



ESCUELA DE NEGOCIOS

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS

**DETECCIÓN DE ANOMALÍAS EN SEÑALES DE OPERACIÓN PARA
TURBINAS DE GENERACIÓN ELÉCTRICA**

Profesor

Ing. Mario Salvador Gonzáles, Ph.D.

Autor

Carlos Alfredo Ríos Mendoza

2023

RESUMEN

En la industria petrolera, la generación eléctrica desempeña un papel crítico para mantener las operaciones en funcionamiento de manera continua y segura. Las turbinas de generación eléctrica, como el modelo GE-LM2500 de combustión dual Diesel/Gas en las plantas petroleras son esenciales, pero están sujetas a un alto desgaste y variabilidad en la demanda. El mantenimiento tradicional basado en intervalos de tiempo no optimiza recursos ni previene interrupciones costosas. Por lo tanto, se plantea desarrollar un sistema de detección de anomalías basado en tecnología avanzada de adquisición y análisis de datos, utilizando el aprendizaje automático con Isolation Forest. Este análisis se estima como de carácter no supervisado, considerando que no se dispone de datos diferenciadores previos que pudieran haber sido etiquetados como una condición habitual y un atípico. Los objetivos incluyen el uso de un sistema de adquisición de datos y entrenar un modelo de aprendizaje automático para identificar patrones anómalos, anticipando fallas, reduciendo los costos de mantenimiento, maximizando así la disponibilidad de energía en un entorno crítico. Este proyecto busca revolucionar la gestión de la energía en la industria petrolera y mejorar la eficiencia operativa.

Palabras clave: Anomalía, Patrones, Mantenimiento, detección

ABSTRACT

In the oil industry, power generation plays a critical role in keeping operations running continuously and safely. Power generation turbines, such as the GE-LM2500 dual combustion Diesel/Gas model in the oil plants are essential, but they are subject to high wear and tear and variability in demand. Traditional interval-based maintenance does not optimize resources or prevent costly outages. Therefore, it is proposed to develop an anomaly detection system based on advanced data acquisition and analysis technology using machine learning with Isolation Forest. This analysis is taken as unsupervised, considering that there is no previous differentiating data available that could make it labeled as a habitual condition and an atypical one. The objectives include the use of a real-time data acquisition system and training a machine learning model to identify anomalous patterns, anticipating failures and reducing maintenance costs, thus maximizing energy availability in a critical environment. This project seeks to revolutionize energy management in the oil industry and improve operational efficiency.

Key Words: Anomaly, Patterns, Maintenance, detection

Índice de Contenidos

1. Introducción	1
2. Revisión de la Literatura	2
2.1 Detección de Anomalías	2
2.2 Categoría de detección de anomalías.....	3
3. Identificación del Objeto de Estudio.....	3
4. Planteamiento del problema	4
5. Objetivo General.....	6
6. Objetivos Específicos.....	6
7. Justificación y aplicación de la Metodología	7
7.1 Recopilación de los datos	7
7.4 Visualización de variables	10
7.5 Modelo Estadístico (Modelo de clasificación).....	11
7.5.1 Isolation Forest	11
8. Resultados.....	12
8.1 Análisis del modelo	12
8.2 Interpretación de resultados.....	15
9. . Discusión de los resultados y propuesta de solución.....	17
9.1 Implicaciones para la organización	18
9.1.1 Adopción de estrategias empresariales	20
9.1.1.1 Estrategia: Responsabilidad social y sostenibilidad.....	20
9.1.1.2 Estrategia: Innovación y mejora continua	21
10. Conclusiones y Recomendaciones.....	22
10.1 Conclusiones	22
10.2 Recomendaciones	23
11. Referencias	23
ANEXOS	26

Índice de Figuras

Figura 1. Modos de detección de anomalías con base en la disponibilidad de etiquetas en los datos. (Goldsteine & Uchida, 2016).....	3
Figura 2. Plot de varibales con datos incompletos.	8
Figura 3. Plot de variables con datos imputados.....	8
Figura 4. Distribución de las variables seleccionadas en diagrama de cajas... ..	10
Figura 5. Detección de anomalías con agregación de outliers aleatorios	13
Figura 6. Clasificación de anomalías con rangos de tolerancia.....	14
Figura 7. Histograma de la Frecuencia para predicciones de anomalías	15
Figura 8. Serie temporal vs datos de dispersión - Tendencia del Voltaje del Generador Eléctrico	16
Figura 9. Semaforización de los datos en la tendencia	16
Figura 10. Scatterplot detección de anomalías bivariado según resultados de anomaly score	17
Figura 11. Scatter semáforo de detecciones bivariado	18

Índice de Tablas

Tabla 1. Descripción de variables seleccionadas para el análisis	9
Tabla 2. Estadísticos descriptivos de las variables seleccionadas	11
Tabla 3. Clasificación de datos en subgrupos según el valor de score dado por Isolation forest	14

1. Introducción

En el dinámico panorama empresarial actual, la adopción estratégica de soluciones novedosas se ha vuelto esencial para abordar los desafíos inherentes a diversas industrias. La industria petrolera, siendo un pilar fundamental de la economía ecuatoriana, no está exenta de obstáculos complejos que requieren una respuesta ingeniosa y efectiva. Uno de estos desafíos se relaciona con la gestión de la energía y el manejo óptimo que se debe llevar a cabo para los equipos de alta criticidad en la operación. Bajo ese contexto, se propone explorar cómo las soluciones y estrategias empresariales basadas en analítica de datos pueden revolucionar la forma en que la industria petrolera gestiona las operaciones de mantenimiento y la disponibilidad energética

Garantizar la operatividad cotidiana e ininterrumpida en una planta de producción de crudo radica en la gestión eficiente de la energía y el manejo óptimo de equipos de alta criticidad en las operaciones, para esta finalidad, la industria petrolera da preferencia a las turbinas de generación eléctrica que son consideradas componentes esenciales por el grado de confiabilidad y el aporte energético que manejan, proporcionando una parte significativa de la energía necesaria para alimentar equipos diversos y sistemas fundamentales en las instalaciones tanto en fondo como en superficie que componen el proceso industrial de extracción. Pese a esta mencionada confiabilidad, todo equipo mecánico y en especial las máquinas giratorias no pueden prescindir de mantenimientos periódicos que permitan gestionar la durabilidad del equipo rotativo y evaluar su condición a lo largo del tiempo, lo que afecta en consecuencia a los índices de disponibilidad de la máquina. Esto por su lado, los cataloga como equipos críticos.

La finalidad primordial de este proyecto es desarrollar un sistema de detección de anomalías con base en tecnologías avanzadas con el uso de la analítica de datos y el aprendizaje automático con la adquisición de datos disponible del sistema de monitoreo que permita optimizar la gestión de estas turbinas, y

determinar cómo el análisis predictivo está relacionado al mejoramiento de su eficiencia, por consecuencia la reducción de los costos de mantenimiento por paradas no programadas y la mejora de los indicadores de rendimiento de despacho energético reflejados en la disponibilidad de energía eléctrica.

