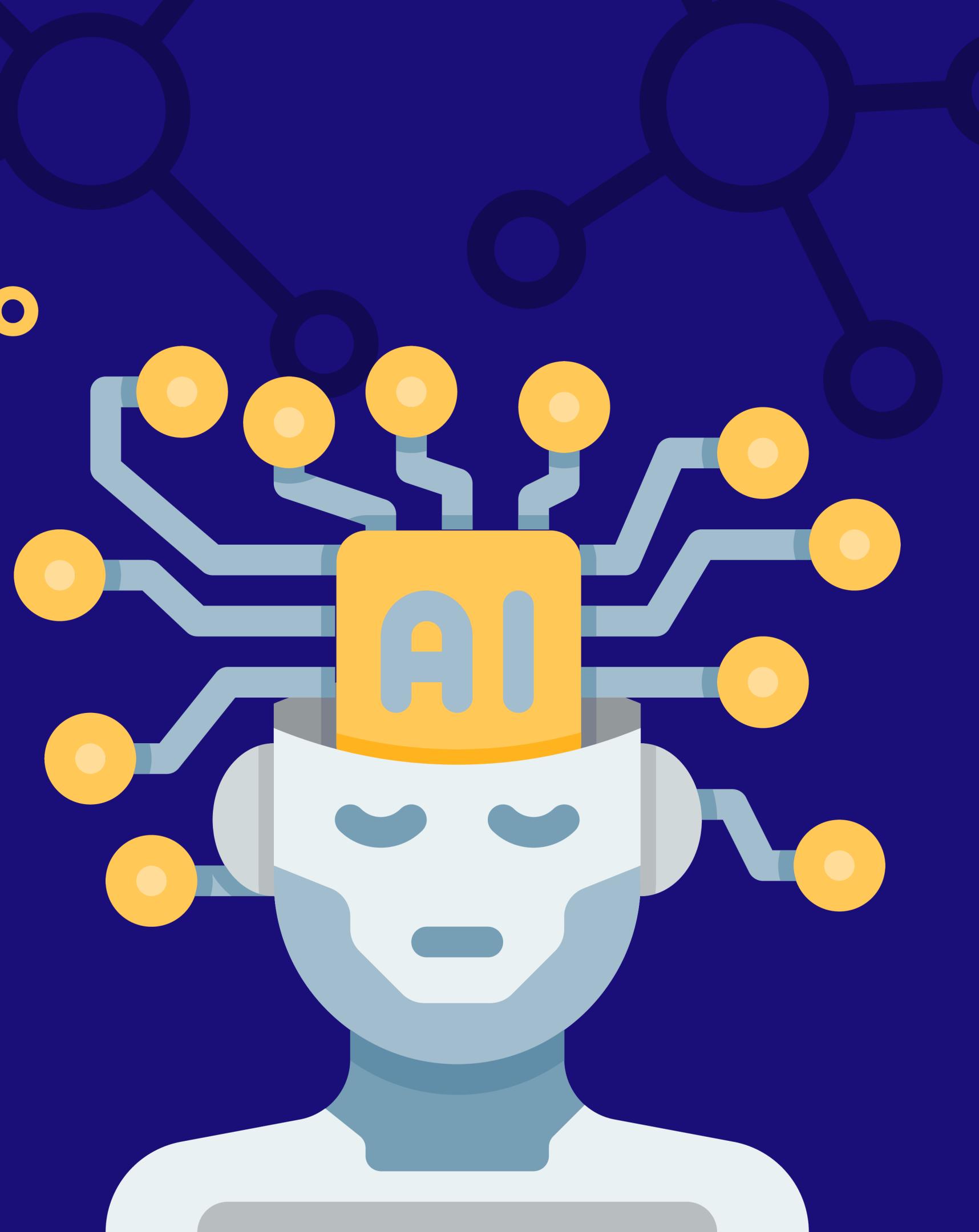


Mantenimiento Predictivo de Transformadores Eléctricos Usando Modelos de Machine Learning

Alumnos :

