en cinco capítulos, cada uno de los cuales aborda diferentes aspectos del desarrollo de la investigación. El Capítulo 1 presenta el protocolo de investigación, donde se discuten los antecedentes, la identificación del problema, la justificación, los objetivos y las hipótesis. Además, se detallan las variables y la metodología empleada en el estudio. El Capítulo 2 está dedicado al marco teórico y

conceptual, donde se exploran los fundamentos del aprendizaje automático, la detección de anomalías, y la aplicación de técnicas específicas como el AMFE en el contexto de transformadores de distribución. En el Capítulo 3, se desarrolla el trabajo de investigación, incluyendo la parametrización de anomalías, el entrenamiento del algoritmo basado en el árbol de decisiones para detectar anomalías y la rentabilidad de la implementación del proyecto. El Capítulo 4 se centra en la discusión de los resultados obtenidos, evaluando la efectividad del modelo propuesto y su impacto en la reducción de riesgos y costos operativos. Finalmente, el Capítulo 5 presenta las conclusiones generales del estudio, así como recomendaciones para futuras investigaciones y la implementación del modelo en un contexto más amplio dentro del sector de distribución eléctrica.

Esta investigación representa una contribución significativa al campo de la ingeniería del mantenimiento eléctrico, al proporcionar una herramienta avanzada para la identificación y gestión de anomalías en transformadores de distribución. La implementación de un modelo de aprendizaje automático no solo mejorará la eficiencia operativa y reducirá los indicadores SAIDI y SAIFI, sino que también contribuirá a la reducción de costos y a la continuidad del servicio eléctrico.

CAPÍTULO I. PROTOCOLO DE LA INVESTIGACIÓN

1.1. Antecedentes de la investigación

Los sistemas de distribución eléctrica producen una gran cantidad de datos derivados de las actividades de mantenimiento predictivo, preventivo y correctivo. En la actualidad, se dispone de una variedad de investigaciones previas en las que procesaron grandes volúmenes de información. A continuación, se detalla las aportaciones de diversos autores en relación con la aplicación de aprendizaje automático a sistemas eléctricos.

En la investigación realizada por De Benedetti et al. (2018)¹, se emplean redes neuronales artificiales para predecir el consumo de energía y detectar anomalías en sistemas fotovoltaicos. En particular, se utiliza un perceptrón multicapa (Multi-Layer Perceptron, MLP), el cual, mediante el aprendizaje supervisado, entrena su modelo con datos históricos de un año para predecir el consumo energético y detectar anomalías con los datos obtenidos en tiempo real, como parte de las actividades de mantenimiento predictivo. Tras aplicar este enfoque en sus labores de monitoreo y control, se demostró la eficacia del algoritmo, alcanzando una tasa de detección de anomalías superior al 90% y un margen del error cuadrático medio (RMSE) de validación igual al 2,3%. Los autores proponen que el trabajo futuro podría incluir el desarrollo de reglas de clasificación de fallas, una vez que el modelo haya alcanzado la madurez y se haya validado su precisión y robustez. Además, podría enfocarse en la identificación de las causas subyacentes de las fallas, proporcionando al operador información valiosa para implementar las acciones correctivas necesarias.

¹ DE BENEDETTI, M., LEONARDI, F., MESSINA, F., SANTORO, C., VASILAKOS, A. (2018). Anomaly detection and predictive maintenance for photovoltaic systems. *Neurocomputing*, 310, 59-68. <https://doi.org/10.1016/j.neucom.2018.05.017>

Wang et al. (2022)² señalaron que el desarrollo económico en China está directamente vinculado con el avance tecnológico, el cual se logra integrando las redes de distribución con la informática y las comunicaciones. No obstante, esta tecnología se ha desarrollado únicamente en generación y transmisión, dejando de lado la distribución. La red neuronal empleada es el auto-codificador (AE), que a través del aprendizaje no supervisado comprime y reconstruye datos, complementándose con un auto-codificador variacional (VAE) que introduce ruido gaussiano para aumentar la robustez del modelo. Para el entrenamiento del modelo, se utilizó información histórica de carga eléctrica desde enero de 2016 hasta enero de 2021. Los datos fueron divididos en 150 épocas con un tamaño de lote de 50 datos. Tras el procesamiento del modelo, se verificó una precisión de 0.998 y una exactitud de 0.889. Además, se realizó el entrenamiento con otros modelos como iForest (muestra máxima de 64 y 50 árboles) y LSTM (150 épocas y un tamaño de lote de 100), obteniendo para el iForest una precisión de 0.750 y una exactitud de 0.997, y para el LSTM una precisión de 0.807 y una exactitud de 0.996. Los autores concluyen que el algoritmo seleccionado es eficaz para identificar datos anómalos en la red de distribución de energía, demostrando una alta precisión y exactitud.

Veerakumar et al. (2023)³ emplean unidades de medición fasorial (PMU) en redes de distribución de 50 kV en los Países Bajos. Los PMU son dispositivos utilizados para medir las ondas de corriente y voltaje en el dominio del tiempo, cuyos datos son enviados a una plataforma web. Esta plataforma dispone de técnicas eficientes para la estimación de estado, así como para la detección, discriminación e identificación de anomalías, tales como cambios súbitos de carga y datos erróneos. Se utilizan dos herramientas

² WANG, Z., DING, Y., ZHANG, T. (2022). Distribution Network Anomaly Detection Algorithm Based on VAE. 11th International Conference of Information and Communication Technology (ICTech), 84-87. <https://doi.org/10.1016/j.neucom.2018.05.017>

³ VEERAKUMAR, N., CETENOVIC, D., KONGURAI, K., POPOV, M., JONGEPIER, A., TERZIJA, V. (2023). PMU-based Real-time Distribution System State Estimation Considering Anomaly Detection, Discrimination and Identification. International Journal of Electrical Power and Energy Systems, 148, 108916. <https://doi.org/10.1016/j.ijepes.2022.108916>

computacionales: la estimación del estado asistida por pronóstico (FASE) para las mediciones de corriente y voltaje en tiempo real, y el filtro de Kalman extendido (EKF) para pronósticos y filtrados rápidos de estado. A partir de estas herramientas, desarrollaron el método de detección, discriminación e identificación de anomalías (ADDI), que integra técnicas de inteligencia artificial. Entre estas técnicas se incluye el análisis de innovación normalizada (LNI), que detecta anomalías basándose en las diferencias entre los valores reales y las predicciones del modelo; la asimetría de las distribuciones de innovaciones normalizadas (NI), que analiza la simetría para identificar el tipo de anomalía, donde una desviación indica un error de datos y una simetría mantenida sugiere un cambio súbito de carga; y el índice de razón de asimetría (SIR), empleado para mejorar la precisión en la discriminación de anomalías. Los autores concluyen que el método utilizado es viable para el procesamiento de datos, permitiendo que la detección, discriminación e identificación de anomalías se realice de manera rápida y confiable, cumpliendo con los requisitos de alta eficiencia computacional. Además, recomiendan que un reto para futuras investigaciones sería ampliar los criterios utilizados para la detección de anomalías sin comprometer la eficiencia en el procesamiento de datos, y también mejorar la capacidad para detectar fallas.

Liu et al. (2023)⁴ abordan los problemas de baja precisión y altos costos en los métodos tradicionales para la detección temprana de fallas en transformadores, proponiendo una metodología que combina parámetros eléctricos y vibraciones. Utilizan clustering jerárquico aglomerativo no supervisado para identificar patrones en las vibraciones de 439 transformadores, lo que permite etiquetar los datos eléctricos de manera efectiva. Posteriormente, aplican un árbol de decisiones a los datos etiquetados, lo que les permite predecir de manera temprana las fallas en los transformadores de

⁴ LIU, X., LI, J., SHAO, L., LIU, H., REN, L., ZHU, L. (2023). Transformer Fault Early Warning Analysis Based on Hierarchical Clustering Combined with Decision Trees. *Energies* 2023, 16 (3). <https://doi.org/10.3390/en16031168>

distribución. Los resultados muestran una precisión del 92% en la detección de fallas mecánicas, además de una reducción significativa en la tasa de falsos positivos, lo que representa una mejora en la fiabilidad del sistema. En conclusión, el enfoque propuesto facilita la detección temprana de fallas en transformadores, mejorando la seguridad operativa y reduciendo los riesgos de interrupciones eléctricas inesperadas. Este método ofrece una solución más eficiente y rentable que los enfoques tradicionales, permitiendo una mejor gestión del mantenimiento predictivo en transformadores de distribución.

Ma et al. (2023)⁵ destacan que, para garantizar la confiabilidad del suministro de energía en la red de distribución, mejorar la satisfacción del cliente en relación con el consumo eléctrico y establecer un monitoreo efectivo de las sobrecargas en transformadores públicos, han propuesto la implementación conjunta de los modelos de aprendizaje automático LSTM y XGBoost. En primer lugar, utilizan el algoritmo LSTM para construir un modelo de predicción de carga que estima la segunda carga más alta del día siguiente. En segundo lugar, extraen veintidós características de los datos históricos de operación y de las condiciones meteorológicas. Posteriormente, con el algoritmo XGBoost, se realiza una predicción a corto plazo de la sobrecarga. Para este proceso, se toma como referencia la información histórica de los tres meses anteriores, dividiéndola en cuatro periodos semanales de medición, utilizando datos de 66,200 transformadores en una zona específica y procesándolos en nueve servidores. Para evaluar el desempeño del modelo LSTM-XGBoost, se compararon otros algoritmos como la retropropagación, los árboles de decisión y los árboles de decisión de aumento de gradiente, obteniendo para el LSTM-XGBoost un valor de precisión del 85.87% y una tasa de recuperación del 72.67%, valores que superaron a los obtenidos por los otros modelos evaluados. Los autores concluyen que el modelo propuesto, además de considerar factores eléctricos, toma en cuenta el día

⁵ MA, H., YANG, P., WANG, F., WANG, X., YANG, D., FENG, B. (2023). Short – Term Heavy Overload Forecasting of Public Transformers Base on Combined LSTM – XGBoost Model. *Energies* 2023, 16 (3). <https://doi.org/10.3390/en16031507>

de la semana, el mes y la temperatura, analizando el impacto de estas variables en la predicción de sobrecargas. Esto se logra combinando una predicción de carga a largo plazo con una predicción de sobrecarga a corto plazo.

Ramesh et al. (2022)⁶ expresan que los transformadores de distribución constituyen un componente fundamental en una red de distribución de energía y en las emergentes redes inteligentes. Debido al incremento en la demanda por parte de los consumidores, aumenta significativamente la probabilidad de fallas por sobrecarga, fallas en las líneas de alimentación y un enfriamiento ineficaz. Los autores han implementado un sistema basado en IoT para recopilar datos de corriente y temperatura del transformador, los cuales son enviados a la nube para su procesamiento. Han aplicado el algoritmo de bosques de aislamiento (Isolation Forest) para la detección de anomalías, logrando identificar posibles fallas con una anticipación de 24 horas. Los resultados obtenidos muestran que se detectó el 100% de las anomalías en un amplio conjunto de datos en tiempo real. Sin embargo, uno de los desafíos que enfrentan es la escasez de datos anómalos en el conjunto de datos.

Moldovan & Buzdugan (2023)⁷ presentan un modelo de redes neuronales artificiales (ANN) diseñado para detectar, localizar, clasificar y predecir fallas, enfocándose particularmente en cortocircuitos en cables eléctricos. Este tipo de fallas impacta de manera significativa en la continuidad del servicio y en el tiempo de duración de la misma. La metodología empleada consistió en modelar en Simulink una red trifásica de 20 kV con diferentes líneas que suman una longitud total de 22 km. Posteriormente, se realizaron 6150 simulaciones de la red para generar el conjunto de datos necesario para el modelo

⁶ RAMESH, J., SHAHRIAR, S., AL-ALI, A., OSMAN, A., SHAABAN, M. (2022). Machine Learning Approach for Smart Distribution Transformers Load Monitoring and Management System. MDPI, 15(21), 7981. <https://doi.org/10.3390/en15217981>

⁷ MOLDOVAN, A., BUZDUGAN, M. (2023). Prediction of Faults Location and Type in Electrical Cables Using Artificial Neural Network. Sustainability (Switzerland), 15(7). <http://doi.org/10.3390/su15076162>

de aprendizaje automático, logrando una tasa de precisión del 98% en la localización de fallas y del 94.7% en la identificación del tipo y ubicación de la falla. Durante el proceso de simulación, se implementaron 12 tipos de fallas a lo largo de la red, colocándolas en 22 posiciones distintas y formando 25 escenarios diferentes. Los autores concluyen que la aplicación de redes neuronales ofrece una gran ventaja en la predicción y localización de fallas, además de reducir significativamente los tiempos de simulación.

Xiang et al. (2020)⁸ explican que los transformadores de distribución son equipos claves en la red de distribución eléctrica. Para la detección de anomalías en los informes de inspección se propone el uso de algoritmos como la distribución gaussiana simple y multivariada, el factor de vecino local (LOF) y la máquina de vectores de soporte de una clase (one-class SVM). El trabajo de investigación introduce y analiza el problema de informes anómalos y para ello se sugiere el uso del aprendizaje automático supervisado para la detección de anomalías. Los resultados que obtuvieron en su trabajo de investigación fueron que one-class SVM es más eficaz para detectar anomalías ya que detectó en siete de nueve conjuntos de datos probados. En caso de no contar con datos suficientes introduce el uso de algoritmos para la construcción de anomalías. Hace uso de los algoritmos propuestos para evaluar cual es el mejor de ellos, obteniendo que el one-class SVM es el idóneo para la detección de anomalías.

Xiao et al. (2018)⁹ señalan que la detección de anomalías en la temperatura del aceite del transformador es un aspecto crucial y valioso para la operación segura de los transformadores y del sistema eléctrico. Los autores proponen un enfoque híbrido de modelamiento que comienza con el uso de un algoritmo de clustering no supervisado para

⁸ XIANG, B., LIU, Z., ZHANG, K. (2020). Flagging Implausible Inspection Reports of Distribution Transformers via Anomaly Detection. IEEE Access, 8, 75798-75808. <http://doi.org/10.1109/ACCESS.2020.2989330>

⁹ XIAO, F., LENG, X., YE, K., HU, Y., LI, X., ZHU, L. (2018). The Research of Anomaly Detection Method for Transformer Oil Temperature Based on Hybrid Model of Non-Supervised Learning and Decision Forests. IOP Conference Series, 192 (1). <http://doi.org/10.1088/1755-1315/192/1/012020>

definir el umbral de temperatura, estableciendo así cuándo una temperatura es considerada anómala. Posteriormente, utilizan bosque de decisión (decision forest) para detectar dichas anomalías. La detección de anomalías se basa en dos reglas: la primera establece que, bajo las mismas condiciones de operación, es poco probable que la temperatura sea anormal; la segunda considera que, aunque la temperatura sea la misma, diferentes condiciones de operación pueden indicar una situación normal o anormal. Los resultados que obtuvieron fueron que se logró una precisión de 90% y una tasa de Recall del 89%, que es superior a la precisión del 86% cuando se usa los métodos clásicos cromatográficos aplicados en el aceite. En el experimento, emplearon datos del transformador número 4 de una subestación en Shanghai, utilizando un conjunto de datos que abarca desde diciembre de 2014 hasta enero de 2016 para entrenar el modelo híbrido de detección de anomalías, validándolo posteriormente con datos del año 2017.

Manivannan et al. (2012)¹⁰ manifiesta que su principal objetivo es describir un nuevo sistema de clasificación en línea, no intrusivo, destinado a identificar y reportar eventos en el sistema de energía que ocurren en un alimentador de distribución, basándose en la causa subyacente, utilizando señales adquiridas en la subestación de distribución. Idealmente, este sistema debe extraer e informar parámetros que ayuden a las empresas de servicios públicos a localizar el componente defectuoso. Para ello, desarrollaron un sistema experto basado en lógica difusa, que permite clasificar los datos de eventos del sistema de potencia. Este clasificador es jerárquico, difuso, escalable y modular. Posteriormente, implementaron un motor de inferencia difusa que soporta una estructura jerárquica para la clasificación de eventos en el sistema de potencia. Además, identificaron los rasgos característicos correspondientes a los componentes del sistema de potencia que están en proceso de fallar o que ya han fallado. Utilizaron información existente de

¹⁰ MUTHU MANIVANNAN, K. (2012). Power distribution system event classification using fuzzy logic. Tesis de doctorado. Texas, EEUU. Texas A&M University Libraries. <https://doi.org/hdl.handle.net/1969.1/158893>

alimentadores donde se produjeron eventos bajo condiciones normales de funcionamiento, empleando algoritmos que permitieron extraer características específicas de dichos eventos. Asimismo, desarrollaron un algoritmo que permitió medir los parámetros de tensión y corriente eficaz, y analizar las formas de ondas utilizando el Dynamic Time Warping (DTW). Este algoritmo facilitó la clasificación de los eventos, reportando al personal aquellos considerados importantes con la información necesaria para diagnosticar, evaluar y actuar. Finalmente, diseñaron un sistema que permite a los operadores acceder a la información en línea. Los autores concluyen que utilizaron el algoritmo de clasificación basado en lógica difusa (FLCA) para clasificar los datos de eventos del sistema de potencia, logrando identificar las características de eventos causados por arco eléctrico de baja corriente, sobrecorrientes y fallas en bancos de condensadores. Además, desarrollaron algoritmos para la extracción de información y utilizaron el software Fuzzy Dynamic Time Warping (FDTW) para localizar físicamente las fallas y los componentes defectuosos. Finalmente, desarrollaron el Intelligent Power System Event Reporting System (IPSERS), un marco de informes inteligentes que permite procesar, clasificar la información y generar reportes de manera oportuna y fácil de usar.

La revisión de la literatura revela una notable diversidad en la aplicación de algoritmos de inteligencia artificial para la detección de anomalías en sistemas eléctricos de distribución. Sin embargo, muchos estudios emplean arquitecturas complejas que demandan equipos de alto rendimiento y elevado costo, lo que incrementa significativamente los costos de implementación. Este factor se torna especialmente relevante en el contexto de una empresa de distribución eléctrica que opera con aproximadamente 15,000 subestaciones y un número equivalente de transformadores. La Tabla 1 muestra los algoritmos de inteligencia artificial implementados previamente. Para implementar estos algoritmos se requiere de instrumentos avanzados como los PMU¹¹. La

¹¹ PMU: Phasor Measurement Unit

implementación de estas tecnologías en todas las subestaciones supondría una inversión difícil de justificar, dada la escala y diversidad de las instalaciones involucradas. Por lo tanto, existe dos desafíos principales, el primer desafío es utilizar equipos de medición más accesibles económicamente; y el segundo desafío es desarrollar algoritmos que puedan adaptarse al volumen de datos recopilados por los equipos de medición de variables eléctricas en un transformador. Resolver estos dos desafíos facilitaría su implementación a gran escala y aseguraría la sostenibilidad a largo plazo de los procesos de detección de anomalías en una empresa de distribución eléctrica.

Tabla 1. Documentos referidos al uso de algoritmos de inteligencia artificial

Item	Autor	Título	AE-VAE	FASE-EKF	Clustering jerárquico aglomerativo	Árbol de decisiones	Random Forest	LSTM-XGBoost	Isolation Forest	ANN	ONE-CLASS SVM	Fuzzy Logic
1	De Benedetti et al.	Anomaly detection and predictive maintenance for photovoltaic systems	-	-	-	-	-	-	-	✓	-	-
2	Wang et al.	Distribution Network Anomaly Detection Algorithm Based on VAE	✓	-	-	-	-	-	-	-	-	-
3	Veerakumar et al.	PMU-based Real-time Distribution System State Estimation Considering Anomaly Detection, Discrimination and Identification	-	✓	-	-	-	-	-	-	-	-
4	Liu et al.	Transformer Fault Early Warning Analysis Based on Hierarchical Clustering Combined with Decision Trees	-	-	✓	✓	-	-	-	-	-	-
5	Ma et al.	Short – Term Heavy Overload Forecasting of Public Transformers Base don Combined LSTM – XGBoost Model	-	-	-	-	-	✓	-	-	-	-
6	Ramesh et al.	Machine Learning Approach for Smart Distribution Transformers Load Monitoring and Management System	-	-	-	-	-	-	✓	-	-	-
7	Moldovan et al.	Prediction of Faults Location and Type in Electrical Cables Using Artificial Neural Network	-	-	-	-	-	-	-	✓	-	-
8	Xiang et al.	Flagging Implausible Inspection Reports of Distribution Transformers via Anomaly Detection	-	-	-	-	-	-	-	-	✓	-
9	Xiao et al.	The Research of Anomaly Detection Method for Transformer Oil Temperature Based on Hybrid Modelo Non-Supervised Learning and Decision Forests	-	-	✓	-	✓	-	-	-	-	-
10	Manivannan	Power distribution system event classification using fuzzy logic	-	-	-	-	-	-	-	-	-	✓

Nota: Fuente propia

En la Tabla 2, se observa que la detección de anomalías en sistemas eléctricos requiere la medición de diversas variables, como la carga eléctrica, energía, corriente, voltaje, temperatura y vibración. No obstante, la medición de múltiples variables incrementa el costo de los dispositivos de medición, ya que se necesitan equipos sofisticados para monitorear estas variables. A medida que se demanda el monitoreo de más variables, el costo de los equipos tiende a aumentar considerablemente. Lo que podría representar una

barrera para la adopción generalizada de estos métodos en entornos de distribución eléctrica, especialmente en aquellos con limitaciones presupuestarias. Por ello, es necesarios enfocar la investigación en la medición de variables eléctricas que afecten directamente a los clientes, como la corriente y el voltaje en baja tensión. Esta estrategia no solo reduciría los costos asociados a la implementación de equipos de medición, sino que también representaría una oportunidad significativa para desarrollar investigaciones que prioricen estas variables esenciales, facilitando así la optimización de los procesos de detección de anomalías y asegurando un enfoque más práctico y accesible para las empresas de distribución eléctrica.

Tabla 2. Variables usadas para el entrenamiento de algoritmos de inteligencia artificial

Item	Autor	Título	Energía	Carga Eléctrica	Corriente	Voltaje	Vibración	Temperatura
1	De Benedetti et al.	Anomaly detection and predictive maintenance for photovoltaic systems	✓	-	-	-	-	-
2	Wang et al.	Distribution Network Anomaly Detection Algorithm Based on VAE	-	✓	-	-	-	-
3	Veerakumar et al.	PMU-based Real-time Distribution System State Estimation Considering Anomaly Detection, Discrimination and Identification	-	-	✓	✓	-	-
4	Liu et al.	Transformer Fault Early Warning Analysis Based on Hierarchical Clustering Combined with Decision Trees	-	-	✓	✓	✓	-
5	Ma et al.	Short – Term Heavy Overload Forecasting of Public Transformers Base don Combined LSTM – XGBoost Model	-	✓	-	-	-	-
6	Ramesh et al.	Machine Learning Approach for Smart Distribution Transformers Load Monitoring and Management System	-	-	✓	-	-	✓
9	Xiao et al.	The Research of Anomaly Detection Method for Transformer Oil Temperature Based on Hybrid Modelo Non-Supervised Learning and Decision Forests	-	-	-	-	-	✓
10	Manivannan	Power distribution system event classification using fuzzy logic	-	-	✓	✓	-	-

Nota: Fuente propia

En la Tabla 3, se observa que los autores han empleado diferentes etapas en el monitoreo y control de condiciones subestándares de los equipos que componen el sistema eléctrico. Estas etapas se inician con la detección de anomalías, luego con el monitoreo de las variables eléctricas, continua con la alerta temprana de fallas y finalmente se remiten los reportes de los eventos generados. Una demora en informar una condición

subestándar, tiene un doble impacto, el primero es en la reducción de los tiempos de atención. Y el segundo, es en el aumento significativo de los costos para adquirir equipos que lo detecten y en la reparación de estas condiciones subestándares. Mostrando de esta manera una relación inversa entre el tiempo de detección temprana de una condición subestándar con el costo asociado a su detección y reparación. Entonces, para implementar una solución generalizada es necesario la detección temprana de las condiciones subestándares; trayendo como consecuencia el uso de equipos de medición de variables eléctricas de menor costo. Esta estrategia no solo impactaría en la reducción de interrupciones del suministro eléctrico sino en la reducción de los gastos asociados al mantenimiento preventivo y correctivo.

Tabla 3. Etapas del monitoreo y control de condiciones subestándares

Item	Autor	Título	Detección de anomalías	Monitoreo de sobrecargas	Alertas tempranas de falla	Reporte Eventos
1	De Benedetti et al.	Anomaly detection and predictive maintenance for photovoltaic systems	✓	-	-	-
2	Wang et al.	Distribution Network Anomaly Detection Algorithm Based on VAE	✓	-	-	-
3	Veerakumar et al.	PMU-based Real-time Distribution System State Estimation Considering Anomaly Detection, Discrimination and Identification	✓	-	-	-
4	Liu et al.	Transformer Fault Early Warning Analysis Based on Hierarchical Clustering Combined with Decision Trees	-	-	✓	-
5	Ma et al.	Short – Term Heavy Overload Forecasting of Public Transformers Base don Combined LSTM – XGBoost Model	-	✓	-	-
6	Ramesh et al.	Machine Learning Approach for Smart Distribution Transformers Load Monitoring and Management System	-	✓	-	-
8	Xiang et al.	Flagging Implausible Inspection Reports of Distribution Transformers via Anomaly Detection	✓	-	-	-
9	Xiao et al.	The Research of Anomaly Detection Method for Transformer Oil Temperature Based on Hybrid Modelo Non-Supervised Learning and Decision Forests	✓	-	-	-
10	Manivannan	Power distribution system event classification using fuzzy logic	-	-	-	✓

Nota: Fuente propia

En la Tabla 4, se observa que la inteligencia artificial se ha aplicado en los tres niveles de tensión de sistemas eléctricos: generación, transmisión y distribución. Los estudios revisados se centran en la detección de anomalías en componentes del sistema eléctrico, especialmente aquellas que pueden afectar la continuidad del servicio en media y alta tensión. Sin embargo, la mayoría de estos trabajos prioriza la identificación de fallas que impactan la infraestructura de transmisión y distribución, sin prestar suficiente atención a cómo estas fallas podrían generar interrupciones directas para los clientes finales, que son los usuarios más afectados en situaciones de falla. Considerando que en media tensión la inversión para asegurar la transferencia de carga en caso de fallas es considerablemente menor que el costo de proveer doble alimentación a los clientes, entonces el transformador se convierte en un componente crítico para mantener la continuidad del servicio.

Tabla 4. Aplicación de algoritmos de inteligencia artificial en instalaciones eléctricas

Item	Autor	Título	Sistema de Distribución	Sistema Potencia	Sistema FotoVoltaico	Mantenimiento Predictivo	Transformador	Red eléctrica	Paneles solares	Red distribución
1	De Benedetti et al.	Anomaly detection and predictive maintenance for photovoltaic systems	-	-	✓	✓	-	-	✓	-
2	Wang et al.	Distribution Network Anomaly Detection Algorithm Based on VAE	✓	-	-	-	-	-	-	✓
3	Veerakumar et al.	PMU-based Real-time Distribution System State Estimation Considering Anomaly Detection, Discrimination and Identification	-	✓	-	✓	-	-	-	✓
4	Liu et al.	Transformer Fault Early Warning Analysis Based on Hierarchical Clustering Combined with Decision Trees	✓	-	-	✓	✓	-	-	-
5	Ma et al.	Short-Term Heavy Overload Forecasting of Public Transformers Based on Combined LSTM-XGBoost Model	✓	-	-	✓	✓	-	-	-
6	Ramesh et al.	Machine Learning Approach for Smart Distribution Transformers Load Monitoring and Management System	✓	-	-	✓	✓	-	-	-
7	Moldovan et al.	Prediction of Faults Location and Type in Electrical Cables Using Artificial Neural Network	✓	-	-	✓	-	✓	-	-
8	Xiang et al.	Flagging Implausible Inspection Reports of Distribution Transformers via Anomaly Detection	✓	-	-	✓	✓	-	-	-
9	Xiao et al.	The Research of Anomaly Detection Method for Transformer Oil Temperature Based on Hybrid Modelo Non-Supervised Learning and Decision Forests	✓	-	-	✓	✓	-	-	-
10	Manivannan	Power distribution system event classification using fuzzy logic	✓	-	-	✓	-	-	-	✓

Nota: Fuente propia

En resumen, la revisión de la literatura evidencia que, aunque los algoritmos de inteligencia artificial han avanzado en la detección de anomalías en sistemas eléctricos, persisten desafíos significativos. Los altos costos asociados con la implementación de

arquitecturas complejas y equipos de medición sofisticados limitan su aplicabilidad a gran escala. Además, la demora en la detección de las condiciones subestándares impacta significativamente en los costos asociados a su reparación y a los equipos necesarios para su detección. Existe la necesidad de desarrollar algoritmos eficientes y económicamente viables que utilicen variables básicas como la corriente y el voltaje en baja tensión. De modo que garanticen una detección de anomalías efectiva y una mejora en la continuidad del servicio eléctrico.

1.2. Identificación y Descripción del Problema de Estudio

La literatura muestra que la gestión del mantenimiento en las redes de distribución eléctrica es un tema esencial e importante para garantizar la confiabilidad del sistema (Mirhosseini et al., 2022)¹². Encuentra una relación estrecha entre la confiabilidad del sistema y la reducción de fallas y del tiempo de interrupción (Mirsaeedi et al., 2018)¹³. La existencia de fallas e interrupciones son generada por la no atención oportuna de anomalías que aparecen los componentes de la instalación eléctrica. Sin embargo, la vasta cantidad de componentes de las redes de distribución limitan la posibilidad de ejecutar mantenimiento en toda la red (Keynia et al., 2022)¹⁴. En ese contexto, el transformador de distribución eléctrica es considerado un equipo clave (Xiang et al., 2020)⁸ para asegurar la continuidad del servicio eléctrico.