2. Revisión de la Literatura

2.1 Detección de Anomalías

La detección de anomalías se puede entender como la actividad de distinguir y discriminar los elementos dentro de un conjunto establecido que no se ajustan a la conducta, tendencia o patrón común que la mayoría de los elementos siguen. Se refiere a la identificación de eventos o registros que son inusuales en el estado de un sistema en comparación con un estándar establecido en términos de sus características.

Para poder diferenciar una anomalía, es necesario contrastar elementos o eventos que se presentan con una alta frecuencia, de todos los elementos que difieren de este comportamiento habitual dentro de un margen de tolerancia.

Dentro del campo del aprendizaje automático y la estadística, la detección de anomalías es una técnica que se utiliza para identificar estos patrones inusuales o atípicos en un marco referencial de un conjunto de datos. Esta técnica es ampliamente utilizada en diversas aplicaciones, como la detección de malfuncionamientos o la identificación de fallas en máquinas y equipos, volviéndose un tema fundamental en el ámbito industrial sobre todo para áreas relacionadas al mantenimiento. Poder anticipar un evento potencial en una organización puede evitar pérdidas y afectaciones al activo, al personal y al medio ambiente.

2.2 Categoría de detección de anomalías

Como (Goldsteine & Uchida, 2016) tratan en su investigación, contrariamente a las tareas de clasificación convencionales, la detección de anomalías generalmente se utiliza en conjuntos de datos no etiquetados, centrándose únicamente en la estructura interna de los datos. Es a este desafío al que se le denomina detección de anomalías no supervisada.

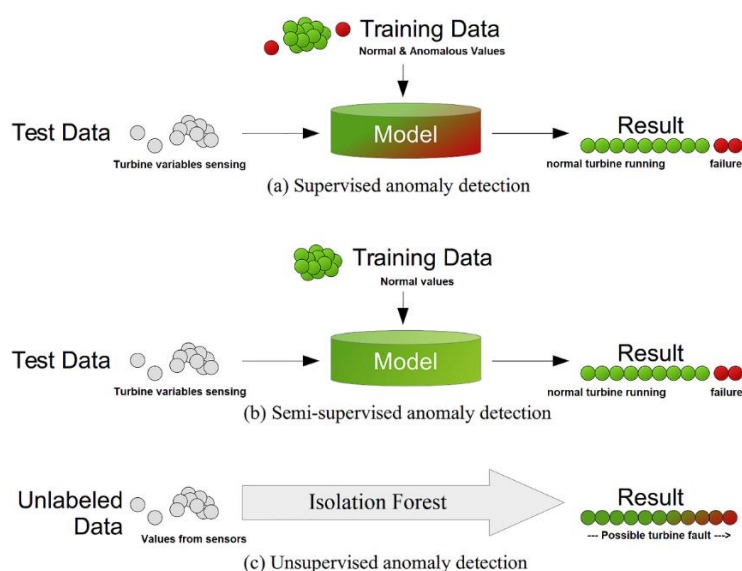


Figura 1. Modos de detección de anomalías con base en la disponibilidad de etiquetas en los datos. (Goldsteine & Uchida, 2016)

3. Identificación del Objeto de Estudio

En la industria petrolera, la generación eléctrica desempeña un papel crítico para mantener las operaciones en funcionamiento de manera continua y segura. Las turbinas de generación eléctrica son componentes esenciales en este sentido, suministrando alrededor de 20.5 MVatios de un total de capacidad dentro de las instalaciones de 120 MVatios, de manera que la energía que proporcionan bordea el 21.25 % de la energía necesaria para alimentar equipos y sistemas fundamentales ligados a la producción. Sin embargo, estas turbinas están sujetas a un alto desgaste debido a las condiciones extremas de funcionamiento

y la variabilidad en la demanda de energía como también subsistemas propios de respaldo, que por factores externos pueden convertirse en puntos de fallo. Prescindir de esta cantidad de energía de forma abrupta, resultaría en un parado parcial o total de planta ya sea por una inadecuada liberación de carga o el modo de configuración de manejo energético activo en determinado momento.

Estas dificultades se abordan mediante la supervisión del personal técnico encargado, la ejecución de maniobras operativas y el establecimiento de condiciones de proceso óptimas y seguras, respaldadas por operaciones de mantenimiento programadas previamente definidas.

El trabajo se centrará en las turbinas de generación eléctrica utilizadas dentro de las operaciones de extracción que se llevan a cabo en las plantas petroleras Bloque 16, facilidades del Sur, específicamente en el modelo GE-LM2500 de combustión dual Diesel/Gas, donde se investigarán y analizarán las características técnicas, el comportamiento operativo, el desgaste y la degradación de elementos asociados a estas turbinas en el contexto de aplicaciones industriales referentes a la industria del petróleo y su aplicación específica en la generación eléctrica con el propósito de analizar en profundidad su desempeño, optimización y posibles sugerencias de mantenimiento. El estudio abordará aspectos clave relacionados con la operación, confiabilidad, eficiencia y disponibilidad de la turbina, considerando su funcionamiento con dos diferentes tipos de combustible, tanto Diesel como Gas, y explorando las mejores prácticas en mantenimiento y estrategias de mejora de su rendimiento.

4. Planteamiento del problema

El mantenimiento de las turbinas de generación eléctrica en plantas petroleras es costoso, tanto en términos de tiempo como de recursos financieros. El enfoque tradicional de mantenimiento preventivo, basado en intervalos de tiempo predefinidos, no optimiza la utilización de los recursos y puede dar lugar a interrupciones no planificadas en la producción. Además, este enfoque no

aprovecha las posibilidades de la tecnología actual, como la monitorización en tiempo real en combinación con el análisis de datos avanzado.

Dentro de las facilidades del sur (SPF- Souther Plant Facilites), el departamento de Generación Eléctrica está a cargo de la operación de la turbina de generación Dual (Diesel/Gas). Por otro lado, existen equipos multidisciplinarios que intervienen en el desarrollo de las actividades de mantenimiento programadas y también de contingencia en caso de fallos aleatorios. Adicionalmente, las paradas programadas para los mantenimientos se tienen contempladas con frecuencias diferenciadas de acuerdo con los subsistemas auxiliares y componentes de seguridad que están asociados al equipo principal (máquina eléctrica).

Durante este periodo de funcionamiento ininterrumpido del equipo, se realiza monitoreo de las variables de proceso y los estatus de permisivos que intervienen en el funcionamiento, así como valores de todos los sensores que sirven de salvaguarda de la máquina para garantizar un apagado controlado y poner tanto la parte del proceso en modo seguro como precautelar la integridad del equipo en sí. En este sentido, se tienen rangos tolerables de operación con alarmas definidas a 4 niveles: bajo-bajo, bajo, alto y alto-alto. Sin embargo, incrementos o disminuciones no habituales de los valores de estas señales pueden estar demostrando un comportamiento anómalo sin necesidad de generar un evento de alarma. Se puede decir incluso que fluctuaciones de este tipo es un indicio del fallo del elemento sensor como tal.

