

**Modelos Predictivos de Clasificación para el Diagnóstico de Disponibilidad en Plantas
Térmicas de Generación de Energía Eléctrica de Colombia**

Elaborado por:
Juan Pablo Marín Vanegas

Universidad Ean
Escuela de Formación en Investigación
Seminario de Investigación
Bogotá
03/12/2024

Resumen

Las plantas térmicas en Colombia son esenciales para la confiabilidad del sistema eléctrico, especialmente en periodos de alta demanda o baja hidrología, pero predecir fallas e indisponibilidades sigue siendo un desafío crítico. El mantenimiento predictivo, basado en análisis exploratorios de datos y modelos clasificatorios de machine learning, permite anticipar problemas al identificar patrones y correlaciones en datos operacionales, facilitando la planificación de mantenimientos preventivos y reduciendo costos operativos y tiempos de inactividad. Estas herramientas optimizan recursos, mejoran la sostenibilidad ambiental y fortalecen la eficiencia operativa, garantizando un suministro continuo y confiable. Aunque los modelos muestran alta precisión, enfrentan retos como el desbalance de clases y la complejidad de los datos, pero su implementación posiciona a estas tecnologías como esenciales para una gestión más eficiente y resiliente de las plantas térmicas, contribuyendo al desarrollo sostenible del sector energético en Colombia.

Palabras clave: mantenimiento predictivo, generación eléctrica, machine learning, análisis de datos, disponibilidades, eficiencia energética.

Problema de Investigación

La generación de energía eléctrica a partir de plantas convencionales, incluyendo plantas térmicas a gas, carbón y de ciclo combinado, enfrenta desafíos técnicos significativos. Las fallas en sus componentes no solo afectan la estabilidad y confiabilidad de la generación, sino también la disponibilidad del recurso, la eficiencia operativa y el impacto ambiental debido al aumento en emisiones que conlleva el uso ineficiente de las máquinas (Xu, 2019)(Holloway & Kerrie, 2018). Estudios previos han identificado que las exigencias operativas intensivas y la complejidad de los sistemas mecánicos empleados aumentan la vulnerabilidad de estos sistemas a condiciones adversas, tales como variaciones de temperatura, presión y humedad (Kurien & Kumar, 2017). Estas condiciones, junto con las características técnicas propias de las plantas, tienden a producir desgastes y sobrecargas en componentes críticos, llevando a estados de inoperatividad inesperada que, si no se anticipan, afectan la continuidad del suministro eléctrico (Hashemian & Bean, 2011) (Khan et al., 2022).

A medida que la cantidad de datos generados por estos sistemas aumenta, el desafío radica en el análisis y clasificación de estos datos de manera eficiente. Los datos incluyen registros históricos de operación, cuyo volumen y complejidad requieren estrategias avanzadas para procesarlos en tiempo real. Sin embargo, muchos operadores carecen de la capacidad tecnológica y de los recursos necesarios para gestionar esta información adecuadamente y prevenir fallas, resultando en tiempos de inactividad y en incertidumbre en la disponibilidad de generación (Dhillon, 2017).

Los síntomas de este problema se manifiestan en forma de inestabilidad en la generación de energía, tiempos de inactividad no programados, y una reducción en el rendimiento operativo, lo cual también compromete la confiabilidad de la red de distribución al no contar con la plena disponibilidad del recurso. Además, los mantenimientos reactivos que se llevan a cabo tras una falla suelen ser costosos e ineficaces, lo que impacta negativamente en la eficiencia económica y en la confianza de los consumidores en el suministro (García Márquez et al., 2012).

El pronóstico de la situación es preocupante si no se adoptan medidas para anticipar los fallos: los periodos de inactividad y pérdidas operativas continuarán aumentando, afectando así tanto a la estabilidad de las redes de distribución como a los costos energéticos para los usuarios. Esta situación también implica un riesgo ambiental debido al uso ineficiente de los recursos energéticos y las emisiones adicionales de gases de efecto invernadero que estas ineficiencias generan (Guevara Carazas, 2010).

Ante esta situación, el control pronóstico propone la implementación de modelos clasificatorios de *machine learning* que puedan identificar patrones de indisponibilidades y fallos a partir de datos operacionales y predictivos (Wang et al., 2019). Los modelos de *machine learning* han demostrado su efectividad en la predicción de fallas en sistemas complejos, permitiendo a los operadores programar mantenimientos preventivos, reducir el tiempo de inactividad y categorizar la indisponibilidad del recurso (Sun et al., 2021) (Guo et al., 2024). Este enfoque permitiría optimizar el uso de los recursos, mejorar la eficiencia operativa y

disminuir el impacto ambiental, resultando en un sistema de generación de energía más confiable y menos costoso (Frootan et al., 2022).

Pregunta de Investigación

¿De qué manera la implementación de modelos clasificatorios de *machine learning* permite identificar de forma temprana los patrones de falla en plantas de generación de energía eléctrica, mejorando la continuidad, disponibilidad y eficiencia del suministro?

Objetivo General

Desarrollar un modelo clasificatorio basado en *machine learning* para la identificación de futuras indisponibilidades de plantas térmicas de generación de energía eléctrica del Sistema eléctrico colombiano.

Objetivos Específicos

1. Identificar y analizar las variables y parámetros operacionales críticos que influyen en las disponibilidades de los sistemas de generación de energía eléctrica en plantas térmicas de Colombia.
2. Implementar técnicas de procesamiento y clasificación de datos para estructurar un conjunto de datos que permita reconocer patrones y anomalías en el funcionamiento de las plantas térmicas de generación.
3. Desarrollar un modelo clasificatorio de *machine learning* capaz de predecir escenarios de indisponibilidad, utilizando datos históricos.
4. Evaluar la efectividad del modelo de *machine learning* en la predicción de indisponibilidades mediante la comparación con métodos tradicionales de mantenimiento y los registros históricos de indisponibilidades.
5. Proponer estrategias de mantenimiento preventivo basadas en las relaciones de las variables y las predicciones del modelo para identificar comportamientos.

Justificación

La investigación sobre la aplicación de modelos clasificatorios de *machine learning* en la identificación temprana de fallas en plantas de generación de energía eléctrica es de gran conveniencia, dado el impacto económico y ambiental que las fallas no controladas tienen en la industria energética. En un contexto donde la eficiencia y la sostenibilidad son prioritarias, contar con una herramienta que permita anticipar problemas críticos en la maquinaria contribuye a la reducción de los periodos de inactividad y los costos asociados a mantenimientos reactivos. Asimismo, la capacidad de optimizar el rendimiento operativo responde a la necesidad creciente de reducir el desperdicio de recursos y las emisiones contaminantes derivadas de la ineficiencia de las máquinas en condiciones subóptimas.

Desde una perspectiva de relevancia social, el proyecto se orienta a fortalecer la estabilidad y confiabilidad en el suministro de energía, lo que tiene efectos positivos en sectores clave como la industria, el comercio y los hogares. La dependencia de la sociedad moderna en un suministro continuo y eficiente de energía implica que cualquier mejora en la confiabilidad del sistema impacta positivamente en la calidad de vida, al mismo tiempo que favorece el desarrollo económico. La implementación de estos modelos predictivos puede, además, contribuir a disminuir la huella de carbono del sector energético al reducir las emisiones vinculadas a los procesos de generación y mantenimiento ineficientes.

Las implicaciones prácticas de esta investigación son significativas. Al identificar patrones de indisponibilidad de manera anticipada, los operadores pueden tomar medidas preventivas antes de que ocurran fallos mayores, mejorando la seguridad operativa y reduciendo el riesgo de paradas imprevistas. Esto permite optimizar los tiempos de mantenimiento y planificar intervenciones en momentos en que la demanda de energía sea baja, evitando así interrupciones en el servicio y garantizando un suministro más estable y rentable. Asimismo, la optimización de recursos generada por esta metodología contribuye a reducir el costo de la energía para los usuarios finales.

