

**Análisis de Confiabilidad y Mantenibilidad con apoyo en Aprendizaje Automático
Aplicados a Transformadores de Distribución MV/LV: Evaluación y Selección de
Metodologías para la Predicción de Fallas.**

Dalmiro de Jesús Barrios Fuentes

Escuela de Ingeniería Mecánica de la Facultad de Ingeniería de la Universidad del Valle

Febrero 2026

Nota del autor

Dalmiro de Jesús Barrios Fuentes; Escuela de Ingeniería Mecánica, Universidad del Valle, Cali, Colombia.

El presente trabajo fue desarrollado bajo la dirección del Doctor Luis Alberto Mora Gutierrez y del M.Sc. Adolfo L. Gomez P.

Para dudas o aclaraciones, enviar correo electrónico a dalmirob@hotmail.com

Trabajo de tesis para optar por el título de Maestría en Gestión de Activos y Mantenimiento.

Resumen

En el marco de la gestión de activos, una estrategia de mantenimiento adecuada permite prevenir el deterioro del desempeño y gestionar el riesgo de fallas a lo largo de la vida útil de los equipos. Estas actividades se financian principalmente con gastos operativos (OPEX) y, en casos específicos, con inversiones de capital (CAPEX), lo que demanda decisiones eficientes en la asignación de recursos para maximizar el valor que generan los activos (Asset Management – an anatomy, 2024).

El presente trabajo sigue el Modelo universal e integral propuesto por Mora Gutiérrez, (2009) para estimar los parámetros de Confiabilidad, Mantenibilidad y Disponibilidad (CMD) aplicadas a una base de datos de 60000 transformadores de distribución MT/BT y a registros de fallas en Colombia desde 2022 y evalúa metodologías de aprendizaje automático. El objetivo es predecir el comportamiento futuro de estos equipos a partir de variables operativas como carga, datos de placa, ubicación y registros de fallas en circuitos para la gestión de mantenimiento de estos activos.

Los resultados se contrastan con criterios de exactitud, precisión, sensibilidad, consistencia y facilidad de implementación, con base en las normas ISO 14224:2016, UNE-EN 17007:2018 y UNE-EN 61703:2021. Finalmente, se recomienda la metodología más adecuada, orientada a fortalecer la planeación presupuestal y la estrategia de mantenimiento predictivo, enfocada en reducir costos, la mejora en la continuidad del servicio y el cumplimiento de indicadores regulatorios en línea con los objetivos organizacionales.

Palabras clave: confiabilidad, mantenibilidad, transformadores de distribución, aprendizaje automático, predicción de fallas, vida útil remanente, mantenimiento predictivo.

Abstract

Within the framework of asset management, an adequate maintenance strategy makes it possible to prevent performance deterioration and manage the risk of failures throughout the service life of equipment. These activities are mainly financed through operational expenditure (OPEX) and, in specific cases, through capital expenditure (CAPEX), which demands efficient decision-making in the allocation of resources to maximize the value generated by assets (Asset Management – an anatomy, 2024).

This study evaluates reliability, maintainability, and machine learning methodologies applied to a database of 60,000 medium-to-low voltage (MV/LV) distribution transformers and failure records collected in Colombia since 2022. The objective is to predict failures and estimate the remaining useful life (RUL) based on operational variables such as load, nameplate data, location, and circuit fault records.

Results are contrasted under criteria of accuracy, consistency, and ease of implementation, with reference to ISO 14224:2016, UNE-EN 17007:2018, and UNE-EN 61703:2021. Finally, the most suitable methodology is recommended to support budget planning and predictive maintenance strategies, aimed to cost reduction, improve service continuity, and ensure compliance with regulatory indicators in alignment with organizational objectives.

Keywords: reliability, maintainability, distribution transformers, machine learning, failure prediction, remaining useful life, predictive maintenance.

Contenido

Resumen	2
Abstract	3
Contenido	4
Figuras	5
Ecuaciones	6
Introducción y Contexto Operacional	7
Planteamiento de problema u oportunidad	9
Límites del problema u oportunidad	10
Justificación	11
Viabilidad de la idea del trabajo de profundización	12
Objetivos.....	13
General	13
Específicos	13
UNO - Datos.....	13
DOS – Patrones.....	13
TRES – Modelos.....	13
CUATRO – Análisis.....	14
CINCO – Proceso.....	14
SEIS – Conclusiones.....	14
Marco teórico.....	15
Marco conceptual	15
Antecedentes de la investigación o del proyecto	29
Estado del arte	31
Marco metodológico	35
Fases o etapas del proyecto	37
Desarrollo	42
Objetivo uno – Datos.....	42
Objetivo 2 – Patrones	62
Objetivo 3 – Modelos	75
Objetivo 4 – Análisis:	85
Objetivo 5 – Proceso.....	96
Objetivo 6 – Conclusiones.....	102
Resultados más relevantes	104
Referencias	106

Figuras

Figura 1 Modelo universal e integral, propuesto para la medición CMD	17
Figura 2 Estados operacionales (disponible/indisponible).....	18
Figura 3 Disponibilidad factible de calcular o deseada de utilizar	19
Figura 4 Curva de confiabilidad y de probabilidad de falla	22
Figura 5 Curva de la bañera o de Davies	22
Figura 6 Proceso ETL para consolidación de la información	44
Figura 7 Criterios de selección de registros	49
Figura 8 Información Dataset completo	50
Figura 9 Estructura descriptiva de los datos de TTF en la etapa de extracción	51
Figura 10 Agrupamiento de datos TTF (MUT) y TTR (MDT) por meses	52
Figura 11 Validación de tamaño muestral de eventos	54
Figura 12 Validación de tamaño muestral de meses	54
Figura 13 Correlación entre variables – Spearman & Pearson	56
Figura 14 Correlograma ACF de eventos y meses para UT y DT	57
Figura 15 Prueba de variabilidad o nivel de UT y DT	59
Figura 16 Perfil de envejecimiento.....	65
Figura 17 Distribución de causa de falla por zona	65
Figura 18 Tiempo hasta la falla vs Año de fabricación	66
Figura 19 TTF por zona	68
Figura 20 Causas de falla según año de fabricación (<2000 vs >2000).....	68
Figura 21 Tiempo hasta la falla según año de fabricación (<2000 vs >2000)	69
Figura 22 TTF por fabricante - Top 10 en número de fallas	70
Figura 23 Distribución mensual de fallas	71
Figura 24 Distribución mensual de fallas por zona	72
Figura 25 Ajuste de distribución de Weibull de 2 parámetros sin censura	75
Figura 26 Ajuste de distribución de Gamma de 2 parámetros - Sin censura.....	77
Figura 27 Comparación de las distribuciones Weibull, Gamma y Lognormal ajustadas a los TTF sin censura	78
Figura 28 Gráfica de probabilidad de Weibull para cada grupo de fabricante	79
Figura 29 Ajuste de distribución Weibull Mixture	82
Figura 30 Densidad de probabilidad y tasa de fallas de ajuste de Weibull Mixture	83
Figura 31 Variables predictoras – Features	87
Figura 32 Matriz de confusión	87
Figura 33 Sensibilidad y especificidad	88
Figura 34 Recall y precisión	90
Figura 35 Resultados de Logistic Regression.....	91
Figura 36 Resultados árbol de decisión	92
Figura 37 Resultados de Random Forest	93
Figura 38 Flujo mensual de trabajo de datos	98
Figura 39 Criterios de evaluación técnico – económica	99

Ecuaciones

Ecuación 1: Disponibilidad.....	16
Ecuación 2 Disponibilidad.....	18
Ecuación 3 Tiempo Promedio Operativo hasta el Fallo	19
Ecuación 4 Tiempo Promedio Fuera de Servicio	19
Ecuación 5 Disponibilidad genérica	20
Ecuación 6: Probabilidad de ocurrencia de un evento (de falla)	21
Ecuación 7: Probabilidad de confiabilidad (éxito).....	21
Ecuación 8: Representación matemática de la función de confiabilidad	22
Ecuación 9: Función de distribución	24
Ecuación 10: Función de confiabilidad.....	24
Ecuación 11: Función de densidad	24
Ecuación 12: Función del riesgo.....	24
Ecuación 13: Distribución de Weibull.....	25
Ecuación 14: Función de densidad de probabilidad de fallas de Weibull	25
Ecuación 15: Función de confiabilidad de Weibull	25
Ecuación 16: Función de tasa de fallas de Weibull	25
Ecuación 17: Función de densidad de probabilidad distribución Logarítmica Normal	26
Ecuación 18 Función de densidad de Gamma	26
Ecuación 19 Función de distribución de Gamma.....	26
Ecuación 20: SAIDI	29
Ecuación 21: SAIFI.....	29
Ecuación 22 MTBF Weibull	83
Ecuación 23 MTTR Weibull	84
Ecuación 24 Exactitud	88
Ecuación 25 Sensibilidad.....	89
Ecuación 26 Sensibilidad.....	89

Introducción y Contexto Operacional

Celsia Colombia S.A. E.S.P., filial del Grupo Argos, desempeña un papel relevante en el sector eléctrico colombiano como operador de red en los departamentos del Valle del Cauca y Tolima, donde atiende a más de 1,3 millones de clientes en 86 municipios. Su infraestructura está compuesta por 11 subestaciones de transmisión, 183 subestaciones de distribución y más de 47 mil kilómetros de redes de distribución en las cuales la compañía realiza inversiones permanentes con el fin de garantizar la continuidad y calidad del servicio (Plan Estratégico de Gestión de Activos de CELSIA, 2024).

La regulación también impulsa estos esfuerzos. La Resolución CREG 015 de 2018 establece que los Operadores de Red deben reducir en al menos un 8 % anual los indicadores de calidad SAIDI y SAIFI, lo que exige no solo optimizar recursos, sino mejorar continuamente la calidad del servicio (Comisión de Regulación de Energía y Gas - CREG 015, 2018).

En este contexto, una buena gestión del mantenimiento es esencial para asegurar la generación valor a los grupos de interés. Por ello, las inversiones deben priorizarse, y dicha priorización debe maximizar el logro del objetivo general de toda la cadena de producción, transmisión y distribución de electricidad: garantizar el suministro a los clientes (Biard & Nour, 2021).

Sin embargo, en la práctica, las empresas eléctricas aún gestionan los transformadores de distribución bajo un esquema de **run-to-failure** (Kabir et al., 2018). Esto significa que las decisiones de reposición se toman a partir de estadísticas simples de fallas del año anterior, sin conocer con precisión el estado real del resto de transformadores ni el número que fallará en el futuro próximo.

Aplicar una estrategia de mantenimiento preventivo a los 60000 transformadores de distribución que posee Celsia no resulta económicamente viable. En su lugar, este trabajo plantea aprovechar la información recopilada durante años de operación e integrar herramientas de la confiabilidad y conceptos de industria 4.0, con el fin de contribuir a la

gestión de activos eléctricos y apoyar los procesos de toma de decisiones orientadas a reducir riesgos con la confiabilidad y el desempeño (Biard & Nour, 2021).

Si bien en transformadores de transmisión existen metodologías avanzadas de diagnóstico, como el análisis de gases disueltos (DGA) y modelos de *machine learning* para monitoreo en línea con resultados efectivos, su aplicación en transformadores de distribución resulta inviable por los altos costos de instalación de sensores y sistemas de adquisición, y por la gran cantidad de equipos a cubrir (Kabir et al., 2018).

Este trabajo se estructura en varias etapas que responden al objetivo de evaluar metodologías de confiabilidad, mantenibilidad y aprendizaje automático aplicadas a transformadores de distribución MV/LV. En primer lugar, se consolida la base de datos de transformadores y circuitos relacionados según la ISO 14224:2016. Luego se analizan patrones históricos de fallas y desempeño, con el fin de identificar variables predictivas.

Posteriormente, se aplican modelos estadísticos de confiabilidad y mantenibilidad, junto con algoritmos de aprendizaje automático, con referencia a la UNE-EN 61703:2021. Los resultados se comparan en términos de precisión y consistencia mediante técnicas estadísticas y estocásticas. Finalmente, se formula un esquema de criterios técnicos alineado con la ISO 14224:2016, la UNE-EN 17007:2018 y la ISO 55001:2024, que permite recomendar la alternativa metodológica más adecuada para la gestión de transformadores MT/BT.

Planteamiento de problema u oportunidad

La disponibilidad de datos se ha convertido en un factor estratégico para la gestión de activos. Las normas ISO 14224:2016 y la UNE-EN 17007:2018 destacan la importancia de estructurar, clasificar y utilizar los registros de fallas y operación como base para análisis de confiabilidad y toma de decisiones en mantenimiento (ISO Norma 14224, 2016).

En el caso de los transformadores de distribución, CELSIA cuenta con información valiosa que incluye históricos de fallas desde 2022, datos de placa, condiciones de carga y localización de casi 60 mil unidades. Sin embargo, dicha información se emplea principalmente para reportar lo que ya pasó, como tableros que muestran los transformadores con mayor recurrencia de fallas o listados de reemplazos anuales. Esta práctica no promueve al análisis de causas profundas, ni permite anticipar la probabilidad de falla, ni estimar la vida útil remanente de la flota.

Frente a esta situación, se identifica una oportunidad: aplicar metodologías de confiabilidad, mantenibilidad y algoritmos de aprendizaje automático para transformar los datos a pronósticos de desempeño y fallas. La confiabilidad, la mantenibilidad y la disponibilidad son prácticamente las únicas medidas técnicas y científicas fundamentadas en cálculos matemáticos, estadísticos y probabilísticos que tiene el mantenimiento para su análisis y su evaluación (Mora Gutiérrez, 2009).

Estudios recientes han demostrado que modelos de *machine learning* incluyendo modelos de Weibull, modelos de support vector machines y modelos de redes neuronales para la estimación de la vida útil real y la capacidad de pronóstico de fallas en equipos eléctricos ofrecen resultados importantes (Cvijic et al., 2023).

La utilización de los registros históricos con un enfoque predictivo abre la posibilidad de planear con mayor precisión el presupuesto de mantenimiento, reducir la ocurrencia de fallas inesperadas y mejorar el desempeño frente a los indicadores SAIDI y SAIFI.

Límites del problema u oportunidad

Este trabajo se centra en los transformadores de distribución MT/BT de Celsia instalados en los departamentos de Tolima y Valle del Cauca, utilizando registros históricos de fallas, datos técnicos y de operación para la aplicación de metodologías estadísticas de confiabilidad y mantenibilidad, así como de algoritmos de aprendizaje automático.

El trabajo no incluye transformadores de transmisión o generación, ni técnicas de monitoreo en línea como el análisis de gases disueltos, debido a que su implementación masiva en distribución resulta inviable. Asimismo, no se desarrolla un modelo nuevo, sino que se evalúan metodologías existentes bajo el marco de las normas ISO 14224:2016, UNE-EN 17007:2018 y UNE-EN 61703:2021, con base en la información histórica disponible en los últimos años.

Justificación

Los transformadores de distribución cumplen un papel importante en la cadena de suministro de energía eléctrica, al ser prácticamente el último eslabón del sistema interconectado nacional frente al usuario final. Con el paso del tiempo, su deterioro es natural y aumenta el riesgo de fallas graves que pueden generar afectaciones a la población y comprometer la continuidad del servicio. En consecuencia, estas fallas impactan directamente en la seguridad y en el cumplimiento de indicadores regulatorios como SAIDI y SAIFI.

Actualmente, Celsia dispone de una gran cantidad de datos operativos y de fallas de transformadores de distribución, lo que representa una oportunidad para no solo desarrollar la dimensión de analítica predictiva sino también para avanzar hacia la dimensión de analítica cognitiva, mediante la integración de técnicas estadísticas de confiabilidad y mantenibilidad junto con algoritmos de inteligencia artificial y aprendizaje automático (Mora Gutiérrez, 2024).

Lo anterior, permite transformar los datos en predicción de fallas y vida útil remanente, lo que a su vez fortalece la capacidad de anticipación y respalda decisiones más efectivas en la gestión de activos eléctricos.

En promedio, las fallas en transformadores de distribución representan 4,54% del total de afectaciones en SAIDI si se cuenta desde hace tres años. Por lo tanto, la explotación de los datos disponibles no solo permite una asignación más eficiente de recursos, sino que también genera un margen adicional para cumplir con la mejora anual del 8% exigida por la CREG en los indicadores de calidad del servicio.

Las fallas en los transformadores implican pérdidas económicas tanto por los gastos asociados al mantenimiento correctivo como por los ingresos que la compañía deja de percibir a causa de la energía no suministrada (ENS). Para mejorar la confiabilidad del sistema y mitigar el riesgo de interrupciones, es esencial identificar las causas y aplicar acciones para disminuir las tasas de falla (Singh et al., 2019).

En los últimos años, las expectativas de los consumidores han aumentado con el desarrollo y avance de las tecnologías modernas. Estos esperan un suministro de mayor calidad, con menor costo y menos fallas. Además, con el envejecimiento, la tasa de fallas de los equipos también aumenta, lo que incrementa de manera considerable tanto la frecuencia de fallas como el costo por interrupciones para los usuarios (Agarwal et al., 2020).

Viabilidad de la idea del trabajo de profundización

El proyecto demuestra alta viabilidad al cumplir con criterios de conveniencia, implicaciones prácticas, valor teórico y utilidad metodológica. Desde el punto de vista de conveniencia, se apoya en la amplia base de datos con que cuenta Celsia, lo cual permite transformar información ya disponible en decisiones estratégicas sin requerir inversiones adicionales en sistemas de monitoreo costosos.

En cuanto a las implicaciones prácticas, la iniciativa aporta a la optimización de la planeación presupuestal, la reducción de intervenciones reactivas y la mejora del desempeño frente a los indicadores SAIDI y SAIFI, en concordancia con la exigencia de mejora anual establecida por la regulación colombiana.

El valor teórico se refleja en la posibilidad de aportar evidencia comparativa sobre la aplicación de metodologías de confiabilidad y aprendizaje automático en transformadores de distribución, además de validar el uso de estándares internacionales como la ISO 14224 y la UNE-EN 17007 en el contexto de distribución de energía eléctrica.

Finalmente, la utilidad metodológica radica en la construcción de un esquema de criterios técnicos que relacione variables operativas, físicas y de entorno con los resultados de predicción. Esto abre la puerta a futuros estudios en gestión de activos eléctricos y aporta al desarrollo de modelos replicables en otras empresas del sector.

Objetivos

General

Evaluar metodologías de confiabilidad de transformadores de distribución MT/BT mediante técnicas estadísticas y modelos de aprendizaje automático para estimar la probabilidad de falla y la vida útil remanente, en apoyo a la toma de decisiones de gestión de activos.

Específicos

UNO - Datos.

Identificar los datos disponibles de los transformadores MT/BT (fabricante, año de fabricación, causas de falla) y de los circuitos adyacentes, con el fin de consolidar una base de datos confiable, con referencia a la ISO 14224:2016 para la estructuración y clasificación de la información. **Nivel 1 – Conocimiento Escala de Benjamín Bloom y Robert Gagné.**

DOS – Patrones.

Interpretar los patrones históricos de fallas y el desempeño de los transformadores de distribución, con el fin de determinar las variables relevantes para los modelos de confiabilidad y los algoritmos de predicción. **Nivel 2 – Comprensión de la Escala de Benjamín Bloom y Robert Gagné.**

TRES – Modelos.

Aplicar modelos probabilísticos de análisis de confiabilidad para estimar la función de confiabilidad, la tasa de falla y la vida útil remanente de los transformadores de distribución MT/BT. **Nivel 3 – Aplicación de la Escala de Benjamín Bloom y Robert Gagné.**

CUATRO – Análisis.

Examinar el desempeño de algoritmos de aprendizaje automático para la clasificación de transformadores con alta probabilidad de falla mediante métricas como precisión, recall y AUC. **Nivel 4 – Análisis de la Escala de Benjamín Bloom y Robert Gagné.**

CINCO – Proceso.

Proponer un proceso de apoyo a la gestión de transformadores de distribución MT/BT que integra los resultados del análisis de confiabilidad y de las técnicas de aprendizaje automático, con énfasis en la estandarización del registro de fallas y en la clasificación de modos y causas de falla. **Nivel 5 – Sintetizar de la Escala de Benjamín Bloom y Robert Gagné.**

SEIS – Conclusiones.

Mostrar los principales hallazgos del trabajo de profundización y sus implicaciones prácticas para la toma de decisiones en la gestión de transformadores de distribución (activos eléctricos).

Marco teórico

Marco conceptual

Análisis CMD

El análisis CMD (Confiabilidad – Mantenibilidad – Disponibilidad) es una herramienta que facilita el control de la gestión del mantenimiento, al tiempo que permite proyectar el comportamiento futuro de corto plazo de los equipos en el corto plazo, en cuanto a fallas, reparaciones y los tiempos de operación útil (Mora Gutiérrez, 2009).

Método de predicción: distribuciones estadísticas

Entre los métodos de predicción, las distribuciones estadísticas permiten modelar con mayor precisión el comportamiento temporal de las variables CMD y de sus parámetros característicos. En términos generales, este enfoque es aplicable tanto a equipos y sistemas reparables como no reparables, aunque algunos autores señalan ciertas limitaciones para los primeros. El método cuenta con amplia aceptación internacional debido a la sencillez de su aplicación práctica (Mora Gutiérrez, 2009).

Modelo universal e integral, propuesto para la medición CMD

El presente trabajo sigue la metodología propuesta por Mora Gutiérrez (2009) para estimar los parámetros de CMD expuesta en la Figura 1, el cual se conforma de seis etapas. La primera define el procedimiento para la obtención, tabulación y tratamiento de los datos. La segunda determina el tipo de disponibilidad según la información disponible y expectativas de la organización (Mora Gutiérrez, 2009).

En la tercera etapa se selecciona entre el método directo de máxima verosimilitud (MLE) o el método de alineación Weibull, que incluye la estimación de los parámetros de no confiabilidad y mantenibilidad mediante técnicas como el i-kaésimo, Benard, Kaplan-Meier o rango de medianas. En ambos casos se validan los resultados mediante la valoración de los índices de bondad de ajuste: correlación (r) y con los coeficientes de determinación (r^2) (Mora Gutiérrez, 2009).

La cuarta etapa aborda la parametrización de Weibull y sus pruebas de bondad de ajuste, aplicando los métodos de Kolmogórov-Smirnov, Anderson-Darling y Chi-cuadrado Ji². La quinta permite la parametrización con funciones alternativas a Weibull y la evaluación de su ajuste, mientras que la sexta integra todos los cálculos CMD con la función que se seleccione y cumpla con todos los ajustes (Mora Gutiérrez, 2009).

Disponibilidad

Probabilidad de que un equipo se encuentre en condiciones para cumplir su función y este cumpla satisfactoriamente esta cuando sea requerido (Mora Gutiérrez, 2009).

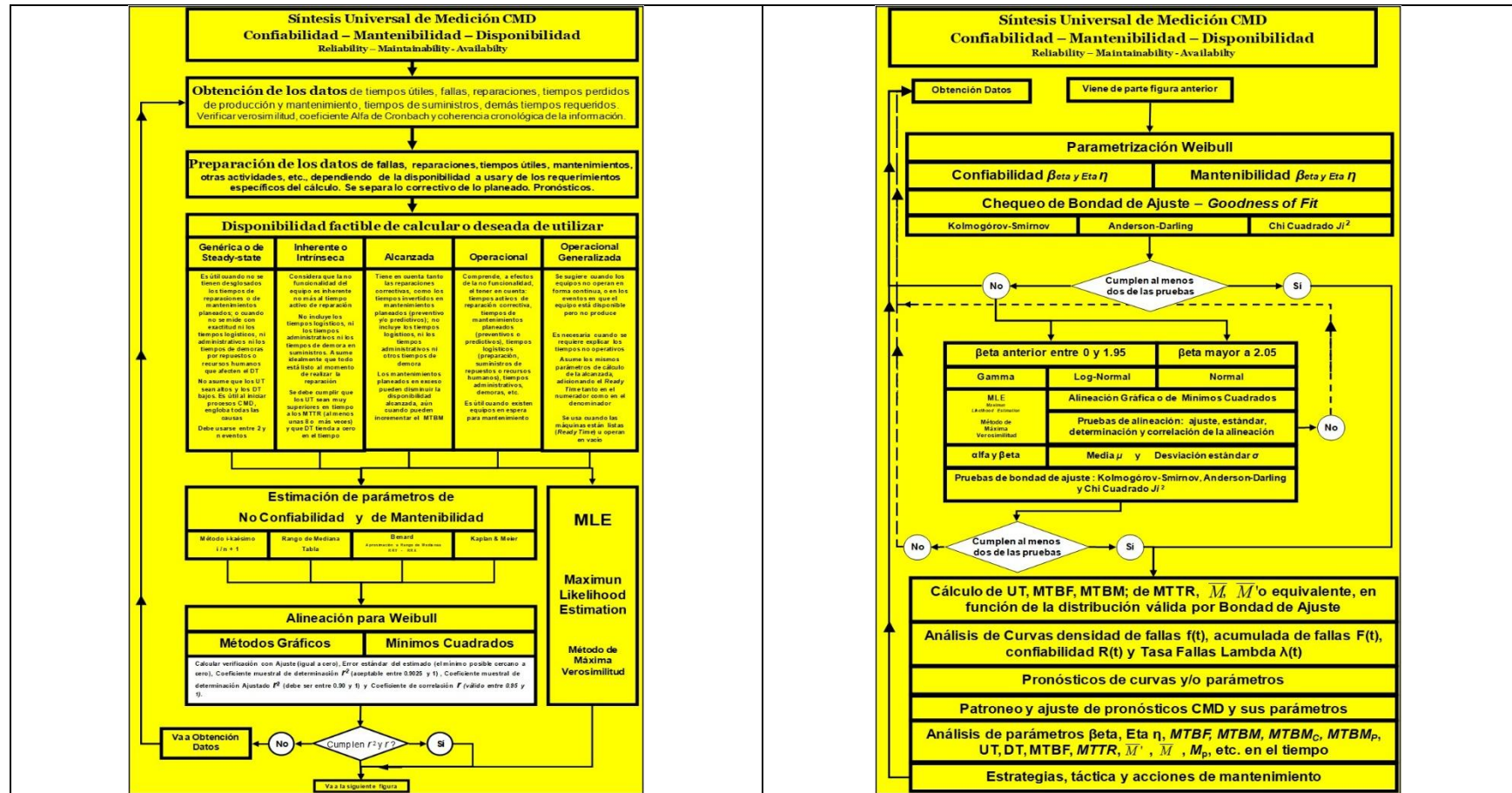
Ecuación 1:

$$Disponibilidad = \frac{Confiabilidad}{Confiabilidad + Mantenibilidad}$$

Nota. La ecuación muestra la relación entre la confiabilidad y la mantenibilidad en el cálculo de la disponibilidad de un equipo o sistema. Fuente: (Mora Gutiérrez, 2009).

Figura 1

Modelo universal e integral, propuesto para la medición CMD



Nota. Modelo universal e integral de medición de CMD propuesto por Mora Gutiérrez (2009) para modelos puntual y de

distribuciones. Establece una metodología única para estimar parámetros, permitiendo comparaciones consistentes entre empresas e industrias. Fuente: (Mora Gutiérrez, 2009).

Según Parra Márquez & Crespo Márquez (2012) la disponibilidad es un indicador que expresa el porcentaje del tiempo total en que un equipo se encuentra apto para cumplir su función prevista, bajo el supuesto de que dispone de todos los recursos externos necesarios para su operación.

