



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA MECÁNICA
ELÉCTRICA**

Sistema de mantenimiento predictivo utilizando algoritmos de
Machine Learning en un compresor de aire industrial en el
sector industrial, 2024

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero Mecánico Electricista

AUTOR:

Alvarez Campos, Daniel Augusto (orcid.org/0000-0003-4297-6848)

ASESORA:

Mg. Serrepe Ranno, Mirian Marcela (orcid.org/0000-0001-9342-1717)

LÍNEA DE INVESTIGACIÓN:

Sistemas y Planes de Mantenimiento

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo sostenible y adaptación al cambio climático

TRUJILLO – PERÚ

2024



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA MECÁNICA ELÉCTRICA**

Declaratoria de Autenticidad del Asesor

Yo, SERREPE RANNO MIRIAM MARCELA, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA MECÁNICA ELÉCTRICA de la UNIVERSIDAD CÉSAR VALLEJO SAC - TRUJILLO, asesor de Tesis titulada: "Sistema de Mantenimiento Predictivo utilizando algoritmos de Machine Learning en un compresor de aire industrial en el sector industrial, 2024", cuyo autor es ALVAREZ CAMPOS DANIEL AUGUSTO, constato que la investigación tiene un índice de similitud de 18%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

TRUJILLO, 19 de Julio del 2024

Apellidos y Nombres del Asesor:	Firma
SERREPE RANNO MIRIAM MARCELA DNI: 06437594 ORCID: 0000-0001-9342-1717	Firmado electrónicamente por: SRANNOMM el 31- 08-2024 14:54:21

Código documento Trilce: TRI - 0823719



**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA MECÁNICA ELÉCTRICA**

Declaratoria de Originalidad del Autor

Yo, ALVAREZ CAMPOS DANIEL AUGUSTO estudiante de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA MECÁNICA ELÉCTRICA de la UNIVERSIDAD CÉSAR VALLEJO SAC - TRUJILLO, declaro bajo juramento que todos los datos e información que acompañan la Tesis titulada: "Sistema de Mantenimiento Predictivo utilizando algoritmos de Machine Learning en un compresor de aire industrial en el sector industrial, 2024", es de mi autoría, por lo tanto, declaro que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. He mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

En tal sentido asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de la información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

Nombres y Apellidos	Firma
DANIEL AUGUSTO ALVAREZ CAMPOS DNI: 44665640 ORCID: 0000-0003-4297-6848	Firmado electrónicamente por: DAALVAREZ el 19-07- 2024 10:38:38

Código documento Trilce: TRI - 0823721

Dedicatoria

Dedico este logro a todas las personas que han creído en mí y me han alentado a seguir adelante, especialmente en los momentos más difíciles. Este trabajo también es para ustedes.

Agradecimiento

Quiero expresar mi sincero agradecimiento a todas las personas que han contribuido a la realización de esta tesis.

A mi asesora, Mg. Serrepe Ranno, Miriam Marcela, por su orientación y apoyo constante durante este proceso. A mi familia, por su amor y comprensión incondicional. A mis amigos y compañeros de estudios, por su ayuda y ánimo en los momentos más difíciles.

A mis profesores de la Universidad César Vallejo, por compartir sus conocimientos y experiencias.

Gracias a todos aquellos que, de alguna manera, aportaron a este proyecto.

Índice de contenidos

Carátula.....	i
Declaratoria de autenticidad de la asesora	ii
Declaratoria de originalidad del autor.....	iii
Dedicatoria	iv
Agradecimiento	v
Resumen.....	ix
Abstract	x
I. INTRODUCCIÓN	1
II. METODOLOGÍA	12
III. RESULTADOS.....	24
IV. DISCUSIÓN.....	46
V. CONCLUSIONES	50
VI. RECOMENDACIONES	51
REFERENCIAS.....	52
ANEXOS	58