Desde el valor teórico, este estudio enriquece el campo de la gestión predictiva en la industria energética, integrando técnicas avanzadas de *machine learning* en un sector donde,

tradicionalmente, las intervenciones se realizan de manera reactiva. La investigación aporta conocimientos sobre la eficacia de los modelos de clasificación para el análisis de datos operacionales y su aplicabilidad en sistemas complejos de generación. Esto puede abrir nuevas vías de investigación en la aplicación de tecnologías emergentes para la gestión de infraestructuras críticas, no solo en el sector energético, sino en otros ámbitos industriales.

Finalmente, en cuanto a la utilidad metodológica, este estudio propone una metodología replicable para la implementación de modelos clasificatorios de *machine learning* en la industria energética. El enfoque propuesto en el procesamiento de datos y en la aplicación de algoritmos clasificatorios para el monitoreo en tiempo real puede ser adoptado en otras plantas de generación o en sectores donde el mantenimiento predictivo es clave para la optimización de recursos. La investigación servirá de base para futuras aplicaciones de inteligencia artificial en el mantenimiento de sistemas complejos, ampliando el uso de estas técnicas en el ámbito industrial.

Marco Teórico

En el contexto de la industria de generación de energía, el monitoreo predictivo de fallas es esencial para optimizar la eficiencia, la disponibilidad, mejorar la confiabilidad y reducir el impacto ambiental. Este enfoque, basado en la detección temprana de problemas potenciales mediante modelos de machine learning, permite una intervención oportuna, maximizando la vida útil de los equipos y minimizando tiempos de inactividad.

1. Concepto de Mantenimiento Predictivo

El mantenimiento predictivo es una estrategia proactiva que utiliza tecnologías avanzadas para monitorear el estado de los activos y predecir fallas antes de que ocurran. Esta práctica combina técnicas de análisis de datos, sensores inteligentes y modelos de aprendizaje automático para identificar patrones y anomalías que podrían indicar problemas futuros. De este modo, permite planificar intervenciones correctivas de manera eficiente y estratégica, optimizando recursos y evitando interrupciones inesperadas.

En el sector energético, caracterizado por la operación continua de equipos críticos, el mantenimiento predictivo es fundamental para garantizar la confiabilidad del sistema eléctrico. Según (García Márquez et al., 2012), esta metodología no solo mejora la eficiencia operativa, sino que también prolonga la vida útil de los equipos y minimiza los costos asociados a fallas imprevistas. Además, el enfoque predictivo contribuye a la sostenibilidad, al reducir el desperdicio de energía y emisiones asociadas con mantenimientos reactivos y paradas inesperadas.

2. Concepto de Disponibilidad

En el sector eléctrico colombiano, la obligación de reportar la indisponibilidad de las plantas de generación se encuentra regulada principalmente por las resoluciones de la Comisión de Regulación de Energía y Gas (CREG) y las disposiciones del operador del sistema eléctrico colombiano, XM. Una de las resoluciones relevantes es la Resolución CREG 025 de 1995, que regula el Reglamento de Operación del Mercado de Energía Mayorista. Esta establece las responsabilidades de los agentes del mercado, incluyendo la obligación de reportar información relacionada con las indisponibilidades programadas y no programadas de las plantas generadoras. También se destacan aspectos relacionados con la confiabilidad y la planificación del sistema.

Adicionalmente, la Resolución CREG 071 de 2006 detalla la obligación de informar los mantenimientos programados y las indisponibilidades para garantizar la seguridad y eficiencia en la operación del sistema interconectado nacional.

Por último, los agentes están sujetos a los procedimientos establecidos en el Código de Operación, que es administrado por XM. Este código especifica los detalles técnicos y operativos del reporte, incluyendo los plazos y los formatos que deben cumplirse para notificar las indisponibilidades.

De esta forma las indisponibilidades reportadas son técnicamente las fallas que las plantas de generación presentan, e imposibilitan su funcionamiento.

3. Definición de Modelos Clasificatorios en Machine Learning

Los modelos clasificatorios de machine learning, como árboles de decisión, redes neuronales, bosques aleatorios y máquinas de soporte vectorial, son fundamentales para identificar patrones de fallas. Estos modelos analizan tanto datos históricos como en tiempo real, asignando etiquetas que predicen condiciones de falla en los componentes críticos (Forootan et al., 2022; Guo et al., 2024).

Puntualmente, los árboles de decisión son útiles en aplicaciones de mantenimiento predictivo debido a su capacidad para manejar conjuntos de datos no lineales y con múltiples variables. Su estructura jerárquica facilita la interpretación de los resultados, lo que los convierte en una herramienta idónea para los sectores industriales. Según (Quinlan, 2006), los árboles de decisión son efectivos en la clasificación de datos complejos gracias a su algoritmo iterativo para dividir los conjuntos de datos en subconjuntos más pequeños basados en atributos relevantes. Además, su implementación es computacionalmente eficiente, lo que los hace ideales para entornos donde los tiempos de respuesta son críticos (Safavian & Landgrebe, 1991).

Cada modelo presenta ventajas específicas en cuanto a precisión y procesamiento de datos complejos. (Sun et al., 2021; Wang et al., 2019) destacan el uso de redes neuronales profundas para capturar patrones complejos, mientras que los bosques aleatorios y los modelos de árboles de decisión se recomiendan para datos que requieren clasificaciones rápidas y precisas.

La precisión de los modelos predictivos depende de su capacidad para analizar datos operacionales y de ambiente como temperatura, presión y vibración, permitiendo identificar condiciones anómalas (Hashemian & Bean, 2011). La integración de machine learning con el monitoreo de condición ayuda a capturar patrones complejos de desgaste y deterioro en componentes críticos.

4. Impacto Ambiental en el Sector Energético

La sostenibilidad del sector energético se enfrenta al desafío de minimizar los impactos ambientales asociados con fallas operativas, como las emisiones de gases de efecto invernadero (GEI). Estas emisiones pueden incrementarse significativamente durante interrupciones inesperadas o paradas forzadas en plantas generadoras. En este contexto, los modelos predictivos juegan un papel crucial al optimizar el rendimiento de los equipos, permitiendo una operación más eficiente y reduciendo el desperdicio de recursos.

Los modelos predictivos anticipan problemas antes de que se conviertan en fallas críticas, lo que permite un uso más eficiente de combustibles y minimiza las emisiones durante eventos no planificados (Wang et al., 2019).

En el marco regulatorio de Colombia, la sostenibilidad ambiental es un pilar clave. Tecnologías de mantenimiento predictivo y análisis de datos en tiempo real son integrales para cumplir con las metas de mitigación de emisiones definidas en las contribuciones determinadas a nivel nacional del país.

Estado del Arte

El monitoreo predictivo de fallas en plantas de generación de energía se enfoca en los avances recientes en machine learning aplicados a la identificación y predicción de fallas en entornos operativos complejos. A continuación, se detallan los estudios clave en cada área.

1. Fallas en Plantas de Generación En las plantas de generación de energía, las fallas más comunes incluyen desgaste mecánico, sobrecalentamiento y fluctuaciones en variables ambientales. Estas condiciones contribuyen a una reducción en la confiabilidad y eficiencia del sistema, además de aumentar las emisiones contaminantes. (Bangert, 2020) y (Dhillon, 2017) destacan cómo el desgaste en componentes como turbinas y generadores afecta la operación y enfatizan la necesidad de herramientas de diagnóstico preventivo en condiciones de alta demanda.