Este proyecto se centrará en abordar el desafío de detectar anomalías en series temporales correspondientes a las señales de variables de proceso asociadas al funcionamiento y monitoreo de la turbina de generación eléctrica en una planta petrolera de deshidratación de crudo mediante un enfoque de detección anómala basada en técnicas de análisis de datos como el aprendizaje automático utilizando Isolation Forest, con el objetivo de mejorar la eficiencia operativa, evitar paradas de planta, reducir los costos y garantizar la continuidad de las operaciones del bloque 16.

Se deben considerar los desafíos específicos de las plantas petroleras, como la presencia de ambientes agresivos y la necesidad de mantener altos estándares de seguridad. Por lo tanto, el problema que se plantea es la necesidad de desarrollar un sistema de detección de anomalías confiable para este equipo en específico debido a su alta criticidad en cuanto al aporte de generación energética que representa. Este sistema deberá utilizar tecnologías de vanguardia, como son los sistemas de adquisición de datos y el análisis de datos avanzado, para encontrar patrones en el funcionamiento habitual de la máquina, predecir los no habituales y de esta manera anticipar y generar una oportunidad de acción temprana y toma de decisiones fuera del cronograma de mantenimiento establecido, minimizando así los costos y maximizando la disponibilidad de la turbina.

5. Objetivo General

Desarrollar un sistema de detección de anomalías basado en tecnologías avanzadas de adquisición y análisis de datos, específicamente utilizando la técnica de aprendizaje automático con Isolation Forest, para mejorar la eficiencia operativa y la disponibilidad de la turbina de generación eléctrica modelo GE-LM2500 de combustión dual Diesel/Gas en la planta petrolera del Bloque 16, facilidades del Sur. Este sistema permitirá anticipar y tomar acciones tempranas para prevenir fallas no planificadas, reducir los costos de mantenimiento y garantizar la continuidad de las operaciones en un entorno crítico y altamente demandante de energía.

6. Objetivos Específicos

- Generar un dataset con las series temporales relevantes, extraídas de la base de datos del sistema de monitoreo y adquisición SCADA de la Turbina para posterior procesamiento de datos.
- Implementar y entrenar un modelo de aprendizaje automático basado en la técnica de Isolation Forest, utilizando los datos recopilados.

- Determinar la efectividad y el ajuste de Isolation Forest con la puesta a prueba utilizando sets distintos de datos.
- Reducir la incidencia anual de paradas no programadas debido a fallos para reducir costos de operación.

7. Justificación y aplicación de la Metodología

7.1 Recopilación de los datos

La obtención de los datos proviene de fuentes primarias, principalmente recopilados mediante los dispositivos de medición, denominados dentro de la industria petrolífera como instrumentos de metrología o simplemente instrumentación. Estos instrumentos están basados en tecnologías que abarcan diferentes tipos de elementos análogos y digitales:

Dentro de lo que respecta a elementos digitales, están los llamados sensores, dispositivos con capacidad de detectar y responder a entradas específicas del entorno y transformar el dato en una señal eléctrica. Estos dispositivos son fundamentales para la recolección de datos y son el elemento básico para alimentar el sistema de adquisición de datos SCADA que reúne las señales de campo del cual ha sido extraída la mayor parte de la data. Otra parte que compone el dataset se ha obtenido de la observancia y registro manual del operador de planta dado que no todas las señales están integradas al sistema de supervisión.

7.2 Limpieza y preprocesamiento de datos

En lo que respecta a algoritmos con árboles de decisión, no es posible procesar datos que estén vacíos, todos los datos determinados como nulos producen un error en el procesamiento.

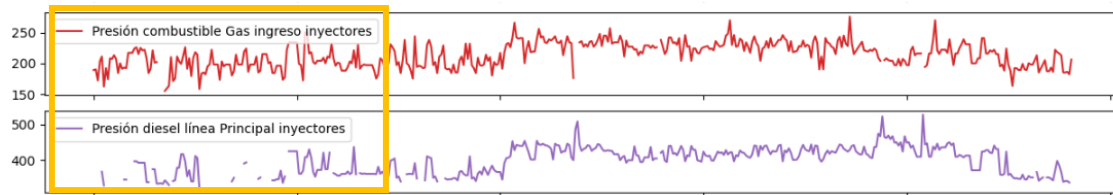


Figura 2. Plot de variables con datos incompletos.

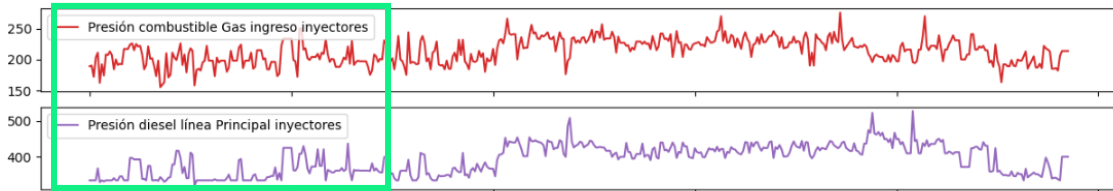


Figura 3. Plot de variables con datos imputados.

Para la depuración de los datos se ha realizado la implementación del algoritmo de imputación KNN basado en los vecinos próximos. El algoritmo asigna cada dato nuevo a un grupo basándose en la proximidad a los k vecinos más cercanos. Esto implica calcular la distancia entre el nuevo elemento y los existentes, ordenar estas distancias de menor a mayor y elegir el grupo al que pertenecerá en función del grupo con la mayor frecuencia y las distancias más cortas. (Merkle España, 2020)

Establecer el parámetro k , puede incurrir en un resultado diferente de imputación de los datos, ya que este algoritmo es sensible a la dimensión de la distancia entre los vecinos. Para determinar un k adecuado se han realizado varios ensayos y observaciones de los resultados.

7.3 Identificación y descripción de las variables

Para la realización del análisis, a lo largo de lo que va del año, se tiene una disponibilidad de 73 variables, que comprenden los puntos de consideración para la operación, control y monitoreo de la turbina.