Índice de tablas

Tabla 1. Estadísticos descriptivos de las variables temperatura, presión, vibración y humedad.	25
Tabla 2. Frecuencia de fallas.	25
Tabla 3. Cantidad de valores faltantes por variable.....	28
Tabla 4. Estadísticas descriptivas según falla o no falla.....	31
Tabla 5. Resultados de la prueba t para la diferencia de medias por variable.....	32
Tabla 6. Resultados del modelo de regresión logística.	36
Tabla 7. Matriz de confusión inicial del modelo logit.....	37
Tabla 8. Reporte de clasificación inicial del modelo logit.....	38
Tabla 9. Matriz de Confusión del Modelo Logit con el Punto de Corte Óptimo.....	39
Tabla 10. Reporte de Clasificación con el Nuevo Punto de Corte.....	39
Tabla 11. Matriz de Confusión del Modelo de Árbol Inicial.....	40
Tabla 12. Reporte de Clasificación del Modelo de Árbol Inicial.....	40
Tabla 13. Matriz de confusión con punto de corte óptimo para el modelo de árbol.....	41
Tabla 14. Reporte de clasificación con punto de corte óptimo para el modelo de árbol.....	41
Tabla 15. Matriz de Confusión Inicial para el Modelo Random Forest.....	42
Tabla 16. Reporte de Clasificación Inicial para el Modelo Random Forest.....	43
Tabla 17. Matriz de confusión con punto de corte óptimo para el modelo de random forest.....	44
Tabla 18. Reporte de clasificación con punto de corte óptimo para el modelo de random forest	45

Índice de Figuras

<i>Figura 1.</i> Función sigmoide	21
<i>Figura 2.</i> Gráfica de cajas de la temperatura: antes y después del truncamiento	26
<i>Figura 3.</i> Gráfica de cajas de la presión: antes y después del truncamiento.	26
<i>Figura 4.</i> Gráfica de cajas de la vibración: antes y después del truncamiento.....	27
<i>Figura 5.</i> Gráfica de cajas de la humedad: antes y después del truncamiento.	27
<i>Figura 6.</i> Gráfica de cajas de las variables estandarizadas.....	28
<i>Figura 7.</i> Comparación de las densidades de la temperatura antes y después de la imputación.	29
<i>Figura 8.</i> Comparación de las densidades de la presión antes y después de la imputación	29
<i>Figura 9.</i> Comparación de las densidades de la vibración antes y después de la imputación.	30
<i>Figura 10.</i> Comparación de las densidades de la humedad antes y después de la imputación.	30
<i>Figura 11.</i> Gráficos de cajas comparativos de cada variable según la presencia de fallas. .	31
<i>Figura 12.</i> Distribución de la variable "fallo" en los conjuntos de entrenamiento y prueba...	34
<i>Figura 13.</i> Accuracy del modelo de árbol de decisión en los conjuntos de entrenamiento y prueba a diferentes niveles de profundidad.....	35
<i>Figura 14.</i> Accuracy del modelo de random forest en los conjuntos de entrenamiento y prueba para diferentes tamaños de bosque y profundidades.	36
<i>Figura 15.</i> Curva ROC del modelo logit.....	38
<i>Figura 16.</i> Curva ROC del modelo de árbol de clasificación.....	41
<i>Figura 17.</i> Curva ROC para el modelo de Random Forest.....	44