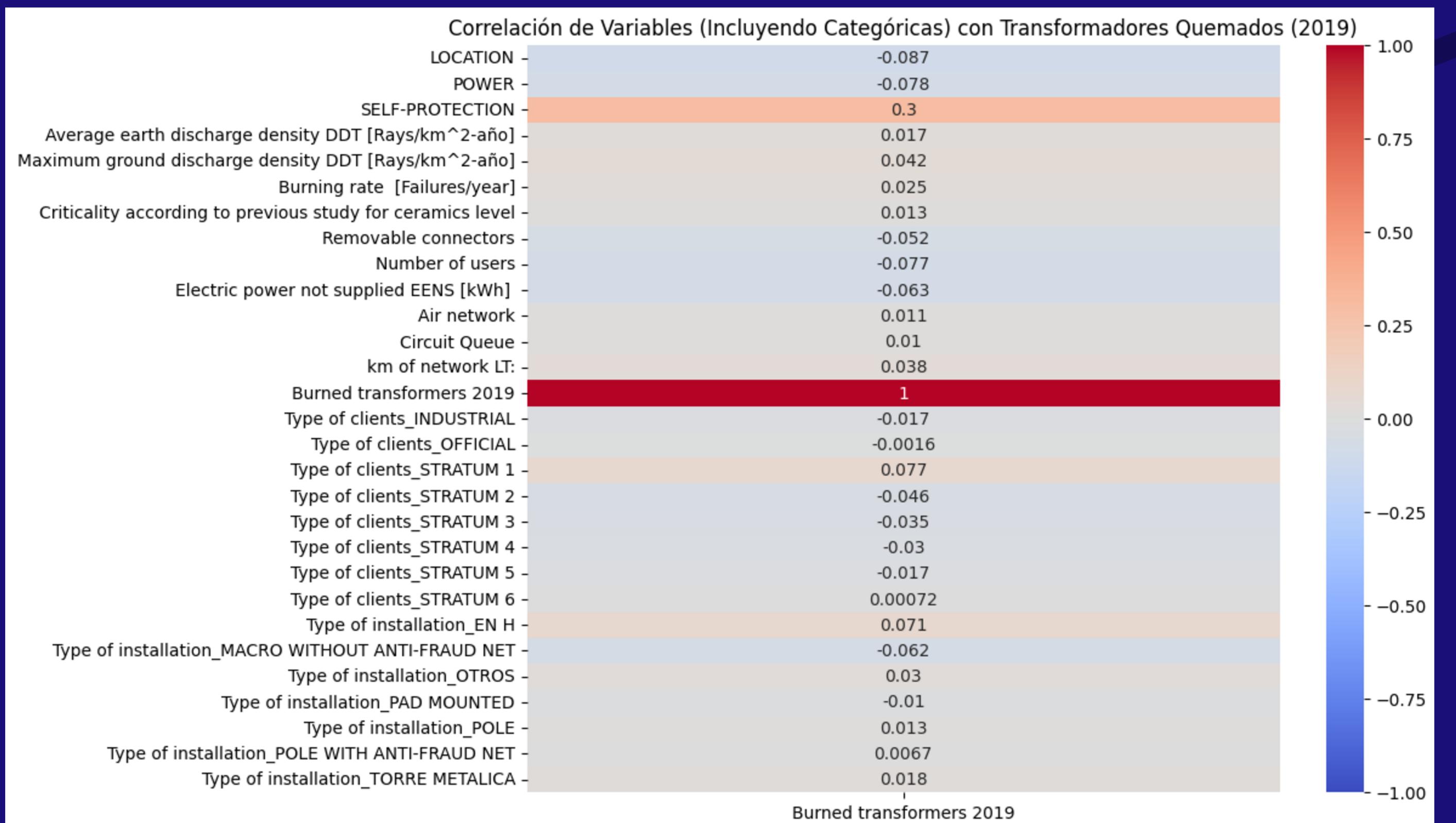
- Guillermo Diego Ojeda Cueto - EMCA-
C.I.: 5354320
- Robert Sebastian Colman Céspedes -
EMCA - C.I.: 5178781



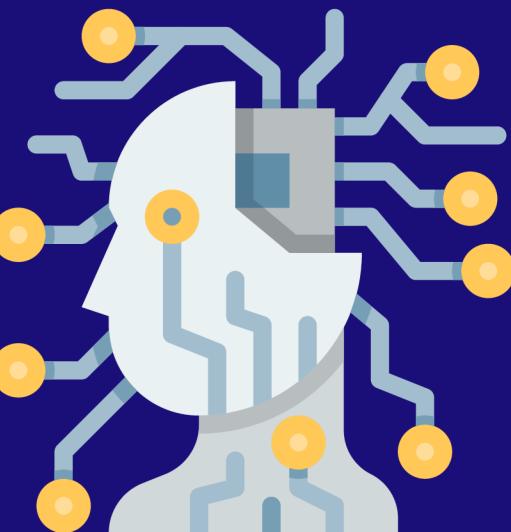
INTRODUCCIÓN

Este trabajo propone un enfoque basado en machine learning para predecir fallos en transformadores eléctricos, utilizando un dataset desbalanceado de la Compañía Energética de Occidente (Colombia). Mediante el desarrollo de características clave, balanceo de datos y optimización de hiperparámetros, se logró mejorar significativamente la predicción de fallos, evaluada con métricas como precisión y F1-score. Los resultados confirman la viabilidad del enfoque y su potencial para sistemas predictivos en infraestructura eléctrica.

Descripción del dataset



Universidad del
Cauca ,Colombia



Selección de Variables

Un análisis de correlación se llevó a cabo para identificar las variables más influyentes en las fallas de los transformadores. Las principales características correlacionadas con los fallos incluyen.

1

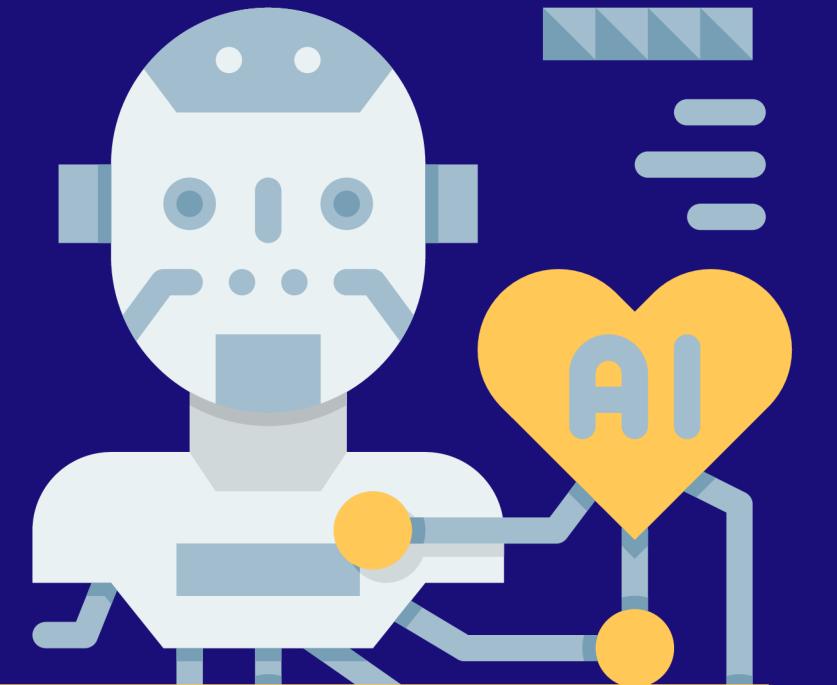
Self-Protection: Transformadores sin protección presentan mayor riesgo de fallos por sobrecargas o cortocircuitos.