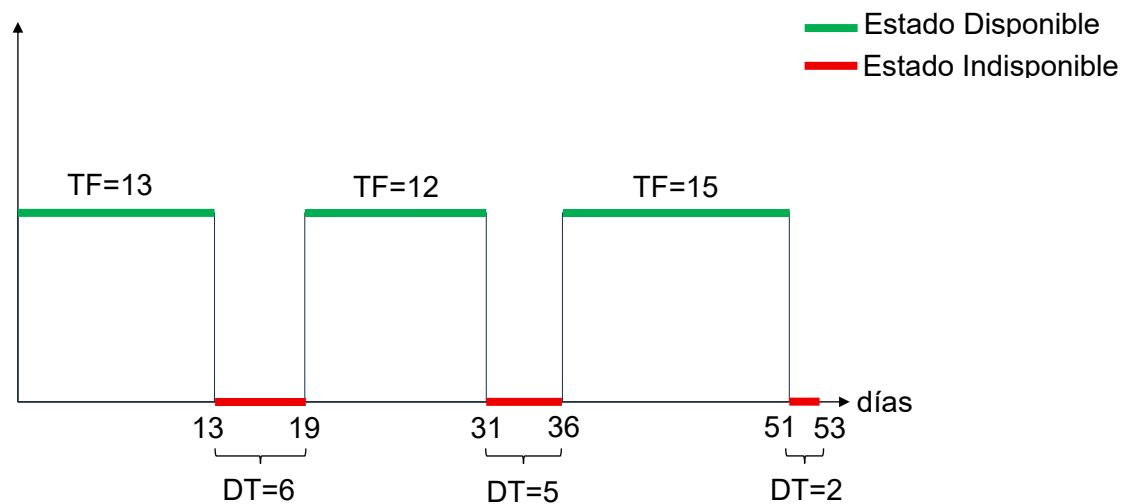
Ecuación 2 Disponibilidad

$$D = \frac{MTTF}{MTTF + MDT} \times 100\%$$

Nota. Este indicador integra los efectos combinados de la confiabilidad (MTTF) y la mantenibilidad (MDT). Lo que proporciona una visión global del desempeño operativo del activo. Un alto valor de disponibilidad indica que el sistema o componente no solo falla con poca frecuencia, sino que además puede ser restablecido con rapidez.

Figura 2

Estados operacionales (disponible/indisponible)



Número de fallas = 3

Nota. En la imagen se observan los tiempos de operación (disponibilidad) y tiempos de inactividad (indisponibilidad) de un activo en producción. Fuente: Adaptada de (Parra Márquez & Crespo Márquez, 2012).

Tiempo promedio operativo hasta el fallo – (MTTF)

Es un indicador de confiabilidad que mide el tiempo promedio que un equipo puede operar sin interrupciones (Parra Márquez & Crespo Márquez, 2012).

Ecuación 3

Tiempo Promedio Operativo hasta el Fallo

$$MTTF = \frac{\sum_{i=1}^{i=n} TTF_i}{n}$$

Nota. Expresa el valor medio de los intervalos de tiempo TTF_i transcurridos entre el inicio de la operación y la ocurrencia de la falla con base en un total de n fallo observados durante el tiempo de estudio. Fuente: (Parra Márquez & Crespo Márquez, 2012).

Tiempo promedio fuera de servicio – (MDT)

Es un indicador que mide el tiempo medio para restaurar un sistema o componente a una condición tal que pueda seguir cumpliendo su función después de un fallo. (Parra Márquez & Crespo Márquez, 2012)

Ecuación 4

Tiempo Promedio Fuera de Servicio

$$MDT = \frac{\sum_{i=1}^{i=n} DT_i}{n}$$

Nota. La ecuación del Tiempo Promedio Fuera de Servicio (MDT) determina el promedio de los intervalos DT_i . El parámetro n indica el número de fallos observados durante el período analizado. Fuente: (Parra Márquez & Crespo Márquez, 2012)

Figura 3

Disponibilidad factible de calcular o deseada de utilizar

Disponibilidad factible de calcular o deseada de utilizar				
Genérica	Inherente o intrínseca	Alcanzada	Operacional	Operacional Generalizada
Es útil cuando no se tienen desglosados los tiempos de reparaciones o de mantenimientos planeados; o cuando no se mide con exactitud ni los tiempos logísticos, ni administrativos ni los tiempos de moras por repuestos o recursos humanos que afecten el DT No asume que los UT sean altos y los DT bajos. Es útil al iniciar procesos CMD, engloba todas las causas Debe usarse entre 2 y en eventos	Considera que la no funcionalidad del equipo es inherente no más al tiempo activo de reparación No incluyen los tiempos logísticos, ni los tiempos administrativos ni los tiempos de demora en suministros. Asume idealmente que todo está listo al momento de realizar la reparación Se debe cumplir que los UT sean muy superiores en tiempo a los MTTR (al menos unas 8 o más veces) y que DT tienda a cero en el tiempo	Tiene en cuenta tanto las reparaciones correctivas, como los tiempos invertidos en mantenimientos planeados (preventivo y/o predictivos); no incluye los tiempos logísticos, ni los tiempos administrativos mora en suministros Los mantenimientos planeados en exceso pueden disminuir la disponibilidad alcanzada, aún cuando pueden incrementar el MTBM	Comprende, a efectos de la no funcionalidad, el tener en cuenta: tiempos activos de reparación correctiva, tiempos de mantenimientos planeados (preventivos o predictivos), tiempos logísticos (preparación, suministros de repuestos o recursos humanos), tiempos administrativos, de moras, etc. Es útil cuando existen equipos en espera para mantenimiento	Se sugiere cuando los equipos no operan en forma continua, o en los eventos en que el equipo está disponible pero no produce Es necesaria cuando se requiere explicar los tiempos no operativos. Asume los mismos parámetros de cálculo de la alcanzada, adicionando el <i>Ready Time</i> tanto en el numerador como en el denominador Se usa cuando las máquinas están listas (<i>Ready Time</i>) u operan en vacío

Nota. Clasificación de los tipos de disponibilidad aplicables en la medición de CMD, según los datos y condiciones operativas del sistema: genérica, inherente, alcanzada, operacional y operacional generalizada. Fuente: (Mora Gutiérrez, 2009).

Disponibilidad genérica - A_G

Ecuación 5

Disponibilidad genérica

$$A_G = \frac{\frac{\sum_{i=1}^m UT_i}{m}}{\frac{\sum_{i=1}^m UT_i}{m} + \frac{\sum_{j=1}^n DT_j}{n}} = \frac{MUT}{MUT + MDT}$$

Nota. La ecuación muestra la relación entre el tiempo útil (UT) y el tiempo de no funcionalidad (DT). Donde m es el número de eventos de UT y n el número de ventos de DT. Fuente: (Mora Gutiérrez, 2009).

A los transformadores de distribución no se les aplica mantenimiento preventivo ni predictivo. Es decir, el activo opera hasta que falle (Run to Failre - RTF). En ese sentido los estos activos se categorizan como no reparables, es decir, su vida útil concluye con un único evento de fallo, lo que implica que deben ser sustituidos. En este contexto, la variable aleatoria

de mayor interés corresponde al tiempo transcurrido hasta la ocurrencia de dicha falla (Parra Márquez & Crespo Márquez, 2012).

Este proyecto se enfoca en transformadores de distribución, por lo tanto, no hay tiempos asociados a mantenimientos planeados dentro de DT. En esa gestión no se desglosan los tiempos por lo tanto solo se miden los tiempos útiles y de no funcionalidad. Por lo anterior, es factible calcular la disponibilidad genérica.

Confiabilidad

Es la capacidad que tiene un sistema o componente para cumplir adecuadamente su función prevista dentro de las condiciones definidas y durante un tiempo determinado. (Liu et al., 2023). Por consiguiente, para determinar si un equipo es confiable, basta con evaluar la frecuencia con la que se presentan sus fallas a lo largo del tiempo.

Para Nachlas (1995), la confiabilidad se entiende como la probabilidad de que un sistema o componente desempeñe satisfactoriamente la función para la cual fue diseñado durante un período de tiempo determinado.

Ecuación 6:

Probabilidad de ocurrencia de un evento (de falla)

$$P_f = \left(\frac{n}{N}\right)$$

Nota. La ecuación define P_f la probabilidad de ocurrencia de un evento; donde n represente el número de eventos de falla (en confiabilidad); N es el número total de eventos posible. Fuente: (Mora Gutiérrez, 2009)

Ecuación 7:

Probabilidad de confiabilidad (éxito o supervivencia)

$$R_a = 1 - P_f$$

Nota. La ecuación define la probabilidad de éxito o supervivencia R_a en un tiempo a . Fuente: (Mora Gutiérrez, 2009)

Ecuación 8:

Representación matemática de la función de confiabilidad

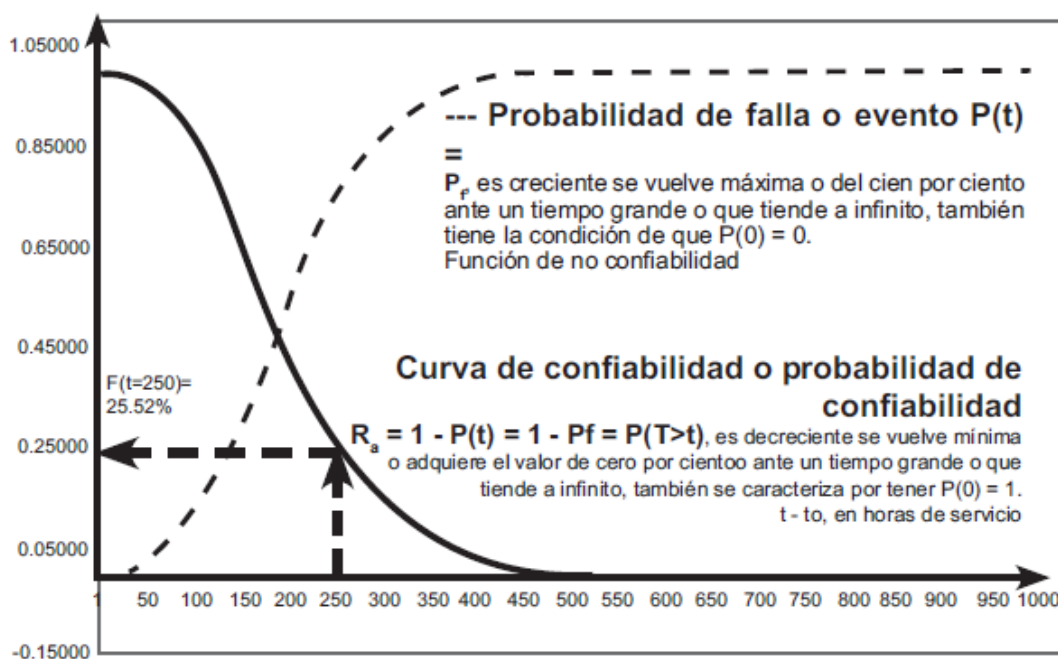
$$R(t) = P[t < T]$$

Nota. $R(t)$ es la función de confiabilidad, T representa el tiempo hasta la falla. Fuente:

(Mora Gutiérrez, 2009)

Figura 4

Curva de confiabilidad y de probabilidad de falla

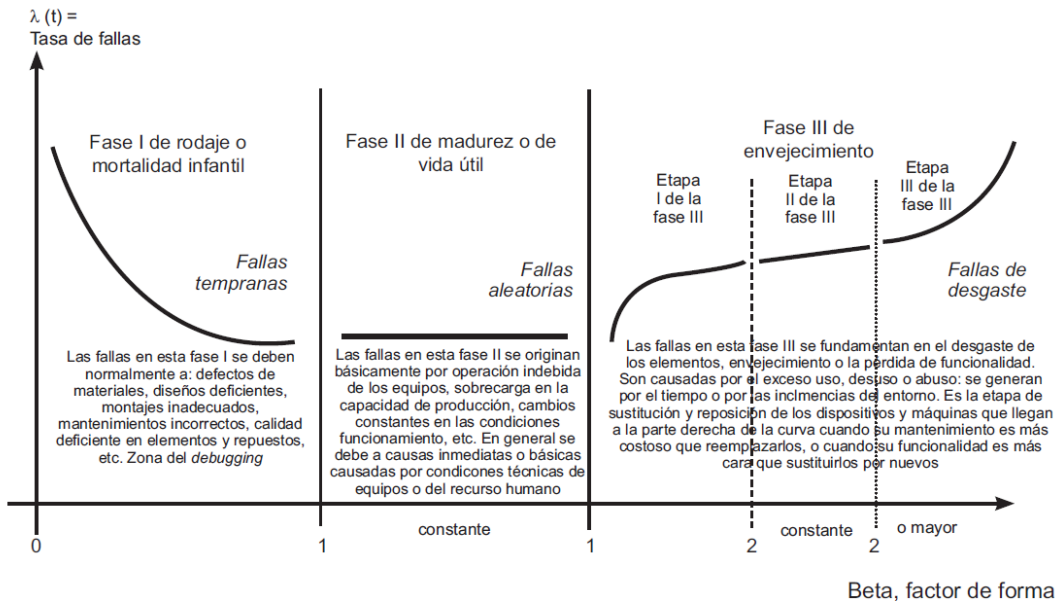


Nota. En la figura se aprecia las curvas de confiabilidad y de probabilidad de un ejercicio donde la probabilidad de que un elemento dure más de 250 horas sin fallar es del 25.52%. También puede leerse para un conjunto de elementos como después de 250 horas deben funcionar correctamente el 25.52%. Fuente: (Mora Gutiérrez, 2009).

Curva de la bañera o de Davies

Figura 5

Curva de la bañera o de Davies



Nota. La figura muestra la evolución de la tasa de fallas en el tiempo y la ubicación del valor del parámetro de forma Beta del elemento evaluado. Fuente: (Mora Gutiérrez, 2009).

A continuación, se presentan las expresiones de las principales distribuciones tradicionales empleadas en análisis de confiabilidad. Estas resultan de especial interés por su capacidad de representar distintos patrones de comportamiento en la tasa de fallas, así como por su flexibilidad y amplia aplicabilidad en el estudio de activos eléctricos.

Distribuciones tradicionales

En teoría, cualquier función de distribución puede emplearse para modelar la duración de los equipos. Sin embargo, en la práctica, resultan más adecuados los modelos cuya función de riesgo mantiene un comportamiento monótono, ya que representan de manera más realista el proceso de falla. Dentro de este grupo, algunas distribuciones se reconocen como las más utilizadas y confiables para describir la fiabilidad de los dispositivos, constituyendo la base de los enfoques tradicionales en análisis de confiabilidad (Nachlas, 1995).

A continuación, para una duración de vida T , las cuatro funciones para describir la duración de vida de dispositivos están dadas por las siguientes expresiones matemáticas:

Ecuación 9:

Función de distribución

$$F(t) = P(T \leq t)$$

Nota. Aquí se tiene la probabilidad de que el tiempo de vida sea menor o igual que t y también se puede definir como probabilidad acumulada de que el activo haya fallado antes de o en t . Fuente: (Nachlas, 1995)

Ecuación 10:

Función de confiabilidad

$$R(t) = \bar{F}(t) = 1 - F(t)$$

Nota. Esta expresión calcula la probabilidad de que el tiempo de vida T sea mayor que t . Fuentes: (Nachlas, 1995)

Ecuación 11:

Función de densidad

$$f(t) = \frac{d}{dt} F(t)$$

Nota. La derivada de la función de distribución es la función de densidad. Fuente: (Nachlas, 1995)

Ecuación 12:

Función del riesgo

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)} = \frac{f(t)}{R(t)}$$

Nota. También conocida como función de intensidad o función de tasa de fallas.

Adicionalmente, se tiene que: $F(t)$ es de tasa de falla creciente (IFR) si $\frac{d}{dt} \lambda(t) \geq 0$; $F(t)$ es de tasa de falla decreciente (DFR) si $\frac{d}{dt} \lambda(t) \leq 0$ y $F(t)$ es de tasa de falla constante (CFR) si $\frac{d}{dt} \lambda(t) = 0$. Fuente: (Nachlas, 1995).

Distribución de Weibull

Otro modelo de distribución de vida bastante utilizado, pues su función de tasa de falla se ajusta a cualquier fase de la curva de la bañera y permite representar mantenibilidad y confiabilidad (Blanco et al., 2018). Se expresa de la siguiente manera:

Ecuación 13:

Distribución de Weibull

$$F(t) = 1 - e^{-\left[\frac{t-\gamma}{\eta}\right]^\beta}$$

Ecuación 14:

Función de densidad de probabilidad de fallas de Weibull

$$f(t) = \frac{\beta}{\eta} \left[\frac{t-\gamma}{\eta}\right]^{\beta-1} e^{-\left[\frac{t-\gamma}{\eta}\right]^\beta}$$

Ecuación 15:

Función de confiabilidad de Weibull

$$R(t) = 1 - F(t) = e^{-\left[\frac{t-\gamma}{\eta}\right]^\beta}$$

Ecuación 16:

Función de tasa de fallas de Weibull

$$\lambda(t) = \frac{\beta}{\eta} \left[\frac{t-\gamma}{\eta}\right]^{\beta-1}$$

Nota. Donde η es el parámetro de escala, mientras que β corresponde al parámetro de forma. Fuente: (Mora Gutiérrez, 2009)

La principal fortaleza de esta distribución radica en su flexibilidad, ya que permite modelar distintos patrones de falla y mantiene una estructura algebraicamente manejable. Al igual que otras distribuciones de dos parámetros, ofrece una representación adecuada de una amplia variedad de situaciones reales (Nachlas, 1995). Al ser tan flexible y adaptarse a la curva de la bañera resulta ser importante considerar analizar los datos desde la perspectiva de esta distribución.

Distribución Logarítmica Normal

Por otra parte, se considera frecuentemente que la distribución logarítmica normal es un modelo representativo de la duración de vida de ciertos componentes electrónicos (Nachlas, 1995). Los valores μ y σ no son la media y la desviación estándar de la variable aleatoria, sino de su logaritmo (Blanco et al., 2018).

Ecuación 17:

Función de densidad de probabilidad distribución Logarítmica Normal

$$f(t) = \frac{e^{\left[-(\ln t - \mu)^2 / 2\sigma^2\right]}}{t\sigma\sqrt{2\pi}}$$

Fuente: (Nachlas, 1995)

Distribución Gamma

Ecuación 18

Función de densidad de Gamma

$$f(t) = \frac{\lambda^\beta}{\Gamma(\beta)} t^{\beta-1} e^{-\lambda t}$$

Fuente: (Nachlas, 1995)

Ecuación 19

Función de distribución de Gamma

$$f(t) = \frac{\lambda^\beta}{\Gamma(\beta)} t^{\beta-1} e^{-\lambda t}$$

Nota. Como en el caso de Weibull, esta distribución muestra un riesgo creciente cuando $\beta > 1$, un riesgo decreciente cuando $\beta < 1$, y un riesgo constante cuando $\beta = 0$. Fuente: (Nachlas, 1995).

Distribuciones no tradicionales

En la literatura se encuentran distribuciones no tradicionales que extienden las capacidades de los modelos clásicos y que pueden resultar de interés para este proyecto. En

particular Blanco et al. (2018), presentan varios modelos que pueden mejorar la representación de los datos de falla en equipos eléctricos. Para el caso del objetivo del presente trabajo, se destacan como candidatas preliminares Burr XII, la Weibull modificada (MWM y NMWM) y la distribución de Hjorth (IDB). Estas ofrecen mayor flexibilidad en la forma de la función de riesgo, permitiendo capturar comportamientos como la curva de la bañera completa, colas largas y alta dispersión en los datos.

El propósito del presente proyecto es evaluar estas herramientas e identificar los modelos que más se ajusten al comportamiento de los transformadores de distribución y faciliten la estimación de vida útil remanente y predicción de fallas. En etapas posteriores del trabajo, al graficar los datos se pueden identificar patrones y seleccionar aquellos modelos que mejor se ajusten al comportamiento real de los activos.

Inteligencia artificial y aprendizaje automático

Para lo que se quiere lograr con este trabajo de profundización y como complemento de la estadística clásica, se requiere explorar los conceptos de inteligencia artificial (IA) y aprendizaje automático (Machine Learning, ML). Dentro de este ámbito, el Aprendizaje Automático (Machine Learning, ML) constituye una de sus ramas más relevantes, al diseñar algoritmos que aprenden patrones a partir de datos históricos y de condición, con el fin de realizar predicciones sobre información no observada previamente (Artasanchez & Joshi, 2020).

En el contexto del presente trabajo, estas herramientas ofrecen un marco para anticipar fallas y estimar la vida útil remanente de transformadores de distribución MT/BT, aprovechando registros operativos bajo lineamientos de confiabilidad establecidos en normas internacionales.

El aprendizaje automático se clasifica en aprendizaje supervisado y no supervisado. El aprendizaje supervisado se fundamenta en el uso de datos de entrenamiento etiquetados, donde cada registro incluye tanto variables de entrada como un valor de salida conocido. Bajo este esquema, los modelos aprenden la relación entre las condiciones del activo y un resultado

esperado, lo que permite aplicar técnicas de clasificación para identificar estados de falla o de regresión para estimar la vida útil remanente (Artasanchez & Joshi, 2020).

Por otra parte, el aprendizaje no supervisado opera sin etiquetas de salida, de modo que el modelo debe descubrir patrones ocultos únicamente a partir de los datos disponibles. En este caso, los algoritmos agrupan registros con características similares o reducen la complejidad del conjunto de datos (Artasanchez & Joshi, 2020).

Por la naturaleza de los datos disponibles, el presente proyecto se orienta hacia el uso de aprendizaje supervisado, el cual se divide en dos tipos de entrenamiento: clasificación y regresión. Como lo indica Sandoval (2018), la clasificación, como técnica que identifica patrones en los datos y luego asigna cada observación a un grupo predefinido, permite categorizar los transformadores de acuerdo con su nivel de riesgo de falla en estados críticos, medios o leves.

Por su parte, la regresión se enfoca en la predicción de valores numéricos, lo que posibilita estimar la probabilidad de falla en un horizonte temporal específico o calcular la vida útil remanente de los equipos.

En la etapa 4 del proyecto se prevé evaluar y contrastar diferentes algoritmos de cada categoría con el fin de validar su coherencia y aplicabilidad frente a los objetivos de predicción de fallas y estimación de vida útil en transformadores de distribución.

Antecedentes de la investigación o del proyecto

El modelo de gestión de mantenimiento de transmisión y distribución de Celsia se estructura en ocho fases alineadas con el ciclo de Deming o mejora continua. Las fases 1, 2 y 3 garantizan la eficacia de la gestión; las fases 4 y 5 comprenden la planificación, programación y ejecución; las fases 6 y 7 se centran en la evaluación y en el análisis de costos de los activos a lo largo del ciclo de vida; y la fase 8 se orienta hacia la mejora continua.

En la fase 6, el equipo de Confiabilidad de Mantenimiento realiza seguimiento a indicadores de calidad del servicio, al número de transformadores de distribución fallados y al desempeño de los activos, entre otros indicadores más.

Los indicadores de calidad de servicio SAIDI (System Average Interruption Duration Index) y SAIFI (System Average Frequency Index), son ampliamente utilizados en el sector eléctrico para medir la duración y frecuencia promedio de las de las interrupciones en la red.

Ecuación 20:

SAIDI

$$SAIDI = \frac{\sum \text{Duración de todas las interrupciones de consumidores}}{\text{Total de consumidores en el sistema}}$$

Nota. Indicador de percepción de duración de interrupciones.

Ecuación 21:

SAIFI

$$SAIFI = \frac{\sum \text{Total de las interrupciones de consumidores}}{\text{Total de consumidores en el sistema}}$$

Nota. Indicador de percepción de frecuencia de interrupciones.

El cumplimiento de estos indicadores se revisa mensualmente y se consolidan en el informe de confiabilidad de redes MT/BT, además de presentarse en un tablero operativo en Power BI que se actualiza diariamente con datos de las plataformas de operación y mantenimiento. Asimismo, el indicador transformadores de distribución fallados cuantifica los

equipos que han salido de servicio por falla o evento propio del activo, registrándose mensualmente en los reportes de confiabilidad.

No obstante, la compañía no calcula actualmente indicadores RAM (Reliability, Availability, Maintainability) específicos para transformadores de distribución. Esto limita la capacidad de evaluar de manera integral la confiabilidad y mantenibilidad de dichos activos, así como de estimar su vida útil remanente.

Al mismo tiempo, la creciente disponibilidad de información operativa y el interés por mejorar la confiabilidad de las redes motivan la necesidad de adoptar metodologías más avanzadas. Entre las cuales se destacan tanto las técnicas estadísticas clásicas como los indicadores RAM propuestos en normas internacionales como la ISO 14224:2016 y la UNE-EN 17007:2018, que pueden complementarse con modelos de aprendizaje automático orientados a la predicción de fallas y de vida útil remanente.

En este contexto, la aplicación de la norma ISO 14224:2016 ofrece un marco para disponer de registros de confiabilidad y mantenimiento con una estructura estandarizada (ISO Norma 14224, 2016). Al integrarse con los procesos e indicadores definidos en la UNE-EN 17007:2018, se asegura la trazabilidad de los datos y cumplimiento de objetivos. A su vez, esta complementariedad se alinea con los requisitos 6.1, 7.6 y 8.1 de la ISO 55001:2024, que establecen la gestión de riesgos y oportunidades, la definición y uso de datos de calidad, y el control de los procesos del ciclo de vida de los activos (BS ISO Norma 55001, 2024).

Estado del arte

Para el desarrollo de este proyecto se considera el estado del arte en metodologías de confiabilidad, mantenibilidad y aprendizaje automático aplicadas al sector eléctrico. El propósito es identificar las técnicas más utilizadas en la industria para el pronóstico de fallas y la estimación de vida útil remanente, así como reconocer tendencias, ventajas y limitaciones de cada enfoque.

Diversos estudios exploran aprendizaje automático para diagnóstico y predicción de fallas en transformadores. Biradar et al. (2024) investigan la aplicación de aprendizaje automático para desarrollar modelos de mantenimiento predictivo en transformadores con datos históricos de pruebas de aceite dieléctrico, temperatura de devanados y registros de corriente y tensión de operación. Los autores concluyen que el XGBoost, Random Forest y Extra Trees ofrecen el mejor equilibrio entre exactitud, precisión y sensibilidad.

Kabir et al. (2018) plantean un modelo de mantenimiento predictivo (PdM) para transformadores de distribución que sustituye la estrategia *run-to-failure*. Su metodología se fundamenta en Random Forest (RF) y Random Under Sampling Boost (RUSBoost), con uso de datos rutinarios y de bajo costo como especificaciones del activo, registros de carga, localización geográfica y condiciones climáticas en lugar de análisis de gases disueltos (DGA).

La validación en más de 700 mil transformadores muestra que ambos algoritmos superan las prácticas convencionales, con un desempeño superior de RUSBoost respecto a RF. Sudrajat et al. (2023) aplican redes neuronales (ANN) y Naive Bayes para modelar la relación entre los valores del índice de salud y estimar la edad de transformadores.

Otros estudios combinan los enfoques de la estadística clásica con aprendizaje automático con el objeto de mejorar la exactitud, como lo mencionan Zhang et al. (2019) al integrar el modelo físico de Eyring con soporte vectorial (SVM) sobre una base probabilística de Weibull e introducen un coeficiente de ponderación que equilibra los aportes del modelo físico y del modelo de machine learning (ML).