En países como el Perú, existen organismos reguladores que supervisan los indicadores de confiabilidad mediante el uso de normativa que miden los valores máximos

¹² MIRHOSSEINI, M., HEYDARI, A., ASTIASO G., D., MANCINI, F., KEYNIA, F. (2022). Reliability based maintenance programming by a new index for electrical distribution system components ranking. *Optimization and Engineering*,23(4), 2315-2333. <https://doi.org/10.1007/s11081-022-09767-8>

¹³ MIRSAEEDI, H., FEREDUNIAN, A., MOHAMMADI-HOSSEININEJAD, S., LESANI, H. (2018). Electricity Distribution System Maintenance Budgeting: A Reliability-Centered Approach. *IEEE Transactions on Power Delivery*. 33(4), 1599-1610. <https://doi.org/10.1109/TPWRD.2017.2772202>

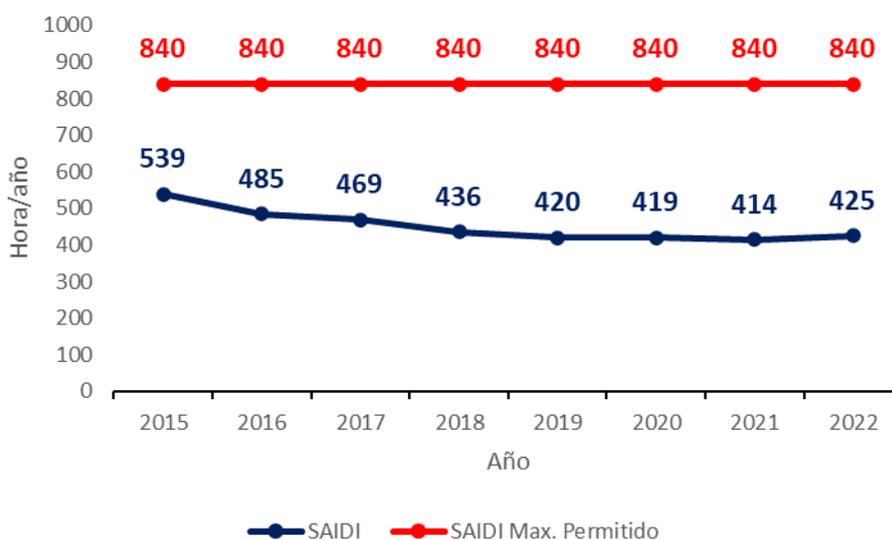
¹⁴ KEYNIA, F., MIHOSSEINI, M. HEYDARI, A., FEKIH, A. (2022). A budget allocation and programming-based RCM approach to improve the reliability of power distribution networks. *Energy Reports*. 8, 5591-5602. <https://doi.org/10.1016/j.egy.2022.04.029>

permitidos de SAIDI y SAIFI Misari et al. (2020)¹⁵. En el Perú, OSINERGMIN realiza esta labor basada en el Decreto Supremo N° 020-97-EM, que establece límites de 8 interrupciones anuales para el SAIFI y 840 minutos por año para el SAIDI. Aunque los estándares nacionales nos permiten controlar la confiabilidad del sistema, el objetivo es mejorar los niveles de confiabilidad hacia estándares internacionales, como los recomendados por la NERC¹⁶, que propone una falla cada 10 años.

A pesar de que la empresa de distribución cumple con los estándares nacionales, como muestra la Figura 1 y

Figura 2. Aún se aprecia que existen márgenes de mejora significativa para poder llegar a cumplir los estándares internacionales. Entonces, la implementación de tecnologías avanzadas, como la inteligencia artificial, se presenta como una alternativa de solución para aumentar la confiabilidad del sistema.

Figura 1. Evolución del SAIDI del 2015 al 2022 de una empresa de Distribución



Nota: Adaptación de la memoria integrada anual 2022, Enel Distribución Perú S.A.A.

¹⁵ MISARI, A., LEITE, J., PIASSON, D., MANTOVANI, J. (2020). Reliability-Centered Maintenance Task Planning for Overhead Electric Power Distribution Networks. Journal of Control, Automation and Electrical Systems, 31(5), 1278-1287. <https://doi.org/10.1007/s40313-020-00606-8>

¹⁶ NERC: North American Electric Reliability Corporation

Figura 2. Evolución del SAIFI del 2015 al 2022 de una empresa de Distribución



Nota: Adaptación de la memoria integrada anual 2022, Enel Distribución Perú S.A.A.

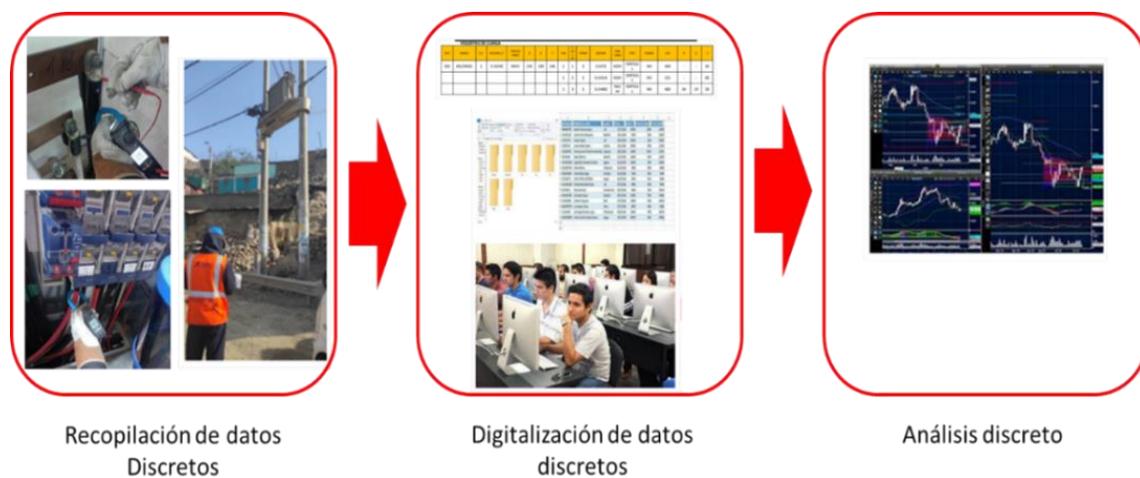
Sin embargo, existen varios obstáculos a nivel local que dificultan la integración efectiva de estas tecnologías. Entre los desafíos identificados para transformadores se tiene:

- Un primer problema que se detecta es que en Perú existe poca especialización profesional en analizar grandes conjuntos de datos.
- Un segundo problema es que se desconoce como estandarizar anomalías en transformadores de distribución eléctrica (transformador).
- Un tercer problema es que la ingeniería peruana tiene poca experiencia en desarrollar algoritmos de aprendizaje automático para detectar anomalías en un transformador de distribución eléctrica.
- Un cuarto problema es que se desconoce la rentabilidad de un proyecto de inteligencia artificial cuando una empresa de distribución eléctrica decida invertir en estas nuevas tecnologías.

Esto queda descrito en la Figura 3, donde se muestra como actualmente el personal técnico lleva a cabo mediciones precisas de variables eléctricas de corriente y voltaje en

los transformadores de distribución durante las actividades de termografía y mantenimiento en caliente. Estas mediciones se registran manualmente y se almacenan en una base de datos, que posteriormente se procesa en hojas de cálculo para su análisis. Este enfoque ha sido efectivo para mantener los indicadores SAIDI y SAIFI por debajo de los límites permitidos, de acuerdo a la normativa peruana. Sin embargo, es evidente que aún existe un margen significativo para continuar reduciendo estos indicadores. Es necesario explorar métodos avanzados para optimizar aún más el desempeño de la red eléctrica, asegurando así un suministro más confiable y eficiente.

Figura 3. Procesamiento de las mediciones puntuales de corriente y voltaje.



Nota: Fuente propia

El tiempo que transcurre entre que el personal técnico de campo envía la información al inspector es de aproximadamente una semana. Esto impacta de manera negativa en el tiempo requerido para la toma de decisiones por parte del inspector de la empresa de distribución. Las consecuencias de la no atención oportuna de una anomalía son la reducción en la venta de energía, reducción de la confiabilidad y aumento en los tiempos de atención. Por ello, es importante detectar a tiempo una sobrecarga para evitar la activación de la protección del transformador y/o pérdidas de aceite por la rotura de la rigidez dieléctrica. Además, una caída de voltaje menor a lo definido por la NTCSE impacta en el funcionamiento de los equipos por parte del cliente. También, la falta de corriente en

un transformador es indicador de una interrupción en los circuitos que alimentan a los clientes. Además, la falta de detección temprana impacta en la imagen de la empresa ya que los clientes hacen uso de diferentes medios sociales para comunicar este inconveniente.

1.3. Formulación del problema

1.3.1. Problema principal

¿De qué manera el aprendizaje automático permitirá detectar anomalías en un transformador de una empresa de distribución eléctrica?

1.3.2. Problemas secundarios

¿De qué manera el análisis de datos históricos permitirá la parametrización de anomalías en un transformador de una empresa de distribución eléctrica?

¿De qué manera la aplicación del algoritmo basado en el árbol de decisiones permitirá detectar anomalías en un transformador de una empresa de distribución eléctrica?

¿De qué manera la evaluación económica permitirá obtener indicadores de rentabilidad al aplicar aprendizaje automático en un transformador de una empresa de distribución eléctrica?

1.4. Justificación e importancia

El presente trabajo de investigación se justifica porque proporcionará una herramienta computacional avanzada para el procesamiento en tiempo real de las variables de corriente y voltaje en baja tensión de los transformadores de distribución. Esta solución optimizará la gestión de la información, reduciendo drásticamente los tiempos de procesamiento para los inspectores y permitiendo una atención más rápida y eficaz a las

anomalías detectadas. La implementación de esta herramienta es crucial para mejorar la continuidad y confiabilidad del servicio eléctrico, garantizando una respuesta ágil y proactiva ante posibles fallas en el sistema.

Además, esta investigación tiene una justificación tecnológica sólida, ya que, como señala (Wang et al., 2022)², el avance de la tecnología y la economía ha permitido la integración de las redes eléctricas con la informática y las comunicaciones. Este progreso es crucial para mejorar la eficiencia operativa. Asimismo, en los trabajos de (Veerakumar et al., 2023)³, (Ramesh et al., 2022)⁶ y (dos Santos et al., 2018)¹⁷, se destaca el impacto positivo del uso de modelos de aprendizaje automático para procesar grandes volúmenes de datos y detectar anomalías de manera oportuna, logrando niveles de precisión superiores al 85%.

Además, tiene una justificación legal, tomando como referencia la Ley N° 31814¹⁸, cuyo objetivo es promover el aprovechamiento de las nuevas tecnologías de la información, la inteligencia artificial, el internet de las cosas y su uso en la industria 4.0. Esta ley es aplicable para el sector privado y público; para el beneficio de la sociedad y promover el desarrollo económico del país.

La importancia de este trabajo de investigación radica en su capacidad para transformar una empresa de distribución eléctrica, que actualmente utiliza equipos de medición puntual, en una entidad que emplea un modelo de aprendizaje automático para la detección proactiva de anomalías en transformadores de distribución. Este enfoque no solo optimiza la identificación y corrección de problemas, sino que también impulsa el

¹⁷ DOS SANTOS, G., DE AQUINO, R., LIRA, M. (2018). Thermography and artificial intelligence in transformer fault detection. *Electrical Engineering*, 100, 1317-1325. <https://doi.org/10.1007/s00202-017-0595-2>

¹⁸ Ley N° 31814 (2023). Ley que promueve el uso de la inteligencia artificial en favor del desarrollo económico y social del país. <https://busquedas.elperuano.pe/download/url/ley-que-promueve-el-uso-de-la-inteligencia-artificial-en-fav-ley-n-31814-2192926-1>

proceso de automatización en el análisis de información, mejorando la eficiencia operativa y fortaleciendo la confiabilidad del servicio eléctrico. Al adoptar estas tecnologías avanzadas, la empresa se posiciona a la vanguardia de la innovación en el sector.

1.5. Objetivos

1.5.1. Objetivo General

Detectar anomalías mediante el aprendizaje automático en un transformador de una empresa de distribución eléctrica.

1.5.2. Objetivos Específicos

La parametrización de anomalías mediante el análisis de datos históricos en un transformador de una empresa de distribución eléctrica.

Detectar anomalías mediante la aplicación del algoritmo basado en el árbol de decisiones en un transformador de una empresa de distribución eléctrica.

Obtener los indicadores de rentabilidad mediante la evaluación económica al aplicar aprendizaje automático en un transformador de una empresa de distribución eléctrica.

1.6. Hipótesis

1.6.1. Hipótesis General

El aprendizaje automático permite detectar anomalías en un transformador de una empresa de distribución eléctrica.

1.6.2. Hipótesis Específicas

Mediante el análisis de datos históricos se logra la parametrización de anomalías en un transformador de una empresa de distribución eléctrica.

Mediante la aplicación del algoritmo basado en el árbol de decisiones se logra detectar anomalías en un transformador de una empresa de distribución eléctrica.

Mediante la evaluación económica se logra obtener los indicadores de rentabilidad al aplicar aprendizaje automático en un transformador de una empresa de distribución eléctrica.

1.7. Variables y Operacionalización de Variables

En la Tabla 5, se pasa a describir como se va a operacionalizar las variables.

Tabla 5. Operacionalización de las variables

VARIABLE	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	OPERACIONALIZACIÓN		ESCALA DE MEDICIÓN
			DIMENSIONES	INDICADORES	
Aprendizaje automático	Conceptualmente, se refiere al conjunto de herramientas utilizadas que permita su implementación para un transformador de una empresa de distribución eléctrica. Incluye aspectos como en el desarrollo del análisis de datos históricos, la construcción de un algoritmo basado en el árbol de decisiones y la evaluación económica	Operacionalmente, esta variable se medirá a través de dimensiones como el análisis de datos históricos (nominal), el algoritmo basado en el árbol de decisiones (nominal) y la rentabilidad del proyecto (nominal).	Análisis de datos históricos	Parametrización de anomalías	100%
				Etiquetado de datos	100%
			Algoritmo basado en árbol de decisiones	Entrenamiento y validación del algoritmo	100%
				Prueba del modelo	100%
Evaluación económica	Rentabilidad del proyecto	100%			
detectar anomalías	Conceptualmente, se refiere a las métricas utilizadas para garantizar el éxito del proyecto de implementación del aprendizaje automático para detectar anomalías en tiempo real. Incluye aspectos como la detección manual de datos anómalos, métricas de desempeño e indicadores de rentabilidad	Operacionalmente, esta variable se medirá a través de dimensiones como a detección manual de anomalías (ordinal), las métricas de desempeño (ordinal) y los indicadores de rentabilidad (ordinal)	Detección manual de datos anómalos	Cantidad de datos anómalos	Numérica
			Métricas de desempeño	Precision	0 a 1
				Recall	0 a 1
				Accuracy	0 a 1
				F1-Score	0 a 1
			Indicadores de rentabilidad	Cálculo del VAN	Soles
Cálculo del TIR	Porcentual				

Nota: Fuente propia

1.8. Diseño Metodológico

1.8.1. Unidades de Análisis

La unidad de análisis es el transformador de distribución escogido para el presente trabajo de investigación. La selección de la muestra es no probabilística, debido a que está asociada de manera directa con el alcance de la investigación. Se va a tomar como muestra los valores de corriente y voltaje medidos en los meses de mayo a julio de 2024.

1.8.2. Fuentes de Información

Las fuentes de información empleadas son artículos científicos, tesis de posgrado, libros especializados. Así como datos obtenidos de mediciones eléctricas realizadas en campo en transformadores, específicamente corriente y voltaje.

1.8.3. Tipo de Investigación

La investigación según su finalidad es del tipo aplicada, ya que vamos a hacer uso de un modelo de aprendizaje automático para la detección de anomalías. El diseño de la investigación es no experimental, ya que se va a ejecutar un entrenamiento supervisado de la arquitectura de Árbol de decisiones (Decision Tree). Tiene un enfoque cuantitativo ya que luego del entrenamiento se van a obtener métricas que van a medir la eficiencia del mismo. Además, va a tener un proceso secuencial, ya que va a iniciar con la recolección de los datos de corriente y voltaje, los cuales van a ser etiquetados. Luego de ello van a servir de insumo para el entrenamiento del modelo escogido y posterior se va a evaluar la viabilidad de la implementación de esta nueva tecnología. Tiene un alcance analítico explicativo ya que el propósito principal de la investigación es evaluar y analizar las relaciones causa – efecto entre las variables “modelo de aprendizaje automático” y “detectar las anomalías”.

1.8.4. Selección de 01 transformador para la instalación del equipo DTM¹⁹

Del universo de transformadores ubicados en el centro histórico de Lima se seleccionó el transformador ubicado en el pasaje Olaya por conveniencia. Por las facilidades técnicas existentes tanto en la disponibilidad de personal técnico como en la menor afectación al SAIDI para su instalación.

1.8.5. Técnicas e Instrumentos de recolección de Datos

Se va a hacer uso de un dispositivo de medición de parámetros eléctricos (DTM¹⁹) para la recolección de los valores de corriente y voltaje en baja tensión.

1.8.6. Técnicas e instrumentos de Análisis y Procesamiento de Datos

Los datos recolectados van a ser procesados en un algoritmo de aprendizaje automático a desarrollarse en el lenguaje Python y el flujo de caja del proyecto se desarrollará en una hoja de cálculo.

¹⁹ DTM: Distribution Transformer Monitoring

1.9. Contribuciones

Las contribuciones destacadas del presente trabajo de investigación en la gestión de mantenimiento son:

- a. Se ha generado un conjunto de datos basado en mediciones eléctricas realizadas en campo, lo que permitirá a futuras investigaciones aplicar y comparar diversas técnicas de aprendizaje automático (*Machine Learning*).
- b. Se ha aplicado aprendizaje automático para la detección de anomalías en la gestión del mantenimiento eléctrico de transformadores de distribución, reduciendo el tiempo en la identificación temprana de fallas.
- c. Se ha demostrado la rentabilidad de estas tecnologías, resaltando su capacidad para reducir costos y mejorar la eficiencia operativa en la gestión de activos eléctricos

Estas contribuciones permiten impulsar la modernización del sector eléctrico peruano mediante enfoques tecnológicos innovadores que promueven mayor confiabilidad y sostenibilidad en las operaciones de mantenimiento.

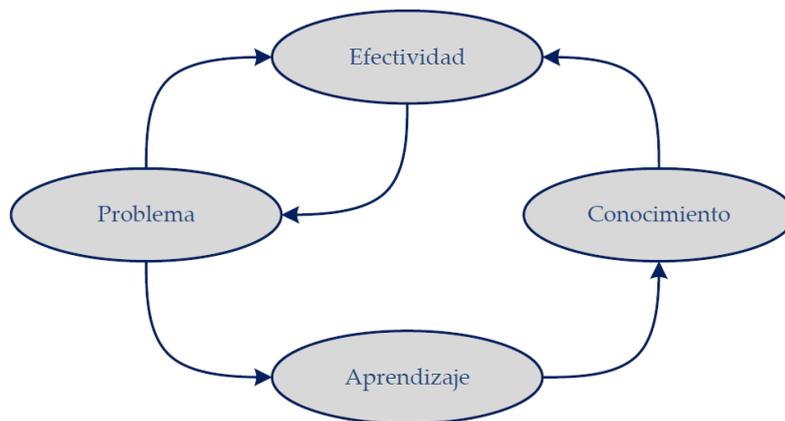
CAPÍTULO II. MARCO TEÓRICO

2.1. Bases Teóricas

2.1.1. Aprendizaje automático (Machine Learning)

Berzal (2018)²⁰ describe que la capacidad de un sistema informático para resolver problemas mediante algoritmos de aprendizaje se basa en procesar un conjunto de datos y realizar iteraciones heurísticas. Estas iteraciones permiten al sistema generar autónomamente el algoritmo adecuado para abordar el problema planteado. Durante este proceso, se llevan a cabo evaluaciones continuas que optimizan el aprendizaje y refinan el conocimiento adquirido hasta alcanzar un nivel de eficiencia predefinido, como se muestra en la Figura 4. Por otro lado, Amr (2020)²¹ señala que los métodos de aprendizaje se dividen en tres categorías: supervisado, semi – supervisado y no supervisado, cada uno adaptado a diferentes escenarios según la cantidad de datos etiquetados disponibles y el grado de intervención humana requerido para guiar el proceso.

Figura 4. Interacción que se genera en un aprendizaje automático



Nota: (Berzal, 2018)²⁰ figura 2, pág. 27

²⁰ BERZAL F. (2018). Redes Neuronales & Deep Learning. Editorial Universidad de Granada, EUG. España.

²¹ AMR, T. (2020). Hands-On Machine Learning with Scikit-Learn and Scientific Python Toolkits: A Practical Guide to Implementing Supervised and Unsupervised Machine Learning Algorithms in Python

2.1.2. Aprendizaje Automático Supervisado

Chatterjee (2022)²² sostiene que un algoritmo supervisado emplea datos etiquetados para construir una función que aproxima los resultados esperados. Los datos de salida generados se comparan con los valores previstos, y en función del error obtenido, el algoritmo ajusta el modelo y continúa con nuevas iteraciones para mejorar su precisión. De manera similar, Bishop & Bishop (2024)²³ explican que, durante el entrenamiento del modelo, se ajustan los hiperparámetros y las características en las capas ocultas, comparándolos constantemente con los valores reales para afinar el modelo y optimizar su rendimiento. Este proceso busca que el modelo, una vez entrenado, pueda manejar eficientemente nuevos datos. Berzal (2018)²⁰ indica que el objetivo del aprendizaje supervisado es entrenar un modelo para que se ajuste a los valores de la función objetivo, permitiéndole luego predecir resultados con datos de entrada no vistos previamente. Finalmente, Amr (2020)²¹ agrega que el aprendizaje supervisado se clasifica en dos categorías principales: clasificación y regresión, cada una dirigida a resolver distintos tipos de problemas.

2.1.3. Tipo de Aprendizaje Automático Supervisado – Clasificación

Phillips (2020)²⁴ indica que este método se utiliza para entrenar un algoritmo, empleando etiquetas que permiten identificarlas y categorizarlas según las categorías previamente definidas en el modelo. Ayyadevara (2018)²⁵ señala que este enfoque predice eventos con pocos resultados posibles, lo que permite al algoritmo identificar y anticipar

²² CHATTERJEE, I. (2022). Machine Learning and Its Application: A Quick Guide for Beginners. Betham Science Publishers Ltd. Busan. South Korea.

²³ BISHOP, C., BISHOP, H. (2024). Foundations and Concepts Deep Learning. Springer, Switzerland.

²⁴ PHILIPS, D. (2020). Jump on the way to the future, discover artificial intelligence and data science. Maximize your business in the modern world mastering Deep Learning, Python and algorithms.

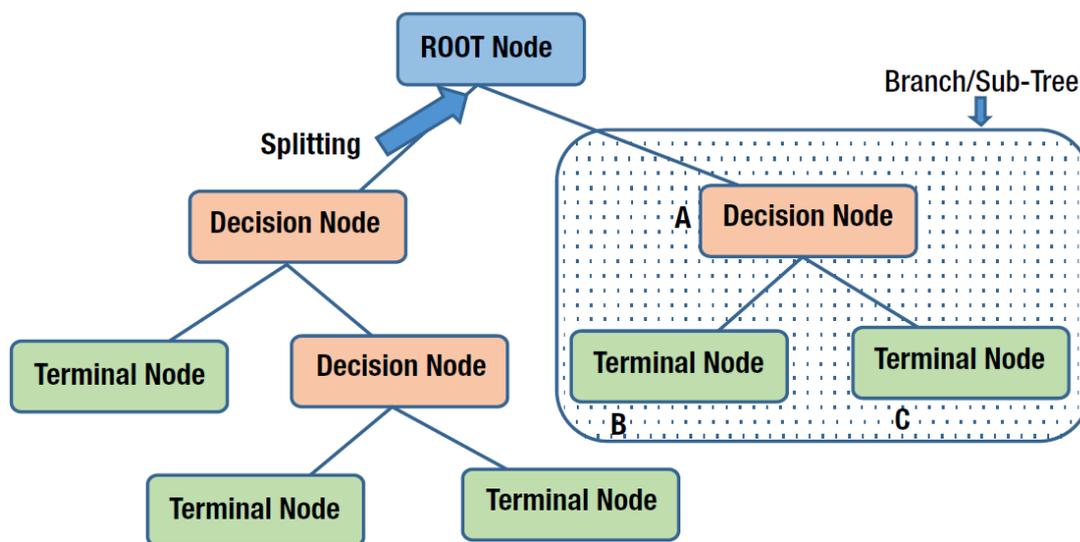
²⁵ AYYADEVARA, V. (2018). Pro Machine Learning Algorithms: A Hands-On Approach to Implementing Algorithms in Python and R. Apress Media LLC. Hyderabad, Andhra Pradesh, India.

resultados específicos con mayor precisión y confiabilidad. Müller & Guido (2017)²⁶ explican que el objetivo de esta técnica es predecir una etiqueta de una lista predefinida, lo cual puede involucrar una clasificación binaria o multiclase según el caso.

2.1.4. Árbol de Decisión (Decision Tree)

Phillips (2020)²⁴ define este método como una forma de aprendizaje supervisado en la que los datos se dividen repetidamente, basándose en un parámetro particular del conjunto de datos. Ayyadevara (2018)²⁵ señala que este algoritmo clasifica eventos o predice valores de salida, basándose en un conjunto de reglas predefinidas, cuyos componentes principales se describen detalladamente en la Figura 5. (Müller & Guido, 2017)²⁶ explican que estos modelos son ampliamente utilizados para tareas de clasificación y regresión, donde el aprendizaje se realiza mediante jerarquías de decisión estructuradas.

Figura 5. Componentes de un árbol de decisiones



Nota: (Ayyadevara, 2018)²⁵ figure 4-2, pág. 73

²⁶ MÜLLER, A., GUIDO, S. (2017). Introduction to Machine Learning with Python: A guide for Data Scientists. O'Reilly Media. United State of America.

2.1.5. Métricas de aprobación del algoritmo Árbol de decisión

Ayyadevara (2018)²⁵ menciona que, en los algoritmos de clasificación, que forman parte del aprendizaje automático supervisado, es esencial evaluar las métricas de Precisión, Recall, Exactitud (Accuracy) y F1-Score. Estas métricas permiten medir la efectividad del modelo al comparar las predicciones realizadas con los valores reales analizando en términos de verdaderos positivos (cuando la predicción y el valor real son verdaderos) como de falsos positivos (cuando la predicción es verdadera, pero el valor real es falso). Cada métrica ofrece una perspectiva diferente sobre el rendimiento del modelo, ayudando a determinar su capacidad para hacer predicciones precisas y confiables.

2.1.6. Anomalías en redes de distribución

El Organismo Supervisor de la Inversión en Energía y Minería (2009)²⁷ define la anomalía como una condición en la que un equipo o componente no cumple con las especificaciones técnicas vigentes. Manivannan et al. (2012)¹⁰ expresan que los sistemas eléctricos de distribución son vulnerables a sobrecargas, envejecimiento y defectos, lo cual se refleja en la detección de anomalías que afectan la eficiencia operativa.

2.1.7. Transformador de distribución

Yebra (2009)²⁸ expresa que es el equipo eléctrico con alta importancia dentro de la red eléctrica de distribución; cuyo uso, es reducir los niveles de tensión para los usuarios finales. Kulkarni & Khaparde (2013)²⁹ manifiestan que reducen los niveles de tensión a un voltaje de uso doméstico y comercial.

²⁷ Organismo Supervisor de la Inversión en Energía y Minería (2009). Resolución de Consejo Directivo OSINERGMIN N° 228-2009-OS/CD. Lima, Lima, Perú. <https://www.osinergmin.gob.pe/newweb/uploads/Publico/RCD.228.2009.OS.CD.pdf>

²⁸ YEBRA JUAN (2009). Sistemas Eléctricos de Distribución. Primera Edición. Reverté Ediciones. México.

²⁹ KULKARNI, S., KHAPARDE, S. (2013). Transformer Engineering: Design, Technology, and Diagnostics. Segunda Edición. CRC Press. Boca Raton.

2.1.8. Modos de Falla y Análisis de Efectos (AMFE)

Acuña (2022)³⁰ manifiesta que es también llamado *Failure Mode Effect Analysis (FMEA)*, es una técnica que permite identificar las causas y efectos de una falla en un equipo o proceso; permitiendo al usuario realizar la identificación, clasificación y prevención de fallas mediante este análisis, plasmando dicha información en un cuadro, tal como se identifica en la Tabla 6. Para utilizar esta herramienta se inicia con la determinación de la función del equipo o proceso, luego se define el modo de falla potencial, posterior a ello, los efectos de falla potencial y las causas que la generan y finalmente los modos de control actuales y las acciones recomendadas de mejora.

Tabla 6. Plantilla para el AMFE

Función	Modo de falla potencial	Efectos de la falla potencial	Causas	Modo de controles actuales	Acciones recomendadas
---------	-------------------------	-------------------------------	--------	----------------------------	-----------------------

Nota: (Acuña, 2022)³⁰ adaptación de figura 6.8, pág. 272

2.1.9. Criterios de aprobación del proyecto

Sapag & Sapag Reinaldo (2012)³¹ expresan que existen dos criterios para evaluar un proyecto, que son el criterio del valor actual neto (VAN) y el criterio de la tasa interna de retorno (TIR). Entonces, para el presente trabajo de investigación vamos a tener otros dos criterios de aprobación que serán el VAN y el TIR. En donde, para el VAN se va a usar una tasa de descuento del 12% (Apuntes de clase) y para un periodo de 05 años.

³⁰ ACUÑA, J. (2022). Ingeniería de Confiabilidad. Segunda Edición. Editorial Tecnológica de Costa Rica. Cartago, Costa Rica.

³¹ SAPAG, N., SAPAG, R. (2012). Preparación y Evaluación de Proyectos. McGraw-Hill Interamericana. México.

2.2. Marco Conceptual

2.2.1. Aprendizaje Automático

Es la capacidad de un sistema computacional para aplicar algún algoritmo a un conjunto de datos para resolver un problema, realizando esta operación de manera iterativa hasta que los resultados se acerquen a la función objetivo.

2.2.2. Aprendizaje Automático Supervisado

Es aquel algoritmo que hace uso de un conjunto de datos etiquetados para realizar los procesos de iteración y obtener una función aproximada que permita ingresar nuevos datos; con ello se logra que cuando se cargue nuevos datos el valor predicho sea lo más cercano al valor real.