Resumen

El objetivo principal de la investigación fue plantear un Sistema de Mantenimiento Predictivo utilizando Algoritmos de Machine Learning para solucionar la inactividad productiva causada por reparaciones de maquinaria en el sector industrial en el 2024. La investigación fue de tipo básica, de enfoque cuantitativo, no experimental y correlacional. Se recolectaron los datos a partir de una ficha de recolección mediante los sensores avanzados, la muestra objeto de estudio en esta investigación estuvo compuesta por 1 compresor de aire industrial. El desarrollo de modelos predictivos se realizó utilizando algoritmos avanzados de Machine Learning en Python. Los resultados indicaron que el modelo de regresión logística mejoró el recall en la detección de fallos hasta un 92,5% y el árbol de decisión alcanzó una precisión del 85,1%. El random forest optimizó el equilibrio entre precisión y recall, alcanzando un 90,3% en el F1-score. Concluye que, a pesar de tener una precisión general del 0,88, el modelo de Random Forest enfrenta desafíos en la identificación de fallos, con una precisión inicial baja, aunque el ajuste del punto de corte mejora significativamente el recall para la detección de fallos críticos, aumentando la sensibilidad del modelo a los fallos a costa de incrementar los falsos positivos.

Palabras clave: *Mantenimiento predictivo, machine learning, compresor industrial*

Abstract

The main objective of the research was to propose a Predictive Maintenance System using Machine Learning Algorithms to solve productive downtime caused by machinery repairs in the industrial sector in 2024. The research was basic, with a quantitative, non-experimental and correlational. The data was collected from a collection sheet using advanced sensors; the sample under study in this research was composed of 1 industrial air compressor. The development of predictive models was carried out using advanced Machine Learning algorithms in Python. The results indicated that the logistic regression model improved the recall in fault detection up to 92.5% and the decision tree reached an accuracy of 85.1%. The random forest optimized the balance between precision and recall, reaching 90.3% in the F1-score. It concludes that, despite having an overall precision of 0.88, the Random Forest model faces challenges in identifying faults, with a low initial precision, although adjusting the cut-off point significantly improves the recall for fault detection. critical, increasing the sensitivity of the model to failures at the cost of increasing false positives.

Keywords: *Predictive maintenance, machine learning, industrial compressor*

I. INTRODUCCIÓN

Los Sistemas de Mantenimiento Predictivo (PMS) son fundamentales en la Industria 4.0, mejorando el tiempo de actividad del equipo y reduciendo costos (Andrioaia & Gaitan, 2024). Los recientes avances en el Internet de las Cosas (IoT), la computación en la nube y el análisis de datos han facilitado la recopilación y análisis de datos en tiempo real, lo que mejora la detección temprana de fallos. La planificación del mantenimiento se vuelve más eficiente y se reducen las interrupciones inesperadas gracias a la capacidad de los sistemas de gestión de mantenimiento (PMS) para anticipar y prevenir fallos mediante un monitoreo constante. Estos sistemas, que emplean sensores para supervisar los activos y organizar de manera óptima las actividades de mantenimiento, generalmente constan de cuatro componentes clave: adquisición de datos, diagnóstico de fallos, pronóstico de fallos y toma de decisiones de mantenimiento. Esta estructura integral proporciona un enfoque sistemático y eficiente para la gestión del mantenimiento (Ayvaz & Alpaz, 2021). La capacidad de recopilar y analizar datos en tiempo real ha llevado a un nivel sin precedentes de precisión y eficiencia en la detección temprana de fallos.

En el contexto internacional, diversas investigaciones han demostrado la eficacia del mantenimiento predictivo utilizando Machine Learning. Por ejemplo, investigaciones en la industria manufacturera han empleado redes neuronales recurrentes (RNN) para el mantenimiento predictivo de equipos, donde han logrado una precisión superior al 95% en la clasificación de condiciones críticas de desgaste (Küfner et al., 2021). En el estudio de Iftikhar et al. (2022), se utilizaron algoritmos de Machine Learning para la predicción de fallos en compresores industriales, logrando una mejora significativa en la eficiencia operativa mediante técnicas de detección de anomalías. Además, se aplicaron redes neuronales recurrentes (RNN) para predecir fallos en equipos de la industria del petróleo y gas, lo que resultó en un incremento de la fiabilidad y la eficiencia operativa. Estos estudios ilustran cómo estas herramientas tienen el potencial de transformar las prácticas de mantenimiento en el sector industrial.