Cvijic et al. (2023) presentan POWERED, un enfoque híbrido que integra modelos físicos, bayesianos y de ML para diagnóstico y estimación de vida útil remanente en transformadores.

Adicionalmente, se encuentran estudios que mantienen un enfoque puramente físico. Syamsiana et al. (2024) usan curvas de carga diarias y estacionales para calcular la pérdida de vida acumulada con el modelo térmico definido en IEC 60076-7, que relaciona la temperatura del punto caliente con la aceleración de la degradación del papel aislante.

La literatura reconoce a los modelos estadísticos como base de los estudios de confiabilidad en transformadores y otros equipos eléctricos. Freeman (1996) demuestra, a partir de sus análisis de la distribución de Weibull y sus parámetros, que una selección incorrecta de la distribución estadística para modelar los datos de fallas incrementa la incertidumbre respecto al momento óptimo de reemplazo e induce a adoptar políticas de mantenimiento equivocadas de manera sistemática.

Lo que plantean Zhou et al. (2012) es un enfoque Bayesiano jerárquico que integra registros históricos de fallas con el conocimiento experto de los ingenieros de mantenimiento. Hong et al. (2009) estudian la problemática recurrente de registros incompletos en el análisis de confiabilidad de transformadores y, con la distribución de Weibull como referencia, formulan un estimador de máxima verosimilitud que incorpora información de equipos fallados y de los que permanecen en servicio.

Ji et al. (2010) desarrollan un esquema de decisión fundamentado en un modelo de riesgos proporcionales de Weibull (PHM), que integra la condición del activo y su edad para estimar la tasa de falla y definir el momento de intervención.

Por otra parte, Sotiropoulos et al. (2007), Jiang et al. (2019) y Selva et al. (2018) proponen el uso de modelos ocultos de Markov (HMM) para estimar indicadores de salud de transformadores a partir de variables de condiciones (DGA, humedad y furanos).

La literatura también aborda el problema desde el punto de vista de la optimización de la gestión de mantenimiento y costos. Huang & Chiu (2024), para optimizar la programación de mantenimiento de transformadores de distribución, incorporan Weibull y condiciones ambientales para alimentar el algoritmo de la planta carnívora (CPA).

Asimismo, Dou et al. (2024) integran vida técnica y vida económica para calcular reemplazo óptimo de transformadores de 10 kV. La vida técnica se estima con LS-SVM optimizada por Grey Wolf, usa variables de operación y condición; la vida económica considera costos de falla, mantenimiento y valor del dinero en el tiempo.

Qiao & Wang (2024) proponen una estrategia de mantenimiento basado en condición (CBM) que combina costo y confiabilidad y se resuelve con un algoritmo Optimización por Enjambre de Partículas (PSO). En el caso de transformadores de distribución ilustra cómo el algoritmo decide entre inspección, mantenimiento o retiro, en función del balance entre crecimiento de la tasa de falla y valor económico remanente.

Agarwal et al. (2020) aplican Weibull para analizar la confiabilidad de un sistema radial de distribución, considerando explícitamente el envejecimiento de los transformadores y calculan Energía No Suministrada (ENS) y el Costo Esperado de Interrupción (ECOST).

Finalmente, la inteligencia artificial (IA) se consolida como un eje transversal en la gestión de activos eléctricos. La incorporación de algoritmos de aprendizaje automático y analítica avanzada permite anticipar fallas potenciales y definir acciones de mantenimiento más eficientes. Así, la IA se posiciona como herramienta estratégica para mejorar confiabilidad y disponibilidad, además de constituir un pilar de la transformación digital en las organizaciones del sector eléctrico (Carolina et al., 2025).

Marco metodológico

El desarrollo del presente proyecto de profundización se fundamenta en los conocimientos adquiridos en las asignaturas de la maestría en Gestión de Activos y Mantenimiento de la Universidad del Valle, las cuales aportan un marco metodológico y conceptual que se puede aplicar para materializar los beneficios del objetivo del presente trabajo.

En la asignatura *Gestión de la información y medición del desempeño e indicadores de SGA* se abordan los procesos de recopilación, almacenamiento y procesamiento de los datos, (ISO Norma 14224, 2016), así como la construcción de indicadores estratégicos, tácticos y operativos en referencia con las normas ISO 14224:2016 y la UNE EN 15341:2020. Lo que facilita la depuración de la información y construcción del dataset e interpretación de resultados en la validación de las estrategias de mantenimiento y confiabilidad propuestas.

La asignatura *Herramientas de confiabilidad* es uno de los ejes de este proyecto, al brindar los fundamentos necesarios para aplicar los modelos de estadística clásica y la comprensión de metodologías avanzadas. Diversos trabajos aplicados en el sector eléctrico han demostrado la utilidad de estas técnicas. Por ejemplo, el uso de modelos de riesgo proporcionales de Weibull (PHM) para predecir fallas en equipos de generación (M. A. T. García, 2025).

El proyecto se beneficia de estas herramientas para analizar datos históricos de fallas, modelar el comportamiento de los activos y cuantificar el riesgo, para conformar los pilares de este estudio.

La asignatura *Mantenimiento dentro de la gestión de activos* proporciona una visión sistémica de los procesos de planeación, programación, ejecución y control del mantenimiento, considerando factores técnicos, humanos y económicos. Esta materia proporciona las herramientas para articular los objetivos de la organización con las decisiones de mantenimiento garantizando la sostenibilidad y el valor generado por los activos a lo largo de

su ciclo de vida ya que se fundamenta en los estándares internacionales ISO 55001, UNE EN 16646 y UNE EN 17007.

El presente proyecto corresponde a una investigación de tipo explicativo–predictivo, dado que busca identificar y correlacionar variables que influyen en la ocurrencia de fallas en transformadores de distribución MV/LV, para posteriormente formular un modelo que permita predecir la probabilidad de falla y estimar la vida útil remanente de los equipos. La información que se utiliza para el desarrollo de este proyecto es principalmente artículos y libros de expertos en las materias de mantenimiento, inteligencia artificial y sector eléctrico.

Adicionalmente, se cuenta con una base de datos desde 2022 con históricos de fallas y datos de operación de casi 60 mil transformadores de distribución, complementados con documentación técnica de fabricantes y normativas internacionales (ISO 14224:2016, UNE-EN 17007:2018, ISO 55001:2014).

El enfoque metodológico combina dos vertientes: análisis estadístico de confiabilidad (distribuciones de vida, MTBF, MTTR, tasas de falla) y algoritmos de machine learning aplicados al mantenimiento predictivo de transformadores (Random Forest, Support Vector Machines, Redes Neuronales, entre otros).

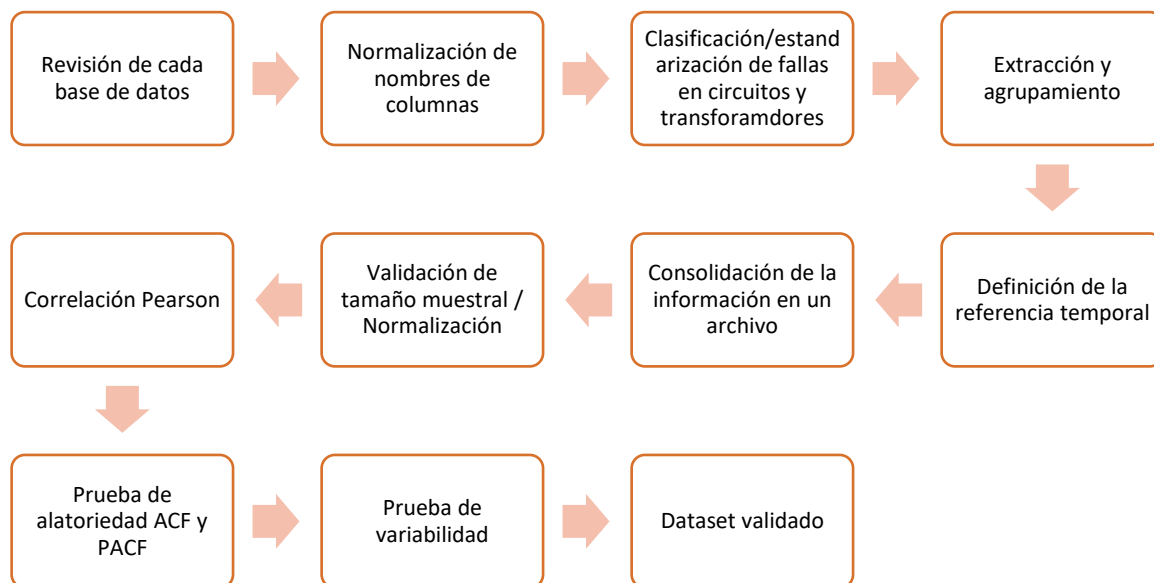
De acuerdo con los objetivos planteados, la metodología se estructura en fases que agrupan actividades específicas.

Fases o etapas del proyecto

Etapas 1 – Preparación y organización de datos (Objetivo 1: datos)

Ilustración 1

Desarrollo Objetivo 1



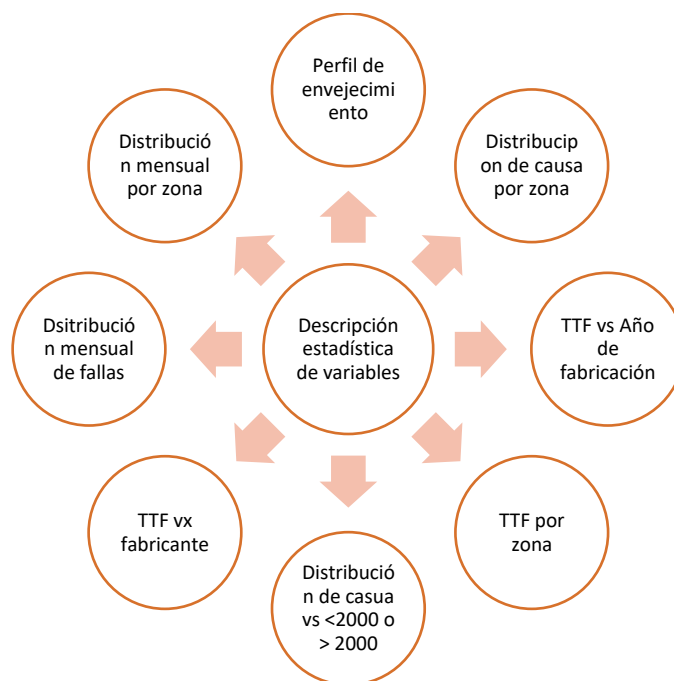
Nota. La ilustración muestra las etapas del desarrollo del objetivo 1. Fuente: Elaboración propia con base en el método de obtención de data burda de (Mora Gutiérrez, 2022)

Resultado esperado: dataset confiable y estandarizado, listo para análisis.

Etapas 2 – Análisis exploratorio y caracterización (Objetivo 2: patrones)

Ilustración 2

Desarrollo Objetivo 2



Resultado esperado: conjunto de variables críticas y criterios de predicción.

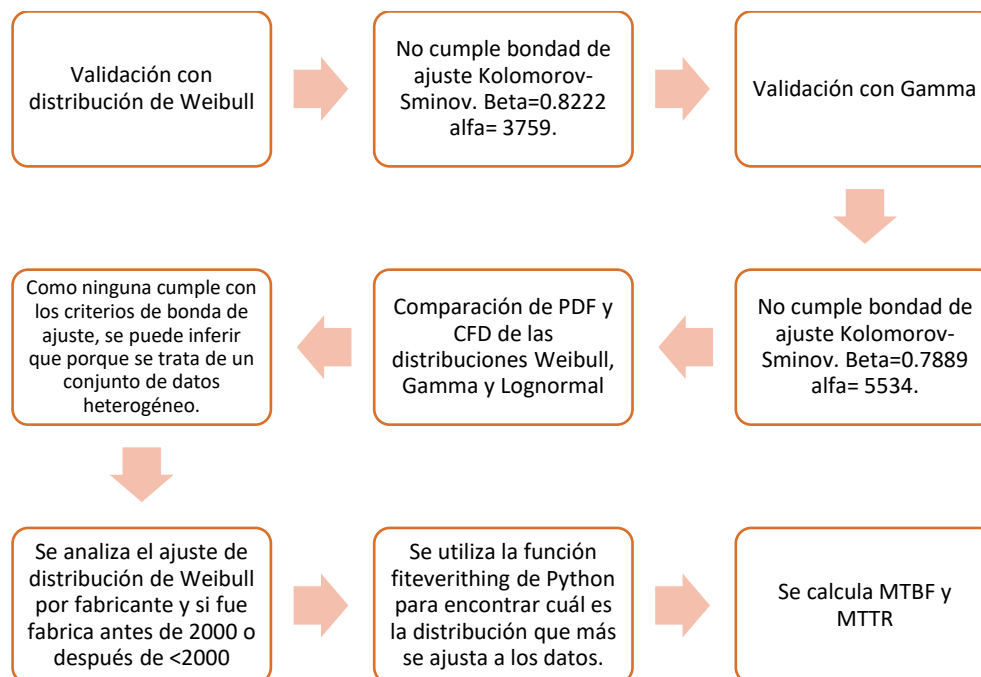
Nota. La ilustración muestra los gráficos analizados en el desarrollo del objetivo 2.

Fuente: Elaboración propia con base en los datos y las recomendaciones de exploración de datos de (Kadre & Reddy Konasani, 2021)

Etapas 3 – Análisis estadístico de confiabilidad (Objetivo 3: modelos)

Ilustración 3

Desarrollo Objetivo 3



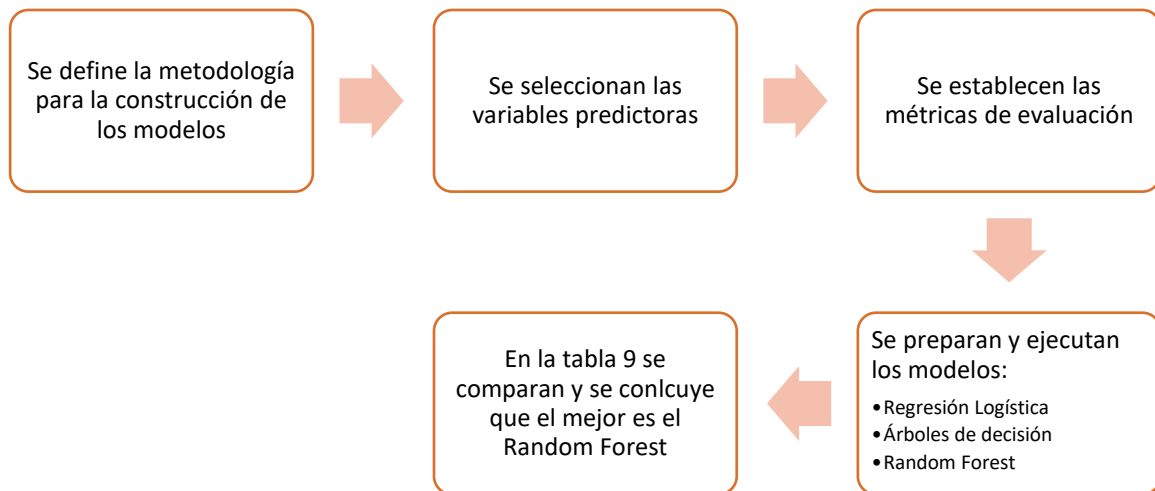
Nota. La ilustración muestra las etapas del desarrollo del objetivo 3. Fuente: Elaboración propia con base en el método de cálculo CMD de Mora Gutiérrez (2009).

Resultado esperado: Modelo probabilístico sólido y parámetros confiables.

Etapas de desarrollo del objetivo 4: análisis

Ilustración 4

Desarrollo Objetivo 4



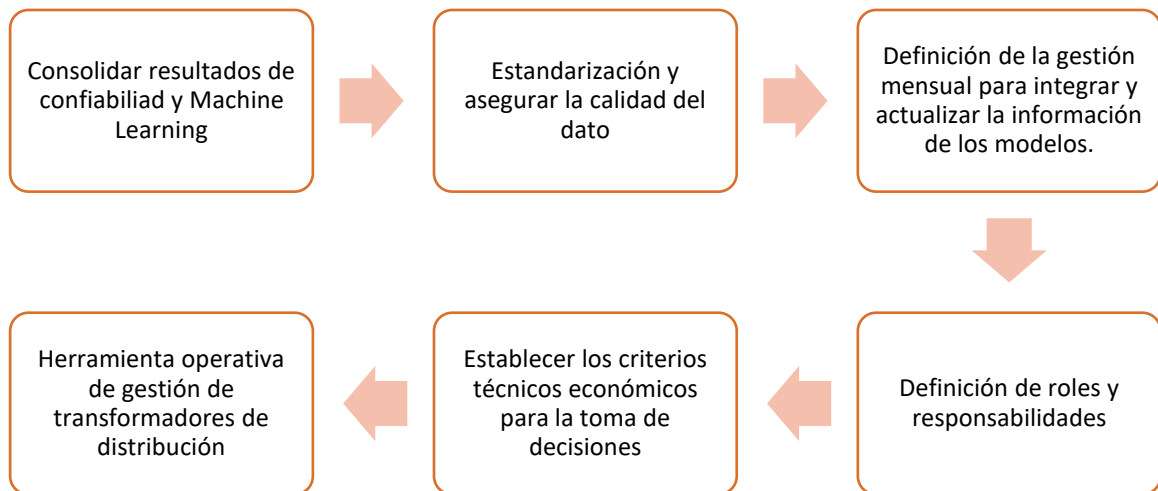
Nota. La ilustración muestra las etapas del desarrollo del objetivo 4. Fuente: Elaboración propia con base en los procedimientos de construcción de modelos de Kadre & Reddy Konasani (2021).

Resultado esperado: Modelo predictivo con desempeño cuantificado y documentado.

Etapas 5 – Propuesta del proceso de apoyo a la gestión de transformadores de distribución MT/BT (Objetivo 5: proceso)

Ilustración 5

Desarrollo Objetivo 4



Nota. La ilustración muestra las etapas del desarrollo del objetivo 5. Fuente: Elaboración propia con base en las buenas prácticas y requisitos de las normas 55001, 17007 y 14224.

Resultado esperado: Proceso estructurado y aplicable por la empresa para gestión predictiva y toma de decisiones.

Desarrollo

Objetivo uno – Datos

Para llevar a cabo un correcto análisis CMD y el cálculo de predicción y probabilidad de fallas en transformadores de distribución, el primer paso consiste en la consolidación de una base de datos técnica confiable y estructurada. La calidad de los resultados en las decisiones de mantenimiento y gestión de activos depende directamente de la calidad, completitud y trazabilidad de los datos recopilados sobre los activos (ISO Norma 14224, 2016).

Por lo tanto, este primer objetivo tiene como propósito identificar, verificar y estructurar los datos disponibles de los transformadores de distribución de media y baja tensión (MT/BT) y de los eventos de los circuitos adyacentes, con coherencia técnica y temporal para los análisis posteriores. Para esto se usa como base el marco metodológico del estadio descriptivo y el proceso de obtención de data burda de A. Mora Gutiérrez (2022).

El resultado corresponde a la creación de una base de datos maestra (base_trafos_consolidada.xlsx), que integra información técnica, operacional y contextual de 17.109 transformadores con eventos registrados entre enero de 2020 y agosto de 2025. Con esto se cumple el requisito de contar con un volumen de información suficiente para que la estimación de los parámetros de confiabilidad se aproxime al valor real (Mora Gutiérrez, 2009).

En la fase inicial del proceso descriptivo se debe determinar si el volumen y la calidad de los datos disponibles son adecuados para garantizar la validez de los análisis. La cantidad mínima de información constituye el primer control sobre la “data bruta” y se define a partir de la disponibilidad real de registros y su correspondencia con los parámetros de tiempo analizados. En consecuencia, se recomienda disponer de al menos 30 registros válidos por grupo o variable analizada, de modo que la media muestral se asemeje a la media poblacional y se garantice un fundamento estadístico confiable (Mora Gutiérrez, 2022).

En este proyecto, esta directriz se aplica para validar la información consolidada en la base base_trafos_consolidada.xlsx, verificando la cantidad de transformadores, fallas y eventos

de circuito disponibles en cada categoría analizada. De esta forma, se asegura que los resultados de los análisis descriptivos y de los posteriores modelos de confiabilidad y aprendizaje automático se sustenten en un conjunto de datos representativo y estadísticamente suficiente.

Para la consolidación del archivo maestro, se integraron tres fuentes principales de información que se muestran en la Tabla 1.

Tabla 1
Descripción de las fuentes de información

Fuente	Descripción	Cobertura temporal	Observaciones
Base de transformadores	Contiene información técnica básica del transformador (fabricante, capacidad, año de instalación o reposición).	Según año de instalación.	Incluye únicamente transformadores en servicio. Se usan solo registros de transformadores instalados desde 2020.
Base de eventos en transformadores	Contiene fecha, modo y causa de la falla, además del fabricante y la capacidad del transformador afectado.	Enero de 2020 a agosto de 2025.	Presenta casos con placas ilegibles y asignaciones de modos y causas de fallas no estandarizadas (39 causas diferentes).
Base de eventos en circuitos	Contiene causas de disparo de los circuitos de MT	Agosto de 2022 a agosto de 2025.	Presenta causas no estandarizadas (80 categorías).

Nota: La tabla muestra la descripción general de las tres fuentes para la consolidación del archivo maestro. Fuente: Elaboración a partir de la Base de transformadores (Celsia Colombia SA ESP, 2025), la Base de eventos en transformadores (Celsia Colombia SA ESP, 2020-2025) y la base de eventos en circuitos de MT (Celsia Colombia SA ESP, 2022-2025).

Para la integración de la información se emplean métodos adecuados de limpieza y transformación bajo un enfoque ETL (Extract, Transform, Load) con ayuda de Python (librerías pandas, numpy, matplotlib y seaborn) como herramienta para el manejo de grandes cantidades de datos dentro del entorno Google Colab. Con esto se logra convertir la data original en la entrada de los modelos de análisis posteriores. (Mora Gutiérrez, 2022)

Cada registro de transformador se relaciona a su respectivo circuito y se añaden nuevas variables para contabilizar los eventos ocurridos en el circuito en ventanas de 1, 3, 6 y 12 meses previos a cada evento del transformador. Con esto se busca encontrar correlación entre el entorno operativo con el comportamiento de cada activo.

Figura 6

Proceso ETL para la integración de la información



Nota. El esquema representa las fases del proceso ETL utilizado para integrar la información de la Base de Transformadores, la Base de Eventos en Transformadores y la Base de Eventos en Circuitos MT. Fuente: Elaboración propia a partir de datos extraídos de la Base de transformadores (Celsia Colombia SA ESP, 2025), la Base de eventos en transformadores

(Celsia Colombia SA ESP, 2020-2025) y la base de eventos en circuitos de MT (Celsia Colombia SA ESP, 2022-2025).

El archivo base_trafos_consolidada.xlsx contiene las siguientes columnas principales:

- **Identificación y operación:** custom_id, ciclo, fecha inicio operación, fecha de falla, fecha de normalización, duración falla (días), días de operación, ¿ha fallado?, censura derecha.
- **Datos técnicos:** fabricante, kva, fecha de fabricación, causa falla - bol, fecha_repos.
- **Ubicación y entorno:** zona, altitud, longitud, latitud, nom_cto, nom_cto_norm.
- **Variables contextuales:** conteos de eventos del circuito en ventanas de 1, 3, 6 y 12 meses (cto_fail_count_1m, cto_fail_count_3m, cto_fail_count_6m, cto_fail_count_12m), desagregados por tipo de causa (vegetación, animal, atmosférica, externa, etc.) como se observa en la **¡Error! No se encuentra el origen de la referencia..**

Esta estructura facilita los análisis CMD, correlaciones entre fallas y entorno operativo, y posteriormente la aplicación de algoritmos de aprendizaje automático supervisado (Objetivos 2 y 3).

Durante la clasificación de las variables se identifican 80 causas distintas de disparos de los circuitos; las cuales se enfocan más en explicar o definir el mecanismo de falla que la misma causa. Por tal motivo, con base en esa información y lo propuesto por ISO Norma 14224 (2016), se agrupan en 10 categorías principales, como se observan en la Tabla 2 y Tabla 3.

Tabla 2

Clasificación de fallas en circuitos eléctricos (adaptada de ISO 14224:2016)

Categoría	Descripción
Fallas eléctricas internas	Defectos eléctricos en elementos principales del sistema (aislamiento, conexiones, sobrecalentamiento)

Fallas mecánicas	Problemas en componentes mecánicos del sistema (soportes, retenidas, crucetas, apoyos)
Fallas en sistemas de protección y control	Mal funcionamiento o coordinación incorrecta en protecciones
Errores humanos u operacionales	Acciones incorrectas de operación o mantenimiento
Actividades de mantenimiento programado o correctivo	Intervenciones planeadas o no planeadas que implican la desconexión del circuito
Factores externos (ambientales y naturales)	Condiciones ambientales, fenómenos atmosféricos o agentes biológicos que afectan el sistema
Interferencias o daños de terceros	Intervenciones externas ajenas al operador del sistema
Fallas en red secundaria o elementos de conexión	Problemas en líneas, ramales, puentes o elementos de seccionamiento.
Problemas en servicios auxiliares o comunicaciones	Fallas en suministro auxiliar o sistemas de supervisión
Causas desconocidas	Causas no identificadas

Nota. La tabla presenta la clasificación de fallas en circuitos eléctricos utilizada para el análisis del estudio. La estructura corresponde a una adaptación operativa de la taxonomía establecida en la ISO 14224:2016 para el registro y categorización de fallas en sistemas eléctricos. Fuente: Elaboración propia a partir de datos de la base de eventos en circuitos MT (2022-2025) y las directrices de clasificación de fallas definidas en la ISO 14224:2016.

Tabla 3

Clasificación de fallas en transformadores de distribución (adaptada de ISO 14224:2016)

Categoría	Descripción
Fallas eléctricas internas del transformador	Deterioro o defecto interno en los devanados, conexiones, bobinas o aislamiento.
Fallas por sobrecarga o envejecimiento	Estrés térmico o degradación por fin de vida útil o por carga excesiva de manera continua.

Factores ambientales y climáticos	Condiciones atmosféricas extremas o eventos naturales con impacto en la integridad del equipo.
Fallas mecánicas o estructurales	Daños en la estructura física que sostiene el transformador o en sus componentes externos.
Fallas externas del sistema eléctrico (red MT/BT)	Defectos o eventos en la red externos al transformador que generan cortocircuitos.
Interferencias o daños de terceros	Manipulación sin autorización o acciones que producen daños en el transformador o en sus elementos.
Defectos de componentes o conexiones	Fallas en bornes, contactos, empalmes u otros accesorios eléctricos del transformador.
Problemas de diseño o fabricación	Defectos de manufactura o errores en el diseño del transformador o de sus partes internas.
Causas desconocidas o no determinadas	Situaciones donde no se pudo identificar una causa raíz específica.
Factores combinados o indeterminados del entorno	Interacciones entre varios factores (ambientales, mecánicos, eléctricos) sin predominancia clara.