2.2.3. Tipo de Aprendizaje Automático Supervisado – Clasificación

Es un tipo de aprendizaje supervisado que luego del entrenamiento la variable de salida se compara con la etiqueta de salida predefinida.

2.2.4. Árbol de decisiones

Es un algoritmo de aprendizaje supervisado donde permite clasificar un evento o predecir la variable de salida y compararlo con respecto a las etiquetas predefinidas.

2.2.5. Transformador de distribución

O también llamado simplemente transformador para el presente trabajo de investigación; es el equipo clave dentro de un sistema de distribución, al cual se le va a instalar un equipo de monitoreo para recopilar y enviar a la nube los valores de corriente y voltaje en baja tensión.

2.2.6. Anomalías en un transformador de distribución

Es toda condición subestándar en la cual está trabajando el transformador y que en el corto o mediano plazo pueda generar una interrupción en el suministro de energía eléctrica a los usuarios finales.

2.2.7. Detectar anomalías

Es encontrar datos registrados de corriente y voltaje anómalos dentro de un conjunto de datos que cumplan una de las tres anomalías definidas para el presente trabajo de investigación.

2.3. Marco Legal

2.3.1. Calidad de Producto – Tensión

De acuerdo a la Norma Técnica de Calidad del Servicio Eléctrico (NTCSE) publicada en el Decreto Supremo DS020-97 (y sus actualizaciones), la tolerancia admitida para los valores nominales de tensión es de hasta $\pm 5,0\%$.

2.3.2. Calidad de Suministro – Interrupciones

De acuerdo a la Norma Técnica de Calidad del Servicio Eléctrico (NTCSE) publicada en el Decreto Supremo DS020-97 (y sus actualizaciones), se define dos indicadores:

Número Total de Interrupciones por Cliente por Semestre (N), que se calcula por el número de interrupciones sucedidas en un semestre

Duración Total Ponderada de Interrupciones por Cliente (D), es la sumatoria de las duraciones ponderadas de las interrupciones sucedidas en el semestre.

CAPÍTULO III. METODOLOGÍA

3.1. Introducción

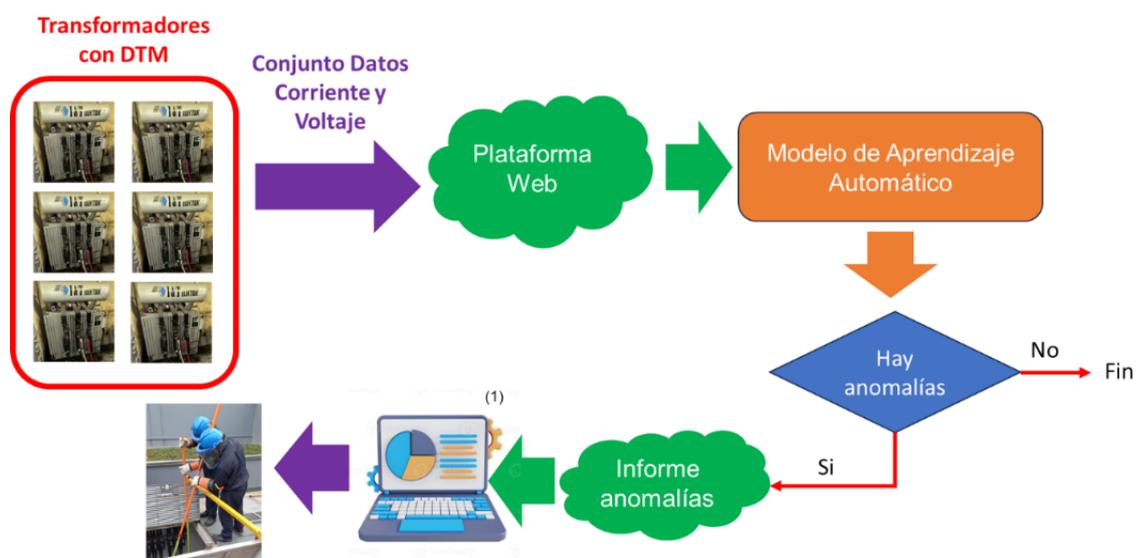
Un transformador de distribución desempeña un papel esencial al disminuir los niveles de tensión, como en el caso de Lima Norte, donde convierte 10,000 voltios en 220 voltios, garantizando así que los clientes de baja tensión reciban la energía necesaria. Este tipo de transformador se encuentra ubicado en un espacio denominado "Subestación Eléctrica de Distribución" (SED). De acuerdo con las regulaciones de OSINERGMIN, el sistema eléctrico debe operar de manera constante y sin interrupciones. Cualquier falla que supere los límites establecidos puede resultar en sanciones para la empresa distribuidora, a menos que se implementen mejoras o actualizaciones en la calidad o tecnología del sistema eléctrico. Por esta razón, las actividades de mantenimiento que se pueden realizar en el transformador son limitadas, restringiéndose principalmente a la limpieza de su exterior y la medición de parámetros críticos como corriente y voltaje, para garantizar su correcto funcionamiento.

Para gestionar estos desafíos de manera efectiva, es fundamental instalar equipos avanzados en cada transformador de la empresa de distribución eléctrica. Estos dispositivos deben registrar y transmitir información en tiempo real a la nube, donde un algoritmo de aprendizaje automático analizará los datos para detectar anomalías de forma inmediata. Ante la detección de irregularidades, el sistema debe notificar al supervisor responsable de manera instantánea, permitiendo una respuesta rápida y eficiente. Además, es crucial evaluar la rentabilidad del proyecto para garantizar que su implementación sea tanto técnicamente viable como económicamente beneficiosa para la empresa. Este enfoque integral no solo optimizará la detección y respuesta ante fallas, sino que también reforzará la confiabilidad y continuidad del servicio eléctrico, preparándose mejor para enfrentar los desafíos operativos y manteniendo altos estándares de eficiencia.

La implementación de estas tecnologías es, por tanto, un paso estratégico necesario para modernizar el mantenimiento en la red de distribución eléctrica.

Como se ilustra en la Figura 6, la solución propuesta es aplicar un modelo de aprendizaje automático para detectar anomalías en los transformadores de una empresa de distribución eléctrica. Este enfoque es crucial para enfrentar los desafíos actuales, permitiendo la identificación proactiva y precisa de fallas antes de que se conviertan en problemas graves. La implementación de este modelo optimizará las operaciones de mantenimiento y fortalecerá la continuidad y confiabilidad del servicio eléctrico, aspectos fundamentales para la empresa y sus clientes. La adopción de tecnologías avanzadas como el aprendizaje automático es esencial para garantizar que la red de distribución funcione de manera eficiente y segura. Esta solución no solo aborda las necesidades inmediatas, sino que también posiciona a la empresa en la vanguardia de la innovación tecnológica en el sector eléctrico, asegurando su competitividad y sostenibilidad a largo plazo.

Figura 6. Nueva forma de trabajar aplicando aprendizaje automático

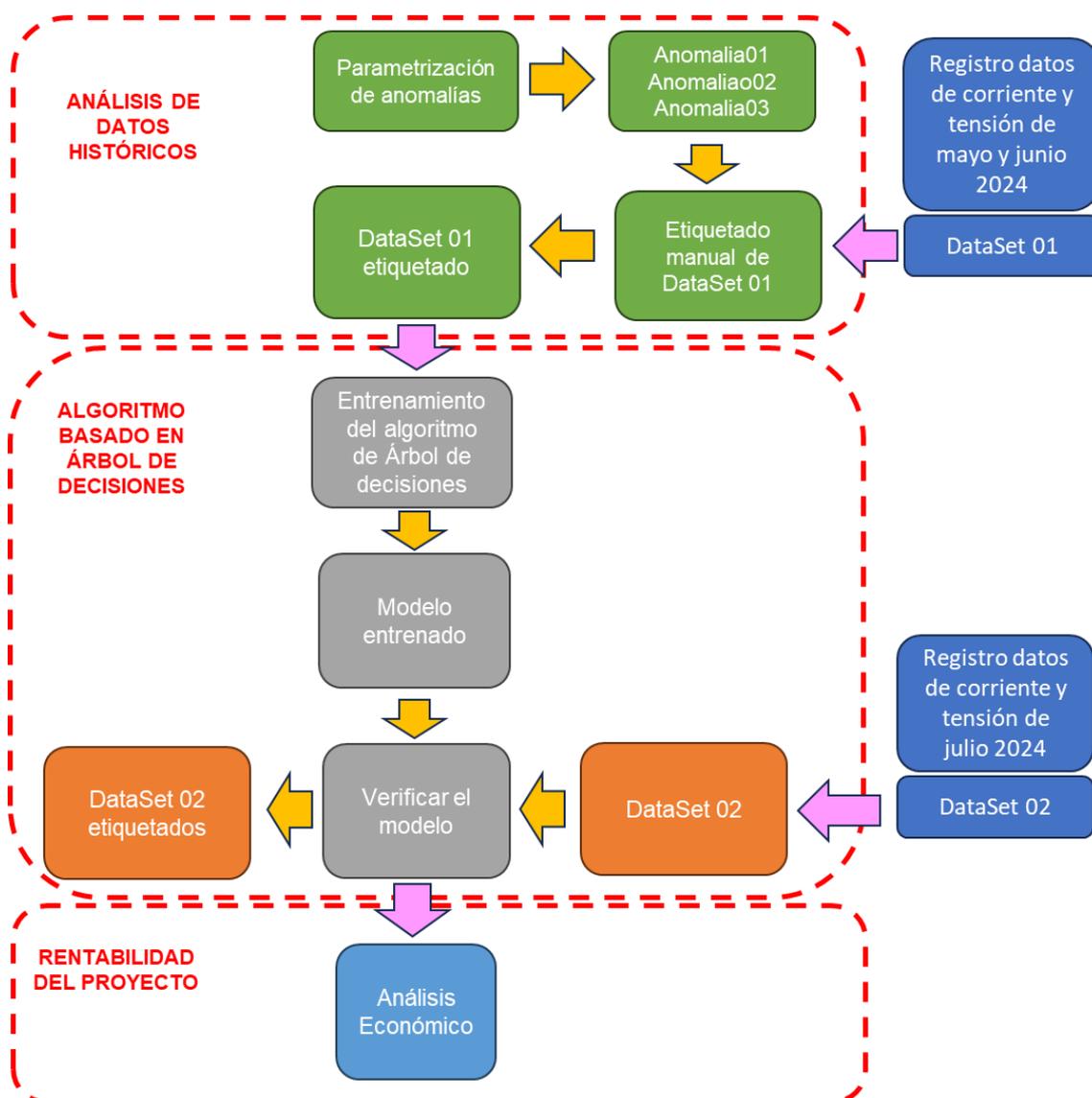


(1) <https://es.vecteezy.com/png/18245113-datos-de-informe-de-computadora-portatil-de-ilustracion-3d>

Nota: Fuente propia

La metodología propuesta se centra en tres objetivos claves, tal como se muestra en la Figura 7, el primero es realizar el análisis de datos históricos, luego desarrollar el algoritmo basado en árbol de decisiones y finalmente evaluar la rentabilidad del proyecto. Para el desarrollo de la metodología se va a utilizar dos conjuntos de datos llamados “DataSet 01” y “DataSet 02”. DataSet 01 es el conjunto de datos conformador por el registro de datos de corriente y voltaje en baja tensión de mayo y junio de 2024. DataSet 02 es el conjunto de datos conformador por el registro de datos de corriente y voltaje en baja tensión de julio de 2024.

Figura 7. Diagrama de flujo de la metodología propuesta



Nota: Fuente propia

Con respecto al análisis de datos históricos, el desarrollo de la metodología inicia con la parametrización de anomalías. En este estudio se propone identificar 03 anomalías asociadas al modo de fallo de un transformador que afecta de manera directa al suministro y calidad del servicio eléctrico a los clientes. Luego de conocer cuáles son estas tres anomalías procedemos a etiquetar de manera manual el DataSet 01. En donde se identifica qué tipo de anomalía está asociada o si no tiene anomalía, que en este caso se usará la etiqueta “normal”.

Con respecto al desarrollo del algoritmo basado en árbol de decisiones, hacemos uso de datos etiquetados manualmente del DataSet 01 para el entrenamiento y prueba del modelo aplicando la validación cruzada estratificada y revisando las métricas de desempeño se verifica la eficiencia del modelo entrenado. Posterior a ello, el modelo entrenado se utiliza para verificar si puede detectar anomalías en otro conjunto de datos que no haya sido utilizado en el entrenamiento, para ello, se utiliza el conjunto de datos DataSet 02.

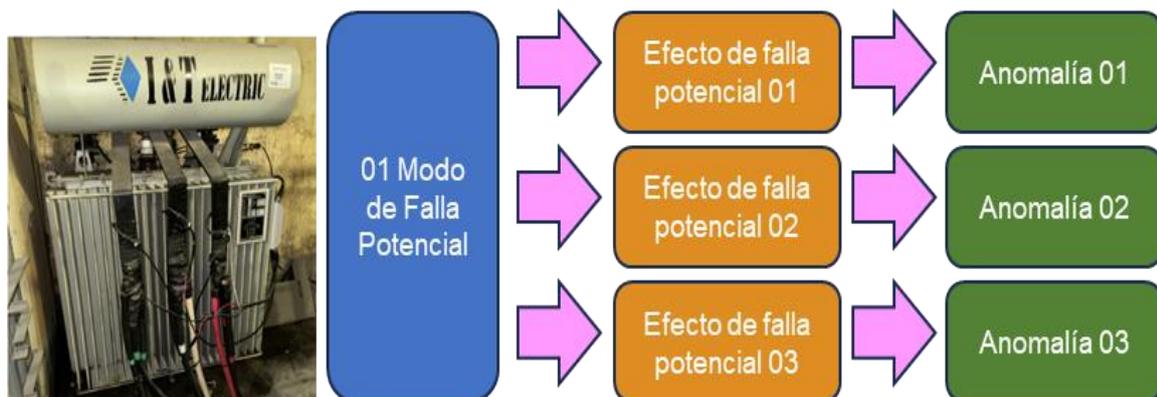
Con respecto a la evaluación económica del proyecto, se verifica si los indicadores financieros cumplen con los criterios de rentabilidad. Se realizará mediante un flujo de caja proyectado a cinco años. Considerando una inversión inicial para la instalación de 20 DTM, con una tasa de descuento del 12%. En el flujo de caja, se registrará el costo de inversión en el año cero y los ahorros generados a partir del año 1 hasta el año 5. Estos ahorros incluyen la reducción en costos de interrupciones del servicio, considerando la cantidad de clientes afectados, el tiempo de interrupción, la cantidad de eventos y el costo promedio de la energía. Además, se incluirá la depreciación de los equipos y el pago del impuesto a la renta del 29.5%. Los indicadores utilizados para evaluar la rentabilidad del proyecto son el Valor Actual Neto (VAN), que se calcula restando la inversión inicial de la suma de los valores presentes de los ahorros futuros, y la Tasa Interna de Retorno (TIR), que se obtiene variando la tasa de descuento hasta que el VAN sea cero. Estos cálculos permitirán determinar la viabilidad económica del proyecto y su retorno sobre la inversión.

3.2. Análisis de datos históricos

3.2.1. Parametrización de anomalías

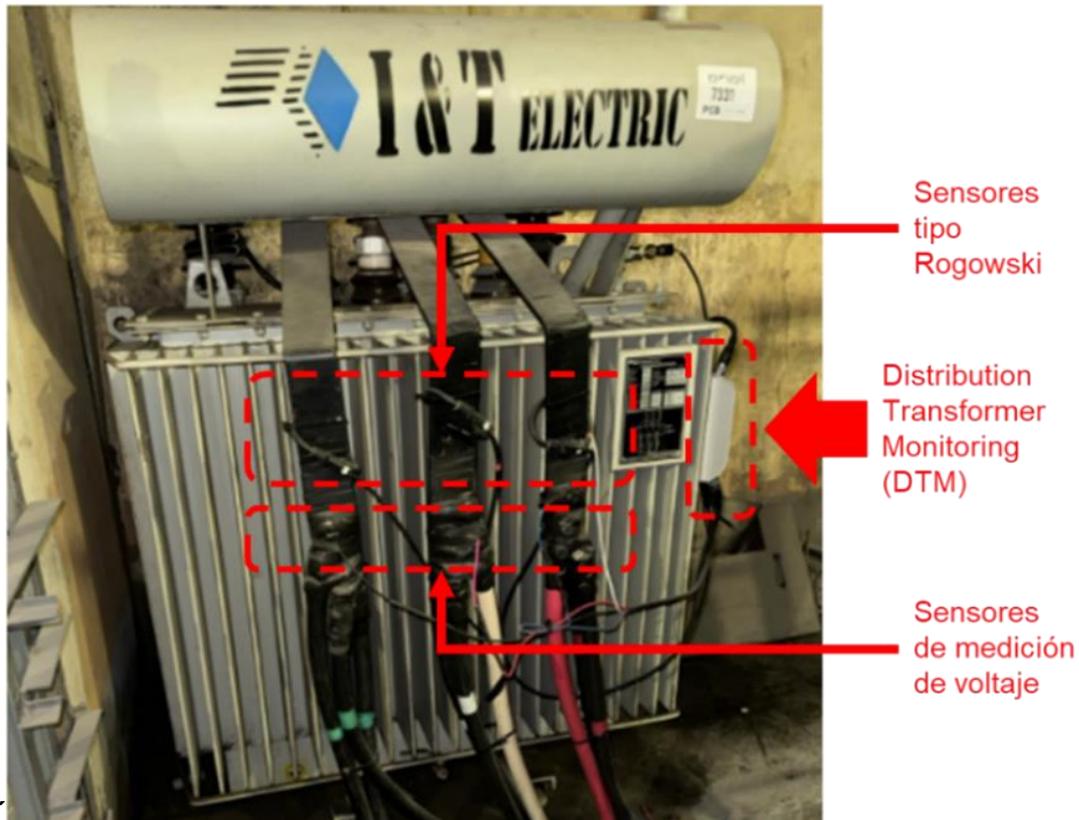
Las anomalías a parametrizar se determinarán a partir del modo de efecto de falla, lo que permitirá identificar un modo de falla potencial. Este modo de falla potencial estará asociado a varios efectos posibles. En esta investigación se plantean tres efectos de falla específicos: interrupción del servicio eléctrico, aumento de la temperatura en el transformador e incumplimiento de la NTCSE por caída de voltaje. Para cada uno de estos efectos, se asociará una causa, que se considerará como una anomalía, tal como se ilustra en la Figura 8. Estas anomalías se identificarán mediante el análisis de datos de voltaje y corriente en baja tensión, medidos por el equipo de monitoreo de transformadores de distribución (DTM). El equipo DTM permitirá registrar los valores de corriente y voltaje en baja tensión durante los meses de mayo, junio y julio de 2024, como se muestra en la Figura 9. Este enfoque garantiza que las anomalías sean detectadas y analizadas con precisión, facilitando la identificación de patrones críticos en el funcionamiento del transformador.

Figura 8. Diagrama de flujo para identificar anomalías



Nota: Fuente propia

Figura 9. DTM instalado en un transformador



Nota: Fuente propia

3.2.2. Etiquetado de datos

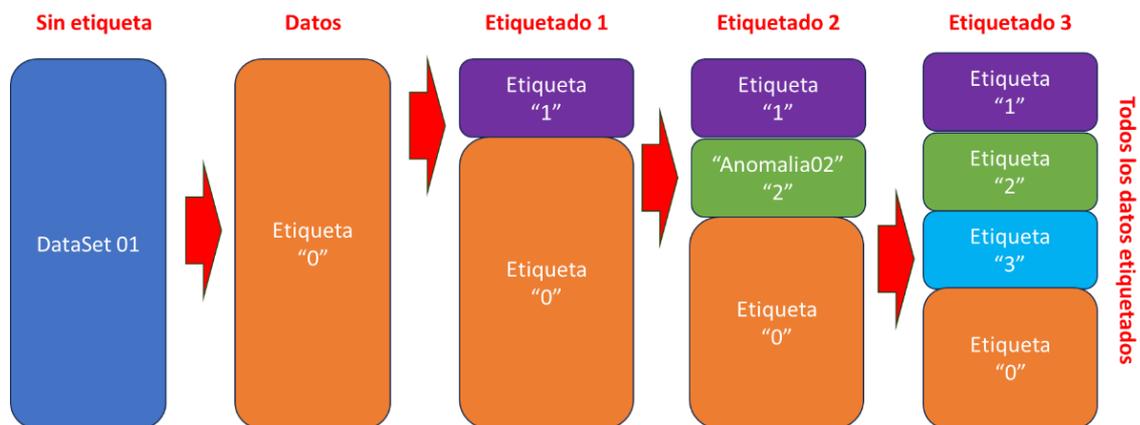
El "DataSet 01" será etiquetado para identificar que datos son anómalos o normales; esto se realiza usando lo descrito en la Tabla 7. Estos criterios permitirán clasificar de manera precisa cada fila de datos, asegurando que las anomalías se identifiquen correctamente y que los datos normales se diferencien adecuadamente. Este proceso es fundamental para preparar los datos de cara al análisis posterior, facilitando la implementación de modelos de aprendizaje automático que detecten y analicen eficientemente las condiciones operativas del transformador.

Tabla 7. Criterios para etiquetado de datos con anomalía

ID	Anomalía	Código	Etiqueta	Criterio
1	Cortocircuito externo	Anomalia01	01	Valor de Corriente igual a cero
2	Sobrecarga	Anomalia02	02	Valor de corriente mayor a 1653.36 A
3	Caída de voltaje	Anomalia03	03	Valor de Voltaje fase a tierra menor a 120.66 V

Nota: Fuente propia

Figura 10. Diagrama de flujo del código de etiquetado de datos



Nota: Fuente propia

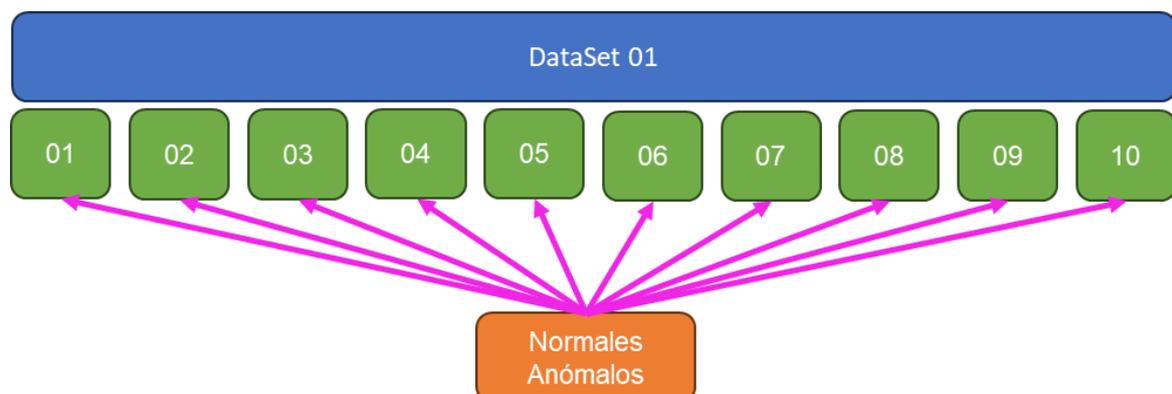
Tal como se muestra en la Figura 10, el proceso de etiquetado comienza asignando a todos los datos la etiqueta "0". Luego, se identifican los datos donde la corriente en las tres fases es igual a cero. A estos datos se les asigna la etiqueta "1". A continuación, se seleccionan los datos que aún tienen el código "Normal" y se identifican aquellos con corriente máxima en las tres fases mayor a 1653.37 A. Estos datos reciben la etiqueta "2". Finalmente, se seleccionan los datos restantes con el código "Normal" y se identifican los que tienen voltaje mínimo en las tres fases menor a 120.66 V. Estos datos reciben la etiqueta "3". Todo este diagrama de flujo de desarrolló usando Python para el etiquetado manual de los datos de entrenamiento y prueba.

3.3. Entrenamiento del modelo basado en el árbol de Decisiones

3.3.1. Preparación de datos

El conjunto de datos DataSet 01 tendrá como función objetivo la columna “etiqueta” que identifica que datos son normales o anómalos y que tipo de anomalía es. Además, se aplica una conversión de datos a formato numérico para eliminar cualquier error que pueda aparecer cuando se genera el DataSet y los datos tengan otro tipo de formato. Luego los datos numéricos son estandarizados para evitar que algunas características sean dominantes sobre otras debido a su escala. Para la división de datos para el entrenamiento y prueba del modelo se va a usar la validación cruzada estratificada 10 veces. Esta técnica realiza la evaluación del modelo del árbol de decisiones 10 veces, para esto divide los datos en 10 partes y en cada evaluación escoge 09 de ellos para el entrenamiento y el restante para la prueba. Esto lo realiza para que cada parte pase a ser prueba dentro del modelo y los 09 restantes pasen a ser entrenamiento del modelo. La estratificación garantiza que cuando se combinen los datos de manera aleatoria para generar las 10 partes exista datos de cada clase dentro de cada subconjunto, tal como se muestra en la Figura 11.

Figura 11. Subconjunto de datos que contienen datos normales y anómalos



Nota: Fuente Propia

3.3.2. Entrenamiento del algoritmo

Como estamos usando validación cruzada estratificada 10 veces, en la primera iteración el algoritmo va a tomar 09 subconjunto para el entrenamiento y el restante para la prueba del modelo. En la segunda iteración escoge otros 09 subconjuntos para el entrenamiento y el restante para la prueba del modelo, teniendo en cuenta que el subconjunto de prueba no se repita. Y esta operación la realiza 10 veces, tal como se muestra en la Figura 12. Además, el algoritmo del aprendizaje automático escogido es el árbol de decisiones.

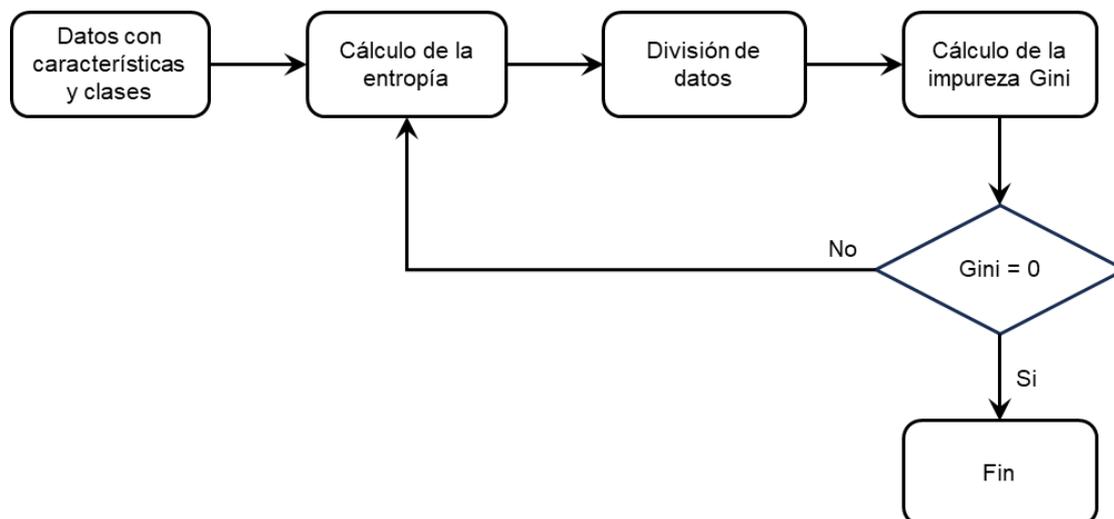
Figura 12. Iteraciones del conjunto de datos para el entrenamiento y prueba



Fuente: Propia

Tal como se muestra en la Figura 13, el algoritmo de árbol de decisión comienza con los datos de entrenamiento. Primero, se calcula la entropía de cada característica, lo que proporciona una medida de la incertidumbre dentro de las distribuciones de clases asociadas a esas características. La entropía se determina utilizando la Ecuación 1, que permite identificar cuál característica ofrece la mayor reducción de incertidumbre cuando se utiliza como nodo raíz. La característica seleccionada actúa como el punto de partida en la estructura del árbol, donde su elección maximiza la ganancia de información, un principio central en la construcción de árboles de decisión eficientes y precisos en aprendizaje automático.

Figura 13. Procesos del algoritmo de árbol de decisiones



Nota: Fuente Propia

Con la característica del nodo raíz seleccionada, el siguiente paso es calcular el índice de impureza Gini utilizando la Ecuación 2. Este cálculo permite dividir los datos en subconjuntos, asignándolos a nodos hijos (izquierdo y derecho) de manera que la impureza en estos subconjuntos sea mínima. El proceso asegura que cada decisión en el árbol contribuya a una clasificación más precisa. Esta metodología es crítica en la creación de árboles de decisión robustos, ya que cada división del árbol busca reducir la variabilidad dentro de los nodos, optimizando así la predicción de clases. El uso del índice Gini es una

práctica común en aprendizaje automático para mejorar la eficacia de los modelos de clasificación.

El algoritmo sigue evaluando la impureza Gini en cada nodo intermedio. Si en un nodo la impureza Gini es cero, significa que todos los datos en ese nodo pertenecen a una única clase, y el nodo se convierte en una hoja, terminando así la partición en esa rama del árbol. Si la impureza no es cero, el proceso se repite: se recalcula la entropía para las características restantes, y se vuelve a calcular el índice Gini para determinar cómo dividir los datos en los nodos siguientes. Este proceso iterativo es fundamental en aprendizaje automático, donde la construcción de árboles de decisión requiere decisiones basadas en la reducción progresiva de la impureza hasta que el árbol esté completamente desarrollado y optimizado

Ecuación 1. Cálculo de la incertidumbre en un árbol de decisiones clasificadorio

$$-p \log_2 p - q \log_2 q$$

Nota: (Ayyadevara, 2018)²⁵ pág. 75

Ecuación 2. Cálculo de la impureza Gini

$$1 - (p^2 + q^2)$$

Nota: (Ayyadevara, 2018)²⁵ pág. 79

3.3.3. Métricas de desempeño

Como el modelo ha realizado 10 iteraciones, entonces se va a contar con 10 valores para las métricas de desempeño y para calcular cada una de ellas se va a calcular el valor promedio de los 10 obtenidos para cada métrica. Para el presente trabajo de investigación se va a tomar cuatro variables para el cálculo de las métricas de desempeño: los verdaderos positivos (TP), los falsos positivos (FP), los falsos negativos (FN) y los verdaderos negativos (TN). En nuestro trabajo de investigación, los verdaderos positivos (TP) son los datos anómalos identificados como anomalías. Los falsos positivos (FP) son los datos normales clasificados como datos anómalos. Los falsos negativos (FN) son

aquellos datos anómalos clasificados como normales. Los Verdaderos negativos (TN) son los datos normales que fueron clasificados como normales. Entonces, para medir el desempeño del algoritmo se va a utilizar cuatro métricas: precisión (precision), sensibilidad (Recall), exactitud (Accuracy) y F1 – Score.

