

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR**  
**FACULTAD DE INGENIERÍA**



TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE MAGISTER EN  
SISTEMAS DE LA INFORMACIÓN MENCIÓN DATA SCIENCE

MODELO DE CLASIFICACIÓN AUTOMÁTICA PARA IDENTIFICAR FALLAS EN  
UNIDADES DE TRANSFORMACIÓN DEL SISTEMA ELÉCTRICO ECUATORIANO

Autor: Ing. Carlos Molina B.

Director: Ph.D. Vladimir Bonilla V.

**Quito, D.M., enero – 2024**

## RESUMEN

---

Este estudio propone el uso de técnicas de aprendizaje automático para automatizar la metodología DGA en la clasificación de fallas en unidades de transformación del sistema eléctrico ecuatoriano.

Se analizó un conjunto de datos compuesto por 1099 registros de concentraciones de gases, abarcando un período de 2014 a 2023. Luego de aplicar la metodología CRISP-DM para el entendimiento del negocio y de los datos, se empleó varios algoritmos de aprendizaje automático, en donde resalta el modelo entrenado en base a bosques aleatorios, como candidato potencial para clasificar y predecir el estado operativo de las unidades de transformación.

Finalmente, este estudio propone a empresas del sector eléctrico ecuatoriano, tanto públicas como privadas, automatizar los procedimientos de Análisis de Gases Disueltos en Aceite Aislante, que permitan mejorar las estrategias de mantenimiento y operación, reduciendo tiempos de análisis e interpretación, y mejorando la confiabilidad y eficiencia del sistema eléctrico ecuatoriano.

## ABSTRACT

---

This study proposes the use of machine learning techniques to automate the DGA methodology in the classification of faults in transformer units of the Ecuadorian electrical system.

A dataset composed of 1099 gas concentration records was analyzed, covering a period from 2014 to 2023. After applying the CRISP-DM methodology to understand the business and the data, several machine learning algorithms were used, where the model trained based on random forests stands out as a potential candidate to classify and predict the operational status of the transformation units.

Finally, this study proposes to companies of the Ecuadorian electric sector, both public and private, to automate the procedures of Dissolved Gas Analysis in Insulating Oil, which will allow improving the maintenance and operation strategies, reducing analysis and interpretation times, and improving the reliability and efficiency of the Ecuadorian electric system.

## ÍNDICE DEL CONTENIDO

---

CAPÍTULO I. Introducción .....	1
1.1. Antecedentes .....	2
1.2. Planteamiento del Problema .....	4
1.3. Objetivos .....	5
1.3.1. Objetivo General.....	5
1.3.2. Objetivos Específicos.....	5
1.4. Alcance .....	6
CAPÍTULO II. Estado del Arte .....	7
2.1 ¿Cuál es el mecanismo de formación de gases disueltos en fallas de unidades de transformación? .....	7
2.2 Procedimiento tradicional para el análisis de gases disueltos – DGA.....	8
2.2.1 Identificación del Estatus DGA.....	8
2.2.2 Clasificación de tipos de falla utilizando el triángulo de Duval.....	10
2.3 Técnicas de aprendizaje automático para el diagnóstico de unidades de transformación .....	11
2.4 Automatización del proceso de clasificación de la condición de unidades de transformación del sistema eléctrico ecuatoriano .....	12
CAPÍTULO III. Metodología de la Investigación.....	15
3.1. Metodología de la investigación y modelado .....	15
3.2. Entendimiento del Negocio.....	15
3.3. Entendimiento de los datos .....	16
3.4. Preparación de los datos.....	29
3.4.1 Depuración de registros duplicados .....	29

3.4.2	Escalamiento de las variables predictoras .....	30
3.4.3	Tratamiento de valores atípicos (outliers).....	30
3.4.4	Transformación de variables categóricas .....	33
3.4.5	Selección de variables .....	34
CAPÍTULO IV. Resultados.....		36
4.1.	Modelamiento.....	36
4.1.1	Entrenamiento del modelo de árboles de decisión .....	37
4.1.2	Entrenamiento del modelo de bosques aleatorios.....	38
4.1.3	Entrenamiento del modelo de máquinas de vectores de soporte (SVM) .....	39
4.1.4	Entrenamiento del modelo k vecinos más cercanos (KNN).....	39
4.2.	Evaluación del modelo de clasificación automática para identificar fallas en unidades de transformación del sistema eléctrico ecuatoriano .....	41
4.3.	Discusión .....	43
CAPÍTULO V. Conclusiones y Recomendaciones .....		46
5.1.	Conclusiones.....	46
5.2.	Recomendaciones.....	48
CAPÍTULO VI. Referencias .....		51

## ÍNDICE DE FIGURAS

---

<b>Figura 1.</b> Distribución de la infraestructura del SNT en territorio ecuatoriano.....	3
<b>Figura 2.</b> Proporción de la condición de transformadores y autotransformadores en el SNT.....	4
<b>Figura 3.</b> Evolución relativa de gas disuelto en aceite mineral como función de la temperatura y el tipo de falla. ....	8
<b>Figura 4.</b> Proceso tradicional de clasificación DGA.....	9
<b>Figura 5.</b> Clasificador de aprendizaje de gran margen.....	12
<b>Figura 6.</b> Propuesta de automatización del proceso de clasificación DGA.....	13
<b>Figura 7.</b> Condición operativa de las unidades de transformación en el SNT.....	18
<b>Figura 8.</b> Proporción del tipo de condición de las unidades de transformación por zona operativa del SNT. ....	19
<b>Figura 9.</b> Distribución de las variables numéricas del conjunto de datos. ....	21
<b>Figura 10.</b> Análisis de densidades de la variable AGE. ....	23
<b>Figura 11.</b> Inversión anual por etapa funcional en el sector eléctrico ecuatoriano.	24
<b>Figura 12.</b> Mapa de calor de las correlaciones entre variables predictoras. ....	26
<b>Figura 13.</b> Varianza explicada por las variables predictoras en PCA. ....	27
<b>Figura 14.</b> Inspección de outliers luego del tratamiento. ....	32
<b>Figura 15.</b> Entrenamiento del modelo de árboles de decisión.....	38
<b>Figura 16.</b> Entrenamiento del modelo de bosques aleatorios. ....	38
<b>Figura 17.</b> Entrenamiento del modelo de máquinas de vectores de soporte (SVM) .....	39
<b>Figura 18.</b> Determinación del número óptimo de clústeres para el algoritmo KNN. ....	40
<b>Figura 20.</b> Curva ROC y AUC .....	42

## ÍNDICE DE TABLAS

---

<b>Tabla 1.</b> Resumen de unidades de transformación en el SNT. ....	2
<b>Tabla 2.</b> Tipos de falla característicos en unidades de transformación. ....	10
<b>Tabla 3.</b> Información de las variables del conjunto de datos.....	17
<b>Tabla 4.</b> Resumen de los estadísticos principales de las variables numéricas. ....	20
<b>Tabla 5.</b> Porción de valores duplicados .....	29
<b>Tabla 6.</b> Resumen de los estadísticos posterior al escalamiento.....	30
<b>Tabla 7.</b> Información descriptiva de valores atípicos. ....	31
<b>Tabla 8.</b> Resumen de los estadísticos posterior al tratamiento de outliers.....	33
<b>Tabla 9.</b> Codificación de la variable categórica.....	33
<b>Tabla 10.</b> Variables seleccionadas para el modelo.....	34
<b>Tabla 11.</b> Porción de valores para entrenamiento y evaluación.....	36
<b>Tabla 12.</b> Hiperparámetros para el entrenamiento del algoritmo de árboles de decisión.....	37
<b>Tabla 13.</b> Hiperparámetros para el entrenamiento del algoritmo de bosques aleatorios.....	38
<b>Tabla 14.</b> Hiperparámetros para el entrenamiento del algoritmo SVM.....	39
<b>Tabla 15.</b> Hiperparámetros para el entrenamiento del algoritmo KNN.....	40
<b>Tabla 16.</b> Métricas de evaluación de los modelos de entrenamiento.....	41
<b>Tabla 17.</b> Matriz de confusión de la evaluación del modelo arboles de decisión ..	43

## **CAPÍTULO I. Introducción**

Los transformadores, autotransformadores y reactores de derivación son componentes esenciales en la cadena de valor de los sistemas eléctricos de potencia – SEP, jugando un rol esencial en la eficiencia del servicio público de energía eléctrica. Sin embargo, a pesar de su avanzada ingeniería y del riguroso mantenimiento y operación, estas unidades no son inmunes a fallos.

Desde una perspectiva técnica, fallas en estos equipos impactan severamente la confiabilidad y disponibilidad del suministro eléctrico, generando además repercusiones económicas significativas debido a la interrupción del servicio y la consecuente necesidad de activar sistemas de generación térmica como respaldo. Más allá de estas implicaciones de orden técnico y económico, las fallas en estas unidades representan un riesgo para la seguridad de las personas e instalaciones, pudiendo causar daños ambientales, en particular por la contaminación con aceite dieléctrico en las áreas afectadas. En Ecuador, estos incidentes acarrearán rigurosas sanciones por parte de la Agencia de Regulación y Control de Energía y Recursos Renovables (ARCERNNR), el Operador Nacional de Electricidad (CENACE) y el Ministerio del Ambiente, Agua y Transición Ecológica (MAATE).

Ante esta necesidad, las áreas especializadas han intensificado sus esfuerzos para mejorar el entendimiento en cuanto a la mecánica de formación de fallas en unidades de transformación. Sin duda alguna, la experiencia acumulada y los avances tecnológicos han permitido realizar avances significativos en la interpretación de la condición de transformadores, autotransformadores y reactores de derivación, basándose en la relación entre las diversas concentraciones de gases proveniente de su aceite dieléctrico.

Frente a los desafíos presentados por la gestión y mantenimiento de transformadores, autotransformadores y reactores de derivación, el Machine Learning emerge como una solución prometedora. Esta tecnología, representativa del avance en técnicas computacionales que simulan la inteligencia humana, ha demostrado ser una herramienta eficaz para el diagnóstico de maquinaria. A través del uso de algoritmos de aprendizaje supervisado, el aprendizaje automático posibilita la detección temprana de fallas en estas unidades críticas, permitiendo intervenciones preventivas y mejorando significativamente la gestión de riesgos en sistemas eléctricos.

## 1.1. Antecedentes

Dentro del sector eléctrico, el transformador, que incluye autotransformadores y reactores de derivación, es reconocido como el componente más crucial y fundamental de un SEP. Su diseño, basado en los principios de inducción electromagnética, permite una eficiente transformación de los parámetros de voltaje y corriente para minimizar las pérdidas por transmisión de energía. Este mecanismo es vital para cumplir con los requerimientos técnicos de las plantas de producción (generadores) y los principales centros de consumo (consumidores).

En este marco, y según el Plan Maestro de Electricidad 2019 – 2027 2027 (Ministerio de Energía y Recursos Naturales no Renovables del Ecuador, s/f, p. 62); para el 2018 se identificó que la infraestructura eléctrica del país constaba de 153 transformadores en el Sistema Nacional de Transmisión – SNT. Estos transformadores, con una capacidad máxima reportada por el CENACE de 14.902,63 MVA, están distribuidos en 50 subestaciones fijas y 4 subestaciones móviles. Este despliegue subraya la escala y la importancia estratégica de esta infraestructura en el contexto nacional.

**Tabla 1.** Resumen de unidades de transformación en el SNT.

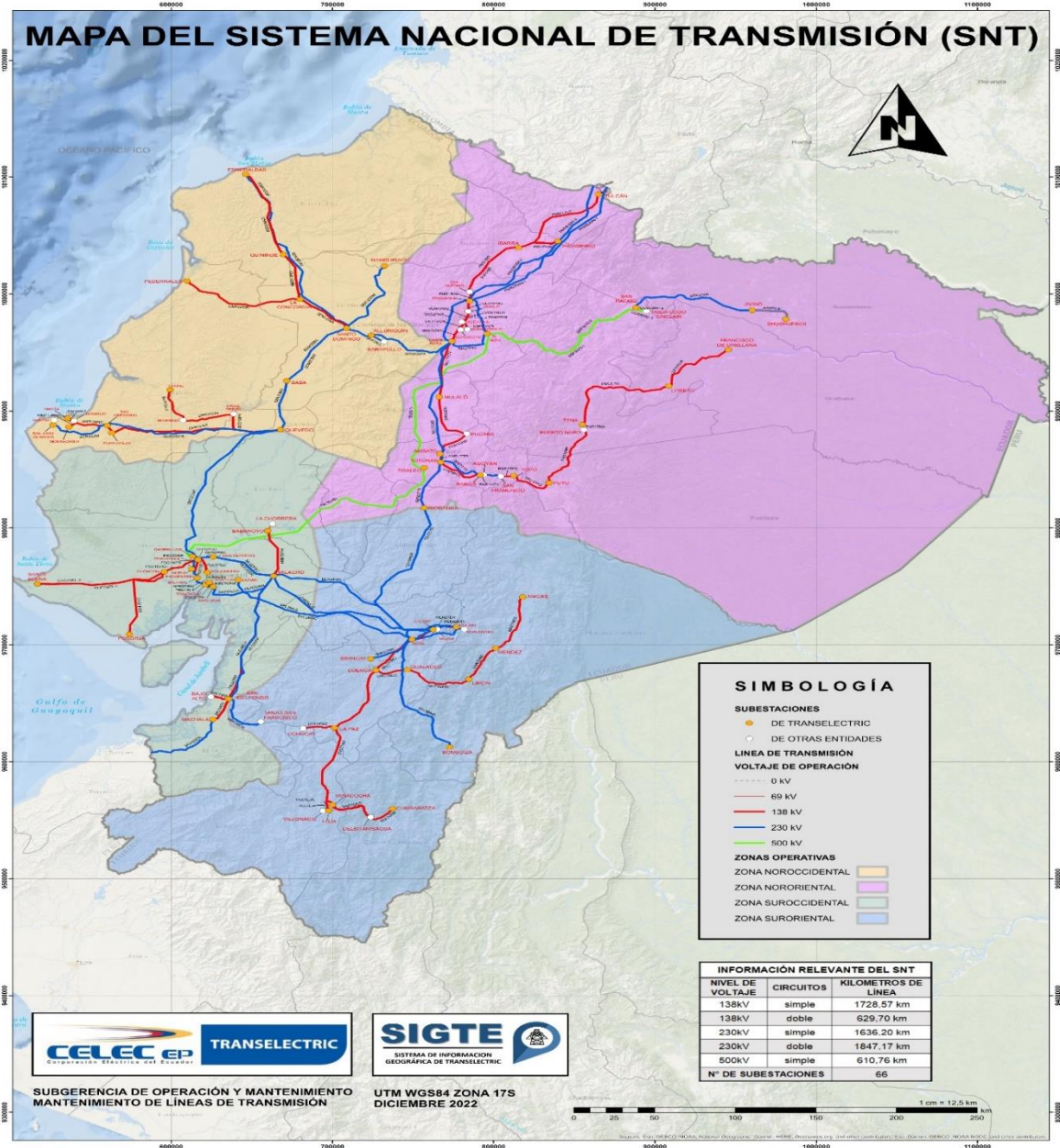
Relación de Transformación [kV]	Unidades trifásicas	Unidades monofásicas
230 / 138	46	7
230 / 69	17	6
138 / 69	38	35
138 / 34.5	-	1
138 / 22	-	1
138 / 13.8	-	2

**Fuente:** Plan Maestro de Electricidad 2019 – 2027.

En términos visuales, la disposición de la infraestructura del SNT, actualizada hasta septiembre de 2023, se ilustra en la Figura 1. Esta visualización proporciona una perspectiva clara y exhaustiva de la distribución de los activos críticos del SNT,

destacando la división estratégica de las zonas operativas, así como, la conectividad de subestaciones y líneas de transmisión en territorio ecuatoriano.

**Figura 1. Distribución de la infraestructura del SNT en territorio ecuatoriano**



**Fuente:** Corporación Eléctrica del Ecuador – Unidad de Negocio Transelectric.

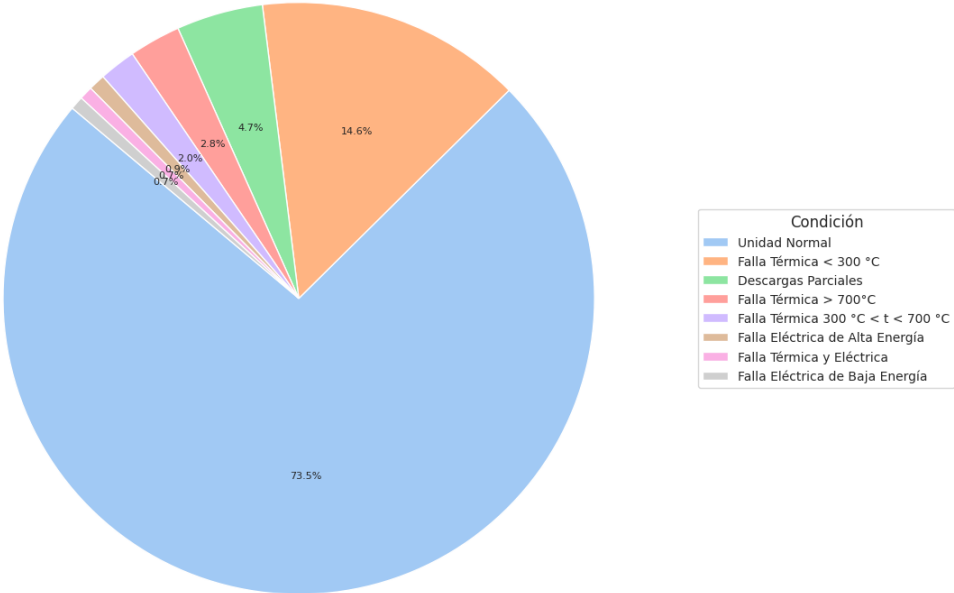
Es fundamental resaltar que, hasta la fecha de publicación del presente trabajo, ni en el Ministerio de Energía y Recursos Naturales No Renovables, ni en las páginas oficiales del Centro Nacional de Control de Energía – CENACE o de la Corporación

Eléctrica del Ecuador – CELEC EP, se han difundido indicadores que reflejen el estado actual del parque de transformadores, autotransformadores y reactores del SNT. Si bien, la ausencia de esta información pública es un aspecto crítico, esta carencia señala una oportunidad para mejorar la transparencia y la gestión de datos en el sector, lo cual es vital para la planificación estratégica y la toma de decisiones informadas en el ámbito de la energía eléctrica.