Nota. La tabla presenta la clasificación de las fallas en transformadores de distribución MT/BT; se agrupan en diez categorías principales. La estructura corresponde a una adaptación técnica de la taxonomía de fallas para equipos eléctricos definida en la ISO 14224:2016

Fuente: elaboración propia con base en información de la base de eventos en transformadores (2020–2025) y en los lineamientos de clasificación de fallas de la ISO 14224:2016.

Este proceso permite homogenizar la variable Causa Falla - Bol, al eliminar ambigüedades semánticas y facilita el análisis posterior. No obstante, esto deja en evidencia la necesidad de reforzar esta parte del proceso según lo propuesto en la etapa 5.

Para la verificación de la calidad y coherencia de los datos, se realiza una comprobación de que los eventos siguen una secuencia lógica. Para la coherencia cronológica se organiza el dataset de tal manera que se muestren los ciclos de vida de cada transformador con base en la información registrada de forma organizada y lógica.

Tabla 4

Estructura del dataset – cronología de eventos

Custom_ID	Ciclo	Fecha inicio operación	Fecha de falla	Fecha de normalización	Duración falla (días)	Fecha de fabricación	Días de operación
210000000547371	1	1/01/2008	25/03/2021	26/03/2021	1	2008	4832
210000000547371	2	26/03/2021	21/06/2021	2/07/2021	11	2019	87
210000000547371	3	2/07/2021				2020	1521

Nota. En la tabla se muestra la estructura del dataset con la que se asegura la coherencia cronológica y lógica de cada ciclo de vida del transformador: Fuente: Elaboración propia con base en información de la Base de Transformadores (2025) y de la Base de Eventos en Transformadores (2020–2025).

Con respecto al tratamiento de censura y datos incompletos, los transformadores que permanecen en servicio sin fallar se consideran censurados a la derecha. En la base consolidada, estos registros se marcan en la columna “Censura Derecha”.

Para determinar qué registros de transformadores son incluidos en este estudio, se utiliza la información de eventos en transformadores que se encuentra disponible a partir de enero de 2020, por lo tanto, se adopta dicho año como punto de referencia temporal para la conformación del conjunto de datos. La Figura 7 presenta de manera esquemática los diferentes escenarios según la disponibilidad de fechas de puesta en operación (FPO) y de falla (FF), así como la clasificación correspondiente de cada caso (válido, censurado o excluido).

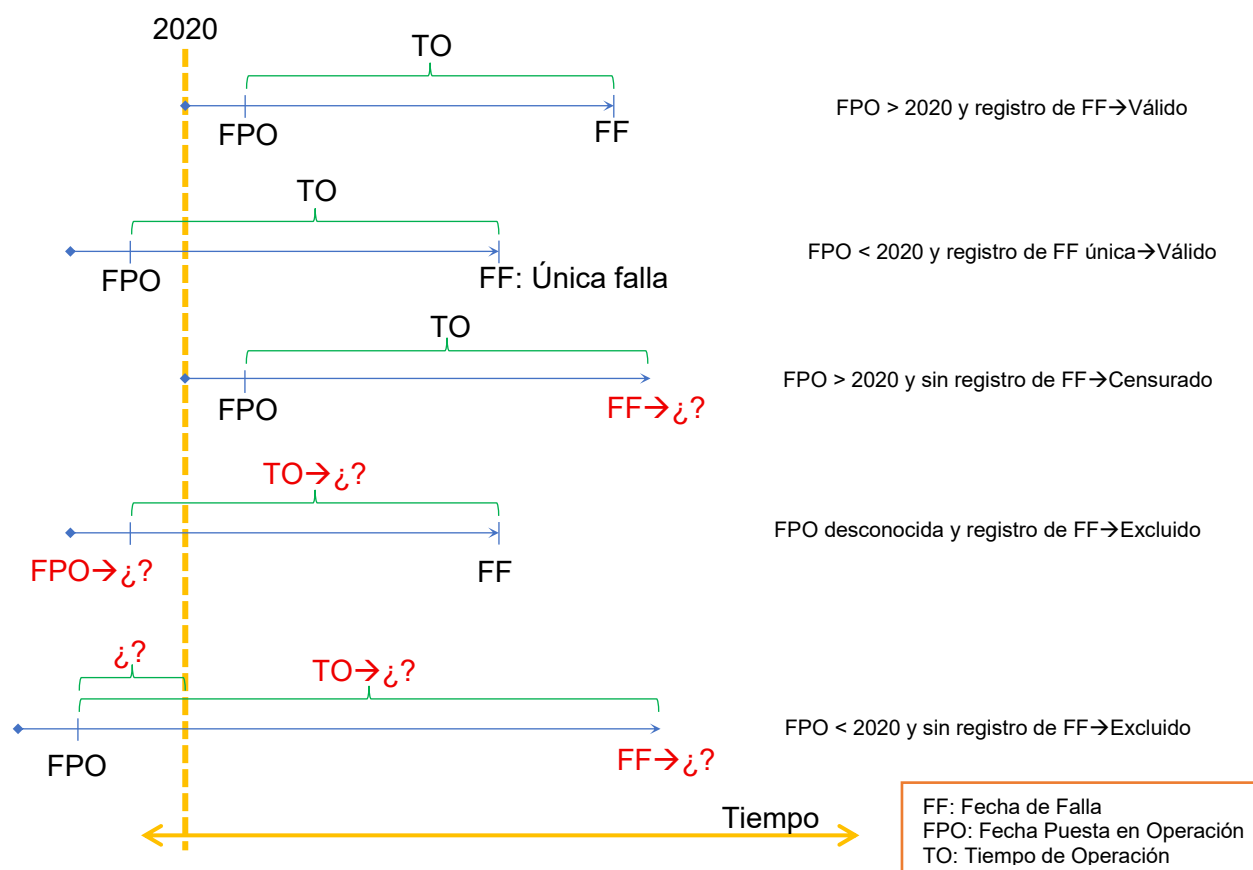
En resumen, los activos con fecha de puesta en operación (FPO) posterior a 2020 y con registro de falla (FF) son válidos, al contar con información temporal completa; mientras que aquellos con FPO posterior a 2020 y sin evidencia de falla se clasifican como censurados a la derecha, al permanecer en servicio durante el periodo de observación.

Los transformadores instalados antes de 2020 se incluyen siempre que presenten solo una falla posterior a esa fecha y conserven coherencia cronológica entre registros. Finalmente,

los casos con FPO desconocida, sin fecha de falla válida o con FPO anterior a 2020 sin información adicional entre esa fecha y 2020 se excluyen del estudio, aunque se mantienen para análisis descriptivos de frecuencia y para la identificación posterior de tendencias generales de fallas, fabricantes y zonas con mayor incidencia.

Figura 7

Criterios de selección de registros



Nota. La figura resume los criterios que se emplean para seleccionar los registros válidos del análisis, se considera el 2020 como el inicio del periodo de observación y las condiciones de censura o exclusión de los datos. Fuente: Elaboración propia con base en información de la base de transformadores (2025) y de la base de eventos en transformadores (2020–2025).

Obtención y limpieza de la data burda

Para proceder con los cálculos de confiabilidad, es necesario aplicar el proceso de obtención de la data burda propuesto por (Mora Gutiérrez (2022) en el estadio descriptivo. A partir de la base consolidada se procede a cuantificar la cantidad de datos. Antes, es importante aclarar que los tiempos a evaluar corresponden a tiempos hasta la falla (TTF) y tiempos de reposición (TTR). En la Tabla 5 se presenta la cantidad de datos de acuerdo con el tipo de registro.

Tabla 5

Resumen de registros

Descripción del dato	Número de datos	Porcentaje de completitud
Total de registros	17109	NA
Registros con fechas de falla	7933	100%
Registros con Duración de falla (DT)	7933	100%
Registros con Tiempo de Operación (UT) (incluye censurados)	16390	96%
Registros con Tiempo de Operación (UT) (sin censura)	7214	90%
Registros de datos con censura	9176	100%
Registros con fabricante	7398	93%
Registros con fecha de fabricación	16319	95%
Registros con capacidad	7918	99.8%
Registros con ubicación geográfica	17106	99.9%

Nota. La tabla presenta el resumen de los registros en la base de datos maestra y muestra el nivel de completitud de cada tipo de información, de acuerdo con los criterios definidos para el análisis. Fuente: Elaboración propia con base en información de la base de transformadores (2025) y de la base de eventos en transformadores (2020–2025).

Figura 8

Información Dataset completo

```

RangeIndex: 17109 entries, 0 to 17108
Data columns (total 70 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   custom_id                             17109 non-null  int64
1   ciclo                                 17109 non-null  int64
2   fecha_inicio_operación                 16390 non-null  datetime64[ns]
3   fecha_de_falla                         7933 non-null   datetime64[ns]
4   fecha_de_normalización                 7933 non-null   datetime64[ns]
5   dt                                     7933 non-null   float64
6   fecha_de_fabricación                  16319 non-null  float64
7   edad                                  16319 non-null  float64
8   ttf                                    16390 non-null  float64
9   ha_fallado?                           17109 non-null  category
10  censura_derecha                       17109 non-null  category
11  fabricante                             7933 non-null   object
12  kva                                    7918 non-null   float64
13  causa_falla_trafo                     17109 non-null  object
14  fecha_repos                            17109 non-null  datetime64[ns]
15  altitud                               17106 non-null  float64
16  zona                                   17109 non-null  object
17  longitud                              17106 non-null  float64
18  latitud                               17106 non-null  float64
19  nt                                     17109 non-null  float64
20  nom_cto                               17109 non-null  object
21  nom_cto_norm                          17109 non-null  object
22  cto_fail_1m                           17109 non-null  category
23  cto_fail_count_1m                     17109 non-null  int64

```

Nota. La figura presenta la salida del método `info()` del paquete `pandas` en `Python`. El resultado muestra las columnas, la cantidad de valores presentes en cada una y el tipo de dato que el entorno asigna durante la carga. Fuente: Elaboración propia a partir del uso del método `info()` de la biblioteca `pandas` en `Python`, con lectura del archivo maestro desde `Google Colab`.

Extracción y agrupamiento

Para verificar la consistencia del dataset, se seleccionan aleatoriamente transformadores y se contrastan los eventos asociados con los registros originales de fallas de circuito. Adicionalmente se realizan las etapas de extracción y agrupamiento como se observa en la Figura 9 y Figura 10.

Figura 9

Estructura descriptiva de los datos de TTF en la etapa de extracción

	A	B	C	D	E	F	G	H	I	J
	Global	Fecha inicio operación	Fecha de Falla	Días de operación		Año Mes	Días de operación		Fecha	Valor días promedio mes
2	1	1/01/1996	2/01/2020	8767		Año 1 - Mes 1	8767		Año 1 - Mes 1	3931,5
3	2	1/01/2019	2/01/2020	366		Año 1 - Mes 1	366		Año 1 - Mes 2	5032,6
4	3	1/01/2016	3/01/2020	1463		Año 1 - Mes 1	1463		Año 1 - Mes 3	4662,3
5	4	1/01/2017	3/01/2020	1097		Año 1 - Mes 1	1097		Año 1 - Mes 4	4597,9
6	5	1/01/2007	4/01/2020	4751		Año 1 - Mes 1	4751		Año 1 - Mes 5	3780,4
7	6	1/01/2018	4/01/2020	733		Año 1 - Mes 1	733		Año 1 - Mes 6	4052,6
8	7	1/01/2019	4/01/2020	368		Año 1 - Mes 1	368		Año 1 - Mes 7	4501,3
9	8	1/01/1997	5/01/2020	8404		Año 1 - Mes 1	8404		Año 1 - Mes 8	4669,8
10	9	1/01/2008	5/01/2020	4387		Año 1 - Mes 1	4387		Año 1 - Mes 9	3820,8
11	10	1/01/2006	6/01/2020	5118		Año 1 - Mes 1	5118		Año 1 - Mes 10	4466,9
12	11	1/01/2017	6/01/2020	1100		Año 1 - Mes 1	1100		Año 1 - Mes 11	4043,4
13	12	1/01/2019	6/01/2020	370		Año 1 - Mes 1	370		Año 1 - Mes 12	4320,7
14	13	1/01/2009	7/01/2020	4023		Año 1 - Mes 1	4023		Año 2 - Mes 1	3105,4
15	14	1/01/1995	8/01/2020	9138		Año 1 - Mes 1	9138		Año 2 - Mes 2	4232,4
16	15	1/01/2001	8/01/2020	6946		Año 1 - Mes 1	6946		Año 2 - Mes 3	3488,9
17	16	1/01/2007	8/01/2020	4755		Año 1 - Mes 1	4755		Año 2 - Mes 4	5158,9
18	17	1/01/2016	8/01/2020	1468		Año 1 - Mes 1	1468		Año 2 - Mes 5	4443,2
19	18	1/01/2016	8/01/2020	1468		Año 1 - Mes 1	1468		Año 2 - Mes 6	4652,5
7201	7200	6/12/2020	19/07/2025	1686		Año 6 - Mes 7	1686		Año 5 - Mes 12	3438,4
7202	7201	1/01/1986	20/07/2025	14445		Año 6 - Mes 7	14445		Año 6 - Mes 1	2725,4
7203	7202	29/05/2024	20/07/2025	417		Año 6 - Mes 7	417		Año 6 - Mes 2	4013,2
7204	7203	14/05/2024	21/07/2025	433		Año 6 - Mes 7	433		Año 6 - Mes 3	3635,3
7205	7204	1/01/2005	23/07/2025	7508		Año 6 - Mes 7	7508		Año 6 - Mes 4	3492,8
7206	7205	1/01/2002	24/07/2025	8605		Año 6 - Mes 7	8605		Año 6 - Mes 5	3947,5
7207	7206	18/02/2023	24/07/2025	887		Año 6 - Mes 7	887		Año 6 - Mes 6	3647,2
7208	7207	20/05/2024	24/07/2025	430		Año 6 - Mes 7	430		Año 6 - Mes 7	3453,9
7209	7208	12/03/2025	24/07/2025	134		Año 6 - Mes 7	134		Tiempo total	276898
7210	7209	30/04/2024	25/07/2025	451		Año 6 - Mes 7	451		MUT	4133
7211	7210	1/01/2024	27/07/2025	573		Año 6 - Mes 7	573		Desv	518
7212	7211	25/10/2024	27/07/2025	275		Año 6 - Mes 7	275		Variabilidad	13%
7213	7212	1/01/1992	28/07/2025	12262		Año 6 - Mes 7	12262			
7214	7213	1/01/2018	28/07/2025	2765		Año 6 - Mes 7	2765			
7215	7214	16/05/2025	28/07/2025	73		Año 6 - Mes 7	73			
7216			Tiempo total	29897845						
7217			MUT	4144						
7218			Desv	4436						
7219			Variabilidad	107%						

Nota. La figura presenta la organización inicial de los datos de TTF y TTR en la actividad de extracción del estadio descriptivo. La estructura ordena fechas, días en operación y ciclos del activo para formar la base descriptiva previa al agrupamiento. Fuente: Elaboración propia con base en datos del archivo maestro del estudio y en la secuencia de etapas del proceso de obtención de la data burda del estadio descriptivo descrito en (A. Mora Gutiérrez, 2022).

Figura 10

Agrupamiento de datos TTF (MUT) y TTR (MDT) por meses

	A	B	C	D	E
1		MUT		MDT	
2		FECHA	VALOR DÍAS PROMEDIO MES	FECHA	VALOR DÍAS PROMEDIO MES
3	1	Año 1 - Mes 1	3931,53	Año 1 - Mes 1	2,74
4	2	Año 1 - Mes 2	5032,58	Año 1 - Mes 2	1,93
5	3	Año 1 - Mes 3	4662,28	Año 1 - Mes 3	3,57
6	4	Año 1 - Mes 4	4597,91	Año 1 - Mes 4	2,46
7	5	Año 1 - Mes 5	3780,36	Año 1 - Mes 5	3,11
8	6	Año 1 - Mes 6	4052,57	Año 1 - Mes 6	2,63
9	7	Año 1 - Mes 7	4501,28	Año 1 - Mes 7	2,52
10	8	Año 1 - Mes 8	4669,83	Año 1 - Mes 8	2,30
11	9	Año 1 - Mes 9	3820,84	Año 1 - Mes 9	2,09
12	10	Año 1 - Mes 10	4466,93	Año 1 - Mes 10	1,94
13	11	Año 1 - Mes 11	4043,35	Año 1 - Mes 11	1,66
14	12	Año 1 - Mes 12	4320,66	Año 1 - Mes 12	2,49
15	13	Año 2 - Mes 1	3105,40	Año 2 - Mes 1	3,02
16	14	Año 2 - Mes 2	4232,37	Año 2 - Mes 2	1,49
17	15	Año 2 - Mes 3	3488,89	Año 2 - Mes 3	2,95
18	16	Año 2 - Mes 4	5158,91	Año 2 - Mes 4	2,00
58	56	Año 5 - Mes 8	3999,21	Año 5 - Mes 8	3,88
59	57	Año 5 - Mes 9	4500,90	Año 5 - Mes 9	3,10
60	58	Año 5 - Mes 10	4144,02	Año 5 - Mes 10	3,50
61	59	Año 5 - Mes 11	3620,05	Año 5 - Mes 11	3,85
62	60	Año 5 - Mes 12	3438,43	Año 5 - Mes 12	2,71
63	61	Año 6 - Mes 1	2725,44	Año 6 - Mes 1	2,54
64	62	Año 6 - Mes 2	4013,17	Año 6 - Mes 2	3,92
65	63	Año 6 - Mes 3	3635,28	Año 6 - Mes 3	4,90
66	64	Año 6 - Mes 4	3492,80	Año 6 - Mes 4	5,05
67	65	Año 6 - Mes 5	3947,48	Año 6 - Mes 5	4,47
68	66	Año 6 - Mes 6	3647,17	Año 6 - Mes 6	4,39
69	67	Año 6 - Mes 7	3453,93	Año 6 - Mes 7	2,81
70		Tiempo total	276898	Tiempo total	243
71		MUT	4133	MDT	4
72		Desv	518	Desv	1
73		Variabilidad	13%	Variabilidad	37%

Nota. En la figura se presentan los valores en agrupación mensual de los parámetros TTF y TTR para 67 meses continuos. Fuente: Elaboración propia con base en datos del archivo maestro del estudio y en la secuencia de etapas del proceso de obtención de la data burda del estadio descriptivo descrito en (A. Mora Gutiérrez, 2022).

Validación del tamaño muestral

Como se puede observar en la Figura 11 y en la Figura 12, en cuanto a la validación de tamaño muestral, los resultados son satisfactorios tanto para la muestra de eventos como de meses. En la validación por eventos, los valores más críticos estimados para el tamaño muestral son del orden de los 3000 y 2500, lo que significa que el conjunto de datos de 7214 y 7933 para UT y DT respectivamente asegura la suficiencia de datos para las actividades estadísticas del proceso analítico (A. Mora Gutiérrez, 2022).

En la validación por meses, los resultados se aproximan al límite poblacional de 67 meses; aun así, los valores más críticos se mantienen por debajo. En conclusión, los tamaños muestrales están correctos y son pertinentes para el análisis estadístico ya que se cumple

incluso con niveles de confianza del 95 % y del 98 %, precisión máxima del 1 % y 2 % y valores de proporcionalidad dentro de los rangos aceptados para las actividades del estadio descriptivo (A. Mora Gutiérrez, 2022).

Figura 11

Validación de tamaño muestral de eventos

	A	B	C	D	E	F	G	H
1		Variable aleatoria		TO		Variable aleatoria		TR
2	1	1598	1	1921		5649	1	1
3	2	4600	1	561		3107	1	1
4	3	224	0	4453		4098	1	1
30	29	5031	1	778		714	1	2
31	30	2413	1	1017		6309	0	5
32	31	1794	1	3790		3571	0	11
33		Media muestral		4438,48				2,81
34		Media poblacional		4144,42				3,65
		Relación Media muestral/Media Poblacional		1,07				0,77
35		Tamaño poblacional		7214				7933
37		Media	0,645	4438,484		0,839		2,806
38		Desv st	0,486	5266,742		0,374		3,535
40		Proporción +	64,52%			83,87%		
41		Proporción -	35,48%			16,13%		
42		Población o univ	7214	7214		7933		7933
43		Error de precisión max deseado	1%	2%		1%		2%
44				177,539355				0,11225806
45		Rango de error permisibles más o menos		4438,483870				2,806451612
				96774 más o menos				90323 más o menos
				177,5393548				0,112258064
46				3871				516129
47		Probabilidad (Distrib Normal)	95%	98%		95%		98%
48		Nivel Z deseado	1,6449	2,0537		1,6449		2,0537
49		n con población finita	4468,424582	6965,45414		2903,597542		4526,17587
50			1,340710117	3,85074673		1,159215069		3,74335034
51			3332,878992	1808,85803		2504,796237		1209,1243

Nota. La figura muestra los resultados de las etapas de categorización y estandarización. Los resultados confirman que el volumen disponible en el dataset supera el mínimo requerido con los más de 7000 eventos registrados. Fuente: Elaboración propia con base en datos del archivo maestro del estudio y en la secuencia de etapas del proceso de obtención de la data burda del estadio descriptivo descrito en (A. Mora Gutiérrez, 2022).

Figura 12

Validación de tamaño muestral de meses

	A	B	C	D	E	F	G	H
1		Variable aleatoria		TO		Variable aleatoria		TR
2	1	65	1	3947,47788		7	1	2,51960784
3	2	29	1	3984,2439		51	1	3,48958333
4	3	13	1	3105,4		44	1	3,33846154
30	29	52	1	3665,3557		27	0	5,9862069
31	30	12	0	4320,66387		8	1	2,30434783
32	31	5	1	3780,36364		10	1	1,94160584
33		Media muestral		4126,59				3,39
34		Media poblacional		4132,81				3,62
		Relación Media muestral/Media Poblacional		1,00				0,94
35		Tamaño poblacional		67				67
36								
37								
38		Media	0,581	4126,594		0,677		3,394
39		Desv st	0,502	501,033		0,475		1,374
40								
41		Proporción +	58,06%			67,74%		
42		Proporción -	41,94%			32,26%		
43		Población o univ	67	67		67		67
44		Error de precisión max deseado	1%	2%		1%		2%
45				165,063774		0,13575709		
		Rango de error permisibles más o menos		4126,594352		3,393927324		
				3357 más o menos		47839 más o menos		
46				165,0637740		0,135757092		
				93428		979136		
47		Probabilidad (Distrib Normal)	95%	98%		95%		98%
48		Nivel Z deseado	1,6449	2,0537		1,6449		2,0537
49		n con población finita	44,14141732	68,8083714		39,61409247		61,7511025
50			0,665427124	1,05339062		0,597855111		0,94805825
51			66,33546442	65,3208508		66,26035591		65,1342918

Nota. La figura muestra los resultados de las etapas de categorización y estandarización. Los resultados confirman que el volumen disponible (67 meses) supera el mínimo requerido. Fuente: Elaboración propia con base en datos del archivo maestro del estudio y en la secuencia de etapas del proceso de obtención de la data burda del estadio descriptivo descrito en (A. Mora Gutiérrez, 2022).

En línea con la fase I-1-F (Normalización de data) descrita por A. Mora Gutiérrez (2022), se realizan pruebas para comprobar la calidad de la información y determinar si las variables del análisis presentan relación significativa para avanzar hacia los estadios posteriores. Para el caso de los tiempos hasta la falla (UT/TTF) y los tiempos de reparación (DT/TTR), se evalúan los coeficientes de correlación de Spearman y Pearson (Mora Gutiérrez, 2022).

Como se puede observar en la Figura 13, los resultados muestran que la prueba de Spearman arroja un p-value = 0.0045 (<0.05) lo que significa que existe correlación entre las dos variables. El signo negativo del coeficiente confirma que, a medida que los tiempos de UT aumentan, los tiempos de DT tienden a disminuir. Por otro lado, con

Pearson, se obtiene un p-value > que 0.05, lo que sugiere que no existe relación entre las variables.

Por lo anterior, se ratifica que la información cumple con las condiciones del estadio descriptivo para correlación y permite avanzar en el proceso hacia las siguientes validaciones de normalización previstas en la metodología propuesto por A. Mora Gutiérrez (2022).

Figura 13

Correlación entre variables – Spearman & Pearson

Correlation Coefficients		Correlation Coefficients	
Pearson	-0,22419243	Pearson	-0,22419243
Spearman	-0,3454164	Spearman	-0,3454164
Kendall	-0,22942411	Kendall	-0,22942411
Pearson's coeff (t test)		Spearman's coefficient (test)	
Alpha	0,05	Alpha	0,05
Tails	2	Tails	2
corr	-0,22419243	rho	-0,3454164
std err	0,12181811	t-stat	-2,94457095
t	-1,84038669	p-value	0,00450352
p-value	0,07034859		
lower	-0,46755209		
upper	0,01916722		

Nota. La figura presenta los coeficientes de correlación de Pearson y Spearman para las variables UT (TTF) y DT (TTR) calculados mediante el complemento Real Statistics de Excel, en el marco de la fase I-1-F normalización del estadio descriptivo. El análisis indica que la correlación de Spearman muestra significancia estadística (p-value = 0.0045), mientras que la prueba de Pearson no arroja significancia (p-value > 0.05). Fuente: Elaboración a partir de datos del archivo maestro del estudio y del procedimiento de correlación estadística definido en (Mora Gutiérrez, 2022), con apoyo del complemento Real Statistics de Excel.

Prueba de aleatoriedad con ACF y PACF

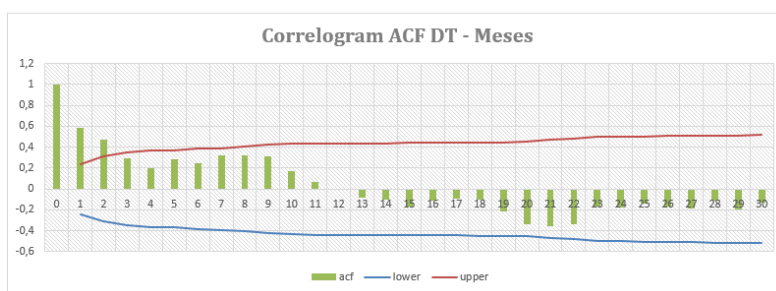
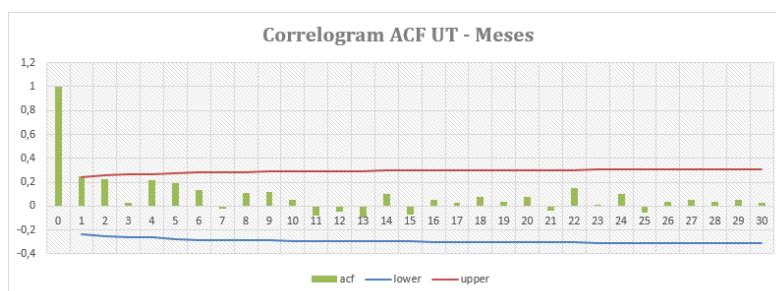
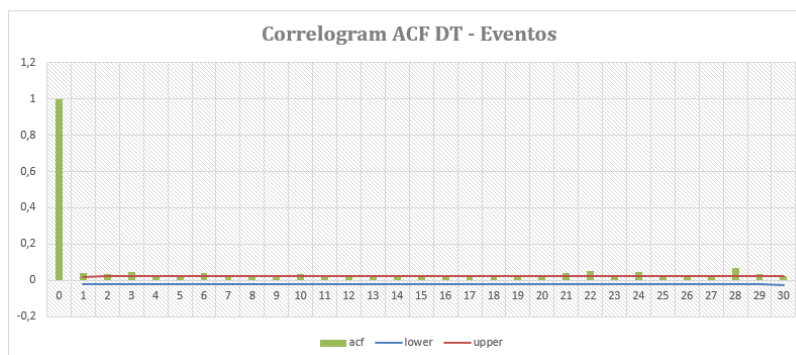
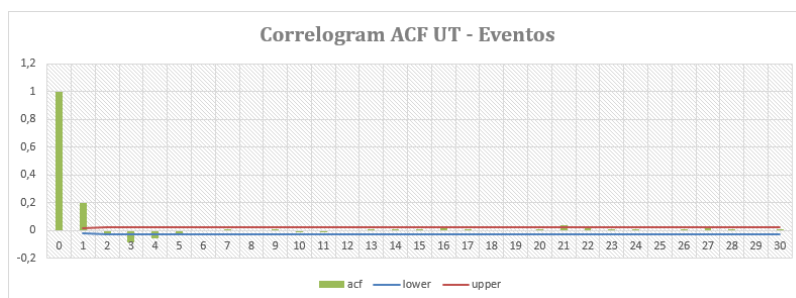
A continuación, se aplica la prueba de autocorrelación mediante ACF y PACF con el fin de determinar si los datos de cada variable presentan dependencia interna o si se comportan de forma aleatoria. Esta exploración permite identificar patrones temporales relevantes y establece si las series poseen estructura suficiente para avanzar hacia esquemas de pronóstico en la zona predictiva del proceso analítico. De acuerdo con el enfoque planteado por A. Mora Gutiérrez, 2022), la revisión de ACF y PACF permite definir si los datos admiten modelos de predicción directa o si requieren metodologías de pronóstico indirecto.