La precisión se calcula usando la Ecuación 3, en donde se divide los verdaderos positivos entre la suma de los verdaderos positivos con los falsos positivos. La sensibilidad se calcula usando la Ecuación 4, en donde se divide los verdaderos positivos entre la suman de los verdaderos positivos con los falsos positivos. La exactitud se calcula usando la Ecuación 5, en donde se divide la suma de los verdaderos positivos con los verdaderos negativos entre el total de datos. La métrica F1 – Score se calcula usando la Ecuación 6, en donde se mide la relación que existe entre la precisión y la sensibilidad. Para que las métricas de desempeño sean aceptables deben ser mayores a 0.9.

Ecuación 3. Cálculo de la precisión del modelo

$$Precision = \frac{TP}{TP + FP}$$

Nota: (Ayyadevara, 2018)²⁵, pág. 7

Ecuación 4. Cálculo de la sensibilidad del modelo

$$Recall = \frac{TP}{TP + FN}$$

Nota: (Ayyadevara, 2018)²⁵, pág. 7

Ecuación 5. Cálculo de Accuracy

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

Nota: (Ayyadevara, 2018)²⁵, pág. 7

Ecuación 6. Cálculo de F1–score

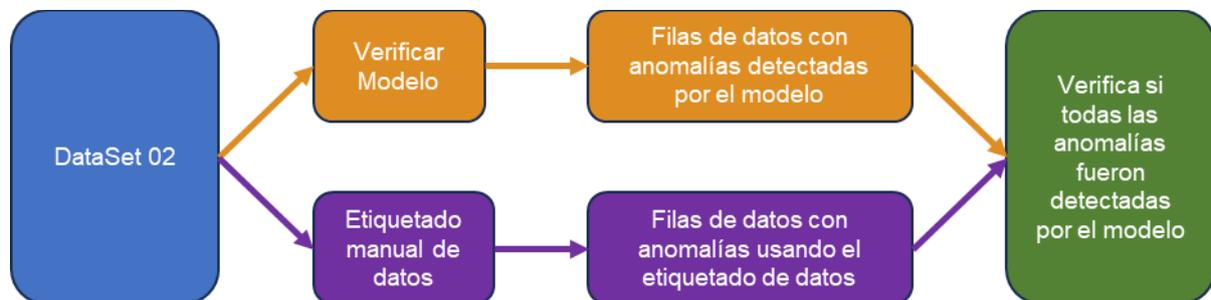
$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Nota: (Ayyadevara, 2018)²⁵, pág. 7

3.3.4. Verificar el modelo basado en el árbol de decisiones

Para verificar la eficiencia del modelo, se ha generado un algoritmo para con el DataSet 02, el cual no se encuentra etiquetado, sea evaluado por el modelo generado durante el entrenamiento. Al final del proceso se van a obtener la lista de anomalías que fueron detectas y se compararán de manera manual haciendo uso una hoja de cálculo la precisión y eficiencia del modelo usando las métricas de aprobación utilizadas durante el entrenamiento y esto lo dejamos descrito en la Figura 14.

Figura 14. Diagrama de flujo para verificar el modelo basado en árbol de decisiones



Nota: Fuente propia

3.4. Evaluación Económica

Para realizar el análisis financiero utilizando el Valor Actual Neto (VAN) y la Tasa Interna de Retorno (TIR), se proyectará un flujo de caja para un periodo de cinco años, considerando la inversión necesaria para la instalación de 20 DTM, con una tasa de descuento del 12%. En el flujo de caja, se incluirá en el año cero el costo de inversión del proyecto, mientras que desde el año 1 hasta el año 5 se considerarán como egresos los ahorros generados por la implementación del proyecto. Además, se incluirá la depreciación de los equipos adquiridos y el pago del impuesto a la renta, que es del 29.5%. Toda esta información se registrará en una hoja de cálculo, siguiendo el formato de la Tabla 8. Este enfoque permitirá evaluar la viabilidad económica del proyecto, asegurando que tanto el VAN como el TIR reflejen con precisión los beneficios financieros y el retorno de la inversión proyectados, facilitando así la toma de decisiones estratégicas para la empresa

Tabla 8. Flujo de caja para el análisis financiero del proyecto

Descripción	0	1	2	3	4	5
Ingresos afectos a impuestos						
Egresos afectos a impuestos (ahorro)						
Δ EBITDA						
Gastos no reembolsables (depreciación)						
Resultado antes de impuestos						
Impuesto de la Renta (29.5%)						
Resultado después de impuesto						
Ajuste por gastos no reembolsables						
Resultado operacional neto						
Inversión						
Ahorro						
Tasa de descuento						
VAN						
TIR						

Nota: Fuente propia

Con los valores de la inversión en el año 0 y los ahorros proyectados desde el año 1 hasta el año 5, se procederá al cálculo del VAN y del TIR del proyecto. Para calcular el VAN, se utilizará la Ecuación 7, donde se restará el valor de la inversión inicial de la sumatoria de los valores presentes de los ahorros futuros. En cuanto al TIR, se construirá una tabla en la que la tasa de descuento variará desde cero hasta encontrar el valor que haga que el VAN sea igual a cero, determinando así el TIR del proyecto. Este proceso permitirá evaluar la rentabilidad del proyecto de manera precisa, asegurando que tanto el VAN como el TIR reflejen la viabilidad económica y el potencial retorno sobre la inversión, lo cual es crucial para tomar decisiones estratégicas informadas

Ecuación 7. Cálculo del VAN del proyecto

$$VAN = INVERSION - \sum_{i=1}^n \frac{VA}{(1+r)^i}$$

Nota: “n” es la cantidad de periodos, “r” es la tasa de descuento y “VA” son los valores de ahorro futuros en cada periodo

Para determinar el monto a invertir, se considerarán los costos asociados al suministro e instalación del equipo de Monitoreo del Transformador de Distribución (DTM), los periféricos necesarios para el análisis de datos, los programas requeridos para la gestión de estos, y el soporte técnico para el mantenimiento de los sistemas computacionales. Para calcular los egresos antes de impuestos, es decir, el ahorro potencial previo a impuestos que se logrará con la instalación del DTM, se tendrán en cuenta la cantidad de clientes afectados, el tiempo de interrupción, la cantidad de eventos y el costo promedio de la energía. Todo este cálculo se llevará a cabo mediante una hoja de cálculo, asegurando que todos los factores relevantes sean considerados para obtener una estimación precisa de la inversión y los ahorros proyectados, facilitando así una evaluación financiera integral del proyecto.

CAPÍTULO IV. RESULTADOS

4.1. Recolección de datos

Ubicquia (2024)³² es una compañía que ha desarrollado un equipo de monitoreo para un transformador de distribución llamado DTM+. El DTM+ es un dispositivo electrónico que mide las variables de voltaje, corriente, presión y temperatura de un transformador de distribución. Sin embargo, como el transformador de distribución de prueba trabaja a presión ambiente, es decir, no ocurre una hermeticidad completa, no es factible medir las variables de presión y temperatura. Entonces, para este caso de estudio solo se cuenta con las variables de voltaje en baja tensión y corriente en media y baja tensión. Y para nuestro trabajo de investigación vamos a hacer uso de las variables de voltaje y corriente en baja tensión.

El DTM+ es un dispositivo electrónico compuesto por un conjunto de sensores tipo Rogowski³³ para la medición de corriente y cables para la medición de voltaje del transformador. Tiene un sistema de imantado que permite adherirlo a la carcasa del transformador permitiendo una instalación fácil y rápida. En caso el equipo se instala en subestaciones aéreas tiene la capacidad de medir las vibraciones que ocurra en tiempo real. Cuenta con la capacidad de enviar los valores de tensión y de corriente cada minuto hacia una plataforma web. Para ello utiliza las comunicaciones LTE privadas o públicas, lo que permite brindar velocidad y conectividad de red necesarias e independiente del operador de telecomunicaciones.

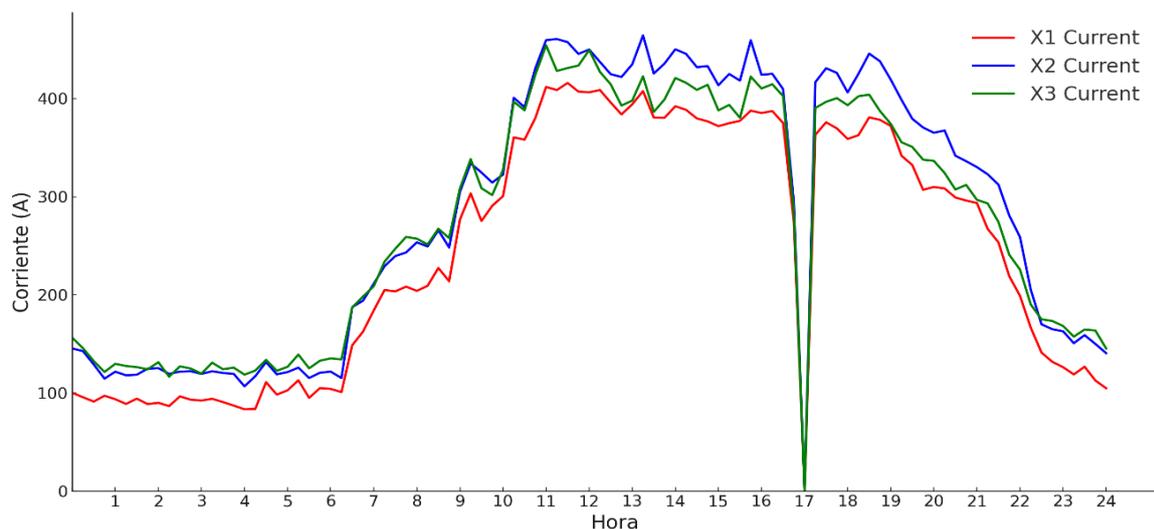
Para la recolección de datos se instaló un equipo de monitoreo de transformador (DTM+) entregado por el proveedor para evaluar si se puede instalar un piloto de 20

³² UBICQUIA (2024). <https://www.ubicquia.com/products/distribution-transformer-monitoring-dtm>. Ubicquia, Inc. 401 E Las Olas Blvd, Suite 800, United States.

³³ Sensores tipo Rogowski: dispositivo en forma de lazo o anillo que se coloca alrededor del cable eléctrico para medir corrientes eléctricas alternas.

unidades. Para el proceso de entrenamiento y validación del modelo se tomó el conjunto de datos “DataSet 01” que es conformado por las mediciones realizadas durante los meses de mayo y junio de 2024. En ese periodo se midió un total de 101 758 registros de datos de corriente y de voltaje en baja tensión. Mientras que, para verificar el modelo obtenido se hará uso del conjunto de datos “DataSet 02” que es conformado por las mediciones realizada durante el periodo julio de 2024. En ese periodo se midió un total de 25 809 registros de datos de corriente y voltaje en baja tensión. En el DataSet 01 se ha detectado que la cantidad de datos anómalos solo representan el 0.04% del total de datos existentes.

Figura 15. Diagrama de carga de las tres fases de corriente durante el día 18/06/2024



Nota: Fuente propia

Del DataSet 01 se ha seleccionado el día 18/06/2024 para visualizar el comportamiento de la carga durante un día cualquiera y visualizar cuando ocurre la hora punta. En la Figura 15 se observa que la carga se eleva a las 6:00 horas hasta las 19:00 horas; entonces, la hora punta del transformador es en dicho periodo de tiempo. Además, se observa una caída abrupta al valor de cero en un punto que puede ser interpretado como una anomalía que apareció durante la medición y que merece ser revisada con mayor detalle. Ahora, para comprobar la idoneidad del DataSet 01 se evalúa la correlación entre corriente y voltaje de baja tensión, y la correlación de las anomalías y estas variables. Además, se

realiza el análisis de histogramas y la densidad de datos de las tres fases de corriente y de voltaje en baja tensión.

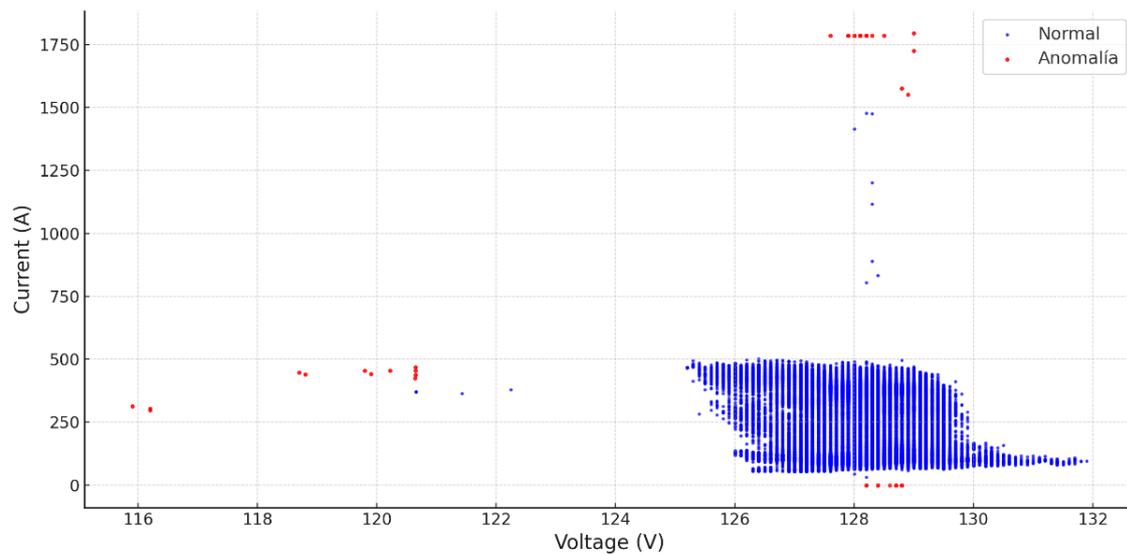
En la Figura 16, se observa una alta correlación entre las tres fases de corriente y entre las tres fases de voltaje. Además, existe una baja correlación inversa entre corriente y voltaje. Entonces, se llega a la conclusión de que se puede tomar como referencia una de las corrientes y uno de los voltajes para realizar el análisis de histogramas y de densidad de datos. Para las corrientes usaremos el valor máximo por registro de corriente y para los voltajes el valor mínimo para mantener la consistencia en las características del modelo entrenado. En la Figura 17, se observa que la mayoría de puntos de color azul están agrupados, mostrando las condiciones operativas del transformador. Los puntos rojos están agrupados en tres zonas, donde la zona inferior se encuentran las anomalías tipo 1, en la parte superior se encuentran las anomalías tipo 2 y en la zona izquierda se encuentran las anomalías tipo 3. Debido a que los puntos anómalos se encuentran fuera de la zona estable se recomienda realizar un mayor estudio para identificar su origen o realizar otras pruebas.

Figura 16. Correlación entre Corriente y Voltaje en baja tensión

	X1 Current (A)	X2 Current (A)	X3 Current (A)	X1 Voltage (V)	X2 Voltage (V)	X3 Voltage (V)
X1 Current (A)	1.00	0.99	0.98	-0.13	-0.15	-0.12
X2 Current (A)	0.99	1.00	0.99	-0.15	-0.17	-0.14
X3 Current (A)	0.98	0.99	1.00	-0.15	-0.16	-0.15
X1 Voltage (V)	-0.13	-0.15	-0.15	1.00	0.99	0.97
X2 Voltage (V)	-0.15	-0.17	-0.16	0.99	1.00	0.98
X3 Voltage (V)	-0.12	-0.14	-0.15	0.97	0.98	1.00

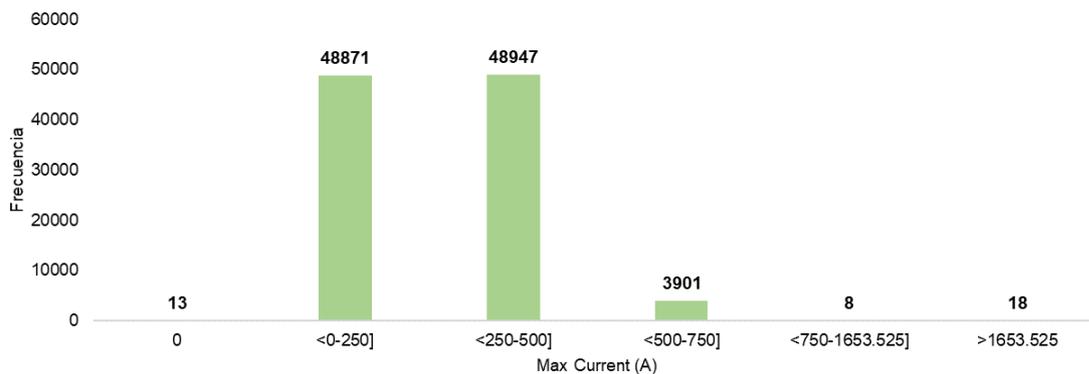
Nota: Fuente propia

Figura 17. Correlación entre anomalías vs Corriente y Voltajes en baja tensión



Nota: Fuente propia

Figura 18. Histograma de corriente en baja tensión

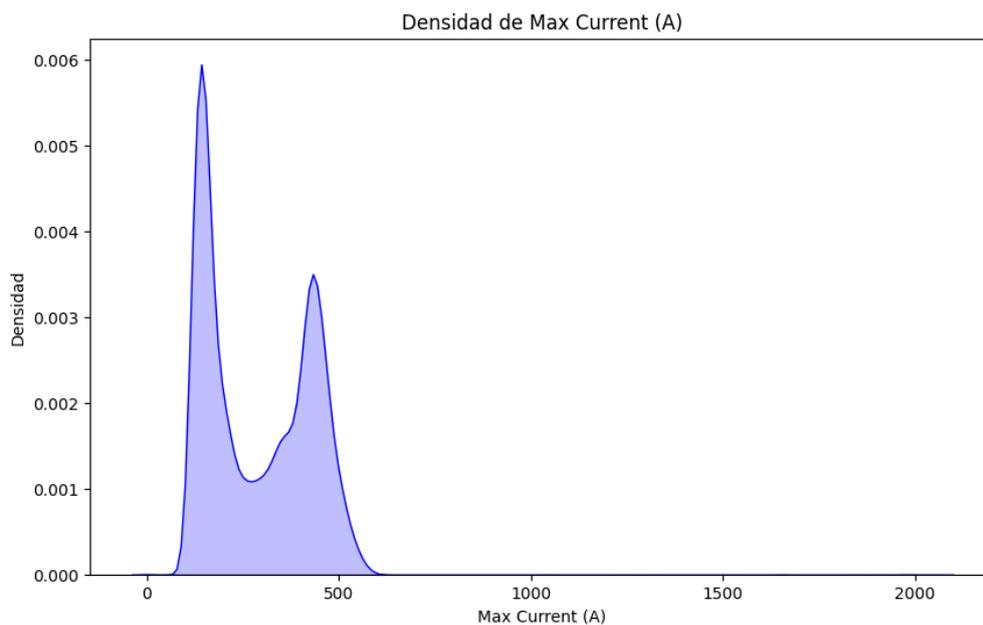


Nota: Fuente propia

Si revisamos el histograma de corriente en la Figura 18, podemos observar que el 48.03% de los datos se encuentra en el rango entre 0A y 250A, 48.10% de los datos está entre el rango de 250A y 500A y el 3.84% de los datos se encuentra entre 500A y 1653.525A. Con esta información se puede confirmar que el 99.97% de los datos se encuentran dentro del rango de trabajo del transformador, el 0.01% de los datos es igual a cero y el 0.02% se encuentra fuera del rango de la corriente nominal del transformador. Si revisamos la densidad de datos de corriente descrito en la Figura 19, se aprecia una alta

concentración de datos cercanos a 0 A y a 500 A, lo que muestra el comportamiento del consumo eléctrico de los clientes tanto en hora punta como en fuera punta. Además, se han encontrado valores mayores a 500 A que son raros para el funcionamiento normal del transformador; lo que es un indicador a que algo está sucediendo en el consumo de energía por parte de los clientes o de las condiciones de la red de baja tensión.

Figura 19. Densidad de datos de corriente en baja tensión

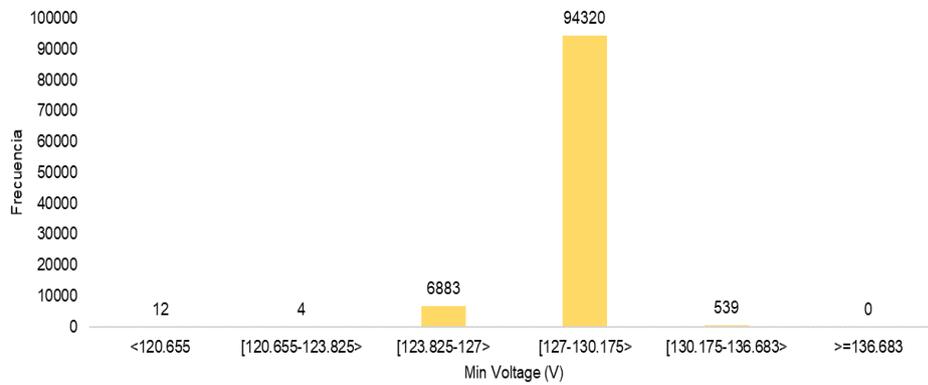


Nota: Fuente propia

Si realizamos el mismo ejercicio para los valores de voltaje en baja tensión, tenemos que en la Figura 20, se observa que el 99.99% de los valores de voltaje se encuentran dentro del rango permitido por la NTCSE y que solo el 0.01% de los datos se encuentran por debajo de dicho valor. Si revisamos la densidad de datos descrito en la Figura 21, podemos observar que hay un pico en 128 V (221.69 V fase a fase) una alta concentración alrededor de dicho valor que muestra la operación normal del transformador. Pero hay una cola hacia la izquierda próximo a 125 V (216.5 V fase – fase) que muestra que una pequeña parte de los datos se aproxima a dicho valor producto del aumento de carga de consumo por parte de los clientes, pero no sale del rango mínimo permitido por

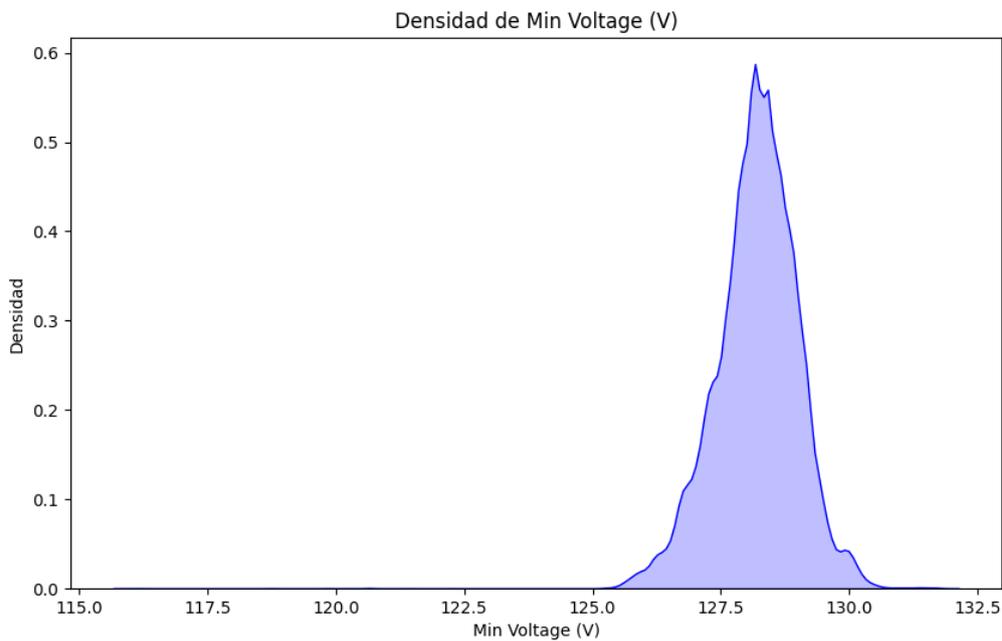
la NTCSE. Además, se muestra raros valores de voltaje menores a 125V y mayores a 130V que podría ser indicado de alguna anomalía en el sistema eléctrico.

Figura 20. Histograma de voltaje en baja tensión



Nota: Fuente propia

Figura 21. Densidad de datos de voltaje en baja tensión



Nota: Fuente propia

Por lo tanto, con lo descrito anteriormente podemos afirmar que los datos considerados como anomalías para este trabajo de investigación se encuentra dentro de los rangos considerados como anómalos para el análisis de datos realizado.

4.2. Resultados de la parametrización de anomalías

La función del transformador de distribución es convertir la energía eléctrica de media tensión a baja tensión para que los clientes puedan hacer uso de la misma. Entonces, podemos afirmar que el modo de falla potencial para el transformado viene a ser que “no transforma la energía eléctrica”. Ahora procedemos a definir tres efectos de falla potencial con su respectiva causa, teniendo en cuenta que hay una diversidad de causas que hayan generado los efectos de la falla potencial. Para el presente trabajo de investigación solo vamos a escoger una de ellas, ya que el objetivo de esta tesis es hacer uso del aprendizaje automático y no de identificar todas las causas que generan los efectos de una falla potencial, por ello en la Tabla 9 se muestra lo siguiente:

Tabla 9. Modos de falla y análisis de efectos de un transformador

Función	Modo de falla potencial	Efectos de la falla potencial	Causas
convertir la energía eléctrica de media tensión (10 000 o 20 000 Volts) a baja tensión (220 Volts)	No transforma la energía eléctrica	Interrupción del servicio eléctrico	Cortocircuito externo
		Aumento de temperatura en el transformador	Sobrecarga del transformador
		Incumplimiento de la NTCSE por caída de voltaje	Caída de voltaje

Nota: Fuente Propia

Un primer efecto de falla potencial, se puede escoger la interrupción del suministro eléctrico que impacta de manera directa al SAIDI y SAIFI. Entonces, podemos afirmar que una causa es un cortocircuito externo en las redes de media tensión que conectan a la subestación eléctrica de transmisión (SET) con la subestación eléctrica de distribución (SED) en donde se ubica el transformador de distribución en estudio. Un segundo efecto de la falla potencial, se puede escoger un aumento de la temperatura en el transformador, ya que en el tiempo deteriora al equipo y al mediano plazo provocaría una interrupción en

el suministro eléctrico e impactar al SAIDI y SAIFI. Entonces, podemos afirmar que una causa es la sobrecarga del transformador, provocado por un mayor consumo por parte de los usuarios finales. Un tercer efecto de la falla potencial, se puede escoger el incumplimiento de la NTCSE³⁴ por caída de tensión. Entonces, podemos afirmar que una causa es la caída de tensión provocada por un ingreso o salida intempestiva de carga. Teniendo en cuenta que una caída en el nivel de tensión impacta de manera directa a los clientes en el aprovechamiento de la energía eléctrica y afectando a sus equipos conectados a la red.

Tabla 10. Modos de controles actuales

Efectos de la falla potencial	Causas	Modo de controles actuales	Acciones recomendadas
Interrupción del servicio eléctrico	Cortocircuito	Interrupción del servicio	Contar con un monitoreo y control en línea de los parámetros de corriente y voltaje haciendo uso del árbol de decisiones para la detección de anomalías por clasificación
Aumento de temperatura en el transformador	Sobrecarga del transformador	Medición puntual de corriente en baja tensión cada vez que se realiza mantenimiento en la subestación (1 vez al año)	
Incumplimiento de la NTCSE por caída de voltaje	Caída de voltaje	Medición puntual de voltaje en baja tensión cada vez que se realiza mantenimiento en la subestación (1 vez al año)	

Nota: Fuente Propia

En la Tabla 10 vamos a describir los modos de controles actuales que pasamos a detallar: un primer modo de control actual asociado al primer efecto de la falla potencial es la recepción de llamadas por parte de los clientes, quienes se quejan de la falta de energía eléctrica y que puede identificarse en el transformador cuando los valores de tensión y/o voltaje son iguales a cero. Un segundo modo de control actual asociado al segundo efecto de la falla potencial son las mediciones puntuales de corriente en baja tensión cada vez

³⁴ NTCSE: Norma Técnica de Calidad de los Servicios Eléctricos. Perú

que se realiza el mantenimiento preventivo sin corte en la subestación eléctrica de distribución (SED). Un tercer modo de control actual asociado al tercer efecto de la falla potencial es la son las mediciones puntuales de voltaje en baja tensión cada vez que se realiza el mantenimiento preventivo sin corte en la subestación eléctrica de distribución (SED).

Tabla 11. Anomalías del presente trabajo de investigación

ID	Anomalía	Código	Etiqueta
1	Cortocircuito externo	Anomalia01	01
2	Sobrecarga	Anomalia02	02
3	Caída de voltaje	Anomalia03	03

Nota: Fuente Propia

Dado que las causas de los efectos de la falla potencial son condiciones subestándares del transformador de distribución, se definirán las anomalías a etiquetar en este trabajo de investigación, como se muestra en la Tabla 11. Estas anomalías representan las condiciones específicas que se analizarán y serán fundamentales para el desarrollo del modelo propuesto. Con esto, las anomalías a considerar en el presente estudio quedan claramente establecidas y servirán como base para el análisis y la evaluación del transformador en cuestión.

