

Universidad Nacional de Asunción



Facultad de Ingeniería

Cátedra:

Inteligencia Artificial Aplicada

Profesores:

- ❖ Ing. Diego Stalder
- ❖ Ing. Christian Torres

Tema:

***Mantenimiento Predictivo de Transformadores
Eléctricos Usando Modelos de Machine
Learning***

Integrantes:

- Colman Cespedes Robert Sebastián
- Ojeda Cueto Guillermo Diego

CI: 5178781

CI: 5354320

San Lorenzo – Paraguay

Título:

Mantenimiento Predictivo de Transformadores Eléctricos Usando Modelos de Machine Learning

Abstract:

Este trabajo presenta un enfoque basado en datos para predecir fallos en transformadores eléctricos utilizando técnicas de machine learning. Se analizó un dataset desbalanceado proporcionado por la Compañía Energética de Occidente en Colombia, que contiene variables operativas y ambientales críticas. Se desarrollaron tres características clave para enriquecer el análisis, optimizando el rendimiento del modelo mediante búsqueda de hiper parámetros y técnicas para balancear los datos. Los resultados muestran una mejora significativa en la predicción de fallos, evaluada a través de métricas como precisión y F1-score, lo que valida la viabilidad del enfoque. Este trabajo ofrece una base sólida para el diseño de sistemas predictivos en infraestructura eléctrica.

Introducción

El mantenimiento predictivo de transformadores eléctricos ha surgido como una estrategia esencial para mitigar fallas inesperadas y optimizar el desempeño de las redes eléctricas. Sin embargo, la identificación temprana de transformadores en riesgo representa un desafío, especialmente en contextos donde las fallas son eventos raros pero críticos. Este trabajo aborda esta problemática mediante el análisis de datos históricos de operación, explorando enfoques avanzados de aprendizaje automático para mejorar la capacidad de predicción y apoyar la toma de decisiones en la gestión de activos eléctricos.

Metodología

Análisis General y Enfoque

Este trabajo se centra en predecir si un transformador eléctrico se quemará o no, utilizando un enfoque de clasificación binaria. El análisis considera cómo las condiciones operativas y el contexto geográfico influyen en la probabilidad de fallo, alineándose con estudios previos que han explorado variables como la ubicación y el tipo de instalación en transformadores de distribución [1, 2].

Dataset y Preprocesamiento

El dataset utilizado fue proporcionado por la Compañía Energética de Occidente, en colaboración con la Universidad del Cauca y la Universidad del Valle, y fue previamente descrito por Bravo et al. [1]. Este incluye datos de transformadores ubicados en zonas urbanas y rurales del Departamento del Cauca, Colombia, con un total de 15,869 registros que abarcan variables categóricas y continuas. El dataset contiene un desbalance significativo, con solo el 5% de los transformadores clasificados como quemados. Para abordar este problema, se utilizó **SMOTE (Synthetic Minority Over-sampling Technique)**, que genera muestras sintéticas en la clase minoritaria para equilibrar las proporciones y mejorar la capacidad predictiva del modelo [3].

Análisis de Correlación y 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 adecuada mostraron una mayor probabilidad de fallo. Esto podría estar relacionado con un posible mal dimensionamiento o ausencia de mecanismos de protección ante sobrecargas o cortocircuitos.
2. **Tipo de Cliente:** Los transformadores que abastecen a estratos sociales bajos tienden a fallar más, posiblemente debido a un bajo mantenimiento o bajadas de líneas energéticas clandestinas
3. **Tipo de Instalación:** Transformadores ubicados en instalaciones aéreas o con estructuras inadecuadas, como postes, mostraron una mayor incidencia de fallas. Esto podría deberse a la susceptibilidad a bajadas de tensión o condiciones ambientales adversas.

Estas variables destacan la importancia de considerar el contexto operativo y la ubicación al desarrollar estrategias de mantenimiento predictivo, lo que se alinea con hallazgos previos sobre la criticidad de los factores ambientales y operativos en el desempeño de transformadores [2, 4].

Modelo Predictivo: Random Forest

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.

Configuración Óptima del Modelo

- **Número de Estimadores:** 100
- **Profundidad Máxima:** 10
- **Criterio de División:** Gini

SMOTE: Manejo del Dataset Desbalanceado

La técnica **SMOTE** fue implementada para abordar el problema de la clase minoritaria (transformadores quemados). Esta técnica genera nuevas instancias sintéticas interpolando entre puntos existentes de la clase minoritaria, lo que mejora la diversidad del dataset y reduce el sesgo hacia la clase mayoritaria [3]. Esto es crucial para garantizar que el modelo no solo prediga correctamente los transformadores que no fallan, sino también los que tienen mayor probabilidad de fallo.