### 1.2. Planteamiento del Problema

Tras analizar 1099 registros de las concentraciones de gases disueltos en aceite aislante, recabados de 152 unidades de transformación existentes SNT, se ha proporcionado insights fundamentales sobre la condición de los transformadores y autotransformadores del sistema eléctrico ecuatoriano. Estos hallazgos son esenciales para comprender y abordar la problemática actual del sistema, así como orientar estratégicamente los objetivos de este trabajo.

**Figura 2.** Proporción de la condición de transformadores y autotransformadores en el SNT.



**Fuente:** Elaboración propia.

Este análisis revela resultados alentadores: un 73.50% de los registros están en dentro de la categoría Unidad Normal – UN, lo que refleja la efectividad de las estrategias operativas y de mantenimiento implementadas. No obstante, el 26.50% restante muestra diversas condiciones que requieren atención.

Conscientes de la importancia crítica de los transformadores en la infraestructura del SNT y su impacto en la operatividad de sectores productivos, así como en la economía y calidad de vida de los usuarios finales, existe una necesidad imperante de explotar otras herramientas de monitoreo confiables para supervisar la condición de mencionados equipos.

Por otro lado, ante la dinámica de expansión del sistema eléctrico ecuatoriano, así como la incorporación progresiva de monitores de gases en línea, surge el desafío de gestionar y procesar grandes volúmenes de datos. La falta actual de departamentos y expertos especializados en el Análisis de Gases Disueltos en Aceite Dieléctrico – DGA, resalta aún más la urgencia de generar metodologías de análisis avanzadas e independientes, que reduzcan la dependencia de la interpretación humana y aumenten la eficiencia operativa de las áreas interesadas.

Esta primera evaluación subraya la importancia de implementar sistemas automatizados para la clasificación del estado de los transformadores, autotransformadores y reactores de derivación, con el fin de reducir los tiempos de análisis e interpretación de los datos.

Partiendo de estos hallazgos, el objetivo central de este estudio es abordar una cuestión crítica: ¿es posible obtener un modelo de aprendizaje automático, capaz de identificar fallas en unidades de transformación del sistema eléctrico ecuatoriano, mejorando los procesos actuales de mantenimiento? La respuesta a esta pregunta tendrá implicaciones significativas en la mejora de las prácticas de mantenimiento y la confiabilidad del sistema eléctrico ecuatoriano.

### **1.3. Objetivos**

#### **1.3.1. Objetivo General**

Desarrollar un modelo de clasificación automática para identificación de fallas en unidades de transformación del sistema eléctrico ecuatoriano, a partir de registros de análisis de gases disueltos (DGA).

#### **1.3.2. Objetivos Específicos**

- Organizar y depurar los datos históricos del análisis de gases disueltos en aceite DGA, de los transformadores, autotransformadores y reactores existentes en el sistema eléctrico ecuatoriano.

- Comparar la precisión de los resultados de los algoritmos de aprendizaje supervisado de clasificación, permitiendo determinar una solución determinante para los problemas de diagnóstico de fallas en unidades de transformación del sistema eléctrico ecuatoriano.

#### **1.4. Alcance**

Dada la singularidad del contexto operativo y de mantenimiento en Ecuador, que no puede ser equiparado directamente con prácticas de otras regiones —factores como la cultura local y la inversión en programas de mantenimiento proactivo como el mantenimiento basado en la condición (RCM) son distintivos—, esta investigación se centra en encontrar un modelo de inteligencia artificial, capaz de clasificar automáticamente la condición y tipo de fallas en las unidades de transformación, adecuándose a las particularidades del sistema eléctrico ecuatoriano.

En el marco de este proyecto, se dispone de 1099 registros de mediciones cromatográficas de gases disueltos en el aceite de 152 transformadores del SNT, correspondientes al período 2014-2023. Estos registros encapsulan una rica diversidad de condiciones operativas, reflejando las particularidades de la infraestructura eléctrica distribuida a lo largo del diverso territorio ecuatoriano, y por tanto, se ven influenciadas por variables críticas como la cargabilidad de las unidades, la periodicidad y la efectividad de las intervenciones de mantenimiento, proporcionando así un panorama integral que respalda la elección del modelo de inteligencia artificial a ser desarrollado, garantizando su adaptabilidad y relevancia en el contexto específico del sistema eléctrico ecuatoriano.

En resumen, este proyecto establecerá un precedente en la integración de sistemas de inteligencia artificial para la gestión de activos en el sector eléctrico ecuatoriano, que se alinea con las necesidades locales y las prácticas de mantenimiento establecidas. Este análisis no solo resaltaré el progreso actual, sino que también identificaré oportunidades para mejoras futuras, sentando así las bases para un mantenimiento predictivo más eficaz y una gestión proactiva de la infraestructura eléctrica.

## CAPÍTULO II. Estado del Arte

### 2.1 ¿Cuál es el mecanismo de formación de gases disueltos en fallas de unidades de transformación?

Durante la década de 1960, la detección de hidrógeno en muestras de aceite sometidas a estrés eléctrico y térmico condujo al establecimiento de la correlación entre la generación de gases combustibles en el aceite dieléctrico y la presencia de anomalías operativas, ya sean fallas activas o incipientes (Sloat et al., 1965, p. 237). Así emergió el Análisis de Gases Disueltos en Aceite Dieléctrico (DGA), que actualmente se reconoce como una metodología económica y confiable para la detección indirecta de estados anormales en transformadores.

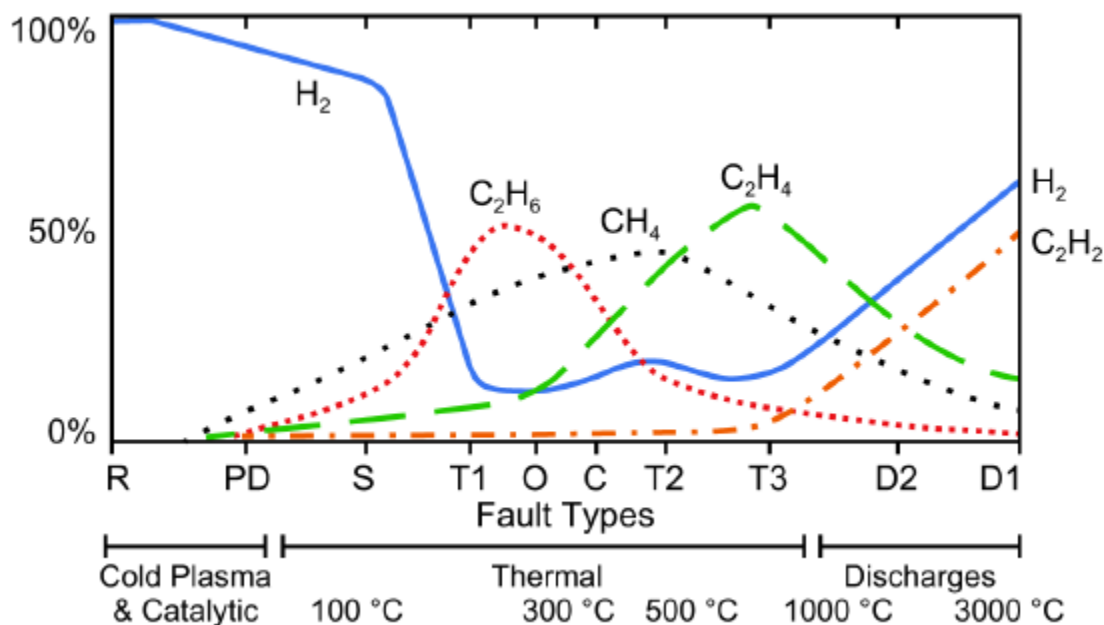
Para elucidar la génesis de los gases disueltos en el aceite del transformador, es esencial comprender que el aceite es una compleja mezcla de hidrocarburos. Bajo esta premisa, se entiende que incidentes térmicos o eléctricos al interior del transformador inducen la transferencia de energía al aceite, provocando la descomposición de enlaces carbono-hidrógeno (C-H) y carbono-carbono (C-C), lo que resulta en la formación de radicales libres y su posterior recombinación, dando lugar a gases o a nuevas moléculas condensables. La extensión de estas reacciones depende de la naturaleza del hidrocarburo presente (nafténico, parafínico, isoparafínico o aromático), la magnitud de la energía liberada y la duración de la exposición del aceite a las condiciones de estrés (Martínez, 2021, p.p. 117-118).

Adicionalmente, la degradación térmica del aislamiento de celulosa impregnada con aceite mineral conduce a la formación de óxidos de carbono (CO y CO<sub>2</sub>), así como a la emisión de cantidades menores de hidrógeno (H<sub>2</sub>) y metano (CH<sub>4</sub>). La tasa de gasificación de estos compuestos incrementa exponencialmente con la temperatura y es directamente proporcional al volumen de material expuesto (IEC, 2022, p.16).

Bajo estas premisas, se infiere que inicialmente (en  $t = 0$ ), el aceite dieléctrico no contiene gases combustibles y tiene concentraciones bajas de óxidos de carbono, asumiendo que el transformador se ha llenado con aceite mediante procesos de vacío. No obstante, en un instante  $t > 0$ , las fallas eléctricas inducen la liberación de energía al aceite, resultando en la producción de gases cuya composición específica es indicativa del tipo de falla (eléctrica, térmica o combinada), y se ve influenciada por la duración de la exposición al evento de falla, parámetros que están intrínsecamente vinculados con el

nivel de energía y la temperatura alcanzada durante el evento de gasificación, como se ilustra a continuación.

**Figura 3.** Evolución relativa de gas disuelto en aceite mineral como función de la temperatura y el tipo de falla.



**Fuente:** IEEE Guía para la interpretación de gases en transformadores embebidos en aceite mineral aislante.

Dentro del ámbito de la ingeniería del mantenimiento de transformadores, el Análisis de Gases Disueltos en Aceite (DGA por sus siglas en inglés), representa un método diagnóstico que permite inferir la naturaleza de las fallas en unidades de transformación mediante la cuantificación y análisis cualitativo de los gases inmersos en el aceite dieléctrico. Esta técnica se basa en la premisa de que los gases generados y sus concentraciones específicas son indicativos de ciertos tipos de anomalías operativas dentro del equipo.

## 2.2 Procedimiento tradicional para el análisis de gases disueltos – DGA

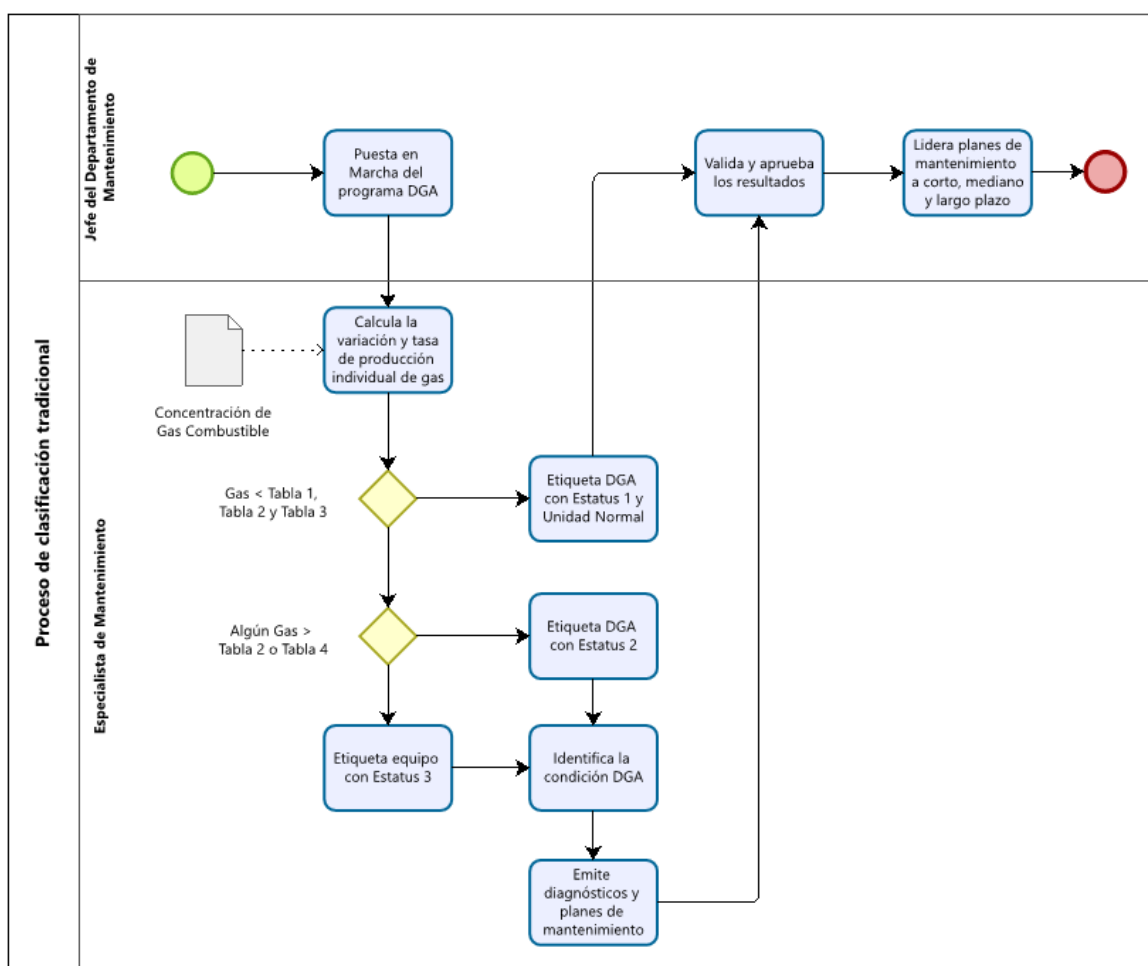
### 2.2.1 Identificación del Estatus DGA

Una vez establecido el mecanismo por el cual se forman los gases disueltos en el aceite dieléctrico, es imperativo avanzar hacia la metodología clásica para la identificación de tipos de falla en transformadores. A tal efecto, se adopta el procedimiento estipulado

por el Instituto de Ingenieros Eléctricos y Electrónicos en la guía estándar IEEE Std. C57.104.

Este método inicia con la determinación de los estados DGA, los cuales se derivan del análisis de la concentración individual, las tendencias temporales y la cinética de producción de los gases combustibles. Este análisis ha permitido establecer a lo largo de estos años, relaciones confiables entre la presencia y dinámica de los gases con las fallas más características en unidades de transformación.

**Figura 4.** Proceso tradicional de clasificación DGA.



**Fuente:** Elaboración Propia.

## 2.2.2 Clasificación de tipos de falla utilizando el triángulo de Duval

El Método de Duval, concebido por Michael Duval de Hydro Koubek en la década de 1960 (Duval, 1989, p. 23), continúa siendo hasta estos días una herramienta valiosa en la identificación y clasificación de los tipos de falla más prevalentes en transformadores. A pesar de la existencia de metodologías más recientes también desarrolladas por Duval, este enfoque original mantiene su relevancia debido a su eficacia, precisión y alto rendimiento, cualidades que han sido ampliamente reconocidas por la comunidad científica, como se evidencia en la guía estándar IEEE Std C57.104 (IEEE Std C57.104, 2019, p.p. 41-42).

Como en cualquier metodología de DGA para la clasificación de tipos de falla en unidades de transformación, es importante señalar que el método del triángulo de Duval no es aplicable en situaciones donde la condición DGA presenta un Estatus DGA 1, lo que indica una condición operativa normal (UN). En tales casos, no es necesario realizar un análisis de fallas, ya que el equipo se encuentra funcionando dentro de los parámetros esperados.