2

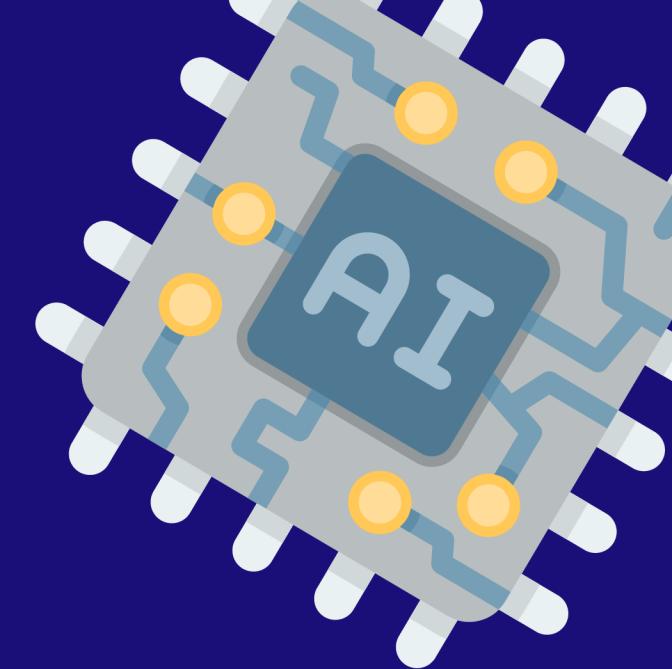
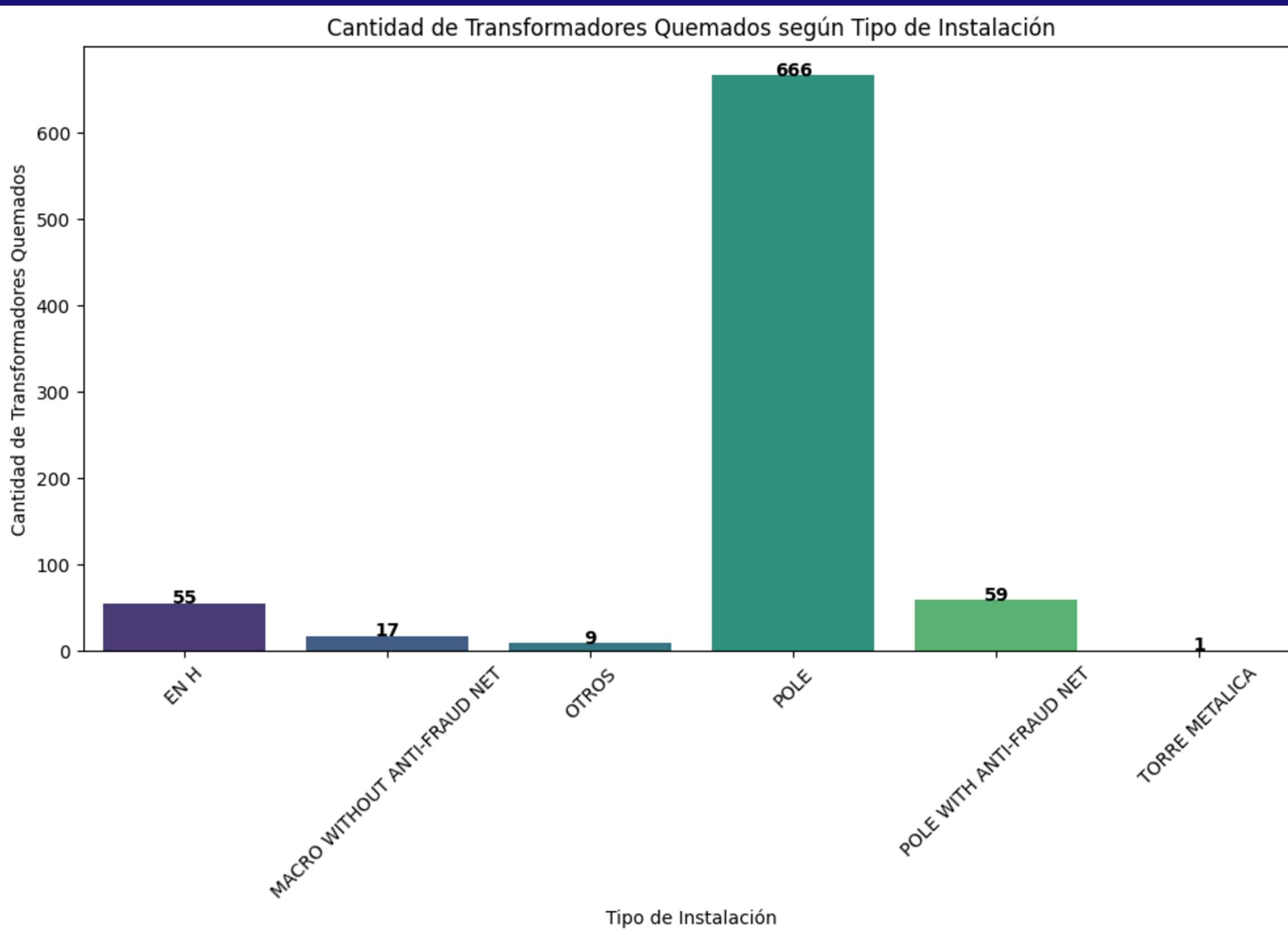
Tipo de Cliente: Transformadores en áreas con bajos recursos tienden a fallar más por menor mantenimiento.