4.3. Resultado del etiquetado de datos

Como resultado, luego de compilar el algoritmo de etiquetado de anomalías, nos da como resultado 43 filas de datos con anomalías, tal como se describe en la Tabla 12. En donde, se detectó que 13 filas de datos están relacionadas con el cortocircuito externo (Anomalia01), se detectó que 18 filas de datos están relacionadas con la sobrecarga (Anomalia02) en el transformador de distribución y se detectó que 12 filas de datos están relacionadas con la caída de tensión (Anomalia03).

Tabla 12. Filas de datos con etiqueta de anomalía

H1 Current	H2 Current	H3 Current	X1 Current	X2 Current	X3 Current	X1 Voltage	X2 Voltage	X3 Voltage	Etiqueta
11.5	9.9	11.9	454.2	563	521.5	120.65	121.2	120.85	3
11.6	9.8	12.1	468.9	565.9	527.4	120.65	121.2	120.85	3
11.5	9.7	11.6	454.8	566.5	507	120.22	120.88	120.43	3
11.8	10.2	11.9	455.3	586	532.7	119.8	120.21	120.02	3
11.2	9.7	11.5	437.3	552.7	512.1	120.65	120.75	120.65	3
10.8	9.7	11.4	425.3	547	517.9	120.64	120.75	120.65	3
11.1	9.6	11.6	441.5	543.3	511.7	119.9	120.3	120.2	3
11	9.1	11.2	439.7	524.9	480.7	118.8	119.2	119.1	3
11.1	9.1	11.2	446.5	526.3	478	118.7	119.2	118.8	3
7.6	6.1	7.6	304.1	357.5	319.5	116.2	116.8	116.1	3
8	6.4	7.9	313.3	381	335.2	115.9	116.2	115.9	3
7.6	6.3	7.6	298	359.4	315.6	116.2	116.8	116.2	3
36.8	31.47	36.63	1575.51	1653.7	1379.51	128.8	129.8	129.1	2
35.48	30.77	31.74	1550.76	1653.7	1240.03	128.9	129.9	129.2	2
40.06	34.7	37.33	1725.03	1850	1439.38	129	129.9	129.2	2
42.59	36.01	33.04	1793.93	1902	1327.69	129	129.9	129.1	2
0	0	0	0	0	0	128.7	129.7	129	1
0	0	0	0	0	0	128.4	129.4	128.7	1
0	0	0	0	0	0	128.2	129.4	128.6	1
0	0	0	0	0	0	128.8	129.8	129.1	1
0	0	0	0	0	0	128.7	129.7	129	1
0	0	0	0	0	0	128.8	129.7	129	1
0	0	0	0	0	0	128.6	129.5	128.8	1
0	0	0	0	0	0	128.7	129.7	129	1
0	0	0	0	0	0	128.4	129.4	128.7	1
0	0	0	0	0	0	128.2	129.4	128.6	1
0	0	0	0	0	0	128.8	129.8	129.1	1
0	0	0	0	0	0	128.7	129.7	129	1
0	0	0	0	0	0	128.8	129.7	129	1
43.1	47.7	43.9	1785.6	1976.1	1817.3	128	128.7	128.1	2
42.5	45.3	42	1785.6	1901.9	1766	128	128.8	128.1	2
41.3	42.7	40.9	1785.6	1845.4	1770.3	128.1	128.9	128.4	2
42.2	47.1	44.3	1785.5	1994.9	1877.6	127.6	128.4	127.7	2
43.3	47.8	45.6	1785.6	1971.6	1880.9	128	128.9	128.2	2
43.6	50.3	47.4	1785.6	2059.3	1943.5	127.9	128.7	128	2
42.5	47.4	45.2	1785.6	1992.4	1898.1	128.2	129.1	128.4	2
42.3	46.6	45.4	1785.4	1966.6	1916.4	128.1	129	128.2	2
43.8	50.2	46.3	1785.6	2046.6	1886.5	128.2	129.1	128.3	2
43.5	48.4	45	1785.5	1986.8	1846.4	128.1	128.9	128.3	2
43.1	48.4	45.2	1785.6	2007.6	1873.6	127.9	128.7	128.1	2
42.4	46.6	46.8	1785.5	1960.7	1844.1	128.3	129	128.5	2
42.6	45.7	43.1	1785.6	1916.2	1807.8	128.2	129.1	128.4	2
43.4	48.2	43.6	1785.6	1982.4	1794.9	128.5	129.4	128.8	2

Nota: Fuente propia

4.4. Resultados del entrenamiento del algoritmo

4.4.1. Preparación de datos

El DataSet 01 cuenta con 101 758 filas y 07 columnas, 03 columnas de corriente, 03 columnas de voltaje y 01 columna de etiqueta. La columna etiqueta clasifica los registros de datos en normales o anómalos, para los datos normales se le asigna el valor de 0 y para los datos anómalos el valor de 1, 2 o 3. Las columnas X1 Current, X2 Current y X3 Current representan las tres fases de corriente y las columnas X1 Voltage, X2 Voltage y X3 Voltage las tres fases de voltaje. El algoritmo utiliza *Pandas*³⁵ para leer el archivo y crea un *DataFrame*³⁶ para organizar los datos en filas y columnas. Luego verifica que no hay valores faltantes en alguna columna, asegurando que el conjunto de datos está limpio y listo para el procesamiento. Posteriormente, realiza la separación de datos en características (x) que son las columnas de corriente y voltaje, y etiquetas (y) que es la función objetivo. Continuamos con la estandarización de datos, que permite eliminar la diferencia de escala que hay entre las columnas de corriente y voltaje. La estandarización utilizada es la *StandardScaler* que permite ajustar los valores de la columna de características (x) a una media de cero y a una desviación estándar de uno.

4.4.2. Entrenamiento del algoritmo

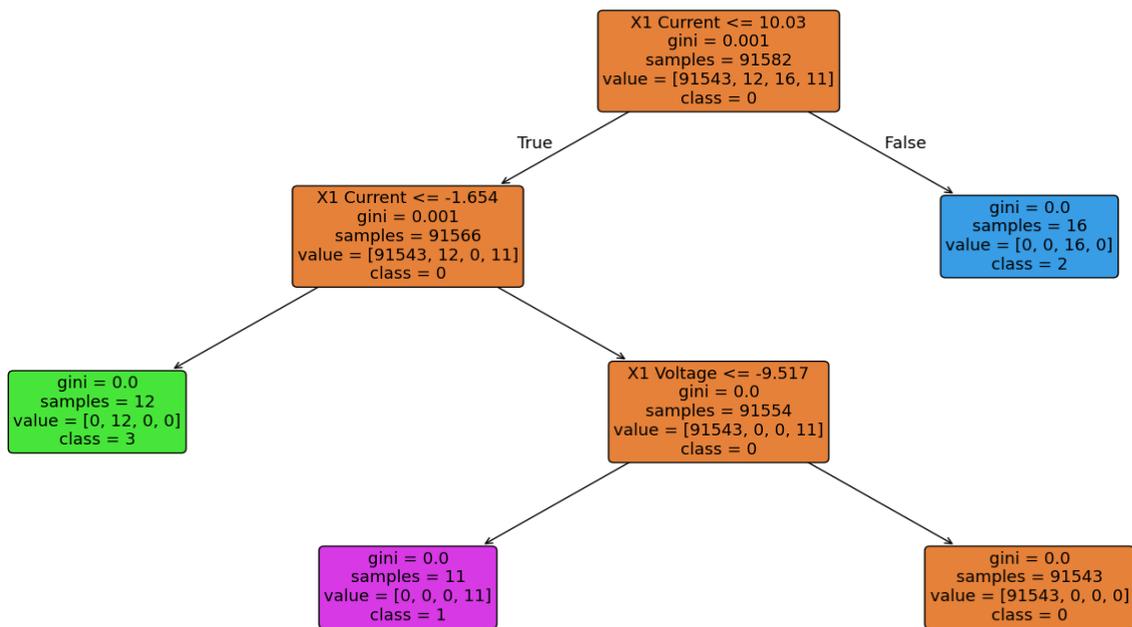
Haciendo uso de la validación cruzada estratificada de 10 iteraciones, el DataSet 01 se ha dividido en 10 subconjuntos, los cuales cuentan con datos normales y anómalos en cada subconjunto. Posteriormente el algoritmo de árbol de decisiones realiza estas 10 iteraciones combinando los 10 subconjuntos, de tal manera que todos los subconjuntos se encuentren tanto en el lado del entrenamiento como de

³⁵ Pandas es una biblioteca de Python utilizada para el análisis y manipulación de datos.

³⁶ DataFrame es una estructura de datos de formato bidimensional de Pandas.

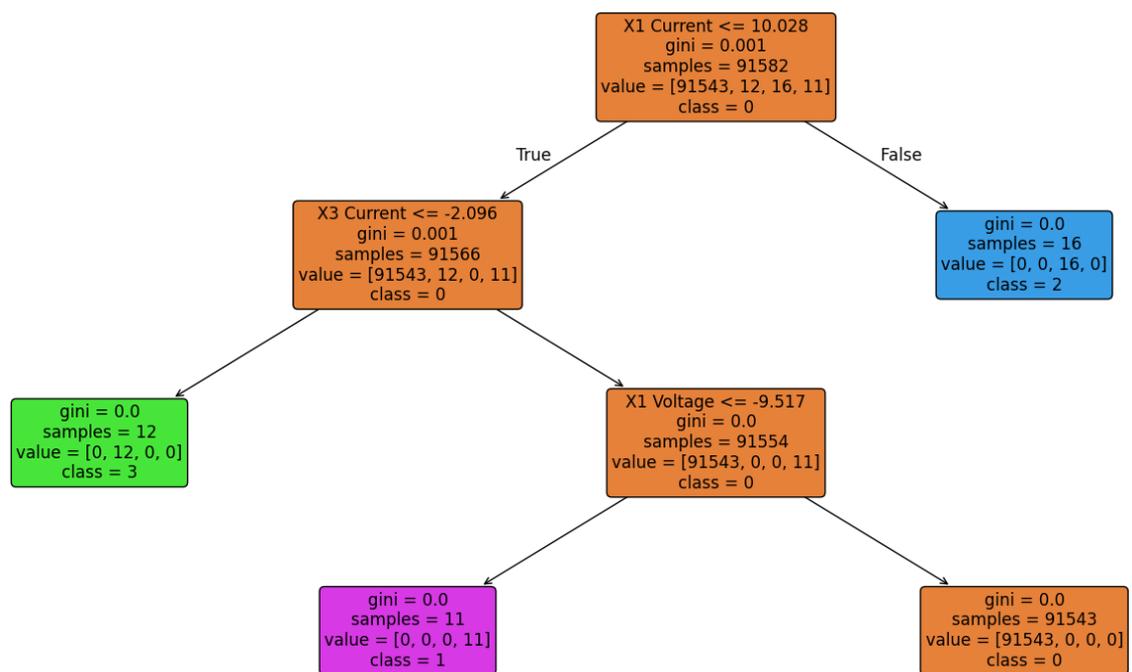
la prueba del modelo. Esto queda descrito en la Figura 22, Figura 23, Figura 24, Figura 25, Figura 26, Figura 27, Figura 28, Figura 29, Figura 30 y Figura 31.