Basándose en la experiencia del comportamiento de la máquina, las variables que representan una alta importancia por su nivel de criticidad y donde amerita aplicar el análisis son las siguientes:

Tabla 1. Descripción de variables seleccionadas para el análisis

Nombre de variable	Descripción	Tipo
Velocidad de turbina de potencia	RPM a las que gira la máquina rotativa. El valor nominal es 3600 rpm para producir una frecuencia eléctrica de 60 Hz. Estándar de frecuencia para la región.	Cuantitativa
Frecuencia del generador	Valor al que oscila la corriente eléctrica. Frecuencia eléctrica dentro de la red.	Cuantitativa
Velocidad Generador de Gas	Velocidad del sistema de gases que permite mantener la frecuencia de giro nominal (3600 rpm) y tomar mayor o menor carga.	Cuantitativa
Voltaje del Generador	Voltaje al que se produce la generación eléctrica hacia la barra de distribución.	Cuantitativa
Temperatura promedio turbina (T5.4)	Temperatura de trabajo de la máquina rotativa considerada como un bloque general.	Cuantitativa
Temperatura de aceite de turbina	Temperatura del aceite lubricante de partes mecánicas en la máquina rotativa (turbina)	Cuantitativa
Temperatura aceite generador	Temperatura del aceite lubricante de partes giratorias en el generador eléctrico.	Cuantitativa
Presión combustible Gas ingreso inyectores	Presión a la que se encuentra el Gas combustible en la entrada de la válvula hacia los inyectores (bifurcación)	Cuantitativa
Presión diesel línea Principal inyectores	Presión a la que se encuentra el Diesel a la entrada de la válvula de combustible hacia los inyectores (bifurcación)	Cuantitativa

Se puede notar que todas las variables son de tipo cuantitativo ya que son provenientes de sensores que toman los valores de variables físicas de proceso típicas como son Temperatura, Presión, y los valores de parámetros eléctricos asociados a la generación de energía eléctrica.

7.4 Visualización de variables

En el análisis descriptivo, se representan las variables seleccionadas para el estudio mediante diagramas de cajas para notar la distribución de los datos e identificar los outliers posibles de cada variable.

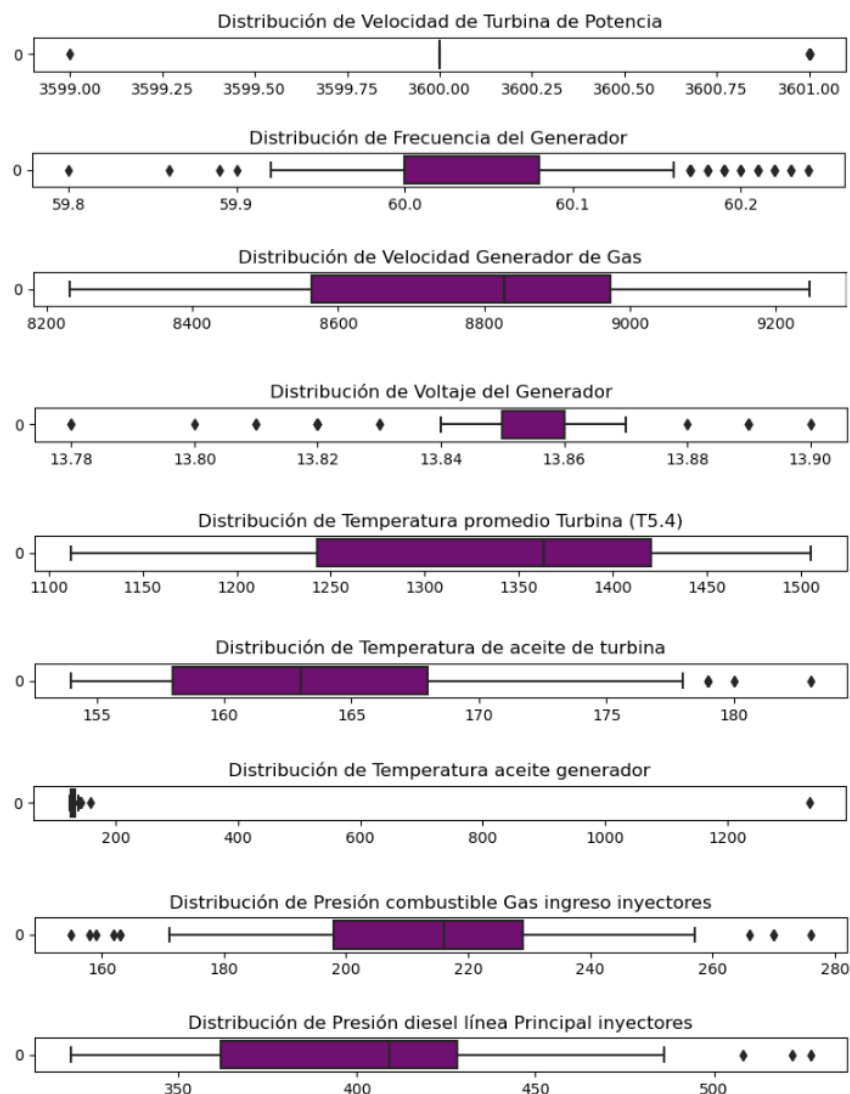


Figura 4. Distribución de las variables seleccionadas en diagrama de cajas.

De la Figura 4, se puede notar la distribución de las variables y los valores mínimos y máximos. Se puede considerar a primera vista que las variables poseen outliers que potencialmente podrían representar anomalías en las señales de los sensores. Partiendo de estos datos muy desviados se pretende determinar grupos de clasificación entre los datos, considerando unos como normales o aceptables y los que se considerarán anómalos, como un indicio de problemas a futuro con la consecuencia de un fallo en el equipo.

A continuación, en la Tabla 2. Se presentan lo estadísticos Descriptivos de las variables seleccionadas para el modelado.

Tabla 2. Estadísticos descriptivos de las variables seleccionadas

	Velocidad de Turbina de Potencia	Frecuencia del Generador	Velocidad Generador de Gas	Voltaje del Generador	Temperatura promedio Turbina (T5.4)	Temperatura de aceite de turbina	Temperatura aceite generador	Presión combustible Gas ingreso inyectores	Presión diesel línea Principal inyectores
count	477.000000	481.000000	477.000000	481.000000	477.000000	481.000000	481.000000	474.000000	423.000000
mean	3600.033543	60.042952	8782.890985	13.854649	1334.184486	163.858628	133.035343	213.683544	400.293144
std	0.191540	0.071449	240.484675	0.012445	101.706818	6.490440	55.040977	20.251199	37.383235
min	3599.000000	59.800000	8232.000000	13.780000	1112.000000	154.000000	125.000000	155.000000	320.000000
25%	3600.000000	60.000000	8564.000000	13.850000	1243.000000	158.000000	127.000000	198.000000	362.000000
50%	3600.000000	60.000000	8828.000000	13.860000	1363.000000	163.000000	129.000000	216.000000	409.000000
75%	3600.000000	60.080000	8974.000000	13.860000	1420.000000	168.000000	133.000000	229.000000	428.000000
max	3601.000000	60.240000	9246.000000	13.900000	1505.000000	183.000000	1334.000000	276.000000	527.000000

7.5 Modelo Estadístico (Modelo de clasificación)

7.5.1 Isolation Forest

Dentro de los modelos estadísticos, se encuentra el caso de algoritmos que permiten la clasificación automática de datos. En lo que a estos respecta, se encuentra el algoritmo Isolation Forest basado en arboles de decisión.