**Tabla 2.** *Tipos de falla característicos en unidades de transformación.*

Identificador del tipo de falla	Descripción ampliada del tipo de falla
PD	Descargas Parciales
T1	Falla Térmica $T < 300^{\circ}\text{C}$
T2	Falla Térmica $300^{\circ}\text{C} < T < 700^{\circ}\text{C}$
T3	Falla Térmica $T > 700^{\circ}\text{C}$
DT	Falla Térmica y Eléctrica
D1	Falla Eléctrica de Bajo Nivel
D2	Falla Eléctrica de Alto Nivel

**Fuente:** IEEE Std. C57.104.

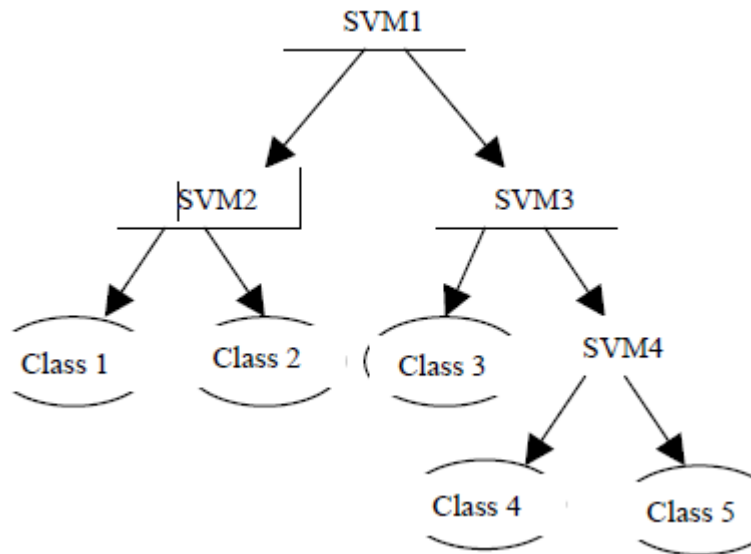
### **2.3 Técnicas de aprendizaje automático para el diagnóstico de unidades de transformación**

En el ámbito del diagnóstico de fallos en transformadores de potencia mediante el Análisis de Gases Disueltos (DGA) y el aprendizaje automático, se han desarrollado varios modelos significativos. En 2017, Roberto Fiallos presentó un modelo basado en la Máquina de Vectores de Soporte con Mínimos Cuadrados (LSSVM), implementado en Matlab (Fiallos, 2017, p.p. 55-60). Este modelo, orientado a predecir la concentración de gases disueltos en transformadores, se fundamenta en datos de cromatografía de tres transformadores entre octubre de 2012 y mayo de 2016. Fiallos aplicó una combinación no lineal y un enfoque de optimización paramétrica, lo que resultó en un modelo adaptable a diversas condiciones operativas y tipos de transformadores, aunque con ciertas limitaciones en cuanto a la generalidad y a la falta clasificación de las condiciones DGA.

Por otro lado, Tao Liu y Zhijie Wang (Liu et al., 2009, p.p. 138-140), propusieron un modelo utilizando Máquinas de Vectores de Soporte (SVM) para el diagnóstico de fallos en transformadores, destacando su superioridad sobre métodos tradicionales como el método de comparación IEC y la detección de red neuronal BP. Para esta investigación, los autores utilizaron 280 muestras de aceite dieléctrico, centrándose en fallos de alta energía y demostrando la capacidad del SVM para analizar datos dinámicos y no lineales.

En contraste, el estudio de Wang, Lu y Huo adoptó un enfoque de clasificador de aprendizaje basado en márgenes amplios, estructurado como un árbol binario con múltiples clasificadores SVM (Wang et al., 2006, p.p. 2886-2891). Aunque su modelo, a diferencia del propuesto por Tao Liu y Zhijie Wang, incluía una gama más amplia de categorías de falla, operaba con un conjunto de datos más limitado, lo que podría comprometer su eficacia en condiciones atípicas de gases disueltos identificados en nuestra región.

**Figura 5.** Clasificador de aprendizaje de gran margen.



**Fuente:** Fault diagnosis of power transformer based on large margin learning classifier.

Las metodologías empleadas en los estudios mencionados, junto con la definición restringida de clases de falla y el alcance limitado de los conjuntos de datos, indican que los modelos propuestos poseen un carácter específico más que general. Esta observación resalta la importancia crítica de aplicar técnicas de aprendizaje automático de manera más extensa y diversificada para abordar las problemáticas asociadas al diagnóstico de fallos en transformadores. La necesidad de ampliar el espectro de las categorías de falla e incrementar la amplitud y diversidad de los datos es evidente, lo que a su vez enfatiza la relevancia de desarrollar modelos de aprendizaje automático más holísticos y adaptativos que puedan generalizarse a una variedad más amplia de escenarios operativos y condiciones de falla del sistema eléctrico ecuatoriano.

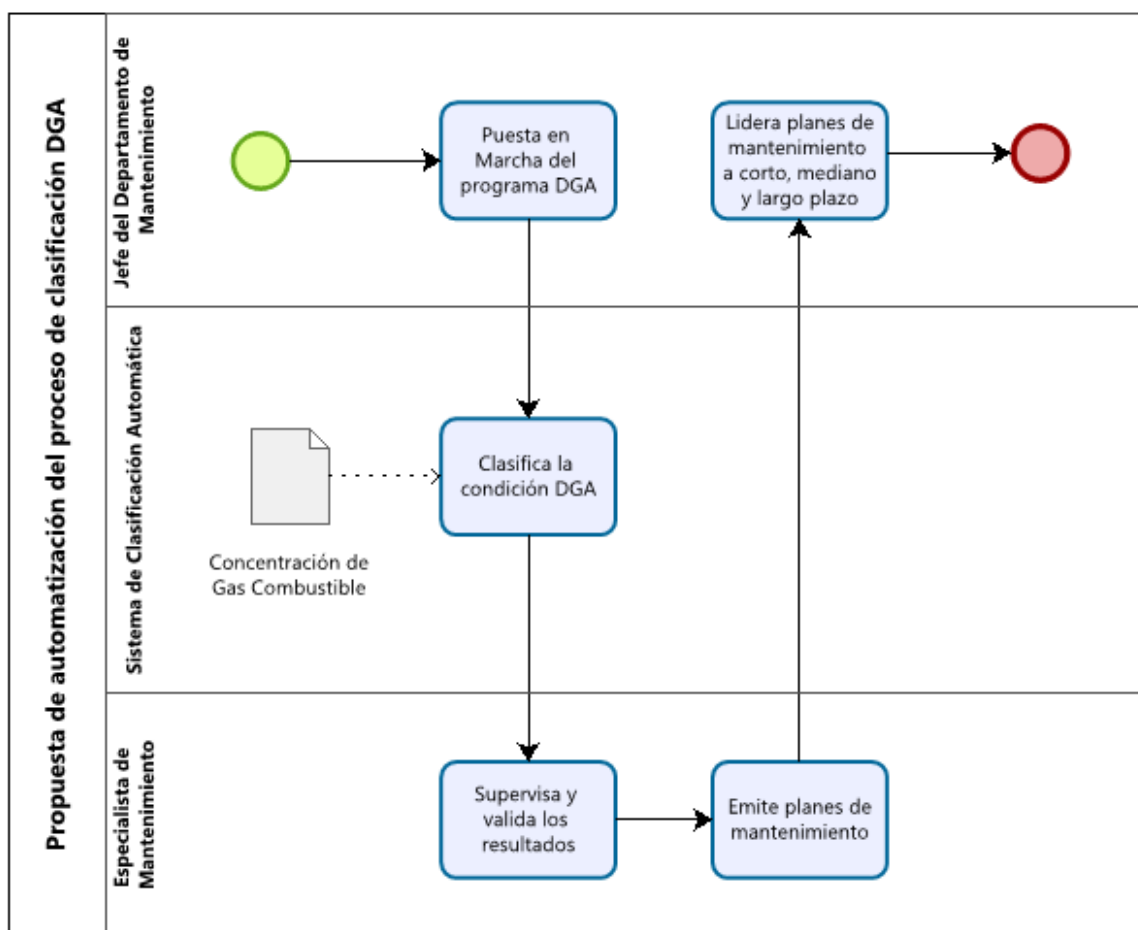
#### **2.4 Automatización del proceso de clasificación de la condición de unidades de transformación del sistema eléctrico ecuatoriano**

Considerando los avances y limitaciones de las metodologías tradicionales para la detección y clasificación automática de fallos en unidades de transformación del sistema eléctrico ecuatoriano, la aplicación de técnicas de aprendizaje automático en conjunto con la metodología estructurada de CRISP-DM puede conducir a un análisis más eficiente y exhaustivo de los datos de concentración de gases combustibles, permitiendo una

interpretación más rápida y precisa de los resultados, mejorando significativamente las prácticas de mantenimiento preventivo y predictivo.

Al personalizar estos modelos para abordar la diversidad y especificidad de las condiciones operativas y fallos en el contexto ecuatoriano, se facilitaría la implementación de una herramienta de inteligencia artificial. Este enfoque no solo reduciría los tiempos de análisis e interpretación de los datos de concentración de gases combustibles en el aceite dieléctrico de transformadores, sino que también aumentaría la confiabilidad y seguridad del sistema eléctrico, contribuyendo a la prevención proactiva de fallos y a la mejora continua en la gestión de activos críticos.

**Figura 6.** Propuesta de automatización del proceso de clasificación DGA.



Powered by  
bizagi  
**Modeler**

**Fuente:** Elaboración Propia.

La propuesta representada en la Figura 6 busca simplificar las operaciones de los departamentos técnicos, transformando a los especialistas en supervisores y analistas de un sistema de diagnóstico automatizado, derivado de la integración de un modelo de inteligencia artificial en el proceso de DGA.

Esta innovación promete no solo agilizar los tiempos de procesamiento y análisis sino también aumentar la capacidad de respuesta frente a incidencias potenciales en transformadores, autotransformadores y reactores de derivación del sistema eléctrico ecuatoriano. El resultado esperado es una mejora sustancial en la eficiencia del mantenimiento preventivo y predictivo, lo que contribuirá significativamente a la confiabilidad y seguridad del sistema eléctrico nacional.

## **CAPÍTULO III. Metodología de la Investigación**

### **3.1. Metodología de la investigación y modelado**

Para resolver la pregunta: ¿es posible obtener un modelo de aprendizaje automático, capaz de identificar fallas en unidades de transformación del sistema eléctrico ecuatoriano, mejorando los procesos actuales de mantenimiento?, resulta esencial seleccionar modelos con capacidad de generalización y sustento teórico. Esto implica establecer un proceso metodológico que transforme los datos históricos obtenidos del análisis de gases disueltos en aceite (DGA) en conocimiento aplicable. El proceso debe culminar con la selección de un algoritmo de clasificación supervisada que se adecue a las condiciones técnicas y operativas del sistema eléctrico ecuatoriano.

Para alcanzar los objetivos de este trabajo, se propone seguir la metodología CRISP-DM, que guiará el desarrollo del proyecto a través de sus fases estructuradas, permitiendo un enfoque sistemático y cíclico. La primera fase del proyecto consistirá en una comprensión profunda del negocio, identificando las necesidades operativas y estratégicas, así como los requerimientos técnicos y las limitaciones del entorno de operación en el sistema eléctrico ecuatoriano.

Dentro de esta fase se realizará una evaluación preliminar de los datos de concentración de gases combustibles, para entender cómo estos pueden ser transformados en información valiosa que alimente el modelo de aprendizaje automático.

### **3.2. Entendimiento del Negocio**

El análisis de las unidades de transformación en el sistema eléctrico ecuatoriano revela una necesidad crítica de monitoreo y diagnóstico eficaz para mantener la confiabilidad y eficiencia del servicio. Tal como se había indicado, los transformadores, autotransformadores y reactores de derivación representan componentes vitales cuyas fallas pueden tener consecuencias significativas que van desde interrupciones del servicio hasta impactos económicos y ambientales severos. Bajo esta premisa, es esencial que las estrategias de mantenimiento y operación estén equipadas con herramientas avanzadas para la detección anticipada de posibles fallas.

En este escenario, el objetivo central del proyecto es determinar la factibilidad de desarrollar un modelo de aprendizaje automático que clasifique el estado de las unidades de transformación. Este modelo buscará mejorar la eficiencia en los tiempos de análisis e

interpretación, asegurándose de que la solución esté adecuadamente adaptada a las condiciones operativas y culturales específicas de Ecuador. La experiencia en la generación exitosa de este modelo beneficiará a las entidades públicas y privadas responsables de la generación, transmisión, subtransmisión y distribución de energía, mejorando sus operaciones y contribuyendo al progreso general del sector energético en el país.

### **3.3. Entendimiento de los datos**

La base de datos comprende 1099 muestras de concentraciones de gases disueltos en aceite aislante, recopiladas entre 2014 y 2023 de 153 unidades de transformación del SNT. Cada muestra contiene 12 atributos, que incluyen la edad del equipo y las concentraciones en partes por millón (ppm) de diversos gases: Hidrógeno ( $H_2$ ), Metano ( $CH_4$ ), Etano ( $C_2H_6$ ), Etileno ( $C_2H_4$ ), Acetileno ( $C_2H_2$ ), Monóxido de Carbono ( $CO$ ), Dióxido de Carbono ( $CO_2$ ), Oxígeno ( $O_2$ ), Nitrógeno ( $N_2$ ), además de la relación ( $O_2/N_2$ ). Estos datos proceden de informes cromatográficos de gases, ejecutados conforme a la normativa ASTM-D3612-C en laboratorios que cuentan con la acreditación ISO/IEC 17025, garantizando así la confiabilidad, trazabilidad y reproducibilidad de las mediciones.

A este conjunto de datos, se ha incorporado la etiqueta de clasificación DGA de cada observación, determinada conforme a los criterios establecidos en la guía estándar IEEE Std. C57.104. La información de las variables del conjunto de datos se describe a continuación:

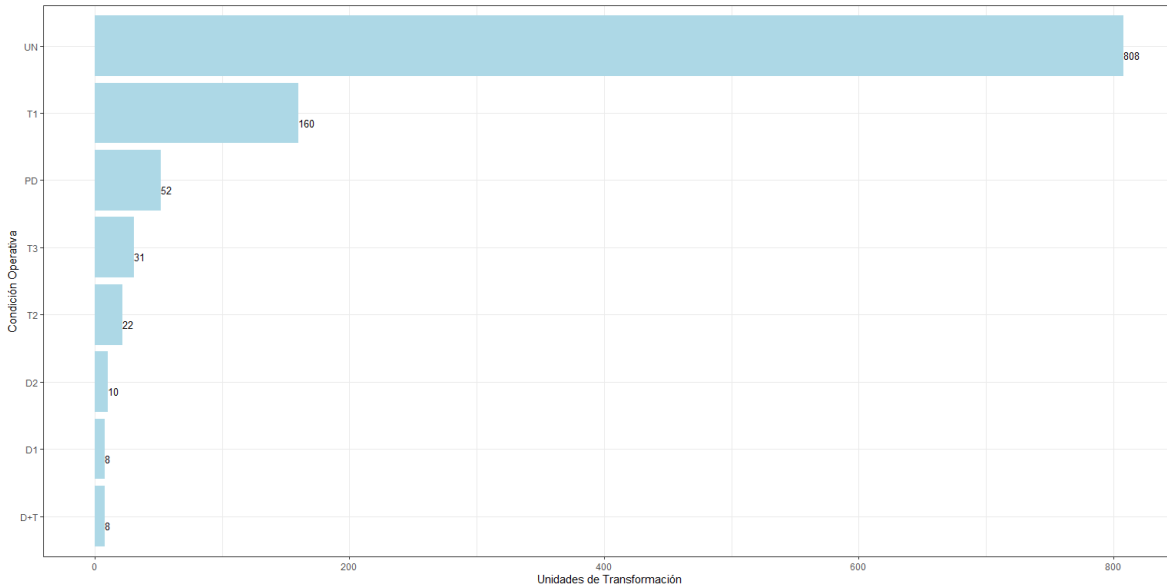
**Tabla 3.** Información de las variables del conjunto de datos.

<b>Atributo</b>	<b>Datos nulos</b>	<b>Tipo de variable</b>
<i>AGE</i>	0	Numérico
<i>O<sub>2</sub>/N<sub>2</sub></i>	0	Numérico
<i>H<sub>2</sub></i>	0	Numérico
<i>CH<sub>4</sub></i>	0	Numérico
<i>C<sub>2</sub>H<sub>6</sub></i>	0	Numérico
<i>C<sub>2</sub>H<sub>4</sub></i>	0	Numérico
<i>C<sub>2</sub>H<sub>2</sub></i>	0	Numérico
<i>CO</i>	0	Numérico
<i>CO<sub>2</sub></i>	0	Numérico
<i>O<sub>2</sub></i>	0	Numérico
<i>N<sub>2</sub></i>	0	Numérico
<i>CONDITION</i>	0	Categorico

**Fuente:** Elaboración Propia.

Dada la falta de indicadores que muestren la condición actual del parque de transformadores, autotransformadores y reactores del SNT, en primer lugar, se realizó un análisis de la distribución de la variable objetivo (*CONDITION*) y de cómo esta varía en función de las distintas zonas operativas del SNT.