Figura 22. Árbol de decisión – Iteración 01



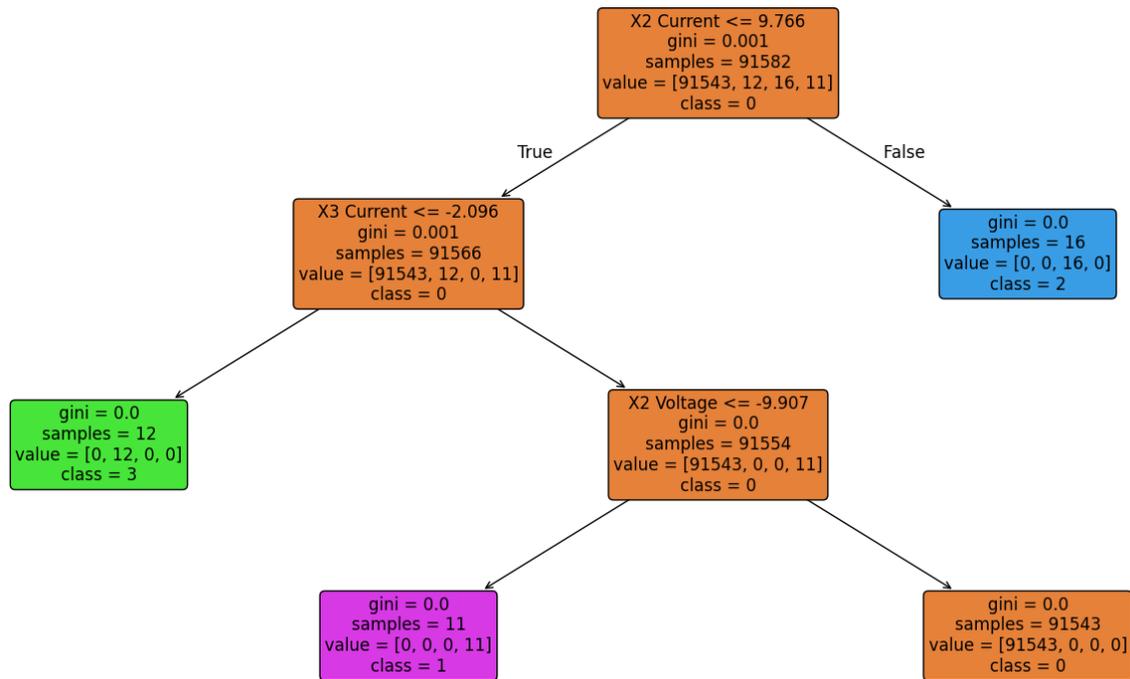
Fuente: Propia

Figura 23. Árbol de decisión – Iteración 02



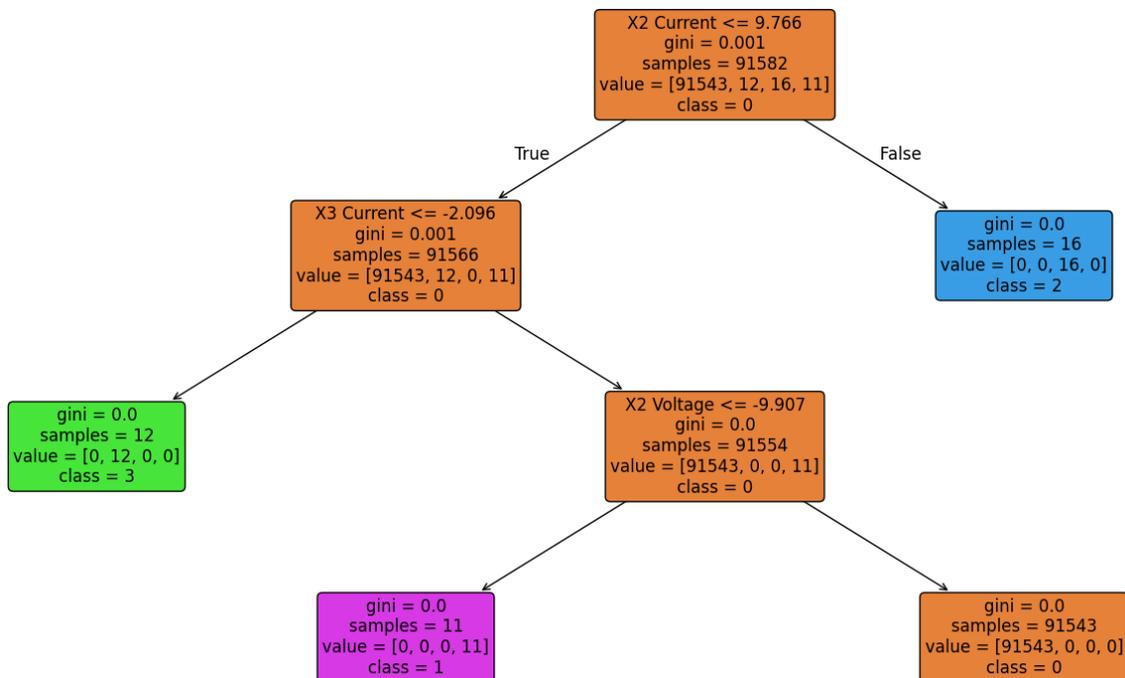
Fuente: Propia

Figura 24. Árbol de decisión – Iteración 03



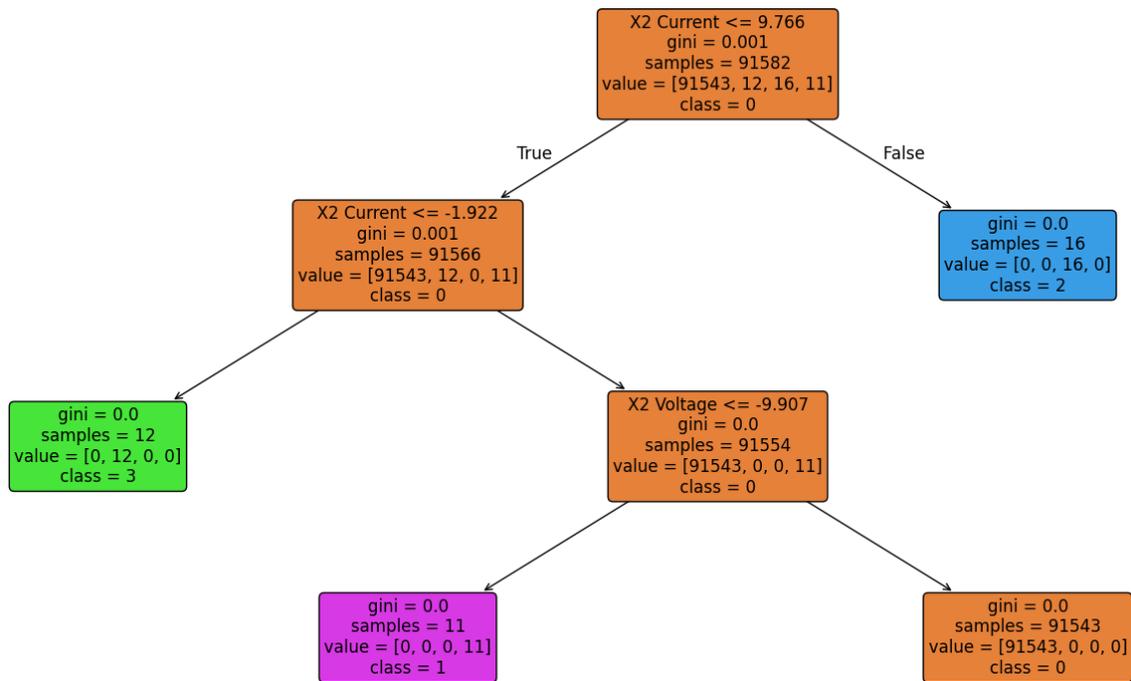
Fuente: Propia

Figura 25. Árbol de decisión – Iteración 04



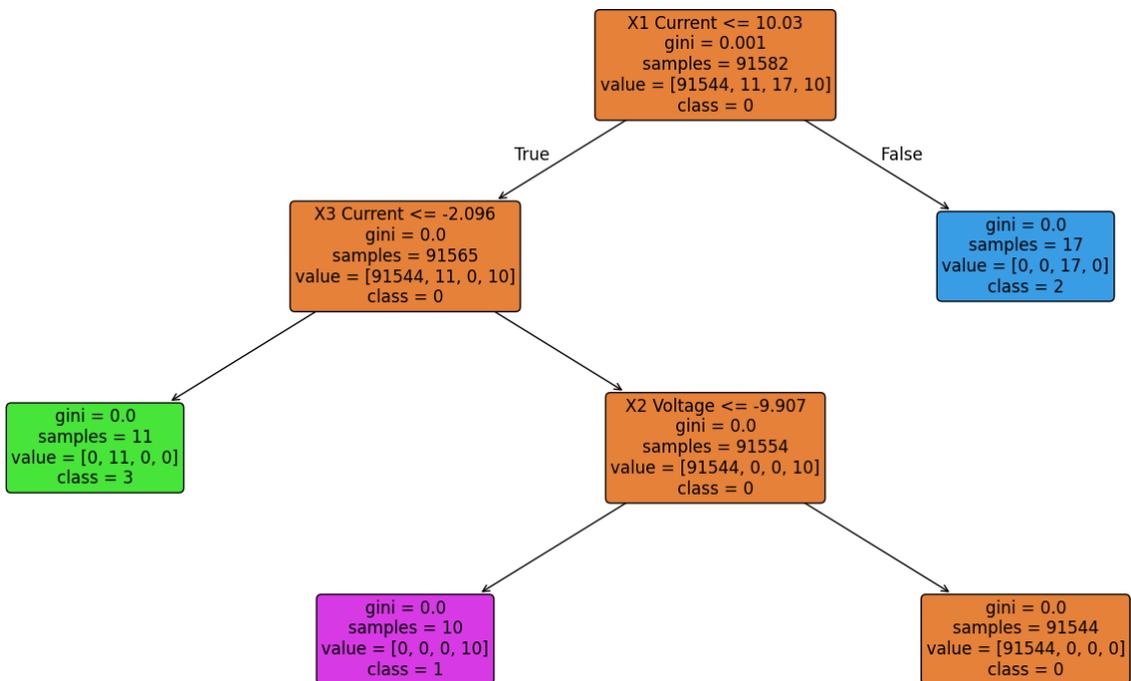
Fuente: Propia

Figura 26. Árbol de decisión – Iteración 05



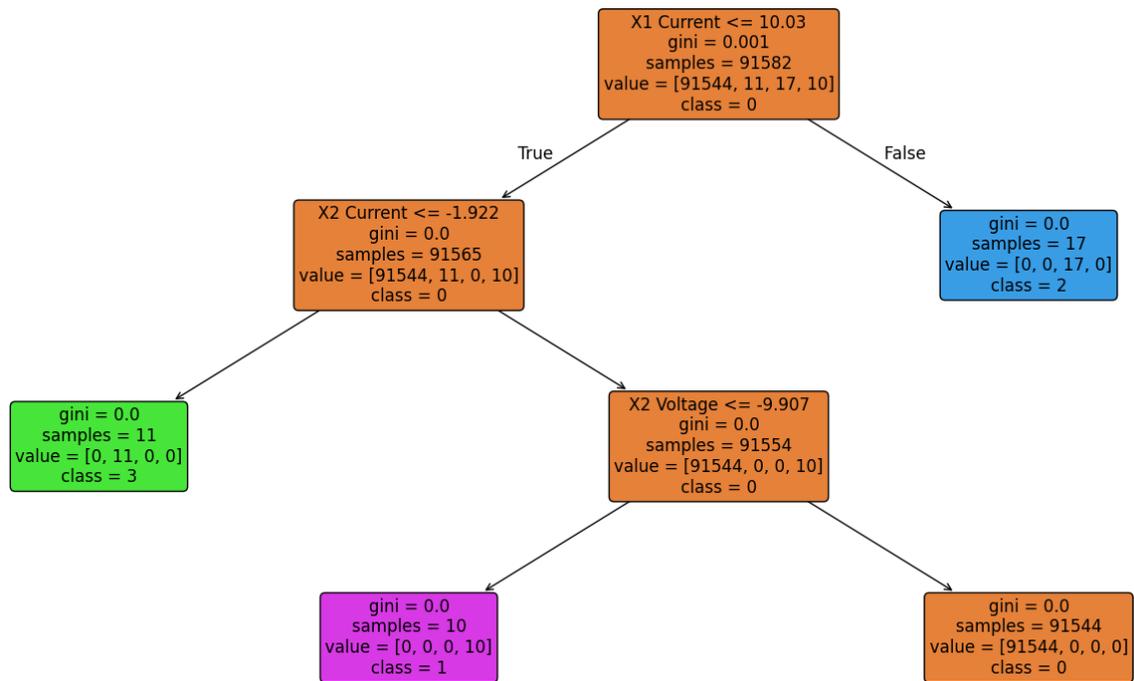
Fuente: Propia

Figura 27. Árbol de decisión – Iteración 06



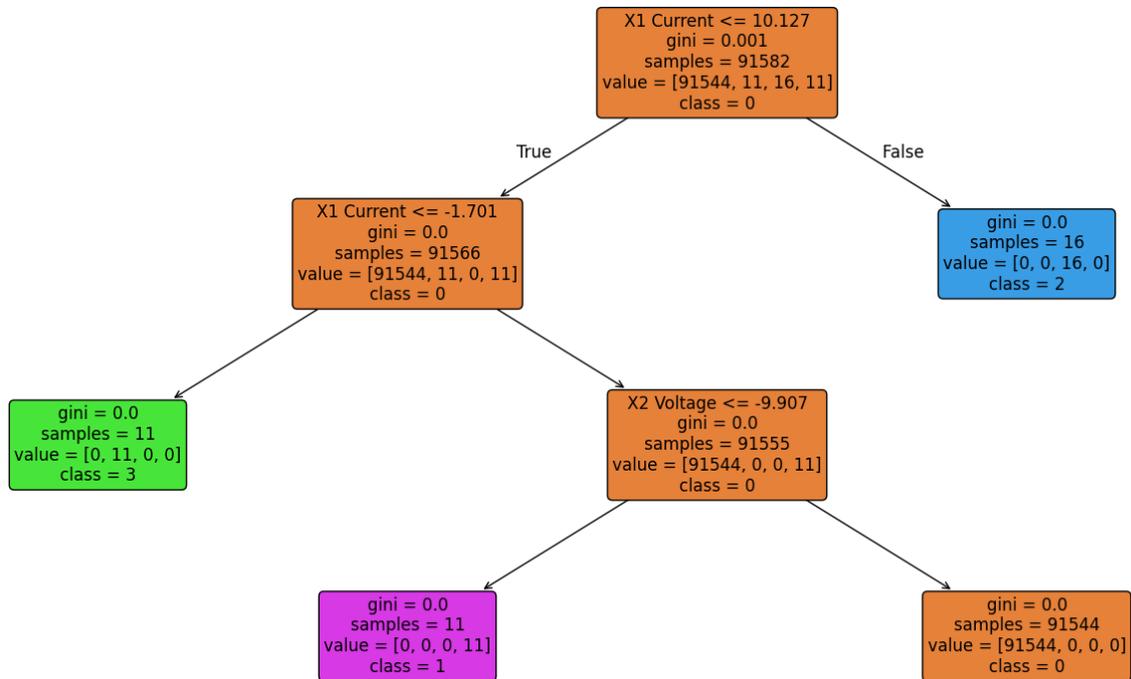
Fuente: Propia

Figura 28. Árbol de decisión – Iteración 07



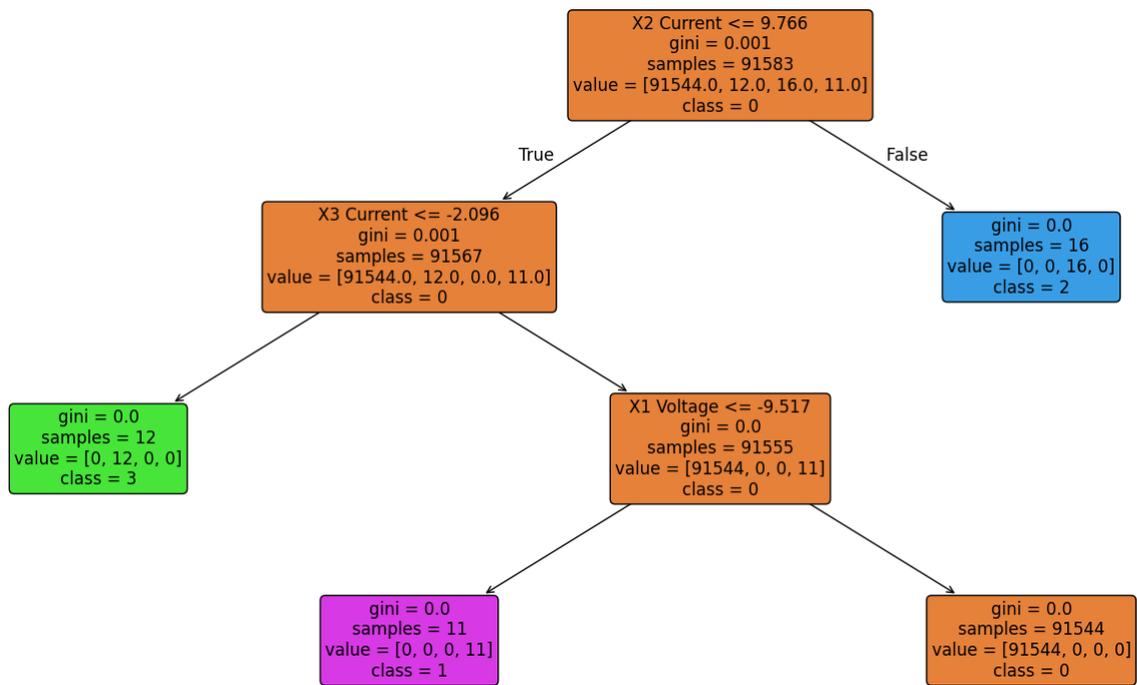
Fuente: Propia

Figura 29. Árbol de decisión – Iteración 08



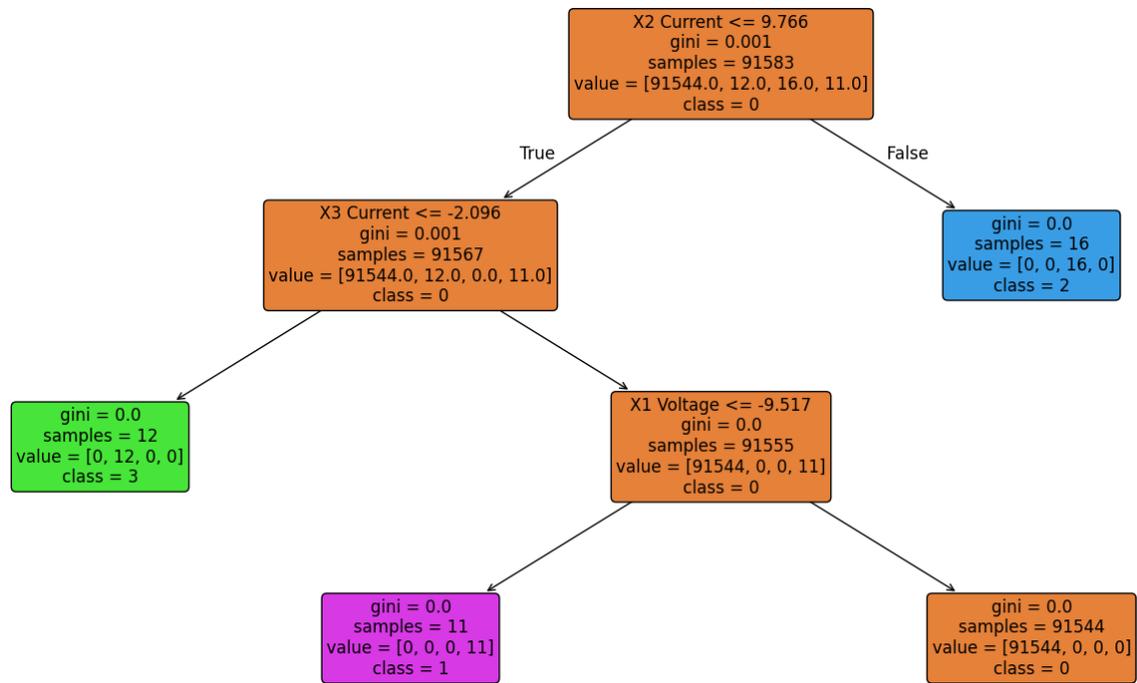
Fuente: Propia

Figura 30. Árbol de decisión – Iteración 09



Fuente: Propia

Figura 31. Árbol de decisión – Iteración 10

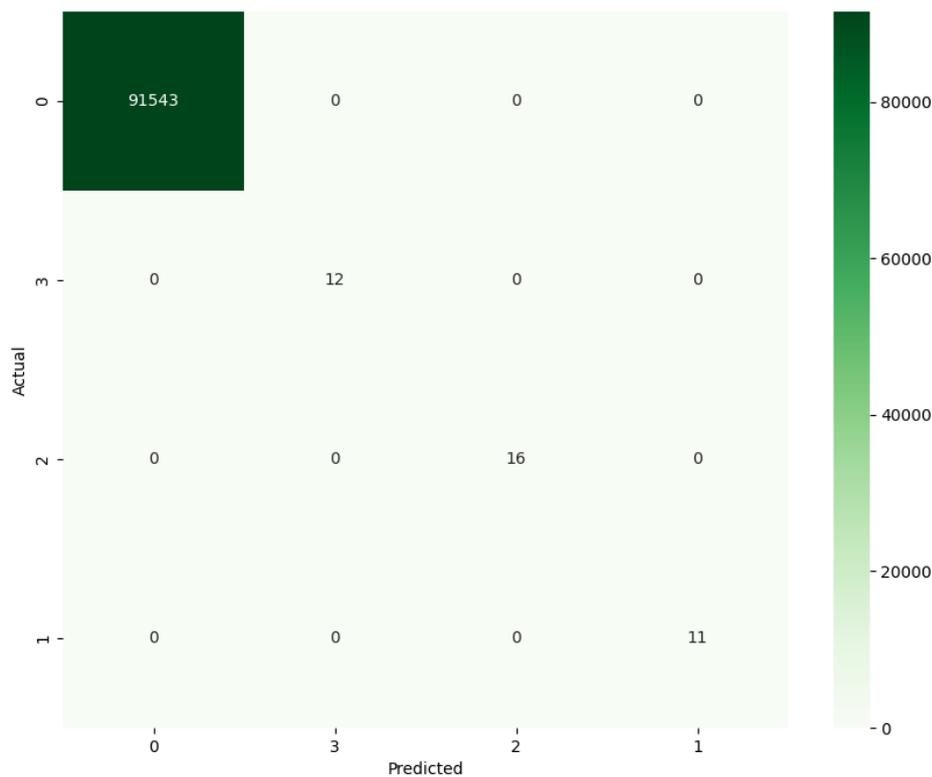


Fuente: Propia

4.4.3. Métricas de desempeño

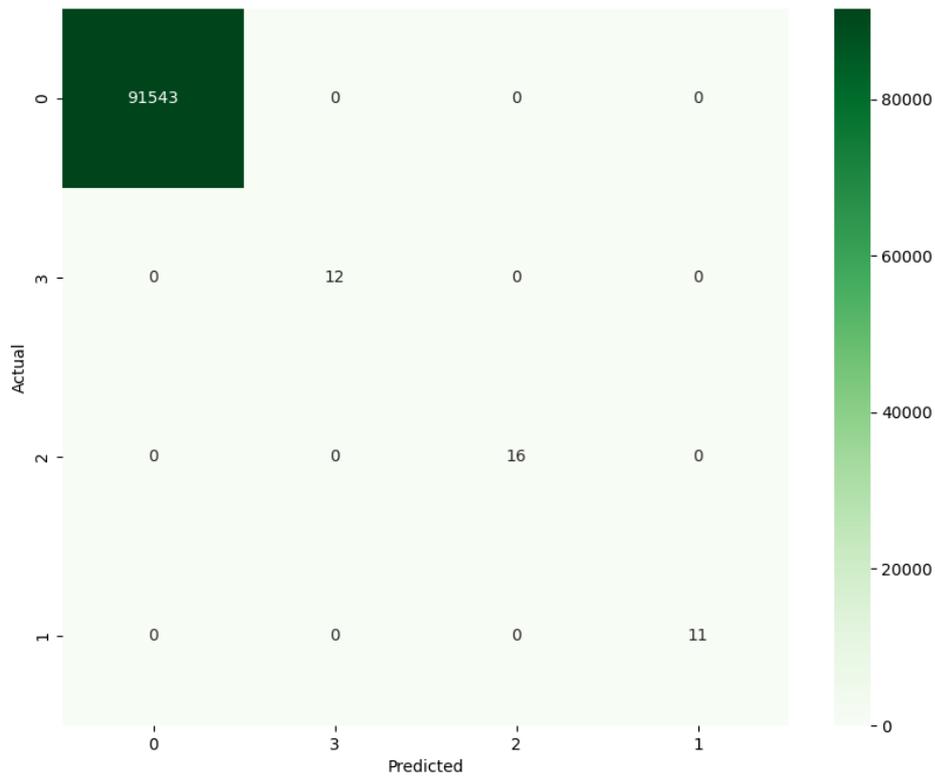
En la Figura 32, Figura 33, Figura 34, Figura 35, Figura 36, Figura 37, Figura 38, Figura 39, Figura 40 y Figura 41, se observan las matrices de confusión del entrenamiento. En la Figura 42, Figura 43, Figura 44, Figura 45, Figura 46, Figura 47, Figura 48, Figura 49, Figura 50 y Figura 51, se observan las matrices de confusión de la prueba. En las matrices se observa que las filas representan la etiqueta actual y las columnas representan la etiqueta predicha. Entonces, como se observa los valores se encuentran concentrados en la diagonal, es decir, las etiquetas predichas son iguales a las etiquetas actuales. Por lo tanto, se aprecia que los datos han sido clasificados de manera correcta, lo que refleja un rendimiento óptimo del modelo durante las fases de entrenamiento y prueba.