Este algoritmo difiere de otros destinados a la detección de anomalías porque principalmente no espera a definir lo que se considera normal para luego reportar todo lo restante como los datos anómalos, sino que se enfoca en aislar las anomalías directamente. (Maklyn, 2022) Funciona construyendo un conjunto de árboles de decisión de forma aleatoria y en lugar de realizar las particiones más

puras (medida mediante Gini o Entropía) utiliza estos árboles para dividir repetidamente el conjunto de datos en subconjuntos.

La forma de definir si un dato es anómalo, es realizando un `anomaly_score` mediante el cálculo (Juaregui, 2020):

$$s(x, n) = 2 - E(h(x)) c(n)$$

En donde:

$h(x)$: es la profundidad (o altura) media de X de los iTrees construidos.

$c(n)$ es la altura media para encontrar un nodo en un Isolation Tree.

n : es el tamaño del dataset.

En lenguaje de programación Python, el modelo se encuentra especificado bajo la librería Scikit-learn.

En el caso de las variables cuantitativas que se manejan en torno a los puntos de monitoreo de la turbina, se tienen solamente valores enteros y flotantes, en adición como ya se mencionó, estos datos no tienen etiqueta alguna que los pueda definir, por lo que, una vez generada la imputación de datos, es posible realizar el modelado de datos directamente y observar cómo responde el modelo y su ajuste.

8. Resultados

8.1 Análisis del modelo

El parámetro principal configurado en el modelo es la contaminación. Se ha determinado en un valor de **contamination** = 0.022, lo que representa que al menos el 2.2% sea marcado como anomalías para el grado de sensibilidad.

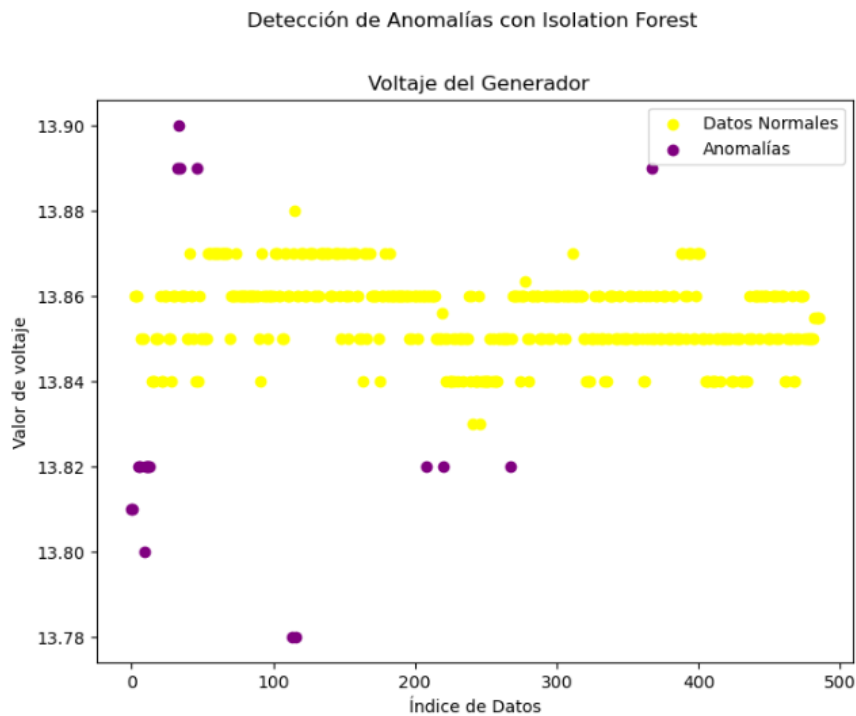


Figura 5. Detección de anomalías con agregación de outliers aleatorios

En la Figura 5. se puede observar cómo el modelo en la clasificación polariza los datos y los determina como anomalías mientras más dispersos se encuentren los puntos, lo cual genera una zona o franja que puede denominarse tolerancia de aceptabilidad. Estas distancias de dispersión en ciertos casos puede que no necesariamente sean anomalías y, por el contrario, generar falsos positivos.

Como el modelo isolation forest maneja una calificación para la clasificación, es posible determinar rangos en función del valor de calificación, en los que el dato clasificado se muestre en uno de estos nuevos grupos catalogados como aceptables y no como anomalía definitiva. Esto ayuda a manejar la ambigüedad en los datos y mejora la capacidad del modelo para generalizar a nuevos datos.

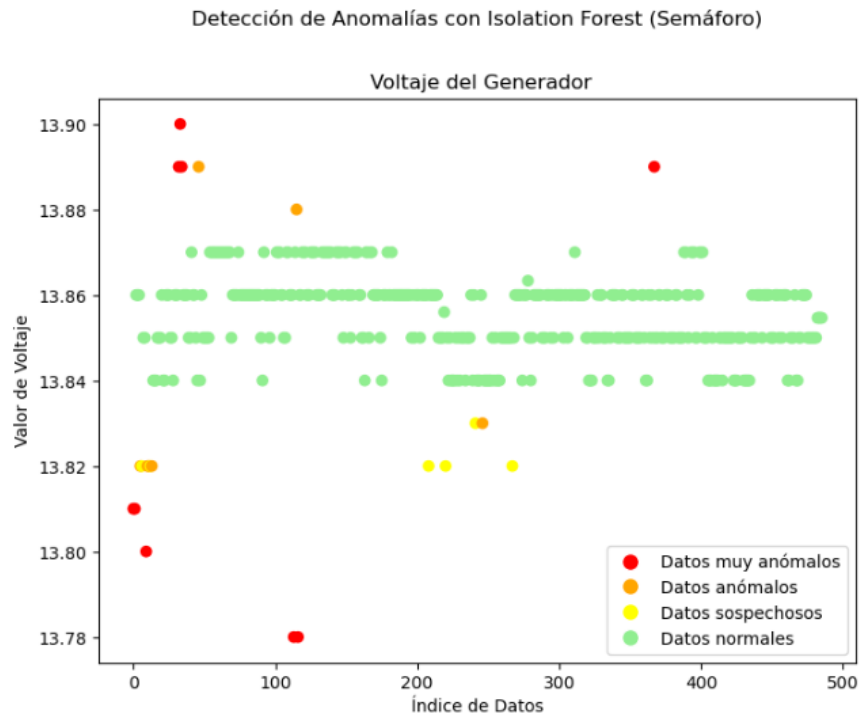


Figura 6. Clasificación de anomalías con rangos de tolerancia

Los subgrupos mostrados en la Figura 6. describen los datos anómalos en dos diferentes grados, y para los datos considerados como normales se añaden los que ya presentan una cierta desviación, pero no necesariamente podrían representar una anomalía.

Tabla 3. Clasificación de datos en subgrupos según el valor de score dado por Isolation forest

Grupos de Clasificación	Score value
Datos muy anómalos	< -0.07
Datos anómalos	< -0.04
Datos sospechosos (borderline)	< -0.005
Datos normales	≤ 0

8.2 Interpretación de resultados

En la clasificación de datos se espera obtener todas las entradas de datos que son anómalas para poder detectar estos comportamientos a tiempo y evitar los paros abruptos del equipo principal.