2. Monitoreo de Condición y Mantenimiento Predictivo El monitoreo de condición es fundamental para la predicción de fallas y el mantenimiento de los equipos. La evolución hacia un mantenimiento predictivo ha sido impulsada por el avance en técnicas de machine learning y el uso de sensores. (García Márquez et al., 2012) describen cómo los datos en tiempo real, obtenidos mediante sensores, permiten detectar anomalías en componentes críticos, mientras que (Kurien & Kumar, 2017) (Angelopoulos & Kontakou, 2020) exploran el uso de algoritmos de machine learning para anticipar fallas, optimizando así los intervalos de mantenimiento y recursos.

3. Teoría del Mantenimiento Predictivo La teoría de mantenimiento predictivo sostiene que mediante el monitoreo continuo de los datos operacionales es posible detectar patrones de desgaste antes de una falla grave. Con la incorporación de machine learning, este proceso se ha automatizado, permitiendo alertas más tempranas y precisas (Dhillon, 2017; Hashemian & Bean, 2011).

4. Modelos de Machine Learning en el Mantenimiento Predictivo Diversos algoritmos de machine learning se han aplicado en el mantenimiento predictivo, mostrando eficacia en la predicción de fallas en sistemas de generación de energía. Los estudios de (Guo et al., 2024) y (Zhao et al., 2019) destacan el uso de redes neuronales profundas para analizar grandes volúmenes de datos operacionales. (Wang et al., 2019) explora los bosques aleatorios, árboles de decisión y máquinas de soporte vectorial para clasificar estados de fallo, resaltando su precisión en la predicción y su capacidad de generalizar en datos complejos. (Al Shorman et al., 2020) muestra cómo la implementación de modelos predictivos reduce los costos operativos al minimizar tiempos de inactividad no planificados.

5. Impacto Ambiental y Eficiencia en Plantas de Energía Convencional La eficiencia en plantas de generación está directamente relacionada con su impacto ambiental. Las fallas operativas aumentan las emisiones y el uso de recursos, mientras que una operación más eficiente contribuye a la sostenibilidad del sector. (Wang et al., 2019) destacan la importancia de reducir interrupciones para minimizar el impacto

ambiental, mostrando que el mantenimiento predictivo contribuye a una operación más sostenible.

6. Impacto Teórico de la Predictibilidad en Sistemas Complejos En sistemas complejos como las plantas de generación, la predictibilidad depende de la calidad de los datos y de la capacidad de los modelos para identificar patrones en entornos con alta variabilidad. (Angelopoulos & Kontakou, 2020) analizan cómo la teoría de la complejidad y el uso de big data pueden mejorar la predicción en sistemas industriales, permitiendo una mayor adaptabilidad y precisión en la toma de decisiones.

7. Modelos de Monitoreo y Diagnóstico de Condición Los modelos de monitoreo de condición utilizan datos en tiempo real para diagnosticar el estado de los equipos. (Kurien & Kumar, 2017) subrayan la utilidad de los análisis de series temporales y algoritmos de detección de anomalías, que ofrecen una base para el mantenimiento predictivo mediante el monitoreo de variables clave como vibración, temperatura y presión.

Metodología

Enfoque, Alcance y Diseño de la Investigación

El enfoque de investigación será cuantitativo, orientado al desarrollo de un modelo de machine learning que permita anticipar la indisponibilidad de la planta en función de variables operacionales y ambientales. Este estudio se enmarca en un diseño no experimental y transversal, con un alcance predictivo y aplicado para la intervención organizacional en plantas térmicas de Colombia. Se trabajará exclusivamente con plantas térmicas debido a sus características técnicas, lo que limita la replicabilidad del modelo en plantas hidráulicas.

Definición de Variables

1. Código

- **Definición Conceptual:** Código de identificación con el cual se referencia a la planta de generación térmica.

- **Definición Operacional:** Es un código de 4 caracteres alfanumérico.
- 2. Fecha**
 - **Definición Conceptual:** Es la fecha de medición de cada una de las demás variables.
 - **Definición Operacional:** Definida como año-mes-día.
- 3. Combustible**
 - **Definición Conceptual:** Tipo de combustible térmico que consume la planta para realizar su operación mecánica de generación de energía.
 - **Definición Operacional:** Diferenciada entre gas y carbón.
- 4. Recurso**
 - **Definición Conceptual:** Nombre con el cual está registrada la planta de generación por parte del agente ante el mercado y sistema.
 - **Definición Operacional:** Nombre comercial de la planta.
- 5. Fecha de Entrada**
 - **Definición Conceptual:** Fecha de entrada en operación de la planta de generación.
 - **Definición Operacional:** Medido en año de entrada.
- 6. Potencia MWh-día**
 - **Definición Conceptual:** Potencia máxima de generación de la planta en 24 horas.
 - **Definición Operacional:** Medido en MWh-día.
- 7. Potencia MWh-mes**
 - **Definición Conceptual:** Potencia máxima de generación de la planta en el mes.
 - **Definición Operacional:** Medido en MWh-mes.
- 8. Disponibilidad declarada por el agente en el mes**
 - **Definición Conceptual:** Los agentes titulares de las plantas de generación indican e informan al sistema su indisponibilidad en el mes.
 - **Definición Operacional:** Medido MWh-mes.

9. Indisponibilidad

- **Definición Conceptual:** Conforme la potencia y la disponibilidad reportada se logra identificar su capacidad de generación.
- **Definición Operacional:** Medido en porcentaje de su disponibilidad reportada y su disponibilidad real.

10. Generación MWh-mes

- **Definición Conceptual:** Generación real en el mes de la planta de generación.
- **Definición Operacional:** Medida en GWh-mes.

11. Precio de Bolsa Promedio Mes

- **Definición Conceptual:** Es el precio promedio de la energía en el mercado eléctrico colombiano.
- **Definición Operacional:** Medido en \$/kWh promedio mes

12. Clasificación de indisponibilidad

- **Definición Conceptual:** Rango de indisponibilidad reportado conforme su potencia esperada para el mes.
- **Definición Operacional:** Categorías de Alta, Moderada y Baja.

Considerando las variables Indisponibilidad y Potencia MWh-mes se ha categorizado el porcentaje de indisponibilidad de cada Recurso para cada Fecha donde de 0 a 0.70 se considera Alta, de 0.71 a 0.90 es Moderada y 0.91 a 1 es Baja, posteriormente se define la variable objetivo a través de la clasificación de las clases Baja y Moderada como 0 y la clase Alta como 1.

Población y Muestra

La población se compone de las 36 plantas de generación térmica de Colombia despachadas centralmente que usan como combustibles gas o carbón, indicando su capacidad instalada para cada mes de cada año, su fecha de entrada en operación, la sumatoria de la generación de cada mes en cada año, la sumatoria de la indisponibilidad reportada de cada mes en cada año y el precio de bolsa promedio de cada mes en cada año, así mismo se identifica el nombre del recurso. El modelo es entrenado y testado con una muestra de 4.300

registros que abarcan el período de 2010-01-01 hasta el 2024-06-30 con el objetivo de clasificar la indisponibilidad del recurso.

Estos datos son públicos y suministrados por XM, el operador del sistema eléctrico colombiano. Por lo anterior, no se requiere la intervención de otros instrumentos de recopilación debido a la disponibilidad y confiabilidad de los datos oficiales y públicos.

Técnicas de Análisis de Datos

El análisis de datos puede abordarse desde diferentes perspectivas. En estudios cuantitativos, el Análisis Exploratorio de Datos (EDA) y los Árboles de Decisión son herramientas clave, proporcionando una base sólida para comprender patrones, relaciones y realizar clasificaciones basadas en datos.