Figura 32. Matriz de confusión para el entrenamiento – iteración 01



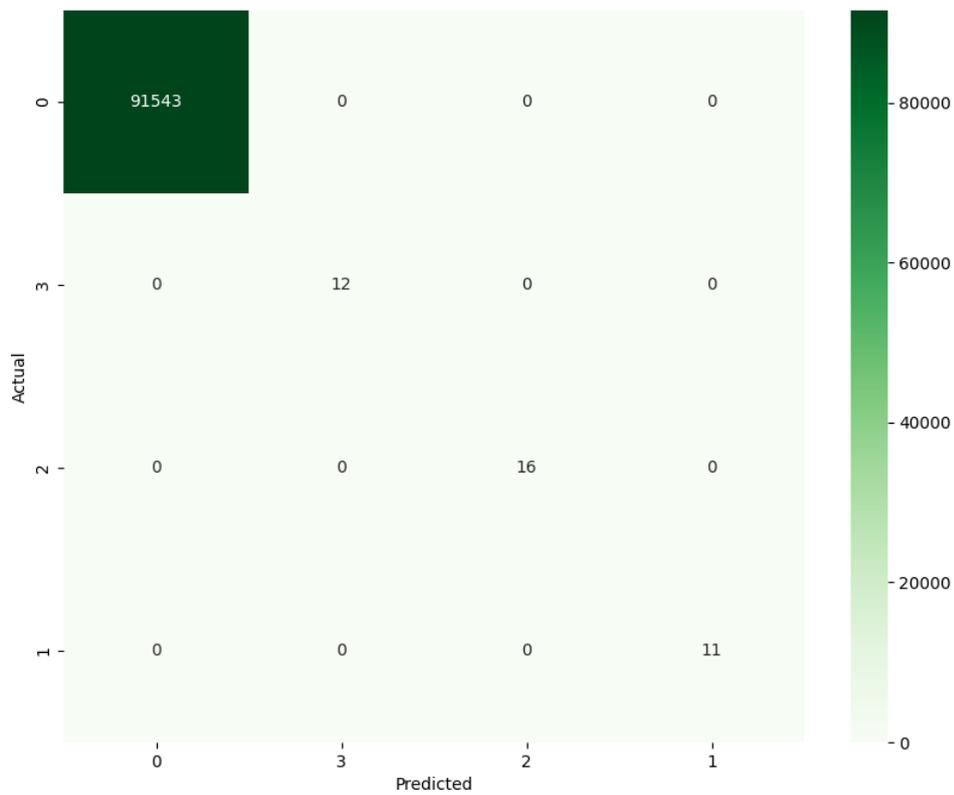
Fuente: Propia

Figura 33. Matriz de confusión para el entrenamiento – iteración 02



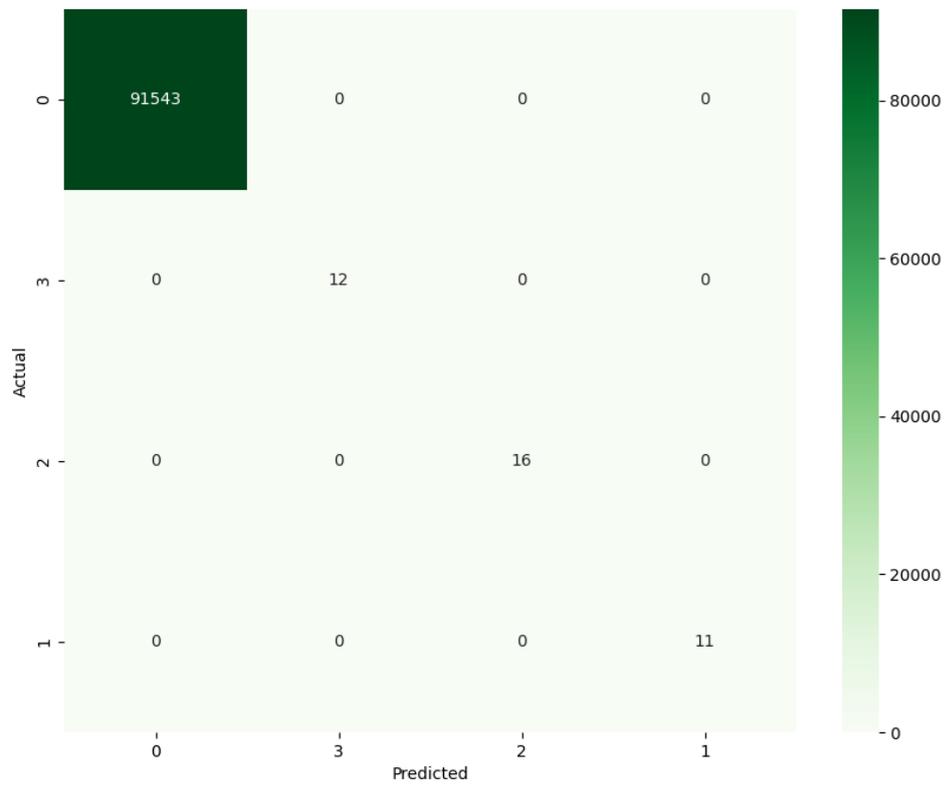
Fuente: Propia

Figura 34. Matriz de confusión para el entrenamiento – iteración 03



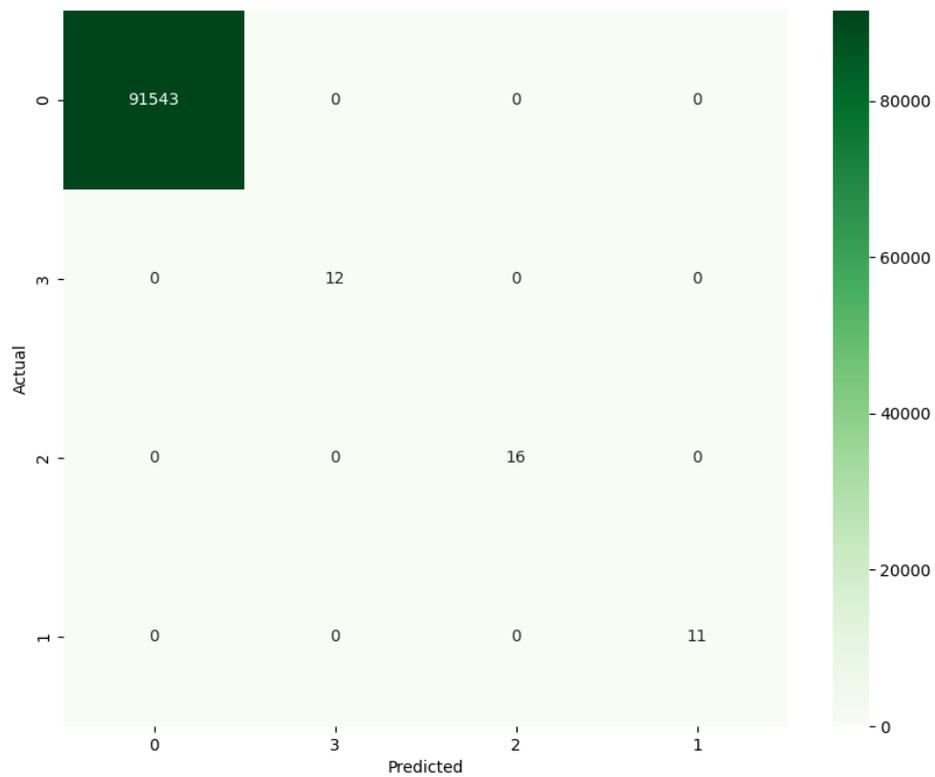
Fuente: Propia

Figura 35. Matriz de confusión para el entrenamiento – iteración 04



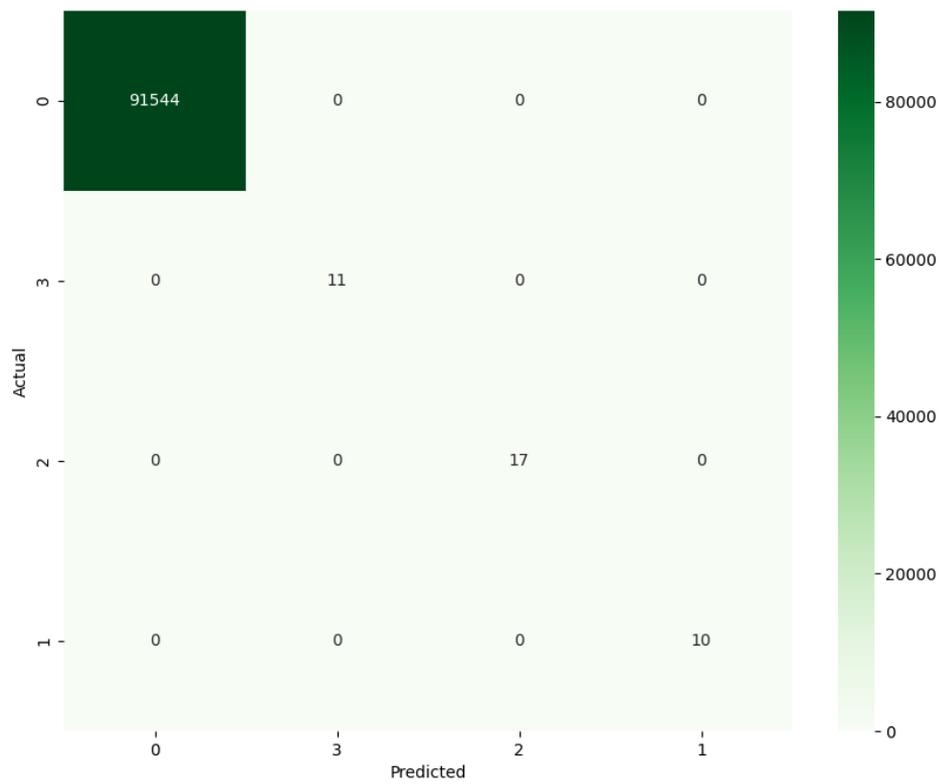
Fuente: Propia

Figura 36. Matriz de confusión para el entrenamiento – iteración 05



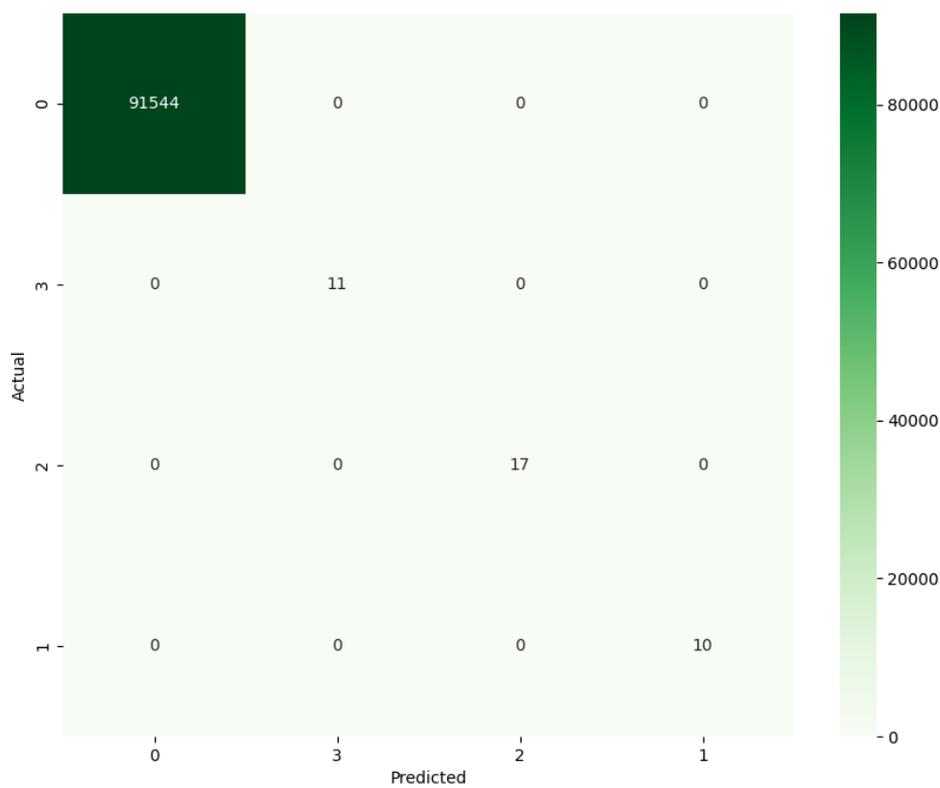
Fuente: Propia

Figura 37. Matriz de confusión para el entrenamiento – iteración 06



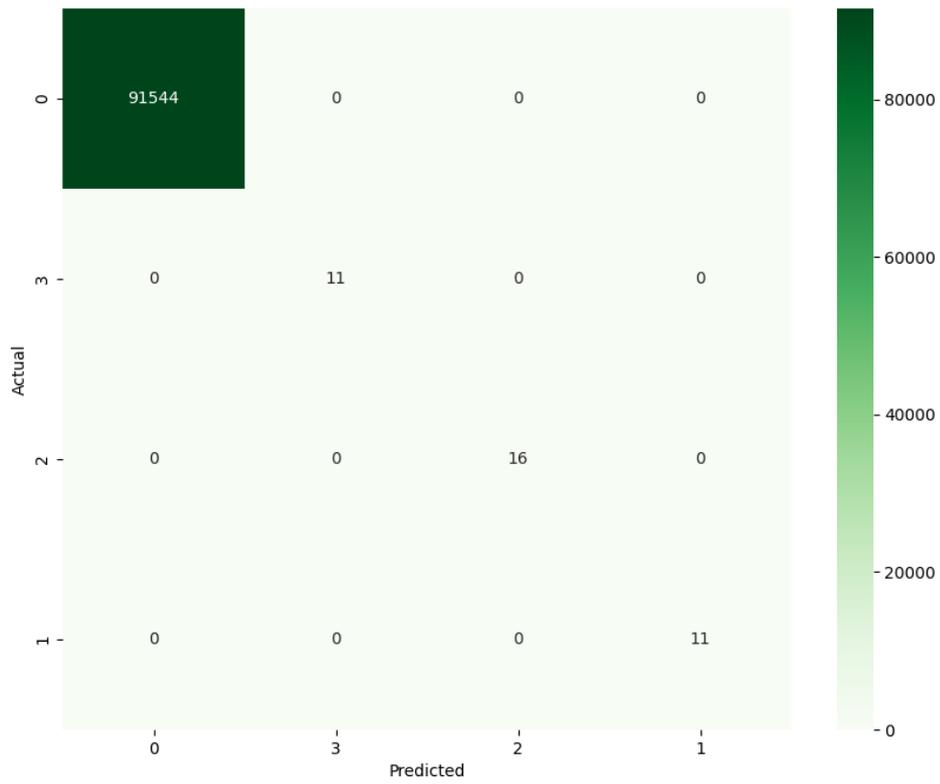
Fuente: Propia

Figura 38. Matriz de confusión para el entrenamiento – iteración 07



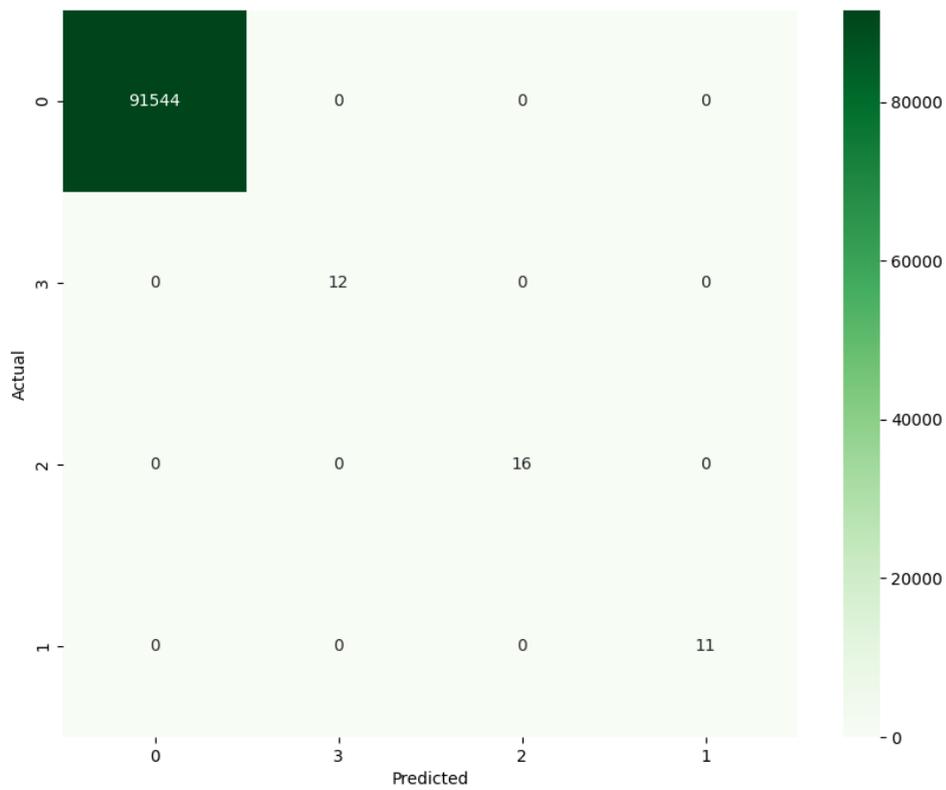
Fuente: Propia

Figura 39. Matriz de confusión para el entrenamiento – iteración 08



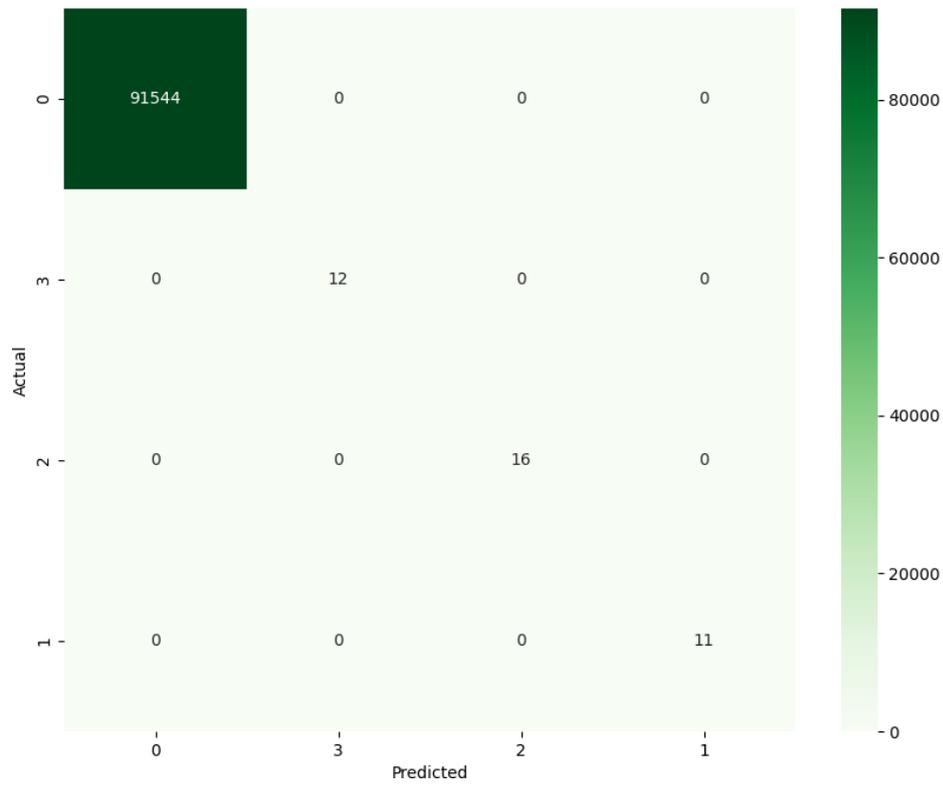
Fuente: Propia

Figura 40. Matriz de confusión para el entrenamiento – iteración 09



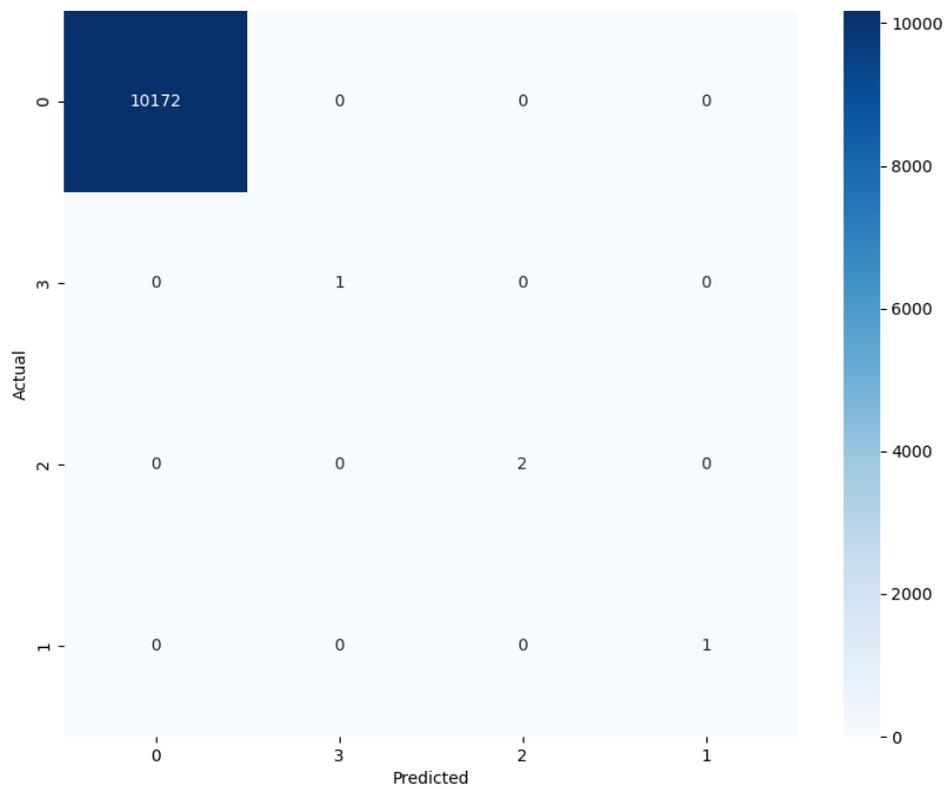
Fuente: Propia

Figura 41. Matriz de confusión para el entrenamiento – iteración 10



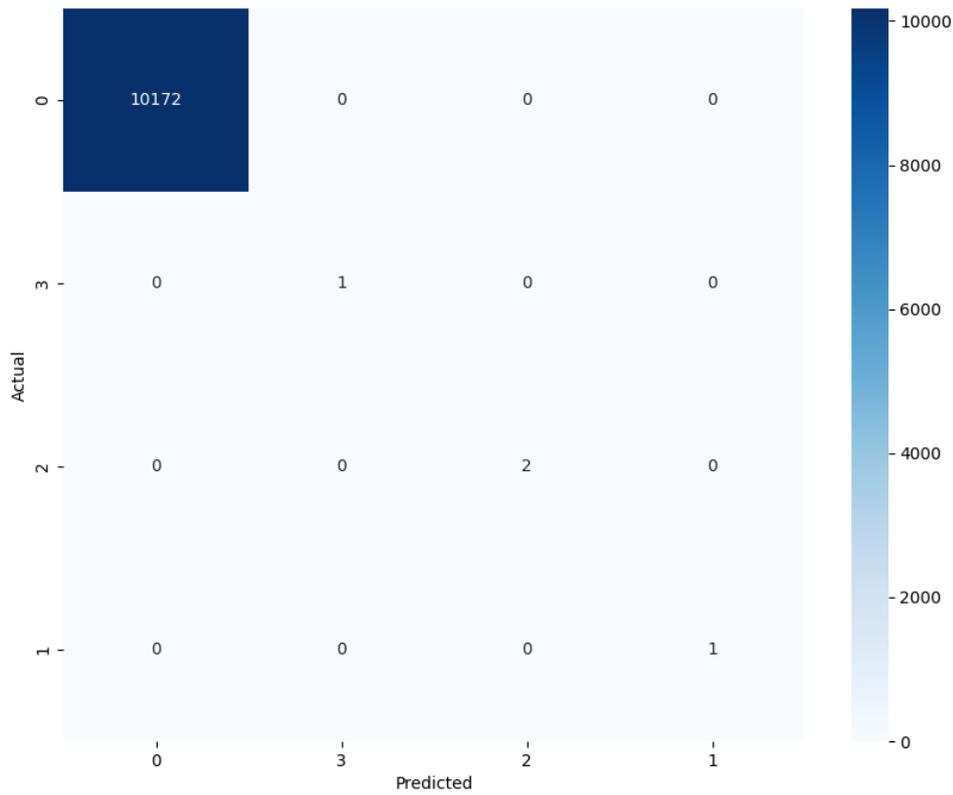
Fuente: Propia

Figura 42. Matriz de confusión para la prueba – iteración 01



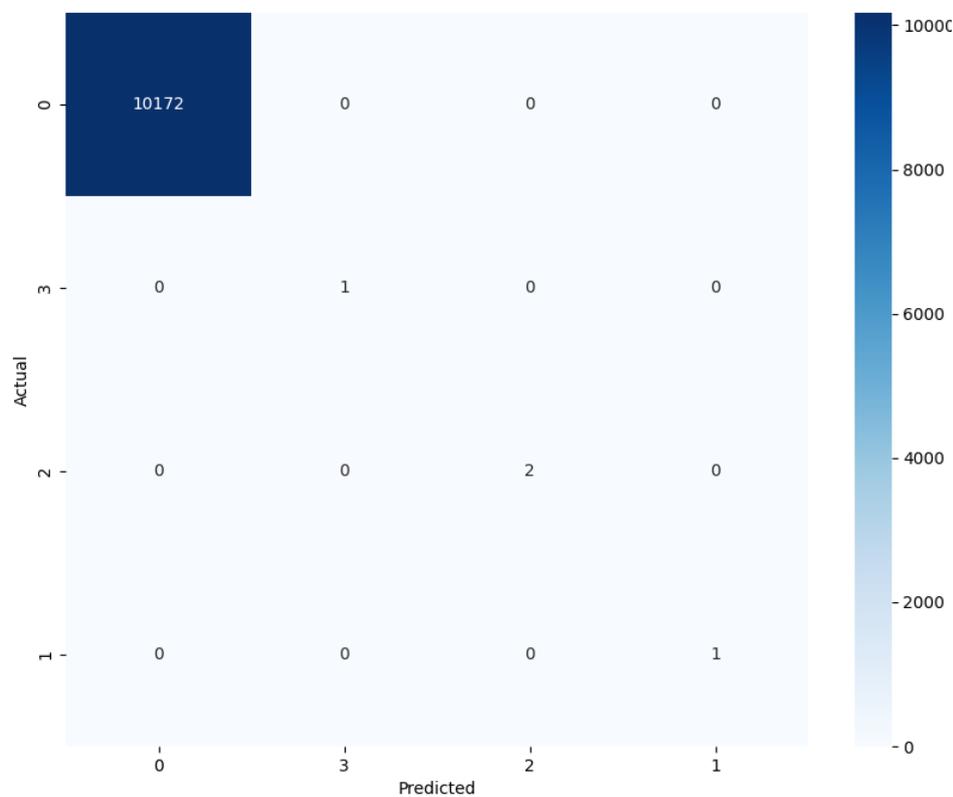
Fuente: Propia

Figura 43. Matriz de confusión para la prueba – iteración 02



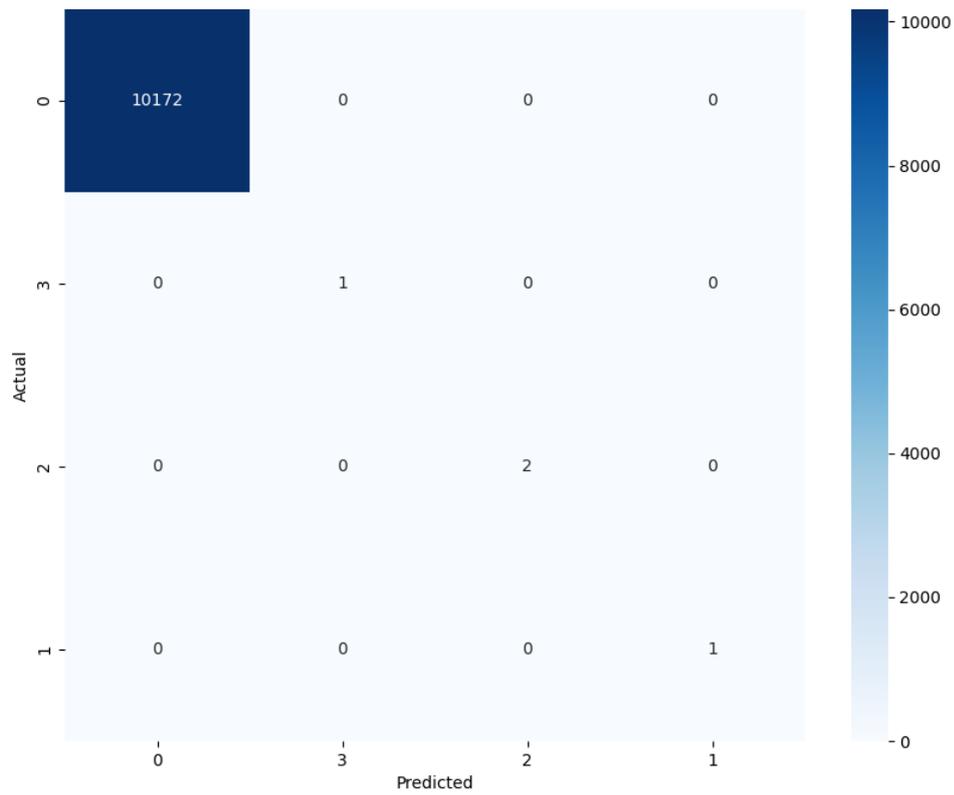
Fuente: Propia

Figura 44. Matriz de confusión para la prueba – iteración 03



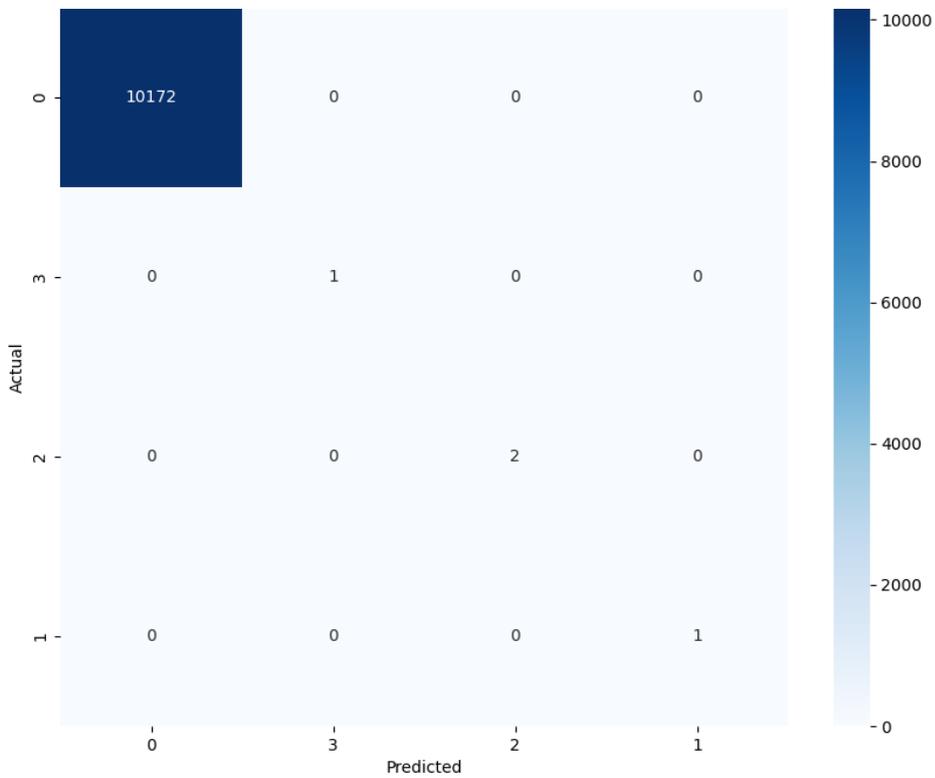
Fuente: Propia

Figura 45. Matriz de confusión para la prueba – iteración 04



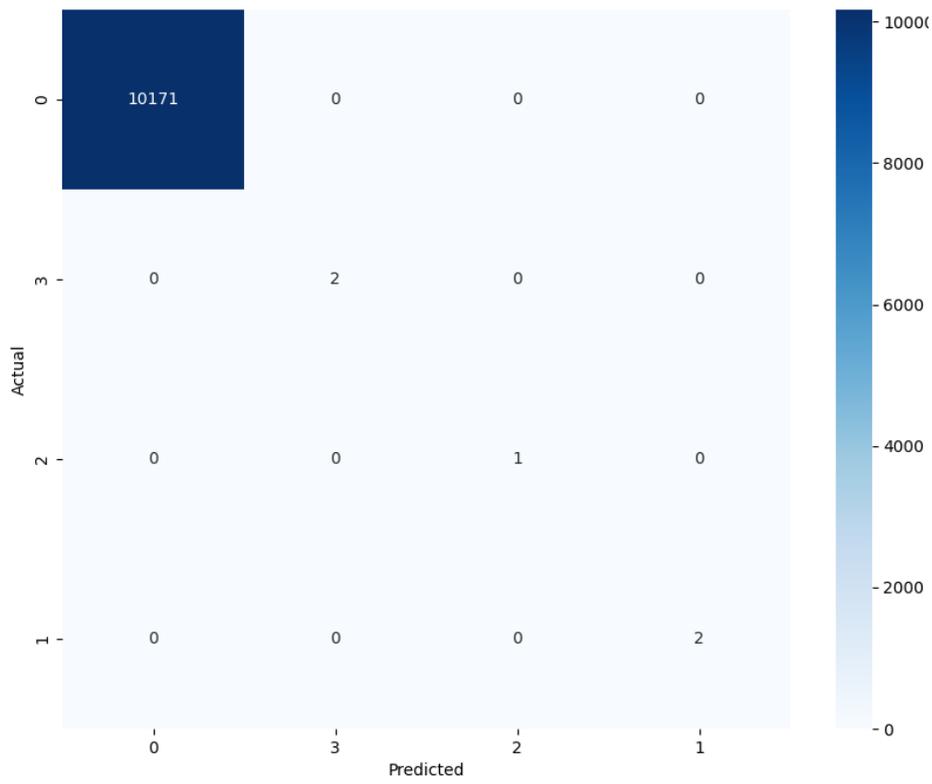
Fuente: Propia

Figura 46. Matriz de confusión para la prueba – iteración 05



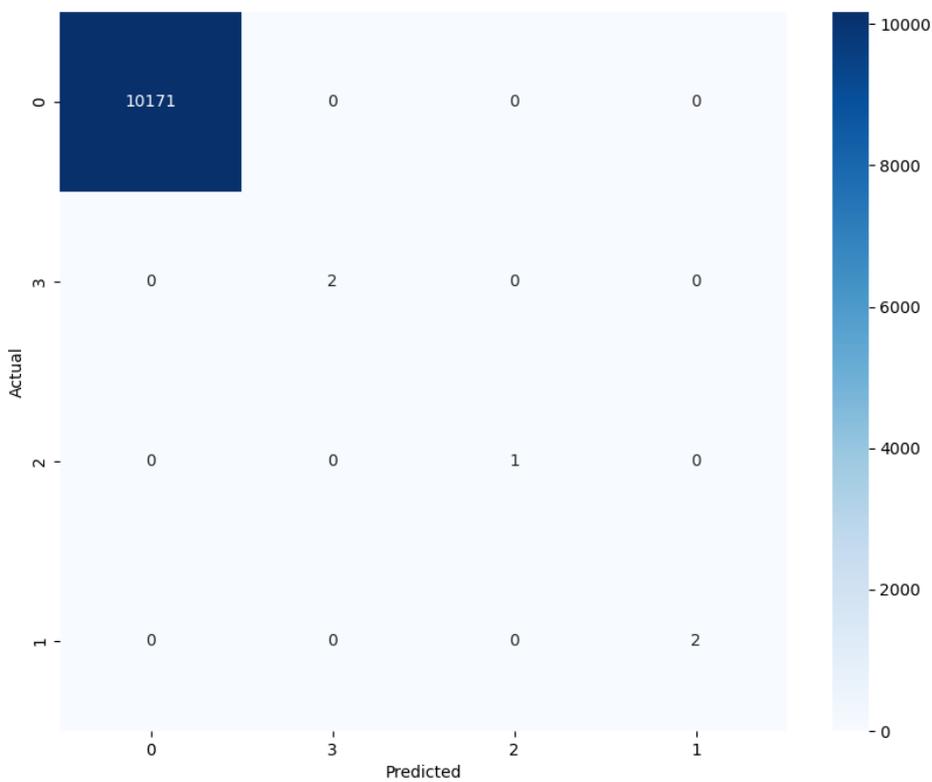
Fuente: Propia

Figura 47. Matriz de confusión para la prueba – iteración 06



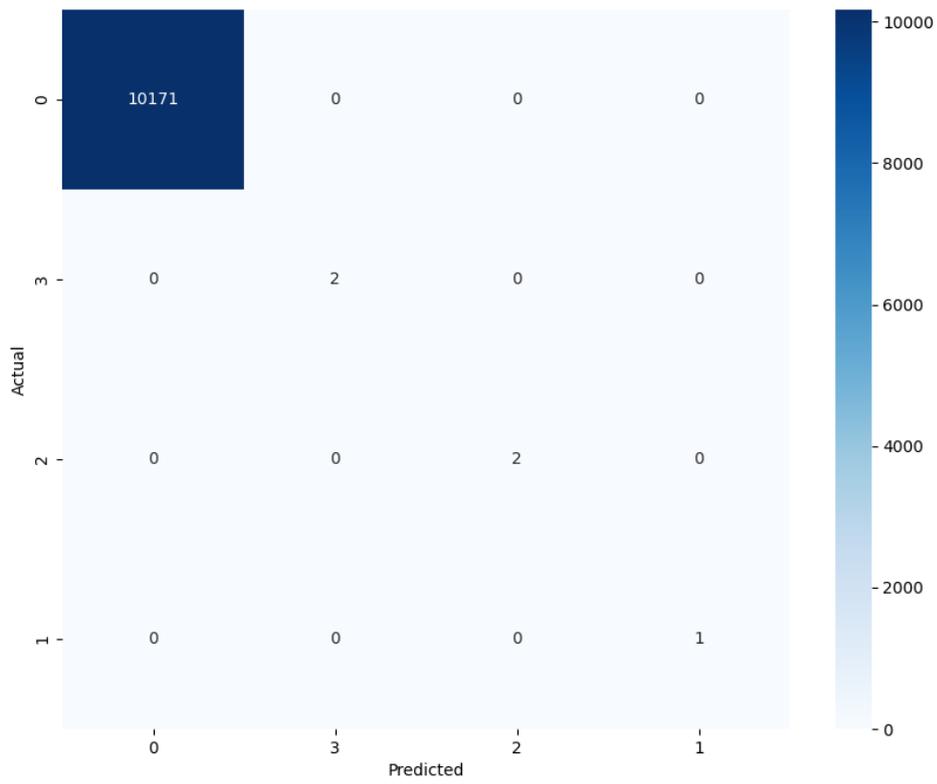
Fuente: Propia

Figura 48. Matriz de confusión para la prueba – iteración 07



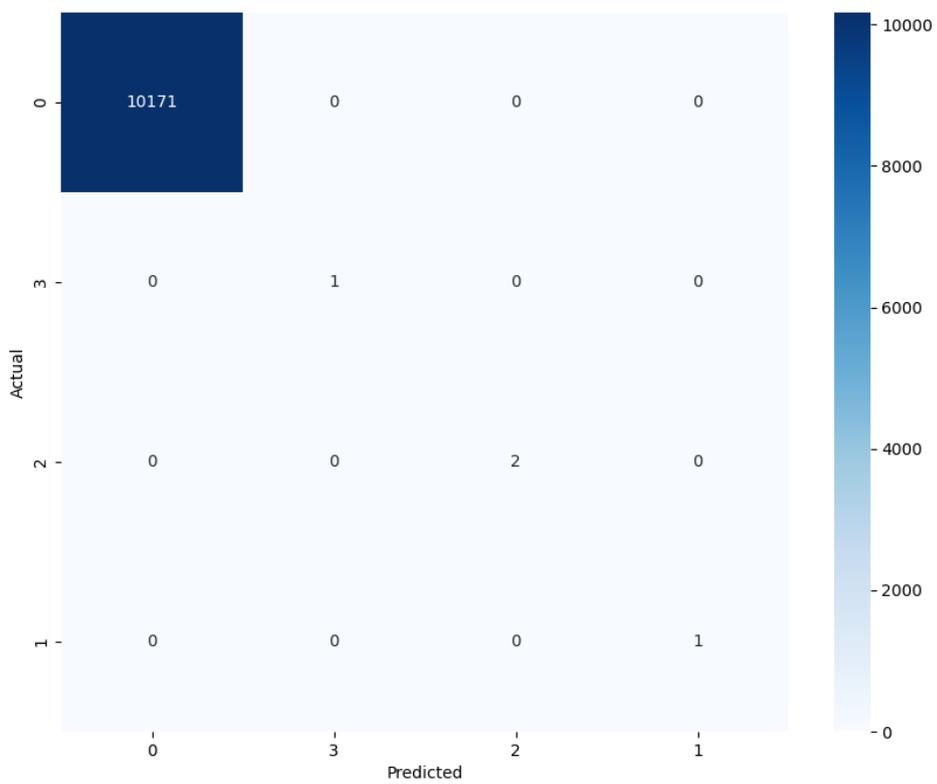
Fuente: Propia

Figura 49. Matriz de confusión para la prueba – iteración 08



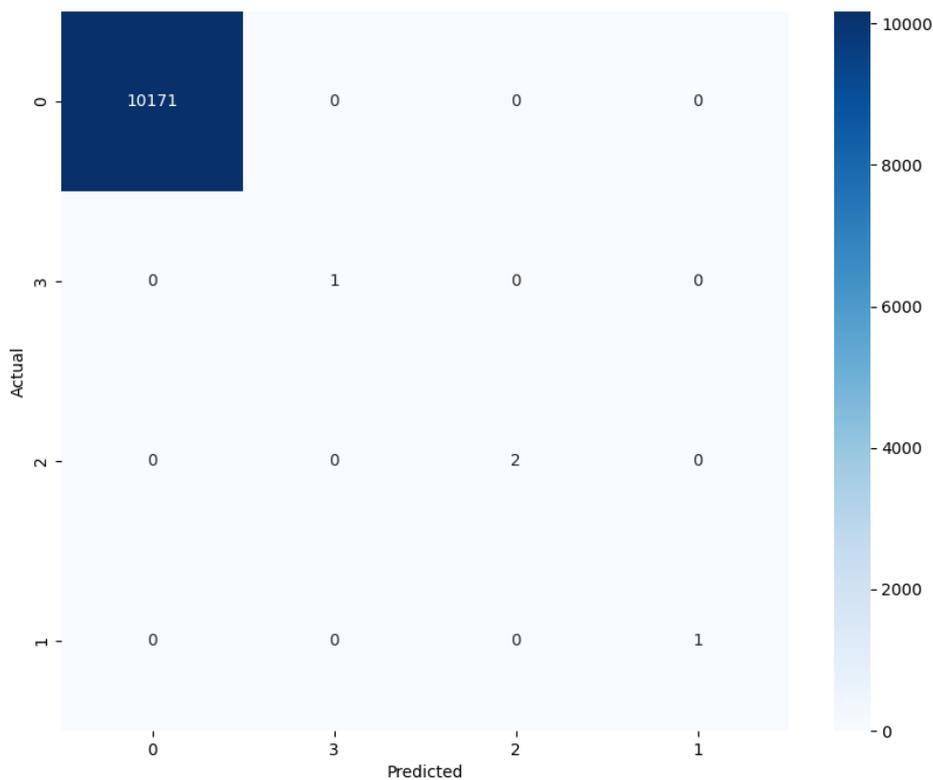
Fuente: Propia

Figura 50. Matriz de confusión para la prueba – iteración 09



Fuente: Propia

Figura 51. Matriz de confusión para la prueba – iteración 10



Fuente: Propia

Tabla 13. Métricas de desempeño en las iteraciones

Iteraciones	Accuracy	Precision	Recall	F1 Score
Iteración 01	1	1	1	1
Iteración 02	1	1	1	1
Iteración 03	1	1	1	1
Iteración 04	1	1	1	1
Iteración 05	1	1	1	1
Iteración 06	1	1	1	1
Iteración 07	1	1	1	1
Iteración 08	1	1	1	1
Iteración 09	1	1	1	1
Iteración 10	1	1	1	1

Nota: Fuente propia

En la Tabla 13, se observa que, en las 10 iteraciones realizadas durante la validación cruzada estratificadas, se ha obtenido un valor de precisión igual a 1 lo que indica que el modelo ha identificado correctamente todas las anomalías, sin falsos positivos. Un valor de sensibilidad igual a 1 significa que el modelo ha detectado todas las anomalías presentes en los datos. Un valor de exactitud igual a 1 refleja que el modelo

distingue correctamente entre datos con y sin anomalías. Finalmente, un F1-Score igual a 1 demuestra un equilibrio perfecto entre precisión y sensibilidad, confirmando que el modelo funciona de manera óptima en la detección y clasificación de anomalías.

4.4.4. Resultados de verificar el modelo basado en el árbol de decisiones

El modelo obtenido durante el entrenamiento y prueba se someterá a una verificación utilizando el conjunto de datos DataSet 02; el cual son datos medidos durante julio de 2024. Para comparar resultados, en la Tabla 15, se observa los datos etiquetados con anomalías de manera manual y guardado en un archivo independiente. Luego el DataSet 02 sin etiquetas será sometido a un código para verificar si el modelo generado durante el entrenamiento puede identificar esas mismas anomalías. Finalmente, en la Tabla 15, se verifica que, si se han identificado los mismos datos con las mismas anomalías, confirmando la eficiencia del modelo.

Tabla 14. Datos etiquetados manualmente del DataSet 02

X1 Current	X2 Current	X3 Current	X1 Voltage	X2 Voltage	X3 Voltage	Codigo	Etiqueta
0	0	0	128.6	129.4	128.7	Anomalia01	1
0	0	0	129.1	129.9	129.2	Anomalia01	1
0	0	0	129.1	129.9	129.3	Anomalia01	1
0	0	0	128.5	129.3	128.6	Anomalia01	1
0	0	0	128.5	129.2	128.7	Anomalia01	1
0	0	0	128.4	129.1	128.5	Anomalia01	1
0	0	0	128.4	129.2	128.4	Anomalia01	1
1350	1588.17	1553	119.4	120.23	119.58	Anomalia03	3
1800.7	1754.55	1855.4	128.2	128.8	128.3	Anomalia02	2
1554.9	1758.77	1493.22	128.4	128.9	128.5	Anomalia02	2
1600.7	1825.73	1541.25	128.4	128.9	128.4	Anomalia02	2
1654.66	1868.89	1499.99	128.3	128.7	128.4	Anomalia02	2
1655.33	1853.53	1576.2	127.9	128.5	128.1	Anomalia02	2
1700.2	1897.66	1568.94	127.7	128.2	127.8	Anomalia02	2
1850.4	2076.83	1656.37	127.9	128.4	127.9	Anomalia02	2
1850.6	2137.83	1788.54	127.5	128	127.6	Anomalia02	2
1850.4	2117.88	1727.66	127.9	128.4	127.9	Anomalia02	2
1700.4	2026.37	1719.26	128	128.4	127.8	Anomalia02	2
1700.4	1905.66	1822.42	127.9	128.4	127.9	Anomalia02	2

Nota: Fuente propia

Tabla 15. Datos detectados como anomalías en el DataSet 02

X1 Current	X2 Current	X3 Current	X1 Voltage	X2 Voltage	X3 Voltage	Predicción_01	Predicción_02	Predicción_03	Predicción_04	Predicción_05	Predicción_06	Predicción_07	Predicción_08	Predicción_09	Predicción_10
0	0	0	128.6	129.4	128.7	1	1	1	1	1	1	1	1	1	1
0	0	0	129.1	129.9	129.2	1	1	1	1	1	1	1	1	1	1
0	0	0	129.1	129.9	129.3	1	1	1	1	1	1	1	1	1	1
0	0	0	128.5	129.3	128.6	1	1	1	1	1	1	1	1	1	1
0	0	0	128.5	129.2	128.7	1	1	1	1	1	1	1	1	1	1
0	0	0	128.4	129.1	128.5	1	1	1	1	1	1	1	1	1	1
0	0	0	128.4	129.2	128.4	1	1	1	1	1	1	1	1	1	1
1350	1588.17	1553	119.4	120.23	119.58	3	3	3	3	3	3	3	3	3	3
1800.7	1754.55	1855.4	128.2	128.8	128.3	2	2	2	2	2	2	2	2	2	2
1554.9	1758.77	1493.22	128.4	128.9	128.5	2	2	2	2	2	2	2	2	2	2
1600.7	1825.73	1541.25	128.4	128.9	128.4	2	2	2	2	2	2	2	2	2	2
1654.66	1868.89	1499.99	128.3	128.7	128.4	2	2	2	2	2	2	2	2	2	2
1655.33	1853.53	1576.2	127.9	128.5	128.1	2	2	2	2	2	2	2	2	2	2
1700.2	1897.66	1568.94	127.7	128.2	127.8	2	2	2	2	2	2	2	2	2	2
1850.4	2076.83	1656.37	127.9	128.4	127.9	2	2	2	2	2	2	2	2	2	2
1850.6	2137.83	1788.54	127.5	128	127.6	2	2	2	2	2	2	2	2	2	2
1850.4	2117.88	1727.66	127.9	128.4	127.9	2	2	2	2	2	2	2	2	2	2
1700.4	2026.37	1719.26	128	128.4	127.8	2	2	2	2	2	2	2	2	2	2
1700.4	1905.66	1822.42	127.9	128.4	127.9	2	2	2	2	2	2	2	2	2	2

Nota: Fuente propia

4.5. Resultados de la evaluación económica

La inversión está asociada a la compra del equipo de Monitoreo del transformador de Distribución (DTM) y los costos de mano de obra (instalación), software, hardware y servicio técnico para el mantenimiento y actualización de la plataforma web a utilizar. Para evaluar la factibilidad del proyecto, se va a tomar como referencia la instalación de estos equipos en 20 transformadores y estos costos de inversión quedan descritos en la Tabla 16, en donde, el costo de inversión es de S/ 330 000.00.

Tabla 16. Costos de inversión del proyecto de instalación de 20 DTM

Descripción	Unidad	C.U.	Cantidad	C.T.
Suministro del DTM	unidad	S/12,000.00	20	S/240,000.00
Instalación del DTM	unidad	S/2,500.00	20	S/50,000.00
Hardware	unidad	S/10,000.00	1	S/10,000.00
Software	anual	S/5,000.00	1	S/5,000.00
Soporte Técnico	anual	S/5,000.00	5	S/25,000.00
Inversión Total				S/330,000.00

Nota: Fuente propia

Ahora procedemos a calcular los costos que se dejarían de incurrir (Ahorro) en caso se implemente este proyecto. La implementación del presente trabajo de investigación va a permitir contar con un monitoreo y control de los valores de corriente y voltaje en baja tensión. Al tener estos datos en línea va a reducir de manera significativa los tiempos de respuesta para la atención de fallas producto de la materialización de las anomalías.

Se ha realizado el análisis de los casos en que se ha generado la interrupción total del suministro de energía eléctrica a todos los clientes de baja tensión del transformador, para ello, se ha realizado el cálculo promedio ponderado entre los clientes conectados a subestaciones convencionales, aéreas y compactas. Además, teniendo un tiempo promedio de respuesta de 10 horas y una cantidad de eventos anual de 06 casos. Y considerando un costo de energía de 0.70 soles por kWh. Esto queda representado en la Tabla 17. El monto que se obtuvo para un transformador es de S/9 072.00; entonces para 20 transformadores sería de 181 440 por año.

Tabla 17. Ahorro promedio anual por transformador

Descripción	Cantidad	Unidad
Clientes afectados	180	clientes
Tiempo de interrupción	12	horas
Cantidad de eventos año	6	eventos
Costo energía	S/0.70	por kwh
Costo Total	S/9,072.00	

Nota: Fuente propia.

En la Tabla 18, se muestra un análisis financiero de un proyecto basado en un flujo de caja proyectado a cinco años. Inicialmente, se realiza una inversión significativa equivalente a S/ 330 000 en el año cero, que es necesaria para poner en marcha el proyecto. A partir del año 1 hasta el año 5, se proyectan ahorros anuales constantes de S/. 181,440.00, que representan los beneficios del proyecto en términos de reducción de costos operativos. Además, hay que adicionar los gastos no reembolsables representados por la depreciación de los equipos adquiridos para el proyecto durante los 05 años de proyección, los pagos por impuestos a la renta (29.5%) asociados al proyecto cada año y el ajuste por gastos no reembolsables representado por la depreciación de los equipos, obteniendo como resultado que cada año se tenga un flujo neto de S/. 112,870.20 por año.