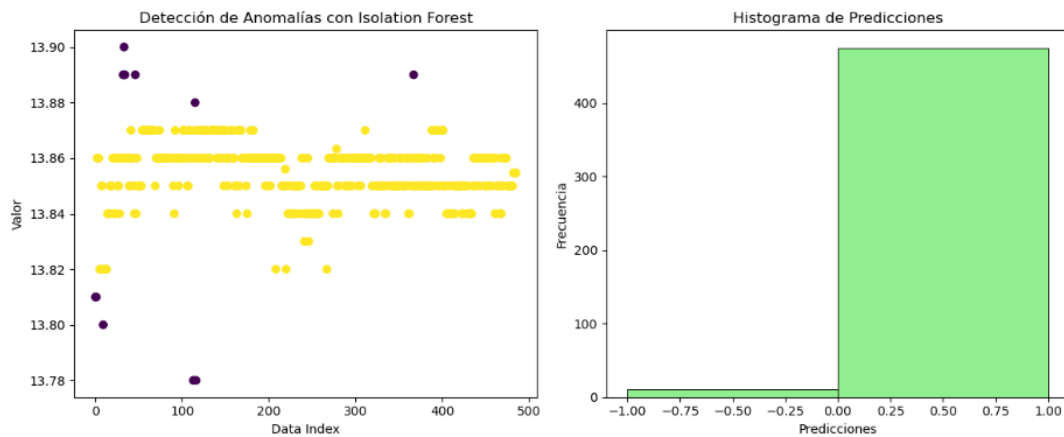


Figura 7. Histograma de la Frecuencia para predicciones de anomalías

En la Figura 7. se muestra la frecuencia con la que se presentaron datos considerados como desviaciones con respecto al grupo de datos total. Se ha tomado en consideración un set de datos de 486 entradas que representan el monitoreo de tres meses consecutivos con un muestreo de cuatro veces por día.

En la secuencia de los datos, resulta difícil la orientación del punto de la trama de datos en que se genera la anomalía detectada, por lo que para poder referenciar estas detecciones es necesario ubicar los datos de interés mediante la fecha en la que fueron registrados.

Para el análisis de fallos en el área de mantenimiento, es habitual presentar tendencias obtenidas de las bases de datos de los historiadores con cortes en periodos de tiempo a conveniencia, para ajustar el despliegue de los datos con las eventualidades suscitadas, de manera que se pueda estimar el comportamiento previo, la detonación del evento y las consecuencias en un

orden cronológico. De esta manera es posible, en un análisis causa raíz, determinar qué pasó y la forma en que ocurrió un fallo de un equipo.

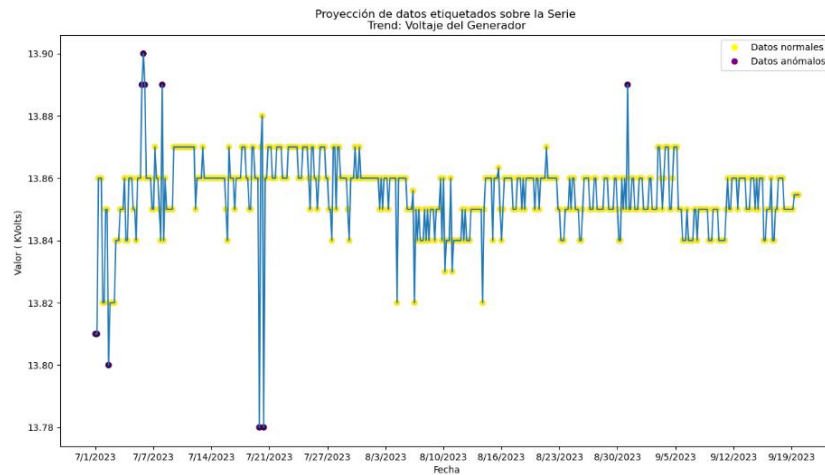


Figura 8. Serie temporal vs datos de dispersión - Tendencia del Voltaje del Generador Eléctrico

De la Figura 8. Se puede observar los eventos en los que los datos son catalogados como anómalos, sin embargo, solamente los eventos suscitados el 20 de julio representaron un apagado de equipos por problemas en la red eléctrica por “Under voltage” producto de una falla del controlador de velocidad.

Los otros datos anómalos presentados, no representan un fallo pese a que son considerablemente desviados respecto de la tendencia. Nuevamente se evidencia que la polarización entre normal y anómalo no permite una determinación irrefutable.

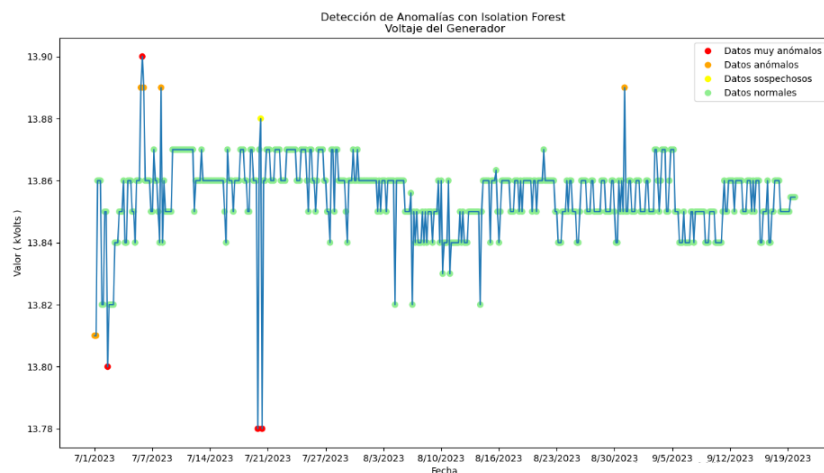


Figura 9. Semaforización de los datos en la tendencia

Se replica el mismo concepto de semaforización para atenuar esta polaridad y enriquecer el ajuste del modelo. De esta manera, en la Figura 9. está representado el fallo real dado por bajo voltaje y los demás eventos son considerados como advertencias donde la tendencia ha presentado picos anómalos que se considerarán como indicios de un posible problema y en el escenario menos favorable, un evento de apagado del equipo.

9. . Discusión de los resultados y propuesta de solución

Dado que el análisis se centra en la definición de los datos en valores de tipo categórico, es posible identificar la etiqueta que el algoritmo terminará decretando para el dato con base en el rango numérico de valores que son asignados a los datos como resultado del procesamiento y calificación que estima el algoritmo (anomaly score value), como se presenta en la Figura 10, y esto contrastado con los rangos que definen los subgrupos presentados en la Tabla 3. Clasificación de datos en subgrupos según el valor de score dado por Isolation forest. De esta manera, la categorización puede tener un grado de congruencia cuantificable que permite evaluar si los límites de cada rango de los subgrupos están ajustados a lo que realmente se deba considerar como anómalo o normal.