En la Figura 14 se aprecia que, tanto para series mensuales como para series por eventos, las pruebas de ACF no se cumplen para UT ni DT. Se observa que los coeficientes de autocorrelación caen rápidamente hacia valores insignificantes y permanecen dentro de las bandas de confianza. Este resultado establece que las variables no poseen autocorrelación significativa y, en consecuencia, su variación no depende de valores pasados.

En el caso de UT, la ausencia de estructura temporal refuerza la interpretación de que los tiempos entre fallas responden a causas dispersas y no sistemáticas, lo que coincide con el comportamiento típico de fallas aleatorias en transformadores de distribución. Este hallazgo será contrastado posteriormente con el análisis de modos y causas de falla para determinar el grado de dispersión de los eventos.

Figura 14

Correlograma ACF de eventos y meses para UT y DT



Nota. La figura presenta los correlogramas ACF por meses y por eventos, donde se aprecia que las dos variables no cumplen con la prueba ACF. Los coeficientes ACF disminuyen y se mantienen dentro de los límites de confianza. Fuente: Elaboración mediante el complemento Real Statistics de Excel a partir de los datos del archivo maestro del estudio y en referencia al esquema de normalización estadística del estadio descriptivo (Mora Gutiérrez, 2022).

Prueba de variabilidad o nivel

En línea con lo propuesto por A. Mora Gutiérrez (2022), se procede con la revisión de la prueba de variabilidad de los datos para las variables UT y D, lo cual es un paso importante para la manipulación y tratamiento estadísticos en los estadios posteriores. Con base en los resultados de la Figura 15, se verifica que las variables UT y DT presentan altos niveles de variabilidad en la serie de eventos (107% y 185%, respectivamente).

Cuando la variabilidad supera el 50%, la serie presenta un comportamiento aleatorio y no estructurado, por lo tanto, no se puede considerar como una serie pivote o relevante para fines predictivos. Esta condición requiere implementar transformaciones para compensar la dispersión (Mora Gutiérrez, 2022).

Por otra parte, en la serie agrupada por meses, los niveles de variabilidad para MUT son significativamente menores (13 %), mientras que para MDT se obtienen niveles moderados (37 %), ambos por debajo del umbral del 50 %. Esto sugiere un comportamiento más estable y adecuado para análisis descriptivos y para su uso en etapas posteriores del proceso CMD. En síntesis, los resultados confirman la naturaleza altamente aleatoria de los eventos individuales y un comportamiento más regular en las series mensuales, cumpliendo así con los criterios de evaluación de variabilidad requeridos en la etapa F – Normalización de data del estadio descriptivo (Mora Gutiérrez, 2022).

Hasta este punto se tienen unos datos validados por metodología científica al seguir el cada una de las etapas del proceso de obtención de la data burda. El resultado de esta etapa es una base de datos consolidada, depurada y normalizada, lista para su utilización en las etapas de análisis estadístico y modelado de fallas.

Figura 15

Prueba de variabilidad o nivel de UT y DT

Eventos

	A	B	C	D
1	Global	UT		DT
2	1	8767		1
3	2	366		2
4	3	1463		2
5	4	1097		2
6	5	4751		5
7	6	733		1
8	7	368		1
9	8	8404		1
10	9	4387		5
11	10	5118		1
12	11	1100		5
13	12	370		3
14	13	4023		3
15	14	9138		1
16	15	6946		1
17	16	4755		1
18	17	1468		4
7925	7924			2
7926	7925			4
7927	7926			3
7928	7927			1
7929	7928			2
7930	7929			2
7931	7930			3
7932	7931			2
7933	7932			1
7934	7933			1
7935	Tiempo total	29897851		28966
7936	MUT	4144,421	MDT	3,651
7937	Desv	4435,959		6,769
7938	Variabilidad	107%		185%

Meses

	A	B	C	D	E
1			MUT		MDT
2		FECHA	VALOR DÍAS PROMEDIO MES	FECHA	VALOR DÍAS PROMEDIO MES
3	1	Año 1 - Mes 1	3931,53	Año 1 - Mes 1	2,74
4	2	Año 1 - Mes 2	5032,58	Año 1 - Mes 2	1,93
5	3	Año 1 - Mes 3	4662,28	Año 1 - Mes 3	3,57
6	4	Año 1 - Mes 4	4597,91	Año 1 - Mes 4	2,46
7	5	Año 1 - Mes 5	3780,36	Año 1 - Mes 5	3,11
8	6	Año 1 - Mes 6	4052,57	Año 1 - Mes 6	2,63
9	7	Año 1 - Mes 7	4501,28	Año 1 - Mes 7	2,52
10	8	Año 1 - Mes 8	4669,83	Año 1 - Mes 8	2,30
59	57	Año 5 - Mes 9	4500,90	Año 5 - Mes 9	3,10
60	58	Año 5 - Mes 10	4144,02	Año 5 - Mes 10	3,50
61	59	Año 5 - Mes 11	3620,05	Año 5 - Mes 11	3,85
62	60	Año 5 - Mes 12	3438,43	Año 5 - Mes 12	2,71
63	61	Año 6 - Mes 1	2725,44	Año 6 - Mes 1	2,54
64	62	Año 6 - Mes 2	4013,17	Año 6 - Mes 2	3,92
65	63	Año 6 - Mes 3	3635,28	Año 6 - Mes 3	4,90
66	64	Año 6 - Mes 4	3492,80	Año 6 - Mes 4	5,05
67	65	Año 6 - Mes 5	3947,48	Año 6 - Mes 5	4,47
68	66	Año 6 - Mes 6	3647,17	Año 6 - Mes 6	4,39
69	67	Año 6 - Mes 7	3453,93	Año 6 - Mes 7	2,81
70		Tiempo total	276898	Tiempo total	243
71		MUT	4133	MDT	4
72		Desv	518	Desv	1
73		Variabilidad	13%	Variabilidad	37%

Nota. La figura muestra la prueba de variabilidad de UT y DT en eventos y meses, calculada con el complemento Real Statistics de Excel. En eventos, los valores superan el 50%, lo que confirma que las series son aleatorias y no constituyen series pivote. En meses, la variabilidad es menor al 50%, lo que indica mayor estabilidad para análisis descriptivos posteriores. Fuente: Elaboración propia mediante el complemento Real Statistics de Excel a partir de los datos del archivo maestro del estudio y en referencia al esquema de normalización estadística del estadio descriptivo (A. Mora Gutiérrez, 2022).

Como conclusión de este objetivo, se valida estadísticamente el dataset conforme a los requerimientos del estadio descriptivo de la metodología que propone A. Mora Gutiérrez (2022). El tamaño muestral disponible supera los mínimos exigidos; la correlación de Spearman evidencia una relación negativa entre UT y DT, mientras que los resultados de Pearson descarta asociación lineal directa; las pruebas de ACF muestran ausencia de autocorrelación, lo que confirma el comportamiento aleatorio de ambas series; adicionalmente, la variabilidad en la serie de eventos refleja la dispersión natural de los tiempos de las fallas en transformadores

MT/BT, en contraste con la menor variabilidad de la serie de meses que da lugar a una base estable para análisis descriptivos posteriores.

Con todo lo anterior, el conjunto de datos presenta estructura adecuada para análisis de confiabilidad. Se evidencian oportunidades de mejora en la estandarización de causas de falla y en la disponibilidad de información sobre mantenimiento, aspectos que requieren atención en las siguientes etapas del trabajo.

Objetivo dos – Patrones

El propósito de este objetivo es explorar, describir y sintetizar la información contenida en la base de datos con el fin de maximizar su comprensión y establecer las bases para el modelado estadístico y predictivo posterior. El análisis busca identificar errores, relaciones entre variables y patrones de comportamiento que permitan caracterizar los transformadores de distribución de acuerdo con sus condiciones operativas, ambientales y de falla.

El análisis parte de la pregunta central: ¿qué transformadores tienen mayor probabilidad de fallar en el corto plazo?

Para dar respuesta a esa pregunta, esta etapa incluye las siguientes fases:

- Reconocimiento estructural del dataset: identificación del tipo de variables (numéricas, categóricas, temporales y binarias), su completitud y distribución estadística.
- Depuración y consistencia de datos: detección de valores atípicos, datos faltantes y errores de codificación.
- Análisis descriptivo y visualización: uso de técnicas estadísticas y gráficas para identificar correlaciones entre variables, tendencias temporales y comportamientos anómalos.
- Identificación de patrones y grupos: exploración de relaciones multivariadas entre variables de ubicación, mantenimiento y fallas.
- Caracterización del riesgo de falla: análisis preliminar de las variables que podrían explicar diferencias en la probabilidad de falla, como edad, fabricante, ubicación o nivel de tensión

Reconocimiento del dataset, estructura y variables

La etapa 2 del trabajo representa el puente entre la caracterización de los datos y la modelación predictiva. Por lo tanto, en primera instancia se verifica la coherencia del tipo de variable que se ha guardado con el tipo del valor correspondiente.

La Tabla 6 presenta la descripción estadística de las variables principales de la base de transformadores que registran eventos de falla y sin considerar observaciones censuradas. Los resultados muestran que el tiempo promedio hasta la falla es de 3843 días (aproximadamente 10,5 años) con una desviación estándar de 12 años, lo cual indica una alta dispersión en la vida útil de los transformadores, muy relacionado con la aleatoriedad de ocurrencia de fallas en este tipo de activos.

En cuanto a la duración promedio de fallas, se observan valores relativamente bajos, lo que sugiere que la mayoría de los eventos son de corta duración, aunque existen casos atípicos como duraciones de hasta 365 días. En cuanto a los datos de la edad, se aprecia que el 50% de las fallas ocurren antes de los 7 años, lo que puede indicar una población con problemas de envejecimiento prematuro o condiciones operativas adversas.

Respecto a la capacidad nominal (KVA), el valor medio de 17.5kVA y una mediana de 15kVA confirman que existe mayoría de transformadores de baja capacidad, característicos de redes de distribución. Finalmente, el análisis del año de fabricación evidencia que la muestra incluye transformadores de diversas generaciones tecnológicas, lo que permite evaluar en las etapas posteriores la incidencia de la antigüedad tecnológica en los patrones de falla.

Tabla 6

Descripción estadística datos filtrados sin censura

	tff	dt	edad	kva	fecha_de_fabricación
Conteo	3547,00	3547,00	3519,00	3547,00	3519,00
Media	3842,68	4,23	12,94	17,49	2012,06
Desv	4355,59	8,04	11,55	20,86	11,55
Mínimo	0,00	1,00	0,00	5,00	1962,00
25%	761,00	1,00	5,00	10,00	2006,00
50%	1822,00	3,00	7,00	15,00	2018,00
75%	5838,00	5,00	19,00	15,00	2020,00
Máximo	22583,00	365,00	63,00	225,00	2025,00

Nota. La tabla contiene el resumen de la descripción estadística de los transformadores que han fallado. Se incluyen las variables: tiempo hasta la falla (tff), duración de la falla (dt), edad al momento de la falla (edad), potencia nominal (kva) y año de fabricación (fecha_de_fabricación). Fuente: Elaboración propia mediante el procesamiento del dataset con Python (uso de librerías pandas/numpy) a partir de los datos del archivo maestro.

A continuación, se presentan las visualizaciones exploratorias correspondientes a las variables del conjunto de datos. Estas permiten examinar la forma, dispersión y estructura interna de los datos, con el fin de identificar patrones iniciales relevantes para el análisis de confiabilidad y para la selección posterior de modelos estadísticos.

La distribución de la Figura 16 muestra que la mayoría de los transformadores fallan en rangos de 5 a 10 años. Lo que coincide con un patrón heterogéneo de la población y sugiera que las fallas no dependen únicamente del envejecimiento, sino también de condiciones operativas y ambientales.

En la Figura 17 se observa que, en todas las zonas, la categoría “Causas desconocidas o no determinadas” (color azul) representa el grupo más amplio de fallas, especialmente en Tolima Centro y Tolima Norte, donde supera el 40 % de los eventos. Esto indica deficiencias en el registro de la causa raíz o limitaciones diagnósticas en campo. Según la (ISO Norma 14224, 2016), este tipo de registros deben ser priorizados para mejorar la trazabilidad y la consistencia de los análisis de confiabilidad.

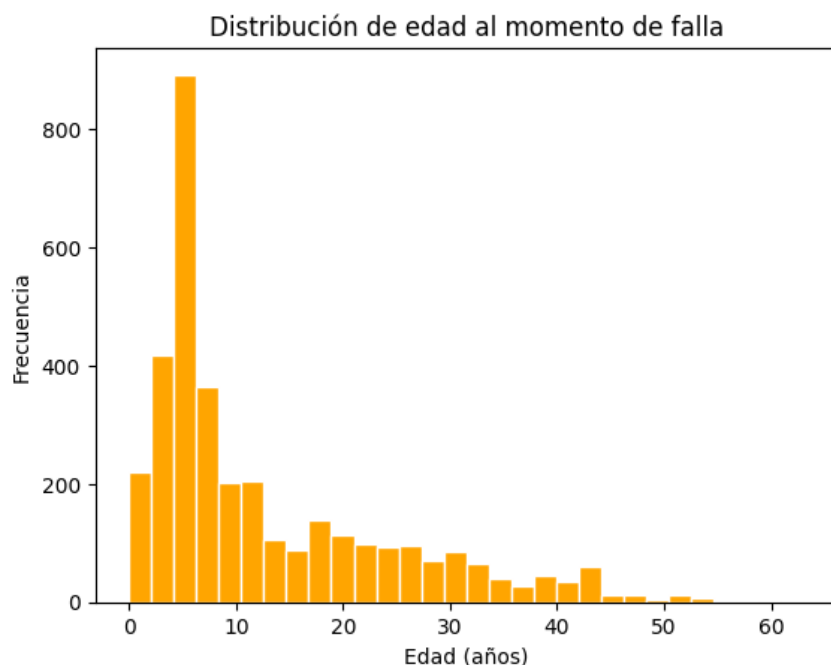
También se observa influencia significativa de factores ambientales en Tolima Oriente y Tolima Sur, los cuales son del 30% - 40% del total. Lo que sugiere una fuerte dependencia del entorno físico, lo que es coherente con zonas geográficas de topografía compleja o rural. Este patrón geográfico confirma la influencia del entorno operativo en el comportamiento de los activos eléctricos.

La distribución observada confirma que las condiciones ambientales y la calidad del registro de fallas son factores críticos para la gestión del mantenimiento de transformadores de

distribución. Por tanto, la variable zona se considera de alta relevancia en la construcción de modelos predictivos de confiabilidad

Figura 16

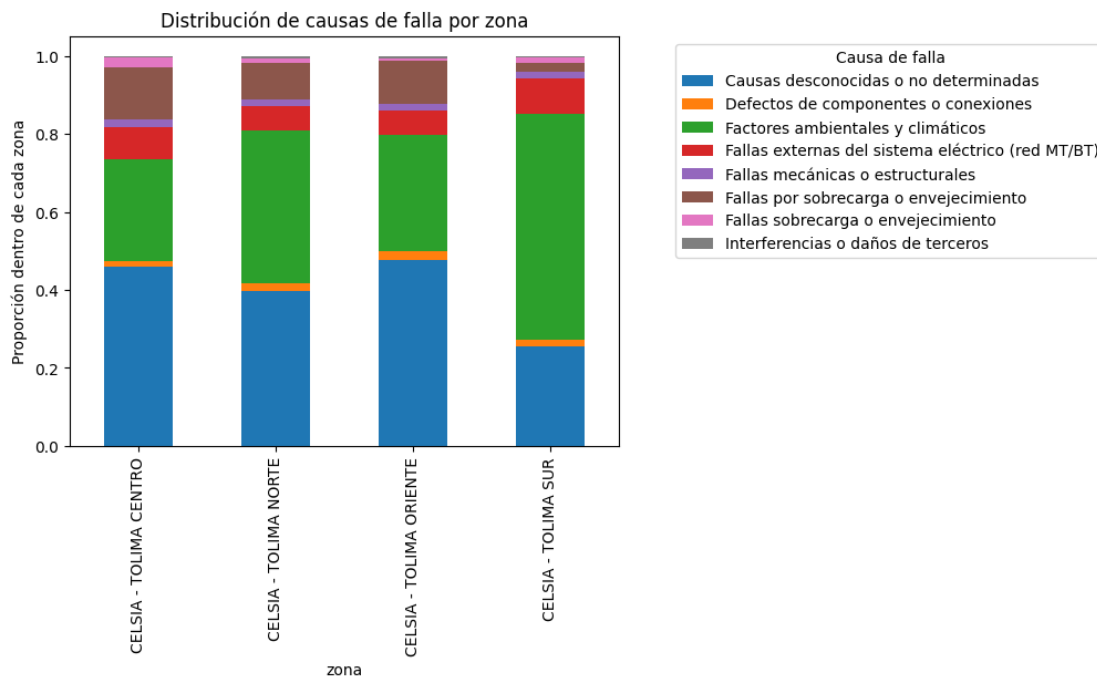
Perfil de envejecimiento



Nota. La gráfica muestra la distribución de la edad de los transformadores MT/BT al momento de la falla asimétrica a la derecha. La mayoría de las fallas ocurren en equipos jóvenes (5-10 años), con cola larga hacia edades altas. Fuente: Elaboración propia a partir de los datos del archivo maestro con librería matplotlib en Python.

Figura 17

Distribución de causa de falla por zona

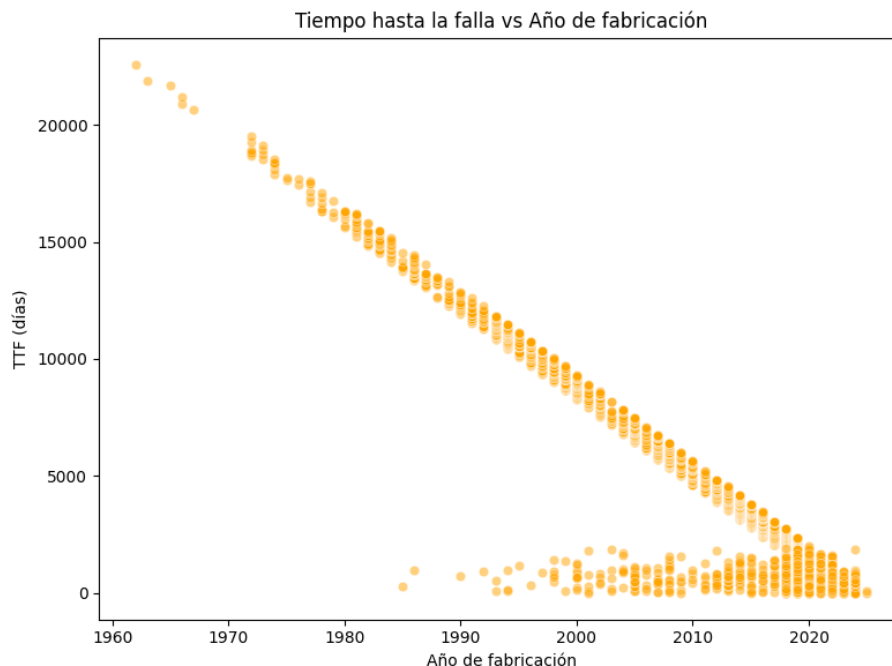


Nota. a figura presenta la proporción relativa de cada categoría de causa de falla dentro de cada zona operativa. El gráfico permite comparar patrones locales de comportamiento y detectar zonas donde ciertas causas tienen mayor peso relativo. Fuente: elaboración propia mediante procesamiento del dataset en Python con librerías pandas y matplotlib.

En la Figura 18 se observa una relación inversa marcada entre el tiempo hasta la falla y el año de fabricación de los transformadores. Los equipos más antiguos, fabricados entre las décadas de 1960 y 1990, presentan tiempos de vida operativa superiores a los 15.000 días ($\approx$ 40 años), mientras que los fabricados después del año 2000 exhiben tiempos hasta la falla inferiores a 5.000 días. Por lo tanto, el año de fabricación se considera una variable altamente correlacionada y complementaria a la edad.

Figura 18

Tiempo hasta la falla vs Año de fabricación



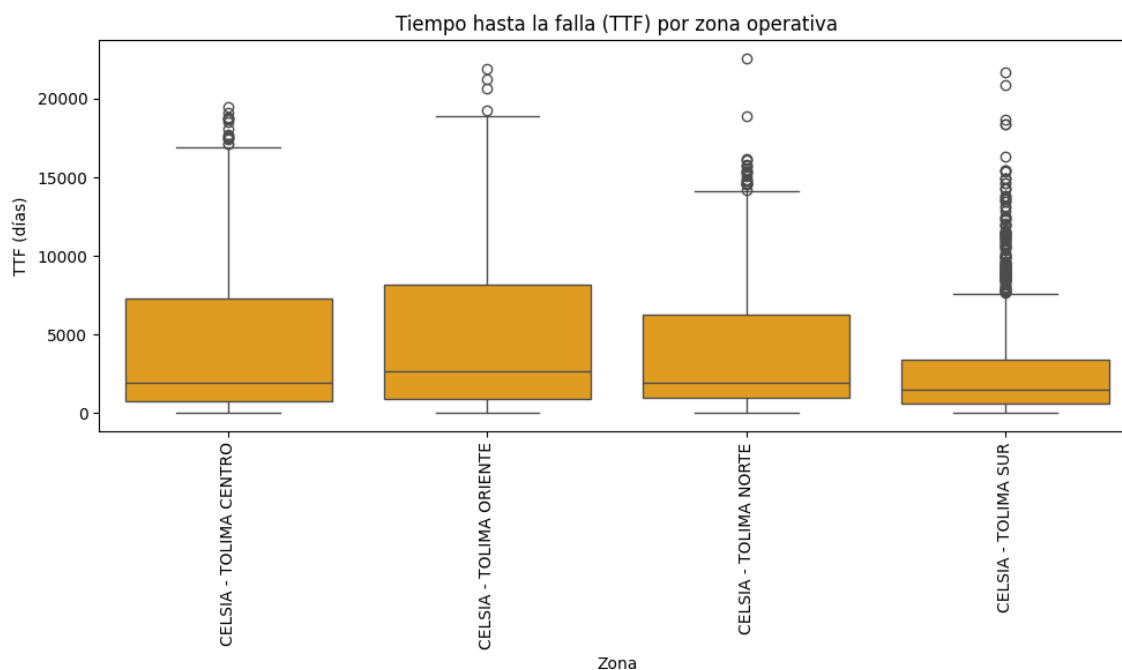
Nota. La figura presenta la relación entre el tiempo hasta la falla (TTF, en días) y el año de fabricación de los transformadores. Cada punto representa una falla de transformador, y muestra la cantidad de días en operación antes de la falla, en función del año en que fue fabricado. Fuente: elaboración propia mediante procesamiento del dataset en Python con librerías pandas y matplotlib.

En la Figura 19 se comparan los tiempos hasta la falla de los transformadores por zona operativa de Celsia Tolima. Se evidencia una diferencia significativa entre regiones, donde las zonas Tolima Centro y Tolima Oriente presentan medianas de TTF superiores, mientras que Tolima Sur tiene los menores valores y mayor concentración de fallas tempranas.

Estos resultados sugieren que el entorno geográfico y las condiciones ambientales influyen de manera determinante en la vida útil de los transformadores. Las zonas con mayor altitud, humedad y dificultad de acceso (como el Sur) presentan una menor confiabilidad, en correspondencia con el patrón de causas de falla dominadas por factores ambientales.

Por lo tanto, la variable zona se considera un indicador indirecto del entorno operativo, con potencial para alimentar los modelos predictivos de TTF.

Figura 19

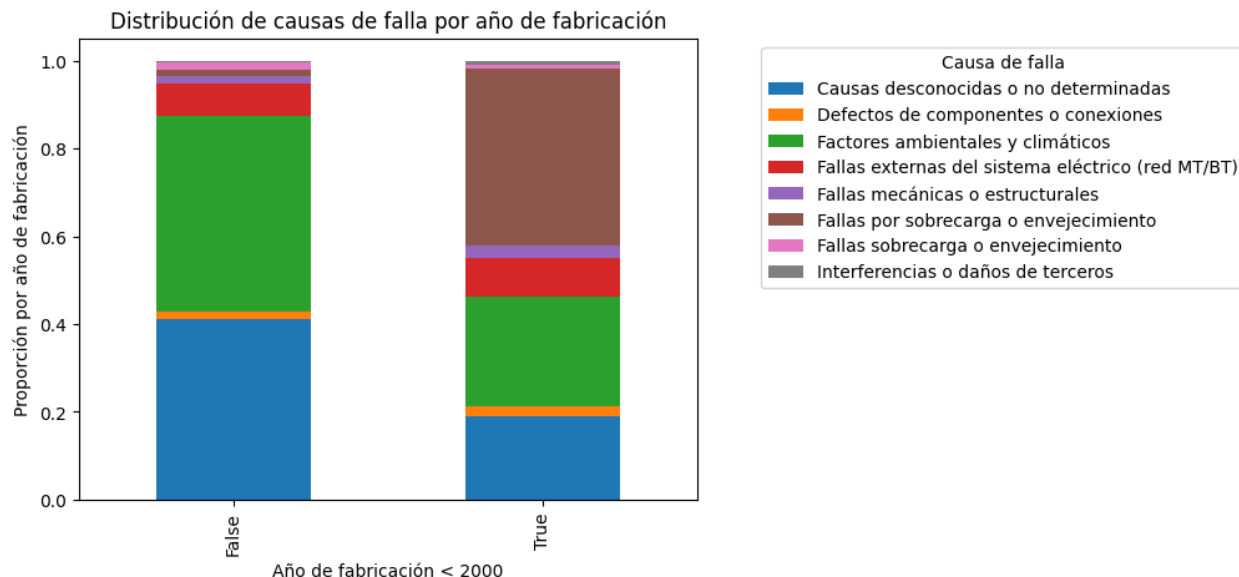
TTF por zona

Nota. La figura presenta la distribución del tiempo hasta la falla (TTF) para los transformadores por zona operativa dentro del área de influencia de Celsia Tolima: Centro, Oriente, Norte y Sur. Cada caja (boxplot) resume la variabilidad del TTF en cada zona. Fuente: elaboración propia mediante procesamiento del dataset en Python con librerías pandas y matplotlib.