3

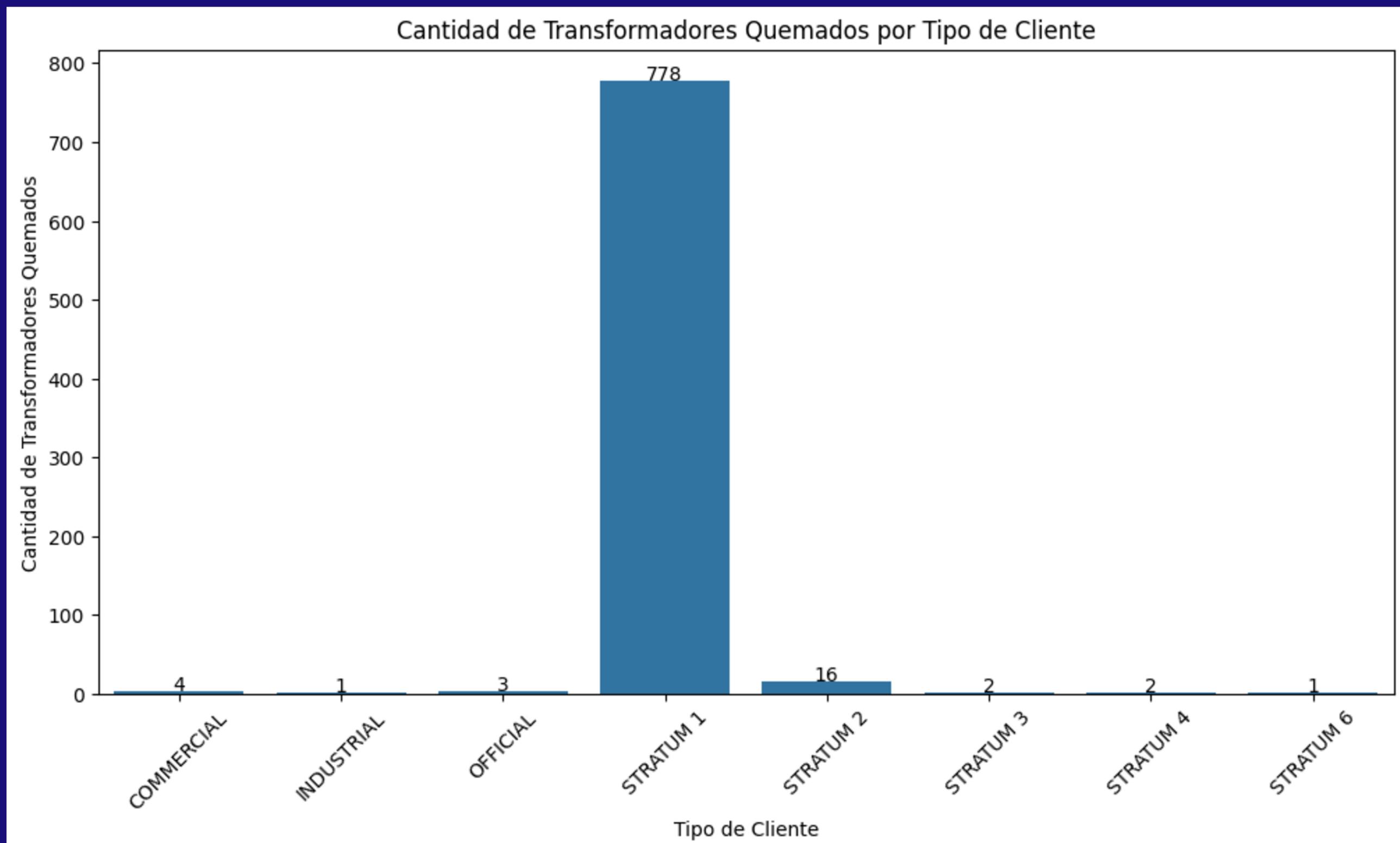
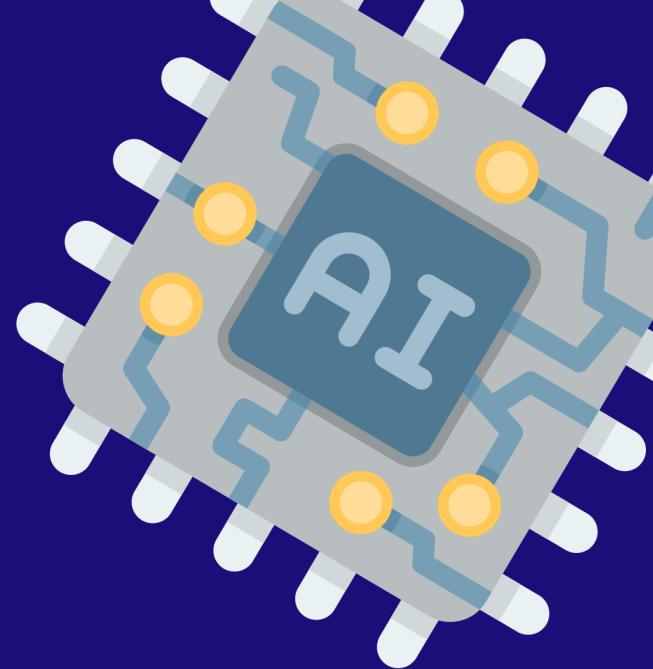
Tipo de Instalación: Instalaciones aéreas o inadecuadas son más vulnerables a fallos por tensión y condiciones ambientales.