1. Análisis Exploratorio de Datos

Es un paso fundamental y primordial en cualquier análisis de datos porque permite conocer las características principales de la muestra de datos, como distribuciones, correlaciones, segmentaciones, estacionalidades y anomalías, antes de aplicar modelos complejos. Esto asegura que las decisiones posteriores se basen en datos limpios y relevantes.

Además, las herramientas gráficas como histogramas, diagramas de dispersión y diagramas de caja resumen información compleja de manera accesible, ayudando a detectar patrones no obvios. De esta forma, la exploración permite identificar las transformaciones necesarias, como la normalización o la codificación de variables, mejorando la calidad de los datos para análisis avanzados.

En el contexto de un estudio cuantitativo, se asegura que los datos estén libres de sesgos y errores, proporcionando una base sólida para análisis posteriores.

2. Árboles de Decisión

Los Árboles de Decisión son una herramienta predictiva clave elegida por su capacidad de interpretación, representando visualmente el proceso de toma de decisiones, lo que facilita la comunicación de hallazgos a audiencias no técnicas.

Una ventaja de este método es su flexibilidad para manejar tanto datos categóricos como continuos, adaptándose a una variedad de problemas analíticos que permiten dividir los

datos basándose en variables explicativas, destacando aquellas con mayor impacto en la variable objetivo. Esto es esencial para estudios que buscan priorizar indicadores o características.

En estudios predictivos, los Árboles de Decisión no solo clasifican con precisión, sino que también identifican qué variables son más influyentes, guiando decisiones estratégicas basadas en datos.

Análisis y Discusión de los Resultados

Desde la instalación de la primera planta térmica en 1963, el Sistema Interconectado Nacional (SIN) de Colombia ha experimentado un crecimiento constante con la entrada de nuevas plantas térmicas, fortaleciendo la confiabilidad y diversificación de la matriz energética del país. Estas plantas, al proporcionar energía en firme mediante combustibles almacenables como carbón y gas, son esenciales para garantizar el suministro durante escenarios críticos, como períodos de baja hidrología.

Sin embargo, identificar y gestionar las indisponibilidades de las plantas térmicas es crucial para mantener la estabilidad del sistema. Estas indisponibilidades, que pueden surgir por fallas técnicas, mantenimiento programado o limitaciones en el suministro de combustibles, representan un desafío significativo para la planeación operativa del SIN. Una gestión efectiva de estas interrupciones permite no solo minimizar riesgos en el suministro, sino también optimizar el despacho de generación para atender la demanda en todo momento.

Por tanto, implementar herramientas analíticas avanzadas, como modelos de machine learning, para predecir indisponibilidades y planificar con antelación, es una estrategia clave. Esto no solo mejora la confiabilidad del sistema, sino que también optimiza los costos operativos y fortalece la capacidad de respuesta del SIN frente a variaciones inesperadas en la generación y la demanda.

Figura 1

Crecimiento Capacidad Instalada Térmica en Colombia

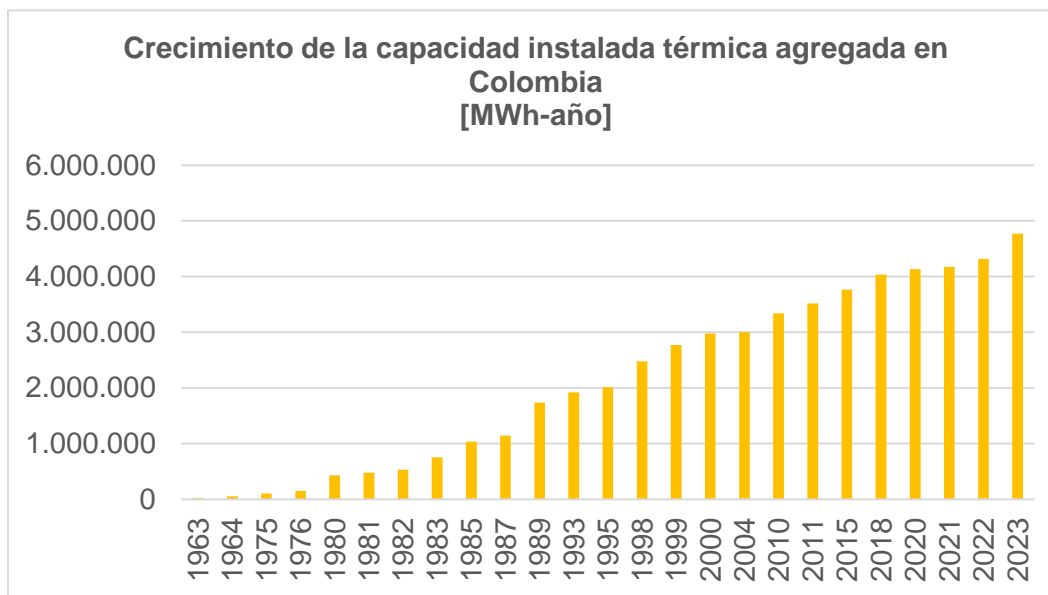
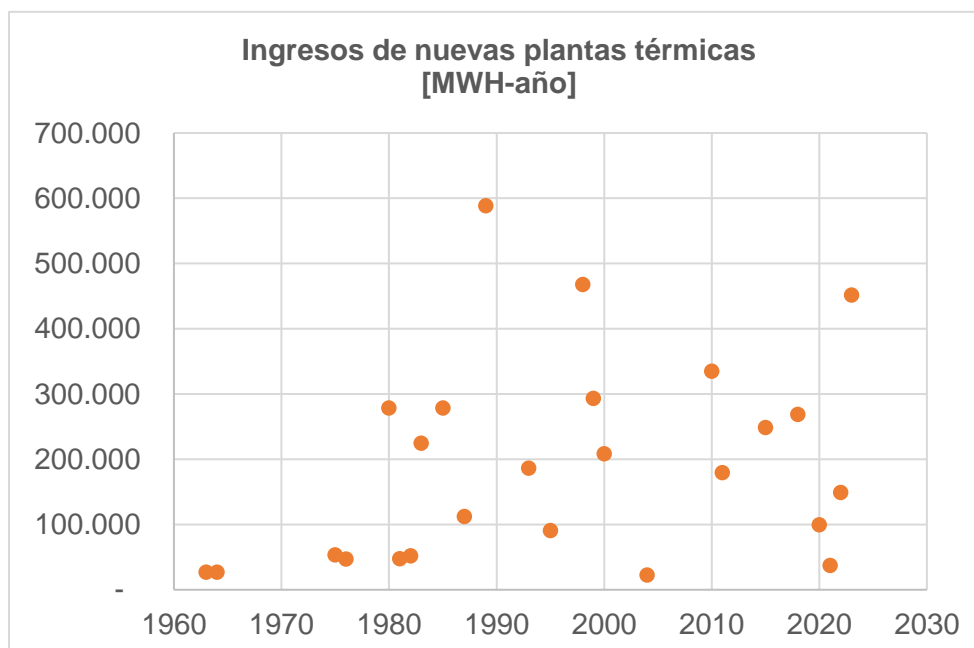


Figura 2

Entrada de Nuevas Plantas Térmicas al Parque de Generación



Esta diversificación se evidencia en la configuración de las plantas térmicas de generación a carbón, gas y ciclo combinado que actualmente operan en el Sistema Interconectado Nacional (SIN). Gracias a esta amplia gama de tecnologías, el sistema no solo cuenta con una mayor flexibilidad, sino que también ha alcanzado su capacidad térmica instalada más alta en la historia. Este hito refleja el compromiso del sector energético con garantizar un suministro confiable, incluso en escenarios de alta exigencia.