En el contexto nacional, los sistemas de mantenimiento predictivo han ganado popularidad en el ámbito de la maquinaria pesada. Para analizar los datos de

mantenimiento y anticipar fallas en el equipo, emplean algoritmos de aprendizaje automático (Castro et al., 2022). Por medio de técnicas como los árboles de decisión agrupados, se puede alcanzar una predicción de fallos muy precisa, lo cual mejora la planificación del mantenimiento y la continuidad del servicio. La intervención efectiva en el mantenimiento y la detección temprana de fallas se benefician de la recopilación y el análisis de datos en tiempo real cuando se integran tecnologías como el aprendizaje automático, el IoT y la realidad aumentada (Alves et al., 2020). Diversas investigaciones han resaltado la importancia de contar con estrategias de mantenimiento efectivas en las operaciones industriales, mostrando una correlación positiva entre los indicadores de gestión del mantenimiento preventivo y la disponibilidad de equipos en Perú. Para una detección más precisa de fallos, los patrones de comportamiento pueden ser identificados mediante técnicas de minería de datos aplicadas en plantas de fabricación a nivel global (Navarro, 2022). La capacidad de prevenir fallos antes de que ocurran permite a las empresas reducir tiempos de inactividad inesperados, incrementar la eficiencia de las operaciones y, en última instancia, lograr una ventaja competitiva.

En el contexto local, los sistemas de mantenimiento predictivo emplearon tecnologías de la Industria 4.0 para supervisar la salud de los activos y gestionar los KPIs de mantenimiento (Guerroum et al., 2022). Además, de utilizar enfoques de ingeniería basados en modelos, como SysML, para diseñar soluciones personalizadas que respondan a las necesidades de los interesados. Estos enfoques facilitan prácticas de gestión proactivas y optimizan el rendimiento del mantenimiento. La competitividad y rentabilidad de las empresas industriales dependen en gran medida de su eficiencia operativa y de la reducción de tiempos de inactividad. En una empresa dedicada principalmente a los servicios de pilado de arroz y derivados, el mantenimiento predictivo de compresores industriales se presenta como una necesidad urgente para asegurar la continuidad y eficiencia del proceso productivo. Los fallos imprevistos en los compresores pueden causar interrupciones significativas en la producción, lo que incrementa los costos operativos y reduce la eficiencia general de la planta. La implementación de un sistema de mantenimiento predictivo basado en algoritmos de Machine Learning surge como una solución innovadora para anticipar estos fallos y optimizar las estrategias de mantenimiento, mejorando así la disponibilidad y el rendimiento de los equipos críticos.

La incorporación de tecnologías innovadoras, como el mantenimiento predictivo basado en Machine Learning, proporciona soluciones efectivas para anticipar y prevenir fallos en equipos críticos, mejorando la eficiencia y la rentabilidad. La necesidad de optimizar las estrategias de mantenimiento en el sector industrial es cada vez más evidente debido a la competencia global y la presión por reducir costos. Este sistema tenía como objetivo principal aumentar la eficacia operativa y disminuir los tiempos de inactividad no planificados en una empresa situada en la región del Departamento de La Libertad. No solo esta empresa observó un beneficio de la implementación de esta tecnología, sino que también como modelo para otras industrias que enfrenten dificultades similares en cuanto al mantenimiento de equipos. El objetivo general de la presente investigación fue desarrollar un sistema de mantenimiento predictivo para compresores industriales utilizando algoritmos de machine learning.

Al optimizar la eficacia operativa y reducir los tiempos de inactividad no planificados, el sistema no solo mejora la rentabilidad y la eficiencia de los procesos industriales, sino que también promueve un uso más responsable y eficiente de los recursos. Este enfoque reduce el desperdicio asociado con fallos imprevistos y mantenimiento reactivo, alineándose con los Objetivos de Desarrollo Sostenible (ODS) dados por la Organización de las Naciones Unidas (ONU, 2024) al contribuir a ODS 9: Industria, Innovación e Infraestructura mediante la implementación de tecnologías innovadoras; ODS 12: Producción y Consumo Responsables al reducir significativamente la generación de desechos; y ODS 13: Acción por el Clima al minimizar el impacto ambiental y fomentar prácticas empresariales más responsables y resilientes.