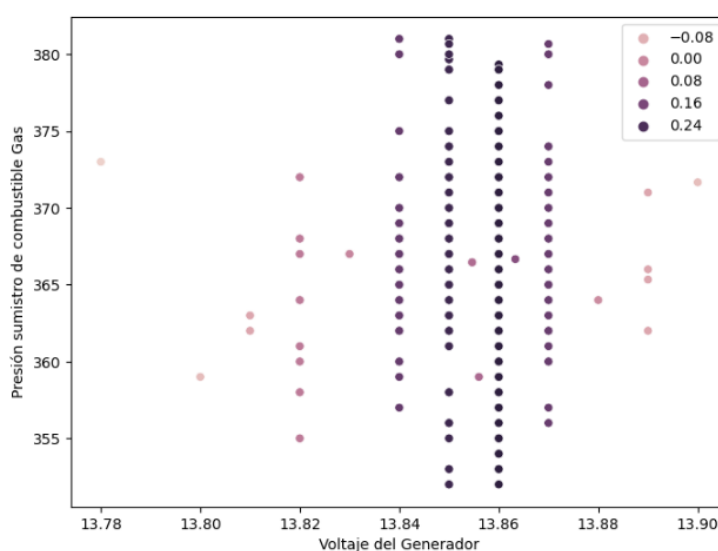


Figura 10. Scatterplot detección de anomalías bivariado según resultados de anomaly score

El matiz que presentan los datos en la Figura 10. parte de un valor neutro correspondiente a cero, para luego degradar o definir en pasos de ± 0.08 respectivamente. Al utilizar este valor como referencia, se puede tener la certeza que, si un dato es evaluado con -0.08 y el subgrupo de muy anómalos datos son todos los datos con score menor a -0.07 , ese dato será decretado efectivamente como muy anómalo y se representará con color rojo. Este proceso se repetirá para los demás subgrupos que fueron definidos con rangos contiguos y se pueden ver representados en la Figura 11.

Con esta información numérica y la comparativa de los valores, es posible llevar la clasificación a un estado más cercano a la realidad.

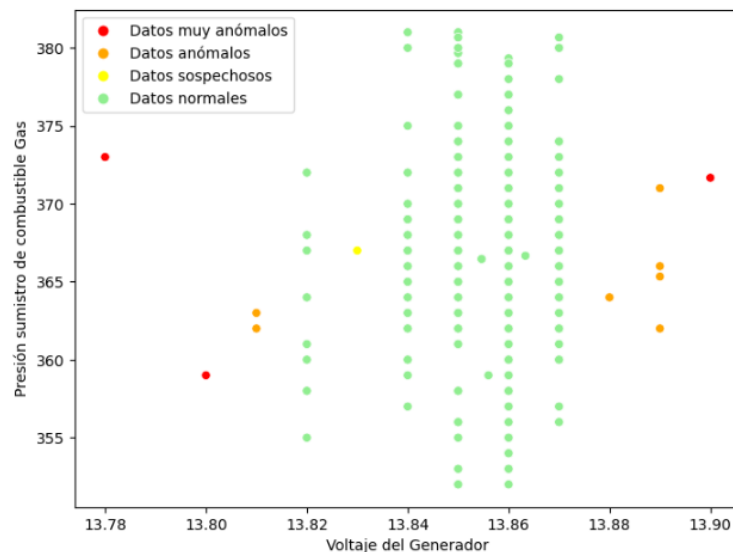


Figura 11. Scatter semáforo de detecciones bivariado

9.1 Implicaciones para la organización

Por medio de este proyecto, se da paso a la optimización, implementando un ciclo de mejora continua basado en datos. Recopila retroalimentación de la operación, analiza datos de rendimiento y ajusta las estrategias en consecuencia para adaptarse a las cambiantes condiciones del sector y las necesidades del proceso.

La empresa tiene el compromiso de optimizar y reducir el consumo innecesario de insumos y repuestos adicionales producto de mantenimientos y paradas no planificadas, con prácticas que reduzcan su impacto ambiental, como la reducción de residuos, el uso eficiente de recursos naturales, la adopción de energías renovables y la minimización de emisiones de carbono; de esta manera, la optimización de la gestión energética y mantenimiento se relacionan y aportan como una base sólida a la mejora de la cadena de valor.

Reducir el número de eventos de apagados forzados de la máquina de generación representa un ahorro cuantioso para la empresa, en cada evento que se produce uno de estos parados se ven involucrados gastos en materiales e insumos, repuestos y mano de obra calificada para la determinación de los fallos. Adicionalmente, el factor de pérdida por inactividad del equipo que está asociada a la pérdida de producción de crudo. Es por consiguiente imperativo desarrollar los mecanismos que permitan estimar problemas en los componentes de la turbina para evitar a toda costa un apagado abrupto. Considerando que la detección de anomalías puede proporcionar una ventana de tiempo de acción, la modalidad de reacción cambia para dar lugar a la toma de decisiones informadas como una nueva posibilidad. Esto da paso a la proactividad dentro de las tareas de mantenimiento.

Se propone dar paso a un esquema de mantenimiento con frecuencias dinámicas basadas en el número de anomalías detectadas en periodos semanales. El acumulado de anomalías durante el periodo significará un acorte o prolongación en acciones de revisión de los instrumentos y subsistemas hasta lograr definir una frecuencia de mantenimiento óptima para cada sección del equipo mayor (turbina).

9.1.1 Adopción de estrategias empresariales

9.1.1.1 Estrategia: Responsabilidad social y sostenibilidad

La extracción de petróleo y gas en áreas sensibles puede tener un impacto significativo en los ecosistemas locales. Derrames de petróleo, contaminación del agua y degradación del suelo son amenazas importantes. La empresa debe tomar medidas para minimizar estos impactos y cumplir con las regulaciones ambientales que garanticen que sus operaciones no afecten negativamente a las especies de flora y fauna presentes en la zona, considerando principalmente que las áreas ambientalmente protegidas suelen albergar una gran biodiversidad, lo que puede requerir medidas adicionales de mitigación y monitoreo. Integrar prácticas empresariales socialmente responsables y sostenibles puede ser una forma de diferenciación y ganar la lealtad de las partes involucradas en el área del petróleo.

A esto se suma que las colonias Waoranis, como comunidad indígena, tienen derechos legales y culturales sobre sus tierras ancestrales, por consiguiente, la empresa debe hacer valer estos derechos y mantener una relación de respeto y consulta con las comunidades locales. Los conflictos con las comunidades pueden resultar en retrasos operativos y daños a la reputación.

Mantener una reputación positiva y relaciones sólidas con las comunidades locales y las partes interesadas, conlleva a realizar actividades de responsabilidad social corporativa que incluyan proyectos de desarrollo sostenible:

- Reducción de huella ambiental

Las empresas pueden implementar prácticas que reduzcan su impacto ambiental, como la reducción de residuos, el uso eficiente de recursos naturales y la minimización de emisiones de carbono. Estas acciones no solo son buenas para el medio ambiente, sino que también pueden generar ahorros a largo plazo.