Como lo recomienda Hong et al. (2009), lo siguiente es la estratificación de la información para encontrar grupos más homogéneos, inicialmente se usan el fabricante y la edad del transformador como variables de estratificación por la diferencia de tecnología de fabricación y construcción entre los transformadores nuevos y antiguos. Se escoge como fecha de corte el año 2000 para definir a transformadores nuevos y antiguos. Posteriormente se hace un análisis de sensibilidad que investiga los efectos del cambio de esa fecha de corte.

Figura 20

Causas de falla según año de fabricación (<2000 vs >2000)

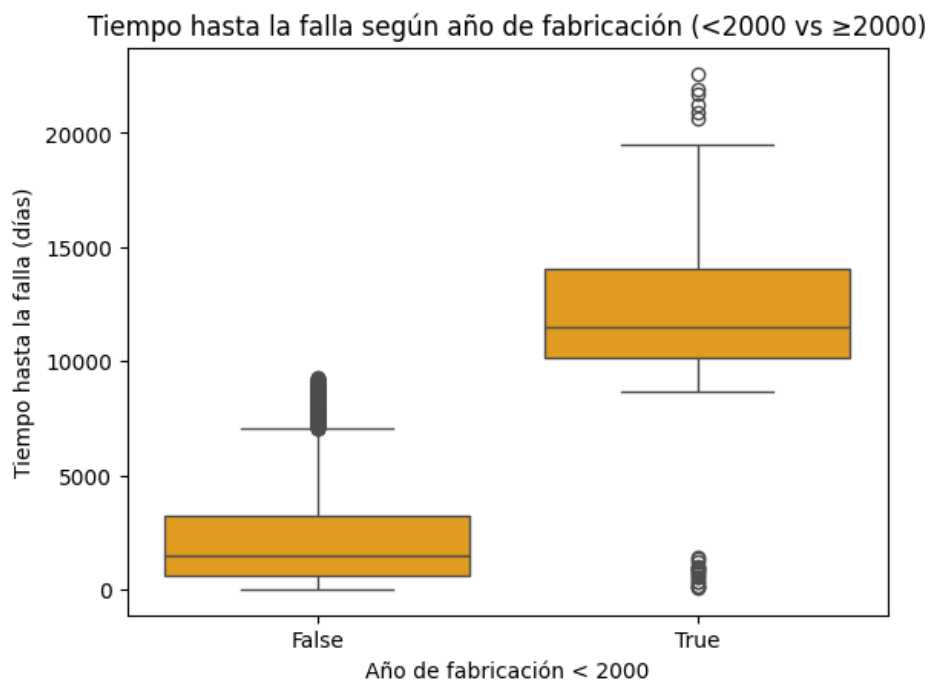


Nota. En la figura se observa la distribución de causas de falla con los datos estratificados según año de fabricación (<2000 vs >2000). Fuente: Fuente: elaboración propia mediante procesamiento del dataset en Python con librerías pandas y matplotlib.

En la *Figura 21* se presenta la comparación del tiempo hasta la falla (TTF) para los dos grupos de transformadores. Los transformadores anteriores al año 2000 presentan mayores tiempos hasta la falla, con una mediana cercana a 11.000 días (≈ 30 años). En contraste, los equipos posteriores al año 2000 muestran una mediana mucho menor, alrededor de 2.000 días ($\approx 5,5$ años). Esto sugiere que los equipos antiguos, aunque tecnológicamente más obsoletos, han mostrado una durabilidad operativa mayor dentro del conjunto analizado.

Figura 21

Tiempo hasta la falla según año de fabricación (<2000 vs >2000)



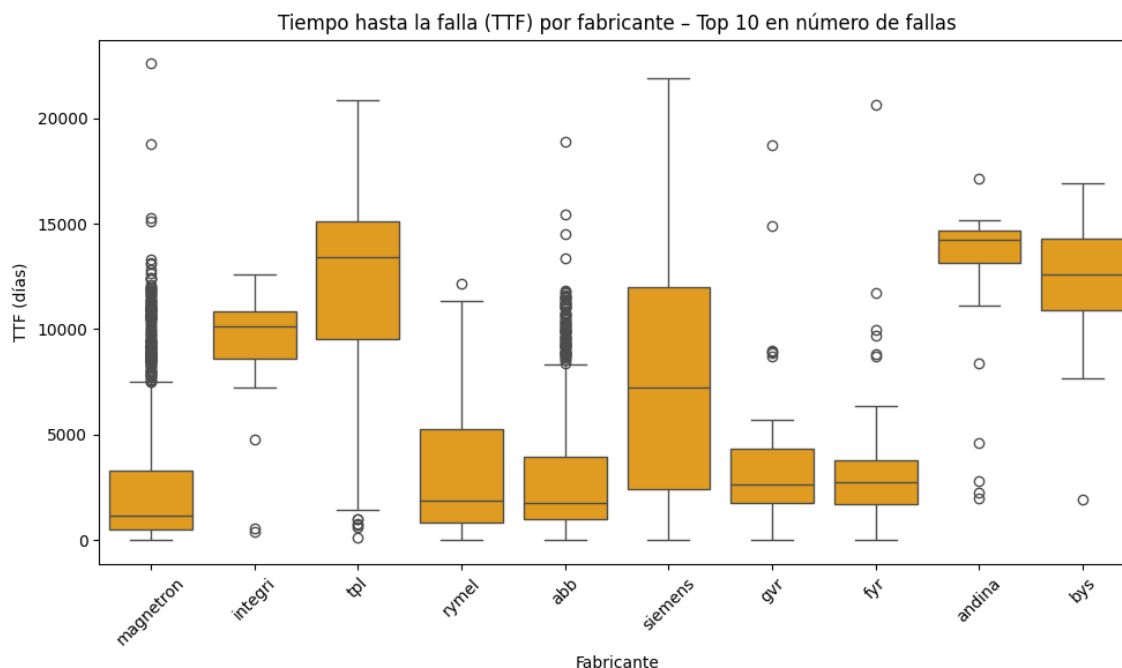
Nota. En la figura se observa cómo la antigüedad tecnológica del transformador influye en el tiempo hasta la falla (TTF). Fuente: Fuente: elaboración propia mediante procesamiento del dataset en Python con librerías pandas y matplotlib.

En la Figura 22 se observa el comportamiento del TTF de los transformadores por fabricante (únicamente los diez con mayor número de fallas registradas). Los resultados evidencian una variabilidad significativa entre fabricantes: los equipos de TPL, Andina, BYS e Integri presentan las mayores medianas de TTF, mientras que Magnetron, ABB, FYR y GYR muestran los valores más bajos. Este comportamiento sugiere diferencias en las prácticas de diseño entre fabricantes.

En términos de gestión de activos, la variable fabricante se considera un indicador técnico relevante que permite identificar riesgos asociados a la procedencia del equipo y contribuye a mejorar la precisión de los modelos predictivos de confiabilidad.

Figura 22

TTF por fabricante - Top 10 en número de fallas

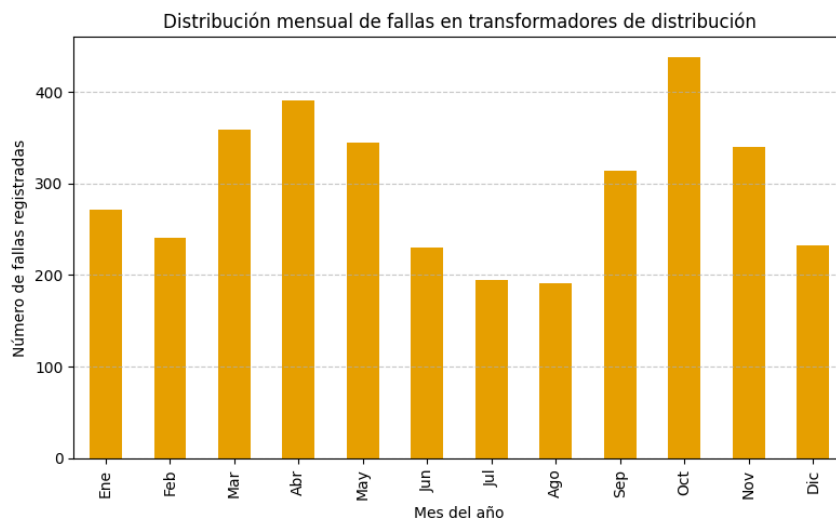


Nota. La figura presenta la comparación del tiempo hasta la falla (TTF) de los transformadores de distribución para los diez fabricantes con mayor número de fallas registradas. Fuente: elaboración propia mediante procesamiento del dataset en Python con librerías pandas y matplotlib.

La distribución mensual de fallas revela un patrón estacional donde se aprecian los picos más altos en abril y octubre; los valores más bajos entre julio y agosto. Estos dos picos coinciden con los períodos de lluvias intensas y alta actividad eléctrica en la región andina colombiana. Lo anterior sugiere una relación directa entre las condiciones ambientales y la probabilidad de falla de los transformadores. La exposición a descargas eléctricas, alta humedad y demás condiciones características de estas épocas pueden acelerar el deterioro del sistema de distribución y aumentar la frecuencia de interrupciones.

Figura 23

Distribución mensual de fallas

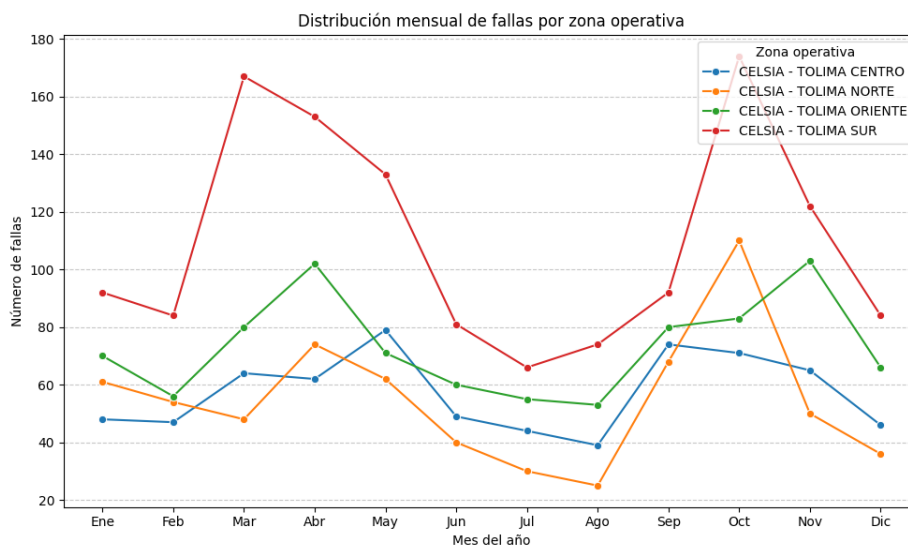


Nota. La figura presenta la frecuencia mensual de fallas registradas en los transformadores de distribución del sistema analizado. Fuente: elaboración propia mediante procesamiento del dataset en Python con librerías pandas y matplotlib.

En la Figura 24 se evidencia una influencia climática y geográfica significativa en la confiabilidad del sistema, otra ya que las zonas con condiciones tropicales más húmedas o con mayor incidencia de tormentas eléctricas tienden a concentraciones más marcadas de fallas en los meses lluviosos.

Figura 24

Distribución mensual de fallas por zona



Nota. La figura muestra la variación mensual del número de fallas en las principales zonas operativas. Fuente: elaboración propia mediante procesamiento del dataset en Python con librerías pandas y matplotlib.

A continuación, se amplía el análisis al conjunto total de transformadores, integrando tanto los registros con falla como los censurados. En total, el dataset contempla 17.109 transformadores, de los cuales 7.933 registran eventos de falla y 9.176 continúan operativos al momento del corte de información. Este enfoque mixto (fallas + censura) es coherente con las recomendaciones de la ISO 14224:2016, que establece la necesidad de conservar ambos tipos de registros para obtener estimaciones no sesgadas de confiabilidad y vida útil.

Tabla 7

Descripción estadística datos filtrados con datos censurados

	tff	dt	edad	kva	fecha_de_fabricación
Conteo	16390,00	7933,00	16319,00	7918,00	16319,00
Media	2405,06	3,65	8,49	17,96	2016,51
Desv	3349,63	6,77	9,67	21,83	9,67
Mínimo	0,00	1,00	0,00	5,00	1948,00
25%	636,00	1,00	3,00	10,00	2018,00
50%	1250,00	2,00	5,00	15,00	2020,00
75%	1950,75	5,00	7,00	15,00	2022,00
Máximo	26902,00	365,00	77,00	300,00	2025,00

Nota. La tabla contiene el resumen de la descripción estadística de todos los transformadores incluidos aquellos que no han fallado. Se incluyen las variables: tiempo hasta la falla (tff), duración de la falla (dt), edad al momento de la falla (edad), potencia nominal (kva) y año de fabricación (fecha_de_fabricación). Fuente: elaboración propia mediante procesamiento del dataset en Python con librería pandas.

De la información de la Tabla 7 se puede concluir que la mayoría de las fallas se normalizan en menos de 5 días, aunque existen eventos extremos de larga duración. La flota tiene una edad tecnológica muy dispersa; predominan transformadores fabricados entre 2018 y

2022, pero coexisten unidades antiguas de más de 70 años. Edad media 8 años, con fuerte asimetría: la mayoría jóvenes (<10 años) y una cola larga de equipos viejos (>40 años). En promedio los transformadores han operado 6.6 años, pero algunos superan los 70 años equivalentes ($\approx 26\,900$ días). La potencia típica es 15 – 25 kVA, típica de transformadores de distribución urbana; pocos outliers llegan a 300 kVA.

El análisis exploratorio determina que las variables con mayor influencia sobre la confiabilidad de los transformadores son la edad, el año de fabricación, el fabricante y la zona operativa. Estas variables describen condiciones de diseño, operación y entorno, y constituyen los principales predictores para los modelos de probabilidad de falla y estimación de vida útil remanente. Además, evidencia un comportamiento estacional en la ocurrencia de fallas, con mayor frecuencia durante meses lluviosos. El estudio confirma la necesidad de fortalecer el registro de datos técnicos y de causas de falla conforme a la (ISO Norma 14224, 2016) y la (UNE-EN Norma 17007, 2018).

Objetivo tres – Modelos

El desarrollo de esta etapa parte de la caracterización estadísticas de los tiempos hasta la falla (TTF) mediante el ajuste de distribuciones teóricas de confiabilidad (UT) y mantenibilidad (DT). Se sigue el modelo método universal de medición CMD desarrollado por Mora Gutiérrez (2009).

El desempeño de cada modelo se verifica mediante indicadores de bondad de ajuste, entre ellos log-likelihood, el criterio de información de Akaike (AIC), el criterio bayesiano de Schwarz (BIC) y el estadístico Anderson–Darling (AD), complementados con la prueba de Kolmogórov–Smirnov (K–S) y Chi – cuadrado para la validación de la hipótesis de distribución.

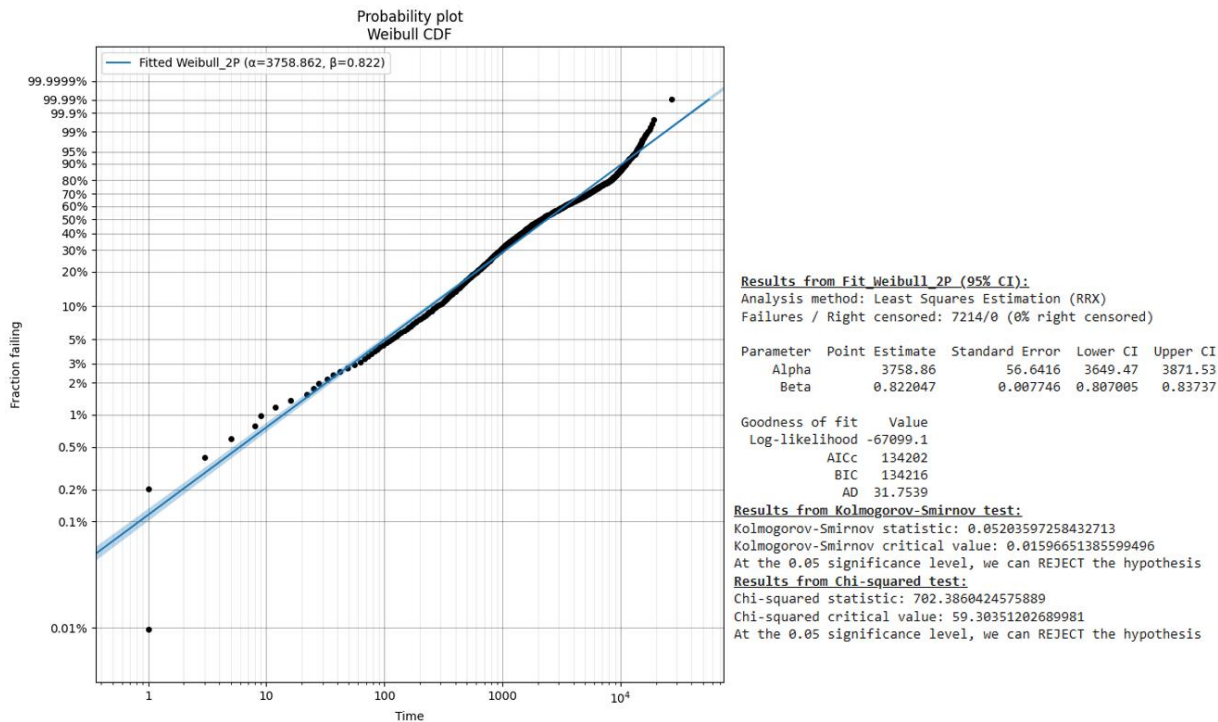
Validación con distribución de Weibull con datos de eventos

De acuerdo con el método universal de medición CMD desarrollado por Mora Gutiérrez (2009), se procede con la alineación para Weibull. Para ello se utiliza la función `Fit_Weibull_2P` de la librería `reliability` en Python.

En la Figura 25 se presentan los resultados del ajuste de las distribuciones Weibull de dos (2P) aplicadas a los tiempos a falla de los transformadores de distribución sin considerar censura. El valor de $\beta < 1$ indica que el conjunto de transformadores se encuentra en una fase de fallas tempranas (mortalidad infantil), donde la tasa de fallas disminuye con el tiempo. Este comportamiento sugiere que las fallas observadas no están asociadas a desgaste progresivo, sino a problemas iniciales de fabricación, instalación o condiciones de puesta en servicio (Mora Gutiérrez, 2009).

Figura 25

Ajuste de distribución de Weibull de 2 parámetros sin censura



Nota. La gráfica corresponde al ajuste de una distribución de Weibull de dos parámetros de los tiempos a falla de los transformadores de distribución, sin considerar datos censurados. Se observa un parámetro de forma $\beta = 0,822$ (<1), característico de una etapa de fallas tempranas, y un parámetro de escala $\alpha = 3759$ días, que representa la vida característica de la población. Fuente: elaboración propia mediante procesamiento del dataset en Python con librería reliability.

Desde el punto de vista de bondad de ajuste, se obtiene no cumple con Anderson–Darling (31,75). Para el caso de Kolmogórov-Smirnov (K-S) y Chi-cuadrado, con la función *KStest* de Python es posible realizar estas pruebas. En la Figura 25 muestra que el resultado de la prueba de bondad de ajuste Kolmogórov-Smirnov (K-S) aplicada a la distribución Weibull de dos parámetros (2P) se rechaza puesto que $D > D_a$, donde $D = 0.052036$ y $D_a = 0.015966$.

Con base en los resultados anteriores, el ajuste de distribución Weibull 2P con los datos de tiempo útil (UT) no cumple con los chequeos de bondad de ajuste que propone el modelo universal de cálculo de CMD propuesto por Mora Gutiérrez (2009).

Validación con distribución Gamma con datos de eventos

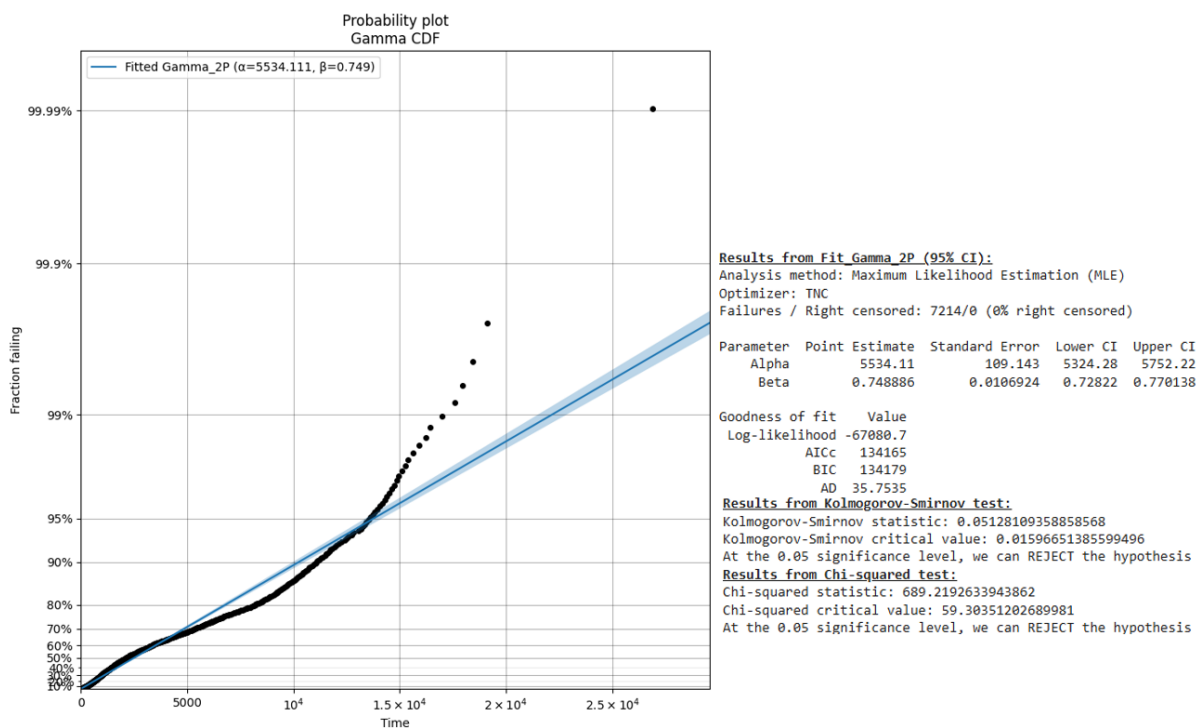
El modelo universal indica que se debe continuar con el proceso de ajuste con otras distribuciones. En este caso que el parámetro de forma β es menor que 1, se debe verificar con Gamma (Mora Gutiérrez, 2009).

La Figura 26 presenta el ajuste de la distribución Gamma de dos (2P). Tanto la distribución Gamma de dos parámetros (2P) como la Weibull de dos parámetros (2P) presentan un comportamiento coherente con la fase de fallas tempranas, con valores de $\beta < 1$ (0,7489 y 0,822, respectivamente).

La Gamma 2P muestra una ligera mejora en la bondad de ajuste, reflejada en menores valores de AIC (134165 frente a 134202), y del estadístico Anderson–Darling (31,75 frente a 35,75). También se observa que no cumple con las pruebas de Kolmogórov–Smirnov (K–S) ni Chi-cuadrado.

Figura 26

Ajuste de distribución de Gamma de 2 parámetros - Sin censura

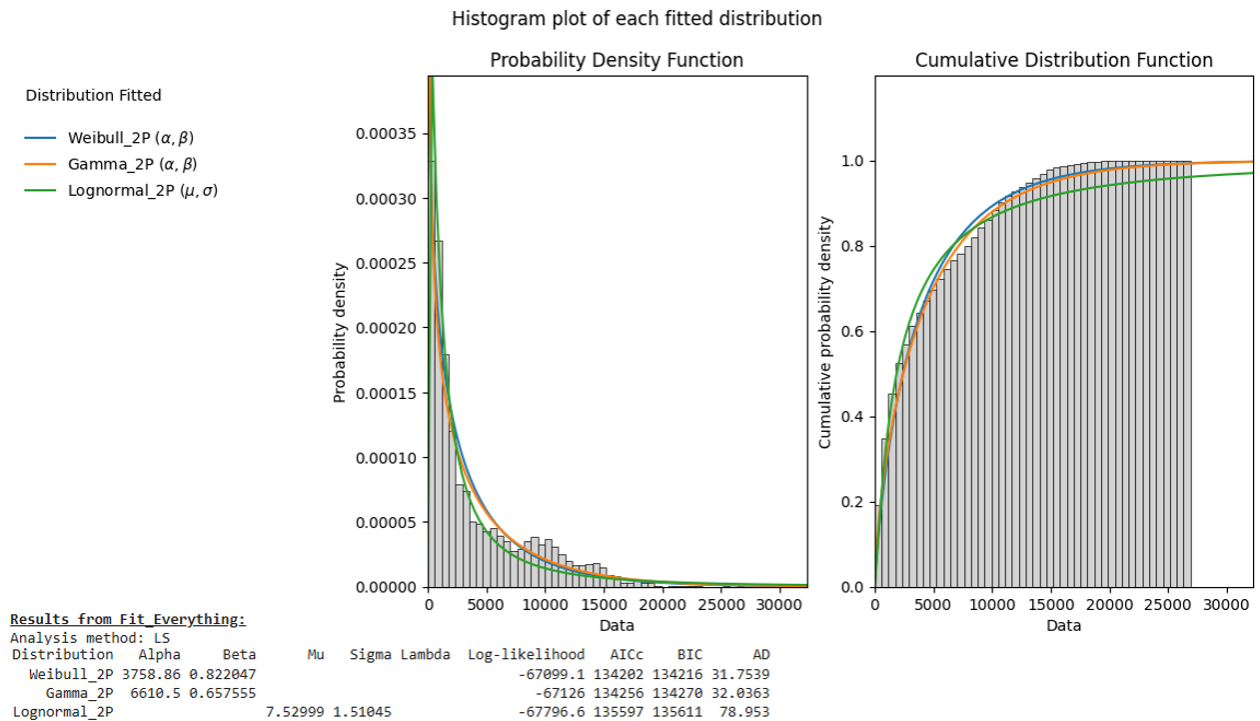


Nota. La figura muestra el ajuste de distribución Gamma de dos parámetros (sin censura) mediante MLE. Se obtiene $\beta = 0,74889 (< 1)$, indicando fallas tempranas, y $\alpha = 5534,11$ días como vida característica. Fuente: elaboración propia mediante procesamiento del dataset en Python con librería reliability.

La Figura 27 presenta la comparación conjunta de las distribuciones Weibull 2P, Gamma 2P y Log-normal 2P ajustadas a los tiempos útiles (UT) sin censura. En la gráfica de densidad de probabilidad (izquierda) se observa que las curvas de Weibull y Gamma, tanto en sus versiones de dos como de tres parámetros, reproducen de manera muy similar la forma empírica de los datos.