**Figura 7.** Condición operativa de las unidades de transformación en el SNT.

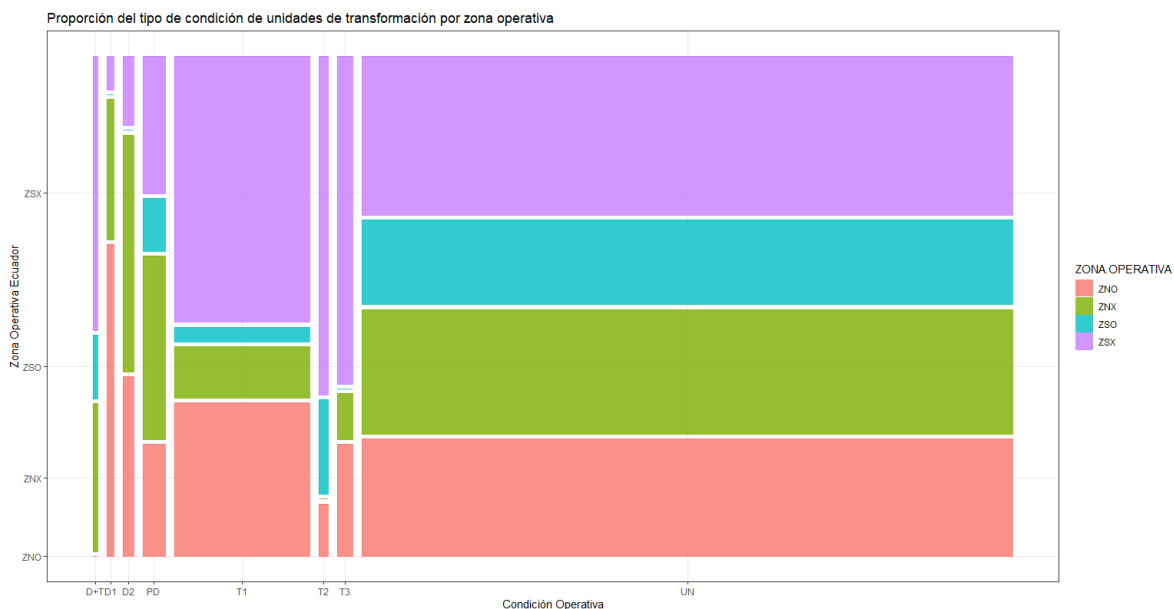


**Fuente:** Elaboración Propia.

La Figura 7 revela que la categoría UN, correspondiente a unidades operativas en estado normal, predomina con 808 observaciones, reflejando que la gran mayoría de las unidades funcionan sin fallas ni anomalías. Esto evidencia que las prácticas de mantenimiento y las condiciones operativas actuales son efectivas, no obstante, las demás categorías indican niveles variados de incidencias.

La categoría T1 señala que las fallas térmicas superiores a 300°C, es la más prevalente después de UN, con 160 muestras afectadas. Por otro lado, aunque las categorías T2, T3, D2 y D+T muestran frecuencias menores, su presencia es significativa ya que implican condiciones internas en las unidades de transformación que requieren grandes cantidades de energía, lo que sugiere la elaboración de planes y acciones preventivas, así como una monitorización constante de la producción de gas para determinar intervenciones correctivas oportunas. Por último, las condiciones PD y D1, aun cuando este tipo de fallas requieren menor cantidad de energía, deben ser objeto de estudio y seguimiento continuo debido a la posibilidad de evolución en sus condiciones DGA.

**Figura 8.** Proporción del tipo de condición de las unidades de transformación por zona operativa del SNT.



**Fuente:** Elaboración Propia.

La Figura 8 ofrece una visión clara sobre la distribución de las condiciones operativas de las unidades de transformación en el SNT, desglosadas por zonas operativas (ZNO, ZNX, ZSO y ZSX). Se destaca que la zona noroccidental muestra una alta proporción de unidades en condición normal, indicativo de un buen rendimiento operativo. En contraste, la zona suroccidental exhibe una variada gama de estados operativos, posiblemente debido a la antigüedad de los equipos en esa área y a la naturaleza de la carga en esa región del país, caracterizada por una intensa actividad industrial, especialmente en las provincias de Guayas y El Oro.

Hecho este primer análisis, a continuación, se detalla un resumen de los estadísticos principales correspondientes a las variables numéricas (predictoras).

**Tabla 4.** Resumen de los estadísticos principales de las variables numéricas.

Variable	Min	Max	Rango	Media	Mediana	Std	Skew	Kurtosis
AGE	0.079399	49.079721	49.000322	23.890526	27.028618	13.116372	-0.196632	-1.370098
O2/N2	0	0.488251	0.488251	0.13343	0.115985	0.103489	0.890529	0.642813
H2	0	3673	3673	10.941128	0	112.34745	31.595826	1028.95001
CH4	0	1037	1037	28.019745	12	50.657496	8.965015	151.282295
C2H6	0	592	592	42.740746	6	90.195774	3.196378	11.019904
C2H4	0	1068	1068	6.924004	1	45.003678	18.825921	397.096838
C2H2	0	200	200	0.343403	0	6.190603	30.819263	987.516316
CO	0	1821	1821	298.360328	200	272.1467	1.63791	3.240754
CO2	29	27237	27208	1951.358508	1550	1917.856083	4.859847	41.005632
O2	0	45533	45533	5424.099181	2939	5905.717311	2.072506	6.316099
N2	1980	133852	131872	44290.19563	40700	26718.14174	0.654425	-0.100745

**Fuente:** Elaboración Propia.

En términos generales, se observa una notable disparidad entre los valores mínimos, máximos y rangos de las variables predictivas. Esto destaca la importancia de normalizar las variables adecuadamente para facilitar la interpretación visual y para evitar que los valores extremos generen sesgos durante el modelado.

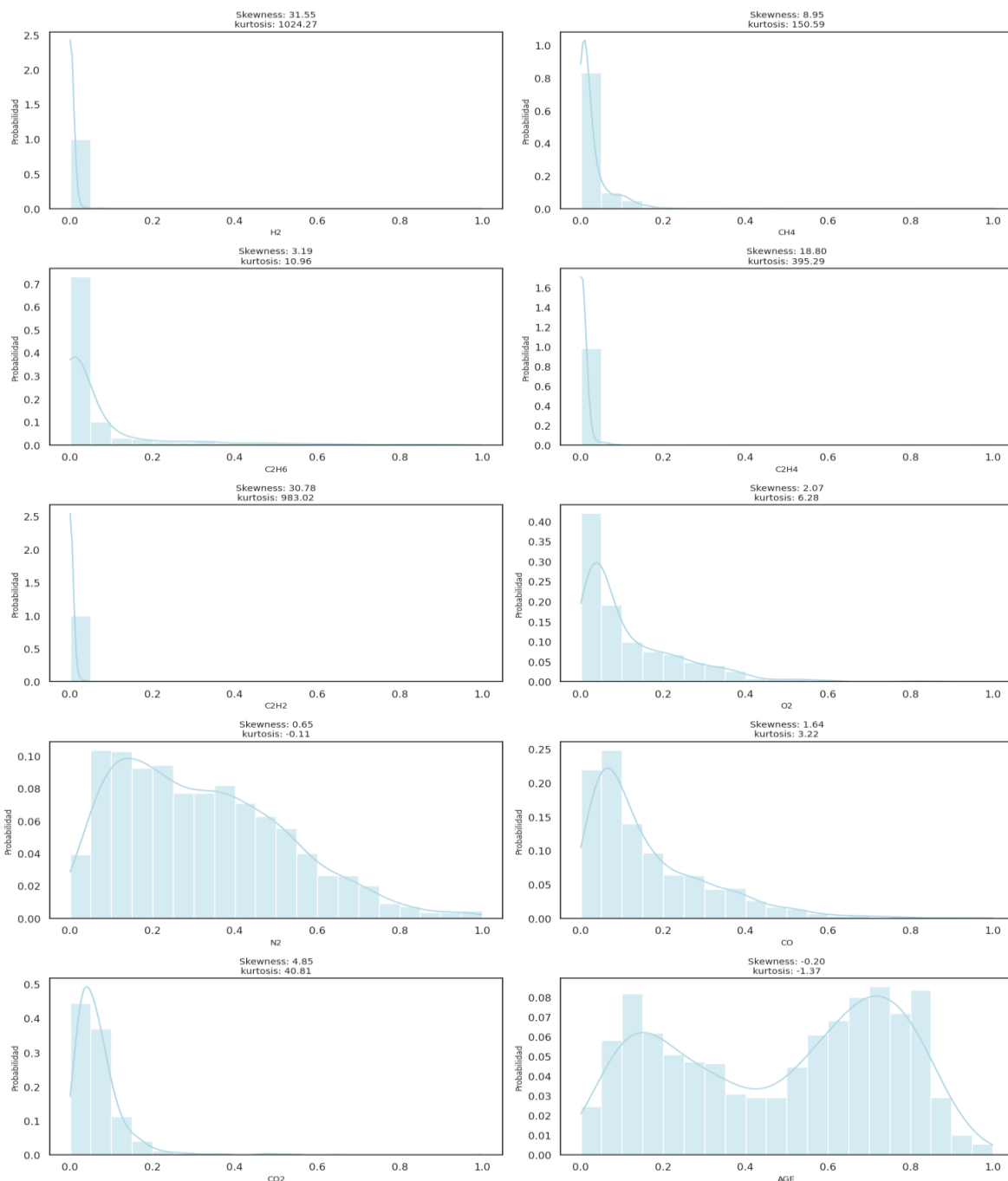
En cuanto a la variable *AGE*, se detecta una amplia gama de edades que van desde poco más de un mes hasta casi 50 años. La mediana excede a la media, indicando una posible inclinación de la distribución hacia unidades más antiguas, aunque la asimetría negativa apunta a una menor presencia de valores extremadamente altos en comparación con los bajos, la curtosis negativa sugiere que la distribución de la edad es relativamente plana, con una menor incidencia de valores extremos que en una distribución normal.

Respecto a los gases disueltos, sus concentraciones presentan un rango amplio, lo cual es coherente con las expectativas de que la producción de gases sea mínima en condiciones normales y aumente con la existencia de problemas internos específicos. Los valores máximos altos reflejan la liberación de energía hacia el aceite y la consiguiente generación de gases debido a fallas eléctricas.

La variable *CO<sub>2</sub>* destaca por no presentar valores mínimos nulos, una condición anticipada en muestras de aceite dieléctrico mineral. No obstante, concentraciones elevadas de dióxido de carbono pueden indicar procesos de oxidación o envejecimiento del papel aislante en las unidades de transformación debido a condiciones sobre todo de alta temperatura en los equipos.

Para profundizar en el análisis estadístico, a continuación, se examina la distribución de las variables numéricas a través de histogramas que ilustran la frecuencia relativa de los diferentes niveles de concentración de los gases disueltos en aceite y la edad de las unidades de transformación.

**Figura 9.** Distribución de las variables numéricas del conjunto de datos.



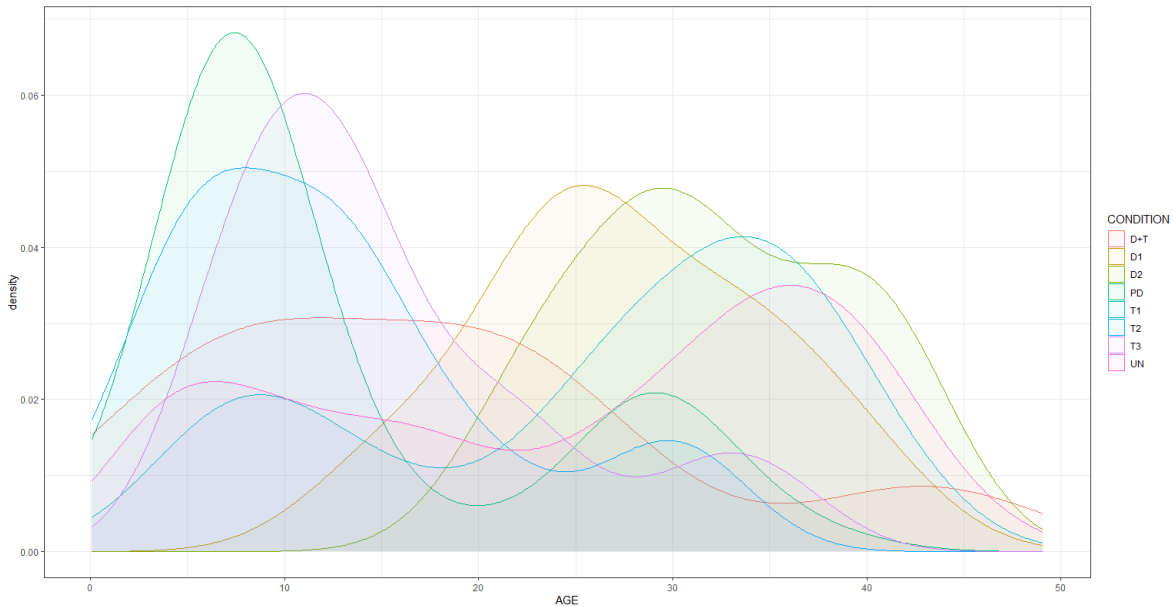
**Fuente:** Elaboración Propia.

Como se ha indicado previamente, las distribuciones presentan una notable diversidad en términos de asimetría y curtosis, lo que evidencia comportamientos marcadamente diferentes entre las variables estudiadas. En lo que respecta a las distribuciones de  $CO$  y  $CO_2$ , ambas exhiben una asimetría positiva y una curtosis significativamente elevada, en particular para el  $CO_2$ . Este patrón podría sugerir que los eventos de alta temperatura, tales como las sobrecargas en transformadores que tienden a incrementar la generación de óxidos de carbono, son eventos poco frecuentes. Este hallazgo es relevante y positivo en relación con la operación del SNT y a la implicación de los agentes de envejecimiento sobre el sistema de aislamiento sólido (papel), ya que altas concentraciones de estos gases podrían indicar procesos de pirólisis.

Respecto a la variable  $AGE$ , los datos reafirman lo anteriormente señalado, es decir, que la curtosis negativa indica una distribución más extendida que la distribución normal, revelando una variabilidad más homogénea en las edades y una menor presencia de valores atípicos. Esto justifica la necesidad de un análisis más detallado para determinar si las edades más elevadas son anomalías puntuales o pertenecen a un subgrupo distinto dentro del conjunto de datos.

Dicho esto, en la Figura 10 muestra un gráfico de densidades que permite examinar con mayor precisión las características intrínsecas de la distribución de la variable  $AGE$ . En este gráfico se puede apreciar cómo se distribuyen las edades de las unidades en estudio, segmentadas por las distintas condiciones de las unidades de transformación del SNT, representados por las curvas de diferentes colores. Este análisis es esencial para comprender los patrones subyacentes que explican esta distribución.

**Figura 10.** Análisis de densidades de la variable AGE.



**Fuente:** Elaboración Propia.

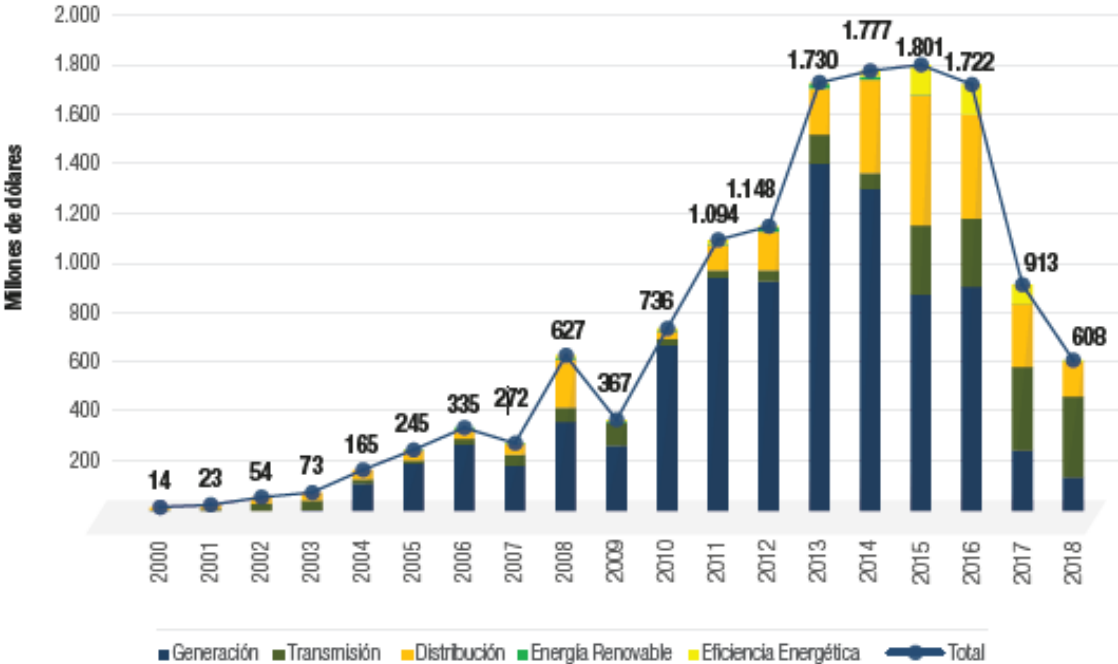
Esta visualización muestra la existencia de dos modas destacadas en la distribución de las edades de los transformadores y autotransformadores, que corresponden con dos subpoblaciones distintas dentro del parque total de equipos. Este fenómeno de binormalidad puede interpretarse a la luz de la evolución histórica del SNT, en donde hasta aproximadamente el año 2013, el sistema se caracterizó por una marcada inercia en términos de incorporación de nuevos proyectos de infraestructura eléctrica (Ministerio de Energía y Recursos Naturales no Renovables del Ecuador, s/f, p. 48); no obstante, a partir de ese momento, el SNT ha evidenciado un vigoroso desarrollo, caracterizado por la constante adición de proyectos de transmisión y, con ello, la integración y puesta en servicio de una nueva generación de unidades de transformación.

Esta dinámica de crecimiento reflejada en la gráfica, ilustra cómo las políticas de inversión y expansión impactan a la composición de la flota de transformadores y autotransformadores. Las dos modas representan, por un lado, las unidades más antiguas que han permanecido desde aproximadamente antes del 2013, y por otro, una población más reciente que ingresó al sistema con la ola de modernización y expansión que siguió.