- Ética en la cadena de suministro

Enfatizar en conseguir un manejo en la ética de la cadena de suministro es fundamental, de tal manera que garantice que los proveedores cumplan con estándares éticos y sostenibles en la producción de materias primas y productos. Esto puede incluir la reducción de la explotación de recursos naturales.

9.1.1.2 Estrategia: Innovación y mejora continua

Analizar la cadena de valor de una empresa es identificar las actividades que agregan valor al servicio y aquellas que no lo hacen o que podrían mejorarse.

Fomentar la innovación en procesos para reducir el impacto ambiental y social permite una mejora considerable en la cadena de valor. Esto implica llevar a cabo actividades basadas en tecnologías recientes que soporten el objetivo de mejorar continuamente:

- Transformación digital

Se adoptan tecnologías digitales para mejorar las operaciones y el modelo de negocio en general. Esta manera de innovación implica la integración de tecnologías digitales con el objetivo de aprovechar al máximo las oportunidades que ofrece la tecnología para ser más eficientes.

- Optimización de la energía y el mantenimiento

Prever fallos en redes eléctricas y otros sistemas críticos para evitar cortes energéticos y mejorar la fiabilidad.

Adicionalmente es posible optimizar la producción y distribución de energía en función de la demanda y las condiciones del proceso industrial.

En una industria con activos costosos, la analítica de datos puede ayudar a predecir fallos y optimizar la programación de mantenimiento preventivo para evitar interrupciones no planificadas. La gestión de redes con analítica de datos busca prevenir problemas antes de que ocurran.

10. Conclusiones y Recomendaciones

10.1 Conclusiones

- El uso de tecnologías como la analítica de datos y el aprendizaje automático demuestra ser efectivo en el análisis de grandes conjuntos de datos provenientes de sensores en equipos industriales. En este caso, el modelo de Isolation Forest empleado para clasificar datos en categorías de anomalías, permite una identificación temprana de posibles problemas en las turbinas de generación eléctrica. No solo tiene un impacto significativo en la reducción de costos operativos al evitar paradas no programadas, sino también mejora la eficiencia en la toma de decisiones para el mantenimiento preventivo, la aplicación de estos sistemas de detección proporciona una herramienta poderosa para la toma de decisiones proactivas.
- La aplicación de estrategias basadas en analítica de datos representa una diferenciación para la organización, al establecer un sistema de clasificación que distingue entre diferentes niveles de anomalías, el personal de mantenimiento puede priorizar y abordar problemas potenciales de manera más efectiva. Además, esta metodología proporciona una base para ajustar las frecuencias de mantenimiento en función de la cantidad y la gravedad de las anomalías detectadas, lo que lleva a una gestión más eficaz de los recursos y una mayor disponibilidad operativa. Estos enfoques no solo son tecnológicamente avanzados, sino que también representan un cambio de paradigma en la gestión de equipos industriales al permitir una toma de decisiones más informada, dando paso a la mejora continua.

10.2 Recomendaciones

- En los modelos de aprendizaje no supervisado, pese a que estos pueden trabajar con grandes volúmenes de datos sin etiqueta y son especializados en segmentación (normal y anómalo), estos corren el riesgo de que descubran patrones o estructuras que son simplemente ruido o variabilidad aleatoria. En este caso en particular, se debe tener en consideración el proceso de imputación de datos en la fase de limpieza, ya que un sobreajuste en el factor k utilizado en el algoritmo KNN puede resultar en una contaminación que degrade la efectividad de la clasificación.
- Pese a que los modelos de detección de anomalías como el Isolation Forest presentan una característica de robustez ante los datos atípicos, que es precisamente lo que los vuelve ideales para estas aplicaciones, necesitan de grandes cantidades de datos para mejorar su precisión y evitar el overfitting. Manejar un volumen considerable de datos al momento de evaluar y realizar el ajuste proporciona mejores resultados en la segmentación.

11. Referencias

- Ariza, J. C. (6 de junio de 2023). *Comparación de Métodos de Aprendizaje Automático para la clasificación*. Obtenido de Universidad de los Andes Colombia: <http://hdl.handle.net/1992/68553>
- Barbado González, A. (14 de Marzo de 2023). *Towards Explainable Machine Learning for Anomaly Detection in Real-World Contexts*. Obtenido de Universidad Politécnica de Madrid Archivo Digital: <https://doi.org/10.20868/UPM.thesis.73247>.
- Bernate, C. T. (2007). *Herramientas de detección de fallas y aislamiento en entornos industriales*. Obtenido de Repositorio UTB: <https://hdl.handle.net/20.500.12585/1210>

- Figuerola, D. A. (2019). *Detección de anomalías en componentes mecánicos en base a deep learning*. Obtenido de Repositorio Universidad de Chile: <https://repositorio.uchile.cl/bitstream/handle/2250/170571/Deteccion-de-anomalias-en-componentes-mecanicos.pdf?sequence=1&isAllowed=y>
- Goldsteine, M., & Uchida, S. (2016). A Comparative Evaluation of Unsupervised. *Plos One*, 1-6.
- Jorge Ivan Romero Gelvez, B. S. (2020). *Aplicación de machine learning en el mantenimiento predictivo industrial con herramientas de código abierto*. Obtenido de Universidad de Bogotá Jorge Tadeo Lozano: <https://expeditiorepositorio.utadeo.edu.co/handle/20.500.12010/10108>
- Juaregui, A. F. (2020). *Isolation Forest en Python*. Obtenido de Andes Fernandez Blog: <https://anderfernandez.com/blog/isolation-forest-en-python/>
- Maklyn, C. (15 de julio de 2022). *Isolation Forest*. Obtenido de Medium: <https://medium.com/@corymaklin/isolation-forest-799fcea44#:~:text=Isolation%20Forest%20is%20an%20unsupervised%20machine%20learning%20algorithm%20for%20anomaly,to%20a%20given%20data%20point.>
- Merkle España. (1 de septiembre de 2020). *El algoritmo K-NN y su importancia en el modelado de datos*. Obtenido de Merkle Blog: <https://www.merkle.com/es/es/blog/algoritmo-knn-modelado-datos>
- Noriega, A. V. (septiembre de 2021). *Aplicación de técnicas de Machine Learning a la predicción de fallos de discos mediante el uso de Spark*. Obtenido de Repositorio abierto de la Universidad de Cantabria: <http://hdl.handle.net/10902/25439>
- Sarmiento Cossio, A. N. (6 de Abril de 2023). *Predicción de la generación de energía eléctrica bruta en bolivia utilizando técnicas de machine learning*.

Obtenido de Documentación Digital de la Universidad Mayor de San
Simón: <http://hdl.handle.net/123456789/37893>

ANEXOS