Figura 27

Comparación de las distribuciones Weibull, Gamma y Lognormal ajustadas a los TTF sin censura



Nota. Comparación de las distribuciones ajustadas (Weibull 2P, Gamma 2P y Lognormal 2P) aplicadas a los tiempos hasta la falla. Las distribuciones Weibull 2P y Gamma

2P presentan la mejor correspondencia global con los datos empíricos. Fuente: elaboración propia mediante procesamiento del dataset en Python con librería reliability.

Los resultados de los ajustes para datos de eventos de las distribuciones Weibull, Gamma y Lognormal no cumplieron los criterios de bondad de ajuste establecidos en el modelo universal de cálculo de CMD (Mora Gutiérrez, 2009). Esta falta de concordancia sugiere que el conjunto de fallas analizado podría no ser estadísticamente homogéneo, sino una mezcla de subpoblaciones con comportamientos de falla distintos, asociados a factores como el fabricante, la antigüedad del equipo o las condiciones ambientales.

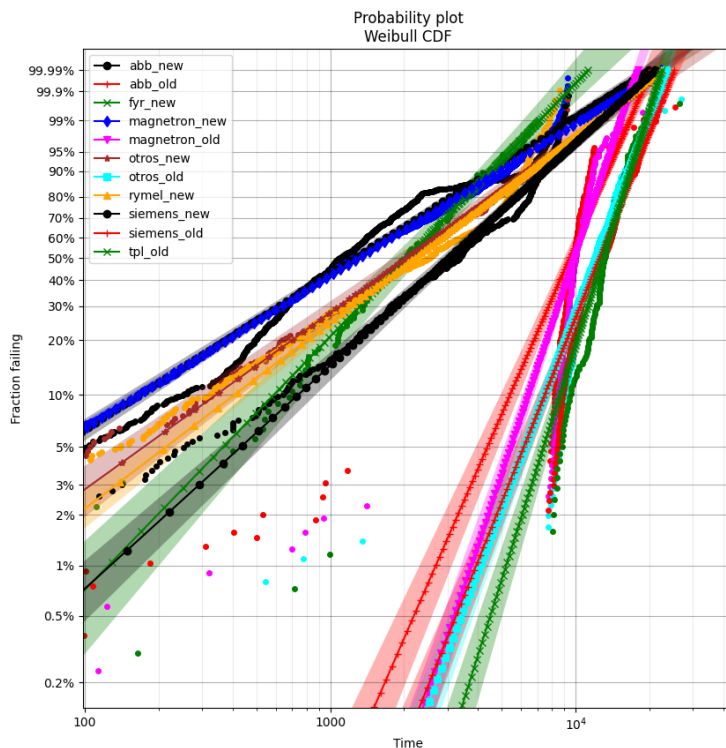
En consecuencia, resulta necesario realizar una revisión estructurada de los datos y proceder a una estratificación del conjunto de observaciones, con el propósito de identificar grupos más uniformes que permitan mejorar la representación estadística de los tiempos hasta la falla y, potencialmente, lograr distribuciones con mayor bondad de ajuste (Hong et al., 2009).

Para el ejercicio de estratificación, se incorporaron variables derivadas, tales como el fabricante agrupado (fab_agr), donde se conservan los siete fabricantes con mayor número de fallas y se reúne el resto bajo la categoría “otros debido a que las muestras eran muy pequeñas en comparación con el resto y el corte de año de fabricación (Pre2000).

En la Figura 28 se presenta el ajuste de distribución Weibull para cada grupo de fabricante, considerando además la variable Pre2000 que diferencia los transformadores fabricados antes y después del año 2000.

Figura 28

Gráfica de probabilidad de Weibull para cada grupo de fabricante



Nota. Ajuste de distribución Weibull por fabricante y periodo de fabricación. Se evidencian diferencias en la pendiente y dispersión entre grupos, donde los equipos antiguos muestran tasas de falla crecientes y los nuevos un comportamiento más estable. Fuente: elaboración propia mediante procesamiento del dataset en Python con librería reliability.

Tabla 8 Parámetros de Weibull para cada grupo de fabricante

Grupo	Alpha	Beta
magnetron_new	1953,92	0,911988
abb_new	1845,37	0,935012
otros_new	2876,44	1,05803
rymel_new	3054,03	1,11724
siemens_new	3988,21	1,33568
fyr_new	2590,92	1,51758
abb_old	10492,4	3,3816
siemens_old	13771,9	3,67113
otros_old	13243,7	3,90753
magnetron_old	10742	4,35625

tpl_old	13918	4,68081
---------	-------	---------

Nota. Parámetros de Weibull para cada grupo de fabricante. Los valores de β reflejan distintas etapas de la curva de la bañera: los grupos new se ubican en fases iniciales, mientras que los old corresponden a la etapa de desgaste. Fuente: elaboración propia.

La comparación de los parámetros α (vida característica) y β (forma) entre los distintos fabricantes también revela diferencias en el desempeño de los equipos, lo cual refuerza la hipótesis de que la población total de transformadores no es estadísticamente homogénea.

Estos resultados explican por qué, en el análisis global previo, ninguna de las distribuciones tradicionales (Weibull, Gamma o Lognormal) logró cumplir los criterios de bondad de ajuste. La mezcla de subpoblaciones con comportamientos de falla distintos genera una dispersión que las distribuciones simples no pueden modelar adecuadamente.

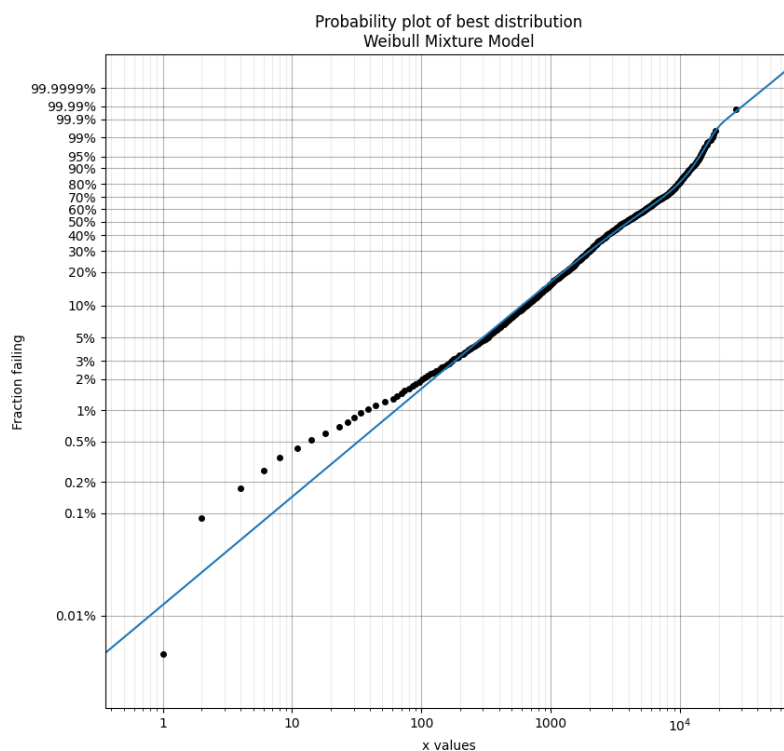
Con base en la función Fit_Everything, el mejor ajuste para el conjunto total de datos de tiempo de falla (TTF) se logra mediante una distribución Weibull Mixture compuesta por dos subpoblaciones. La Figura 29 muestra el gráfico de probabilidad correspondiente al modelo de mezcla, la cual evidencia un ajuste notablemente mejor respecto a las distribuciones simples (Weibull, Gamma y Lognormal).

Este resultado confirma que la población presenta un comportamiento de una combinación de dos grupos de transformadores con patrones de falla distintos: una primera subpoblación con parámetros $\alpha_1 = 3805.6$ y $\beta_1 = 1.05$, asociada a fallas tempranas o aleatorias (fase inicial o de vida útil estable), y una segunda con $\alpha_2 = 11797.5$ y $\beta_2 = 3.46$, característica de la etapa de desgaste o envejecimiento.

La proporción para la primera componente ($p_1 = 0.7545$) indica que aproximadamente el 75 % de los transformadores pertenece a la subpoblación más joven o con fallas aleatorias, mientras que el 25 % restante corresponde a equipos más antiguos.

De acuerdo con la Ecuación 22, se tiene que entonces el MTBF es el valor esperado de la mezcla de las distribuciones.

Figura 29

Ajuste de distribución Weibull MixtureResults from Fit Everything:

Analysis method: MLE

Failures / Right censored: 7208/9176 (56.00586% right censored)

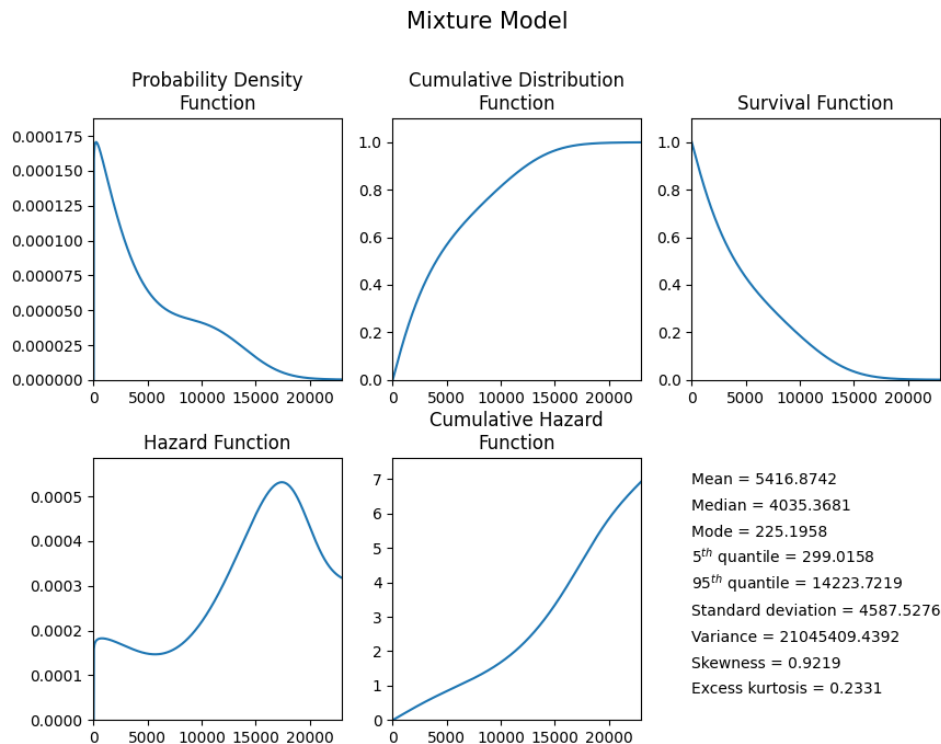
Distribution	Alpha	Beta	Gamma	Alpha 1	Beta 1	Alpha 2	Beta 2	Proportion 1
Weibull_Mixture			3805.57	1.05353	11797.5	3.46146		0.754525

Nota. Ajuste de distribución Weibull Mixture. Se observa un ajuste bimodal que representa adecuadamente la combinación de subpoblaciones con comportamientos de falla distintos, mejorando la correspondencia con los datos reales. Fuente: elaboración propia mediante procesamiento del dataset en Python con librería reliability.

La Figura 30 presenta las funciones características del modelo ajustado. En la parte superior se observan las funciones de densidad de probabilidad (PDF), distribución acumulada (CDF) y supervivencia (SF). La PDF evidencia una distribución bimodal, con un primer pico por fallas tempranas y un segundo que describe el desgaste progresivo de los equipos, lo que confirma la coexistencia de dos subpoblaciones con comportamientos distintos.

Figura 30

Densidad de probabilidad y tasa de fallas de ajuste de Weibull Mixture



Nota. La figura muestra las funciones de confiabilidad del modelo Weibull Mixture. La forma no monótona de la tasa de fallas evidencia la coexistencia de dos etapas de la curva de la bañera: una inicial de fallas aleatorias y otra final de envejecimiento. Fuente: elaboración propia mediante procesamiento del dataset en Python con librería reliability.

La CDF y la SF muestran, respectivamente, el crecimiento acumulado de la probabilidad de falla y la probabilidad de supervivencia a lo largo del tiempo, indicando que alrededor del 95% de los transformadores ha fallado antes de los 14 200 días.

Ecuación 22

MTBF Weibull

$$MTBF = \eta * \Gamma\left(1 + \frac{1}{\beta}\right)$$

Nota. La ecuación define la relación entre el tiempo medio entre fallas y los parámetros de Weibull. Fuente: (Mora Gutiérrez, 2009)

Ecuación 23

MTTR Weibull

$$MTTR = \eta * \Gamma\left(1 + \frac{1}{\beta}\right)$$

Nota. La ecuación define la relación entre el tiempo medio de reparaciones y los parámetros de Weibull. Fuente: (Mora Gutiérrez, 2009)

En ese sentido,

$$MTBF = p \times \eta_1 \times \Gamma\left(1 + \frac{1}{\beta_1}\right) + (1 - p) \times \eta_2 \times \Gamma\left(1 + \frac{1}{\beta_2}\right)$$

De acuerdo con los parámetros del ajuste $p = 0.754525$, $\eta_1 = 3805.57$, $\beta_1 = 1.05353$, $\eta_2 = 11797.5$ y $\beta_2 = 3.46146$. Finalmente, esto resulta en

$$MTBF = 0.754525 \times 3805.57 \times \Gamma\left(1 + \frac{1}{1.05353}\right) + 0.245475 \times 11797.5 \times \Gamma\left(1 + \frac{1}{3.46146}\right)$$

$$MTBF = 5416.9 \approx 14.8 \text{ años}$$

A partir de los parámetros del modelo, el MTBF estimado para la mezcla es de aproximadamente 5417 días ($\approx 14,8$ años), lo cual representa la vida media esperada del conjunto global de transformadores, considerando simultáneamente los dos comportamientos de envejecimiento.

Los resultados del análisis de confiabilidad permiten establecer que la población global de transformadores presenta una estructura no homogénea y responde a comportamientos de falla distintos. Las distribuciones simples (Weibull, Gamma y Lognormal) no cumplen los criterios de bondad de ajuste, lo que confirma la presencia de dos grupos con patrones divergentes de riesgo. El modelo Weibull Mixture define con claridad estas subpoblaciones: una fracción mayoritaria con fallas tempranas o aleatorias y otra con fallas propias de desgaste.

Objetivo cuatro – Análisis:

El desarrollo de modelos predictivos requiere más de aplicar un algoritmo. En las etapas anteriores se detalla la revisión y evaluación inicial del conjunto de datos y se define la estructura del problema. A partir de este punto se requiere la selección adecuada de las variables y una validación formal que garantice confiabilidad antes de utilizar cualquier resultado en decisiones de mantenimiento. La literatura subraya que la construcción de un modelo sin una revisión inicial de consistencia y sin una validación progresiva conduce a resultados pocos confiables, incluso cuando se dispone de algoritmos robustos (Kadre & Reddy Konasani, 2021).

Con este propósito, el objetivo 4 incorpora técnicas de clasificación supervisada para estimar el riesgo de falla de los transformadores MT/BT. Se definen los predictores, se construyen los conjuntos de entrenamiento y prueba y se evalúan varios modelos con métricas como la matriz de confusión, precisión, exhaustividad y F1-score. El propósito central consiste en establecer un modelo funcional y práctico que permita identificar transformadores en condición de riesgo y ofrecer una herramienta útil para la toma de decisiones de mantenimiento y reemplazo.

Metodología para la construcción del modelo predictivo

Una vez completadas las etapas de exploración, validación y preparación del dataset en los objetivos previos, la siguiente fase consiste en la construcción, validación y preparación de los modelos de clasificación que permiten estimar el riesgo de falla de los transformadores. Según las guías metodológicas de proyectos de analítica y aprendizaje automático, el proceso se estructura en tres pasos esenciales: modelado, validación y preparación para despliegue (Kadre & Reddy Konasani, 2021).

En la fase de modelado, se seleccionan técnicas adecuadas para problemas de clasificación binaria cuyo propósito es estimar la probabilidad de falla (fallo = 1) a partir de variables operativas y ambientales. Esta fase incluye la definición de los algoritmos candidatos,

el ajuste de sus parámetros y la comparación sistemática de su desempeño. Se evalúan modelos como regresión logística, árboles de decisión, bosques aleatorios y gradiente boosting, siguiendo las recomendaciones para contrastar múltiples enfoques antes de seleccionar uno definitivo Kabir et al. (2018).

En la fase de validación, el objetivo es asegurar que el modelo generaliza adecuadamente y mantiene consistencia frente a datos no usados en el entrenamiento. Para ello se aplican técnicas como train-test split, junto con métricas propias de clasificación: matriz de confusión, exactitud, precisión, sensibilidad, especificidad y el análisis de la curva ROC-AUC. Esta etapa permite identificar posibles fallos del modelo, sesgos o pérdida de capacidad predictiva, y establece el grado de confiabilidad del algoritmo (Kadre & Reddy Konasani, 2021).

Finalmente, como se describe en la etapa 5, en la fase de preparación para despliegue, se estructura la lógica del modelo para que pueda integrarse como herramienta periódica de gestión de activos. Esto incluye la definición del conjunto de features requeridos, el procedimiento de actualización mensual del dataset y la generación automática de un semáforo de riesgo para apoyar la toma de decisiones sobre intervención o sustitución de transformadores. La metodología asegura que el modelo no solo sea preciso, sino también operativo, interpretable y aplicable dentro del ciclo de planificación de mantenimiento.

Selección de variables predictoras

El conjunto de datos incluye tiempo hasta la falla (TTF), si el transformador fue fabricado o no antes del 2000 y eventos de descarga atmosférica en los circuitos. Previo a la aplicación del modelo se verifica consistencia en los datos y ausencia de valores nulos, así como también el formato para integración de las otras partes del modelo.

Para este análisis se utilizan los siguientes predictores:

- <2000: indicador de antigüedad para diferenciar transformadores con año de fabricación anterior a 2000.
- ut: tiempo de uso expresado en meses.

- 1M_DA, 3M_DA, 6M_DA, 12M_DA: variables que indican presencia o ausencia de descargas atmosféricas en ventanas de 1, 3, 6 y 12 meses previos al punto de análisis, conforme a la estructura definida en el Objetivo 2.

Estas variables cumplen los criterios de disponibilidad mensual, relevancia física y ausencia de fuga de información, y permiten examinar la influencia del entorno atmosférico en el riesgo de falla.

Figura 31

Variables predictoras - Features

```
7 features = ['<2000', 'ut', '1M_DA', '3M_DA', '6M_DA', '12M_DA']
8
9 X = dataset[features]
10 y = dataset['fallo']
```

Nota. La imagen presenta la lista features que contiene las variables predictoras (features) y también se observa la variable objetivo (target) para el modelo supervisado. Donde ut refleja el tiempo de operación, <2000 refleja antigüedad de fabricación, censura identifica los datos que siguen en operación y XM_DA relaciona identifica si hubo descargas atmosféricas previas en el circuito. Fuente: Elaboración propia a partir del dataset y Python en el entorno de Google Colab.

Métricas de evaluación

Para comparar los modelos se utilizan las siguientes métricas recomendadas y ampliamente utilizadas por distintos autores. Cada métrica ofrece una información específica sobre el desempeño de cada modelo frente al conjunto de datos y el objetivo de lograr discriminar unidades con mayor riesgo de falla.

- Matriz de confusión: Se utiliza para calcular el nivel de precisión del modelo. En la Figura 32 se muestra la estructura de esta tabla y las categorías de los aciertos o desaciertos.

Figura 32

Matriz de confusión

Table 5.2 *Confusion Matrix*

	Predicted Classes	
	0	1
Actual Classes	0	Number of times 0 is predicted as 0
	1	Number of times 1 is predicted as 0
	0	Number of times 0 is predicted as 1
	1	Number of times 1 is predicted as 1

Nota. La figura presenta la estructura de una matriz de confusión empleada en problemas de clasificación binaria. La tabla resume la comparación entre las clases reales y las clases predichas por un modelo. Fuente: Tomado del capítulo 5 de (Kadre & Reddy Konasani, 2021).

- Accuracy: Es la relación entre el número de clasificaciones correctas y el total de observaciones evaluadas. La fórmula utiliza los elementos de la matriz de confusión.

Ecuación 24

Exactitud

$$\text{Exactitud} = \frac{\text{Clasificaciones correctas}}{\text{Total de registros}} = \frac{cm[0,0] + cm[1,1]}{cm[0,0] + cm[0,1] + cm[1,0] + cm[1,1]}$$

Nota. La ecuación muestra el cálculo de la exactitud (accuracy) en modelos de clasificación binaria. Fuente: Tomado del capítulo 5 de (Kadre & Reddy Konasani, 2021).

- Sensibilidad y especificidad: corresponden a la proporción de casos de la clase 0 correctamente identificado como 0 y a la proporción de casos de la clase 1 correctamente identificados como 1 respectivamente.

Figura 33

Sensibilidad y especificidad

Table 5.9 *Sensitivity and Specificity*

	Predicted Classes			Classwise Accuracy
	0 – Bad Customer	1 – Good Customer		
Actual Classes	0 – Bad Customer	Model is predicting bad customer as bad	Model is predicting bad customer as good	<i>Sensitivity</i>
	1 – Good Customer	Model is predicting good customer as bad	Model is predicting good customer as good	<i>Specificity</i>

Nota. La figura presenta un ejemplo del uso de los conceptos de sensibilidad y especificidad dentro del contexto de clasificación binaria. Fuente: Tomado del capítulo 5 de (Kadre & Reddy Konasani, 2021).

Ecuación 25

Sensibilidad

$$\text{Sensibilidad} = \frac{\text{Número de veces que 0 se predice como 0}}{\text{Total de ocurrencias reales de 0}} = \frac{TP}{TP + FN}$$

Nota. La ecuación presenta la definición de sensibilidad como métrica de evaluación para modelos de clasificación binaria. La sensibilidad mide la proporción de casos reales de la clase 0 que el modelo identifica correctamente como 0. Fuente: Tomado del capítulo 5 de (Kadre & Reddy Konasani, 2021).

Ecuación 26

Sensibilidad

$$\text{Especificidad} = \frac{\text{Número de veces que 1 se predice como 1}}{\text{Total de ocurrencias reales de 1}} = \frac{TN}{FP + TN}$$

Nota. La ecuación presenta la definición de especificidad como métrica de evaluación para modelos de clasificación binaria. La especificidad mide la proporción de casos reales de la clase 1 que el modelo identifica correctamente como 1. Fuente: Tomado del capítulo 5 de (Kadre & Reddy Konasani, 2021).

- F1 Score y precisión: medidas a partir de las cuales se puede conocer el grado de certeza del modelo. Se considera una extensión de la sensibilidad y la especificidad, dicha relación se observa en la Figura 34.

Figura 34 Recall y precisión

Table 5.18 *Recall and Precision*

Actual Classes	Predicted Classes			Classwise Accuracy
		Positive	Negative	
	Positive	True Positives (TP)	False Negatives (FN)	
	Negative	False Positives (FP)	True Negatives (TN)	
		Precision = $\frac{TP}{TP + FP}$		
				Recall = $\frac{TP}{TP + FN}$

Nota. La tabla presenta la estructura conceptual de las métricas recall y precisión. El recall expresa la proporción de casos positivos que el modelo identifica correctamente, mientras que la precisión indica la proporción de predicciones positivas que corresponden a fallas reales. Fuente: Tomado del capítulo 5 de (Kadre & Reddy Konasani, 2021).

Preparación de los modelos

Se utiliza la función `train_test_split` de la librería `sklearn` para separar el conjunto en entrenamiento y prueba, en proporción 80/20, con estratificación por clase. Esto garantiza que ambas particiones tengan una distribución representativa de transformadores que han fallado y los que no. (Kadre & Reddy Konasani, 2021)

Modelo 1: Regresión Logística

Ya que se trata de una clasificación binaria, se usa inicialmente el modelo Logistic Regression. Este modelo permite estimar la probabilidad de que un transformador ingrese al estado de falla (fallo =1). En la Figura 35 se observa el resultado con las métricas. En los resultados del entrenamiento se identifica correctamente el 83% de los transformadores sin falla y el 73% de los transformadores con falla. El balance entre precisión y recal indica estabilidad del modelo y ausencia de sobreajuste (Kadre & Reddy Konasani, 2021).

En cuanto al rendimiento en prueba, se mantiene la exactitud del 80%, la cual es aceptable en la industria, un recall de 75% para la clase 1 (fallo), lo que indica que detecta el

75% de los fallos reales y precisión de 71%. La regresión logística ofrece un desempeño sólido y estable, con una capacidad predictiva del 75% para las fallas reales y un 80% de exactitud global. El modelo ofrece interpretabilidad y permite estimar la probabilidad de falla de cada transformador, insumo clave para un sistema de alertas tipo semáforo que prioriza los activos con mayor nivel de riesgo.

Este modelo sirve como referencia base para comparar métodos más complejos en las siguientes etapas, donde se incluyen árboles de decisión y Random Forest.

Figura 35

Resultados de Logistic Regression

```
Confusion_matrix Train:
[[4459  883]
 [ 805 2173]]

Classification Report para Train data:
              precision    recall  f1-score   support

    0           0.85         0.83         0.84         5342
    1           0.71         0.73         0.72         2978

 accuracy          0.80         0.80         0.80         8320
 macro avg          0.78         0.78         0.78         8320
weighted avg          0.80         0.80         0.80         8320

Confusion_matrix Test:
[[1106  230]
 [ 186  559]]

Classification Report para Test data:
              precision    recall  f1-score   support

    0           0.86         0.83         0.84         1336
    1           0.71         0.75         0.73          745

 accuracy          0.80         0.80         0.80         2081
 macro avg          0.78         0.79         0.79         2081
weighted avg          0.80         0.80         0.80         2081
```

Nota. La figura muestra las matrices de confusión y los informes de clasificación para los conjuntos de entrenamiento y prueba. El modelo predice la variable fallo (1 = hubo falla, 0 = no hubo). Fuente: elaboración mediante procesamiento del dataset en Python con librería sklearn.

Modelo 2: Árboles de decisión

El árbol de decisión se incluye como segundo modelo debido a su capacidad para generar reglas explícitas que permiten interpretar el comportamiento de los transformadores a partir de las variables operativas. Esta técnica clasifica los activos mediante una estructura jerárquica de decisiones y facilita la identificación de relaciones no lineales entre las variables predictoras y el estado de falla (fallo = 1).

El entrenamiento se realiza con el mismo conjunto de características utilizado en el modelo anterior, conformado por: <2000, ut, censura, 1M_DA, 3M_DA, 6M_DA y 12M_DA, y una partición del 80% para entrenamiento y 20% para prueba.

El árbol clasifica correctamente la totalidad de las observaciones de la clase 0 (*transformadores sin falla*), lo cual refleja una separación muy marcada entre las dos clases cuando la variable *fallo* toma el valor 0. Esto se evidencia en el recall = 1.00 para la clase 0, tanto en entrenamiento como en prueba. Para la clase 1 (*transformadores con falla*), el modelo logra identificar entre el 60% y 62% de los casos.

Los resultados en entrenamiento y prueba son muy similares, lo cual sugiere que el modelo se encuentra estable y no presenta sobreajuste significativo. La exactitud global permanece constante en 0.86, con variaciones mínimas entre f1-scores de ambas particiones.