Por otro lado, del análisis detallado de la primera moda en la gráfica de densidades AGE refleja una preocupante prevalencia de condiciones adversas como PD, D+T, T2 y T3 en la nueva generación de unidades de transformación. Dicha prevalencia

genera signos de alarma, considerando que la esperanza de vida estimada para este tipo de equipos es de alrededor de 25 años. Una posible explicación a esta situación remonta nuevamente al año 2013, momento en el cual el sistema experimentó una significativa expansión como se explicó anteriormente, y en el cual se introdujeron nuevas unidades de transformación al parque existente bajo procedimientos del, a ese entonces, recién creado Sistema Nacional de Contratación Pública – SERCOP.

**Figura 11.** Inversión anual por etapa funcional en el sector eléctrico ecuatoriano.



**Fuente:** Plan Maestro de Electricidad 2019 – 2027.

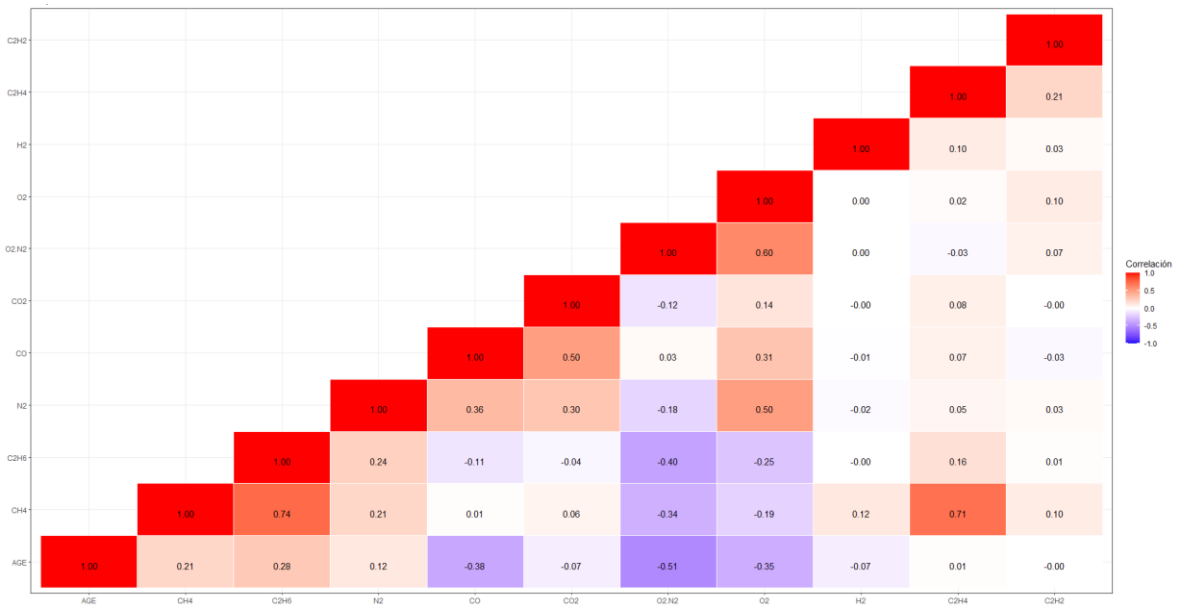
A pesar de que uno de los principios de la Contratación Pública es la optimización de los recursos económicos, las estimaciones de costo-beneficio no parecen haber profundizado en el impacto a largo plazo al haber flexibilizado la incorporación de equipos de menor calidad en el sistema eléctrico ecuatoriano, considerando que desde el nacimiento de la Ley Orgánica del Sistema Nacional de Contratación Pública – LOSNCP, se promovió la participación del mayor número de personas naturales y jurídicas en los procesos contractuales del Estado Ecuatoriano, sin una selección técnica previa de marcas y/o fabricantes de equipos, componentes y accesorios de comprobada trayectoria nacional e internacional (Ley Orgánica Del Sistema Nacional de Contratación Pública, n.d., p. 2). A esta problemática se suma la limitada experiencia del personal de las

entidades contratantes en la formulación de especificaciones técnicas más robustas y detallados para obtener equipos de calidad, derivando en problemas relacionados a los efectos de la obsolescencia programada en la nueva generación de unidades de transformación.

Por otro lado, la segunda moda observada en la Figura 10, alberga la mayor concentración de unidades en estado de condición normal, lo cual subraya la robustez y confiabilidad técnica de transformadores y autotransformadores de marcas empleadas hasta el año 2013, que inclusive han superado su proyección de vida útil. No obstante, es esencial mantener un monitoreo continuo de los equipos clasificados dentro de las condiciones T1, D1 y D2. Esta vigilancia es crítica, ya que tales condiciones pueden obedecer a procesos de envejecimiento normal en curso, señalando la necesidad de establecer estrategias de mantenimiento preventivo o planes de renovación que garanticen la continuidad y eficiencia del servicio público de energía eléctrica.

Tras el análisis de la distribución de la variable *AGE* y su influencia en el comportamiento de los transformadores y autotransformadores del SNT, se procede a explorar más profundamente las interacciones entre las variables del estudio. Para ello, se presenta un mapa de calor de las variables predictoras, herramienta estadística clave que facilita la visualización de las correlaciones entre las variables. Este análisis es fundamental para identificar posibles colinealidades, es decir, situaciones donde dos o más variables están altamente correlacionadas y podrían afectar la precisión de los modelos predictivos.

**Figura 12.** Mapa de calor de las correlaciones entre variables predictoras.



**Fuente:** Elaboración Propia.

Del análisis de la Figura 12, se destaca una marcada correlación entre metano ( $CH_4$ ) y etano ( $C_2H_6$ ), lo que en el ámbito de diagnóstico de transformadores eléctricos suele estar vinculado con fallas térmicas que se producen a temperaturas inferiores a  $300^\circ C$ , como se ha indicado en la Figura 3 de este documento. Esta coocurrencia señala la posibilidad de un calentamiento inicial o problemas en el aislamiento dentro de las unidades de transformación, identificados como T1, que merecen ser analizados y tratados como parte del mantenimiento predictivo-preventivo.

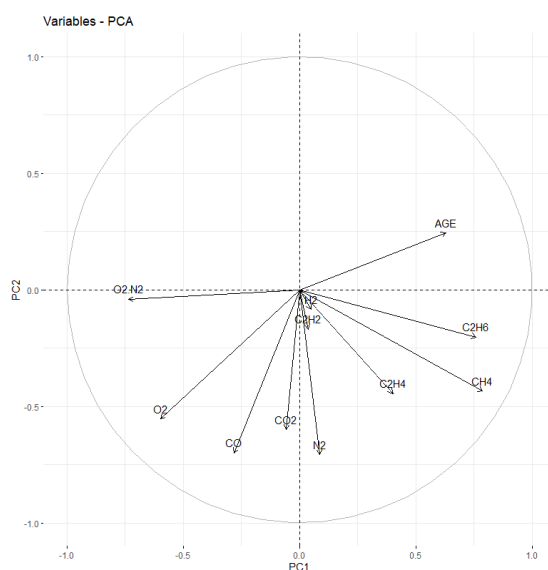
En paralelo, también se observa una correlación significativa entre metano ( $CH_4$ ) y etileno ( $C_2H_4$ ). Esta asociación es particularmente relevante ya que indica la presencia de fallas térmicas de una categoría más severa, identificadas como T2, que se presentan en un rango de temperatura entre  $300^\circ C$  y  $700^\circ C$ . El monitoreo de la relación entre estos dos gases es igualmente importante y debe incorporarse en el análisis detallado de la condición del equipo, para implementar acciones correctivas o preventivas en el marco de un programa de mantenimiento basado en la condición de los activos.

Si bien, desde la perspectiva de la ciencia de datos, la evidencia de correlaciones marcadas subraya la importancia de incorporar la interdependencia entre variables

específicas en el desarrollo de modelos analíticos que clasifiquen el estado operativo de las unidades de transformación. En términos prácticos, se ha corroborado que la presencia conjunta de diversos gases disueltos en el aceite de los transformadores constituye un indicador crítico para la detección y categorización de diferentes tipos de fallas, siendo su comprensión multifacética vital para el diseño de sistemas de diagnóstico precisos y para la mejora de estrategias de mantenimiento predictivo.

A continuación, se muestra un análisis de componentes principales (PCA), técnica estadística empleada para resaltar correlaciones y patrones subyacentes en un conjunto de datos con múltiples variables. Esta representación permite identificar las direcciones de máxima variabilidad y, por ende, descubrir las relaciones más significativas entre las variables, obteniendo una nueva perspectiva de las correlaciones.

**Figura 13.** Varianza explicada por las variables predictoras en PCA.



**Fuente:** Elaboración Propia.

El análisis de la Figura 13 muestra claramente las relaciones entre las variables a través del Análisis de Componentes Principales (PCA). Es evidente que el metano ( $CH_4$ ) muestra una fuerte correlación con el etano ( $C_2H_6$ ) y el etileno ( $C_2H_4$ ), como lo indican las direcciones similares de sus vectores en el gráfico. Este comportamiento concuerda con la comprensión previa de que la presencia combinada de estos gases puede ser sintomática de ciertos tipos de fallos térmicos. A su vez, la edad ( $AGE$ ) indica una asociación menos marcada con estos gases, lo que sugiere que, aunque contribuye a la generación de

gases a lo largo del tiempo, existen otros factores (fallas eléctricas, térmicas o combinadas) determinantes en su producción.

En cuanto a la correlación entre el acetileno ( $C_2H_2$ ) y el nitrógeno ( $N_2$ ), el gráfico de PCA indica una posible relación. Sin embargo, dada la función del nitrógeno como gas inerte utilizado para proteger el aceite de la oxidación dentro de los transformadores, una correlación directa con el acetileno como subproducto de fallas eléctricas severas, no sería intuitiva desde el punto de vista de la química de los procesos internos del transformador. En este sentido, cualquier correlación observada en el PCA entre  $C_2H_2$  y  $N_2$  podría no ser indicativa de una relación causal, sin embargo, merece ser analizada.

En consecuencia de este análisis, se plantea la estrategia de consolidar o excluir variables con el fin de depurar los modelos de clasificación. Este proceso implica discernir y prescindir de variables redundantes que aportan escasamente a la variabilidad global del conjunto de datos, mejorando así la eficiencia y claridad del modelo predictivo.

Finalmente, en el Anexo 3 se presenta una matriz de dispersión que ofrece una representación visual comprensiva de las interacciones entre múltiples variables del conjunto de datos, proporcionando una visión detallada de las correlaciones potenciales y las tendencias predominantes. Este análisis gráfico señala en su parte pertinente que combinaciones de variables como metano ( $CH_4$ ) con etano ( $C_2H_6$ ) y metano ( $CH_4$ ) con etileno ( $C_2H_4$ ) tienden a formar conglomerados específicos asociados a ciertas condiciones de falla, patrones que refuerzan la noción de que estas variables, en conjunto, pueden servir como predictores potentes para clasificar tipos concretos de fallas.

En conclusión, en esta fase de entendimiento de los datos, uno de los descubrimientos más notables para el aprendizaje automático ha sido la correlación significativa detectada entre el metano ( $CH_4$ ) y etano ( $C_2H_6$ ), así como el metano ( $CH_4$ ) y el etileno ( $C_2H_4$ ), siendo estas combinaciones indicadores confiables de fallas térmicas específicas, convirtiéndolos en predictores potenciales clave para la clasificación de tipos de fallas en unidades de transformación.

Con respecto a la variable de edad ( $AGE$ ), esta también tiene una influencia significativa, aunque su correlación con la producción de gases es menor en comparación con las correlaciones entre los gases antes indicados. Un aspecto importante en la distribución de la edad es que esta no se ajusta a un modelo uniforme, sin embargo,

presenta una variabilidad que podría afectar la capacidad predictiva de los modelos. Esto subraya la importancia de integrar la edad en la elaboración de modelos predictivos.

Finalmente, con respecto a la no normalidad de las variables predictoras, se ve necesario implementar estrategias como la transformación de variables y la implementación de modelos tolerantes a la no normalidad, éstos pueden ser: Árboles de Decisión y Bosques Aleatorios.

### 3.4. Preparación de los datos

Avanzando en el ciclo del modelo CRISP-DM, se realiza la preparación de los datos, en donde se ejecutarán acciones requeridas, que son motivadas por los descubrimientos de la etapa previa de entendimiento de los datos. Las actividades esenciales para llevar a cabo en esta fase incluyen:

- a) Depuración de registros duplicados para garantizar la integridad y unicidad de la información.
- b) Escalamiento de las variables predictoras.
- c) Identificación y manejo adecuado de valores atípicos, asegurando que las anomalías no distorsionen los patrones genuinos de los datos.
- d) Transformación de variables categóricas mediante técnicas de codificación para facilitar su uso en algoritmos analíticos.
- e) Selección cuidadosa de las variables más relevantes.

#### 3.4.1 Depuración de registros duplicados

La Tabla 5 detalla que los valores duplicados del volumen total del conjunto de datos original.

**Tabla 5.** *Porción de valores duplicados*

Parámetro	Valores	Porcentaje
Únicos	1090	99.18%
Duplicados	9	0.82%
Total	1099	100%

**Fuente:** Elaboración Propia.

Dado que los valores duplicados representan el 0.82% del total del conjunto de datos original, se eliminan estos registros redundantes. Esta acción es necesaria para prevenir cualquier asignación errónea de clases o impacto negativo en la precisión al aplicar los algoritmos de clasificación.

### 3.4.2 Escalamiento de las variables predictoras

En la fase de entendimiento de los datos, se reveló una disparidad significativa en los valores mínimos, máximos y los rangos de las variables predictoras, lo que puso en manifiesto la necesidad de normalizar o estandarizar los datos para asegurar la coherencia y comparabilidad a lo largo del conjunto de datos.

**Tabla 6.** Resumen de los estadísticos posterior al escalamiento.

Variable	Min	Max	Rango	Media	Mediana	Std	Skew	Kurtosis
AGE	0	1	1	0.485938	0.54998	0.267679	-0.196632	-1.370098
O2/N2	0	1	1	0.273282	0.237552	0.211959	0.890529	0.642813
H2	0	1	1	0.002979	0	0.030587	31.595826	1028.95001
CH4	0	1	1	0.02702	0.011572	0.04885	8.965015	151.282295
C2H6	0	1	1	0.072197	0.010135	0.152358	3.196378	11.019904
C2H4	0	1	1	0.006483	0.000936	0.042138	18.825921	397.096838
C2H2	0	1	1	0.001717	0	0.030953	30.819263	987.516316
CO	0	1	1	0.163844	0.10983	0.149449	1.63791	3.240754
CO2	0	1	1	0.070654	0.055903	0.070489	4.859847	41.005632
O2	0	1	1	0.119125	0.064547	0.129702	2.072506	6.316099
N2	0	1	1	0.320843	0.293618	0.202607	0.654425	-0.100745

**Fuente:** Elaboración Propia.

En la Tabla 6 se presenta los resultados del proceso de escalado a través de la presentación de los estadísticos descriptivos finales. Es relevante señalar que se empleó la librería MinMaxScaler de Scikit-learn en Python para llevar a cabo esta actividad, garantizando así que todas las variables se ajusten a una escala común que es esencial para los análisis subsiguientes y la modelización predictiva.

### 3.4.3 Tratamiento de valores atípicos (outliers)

La Tabla 7 detalla la frecuencia y proporción de valores atípicos en el conjunto de datos, tras haber efectuado la eliminación de entradas duplicadas. Esta tabla ilustra el

conteo y porcentaje de outliers por variable, ofreciendo una perspectiva integral de la incidencia de datos atípicos.

**Tabla 7.** Información descriptiva de valores atípicos.

Variable	Outliers	Porcentaje
$C_2H_6$	152	13.830755
$CH_4$	128	11.646952
$C_2H_4$	109	9.918107
$H_2$	80	7.279345
$C_2H_2$	65	5.914468
$CO_2$	57	5.186533
$O_2$	39	3.548681
$CO$	38	3.457689
$O_2/N_2$	15	1.364877
$N_2$	9	0.818926

**Fuente:** Elaboración Propia.

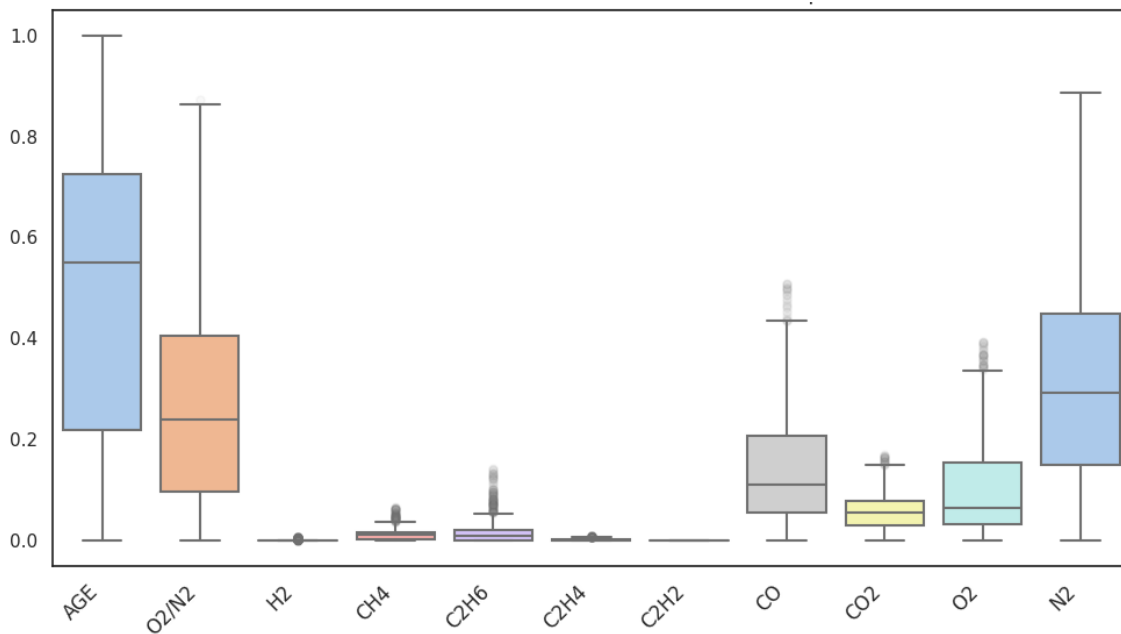
Para abordar esta problemática, se decidió implementar un método en el cual los valores que se encontraban más allá de tres desviaciones estándar de la mediana fueron reemplazados por la misma mediana. Esta decisión se fundamentó en las siguientes razones:

- a) Se reconoció que los valores atípicos reflejan condiciones reales y no errores de muestreo, proporcionando información vital para el entrenamiento del modelo de clasificación.
- b) La mediana, al ser menos susceptible a los valores extremos en comparación con la media, representa de manera más confiable el punto central de los datos, especialmente en distribuciones con presencia de atípicos. Este enfoque permitió preservar la integridad del conjunto de datos, manteniendo una medida fuerte de la tendencia central, lo cual es particularmente valioso en conjuntos con un número limitado de observaciones.