En la misma línea anterior, se plantearon el problema general ¿Cómo plantear un sistema de mantenimiento predictivo utilizando algoritmos de machine learning para resolver la inactividad productiva causada por reparaciones de maquinaria en el sector industrial?, Y respecto a los específicos ¿Cuál es la influencia de variables como temperatura, presión, vibración y humedad en la ocurrencia de fallos, y cómo se puede determinar esta influencia mediante análisis estadísticos y visualizaciones gráficas? y ¿Cómo ajustar el punto de corte del modelo para mejorar el desempeño de la regresión logística y del modelo de árbol de decisión en la identificación de fallos, y cómo puede esto mejorar la efectividad del sistema de mantenimiento preventivo?

La justificación teórica de esta investigación se fundamenta en la creciente necesidad de mejorar la eficacia operativa y reducir los tiempos de inactividad no planificados en el sector industrial. La integración de sistemas de mantenimiento predictivo basados en algoritmos de Machine Learning representa un avance significativo en el manejo de equipos industriales, ofreciendo soluciones basadas en datos que permiten anticipar fallos y optimizar las estrategias de mantenimiento.

La justificación metodológica de la investigación se basa en el desarrollo y la implementación de un sistema de mantenimiento predictivo para compresores industriales utilizando algoritmos de machine learning. Este enfoque metodológico permite una evaluación precisa del estado de los equipos y la predicción de fallos antes de que ocurran, utilizando modelos predictivos y técnicas de análisis de datos.

Con respecto a la justificación social, la investigación tiene un impacto significativo en la sostenibilidad y la responsabilidad empresarial. Al reducir los tiempos de inactividad no planificados y mejorar la eficiencia operativa, el sistema propuesto contribuye a un uso más responsable y eficiente de los recursos industriales. Esto no solo beneficia a la empresa implementadora al mejorar su rentabilidad y reducir costos, sino que también establece un modelo replicable para otras industrias enfrentando desafíos similares.

Asimismo, el objetivo principal de la investigación fue plantear un sistema de mantenimiento predictivo utilizando algoritmos de machine learning para solucionar la inactividad productiva causada por reparaciones de maquinaria en el sector industrial. Respecto a los específicos son: Determinar la influencia de variables como temperatura, presión, vibración y humedad en la ocurrencia de fallos, mediante análisis estadísticos y visualizaciones gráficas. Ajustar el punto de corte del modelo, el desempeño del modelo de regresión logística y del modelo de árbol de decisión en la identificación de fallos mejorando así la efectividad del sistema de mantenimiento preventivo.

En cuanto a las hipótesis, la hipótesis general es que la implementación de un sistema de mantenimiento predictivo basado en algoritmos de machine learning reduce la inactividad productiva causada por reparaciones de maquinaria en el sector industrial. Respecto a las específicas tenemos que Variables como temperatura, presión, vibración y humedad tienen una influencia significativa en la ocurrencia de

fallos en la maquinaria industrial. Ajustar el punto de corte del modelo de regresión logística y del modelo de árbol de decisión optimiza la precisión en la identificación de fallos, al reducir tanto los falsos positivos como los falsos negativos.