Figura 36

Resultados árbol de decisión

```
[[5341  1]
 [1204 1774]]
```

Classification Report para Train data:

	precision	recall	f1-score	support
0	0.82	1.00	0.90	5342
1	1.00	0.60	0.75	2978
accuracy			0.86	8320
macro avg	0.91	0.80	0.82	8320
weighted avg	0.88	0.86	0.84	8320

```
[[1336  0]
 [ 284 461]]
```

Classification Report para Train data:

	precision	recall	f1-score	support
0	0.82	1.00	0.90	1336
1	1.00	0.62	0.76	745
accuracy			0.86	2081
macro avg	0.91	0.81	0.83	2081
weighted avg	0.89	0.86	0.85	2081

Nota. La figura muestra los resultados del modelo árboles de decisión. Fuente:

Elaboración propia mediante la librería scikit-learn en Python, a partir del conjunto de datos procesado del archivo maestro.

Modelo 3: Random Forest:

El modelo logra un 89 % de precisión en datos de prueba (test). Par la clase 0 la precisión del 90% y un recall de 94% evidencia que identifica bien los transformadores que permanecen en operación. Este comportamiento es similar al de la clase 1 donde el modelo detecta correctamente alrededor del 81 % de los transformadores que fallan, lo que es aceptable para un primer modelo en aplicaciones de mantenimiento. Adicionalmente, la precisión del 88 % significa que cuando el modelo predice fallo, suele acertar.

El modelo permite clasificar transformadores en riesgo de falla con un nivel de acierto apropiado para tareas de priorización en mantenimiento preventivo y predictivo. El rendimiento es superior al del árbol de decisión individual y más estable que la regresión logística debido a las relaciones no lineales entre la edad útil, censura y descargas atmosféricas.

Figura 37

Resultados de Random Forest

```
[[5261  81]
 [ 416 2562]]

Classification Report para Train data:
              precision    recall  f1-score   support

      0       0.93       0.98       0.95       5342
      1       0.97       0.86       0.91       2978

   accuracy       0.94       0.94       0.94       8320
  macro avg       0.95       0.92       0.93       8320
weighted avg       0.94       0.94       0.94       8320

[[1257  79]
 [ 144 601]]

Classification Report para Test data:
              precision    recall  f1-score   support

      0       0.90       0.94       0.92       1336
      1       0.88       0.81       0.84        745

   accuracy       0.89       0.89       0.89       2081
  macro avg       0.89       0.87       0.88       2081
weighted avg       0.89       0.89       0.89       2081
```

Nota. La figura presenta las matrices de confusión y las métricas de desempeño del modelo Random Forest para los conjuntos de entrenamiento y prueba. Se incluyen los indicadores estándar de clasificación supervisada: exactitud (accuracy), precisión (precision), sensibilidad (recall) y puntaje F1 (f1-score). Fuente: Elaboración propia mediante modelado en Python (scikit-learn), con base en las métricas definidas en la literatura de aprendizaje automático para clasificación.

Tabla 9

Tabla comparativa de modelos

Modelo	Accuracy	Precision (1)	Recall (1)	F1-score (1)	Conclusión técnica
Regresión Logística	0.80	0.71	0.75	0.73	Estable, lineal, capta relaciones simples; sirve como modelo base.
Árbol de Decisión	0.86	1.00	0.62	0.76	Clasifica bien la clase 0, pero pierde capacidad en la clase

					1; riesgo de sobreajuste.
Random Forest	0.89	0.88	0.81	0.84	Mejor balance global; mayor capacidad de generalización; adecuado para predicción operacional.

Nota. La tabla presenta la comparación de tres modelos de clasificación aplicados al dataset de transformadores: Regresión Logística, Árbol de Decisión y Random Forest. Se muestran las métricas de desempeño más utilizadas en problemas de clasificación binaria: accuracy, precision, recall y F1-score para la clase minoritaria (fallo = 1). Fuente: Elaboración propia con Python mediante el uso de las librerías scikit-learn y pandas.

El desarrollo del objetivo permitió evaluar distintos algoritmos de clasificación para estimar el riesgo de falla de transformadores MT/BT con base en las variables operativas definidas. Los resultados muestran que la regresión logística cumple la función de modelo base y permite identificar relaciones lineales entre los predictores y la variable objetivo. El árbol de decisión incrementa el desempeño global, pero reduce la capacidad de identificar correctamente la clase de interés (fallo = 1), lo cual limita su utilidad para decisiones de mantenimiento.

El modelo de *Random Forest* presenta el mejor equilibrio entre exactitud, sensibilidad y precisión, y ofrece la mayor capacidad de generalización sobre datos no vistos. Con este comportamiento, *Random Forest* se valida como el modelo más adecuado para implementar un sistema de alerta temprana que priorice transformadores con mayor probabilidad de falla futura. Este modelo constituye la base técnica para el semáforo de riesgo propuesto en el siguiente objetivo.

Objetivo cinco – Proceso

El Objetivo 5 consolida e integra los resultados de las etapas anteriores del estudio. Los análisis efectuados permitieron depurar y validar la información disponible (Objetivo 1), caracterizar los patrones operativos y ambientales que influyen en el desempeño de los transformadores (Objetivo 2), estimar parámetros de confiabilidad y vida útil a través de modelos probabilísticos (Objetivo 3) y desarrollar modelos de clasificación basados en aprendizaje automático para estimar el riesgo de falla (Objetivo 4).

No obstante, la utilidad práctica de estos resultados depende de su incorporación dentro de un proceso operativo formal, que asegure consistencia en el registro de fallas, facilite el análisis periódico de criticidad y respalde la toma de decisiones de mantenimiento y de gestión de activos. Este objetivo propone un proceso integral que articula:

- Los criterios de confiabilidad estimados;
- Los resultados de los modelos de predicción;
- Las directrices establecidas por normas internacionales como ISO 14224, UNE-EN 17007, UNE-EN 16646 e ISO 55000;
- Los roles, interfaces y responsabilidades de los equipos operativos.

El propósito final consiste en estructurar un sistema estandarizado que permita evaluar mensualmente el riesgo asociado a cada transformador, priorizar intervenciones y apoyar la planificación técnico-económica del ciclo de vida de los activos.

Integración de los resultados de confiabilidad y aprendizaje automático

Los resultados obtenidos en los Objetivos 3 y 4 constituyen la base técnica del proceso propuesto. Desde la perspectiva de confiabilidad, la población global se comporta como una mezcla de subpoblaciones con características diferenciadas de riesgo, lo que se confirma mediante la estimación de modelos Weibull mixtos. La confiabilidad decrece de forma marcada después de la vida media de fallo, y la edad, la zona operativa y la exposición ambiental se reconocen como factores relevantes.

Por su parte, los modelos desarrollados en el Objetivo 4 permiten clasificar transformadores según su probabilidad de falla. Entre los algoritmos evaluados, Random Forest evidencia el mejor desempeño, con una exactitud cercana al 89%, un buen nivel de recuperación para la clase de falla y un equilibrio adecuado entre precisión y capacidad predictiva. Esto lo ubica como el candidato más apropiado para el uso operativo mensual.

El proceso integra ambos enfoques mediante un índice compuesto de riesgo que considera:

- la probabilidad de falla estimada por el modelo predictivo;
- la confiabilidad residual $R(t)$;
- la vida útil remanente;
- factores operativos como zona, condiciones ambientales y fabricación.

Este índice alimenta la clasificación final tipo semáforo propuesta en este objetivo.

Pilar 1: Estandarización y calidad del dato

El proceso parte de la necesidad de contar con información estructurada, completa y confiable. Las normas ISO 14224, UNE-EN 17007 y UNE-EN 16646 establecen lineamientos para el registro sistemático de fallas y para el aseguramiento de la calidad del dato. Bajo estas directrices se propone:

- Estandarizar el registro de fallas bajo ISO 14224: Todo evento debe registrar, como mínimo: modo de falla; mecanismo o causa raíz; componente afectado; condiciones ambientales y operativas, información complementaria, como la presencia de descargas atmosféricas en ventanas de 1, 3, 6 y 12 meses.
- Asegurar calidad de la información: El proceso debe incluir registro de la falla por parte de la brigadas, verificación técnica del supervisor, revisión y confirmación por analistas de mantenimiento y validación de consistencia estadística por el área de evaluación.

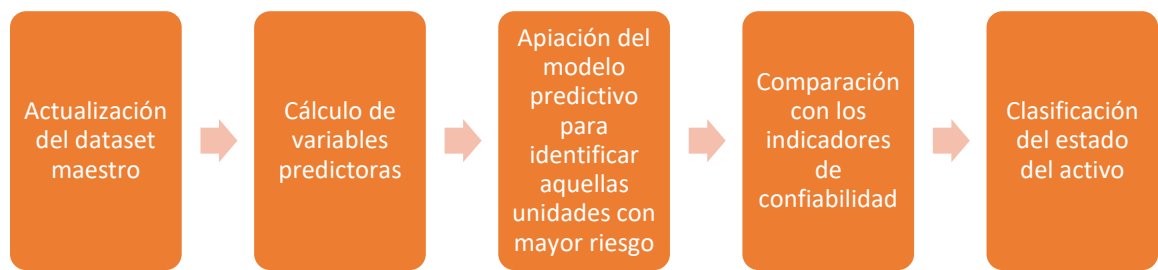
- Efectuar una gestión del cambio adecuada: La implementación requiere que se realicen jornadas de capacitación técnica de brigadas y supervisores, definir un manual de taxonomía y lineamientos de clasificación, auditorías periódicas de calidad del registro y control documental de procedimientos y versiones.

Pilas 2: Integración analítica mensual

La operación del proceso se apoya en un flujo mensual que consolida, transforma y analiza la información:

Figura 38

Flujo mensual de trabajo de datos



Nota. La figura preenta el flujo de actividades que se deben formalizar para garantizar una debida actualización y mantenimiento de los modelos predictivos. Fuente: elaboración propia a partir de los resultados de los objetivos 3 y 4.

La salida del proceso mensual sirve como insumo para priorizar intervenciones y para retroalimentar el modelo cuando se incorporen nuevos patrones de comportamiento en el tiempo.

Pilar 3: Roles y responsabilidades

Es necesario que se definan roles claros dentro de la organización para que este proceso fluya:

Tabla 10

Roles y responsabilidades del proceso

Actor	Responsabilidades
Brigadas de campo	Registro técnico estandarizado de fallas y evidencias.
Supervisores zonales	Validación operativa y verificación de consistencia.
Analistas de mantenimiento	Cálculo mensual del índice de riesgo y revisión técnica.
Gestión de activos	Priorización técnica-económica, aprobación de planes de intervención.
Planeación y programación	Asignación de recursos y secuenciación de trabajos.
Dirección o gerencia	Evaluación estratégica del desempeño y aprobación presupuestal.

Nota. La tabla muestra los roles y responsabilidades que se requieren para garantizar que el proceso se desarrolle y cierre el ciclo de mejora continua. Fuente: elaboración propia a partir de los resultados de los objetivos 3 y 4.

Pilas 4: Criterios de evaluación técnico – económica

Figura 39

Criterios de evaluación técnico - económica

Criterios técnicos	Criterios económicos	Accesibilidad
• Probabilidad de falla estimada. • Tendencia operacional de la zona. • Eventos de descargas atmosféricas. • Época del año.	• ENS esperada. • CAPEX de reemplazo. • OPEX asociado al mantenimiento. • Costos regulatorios por indisponibilidad. • Criterios operativos	• Naturaleza del circuito. • Impacto sobre la continuidad del servicio. • Orden público.

Nota. La figura presenta los criterios utilizados para priorizar la intervención de transformadores dentro del proceso de gestión de activos. Fuente: elaboración propia con base en los resultados del trabajo y de lineamientos de gestión señalados en ISO 55000, ISO 14224 y UNE-EN 16646.

Producto final: herramienta operativa de gestión de transformadores de distribución

El proceso permite consolidar una herramienta integral de gestión de riesgo para transformadores de distribución MT/BT. Esta herramienta articula los elementos analíticos, operativos y normativos definidos en etapas previas, de forma que la organización disponga de un sistema replicable, auditado y alineado con los principios de gestión de activos de ISO 55000, mantenimiento según UNE-EN 16646 y captura estructurada de datos conforme a ISO 14224.

La herramienta se compone de tres elementos principales: un dataset confiable y actualizado con información estructurada según ISO 14224, un modelo predictivo integrado al ciclo mensual de análisis y un tablero de visualización para la toma de decisiones. Estos tres

componentes fortalecen la planificación del ciclo de vida, la gestión del riesgo, la mejora continua y la toma de decisiones con base en datos de calidad.

El proceso integra los resultados analíticos del estudio con los procedimientos operativos y las prácticas recomendadas por normas internacionales de confiabilidad y gestión de activos. La estandarización del registro de fallas mejora la calidad y trazabilidad del dato, mientras que la combinación del análisis de confiabilidad y los modelos de aprendizaje automático permite estimar de forma consistente el riesgo asociado a cada transformador.

El tablero de visualización facilita una socialización oportuna, transparente y reproducible. En conjunto, el proceso fortalece la capacidad institucional para reducir fallas, asignar recursos de manera eficiente y mejorar la continuidad del servicio.

Objetivo seis – Conclusiones

El presente trabajo logra integrar, de manera estructurada y metodológicamente consistente, técnicas de confiabilidad y aprendizaje automático para el análisis y gestión de transformadores de distribución MT/BT. Desde la consolidación y validación científica del dataset (Objetivo 1) hasta la propuesta de un proceso de gestión operacional (Objetivo 5), los resultados permiten concluir que el uso combinado de análisis estadísticos clásicos y modelos predictivos proporciona un marco robusto para la toma de decisiones en mantenimiento y gestión de activos.

En primer lugar, la etapa de preparación y validación de datos permitió construir una base de información sólida, depurada y estadísticamente coherente, cumpliendo las condiciones del estadio descriptivo. Se demostró que el volumen y la estructura del dataset son adecuados para la estimación confiable de parámetros de vida útil y para la aplicación de modelos de predicción. Asimismo, se identificaron brechas importantes en la calidad del dato, particularmente en la estandarización de causas de falla, lo que refuerza la necesidad de adoptar un esquema alineado con la ISO 14224:2016 en las operaciones de registro en campo.

En segundo lugar, el análisis exploratorio permite caracterizar los patrones de comportamiento de la flota: la edad, el año de fabricación, el fabricante, la zona operativa y las condiciones ambientales se confirmaron como variables críticas para explicar diferencias significativas en el riesgo de falla. La presencia de estacionalidad —con picos de fallas en periodos lluviosos— y la influencia geográfica demostraron que el entorno operativo es determinante en la confiabilidad de los activos.

En tercer lugar, la aplicación de modelos estadísticos de confiabilidad deja en evidencia que la población de transformadores no es homogénea. Las distribuciones clásicas (Weibull, Gamma y Lognormal) no cumplieron los criterios de bondad de ajuste establecidos en el método universal CMD, lo cual condujo a la identificación de una mezcla de subpoblaciones con comportamientos de falla distintos. El modelo Weibull Mixture de dos componentes fue el

que mejor representó la realidad operativa, permitiendo estimar un MTBF aproximado de 14,8 años y diferenciando claramente entre transformadores en fase de fallas aleatorias y aquellos en etapa de desgaste. Este hallazgo constituye un insumo clave para decisiones de reemplazo y planificación de inversiones.

En cuarto lugar, la evaluación de modelos de aprendizaje automático complementa de manera efectiva el análisis de confiabilidad. Random Forest ofrece el equilibrio más favorable entre precisión, recall y capacidad de generalización, alcanzando exactitudes cercanas al 90% en los conjuntos de prueba. Los modelos de machine learning permiten priorizar activos con mayor riesgo de falla, siempre y cuando exista un proceso adecuado de actualización y validación periódica del dataset.

Finalmente, la propuesta del proceso integral de gestión de transformadores constituye la síntesis operativa del estudio. Este proceso integra de manera coherente el uso sistemático del modelo predictivo con la estandarización del registro de fallas conforme a ISO 14224, la clasificación estructurada de modos y causas, la incorporación simultánea de criterios técnicos, económicos y de accesibilidad para priorizar activos, la construcción de un tablero de seguimiento en Power BI y la definición de lineamientos operativos para brigadas, supervisores y analistas, para consolidar así un enfoque integral y aplicable para la gestión de transformadores MT/BT.

En conjunto, las conclusiones demuestran que el uso integrado de analítica avanzada y técnicas de confiabilidad representa una oportunidad significativa para mejorar la eficiencia operativa, optimizar recursos y fortalecer la continuidad del servicio en sistemas de distribución eléctrica. El trabajo establece las bases para un proceso de gestión escalable, replicable y alineado con buenas prácticas internacionales, que puede servir como referencia para la implementación de modelos de mantenimiento predictivo y estrategias de gestión de activos bajo un enfoque moderno, técnico y basado en datos.

Resultados más relevantes

Logro 1: Dataset científico y validado

El trabajo consolida un dataset robusto y científicamente validado para la flota MT/BT, resultado de un proceso riguroso de depuración y normalización que incluye limpieza, imputación controlada, estandarización según principios del estadio descriptivo de (Mora Gutiérrez, 2022) y aplicación de pruebas estadísticas como correlación, variabilidad y autocorrelación (ACF/PACF). Este procedimiento asegura coherencia interna, elimina inconsistencias y define adecuadamente la censura, lo que permite contar con una base sólida y apta para análisis de confiabilidad y construcción de modelos predictivos.

Logro 2: Selección del modelo probabilístico que mejor representa la dinámica real de fallas

El estudio determina que la flota se comporta como una mezcla de poblaciones y que el modelo Weibull Mixture de dos componentes ofrece la representación más coherente desde el punto de vista físico y estadístico. Este logro proviene de la evaluación comparada de distribuciones clásicas mediante criterios del método universal CMD, análisis de tasas de falla, validación de formas beta y pruebas de bondad de ajuste, y el uso de librerías de Python lo que permite estimar la confiabilidad residual, la tasa de falla y la vida útil remanente con fundamento técnico.

Logro 3: Implementación y validación de modelos de aprendizaje automático para clasificación de riesgo

El estudio valida la capacidad predictiva de la Regresión Logística, el Árbol de Decisión y el Random Forest, y demuestra que este último ofrece el equilibrio más robusto entre precisión, recall y exactitud para clasificar transformadores en riesgo. Este logro proviene de la definición de un conjunto adecuado de variables predictoras, la división estratificada del dataset, la evaluación mediante métricas formales y la comparación transparente de los

resultados en el conjunto de prueba, lo que habilita el uso de un modelo fiable dentro del proceso mensual de priorización.

Referencias

- Agarwal, U., Jain, N., Kumawat, M., & Maherchandani, J. K. (2020). Weibull distribution based reliability analysis of radial distribution system with aging effect of transformer. In *2020 21st National Power Systems Conference, NPSC 2020*. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/NPSC49263.2020.9331850>
- Artasanchez, Alberto., & Joshi, Prateek. (2020). *Artificial intelligence with Python : your complete guide to building intelligent apps using Python 3.x and TensorFlow 2*. Packt Publishing.
- Asset Management – an anatomy. (2024). *IAM-ASSET MANAGEMENT 2024*.
- Biard, G., & Nour, G. A. (2021). Industry 4.0 Contribution to Asset Management in the Electrical Industry. *Sustainability 2021, Vol. 13, Page 10369, 13(18)*, 10369. <https://doi.org/10.3390/SU131810369>
- Biradar, V., Kakeri, D., & Agasti, A. (2024). Machine Learning based Predictive Maintenance in Distribution Transformers. *2024 8th International Conference on Computing, Communication, Control and Automation, ICCUBEA 2024*. <https://doi.org/10.1109/ICCUBEA61740.2024.10774993>
- Blanco, E. E., Mora, L. A., & Solano, F. (2018). *DISTRIBUCIONES NO TRADICIONALES PARA MEDIR CONFIABILIDAD, MANTENIBILIDAD Y DISPONIBILIDAD (CMD), QUE SE AJUSTAN A VARIAS FASES DE LA CURVA DE DAVIES* (A. García, Ed.; 1st ed.).
- BS ISO Norma 55001. (2024). *BS ISO 55001:2024 Standards Publication Asset management-Asset management system-Requirements*.
- Carolina, A., Vargas, G., & Fandiño Olaya, G. (2025). *Modelo de Diagnóstico Predictivo con Inteligencia Artificial para Fallas en Interruptores de Alta Tensión*.
- Comisión de Regulación de Energía y Gas - CREG 015. (2018). *Resolución CREG 015 - 2018*. https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion_creg_0015_2018.htm

- Cvijic, S., Gupta, N., & Lux, S. (2023). Need for AI in transformer diagnostics and prognostics. *Proceedings - Annual Reliability and Maintainability Symposium, 2023-January*.
<https://doi.org/10.1109/RAMS51473.2023.10088199>
- Dou, D., Wang, Y., Wang, Z., Gao, Z., Na, T., & Wang, T. (2024). Optimal Operation Life Decision Model for Electrical Equipment in the New Power Systems Considering Technical and Economic Factors. *2024 4th International Conference on Intelligent Power and Systems, ICIPS 2024*, 617–627. <https://doi.org/10.1109/ICIPS64173.2024.10900081>
- Freeman, J. M. (1996). Analysing equipment failure rates. *International Journal of Quality and Reliability Management*, 13(4), 39–49. <https://doi.org/10.1108/02656719610114380>
- García, M. A. T. (2025). *Análisis de política de mantenimiento, con modelo predictivo de Riesgos Proporcionales de Weibull PHM y con modelo preventivo a edad constante, caso de estudio en planta Generación Eléctrica*.
- Hong, Y., Meeker, W. Q., & McCalley, J. D. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *Annals of Applied Statistics*, 3(2), 857–879. <https://doi.org/10.1214/00-AOAS231>
- Huang, S. J., & Chiu, P. S. (2024). Enhancement of Maintenance Scheduling of Distribution Transformers Using Carnivorous Plant Algorithm Based Optimization Approach. *IEEE Access*, 12, 51307–51318. <https://doi.org/10.1109/ACCESS.2024.3383856>
- ISO Norma 14224. (2016). *NORMA BS EN-ISO 14224 2016 Industrias de petróleo, petroquímica y gas natural — recolección e intercambio de datos de confiabilidad y mantenimiento de equipos*. British Standards Institution.
- Ji, H. X., Zhang, J. Q., Liu, Z. Y., Liang, G. S., & Zhao, H. S. (2010). Optimal maintenance decision of power transformers. In *Proceedings - International Conference on Electrical and Control Engineering, ICECE 2010*. <https://doi.org/10.1109/ICECE.2010.960>
- Jiang, J., Chen, R., Chen, M., Wang, W., & Zhang, C. (2019). Dynamic Fault Prediction of Power Transformers Based on Hidden Markov Model of Dissolved Gases Analysis. *IEEE*

Transactions on Power Delivery, 34(4), 1393–1400.

<https://doi.org/10.1109/TPWRD.2019.2900543>

Kabir, F., Foggo, B., & Yu, N. (2018). Data driven predictive maintenance of distribution transformers. *China International Conference on Electricity Distribution, CICED*, 312–316.
<https://doi.org/10.1109/CICED.2018.8592417>

Kadre, Shailendra., & Reddy Konasani, Venkata. (2021). *Machine Learning and Deep Learning Using Python and TensorFlow*. McGraw-Hill Education.

Liu, Y., Huang, H.-Z., & Jiang, T. (2023). *Selective Maintenance Modelling and Optimization Basic Methods and Some Recent Advances* (H. Pham, Ed.). Springer.
<https://doi.org/https://doi.org/10.1007/978-3-031-17323-3>

Mora Gutiérrez, A. (2009). *Mantenimiento : planeación, ejecución y control*. Alfaomega Grupo Editor.

Mora Gutiérrez, A. (2022). *Mantenimiento Vanguardia Integral & Analytics, generador de competitividad* (CIMPRO SAS, Ed.; 1st ed.). CIMPRO SAS.

Mora Gutiérrez, A. (2024). *Analytics aplicada a mantenimiento y gestión de activos* (S. Córdoba & C. Vega, Eds.; 1st ed.). Alpha Editorial S.A.

Nachlas, J. A. (1995). *Confiabilidad* (1st ed.).

Parra Márquez, C. A., & Crespo Márquez, A. (2012). *Ingeniería de Mantenimiento y Fiabilidad Aplicada en la Gestión de Activos* (INGEMAN, Ed.; 1st ed.). INGEMAN. www.mgeman.net

Plan Estratégico de Gestión de Activos de CELSIA. (2024). *PEGA Plan Estratégico de Gestión de Activos*.

Qiao, W., & Wang, R. (2024). Optimization Strategy for Condition-Based Maintenance of Aging Equipment Considering Service Life Retirement. *IEEE Information Technology, Networking, Electronic and Automation Control Conference, ITNEC, 2024*, 1236–1245.
<https://doi.org/10.1109/ITNEC60942.2024.10733215>

- Sandoval, J. (2018). *ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS* (Vol. 11).
- Selva, A. M., Yahaya, M. S., Azis, N., Ab Kadir, M. Z. A., Jasni, J., & Yang Ghazali, Y. Z. (2018). Estimation of Transformers Health Index Based on Condition Parameter Factor and Hidden Markov Model. *2018 IEEE 7th International Conference on Power and Energy, PECon 2018*, 288–292. <https://doi.org/10.1109/PECON.2018.8684158>
- Singh, J., Singh, S., & Singh, A. (2019). Distribution transformer failure modes, effects and criticality analysis (FMECA). *Engineering Failure Analysis*, 99, 180–191. <https://doi.org/10.1016/J.ENGFAILANAL.2019.02.014>
- Sotiropoulos, F., Alefragis, P., & Housos, E. (2007). A Hidden Markov Models tool for estimating the deterioration level of a power transformer. *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, 784–787. <https://doi.org/10.1109/EFTA.2007.4416857>
- Sudrajad, G. A., Suwarno, & Subekti, G. A. (2023). Prediction Health Index Using Machine Learning and Its Correlation with Transformer Age on Historical Data. *Proceedings of 2023 4th International Conference on High Voltage Engineering and Power Systems, ICHVEPS 2023*, 188–193. <https://doi.org/10.1109/ICHVEPS58902.2023.10257560>
- Syamsiana, I. N., Febriani, N. A., Sutjipto, R., Sumari, A. D. W., Amalia, R. N., & Nafisah, N. (2024). Remaining Life Prediction of Transformers at Medium Voltage Distribution Substations Based on Load Curve. *ICEECIT 2024 - Proceedings: 2nd International Conference on Electrical Engineering, Computer and Information Technology 2024*, 101–106. <https://doi.org/10.1109/ICEECIT63698.2024.10859583>
- UNE-EN Norma 17007. (2018). *NORMA UNE-EN 17007 Mayo 2018 Proceso de mantenimiento e indicadores asociados*. <https://doi.org/info@une.org>
- Zhang, Y., Zhang, S., & Wang, L. (2019). A Weighted Residual Useful Life Prediction Method for Weibull Distribution Model under Multiple Stress. *2019 Prognostics and System Health*

Management Conference, PHM-Qingdao 2019. <https://doi.org/10.1109/PHM-QINGDAO46334.2019.8942848>

Zhou, D., Li, C., & Wang, Z. (2012). Power transformer lifetime modeling. *Proceedings of IEEE 2012 Prognostics and System Health Management Conference, PHM-2012.* <https://doi.org/10.1109/PHM.2012.6228952>