- c) Si bien, inicialmente fue utilizada la metodología de imputación a partir del IQR para detectar atípicos, se determinó que no es el más adecuado para este estudio específico. Los valores atípicos, lejos de ser errores, son indicadores de eventos reales, como fallas en los transformadores. La eliminación de estos mediante el método IQR podría resultar en la pérdida de información crucial para la clasificación.

Por tanto, se justifica la adopción de la técnica de reemplazo utilizando tres desviaciones estándar de la mediana. Esta metodología se basa en la distribución normal, que postula que casi todos los valores (99.7%), deben ubicarse dentro de las tres desviaciones estándar de la media, siendo beneficioso en este caso de estudio, garantizando que se conserven las observaciones esenciales para un modelo de clasificación adaptativo para la condición operativa de la infraestructura del sistema eléctrico ecuatoriano.

**Figura 14.** Inspección de outliers luego del tratamiento.



**Fuente:** Elaboración Propia.

La Figura 14 muestra un diagrama de caja de las variables predictoras después del tratamiento para valores atípicos. Es notorio que las distribuciones de estas variables presentan un sesgo reducido como resultado de la reducción significativa de valores extremos. A continuación, se detallan los estadísticos correspondientes.

**Tabla 8.** Resumen de los estadísticos posterior al tratamiento de outliers.

Variable	Min	Max	Rango	Media	Mediana	Std	Skew	Kurtosis
AGE	0	1	1	0.486306	0.54998	0.268502	-0.200549	-1.376939
O2/N2	0	0.871839	0.871839	0.264441	0.239002	0.194797	0.625694	-0.246067
H2	0	0.075415	0.075415	0.002041	0	0.005511	7.378151	78.133883
CH4	0	0.155256	0.155256	0.023434	0.011572	0.031363	2.002853	3.470741
C2H6	0	0.461149	0.461149	0.04484	0.010135	0.086383	2.819044	7.889005
C2H4	0	0.097378	0.097378	0.0041	0.000936	0.009467	4.896567	29.192015
C2H2	0	0.055	0.055	0.000627	0	0.0038	8.023661	75.874231
CO	0	0.556288	0.556288	0.151828	0.110516	0.1257	1.140689	0.523491
CO2	0	0.267054	0.267054	0.06385	0.055903	0.044462	1.395825	2.40741
O2	0	0.454769	0.454769	0.107	0.064805	0.103606	1.225306	0.562512
N2	0	0.885381	0.885381	0.314466	0.291819	0.194301	0.547247	-0.450568

**Fuente:** Elaboración Propia.

La comparativa entre los datos descriptivos previos y posteriores al manejo de valores atípicos, tal como se detalla en las Tablas 6 y 8, revela una evolución significativa en la estructura de las distribuciones de las variables.

La desviación estándar, ahora más baja en todas las variables, denota una agrupación de datos más próxima a la media. Este cambio muestra que la dispersión, anteriormente amplificada por la presencia de valores atípicos, ha sido efectivamente contenida tras la aplicación del tratamiento específico.

La asimetría, ahora atenuada en la mayoría de las variables, indica una transición hacia distribuciones más equilibradas y simétricas, alejándose de la influencia de valores extremos. Además, un descenso en la curtosis subraya la moderación de las colas distributivas, lo cual implica que la incidencia de datos extremos, potenciales perturbadores del análisis, se ha reducido significativamente.

#### **3.4.4 Transformación de variables categóricas**

En la sección correspondiente a la transformación de variables categóricas, se procedió a convertir los valores categóricos de la columna 'CONDITION' en una representación numérica. Esta codificación es fundamental para que los algoritmos de aprendizaje automático puedan procesar la información, ya que muchos de ellos requieren entradas numéricas.

**Tabla 9.** Codificación de la variable categórica.

CONDITION	CONDITION_CODE	Descripción ampliada del tipo de falla
D+T	0	Falla Térmica y Eléctrica
D1	1	Falla Eléctrica de Bajo Nivel
D2	2	Falla Eléctrica de Alto Nivel
PD	3	Descargas Parciales
T1	4	Falla Térmica $T < 300^{\circ}\text{C}$
T2	5	Falla Térmica $300^{\circ}\text{C} < T < 700^{\circ}\text{C}$
T3	6	Falla Térmica $T > 700^{\circ}\text{C}$
UN	7	Unidad Normal

**Fuente:** Elaboración Propia.

Para llevar a cabo esta tarea se utilizó la librería Scikit-learn de Python, específicamente el método LabelEncoder.

### 3.4.5 Selección de variables

Considerando el conocimiento adquirido en la fase de entendimiento de los datos y con el objeto de depurar los modelos de clasificación a partir de discernir y prescindir de aquellas variables redundantes que aportan escasamente a la variabilidad global del conjunto de datos, mejorando así la eficiencia y claridad del modelo predictivo, a continuación, se presenta la estructura del conjunto de datos final que se utilizará en la etapa de modelado.

**Tabla 10.** Variables seleccionadas para el modelo.

Atributo	Tipo de variable
<i>AGE</i>	Predictor
$O_2/N_2$	Predictor
$H_2$	Predictor
$CH_4$	Predictor
$C_2H_6$	Predictor

$C_2H_4$	Predictor
$C_2H_2$	Predictor
$CO_2$	Predictor
<i>CONDITION_CODE</i>	Objetivo

**Fuente:** Elaboración Propia

Para la elección de variables se ha guiado principalmente por los criterios establecidos en la guía estándar IEEE Std. C57.104, la experiencia consolidada en el diagnóstico y mantenimiento en unidades de transformación, así como por las valiosas percepciones obtenidas durante la fase de comprensión de los datos.

En este capítulo, se ha logrado extraer conocimientos de gran valor que son fundamentales para la toma de decisiones en la gestión del negocio y esenciales para la correcta preparación de los datos. La última sección del capítulo estuvo dedicada a la preparación y estructuración de los datos, asegurando que estén acondicionados para la subsiguiente fase de modelado.

Teniendo en cuenta las características particulares de la distribución de los datos, se ha hecho una selección cuidadosa de modelos analíticos, cuya eficacia no se ve comprometida por la ausencia de normalidad en los datos. Este paso garantizará que las técnicas de modelado aplicadas sean las más adecuadas para abordar la complejidad y la singularidad del conjunto de datos, obteniendo resultados aceptables para la clasificación de las condiciones operativas de las unidades de transformación en el sistema eléctrico ecuatoriano.

## CAPÍTULO IV. Resultados

En este capítulo, se detalla el proceso de modelado, evaluación y selección de un modelo de aprendizaje automático, capaz de identificar fallas en unidades de transformación del sistema eléctrico ecuatoriano.

Siguiendo las observaciones previas en el Capítulo 3 acerca de la no normalidad de las variables predictoras, se requirió la implementación de métodos de modelado estadístico tolerantes a distribuciones no normales. En este contexto, se han seleccionado y aplicado los siguientes modelos de aprendizaje automático:

- a) Árboles de Decisión: Utilizados por su simplicidad y eficacia en la clasificación.
- b) Bosques Aleatorios: Una extensión de los árboles de decisión, conocido por su mayor precisión y resistencia al sobreajuste
- c) Máquina de Vectores de Soporte (SVM): Elegido por su eficiencia en espacios de alta dimensión y su versatilidad en diferentes tipos de datos.
- d) Modelo de K-vecinos más cercanos (KNN): Incluido por su capacidad para realizar clasificaciones basadas en la similitud de las muestras.

La implementación de estos modelos se ha llevado a cabo en Python, utilizando el entorno de Google Colaboratory, dada su accesibilidad y amplias capacidades para la computación científica y el análisis de datos.

### 4.1. Modelamiento

En principio, con un total de 1099 observaciones, se ha establecido una partición del conjunto de datos, asignando un 70% para el entrenamiento y un 30% para la evaluación. Para asegurar la reproducibilidad de los resultados, se ha fijado el parámetro `random_state` en 42; garantizando la consistencia en la selección aleatoria de los datos entre las distintas ejecuciones del modelo.

**Tabla 11.** Porción de valores para entrenamiento y evaluación.

Conjunto	Valores
Entrenamiento	763 x 8
Evaluación	327 x 8

**Fuente:** Elaboración Propia.

Para abordar el desequilibrio en la distribución de la variable objetivo CONDITION\_CODE, detectado en la etapa de análisis exploratorio de datos y evidenciado por una dominancia de la categoría UN en contraste con otras categorías menos representadas, se optó por la implementación de la técnica Synthetic Minority Over-sampling Technique (SMOTE). El procedimiento SMOTE se integró a la fase de modelado para rectificar este balance de clases dentro del conjunto de datos, incrementando efectivamente el número de instancias de las categorías minoritarias. La aplicación de esta función permite reducir los efectos de un sesgo del modelo hacia la clase preponderante y para incrementar la precisión en la detección de las demás condiciones, que son menos frecuentes pero críticos en la predicción de fallos en unidades de transformación.

#### 4.1.1 Entrenamiento del modelo de árboles de decisión

En la fase de entrenamiento del modelo basado en árboles de decisión, se llevó a cabo una búsqueda de los hiperparámetros mediante el proceso de búsqueda en cuadrícula, conocido como Grid Search, combinado con una técnica de validación cruzada de 5 pliegues. Este enfoque sistemático permite evaluar múltiples combinaciones de hiperparámetros para determinar el conjunto óptimo que mejora el rendimiento del modelo.

**Tabla 12.** Hiperparámetros para el entrenamiento del algoritmo de árboles de decisión.

Hiperparámetro	Ajuste
Criterio	gini
Máxima Profundidad	ninguna
Muestras mínimas	1
Muestras mínimas divididas	2

**Fuente:** Elaboración Propia.

Una vez determinados los hiperparámetros óptimos mediante la validación cruzada y la búsqueda en cuadrícula, se procedió al entrenamiento del modelo de árbol de decisión. La exactitud del modelo, sobre el conjunto de entrenamiento, arrojó un valor de 98.63%.

**Figura 15.** Entrenamiento del modelo de árboles de decisión.

```
▼ DecisionTreeClassifier  
DecisionTreeClassifier()
```

**Fuente:** Elaboración Propia.

#### 4.1.2 Entrenamiento del modelo de bosques aleatorios

El proceso de afinación de hiperparámetros para el algoritmo en cuestión se llevó a cabo también utilizando el método de búsqueda exhaustiva Grid Search. Como resultado de esta metodología, se identificaron los ajustes óptimos que permitieron el entrenamiento del modelo de bosques aleatorios con las configuraciones más adecuadas para maximizar su rendimiento.

**Tabla 13.** Hiperparámetros para el entrenamiento del algoritmo de bosques aleatorios.

Hiperparámetro	Ajuste
Criterio	gini
Máxima Profundidad	ninguna
Muestras mínimas	1
Muestras mínimas divididas	2
Número de estimadores	200

**Fuente:** Elaboración Propia.

Tras aplicar los ajustes señalados, el modelo fue entrenado, alcanzando una exactitud del 98.82% en la clasificación con el conjunto de datos de entrenamiento.

**Figura 16.** Entrenamiento del modelo de bosques aleatorios.

```
▼ DecisionTreeClassifier  
DecisionTreeClassifier()
```

**Fuente:** Elaboración Propia.

### 4.1.3 Entrenamiento del modelo de máquinas de vectores de soporte (SVM)

Continuando con la metodología para determinar los ajustes óptimos, se exponen a continuación los hiperparámetros ajustados para el entrenamiento del modelo basado en máquinas de vectores de soporte (SVM).

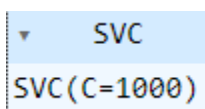
**Tabla 14.** Hiperparámetros para el entrenamiento del algoritmo SVM.

Hiperparámetro	Ajuste
C	1000
gamma	Scale
kernel	rbf

**Fuente:** Elaboración Propia.

La implementación de los hiperparámetros óptimos resultó en una exactitud del 97.16% en el conjunto de datos de entrenamiento.

**Figura 17.** Entrenamiento del modelo de máquinas de vectores de soporte (SVM)

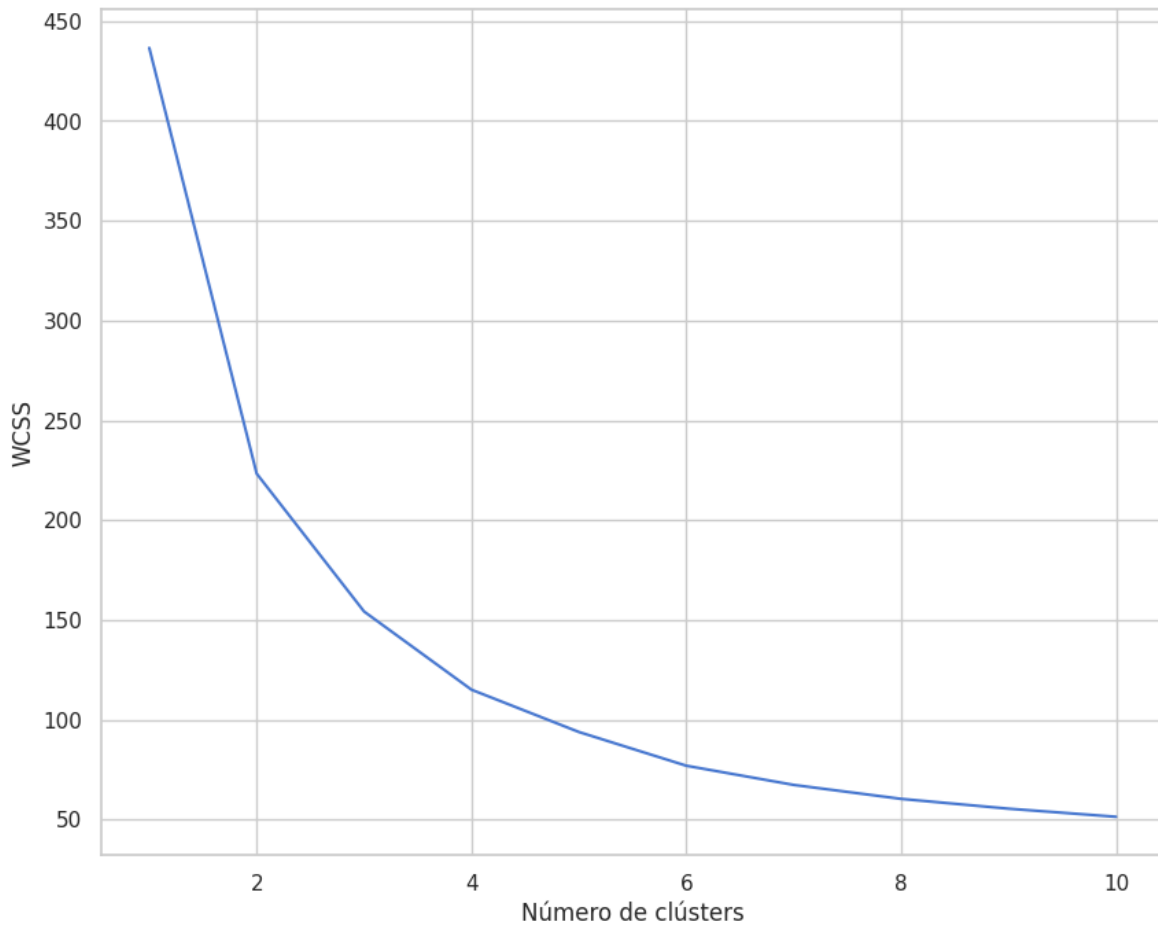


**Fuente:** Elaboración Propia.

### 4.1.4 Entrenamiento del modelo k vecinos más cercanos (KNN)

Para este modelamiento se realizó una evaluación preliminar para determinar la cantidad idónea de clústeres aplicando el método del codo. La inspección de la gráfica revela un punto de inflexión en cuatro clústeres, donde la tendencia de la suma de cuadrados de las distancias dentro de los clústeres (WCSS) comienza a atenuarse, indicando que este es el número óptimo de clústeres para la agrupación de los datos.