Métricas de Evaluación

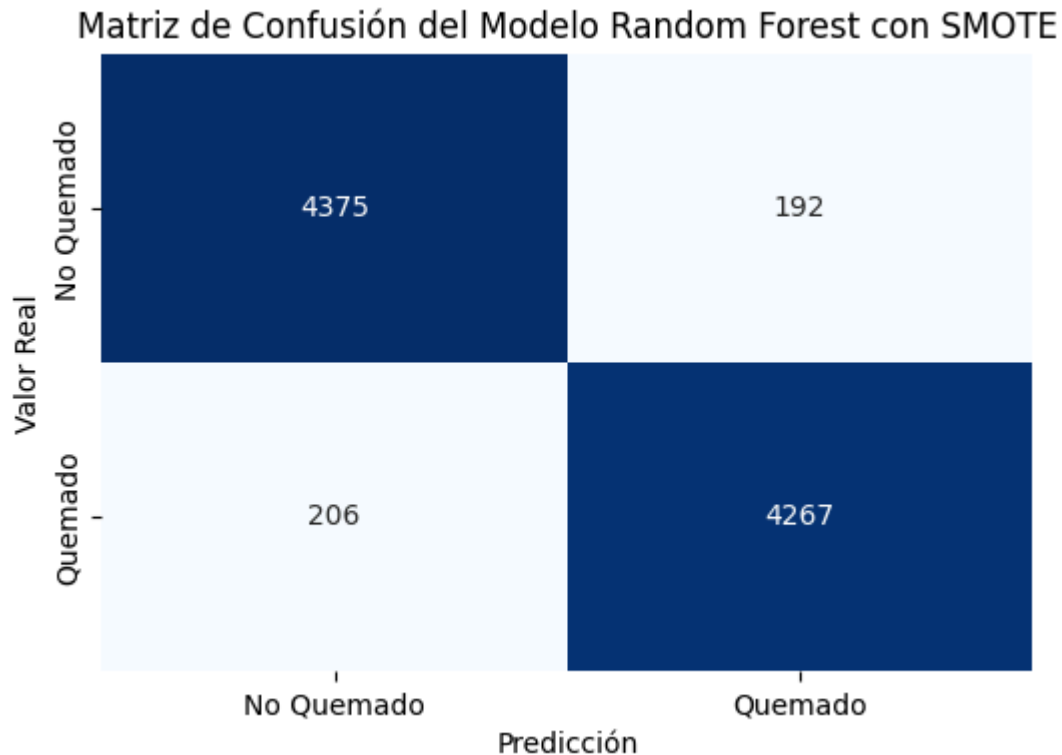
Para evaluar el desempeño del modelo, se seleccionaron métricas específicas que permiten medir la capacidad del modelo de clasificar correctamente los transformadores quemados y no quemados. Estas métricas son clave para validar el enfoque propuesto en contextos de datos desbalanceados.

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: Visualización que muestra la relación entre predicciones correctas e incorrectas, desglosando verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos.



```
Evaluación del Modelo Random Forest con SMOTE:  
Accuracy: 0.96  
Precision: 0.96  
Recall: 0.95  
F1-Score: 0.96
```

Conclusiones

Los resultados obtenidos demuestran que el enfoque basado en Random Forest combinado con SMOTE es eficaz para predecir fallas en transformadores eléctricos, incluso en contextos de datos desbalanceados.

Las principales contribuciones de este trabajo incluyen:

- La identificación de variables clave que afectan directamente las fallas, como la capacidad del transformador y la densidad de descargas atmosféricas.
- La implementación de SMOTE como herramienta para equilibrar las clases y mejorar la capacidad predictiva del modelo.
- Un modelo optimizado que alcanzó métricas de evaluación sólidas, con una precisión del 96% y un F1-Score elevado.

Limitaciones y Trabajo Futuro

Limitaciones:

El presente estudio se enfocó únicamente en el análisis de datos históricos estáticos correspondientes al año 2019, sin considerar un enfoque de series temporales o forecasting. Esto limita la capacidad del modelo para identificar tendencias a lo largo del tiempo, como patrones estacionales o degradación progresiva en los transformadores. Además, el análisis estuvo restringido a un dataset específico de una región de Colombia, lo que podría dificultar la generalización del modelo a otras regiones o condiciones operativas. Por último, no se consideraron otros métodos de análisis complementarios, como el uso de datos de cromatografía para evaluar la composición del aceite dieléctrico de los transformadores.

Trabajo Futuro:

Desarrollo de un Dataset Temporal: Como trabajo futuro, se podría diseñar un dataset de series temporales que capture mediciones continuas de variables críticas en una subestación en Paraguay. Esto permitiría aplicar técnicas de forecasting para identificar patrones de fallas en el tiempo y anticiparse a posibles interrupciones.

Análisis Cromatográfico: Otro enfoque complementario sería incorporar datos de análisis cromatográfico del aceite dieléctrico de los transformadores. Este tipo de análisis permite evaluar el estado interno del equipo y podría enriquecer el modelo predictivo con nuevas características relacionadas con la degradación química de los componentes.

Estas extensiones no solo mejorarían la precisión de las predicciones, sino que también aportarían una perspectiva más integral para la gestión del mantenimiento predictivo en transformadores eléctricos.

Referencias

1. Bravo, D.-A., Alvarez, L.-I., & Lozano, C.-A. Dataset of distribution transformers for predictive maintenance. *Data in Brief*, vol. 38, 2021. DOI: 10.1016/j.dib.2021.107454.
2. Zhou, X., Lin, J., Xu, X., Wang, Z., & Zhang, Y. Data-driven predictive maintenance for a fleet of high-speed trains: A deep learning approach. *Data in Brief*, vol. 38, 2021. DOI: 10.1016/j.dib.2021.107380.
3. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002. DOI: 10.1613/jair.953.

4. Silva, E., et al. Predictive maintenance in power distribution systems: A review and a roadmap. *Electric Power Systems Research*, vol. 189, 2020. DOI: 10.1016/j.epsr.2020.106645.