La capacidad térmica instalada, en sus máximos históricos, no solo fortalece la resiliencia del sistema frente a variaciones en la oferta y la demanda, sino que también destaca la importancia de identificar y gestionar oportunamente las indisponibilidades en estas plantas. Al mantener operativas estas unidades clave, se asegura la disponibilidad de energía en firme, necesaria para complementar la generación hidráulica predominante en Colombia y atender la demanda energética del país en cualquier circunstancia.

Figura 3

Plantas Térmicas por Tipo de Combustible

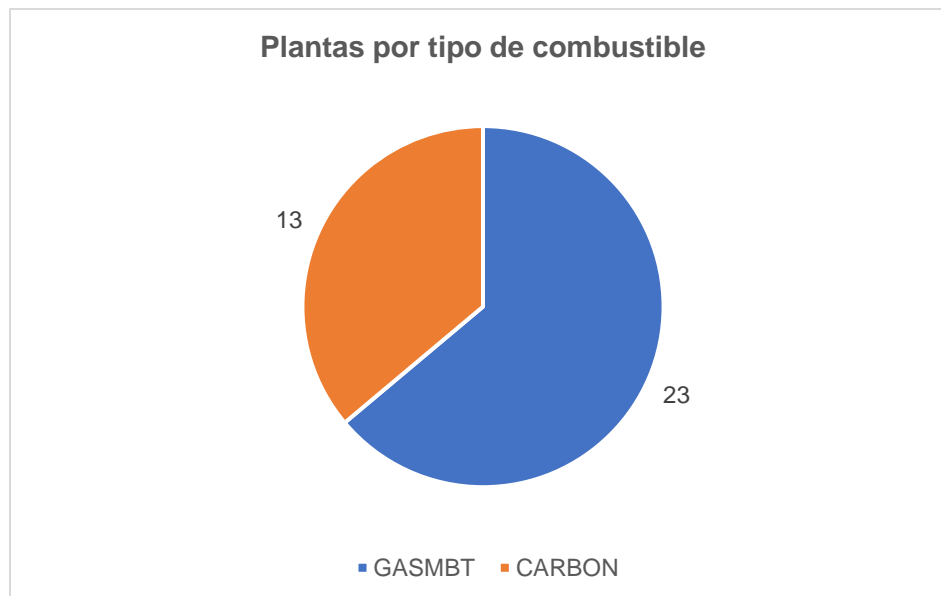
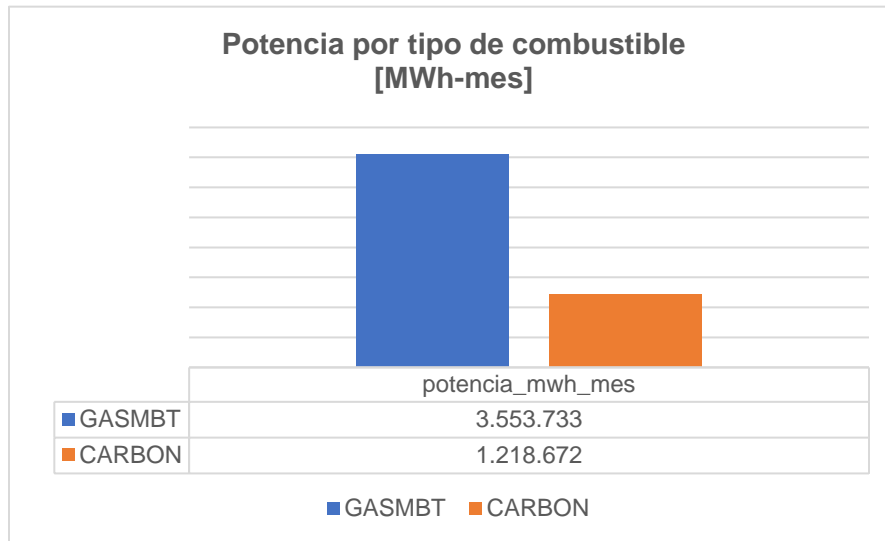


Figura 4

Capacidad Instalada por Tipo de Combustible



El análisis de la disponibilidad térmica en Colombia entre 2010 y 2024 revela tendencias significativas sobre el desempeño de las plantas de generación a carbón, gas y ciclo combinado en el Sistema Interconectado Nacional (SIN). A continuación, se destacan los principales hallazgos:

Figura 5

Indisponibilidad Reportada por Planta Térmica a Carbón

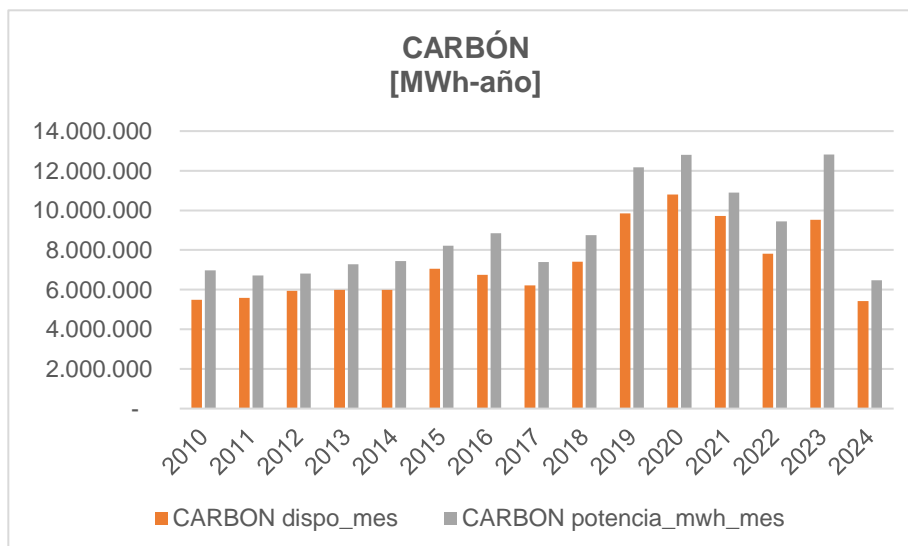


Figura 6

Indisponibilidad Reportada por Planta Térmica a Gas

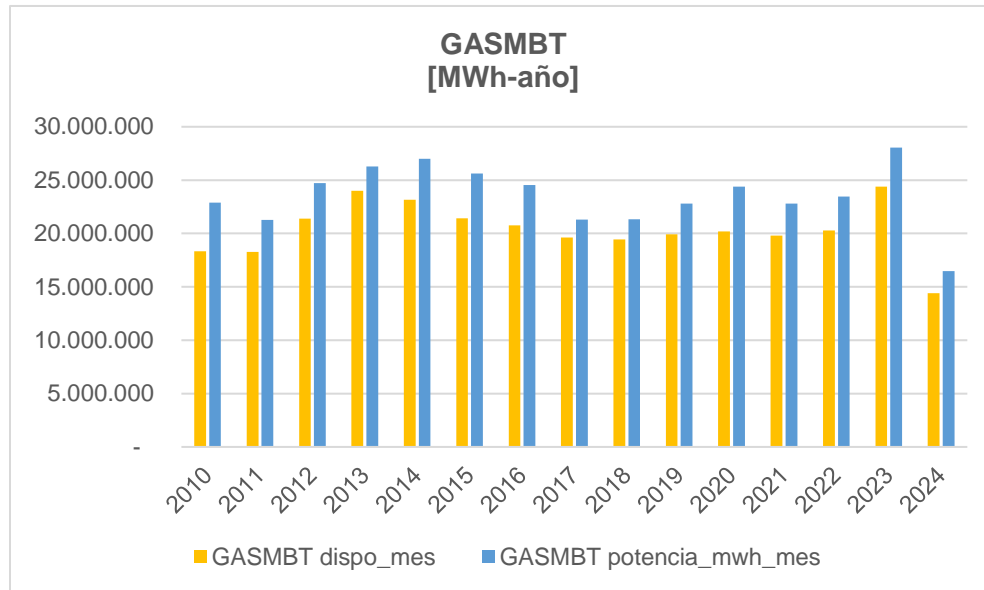


Figura 7

Indisponibilidad Reportada por Planta Térmica

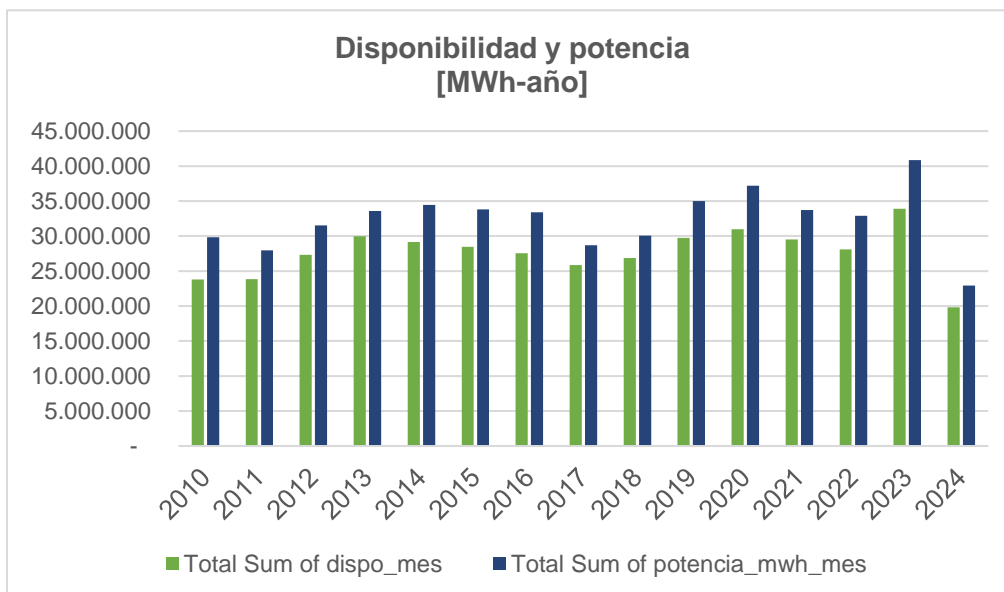


Tabla 1

Porcentaje de Indisponibilidad por Tipo de Tecnología

Año	Disponibilidad Carbón	Disponibilidad Gasmbt	Disponibilidad Térmica
2010	78,82%	80,05%	79,76%
2011	83,30%	85,89%	85,27%
2012	87,27%	86,51%	86,67%
2013	82,16%	91,34%	89,35%
2014	80,43%	85,73%	84,58%
2015	85,87%	83,62%	84,17%
2016	76,27%	84,64%	82,42%
2017	84,02%	92,09%	90,01%
2018	84,61%	91,25%	89,32%
2019	80,82%	87,29%	85,04%
2020	84,47%	82,72%	83,32%
2021	89,25%	86,79%	87,59%
2022	82,72%	86,43%	85,37%
2023	74,32%	86,94%	82,98%
2024	83,94%	87,44%	86,46%
Total	82,36%	86,52%	85,38%

Tendencias generales

1. Disponibilidad promedio general

- La disponibilidad térmica total promedia un 85,38%, lo que indica que el sistema logra mantener operativa una porción considerable de su capacidad térmica instalada.
- Las plantas de gas presentan la mayor disponibilidad promedio (86,52%), seguidas por las de carbón (82,36%).

2. Evolución anual

- Los años 2012 y 2013 destacan como períodos de alta disponibilidad térmica, alcanzando valores de 86,67% y 89,35%, respectivamente. Esto puede reflejar mejoras en mantenimiento, operación o en la gestión de los suministros.
- En contraste, 2016 y 2023 muestran niveles de disponibilidad más bajos (82,42% y 82,98%, respectivamente), lo que puede estar relacionado con eventos puntuales, problemas técnicos o dificultades en el suministro de combustibles.

Comparación por tipo de tecnología

a. Plantas a carbón

- Aunque presentan la menor disponibilidad promedio (82,36%), lograron su mejor desempeño en 2021 con un 89,25%, probablemente gracias a una gestión eficiente del mantenimiento y suministro.
- Sin embargo, su nivel más bajo se registra en 2023 (74,32%), reflejando posibles desafíos operativos en ese año.

b. Plantas a gas

- Estas plantas muestran la mayor disponibilidad promedio, destacándose especialmente en 2017 y 2018, con valores de 92,09% y 91,25%, respectivamente.
- Aunque su desempeño es consistente, su nivel más bajo se registra en 2020 (82,72%), posiblemente debido a restricciones en el suministro de gas o interrupciones no programadas.

Implicaciones operativas

1. Confiabilidad del sistema

- La alta disponibilidad térmica en promedio refuerza el papel estratégico de las plantas térmicas como respaldo en momentos de baja hidrología o alta demanda.
- Sin embargo, la variabilidad entre años y tecnologías enfatiza la necesidad de una gestión diferenciada y adaptativa para maximizar su aporte.

2. Gestión de mantenimientos

- Los datos sugieren que los períodos de alta disponibilidad coinciden con años de mejoras operativas y posiblemente inversiones en mantenimiento preventivo. Es crucial priorizar estas estrategias, especialmente en tecnologías con menor desempeño promedio, como el carbón.

3. Diversificación energética

- Aunque las plantas térmicas han mostrado una disponibilidad promedio alta, es necesario complementar su operación con fuentes renovables para mitigar riesgos asociados a interrupciones de suministro o fluctuaciones en precios de combustibles fósiles.

Al analizar los resultados de la correlación entre la indisponibilidad agregada reportada de las plantas térmicas por mes y el precio de bolsa promedio del mes para cada uno de los años de la muestra, se puede observar una serie de patrones y puntos clave que se pueden discutir. El valor de R^2 varía considerablemente a lo largo de los años, lo que indica que la relación entre ambas variables no es constante, y hay ciertos factores que parecen influir en la variabilidad de los resultados.

Tabla 2

Resultados de Correlación entre Indisponibilidad – Precio de Bolsa Promedio Mes

Año	R2
2010	0,0314
2011	0,1521
2012	0,0348
2013	0,0551
2014	0,035
2015	0,1534
2016	0,4253
2017	0,4084
2018	0,0673
2019	0,2186
2020	0,5785
2021	0,2254
2022	0,5218
2023	0,0055
2024	0,5782

Puntos clave

1. Bajos valores de R^2 en ciertos años: En años como 2010, 2012, 2013, 2014, 2015, 2018, 2023 y algunos otros, el valor de R^2 es bajo (por debajo de 0.15 en varios de estos años), lo que indica que la relación entre la indisponibilidad y el precio de bolsa promedio es débil. Esto sugiere que, en esos años, las fluctuaciones en los precios no fueron bien explicadas por la variabilidad de la indisponibilidad de la planta, o que hubo otros factores que influyeron más en los precios de bolsa.

2. Incremento en R^2 a partir de 2016: A partir de 2016, se observa un aumento notable en el valor de R^2 , alcanzando un máximo de 0.5785 en 2020. Este es un valor moderado que sugiere que, en estos años, la indisponibilidad de la planta tiene una relación más fuerte con el precio de bolsa promedio, y el modelo explica una proporción considerable de la variabilidad en los precios. Este aumento puede estar asociado con eventos específicos del mercado eléctrico o situaciones de generación de energía en esos años que hicieron que la relación fuera más directa.