**Figura 18.** Determinación del número óptimo de clústeres para el algoritmo KNN.



**Fuente:** Elaboración Propia.

Tras establecer el número de clústeres, el siguiente paso consistió en refinar la configuración del modelo mediante búsqueda de los hiperparámetros residuales. Para ello, se empleó el método de búsqueda en cuadrícula (Grid Search). Los resultados se presentan a continuación.

**Tabla 15.** Hiperparámetros para el entrenamiento del algoritmo KNN

Hiperparámetro	Ajuste
Clústeres	4
Algoritmo	Automático
Tamaño de hoja	20
Método de distancia	Manhattan (p=1)

Hiperparámetro	Ajuste
Peso	Distancia

**Fuente:** Elaboración Propia.

El modelo, una vez entrenado, alcanzó una exactitud del 94.31% al ser evaluado con el conjunto de datos de entrenamiento.

#### 4.2. Evaluación del modelo de clasificación automática para identificar fallas en unidades de transformación del sistema eléctrico ecuatoriano

Luego de la fase de entrenamiento, los modelos se sometieron a evaluación utilizando el conjunto de datos de prueba. Este procedimiento permite medir el desempeño de los algoritmos frente a datos no expuestos previamente durante el entrenamiento, proporcionando así una valoración objetiva de su capacidad predictiva general. Los resultados de esta evaluación fueron cuantificados mediante métricas de precisión (accuracy), sensibilidad (recall) y puntuación F1 (F1-Score).

**Tabla 16.** Métricas de evaluación de los modelos de entrenamiento

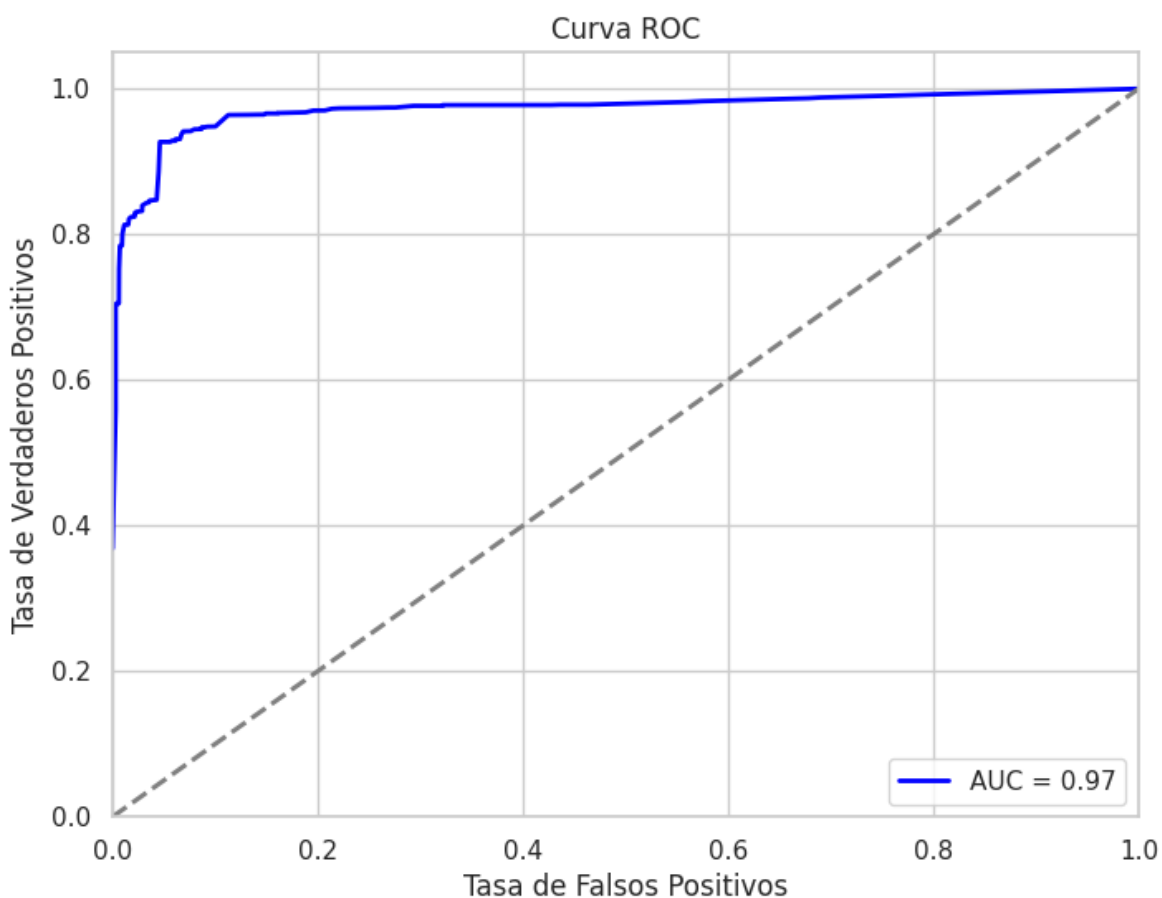
Modelo	Accuracy	Recall	F1-Score
Árbol de Decisión	0.813456	0.813456	0.821986
Bosques Aleatorios	0.911315	0.911315	0.911212
SVM	0.859327	0.859327	0.859106
KNN	0.718654	0.718654	0.770711

**Fuente:** Elaboración Propia.

La Tabla 16 proporciona una comparativa detallada de las métricas de evaluación de cuatro modelos de aprendizaje automático, evidenciando que el modelo de bosques aleatorios (Random Forest) sobresale con una puntuación superior en todas las métricas: exactitud (accuracy), sensibilidad (recall) y la puntuación F1 (F1-Score), con valores aceptables, consistentes y coherentes. Este resultado indica que el modelo de bosques aleatorios no solo predice con mayor precisión las clases correctas, sino que también mantiene un equilibrio entre la precisión y la predicción de clases positivas, lo cual es fundamental en contextos donde las consecuencias de la predicción errónea pueden ser críticas.

En contra parte, el algoritmo KNN muestra un rendimiento significativamente inferior en el conjunto de datos de prueba, lo cual sugiere que su capacidad de generalización es limitada en comparación con los demás modelos, convirtiéndose en la opción menos viable para la clasificación automática de las condiciones de las unidades de transformación en el sistema eléctrico ecuatoriano, especialmente en un entorno operativo donde se pueden encontrar múltiples y variadas configuraciones de los datos.

**Figura 19.** Curva ROC y AUC



**Fuente:** Elaboración Propia.

La Curva Característica Operativa del Receptor (ROC) presentada, ostentando un Área Bajo la Curva (AUC) de 0.97, indica una excelente capacidad del modelo de árboles de decisión para distinguir entre las diferentes condiciones operativas de las unidades de transformación del sistema eléctrico ecuatoriano. Un AUC aproximándose al valor unitario denota una elevada sensibilidad del modelo, implicando que tiene una gran probabilidad de identificar correctamente los casos positivos, mientras simultáneamente exhibe una

elevada especificidad, reflejada en una tasa reducida de falsos positivos, lo que valida las métricas obtenidas en la Tabla 16.

**Tabla 17.** Matriz de confusión de la evaluación del modelo arboles de decisión

	D+T	D1	D2	PD	T1	T2	T3	UN
D+T	1	0	0	0	0	1	0	0
D1	0	1	1	0	0	0	0	0
D2	0	0	1	0	0	0	0	1
PD	0	0	0	14	0	0	0	2
T1	0	0	0	0	43	0	0	8
T2	0	0	0	0	1	6	0	1
T3	0	0	0	0	1	0	5	0
UN	1	0	0	4	7	0	1	227

**Fuente:** Elaboración Propia.

La matriz de confusión contenida en la Tabla 17 evidencia un rendimiento competente del modelo de árboles de decisión, especialmente en la categorización precisa de la condición operativa Normal (UN), donde se observa una asignación correcta en 227 ocasiones. Aunque la clasificación de algunas condiciones de falla por parte del modelo no alcanza el mismo nivel de exactitud, esto no desmerece la capacidad general del modelo. Las discrepancias en el rendimiento para estas categorías están relacionadas con una representación insuficiente de ejemplos en el conjunto de entrenamiento. Cabe señalar que, pese a estos desafíos, el modelo ha demostrado ser eficaz en la identificación de estados operativos normales, lo cual es corroborado por un AUC de 0.97 en la curva ROC, reflejando una alta capacidad de discriminación general.

#### 4.3. Discusión

**Organizar y depurar los datos históricos del análisis de gases disueltos en aceite DGA, de los transformadores, autotransformadores y reactores existentes en el sistema eléctrico ecuatoriano.**

Durante la fase de recolección de datos, se obtuvo información detallada sobre la concentración de gases disueltos en el aceite dieléctrico de 153 unidades de transformación pertenecientes al SNT, durante el período comprendido entre 2014 y 2023. Esta muestra es representativa del sistema eléctrico ecuatoriano, ya que abarca instalaciones distribuidas en las cuatro regiones geográficas del país. Tal distribución geográfica asegura que el conjunto de datos refleja las diversas prácticas de mantenimiento y tipo de carga característica de cada región del país, proporcionando así una visión integral de las condiciones a lo largo del territorio ecuatoriano.

Como resultado de esta fase investigativa, se consolidó y depuró la información recopilada, culminando en la creación de una base de datos. Los datos se extrajeron de informes cromatográficos generados por laboratorios acreditados bajo ISO/IEC 17025. Estos informes ofrecen un registro de las concentraciones de gases disueltos, proporcionando un conjunto de datos confiable y detallado para posteriores análisis.

**Comparar la precisión de los resultados de los algoritmos de aprendizaje supervisado de clasificación, permitiendo determinar una solución determinante para los problemas de diagnóstico de fallas en unidades de transformación del sistema eléctrico ecuatoriano.**

Para este punto, y siguiendo las recomendaciones del Capítulo 3, se procedió a la selección y aplicación de cuatro técnicas de modelado estadístico, todas reconocidas por su tolerancia ante distribuciones no gaussianas. Los algoritmos de Árboles de Decisión, Bosques Aleatorios, Máquina de Vectores de Soporte (SVM) y K-Vecinos más Cercanos (KNN) fueron sometidos a procesos de entrenamiento y evaluación, con el fin de identificar el modelo más eficiente en términos de precisión, sensibilidad y puntuación F1. El modelo basado en Bosques Aleatorios se distinguió por obtener los mejores resultados en estas métricas.

A pesar de que se observó una tendencia del modelo entrenado por el algoritmo de bosques aleatorios a clasificar incorrectamente ciertas fallas como condiciones normales, un comportamiento que se atribuye a la preponderancia de la categoría normal (UN) en el conjunto de datos (73.50% de las instancias, como se indica en la Figura 7), la generalidad y eficiencia de este enfoque de clasificación automática son aceptables.

No obstante, pese a estas limitaciones observadas, la habilidad del modelo para automatizar la detección de fallas en las unidades de transformación es notable y presenta un potencial de mejora considerable. Este potencial se manifiesta en la posibilidad de incrementar la precisión y sensibilidad del modelo con respecto a las fallas, mediante la ampliación del volumen y la diversidad de datos de entrenamiento para las clases de falla menos representadas. Integrar un conjunto de datos de entrenamiento más equilibrado, particularmente aumentando los ejemplos de condiciones anómalas, mejoraría la capacidad del modelo para clasificar con eficacia un espectro más amplio de escenarios operativos, reduciendo así los efectos del sesgo hacia la condición normal y reforzando la confiabilidad de las predicciones en situaciones críticas.

**¿Es posible obtener un modelo de aprendizaje automático, capaz de identificar fallas en unidades de transformación del sistema eléctrico ecuatoriano, mejorando los procesos actuales de mantenimiento?**

Los hallazgos preliminares indican que, aunque existen desafíos inherentes relacionados con el desequilibrio de clases y la generalización de los modelos, la aplicación de algoritmos de aprendizaje automático demuestra una capacidad significativa para automatizar y afinar la detección de estados anómalos. Los modelos estudiados, particularmente el de bosques aleatorios, han mostrado promesa en alcanzar altos niveles de exactitud, aunque su rendimiento varía al enfrentar categorías de falla menos representadas en los datos de entrenamiento. Por tanto, se deduce que la viabilidad de alcanzar un modelo óptimo reside en la mejora continua del entrenamiento de estos algoritmos, enfocándose en la inclusión de un espectro más amplio de datos de fallas en unidades de transformación.

En resumen, este estudio subraya que, con este procedimiento y un conjunto de datos bien constituido, es viable implementar un sistema de aprendizaje automático efectivo que no solo identifique con precisión las fallas en las unidades de transformación, sino que también ofrezca una mejora sustancial en las prácticas actuales de mantenimiento del sistema eléctrico ecuatoriano.

## CAPÍTULO V. Conclusiones y Recomendaciones

### 5.1. Conclusiones

- Comparado con los algoritmos de clasificación sugeridos por varios autores para resolver la problemática, el modelo de Bosques Aleatorios ha evidenciado una superioridad, con una precisión y sensibilidad aproximadas del 91.13%, en la tarea de clasificación automática para la detección de fallos en unidades de transformación. Esta eficacia se atribuye al entrenamiento y evaluación con una población más significativa de muestras y a la adaptación a las peculiaridades de los datos del sistema eléctrico ecuatoriano, los cuales presentaron desafíos específicos en términos de normalidad estadística y desbalance de clases. A pesar de estas complejidades, el modelo se posiciona como un candidato potencial para su implementación en el sistema eléctrico ecuatoriano.
- A partir de la experimentación individual con varias observaciones de concentración de gases de las unidades de transformación del SNT, la implementación del modelo entrenado para automatizar el proceso de DGA ha demostrado ser eficiente, logrando una reducción de tiempo en el procesamiento de datos de aproximadamente el 62% en comparación con métodos tradicionales. Este ahorro de tiempo permitirá enfocar los recursos a la elaboración de planes de mantenimiento más detallados y estructurados para equipos con fallos identificados, reforzando así las hipótesis iniciales en donde se relacionó la implementación de modelos de aprendizaje automático con la mejora de la gestión de activos del sistema eléctrico ecuatoriano.
- Frente a la falta de indicadores que reflejen el estado actual del parque de transformadores, autotransformadores y reactores del SNT, la fase de entendimiento de los datos seguido por la metodología CRISP-DM, reveló una amplia gama de condiciones operativas en las unidades de transformación, reflejando efectividad en las prácticas de mantenimiento, pero también evidenciando la presencia de diversas fallas que requieren atención.

- Otro hallazgo en torno a la fase de comprensión de los datos de este estudio ha sido la identificación de una relación estadísticamente significativa entre la presencia de metano ( $CH_4$ ) y etano ( $C_2H_6$ ), así como entre el metano ( $CH_4$ ) y el etileno ( $C_2H_4$ ). Si bien, para la ciencia de datos estas correlaciones podrían sugerir la eliminación de variables redundantes, en el ámbito de la ingeniería del mantenimiento de transformadores, este vínculo entre los gases se presenta como un indicador confiable de ciertos tipos de fallas térmicas. Esta evidencia subraya la importancia de integrar análisis de datos multidisciplinar en la mejora de estrategias de mantenimiento, permitiendo un diagnóstico más preciso y eficaz de las fallas en transformadores eléctricos.
- El tratamiento de valores atípicos y la normalización de las variables predictoras han demostrado ser esenciales para mejorar la precisión y la capacidad de generalización en los modelos de clasificación. En particular, en el modelo de Bosques Aleatorios (Random Forest), las técnicas de transformación aplicadas en este estudio han resultado en un rendimiento aceptable, evidenciado por las métricas de evaluación.
- Por otro lado, a pesar de la implementación de la técnica Synthetic Minority Over-sampling Technique (SMOTE) para mitigar el desequilibrio observado en la distribución de la variable objetivo `CONDITION_CODE`, persiste la necesidad de abordar de manera más efectiva la desigualdad de clases en el conjunto de datos de entrenamiento. Para mejorar la eficacia de estos modelos, es esencial ampliar y diversificar el espectro de datos relacionados con fallas, permitiendo que los algoritmos se entrenen de una variedad más rica y representativa de escenarios de fallas, aumentando su capacidad para identificar y clasificar con precisión las diversas condiciones operativas y de falla en las unidades de transformación.
- En el ámbito del aprendizaje automático aplicado a sistemas eléctricos, los modelos evaluados, y en particular el algoritmo de Bosques Aleatorios ha demostrado una capacidad destacada para identificar las condiciones

operativas normales en las unidades de transformación. Como ya se había indicado, este estudio revela que enriquecer el conjunto de datos con un mayor número y diversidad de ejemplos en las categorías de fallas menos representadas podría incrementar de manera notable tanto la precisión como la capacidad de detección de anomalías del modelo.

- A pesar de estas consideraciones, se resalta que, con la aplicación de esta metodología y con un conjunto de datos exhaustivamente preparado, es factible el desarrollo de un sistema de aprendizaje automático eficiente para la identificación y clasificación de fallas en las unidades de transformación del sistema eléctrico ecuatoriano. Este avance representa una contribución significativa a la mejora de las estrategias de mantenimiento predictivo y preventivo, subrayando el potencial de estas tecnologías para mejorar la gestión de activos y aumentar la fiabilidad y seguridad de las operaciones en el sector eléctrico.