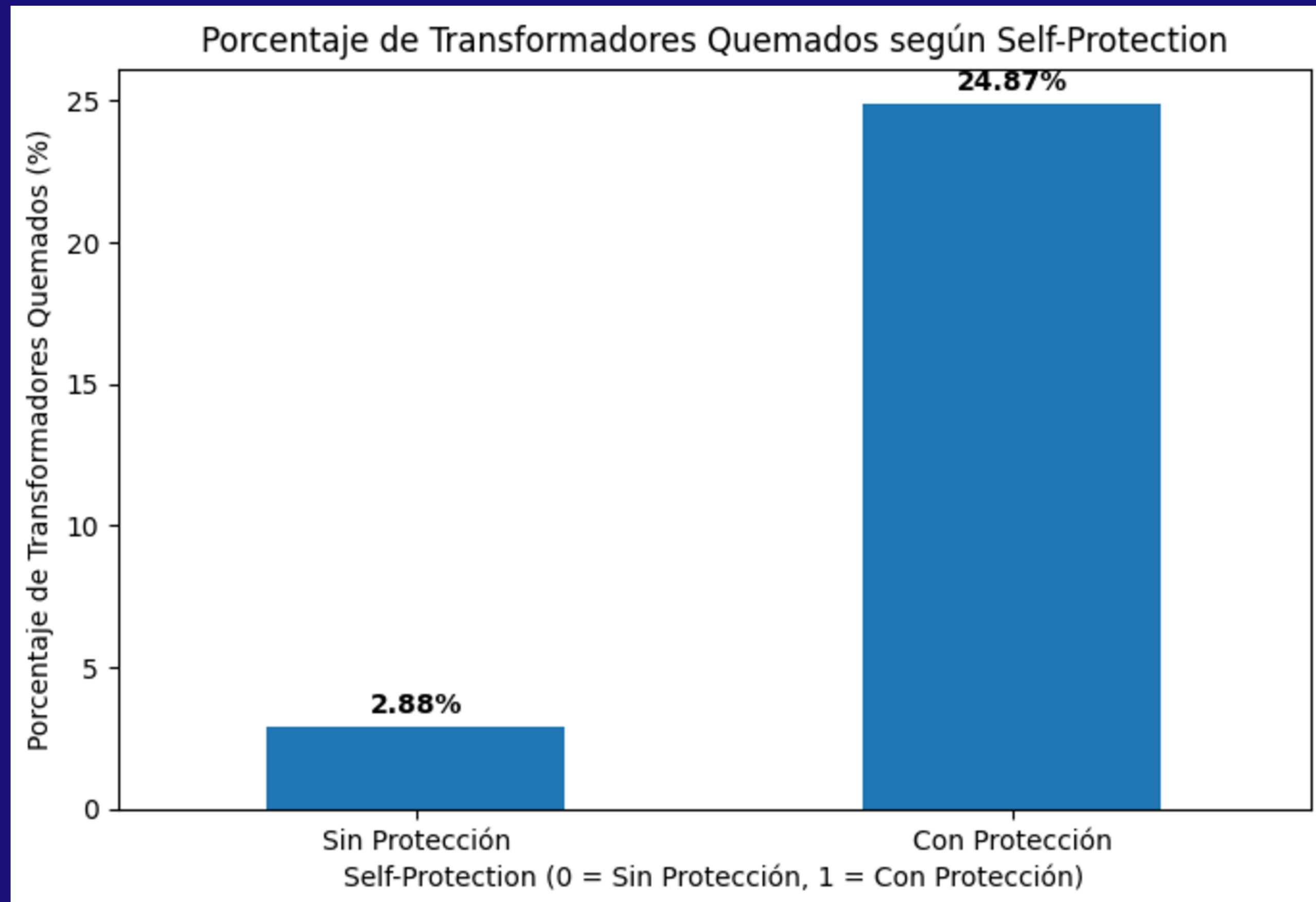
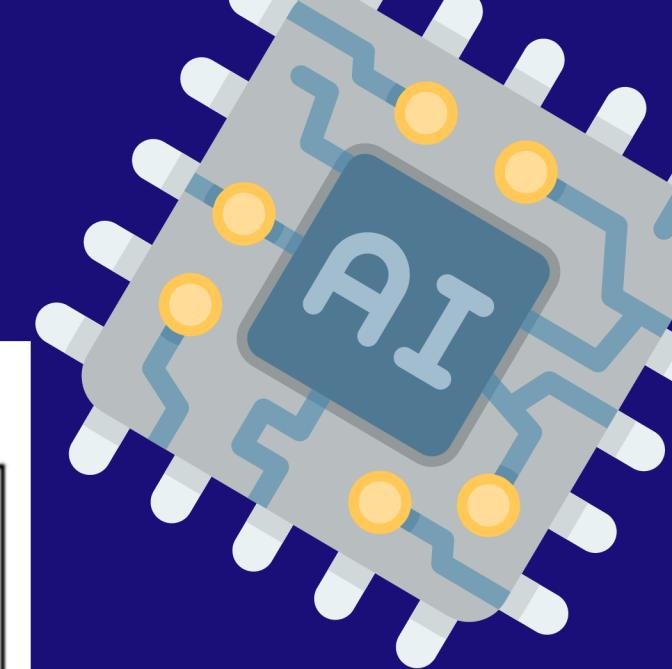
TIPO DE INSTALACION