3. Valor consistente de R^2 en 2024: El valor de 2024 (0.5782) es similar al de 2020, lo que indica que la relación observada en este año sigue siendo fuerte. Esto podría indicar que las condiciones que favorecieron una relación más estrecha entre la indisponibilidad y los precios en 2020 siguen siendo relevantes este año.

Posibles factores que afectan la correlación

Los valores de indisponibilidad y precio de bolsa pueden estar influenciados por varios factores externos como: política energética, cambios en la oferta y demanda de energía, fluctuaciones macroeconómicas, condiciones climáticas que afecten la disponibilidad de las plantas, entre otros.

En algunos años, es posible que la indisponibilidad de las plantas haya tenido un impacto limitado en los precios debido a que otras variables (por ejemplo, generación por fuentes alternativas o importaciones de energía) jugaron un papel más relevante en la determinación de los precios.

Discusión

Variabilidad en los resultados: El análisis muestra una gran variabilidad en la relación entre la indisponibilidad y el precio de bolsa a lo largo del tiempo. Esto puede deberse a factores de mercado específicos o a cambios en la estructura del sistema eléctrico que afectaron la dinámica entre estas dos variables.

Años con mayor ajuste: Los años con valores de R^2 más altos (2016-2024) son aquellos en los que la relación entre la indisponibilidad y los precios parece ser más

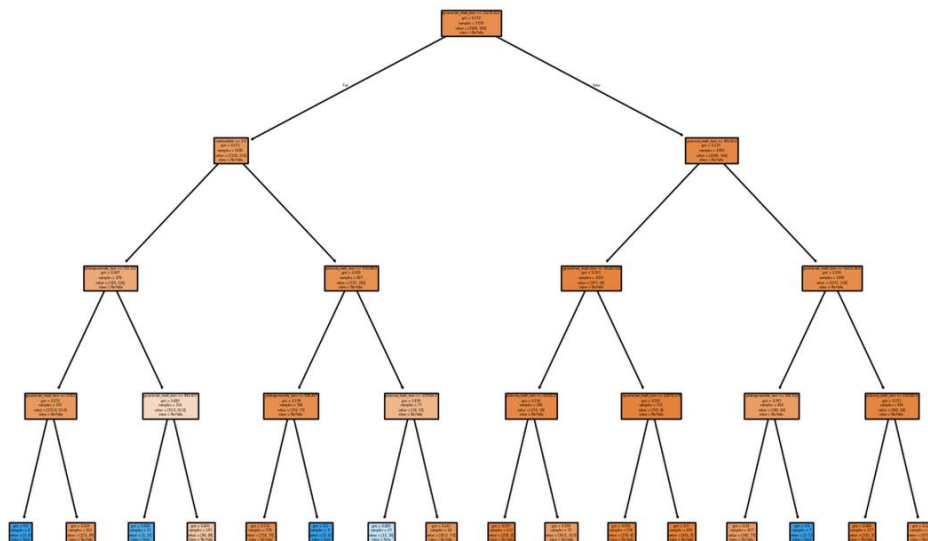
significativa, lo cual podría reflejar un mayor impacto de la generación térmica en la determinación de los precios en esos años. En estos casos, la indisponibilidad de plantas térmicas podría haber tenido una relación más directa con el precio de bolsa debido a una menor capacidad disponible para generar electricidad.

Bajos valores de R^2 : En los años con valores más bajos, se podría investigar si existen factores adicionales que puedan explicar mejor los precios de bolsa, como por ejemplo cambios en la demanda, políticas gubernamentales, eventos imprevistos (desastres naturales, problemas en el sistema interconectado), o incluso cambios en la oferta de energía desde fuentes renovables o importadas.

Métricas de Desempeño del Árbol de Decisión

Figura 8

Árbol de Decisión



Precisión

- **Entrenamiento:** 0.9118
- **Prueba:** 0.8441

El modelo presenta un buen nivel de precisión en ambas fases. El rendimiento en entrenamiento (91.18%) es superior al de prueba (84.41%), lo que podría indicar cierto nivel de sobreajuste. Sin embargo, la diferencia no es excesiva, lo que sugiere un modelo razonablemente generalizable.

Matriz de Confusión

Tabla 3

Matriz de Confusión Sobre la Variable Objetivo

Verdadero	Predicho 0	Predicho 1
Clase 0	448	53
Clase 1	36	34

- **Clase 0:** El modelo clasifica correctamente el 89.4% (448/501) de los casos. La mayoría de los errores provienen de falsos positivos (53 casos).
- **Clase 1:** La capacidad del modelo para identificar fallos (Clase 1) es limitada, con un recall de 0.49, indicando que aproximadamente la mitad de los fallos reales son detectados.

Reporte de clasificación

Tabla 4

Evaluación del Modelo

Métrica	Clase 0	Clase 1	Promedio Macro	Promedio Ponderado
Precisión	0.93	0.39	0.66	0.86
Recall	0.89	0.49	0.69	0.84
F1-Score	0.91	0.43	0.67	0.85

El modelo es altamente preciso en predecir la Clase 0 (No Falla), pero tiene dificultades para predecir correctamente la Clase 1 (Falla). Esto se refleja en la baja precisión (0.39) y F1-score (0.43) para la Clase 1. Estas limitaciones podrían estar relacionadas con el desbalance de clases de los datos.

AUC-ROC

El AUC-ROC obtenido es 0.8275, lo que indica una buena capacidad discriminativa del modelo para diferenciar entre las clases. Sin embargo, un mayor enfoque en mejorar el recall para la Clase 1 podría incrementar la utilidad del modelo en escenarios críticos.

Importancia de las Variables

El análisis de las importancias de las variables muestra los predictores más relevantes para el modelo:

Tabla 5

Importancia de las Variables del Modelo

Variable	Importancia
Generación MWh-mes	0.4451
Precio Promedio de Bolsa (Mes)	0.2455
Potencia MWh-mes	0.2159
Combustible	0.0934
Fecha Entrada	0.0000

La variable **Generación MWh-mes** tiene la mayor influencia en el modelo, seguida de **Precio Promedio de Bolsa (Mes)** y **Potencia MWh-mes**. Estas tres variables explican una proporción significativa de las decisiones del modelo. Por otro lado, la variable **Fecha Entrada** no aporta información predictiva según el modelo.

Discusión de Resultados

1. **Desempeño General:** Aunque el modelo logra un buen nivel de precisión general y un AUC-ROC razonable, su desempeño en la predicción de fallos (Clase 1) es limitado. Esto puede deberse al desbalance de clases presente en los datos, donde los casos de falla representan una proporción pequeña.

2. **Importancia de Variables:** La influencia predominante de **Generación MWh-mes, Precio Promedio de Bolsa (Mes) y Potencia MWh-mes** sugiere que estas dimensiones son críticas para el análisis. Esto subraya la necesidad de monitorear estos indicadores en tiempo real para prevenir fallos.

3. **Limitaciones:**

- **Desbalance de Clases:** La baja representación de la Clase 1 afecta negativamente el recall y la precisión en la predicción de fallos. Estrategias como sobremuestreo o ajustes en los pesos de las clases podrían mejorar estos resultados.
- **Generalización:** Aunque el modelo no muestra un sobreajuste evidente, futuros experimentos con validación cruzada podrían confirmar su robustez.