Tabla 18. Flujo de caja durante 05 años

Descripción	0	1	2	3	4	5
Ingresos afectos a impuestos						
Egresos afectos a impuestos (ahorro)		-S/181,440.00	-S/181,440.00	-S/181,440.00	-S/181,440.00	-S/181,440.00
Δ EBITDA		-S/181,440.00	-S/181,440.00	-S/181,440.00	-S/181,440.00	-S/181,440.00
Gastos no reembolsables (depreciación)		-S/51,000.00	-S/51,000.00	-S/51,000.00	-S/51,000.00	-S/51,000.00
Resultado antes de impuestos		-S/232,440.00	-S/232,440.00	-S/232,440.00	-S/232,440.00	-S/232,440.00
Impuesto de la Renta (29.5%)		S/68,569.80	S/68,569.80	S/68,569.80	S/68,569.80	S/68,569.80
Resultado después de impuesto		-S/163,870.20	-S/163,870.20	-S/163,870.20	-S/163,870.20	-S/163,870.20
Ajuste por gastos no reembolsables		S/51,000.00	S/51,000.00	S/51,000.00	S/51,000.00	S/51,000.00
Resultado operacional neto		-S/112,870.20	-S/112,870.20	-S/112,870.20	-S/112,870.20	-S/112,870.20
Inversión	-S/330,000.00					
Ahorro		S/112,870.20	S/112,870.20	S/112,870.20	S/112,870.20	S/112,870.20
Tasa de descuento	12%					
VAN	S/76,871.81					
TIR	21.036%					

Nota: Fuente propia

Luego de ello, se procede al cálculo del Valor Actual Neto (VAN) y la Tasa Interna de Retorno (TIR), que son métricas clave para evaluar la viabilidad financiera del proyecto. El VAN se calcula descontando los flujos de caja futuros a una tasa del 12%, lo que permite evaluar si los ingresos proyectados superan la inversión inicial. Un VAN positivo, como se

muestra en el cuadro, indica que el proyecto no solo recupera la inversión inicial, sino que también genera un beneficio adicional. Por su parte, la TIR es la tasa que hace que el VAN sea igual a cero, lo que en este caso indicaría la rentabilidad anual promedio esperada del proyecto. Si la TIR resultante es superior al 12%, el proyecto se considera rentable, lo que sugiere que la inversión no solo es recuperable, sino que también ofrece un retorno atractivo. En conjunto, estos resultados financieros demuestran que el proyecto es viable y rentable, justificando su implementación

4.6. Discusión

En esta investigación se desarrolló un modelo de aprendizaje automático basado en un árbol de decisiones para detectar anomalías en un transformador de una empresa de distribución eléctrica. La metodología incluyó la parametrización de anomalías, el etiquetado manual de datos históricos de corriente y voltaje, el entrenamiento del modelo con esos datos, y la verificación del modelo utilizando un conjunto de datos independiente. Además, se evaluó la rentabilidad de la implementación del proyecto mediante indicadores financieros como el VAN y la TIR.

Los resultados obtenidos demuestran que el modelo de árbol de decisiones implementado posee una alta precisión, sensibilidad y exactitud en la detección de anomalías, superando los umbrales de aprobación establecidos (superiores a 0.9). Durante la verificación del modelo entrenado, el modelo logró identificar correctamente las anomalías, confirmando su efectividad para aplicaciones en tiempo real. La evaluación financiera indicó que el proyecto es económicamente viable, con un VAN igual a S/ 76 871.81 y un TIR igual al 21.036%, lo cual indica que es rentable.

Al comparar con estudios previos como el de (Ramesh et al., 2022)⁶, que utilizaron Random Forest para identificar anomalías y enfrentaron problemas debido a la escasez de datos anómalos, o el de (Liu et al., 2023)⁴, que emplearon clustering jerárquico para un aprendizaje no supervisado antes de aplicar un árbol de decisiones, observamos que estos

enfoques usaron algoritmos más complejos. Nuestro enfoque, basado en un árbol de decisiones junto con una parametrización de anomalías, logró resultados competitivos en términos de precisión y exactitud.

Aunque los resultados obtenidos son satisfactorios, futuras investigaciones podrían enfocarse en integrar el conocimiento técnico del personal experimentado durante el entrenamiento del algoritmo. Así como en parametrizar más anomalías y utilizar un conjunto de datos que abarque varios transformadores y un período de medición promedio de 12 meses. Esto permitiría cubrir un mayor rango de anomalías y mantener la precisión del modelo cuando se ingresen nuevos parámetros de anomalías. También sería beneficioso introducir variables independientes adicionales como presión, temperatura y vibración del transformador para predecir anomalías que no pueden ser detectadas únicamente con los valores de corriente y voltaje en baja tensión. Además, la exploración de técnicas de aprendizaje profundo podría ser valiosa para manejar datos más complejos y detectar anomalías más difíciles de identificar. Finalmente, se sugiere expandir el estudio a otros componentes críticos de la red de distribución para evaluar la aplicabilidad y efectividad del modelo en un contexto más amplio, lo que podría fortalecer la continuidad y confiabilidad del servicio eléctrico en la empresa de distribución.

CONCLUSIONES

Se demuestra la efectividad y precisión del modelo basado en árbol de decisiones, demostrando ser una herramienta eficaz y precisa para optimizar el mantenimiento predictivo y preventivo. La aplicación del modelo contribuye en reducir los tiempos de procesamiento y atención de anomalías por parte del personal técnico. Además, impulsa la transición hacia una gestión más eficiente y proactiva de los recursos eléctricos y el uso de tecnologías de la cuarta revolución industrial, como el aprendizaje automático. El aprendizaje automático es crucial para elevar los estándares de confiabilidad en el sector de distribución eléctrica y asegurar el cumplimiento de normativas tanto a nivel nacional como internacional.

La parametrización de anomalías en este estudio, basada en modos de falla específicos del transformador de distribución, ha permitido identificar de manera precisa las condiciones que podrían afectar la calidad del suministro eléctrico. La metodología aplicada, que incluye la clasificación de anomalías por tipo y la asignación de etiquetas correspondientes, garantiza que los datos recopilados reflejen fielmente las situaciones críticas que el modelo debe detectar. Este enfoque es fundamental para asegurar que el algoritmo de aprendizaje automático tenga un entrenamiento robusto, capaz de identificar patrones anómalos con alta precisión, lo que se traduce en una mejora significativa en la gestión de la operación del transformador.

El proceso de entrenamiento del modelo basado en el algoritmo de árbol de decisiones, utilizando datos etiquetados cuidadosamente para reflejar tanto condiciones normales como anómalas, ha mostrado ser altamente efectivo. La utilización de métricas de desempeño como la precisión, la sensibilidad, y el F1-score ha permitido validar el modelo con un alto grado de confiabilidad. Este entrenamiento riguroso asegura que el modelo no solo se adapte a un conjunto de datos para el entrenamiento y prueba, sino que tenga la capacidad de generalizar su aprendizaje con otro conjunto de datos. Demostrando

la capacidad del modelo de predecir anomalías en nuevos conjuntos de datos, contribuyendo así a una detección temprana y eficaz de fallas en el sistema eléctrico.

La evaluación económica del proyecto, realizada mediante un análisis del Valor Actual Neto (VAN) y la Tasa Interna de Retorno (TIR), ha confirmado la viabilidad y rentabilidad de la implementación del modelo de aprendizaje automático en la detección de anomalías en transformadores de distribución. Los resultados del flujo de caja proyectado a cinco años, considerando una tasa de descuento del 12%, evidencian que los ahorros generados por la reducción en costos de interrupciones y el mantenimiento preventivo superan significativamente la inversión inicial. Este análisis económico respalda la adopción de tecnologías avanzadas como una decisión estratégica que no solo mejora la operación del sistema eléctrico, sino que también garantiza un retorno positivo sobre la inversión.

RECOMENDACIONES

Se recomienda la implementación del modelo de aprendizaje automático basado en árboles de decisión en la red de distribución eléctrica a nivel nacional. Su capacidad demostrada para detectar anomalías con alta precisión y su bajo requerimiento de recursos computacionales lo hacen ideal para su integración en el sistema operativo existente, contribuyendo a la mejora de la continuidad del servicio eléctrico.

Se recomienda explorar la inclusión de variables adicionales como presión, temperatura y vibración del transformador en futuros desarrollos del modelo. Estas variables podrían ayudar a detectar anomalías que no son evidentes a partir de los valores de corriente y voltaje, ampliando la aplicabilidad del modelo a una gama más amplia de condiciones operativas.

Se recomienda complementar a la inclusión de nuevas variables con la ampliación del conjunto de datos, incorporando datos de varios transformadores y extendiendo el período de medición a 12 meses. Esto permitirá que el modelo cubra un mayor rango de anomalías, y así mantener su capacidad de detección en diferentes condiciones operativas.

Para mejorar la precisión y reducir los tiempos de entrenamiento, se recomienda integrar el conocimiento técnico del personal experimentado en la fase de entrenamiento del modelo. La experiencia práctica del personal puede proporcionar conocimiento valioso que complementen los datos históricos y optimicen el rendimiento del modelo.

La viabilidad económica del proyecto fue confirmada mediante un análisis financiero que mostró un VAN positivo y una TIR superior a la tasa de descuento utilizada. Entonces, se recomienda la implementación del modelo en un contexto más amplio, potencialmente a nivel nacional.

En caso de requerir hacer un pronóstico de la demanda en un transformador de distribución, se recomienda recopilar las variables de tensión y de corriente en baja tensión durante un periodo no menor de 12 meses. Además, de incluir otros parámetros que impactan en dicho pronóstico, como las fallas que ocurran en la red, el ingreso de nuevos clientes, modificaciones de la red. Realizar un pronóstico de la demanda permitirá implementar nuevos proyectos eléctricos en las redes de media tensión y de detectar hurtos de energía.

BIBLIOGRAFÍA

- Acuña, J. (2022). *Ingeniería de Confiabilidad* (Segunda). Editorial Tecnológica de Costa Rica.
- Amr, T. (2020). *Hands-on machine learning with scikit-learn and scientific Python toolkits : a practical guide to implementing supervised and unsupervised machine learning algorithms in Python*. Packt Publishing Ltd.
- Ayyadevara, V. K. (2018). Pro Machine Learning Algorithms: A Hands-On Approach to Implementing Algorithms in Python and R. In *Pro Machine Learning Algorithms: A Hands-On Approach to Implementing Algorithms in Python and R*. Apress Media LLC. <https://doi.org/10.1007/978-1-4842-3564-5>
- Berzal, F. (2018). *REDES NEURONALES & DEEP LEARNING*. EUG, Editorial Universidad de Granada.
- Bishop, C. M., & Bishop, H. (2024). *Foundations and Concepts Deep Learning* (Primera). Springer.
- Chatterjee, I. (2022). *Machine Learning and Its Application: A Quick Guide for Beginners* (1st ed.). Betham Science Publishers Ltd.
- De Benedetti, M., Leonardi, F., Messina, F., Santoro, C., & Vasilakos, A. (2018). Anomaly detection and predictive maintenance for photovoltaic systems. *Neurocomputing*, 310, 59–68. <https://doi.org/10.1016/J.NEUCOM.2018.05.017>
- dos Santos, G. M., de Aquino, R. R. B., & Lira, M. M. S. (2018). Thermography and artificial intelligence in transformer fault detection. *Electrical Engineering*, 100(3), 1317–1325. <https://doi.org/10.1007/s00202-017-0595-2>

- Keynia, F., Mirhosseini, M., Heydari, A., & Fekih, A. (2022). A budget allocation and programming-based RCM approach to improve the reliability of power distribution networks. *Energy Reports*, 8, 5591–5602. <https://doi.org/10.1016/j.egy.2022.04.029>
- Kulkarni, S. V., & Khaparde, S. A. (2013). *Transformer Engineering: Design, Technology, and Diagnostics* (Second). CRC Press.
- Liu, X., Li, J., Shao, L., Liu, H., Ren, L., & Zhu, L. (2023). Transformer Fault Early Warning Analysis Based on Hierarchical Clustering Combined with Decision Trees. *Energies*, 16(3). <https://doi.org/10.3390/en16031168>
- Ma, H., Yang, P., Wang, F., Wang, X., Yang, D., & Feng, B. (2023). Short-Term Heavy Overload Forecasting of Public Transformers Based on Combined LSTM-XGBoost Model. *Energies*, 16(3). <https://doi.org/10.3390/en16031507>
- Manivannan, K. M., Langari, R., & Narayanan, K. (2012). *POWER DISTRIBUTION SYSTEM EVENT CLASSIFICATION USING FUZZY LOGIC A Dissertation*. <https://doi.org/https://hdl.handle.net/1969.1/158893>
- Mirhosseini, M., Heydari, A., Astiaso Garcia, D., Mancini, F., & Keynia, F. (2022). Reliability based maintenance programming by a new index for electrical distribution system components ranking. *Optimization and Engineering*, 23(4), 2315–2333. <https://doi.org/10.1007/s11081-022-09767-8>
- Mirsaeedi, H., Fereidunian, A., Mohammadi-Hosseininejad, S. M., & Lesani, H. (2018). Electricity Distribution System Maintenance Budgeting: A Reliability-Centered Approach. *IEEE Transactions on Power Delivery*, 33(4), 1599–1610. <https://doi.org/10.1109/TPWRD.2017.2772202>
- Misari, A. R., Leite, J. B., Piasson, D., & Mantovani, J. R. S. (2020). Reliability-Centered Maintenance Task Planning for Overhead Electric Power Distribution Networks.

Journal of Control, Automation and Electrical Systems, 31(5), 1278–1287.
<https://doi.org/10.1007/s40313-020-00606-8>

Moldovan, A. M., & Buzdugan, M. I. (2023). Prediction of Faults Location and Type in Electrical Cables Using Artificial Neural Network. *Sustainability (Switzerland)*, 15(7).
<https://doi.org/10.3390/su15076162>

Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python A GUIDE FOR DATA SCIENTISTS* (First). O'Reilly Media, Inc.

Organismo Supervisor de la Inversión en Energía y Minería. (2009). *Resolución de Consejo Directivo OSINERGMIN N° 228-2009-OS/CD*.
<https://www.osinergmin.gob.pe/newweb/uploads/publico/RCD.228.2009.OS.CD.pdf>.

Phillips, D. (2020). *Machine Learning: Jump on the Way to the Future, Discover Artificial Intelligence and Data Science. Maximize your Business in the Modern World Mastering Deep Learning, Python and Algorithms*.

Ramesh, J., Shahriar, S., Al-Ali, A. R., Osman, A., & Shaaban, M. F. (2022). Machine Learning Approach for Smart Distribution Transformers Load Monitoring and Management System. *Energies*, 15(21). <https://doi.org/10.3390/en15217981>

Sapag, N., & Sapag Reinaldo. (2012). *Preparación y Evaluación de Proyectos* (L. Solano, Ed.; Quinta). McGraw-Hill Interamericana.

Veerakumar, N., Četenović, D., Kongurai, K., Popov, M., Jongepier, A., & Terzija, V. (2023). PMU-based Real-time Distribution System State Estimation Considering Anomaly Detection, Discrimination and Identification. *International Journal of Electrical Power & Energy Systems*, 148, 108916. <https://doi.org/10.1016/j.ijepes.2022.108916>

- Wang, Z., Ding, Y., & Zhang, T. (2022). Distribution Network Anomaly Detection Algorithm Based on VAE. *Proceedings - 2022 11th International Conference of Information and Communication Technology, ICTech 2022*, 84–87. <https://doi.org/10.1109/ICTech55460.2022.00025>
- Xiang, B., Liu, Z., & Zhang, K. (2020). Flagging Implausible Inspection Reports of Distribution Transformers via Anomaly Detection. *IEEE Access*, 8, 75798–75808. <https://doi.org/10.1109/ACCESS.2020.2989330>
- Xiao, F., Leng, X., Ye, K., Hu, Y., Li, X., & Zhu, L. (2018). The Research of Anomaly Detection Method for Transformer Oil Temperature Based on Hybrid Model of Non-Supervised Learning and Decision Forests. *IOP Conference Series: Earth and Environmental Science*, 192(1). <https://doi.org/10.1088/1755-1315/192/1/012020>
- Yebra, J. (2009). *Sistemas Eléctricos de Distribución* (Primera Edición). Reverté Ediciones.

ANEXO A. MATRIZ DE CONSISTENCIA

TITULO: Modelo de aprendizaje automático para detectar anomalías en un transformador de una empresa de distribución eléctrica				
PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLES	METODOLOGÍA
<u>Problema general</u>	<u>Objetivo general</u>	<u>Hipótesis general</u>	<u>Variable independiente (X)</u>	<u>Tipo de investigación:</u> Aplicada
¿De qué manera el aprendizaje automático permitirá detectar anomalías en un transformador de una empresa de distribución eléctrica?	Detectar anomalías mediante el aprendizaje automático en un transformador de una empresa de distribución eléctrica	El aprendizaje automático permite detectar anomalías en un transformador de una empresa de distribución eléctrica	Aprendizaje automático	<u>Nivel de investigación:</u> Explicativo <u>Método:</u> Detección <u>Diseño:</u> Cuantitativo
<u>Problemas específicos</u>	<u>Objetivos específicos</u>	<u>Hipótesis específicas</u>	<u>Variable dependiente (Y)</u>	<u>Selección de 01 transformador para la instalación del equipo DTM:</u>
P1: ¿De qué manera el análisis de datos históricos permitirá la parametrización de anomalías en un transformador de una empresa de distribución eléctrica?	O1: La parametrización de anomalías mediante el análisis de datos históricos en un transformador de una empresa de distribución eléctrica.	H1: Mediante el análisis de datos históricos se logra la parametrización de anomalías en un transformador de una empresa de distribución eléctrica.	Detectar anomalías	Del universo de transformadores ubicados en el centro histórico de Lima se seleccionó el transformador ubicado en el pasaje Olaya por conveniencia. Por las facilidades técnicas existentes tanto en la disponibilidad de personal técnico como en la menor afectación al SAIDI para su instalación. <u>Técnica e instrumentos de recolección de datos:</u> Dispositivo de medición de parámetros del transformador (DTM) para la recolección de los valores de corriente y voltaje en baja tensión
P2: ¿De qué manera la aplicación del algoritmo basado en el árbol de decisiones permitirá detectar anomalías en un transformador de una empresa de distribución eléctrica?	O2: Detectar anomalías mediante la aplicación del algoritmo basado en el árbol de decisiones en un transformador de una empresa de distribución eléctrica.	H2: Mediante la aplicación del algoritmo basado en el árbol de decisiones se logra detectar anomalías en un transformador de una empresa de distribución eléctrica.		
P3: ¿De qué manera la evaluación económica permitirá obtener indicadores de rentabilidad al aplicar aprendizaje automático en un transformador de una empresa de distribución eléctrica?	O3: Obtener los indicadores de rentabilidad mediante la evaluación económica al aplicar aprendizaje automático en un transformador de una empresa de distribución eléctrica.	H3: Mediante la evaluación económica se logra obtener los indicadores de rentabilidad al aplicar aprendizaje automático en un transformador de una empresa de distribución eléctrica.		
				<u>Técnicas e instrumentos de análisis y procesamiento de datos:</u> Lenguaje Python. Hoja de cálculo

ANEXO B. ALGORITMOS

B.1. Algoritmo para el etiquetado de datos

```
import pandas as pd
# Cargar el archivo Excel
file_path = 'TablaDatos202405_06.xlsx'
df = pd.read_excel(file_path)
# Inicializar las columnas 'Codigo' y 'Etiqueta'
df['Codigo'] = 'Normal'
df['Etiqueta'] = 0
# Todos los datos tienen código "Normal" y etiqueta "0"
# Definir las condiciones de anomalía 01
Anomalia01 = (df['X1 Current'] == 0) & (df['X2 Current'] == 0) & (df['X3 Current'] == 0)
df.loc[Anomalia01, 'Codigo'] = 'Anomalia01'
df.loc[Anomalia01, 'Etiqueta'] = 1
# Datos con código "Normal" y etiqueta "0"
# Datos con código "Anomalia01" y etiqueta "1"
# Definir Anomalia02 en datos con etiqueta "0"
Anomalia02 = (df['Etiqueta'] == 0) & (df[['X1 Current', 'X2 Current', 'X3 Current']].max(axis=1) > 1653.36)
df.loc[Anomalia02, 'Codigo'] = 'Anomalia02'
df.loc[Anomalia02, 'Etiqueta'] = 2
# Datos con código "Normal" y etiqueta "0"
# Datos con código "Anomalia01" y etiqueta "1"
# Datos con código "Anomalia02" y etiqueta "2"
# Definir Anomalia03 en datos con etiqueta "0"
Anomalia03 = (df['Etiqueta'] == 0) & (df[['X1 Voltage', 'X2 Voltage', 'X3 Voltage']].min(axis=1) < 120.66)
df.loc[Anomalia03, 'Codigo'] = 'Anomalia03'
df.loc[Anomalia03, 'Etiqueta'] = 3
# Datos con código "Normal" y etiqueta "0"
# Datos con código "Anomalia01" y etiqueta "1"
# Datos con código "Anomalia02" y etiqueta "2"
# Datos con código "Anomalia03" y etiqueta "3"
# Guardar el dataframe etiquetado en un nuevo archivo Excel
output_file_path = 'TablaDatos202405_06_etiquetado.xlsx'
df.to_excel(output_file_path, index=False)
print(f'Datos etiquetados y guardados en {output_file_path}')
```

B.2. Algoritmo para el modelamiento del árbol de decisiones

```
import pandas as pd
from sklearn.model_selection import StratifiedKFold
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
from sklearn.preprocessing import StandardScaler
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os
import joblib

# Crear directorios para guardar árboles, modelos y matrices de confusión
if not os.path.exists("decision_trees"):
    os.makedirs("decision_trees")
if not os.path.exists("saved_models"):
    os.makedirs("saved_models")
if not os.path.exists("confusion_matrices"):
    os.makedirs("confusion_matrices")

# Cargar los datos desde el único archivo
data = pd.read_excel('DataSet01.xlsx') # Cambia el nombre del archivo según corresponda

# Separar las etiquetas
X = data.drop(columns='Etiqueta') # Suponiendo que 'Etiqueta' es la columna de etiquetas
y = data['Etiqueta']
```

```

# Filtrar y guardar registros con etiquetas 1, 2 y 3
filtered_data = data[data['Etiqueta'].isin([1, 2, 3])]
filtered_data.to_excel('FilteredData_Labels_1_2_3.xlsx', index=False)

# Confirmar que todas las columnas en X sean numéricas
X = X.apply(pd.to_numeric, errors='coerce').fillna(0)

# Estandarización de los datos
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X) # Aplicar la estandarización

# Configuración del modelo de árbol de decisión
model = DecisionTreeClassifier()

# Configuración de K-Fold estratificado para mantener la proporción de anomalías en cada fold
skf = StratifiedKFold(n_splits=10, shuffle=True, random_state=42)

# Listas para guardar las métricas de cada fold
accuracy_scores = []
precision_scores = []
recall_scores = []
f1_scores = []

# Matriz de confusión acumulativa para el conjunto de prueba
cumulative_conf_matrix = np.zeros((len(y.unique()), len(y.unique())), dtype=int)

# Diccionario para almacenar las matrices de confusión por fold
confusion_matrices_dict = {}

# Realizar el entrenamiento y evaluación en cada fold
for fold, (train_index, test_index) in enumerate(skf.split(X_scaled, y), 1):
    X_train, X_test = X_scaled[train_index], X_scaled[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]

# Entrenar el modelo
model.fit(X_train, y_train)

# Predicciones para el conjunto de prueba
y_pred = model.predict(X_test)

# Predicciones para el conjunto de entrenamiento
y_train_pred = model.predict(X_train)

# Calcular métricas en el conjunto de prueba usando promedio 'weighted'
accuracy_scores.append(accuracy_score(y_test, y_pred))
precision_scores.append(precision_score(y_test, y_pred, average='weighted'))
recall_scores.append(recall_score(y_test, y_pred, average='weighted'))
f1_scores.append(f1_score(y_test, y_pred, average='weighted'))

# Calcular matrices de confusión para prueba y entrenamiento
test_conf_matrix = confusion_matrix(y_test, y_pred)
train_conf_matrix = confusion_matrix(y_train, y_train_pred)
cumulative_conf_matrix += test_conf_matrix # Sumar matriz de prueba al acumulado

# Guardar matrices de confusión en el diccionario
confusion_matrices_dict[f'Test Fold {fold}'] = test_conf_matrix
confusion_matrices_dict[f'Training Fold {fold}'] = train_conf_matrix

```

```

# Graficar y guardar la matriz de confusión del conjunto de prueba
plt.figure(figsize=(10, 8))
sns.heatmap(test_conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=y.unique(), yticklabels=y.unique())
plt.title(f'Matriz de Confusión - Prueba Fold {fold}')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.savefig(f"confusion_matrices/confusion_matrix_test_fold_{fold}.png")
plt.close()

# Graficar y guardar la matriz de confusión del conjunto de entrenamiento
plt.figure(figsize=(10, 8))
sns.heatmap(train_conf_matrix, annot=True, fmt='d', cmap='Greens', xticklabels=y.unique(), yticklabels=y.unique())
plt.title(f'Matriz de Confusión - Entrenamiento Fold {fold}')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.savefig(f"confusion_matrices/confusion_matrix_train_fold_{fold}.png")
plt.close()

# Graficar y guardar el árbol de decisión para cada fold
plt.figure(figsize=(20, 10))
plot_tree(model, filled=True, feature_names=X.columns, class_names=[str(cls) for cls in y.unique()], rounded=True, fontsize=12)
plt.title(f'Árbol de Decisión - Fold {fold}')
plt.savefig(f"decision_trees/decision_tree_fold_{fold}.png")
plt.close()

# Guardar el modelo entrenado en cada fold
joblib.dump(model, f"saved_models/decision_tree_fold_{fold}.joblib")

# Guardar la matriz de confusión acumulativa (prueba)
confusion_matrices_dict['Cumulative Test'] = cumulative_conf_matrix

# Crear un archivo Excel con las matrices de confusión
with pd.ExcelWriter('ConfusionMatrices.xlsx') as writer:
    for key, matrix in confusion_matrices_dict.items():
        df_matrix = pd.DataFrame(matrix, index=y.unique(), columns=y.unique())
        df_matrix.to_excel(writer, sheet_name=key)

# Graficar la matriz de confusión acumulativa (prueba)
plt.figure(figsize=(10, 8))
sns.heatmap(cumulative_conf_matrix, annot=True, fmt='d', cmap='Purples', xticklabels=y.unique(), yticklabels=y.unique())
plt.title('Matriz de Confusión Acumulativa - Prueba')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.savefig("confusion_matrices/cumulative_confusion_matrix_test.png")
plt.show()

# Crear un DataFrame con los resultados de las métricas
results_df = pd.DataFrame({
    'Accuracy': accuracy_scores,
    'Precision': precision_scores,
    'Recall': recall_scores,
    'F1 Score': f1_scores
})

# Guardar los resultados en un archivo Excel
results_df.to_excel('DecisionTree_CrossValidation_Metrics.xlsx', index=False)

# Imprimir el DataFrame de resultados en consola
print("Resultados de Cross Validation con Decision Tree:\n", results_df)

# Graficar el gráfico de cajas para cada métrica
plt.figure(figsize=(12, 8))
results_df.boxplot(column=['Accuracy', 'Precision', 'Recall', 'F1 Score'])
plt.title("Resultados de Cross Validation (10-Fold) para Decision Tree")
plt.ylabel("Score")
plt.xticks(rotation=45)
plt.show()

```

B.3 Algoritmo para el uso del modelo con nuevo conjunto de datos

```
import pandas as pd
import joblib
from sklearn.preprocessing import StandardScaler
import numpy as np

# Cargar el nuevo conjunto de datos (sin columna objetivo)
new_data = pd.read_excel('DataSet02.xlsx') # Cambia el nombre del archivo según corresponda

# Confirmar que todas las columnas en X sean numéricas
X_new = new_data.apply(pd.to_numeric, errors='coerce').fillna(0)

# Verificar las columnas del nuevo conjunto de datos
print("Columnas del nuevo conjunto de datos:", X_new.columns)

# Asegúrate de que las columnas del nuevo conjunto coincidan con las del conjunto de entrenamiento
required_columns = ['X1 Current', 'X2 Current', 'X3 Current', 'X1 Voltage', 'X2 Voltage', 'X3 Voltage']

# Filtrar las columnas que deben coincidir con las que se usaron en el entrenamiento
X_new = X_new[required_columns]

# Estandarización de los datos (usar el mismo proceso de entrenamiento)
scaler = StandardScaler() # Creación del escalador
X_new_scaled = scaler.fit_transform(X_new) # Aplicamos fit_transform para estandarizar

# Cargar el modelo entrenado de cada fold y hacer predicciones
y_preds = []
for fold in range(1, 11): # Para cada fold, cargar el modelo y hacer la predicción
    model = joblib.load(f"saved_models/decision_tree_fold_{fold}.joblib") # Cambia el nombre del archivo según corresponda
    y_pred_fold = model.predict(X_new_scaled)
    y_preds.append(y_pred_fold) # Guardar las predicciones de cada fold

    # Añadir las predicciones al DataFrame original
    new_data[f'Predicted_Label_Fold_{fold}'] = y_pred_fold

    # Filtrar las filas con etiquetas 1, 2 o 3 para este fold
    filtered_data_fold = new_data[new_data[f'Predicted_Label_Fold_{fold}'].isin([1, 2, 3])]

# Convertir las predicciones en un solo array (promedio de las predicciones por fold)
y_preds = np.array(y_preds)
# Tomar el valor más frecuente (moda) para cada fila de datos
final_predictions = [np.bincount(y_preds[:, i]).argmax() for i in range(y_preds.shape[1])]

# Añadir las predicciones al DataFrame original (como la columna objetivo predicha)
new_data['Predicted_Label'] = final_predictions

# Filtrar las filas con etiquetas 1, 2 o 3 en el conjunto final
filtered_data_final = new_data[new_data['Predicted_Label'].isin([1, 2, 3])]

# Guardar las filas filtradas finales en un archivo Excel
filtered_data_final.to_excel('FilteredData_Labels_1_2_3_Final.xlsx', index=False)
```