TIPO DE CLIENTE



SELF-PROTECTION



Modelo Utilizado

Se utilizó **Random Forest** como modelo base, debido a su capacidad para manejar datos heterogéneos y proporcionar resultados robustos en problemas de clasificación binaria. La optimización del modelo se realizó mediante **Grid Search**, ajustando hiper parámetros clave como el número de estimadores, la profundidad máxima y el criterio de división.

¿QUÉ ES RANDOM FOREST?

Random Forest es un tipo de modelo de machine learning que se utiliza para hacer predicciones y clasificaciones. La idea principal es combinar el poder de muchos modelos sencillos (llamados árboles de decisión) para crear un modelo mucho más robusto y preciso.

¿QUÉ ES SMOTE?

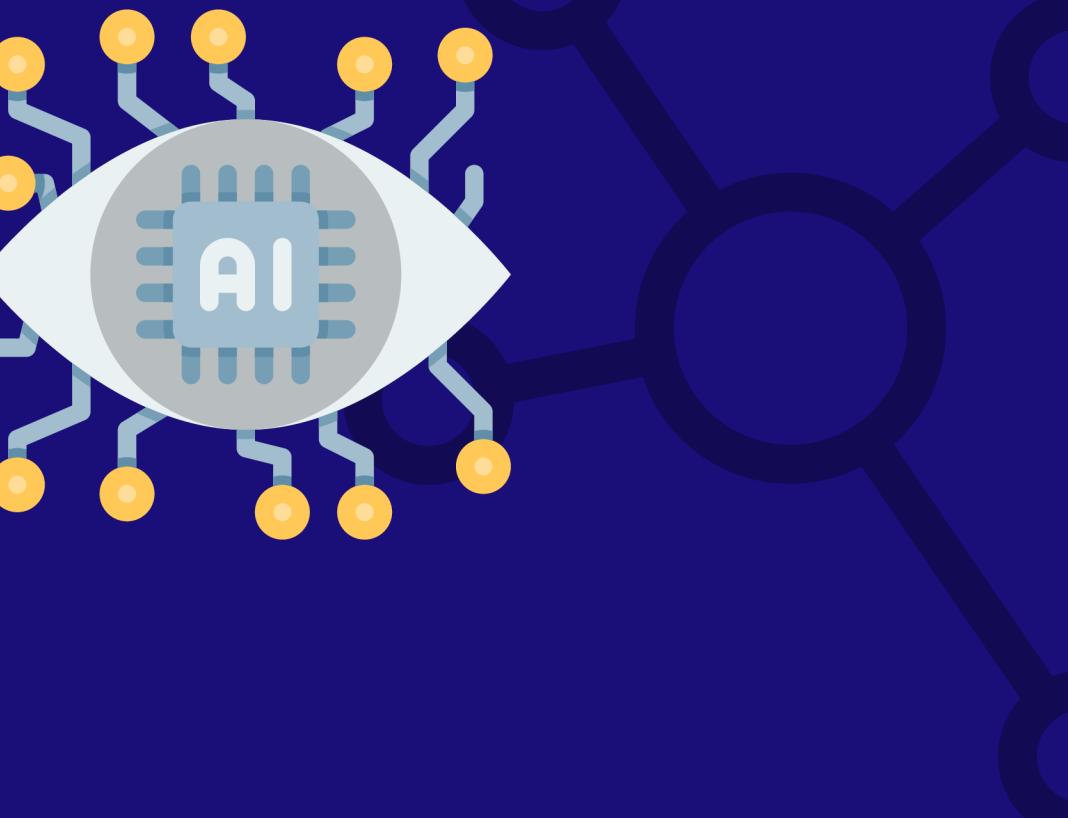
SMOTE (Synthetic Minority Over-sampling Technique) es una técnica que genera ejemplos sintéticos de la clase minoritaria para equilibrar conjuntos de datos desbalanceados, mejorando el rendimiento del modelo.

RANDOM FOREST + SMOTE: ¿POR QUÉ JUNTOS?

Al combinar **Random Forest** con **SMOTE**, podemos crear un modelo que:

1. Aprovecha la precisión y estabilidad de Random Forest.
2. Maneja el problema de datos desequilibrados, lo que mejora su capacidad de reconocer correctamente las categorías minoritarias.

Metricas de Evaluación



Precisión (Accuracy):

Proporción de predicciones correctas sobre el total de predicciones. Es útil para evaluar el rendimiento global.

F1-Score:

Métrica que combina precisión y sensibilidad en un único valor, siendo especialmente útil en casos de clases desbalanceadas, como el presente.