Conclusiones

Confiabilidad y rol estratégico de las plantas térmicas

Las plantas térmicas en Colombia juegan un papel crítico para garantizar la confiabilidad del sistema interconectado nacional, especialmente en periodos de baja hidrología. Con una disponibilidad promedio del 85.38%, estas instalaciones representan un respaldo clave para la estabilidad del suministro eléctrico. Sin embargo, las variaciones anuales y las diferencias de desempeño entre tecnologías (carbón y gas) destacan la necesidad de un enfoque diferenciado en la gestión y planificación de los mantenimientos.

Eficiencia operativa mediante mantenimiento predictivo

La implementación de modelos de machine learning permite una gestión proactiva y eficiente de los recursos energéticos, anticipando fallas antes de que se materialicen. Esto no

solo minimiza los costos operativos y las interrupciones imprevistas, sino que también optimiza los tiempos de mantenimiento, permitiendo planificar intervenciones en momentos de menor demanda energética. Esto es especialmente valioso en un contexto donde la demanda energética fluctúa y la capacidad de respuesta es crucial.

Impacto en la sostenibilidad ambiental

El mantenimiento predictivo contribuye significativamente a la sostenibilidad al reducir las emisiones de gases de efecto invernadero derivadas de mantenimientos reactivos y paradas no programadas. Al mejorar el desempeño operativo de las plantas térmicas, se reduce el desperdicio de recursos fósiles, reforzando el cumplimiento de metas de mitigación climática en el sector energético colombiano. Este enfoque integra las perspectivas ambiental y económica, promoviendo un desarrollo sostenible del sistema eléctrico.

Desempeño y limitaciones del modelo predictivo

Si bien el modelo desarrollado mostró un rendimiento general satisfactorio, con una precisión de 84.41% y un AUC-ROC de 0.8275, su capacidad para identificar fallas (Clase 1) fue limitada debido al desbalance de clases en los datos. La baja proporción de fallos reales en la muestra afecta métricas como el recall y la precisión. Para abordar estas limitaciones, se sugiere aplicar técnicas de balanceo de datos, como el sobremuestreo de la Clase 1, ajustes en los pesos del modelo o la integración de algoritmos especializados para clases desbalanceadas.

Relación entre indisponibilidad y precios de energía

Aunque se identificaron fluctuaciones en la correlación entre indisponibilidad y precios del mercado eléctrico, los años recientes, como 2020 y 2024, evidenciaron una relación más fuerte. Esto sugiere que las condiciones de disponibilidad de las plantas térmicas tienen un impacto significativo en la formación de precios en ciertos contextos. Sin embargo, factores externos, como políticas energéticas, variabilidad climática y disponibilidad de fuentes renovables, también deben ser considerados para interpretar esta dinámica.

Relevancia de la diversificación energética

Aunque las plantas térmicas son esenciales, la diversificación hacia fuentes renovables debe ser una prioridad para garantizar un suministro resiliente y sostenible. Estas fuentes pueden complementar la generación térmica, mitigando riesgos asociados a interrupciones en el suministro de combustibles fósiles y reduciendo la dependencia del mercado eléctrico en tecnologías menos sostenibles.

Contribución teórica y metodológica

Este estudio enriquece el campo del mantenimiento predictivo, demostrando la aplicabilidad de modelos clasificatorios de machine learning en sistemas energéticos complejos. Desde una perspectiva metodológica, proporciona una hoja de ruta replicable para abordar problemas similares en otros contextos industriales. Adicionalmente, promueve la integración de tecnologías emergentes en sectores donde la gestión predictiva es clave para mejorar la eficiencia operativa.

Referencias Bibliográficas

- Al Shorman, O., Irfan, M., Saad, N., & Zhen, D. (2020). *A Review of Artificial Intelligence Methods for Condition Monitoring and Fault Diagnosis of Rolling Element Bearings for Induction Motor*.
https://www.researchgate.net/publication/346706329_A_Review_of_Artificial_Intelligence_Methods_for_Condition_Monitoring_and_Fault_Diagnosis_of_Rolling_Element_Bearings_for_Induction_Motor
- Angelopoulos, M. K., & Kontakou, C. (2020). *Big Data Analytics for Energy Efficiency*.
https://www.researchgate.net/publication/346350792_Big_Data_Analytics_for_Energy_Efficiency
- Bangert, P. (2020). *Predictive Maintenance for Gas Turbines*.
https://www.researchgate.net/publication/338533106_Predictive_Maintenance_for_Gas_Turbines
- Dhillon, B. . (2017). *Engineering Systems Reliability, Safety, and Maintenance*.
<https://www.taylorfrancis.com/books/mono/10.1201/9781315160535/engineering-systems->

reliability-safety-maintenance-dhillon

Forootan, M., Larki, I., Zahedi, R., & Ahmadi, A. (2022). *Machine Learning and Deep Learning in Energy Systems: A Review*.

García Márquez, F. P., Tobias, A. M., Pinar Pérez, J. M., & Papaelias, M. (2012). *Condition monitoring of wind turbines: Techniques and methods*.

https://www.researchgate.net/publication/257415074_Condition_Monitoring_of_Wind_Turbines_Techniques_and_Methods

Guevara Carazas, F. J. (2010). *Risk-based decision making method for maintenance policy selection of thermal power plant equipment*.

https://www.researchgate.net/publication/223400225_Risk-based_decision_making_method_for_maintenance_policy_selection_of_thermal_power_plant_equipment

Guo, J., Yang, Y., Li, H., Dai, L., & Huang, B. (2024). *A parallel deep neural network for intelligent fault diagnosis of drilling pumps*.

<https://www.sciencedirect.com/science/article/pii/S095219762400229X>

Hashemian, H. M., & Bean, W. C. (2011). *State-of-the-Art Predictive Maintenance Techniques*.

https://www.researchgate.net/publication/273563917_State-of-the-Art_Predictive_Maintenance_Techniques

Holloway, J., & Kerrie, M. (2018). *Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review*.

Khan, N., Shahid, Z., Mansoor Alam, M., Abu Bakar Sajak, A., Mazliham, M. S., Khan, T. A., & Rizvi, S. (2022). *Energy Management Systems Using Smart Grids: An Exhaustive Parametric Comprehensive Analysis of Existing Trends, Significance, Opportunities, and Challenges*.

https://www.researchgate.net/publication/364629972_Energy_Management_Systems_Using_Smart_Grids_An_Exhaustive_Parametric_Comprehensive_Analysis_of_Existing_Trends_Significance_Opportunities_and_Challenges

Kurien, C., & Kumar, A. (2017). *Condition monitoring of systems in thermal power plant for vibration, motor signature, noise and wear debris analysis*.

https://www.researchgate.net/publication/321874388_Condition_monitoring_of_systems_in_thermal_power_plant_for_vibration_motor_signature_noise_and_wear_debris_analysis

- Quinlan, J. . R. (2006). Journal of Artificial Intelligence. *Journal of Artificial Intelligence, University of Sidney*.
- Safavian, S. ., & Landgrebe, D. (1991). *A survey of decision tree classifier methodology*.
- Sun, T., Yu, G., Gao, M., & Zhao, L. (2021). *Fault Diagnosis Methods Based on Machine Learning and its Applications for Wind Turbines: A Review*.
https://www.researchgate.net/publication/355750926_Fault_Diagnosis_Methods_Based_on_Machine_Learning_and_its_Applications_for_Wind_Turbines_A_Review
- Wang, J., Dou, Y., Wang, Z., & Jiang, D. (2019). *Multi-fault diagnosis method for wind power generation system based on recurrent neural network*.
- Xu, Z. (2019). *Plant functional diversity modulates global environmental change effects on grassland productivity*. https://www.researchgate.net/publication/336837635_Xu_et_al-2018-Journal_of_Ecology
- Zhao, Y., Li, T., Zhang, X., & C, Z. (2019). *Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future*.
-