## 5.2. Recomendaciones

- Para mejorar la precisión y sensibilidad de los modelos de aprendizaje automático, es recomendable incrementar la cantidad de ejemplos de fallos en el conjunto de datos, poniendo especial énfasis en las categorías de fallas que están menos representadas. Esta estrategia resulta especialmente útil tanto en empresas públicas como privadas dedicadas a la generación y distribución de energía, permitiendo así abordar de manera efectiva el desequilibrio de clases y mitigar los efectos del sesgo en la clasificación. Sin embargo, la implementación de este procedimiento contribuirá a la creación de modelos más confiables, capaces de reflejar con mayor precisión la diversidad de condiciones operativas y fallas en unidades de transformación del sistema eléctrico ecuatoriano.
- Es recomendable profundizar en los estudios sobre la correlación observada entre el Acetileno ( $C_2H_2$ ) y el Nitrógeno ( $H_2$ ). Aunque, como se menciona en el Capítulo III, el nitrógeno se utiliza habitualmente como gas inerte para prevenir la oxidación del aceite en los transformadores, su correlación directa con el acetileno (un subproducto típico de fallas

eléctricas severas), se podría intuir como una interacción química no intuitiva en el contexto de los procesos internos del transformador. Este fenómeno inesperado podría señalar un avance significativo en la ingeniería y mantenimiento de transformadores de potencia, que merece una investigación detallada, lo cual podría aportar nuevas perspectivas en la detección y prevención de fallos, y potencialmente redefinir las prácticas actuales en el mantenimiento de transformadores.

- Con base en los resultados obtenidos del análisis de la distribución de la variable *AGE*, que representa la edad de los transformadores y autotransformadores, se sugiere que el SERCOP implemente un mecanismo de precalificación de proveedores a nivel estatal para los procesos contractuales del Estado Ecuatoriano. Este mecanismo debe incluir un análisis técnico detallado de los requerimientos de equipos y componentes industriales, tomando en cuenta su trayectoria y referencias nacionales e internacionales, previo a proceder con su evaluación de conformidad legal. Adicionalmente, se recomienda a los encargados de la elaboración de los documentos preparatorios y comisiones técnicas de los procedimientos de adquisición, establezcan criterios técnicos más rigurosos y exhaustivos. Estos criterios deben asegurar la adquisición de equipos de transformación de alta calidad, incluyendo especificaciones detalladas que consideren la longevidad operativa de los equipos y su resistencia frente a la obsolescencia programada, acciones que no solo mejoraría la eficiencia y la calidad de los equipos adquiridos, sino que también contribuiría a la real optimización de los recursos públicos en el sector energético ecuatoriano.
- Finalmente, es importante destacar la necesidad de automatizar procesos de gestión de activos de la infraestructura del sector eléctrico ecuatoriano. La implementación de la propuesta de automatización indicada en la Figura 6, apuesta por convertir a las áreas especializadas de mantenimiento en supervisores del proceso de Análisis de Gases Combustibles (DGA), reduciendo los tiempos de clasificación de fallas y enfocando sus recursos al planeamiento de estrategias dedicadas de mantenimiento sobre activos,

lo cual es una mejora evidente a los procesos tradicionales de evaluación, en donde, es común que las áreas de mantenimiento del sector eléctrico ecuatoriano archiven los datos por falta de recursos (horas-hombre de personal experimentado), lo que a mediano o largo plazo generan salidas no programadas de bahías de transformación de subestaciones o patios de elevación, debido a la falta de interpretación de resultados.

## CAPÍTULO VI. Referencias

- Ley Orgánica del Sistema Nacional de Contratación Pública, Pub. L. No. Segundo Suplemento del Registro Oficial 392, 17-II-2021.
- Duval, M. (1989). Dissolved gas analysis: It can save your transformer. *IEEE Electrical Insulation Magazine*, 5(6), 22–27. <https://doi.org/10.1109/57.44605>
- Ernesto Gallo Martínez. (2021). *Diagnóstico y mantenimiento a transformadores en campo* (Vol. 3).
- Fiallos, R. J. (2017). Dissolved gas content forecasting in power transformers based on Least Square Support Vector Machine (LSSVM). *Latin American Journal of Computing*, IV(3), 55–60.
- IEEE Std C57.104. (2019). IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers, 2019-06-13. In *IEEE* (Vol. 2019). <https://ieeexplore.ieee.org/document/8890040>
- INTERNATIONAL ELECTROTECHNICAL, & COMMISSION. (2022). Mineral oil-filled electrical equipment in service - Guidance on the interpretation of dissolved and free gases analysis. *IEC 60599*, 4.
- Liu, T., & Wang, Z. (2009). Design of power transformer fault diagnosis model based on support vector machine. *2009 International Symposium on Intelligent Ubiquitous Computing and Education, IUCE 2009*, 137–140. <https://doi.org/10.1109/IUCE.2009.59>
- Ministerio de Energía y Recursos Naturales no Renovables del Ecuador. (n.d.). *Plan Maestro de Electricidad*.
- Sloat, T. K., Johnson, J. L., & Sommerman, G. M. L. (1965). Gas evolution from transformer oils under high-voltage stress. *EIC 1965 - Proceedings of the 6th Electrical Insulation Conference*, 234–238. <https://doi.org/10.1109/EIC.1965.7461236>
- Wang, X. Z., Lu, M. Z., & Huo, J. B. (2006). Fault diagnosis of power transformer based on large margin learning classifier. *Proceedings of the 2006 International Conference on*

*Machine Learning and Cybernetics*, 2006, 2886–2891.  
<https://doi.org/10.1109/ICMLC.2006.259075>

**ANEXO 1: Tablas 1, 2, 3 y 4 de la guía estándar IEEE Std. C57.104**

**Tabla 1-A.** Límites de concentración de gases del percentil 90 en función de la relación O<sub>2</sub> / N<sub>2</sub> y la edad del transformador.

		Relación O <sub>2</sub> / N <sub>2</sub> ≤ 0.2				Relación O <sub>2</sub> / N <sub>2</sub> > 0.2			
		Edad del Transformador [años]				Edad del Transformador [años]			
		No conocido	1 - 9	10 - 30	> 30	No conocido	1 - 9	10 - 30	> 30
GAS	Hidrógeno (H <sub>2</sub> )	80	75		100	80	40	40	
	Metano (CH <sub>4</sub> )	90	45	90	110	90	45	20	
	Etano (C <sub>2</sub> H <sub>6</sub> )	90	30	90	150	90	30	15	
	Etileno (C <sub>2</sub> H <sub>4</sub> )	50	20	50	90	50	20	25	60
	Acetileno (C <sub>2</sub> H <sub>2</sub> )	1	1			2	2		
	Monóxido de Carbono (CO)	900	900			500	500		
	Dióxido de Carbono (CO <sub>2</sub> )	9000	5000	10000		5000	3500	5500	

**Fuente:** IEEE Std. C57.104 (Tabla 1).

**Tabla 1-B.** Límites de concentración de gases del percentil 95 en función de la relación O<sub>2</sub> / N<sub>2</sub> y la edad del Transformador.

		Relación O <sub>2</sub> / N <sub>2</sub> ≤ 0.2				Relación O <sub>2</sub> / N <sub>2</sub> > 0.2			
		Edad del Transformador [años]				Edad del Transformador [años]			
		No conocido	1 - 9	10 - 30	> 30	No conocido	1 - 9	10 - 30	> 30
GAS	Hidrógeno (H <sub>2</sub> )	200	200			90	90		
	Metano (CH <sub>4</sub> )	150	100	150	200	50	60	30	
	Etano (C <sub>2</sub> H <sub>6</sub> )	175	70	175	250	40	30	40	
	Etileno (C <sub>2</sub> H <sub>4</sub> )	100	40	95	175	100	80	125	
	Acetileno (C <sub>2</sub> H <sub>2</sub> )	2	2		4	7	7		
	Monóxido de Carbono (CO)	1100	1100			600	600		
	Dióxido de Carbono (CO <sub>2</sub> )	12500	7000	14000		7000	5000	8000	

**Fuente:** IEEE Std. C57.104 (Tabla 2).

**Tabla 1-C.** Valores del percentil 95 para variaciones de nivel absoluto (delta), entre sucesivas muestras de DGA

		Variación Máxima $\mu\text{L/L}$ (ppm) entre muestras de laboratorio DGA consecutivas	
		Relación $\text{O}_2 / \text{N}_2 \leq 0.2$	Relación $\text{O}_2 / \text{N}_2 > 0.2$
GAS	Hidrógeno ( $\text{H}_2$ )	40	25
	Metano ( $\text{CH}_4$ )	30	10
	Etano ( $\text{C}_2\text{H}_6$ )	25	7
	Etileno ( $\text{C}_2\text{H}_4$ )	20	
	Acetileno ( $\text{C}_2\text{H}_2$ )	Cualquier Incremento	
	Monóxido de Carbono ( $\text{CO}$ )	250	175
	Dióxido de Carbono ( $\text{CO}_2$ )	2500	1750

Fuente: IEEE Std. C57.104 (Tabla 3).

**Tabla 1-D.** Valores del percentil 95 del análisis de velocidad multipuntos (3-6 puntos) de muestras de DGA

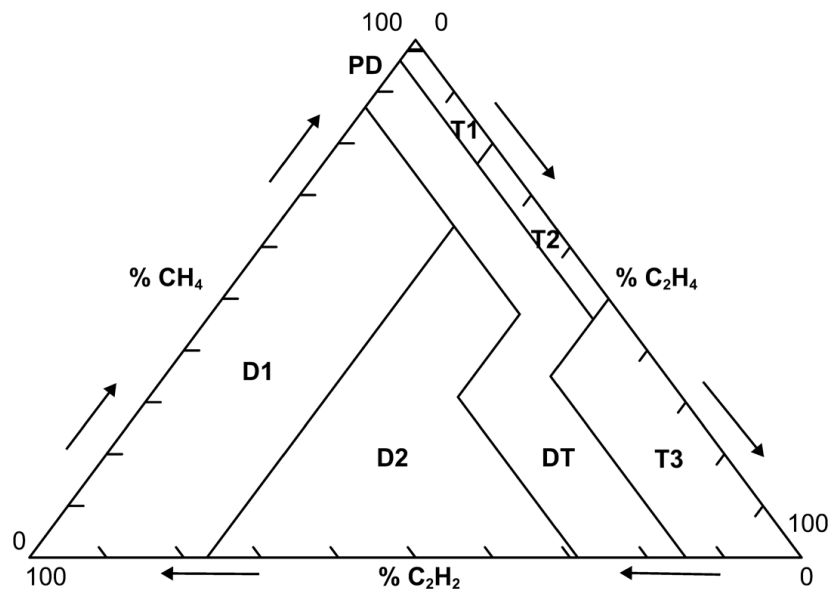
		Período entre la primera y la última muestra de la serie $\mu\text{L/L/año}$ (ppm/año)			
		Relación $\text{O}_2 / \text{N}_2 \leq 0.2$		Relación $\text{O}_2 / \text{N}_2 > 0.2$	
		4 - 9 meses	10 - 24 meses	4 - 9 meses	10 - 24 meses
GAS	Hidrógeno ( $\text{H}_2$ )	50	20	25	10
	Metano ( $\text{CH}_4$ )	15	10	4	3
	Etano ( $\text{C}_2\text{H}_6$ )	15	9	3	2
	Etileno ( $\text{C}_2\text{H}_4$ )	10	7	7	5
	Acetileno ( $\text{C}_2\text{H}_2$ )	Cualquier tasa de incremento		Cualquier tasa de incremento	
	Monóxido de Carbono ( $\text{CO}$ )	200	100	100	80
	Dióxido de Carbono ( $\text{CO}_2$ )	1750	1000	1000	800

Fuente: IEEE Std. C57.104 (Tabla 4).

## ANEXO 2: Metodología del Triángulo de Duval

El Método del Triángulo Duval se ilustra en la Figura 2-A:

**Figura 2-A.** Método del triángulo de Duval.

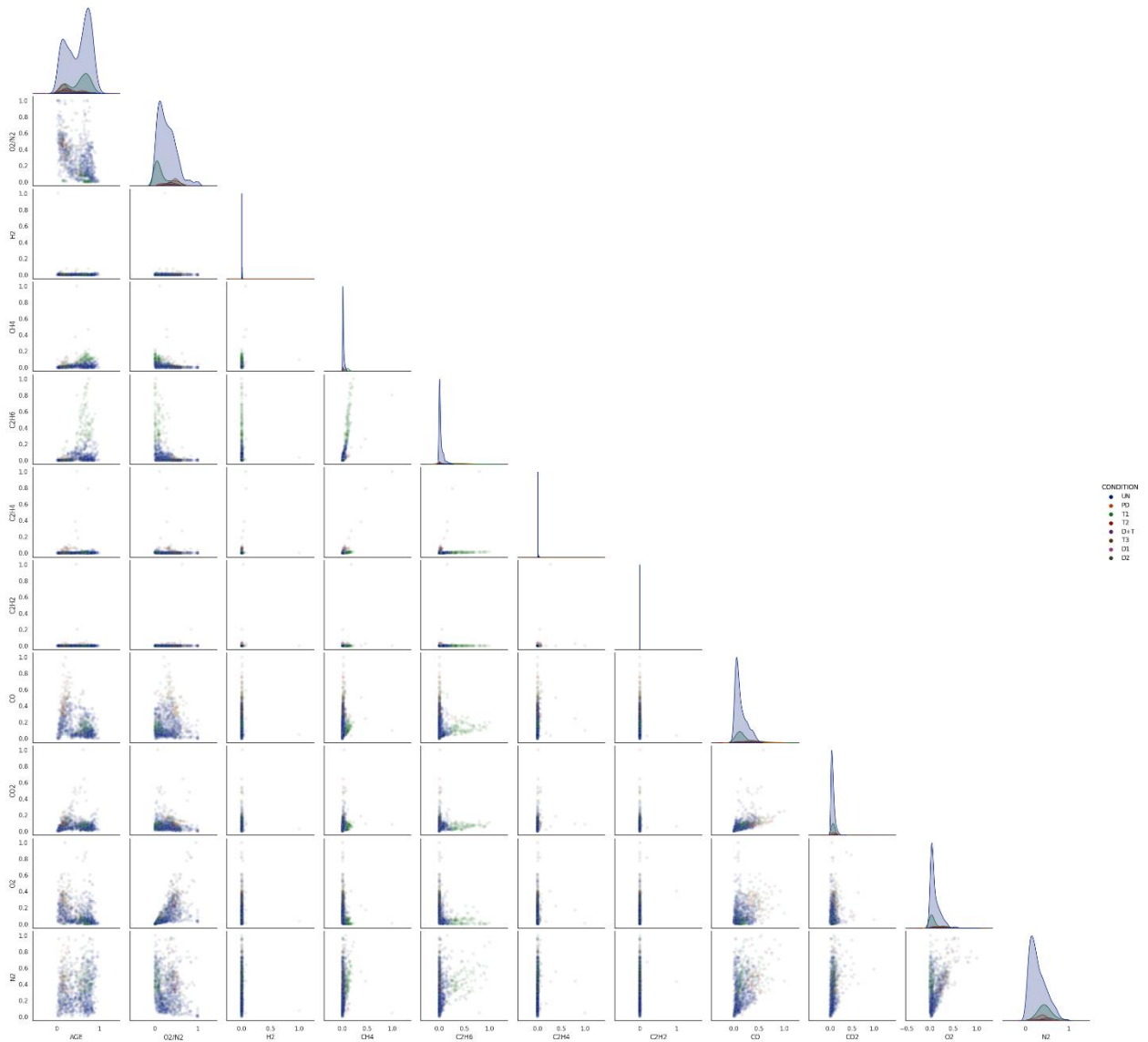


**Fuente:** IEEE Std. C57.104.

El Método del Triángulo Duval utiliza tres gases que corresponden al contenido energético o la temperatura crecientes de las fallas: metano ( $CH_4$ ) para las fallas de baja energía/temperatura, etileno ( $C_2H_4$ ) para las fallas de alta temperatura y acetileno ( $C_2H_2$ ) para las fallas de muy alta temperatura/energía/arqueo. En cada lado del triángulo se representan los porcentajes relativos de estos tres gases.

### ANEXO 3: Matriz de Dispersión

**Figura 3-A.** Matriz de dispersión entre variables predictoras



**Fuente:** Elaboración Propia.

La matriz de dispersión presentada proporciona una visualización de las relaciones entre pares de variables en relación con distintos tipos de fallas en transformadores, permitiendo observar que combinaciones de variables como metano ( $CH_4$ ) con etano ( $C_2H_6$ ) y metano ( $CH_4$ ) con etileno ( $C_2H_4$ ) tienden a formar conglomerados específicos

asociados a ciertas condiciones de falla. Este patrón indica que estas variables, en conjunto, pueden servir como predictores potentes para clasificar tipos concretos de fallas. La presencia de clústeres definidos indica la viabilidad de desarrollar un modelo de clasificación que pueda discriminar eficazmente entre las distintas fallas basándose en estas variables.

Es necesario prestar especial atención a la naturaleza de las distribuciones, ya que su falta de uniformidad puede sugerir la implementación de estrategias para adecuar los datos previo al modelado. Esto podría incluir la aplicación de transformaciones estadísticas que normalicen las distribuciones o seleccionar modelos de clasificación automática poco sensibles a la no normalidad de los datos como son los Árboles de Decisión, Random Forest y Máquinas de Soporte Vectorial (SVM) con kernels no lineales.

## **ANEXO 4: Script Google Colaboratory**

---