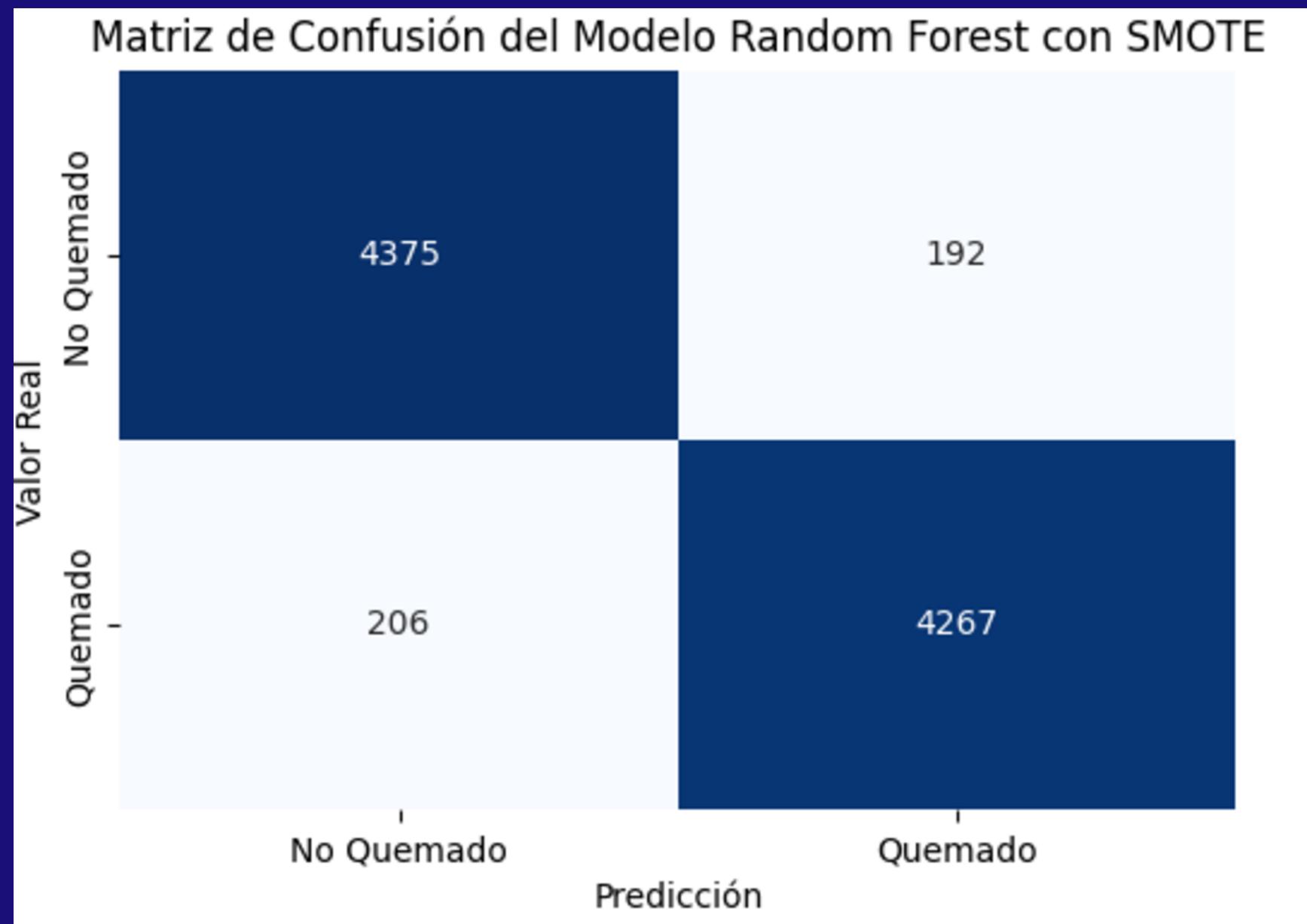
Recall (Sensibilidad):

Proporción de verdaderos positivos sobre todos los casos reales positivos. Mide la capacidad del modelo para identificar correctamente los transformadores quemados.

Matriz de Confusión:

Proporción de predicciones correctas sobre el total de predicciones. Es útil para evaluar el rendimiento global.

MATRIZ DE CONFUSIÓN



PRECISION , RECALL Y ACCURACY

Evaluación del Modelo Random Forest con SMOTE:

Accuracy: 0.96

Precision: 0.96

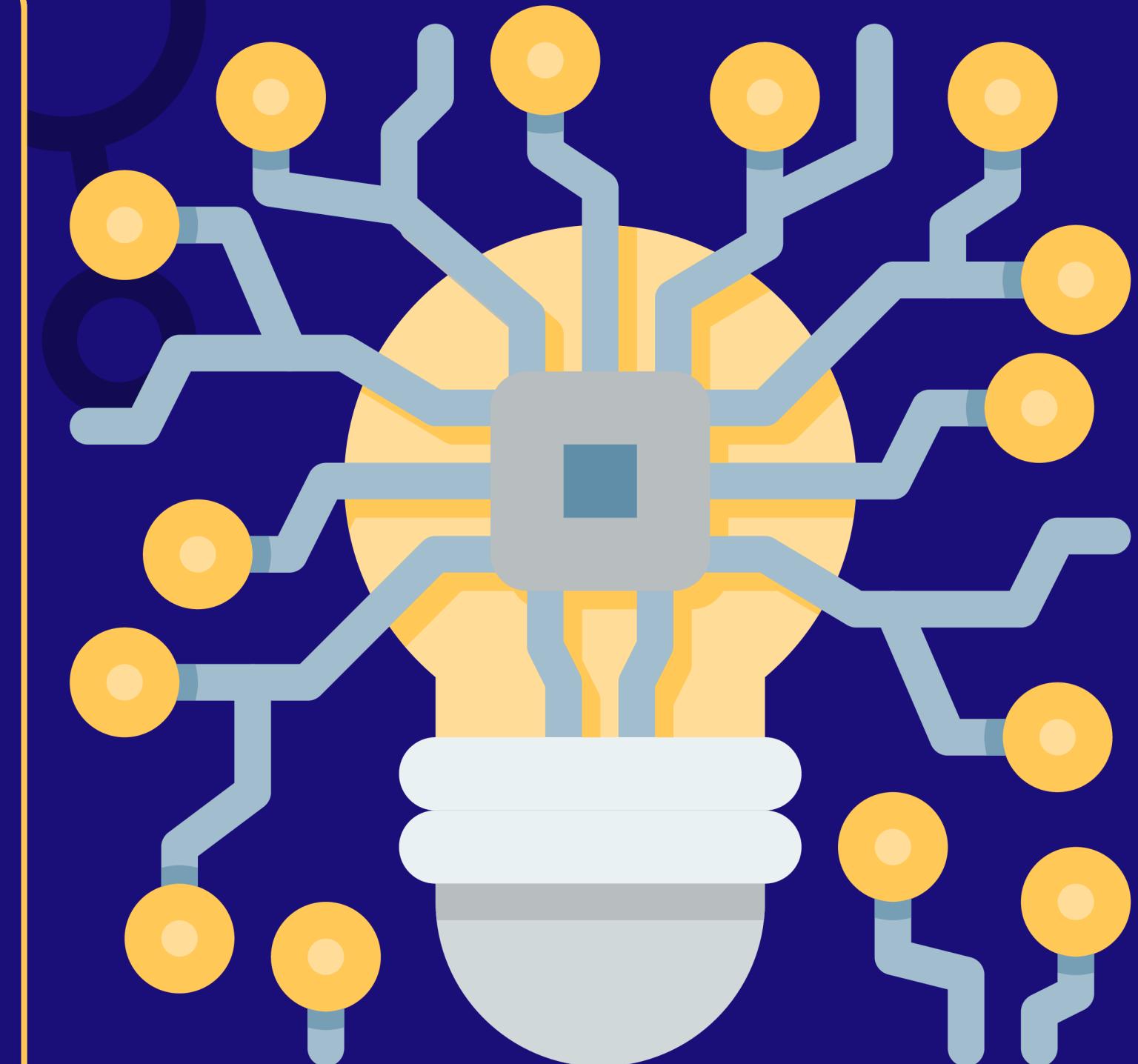
Recall: 0.95

F1-Score: 0.96

Conclusiones

El enfoque basado en Random Forest combinado con SMOTE demostró ser efectivo para predecir fallas en transformadores eléctricos en datos desbalanceados. Las principales contribuciones son:

- **Identificación de variables clave**, como capacidad del transformador y densidad de descargas atmosféricas.
- Uso de SMOTE para equilibrar clases y **mejorar la predicción**.
- Modelo optimizado con una **precisión** del **96%** y un **F1-Score elevado**.



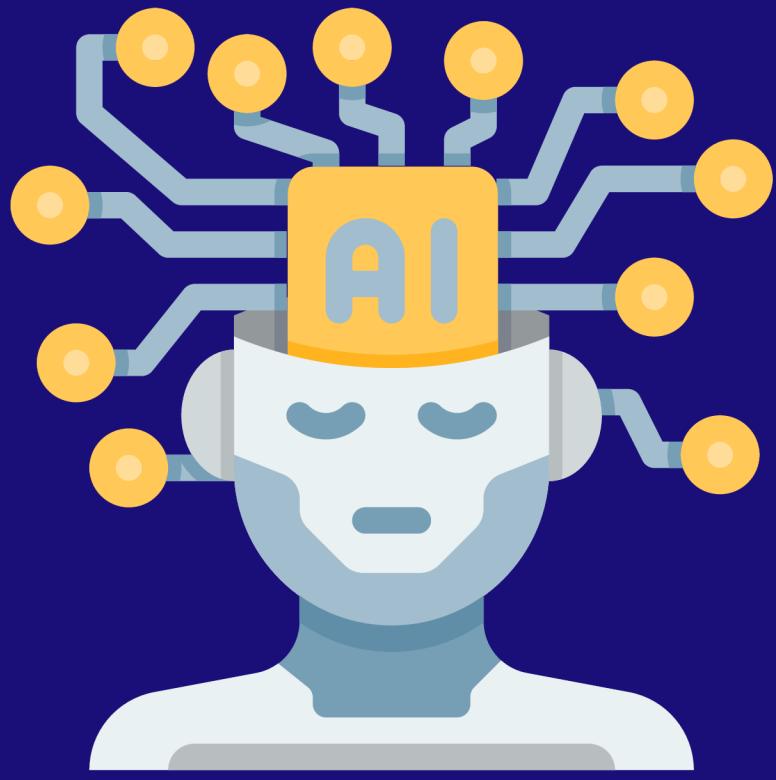
Limitaciones y Trabajo Futuro

Limitaciones:

- El análisis se basó en datos históricos estáticos de 2019, sin incorporar series temporales o forecasting, limitando la detección de tendencias a largo plazo.
- Restricción a un dataset específico de una región en Colombia, dificultando la generalización a otras áreas o condiciones operativas.
- No se incluyó el análisis cromatográfico del aceite dieléctrico, que podría aportar información clave sobre el estado interno de los transformadores.

Trabajo Futuro:

- Dataset Temporal: Diseñar un dataset de series temporales en una subestación en Paraguay para identificar patrones y aplicar forecasting.
- Análisis Cromatográfico: Incorporar datos del aceite dieléctrico para evaluar la degradación interna de los transformadores y enriquecer el modelo predictivo.



Gracias