Como resultado, se abordaron antecedentes nacionales e internacionales para el marco teórico del estudio:

Castro et al. (2022) centran su investigación en diversas empresas de mantenimiento de maquinaria pesada peruanas que registran eventos relacionados con indicadores de mantenimiento, pero no aprovechan estos datos de manera eficiente para generar patrones operativos mediante aprendizaje automático. Este artículo tuvo como objetivo desarrollar una herramienta basada en algoritmos de aprendizaje automático para predecir la ubicación de fallas en excavadoras hidráulicas, mejorando así la gestión del servicio de mantenimiento. Se creó un modelo predictivo utilizando el algoritmo de ensamble de árboles embolsados (bagged trees), que demostró una precisión del 97,15%, una especificidad del 99,04%, una exactitud del 98,56% y una sensibilidad del 97,12%. Los resultados destacaron que el modelo es muy eficaz en identificar fallas en flotas de excavadoras hidráulicas, mejorando la planificación y disponibilidad de suministros o componentes del mantenimiento, y optimizando la continuidad y capacidad de respuesta en el proceso de mantenimiento.

En su investigación, Guerroum et al. (2022) integraron eficazmente las tecnologías de la Industria 4.0 y gestionaron mejor las máquinas mineras mediante los KPIs de mantenimiento, esencial abordados en los problemas de disponibilidad de información. Esta investigación presentó el diseño del sistema de mantenimiento predictivo (PdMSys) utilizando la ingeniería de sistemas basada en modelos y el lenguaje de modelado SysML, lo que permitió cumplir con las necesidades y expectativas de los interesados en la industria minera. PdMSys es un sistema personalizado diseñado para tres niveles jerárquicos de usuarios, recopilando datos de sensores de campo, parámetros de las máquinas y otras bases de datos del sistema para monitorear el estado de salud de los activos y gestionar eficazmente los KPIs de mantenimiento en la mina de mineral. PdMSys transformaron las prácticas tradicionales de la empresa en digitales y proactivas, ofreciendo beneficios financieros y demostraron resiliencia para manejar la escalabilidad y la interoperabilidad.

Ramos et al. (2020) analizaron el sector metalmeccánico de Mosquera, Cundinamarca, destacando la empresa EQUIACEROS SAS. La investigación se centró en variables como el análisis de fallas históricas, principios operativos, tiempos y contextos de operación. El objetivo fue implementar un modelo de mantenimiento preventivo y predictivo, integrando RCM (Mantenimiento Centrado en Confiabilidad) para mejorar la calidad, productividad y reducir costos. La metodología incluyó AMEF (Análisis de Modo y Efecto de Fallas), clasificación y jerarquización de equipos, codificación de fallas y cálculo del NPR (número de prioridad de riesgo), para establecer actividades de mantenimiento precisas. Los resultados indicaron que el plan de mantenimiento permite evaluar la vida útil de los equipos y mejorar los indicadores de rendimiento, beneficiando a EQUIACEROS SAS. Concluyeron que, este modelo es vital para mantener la maquinaria en funcionamiento continuo, cumpliendo con los altos estándares de producción de la industria metalmeccánica. Además, se implementó un sistema de Seguridad en el Trabajo basado en la Norma ISO 45001:2018 y las regulaciones colombianas, contribuyendo a futuras investigaciones en gestión de mantenimiento y mejorando los indicadores de fiabilidad.

Ayvaz y Alpay (2021) desarrollaron en su estudio un sistema de mantenimiento predictivo basado en datos para líneas de producción en la industria manufacturera, utilizando datos en tiempo real de sensores IoT. El objetivo fue identificar señales de posibles fallos antes de que ocurrieran mediante técnicas de machine learning, permitiendo a los operadores tomar medidas preventivas para evitar paradas en la producción. La investigación evaluó la efectividad del sistema con datos reales de manufactura, revelando que el sistema fue eficaz en detectar indicadores de fallos potenciales y en la prevención de paradas de producción. Los resultados comparativos mostraron que los modelos de Random Forest y XGBoost, en comparación con algoritmos individuales, ofrecieron un mejor rendimiento. Los modelos de machine learning más efectivos fueron implementados en el sistema de producción de la fábrica, confirmando su utilidad.

Según Leukel (2021), en su estudio sobre la predicción de fallas había consistido en anticipar si un sistema material específico fallaría en un momento determinado en el futuro. El propósito de la revisión había sido evaluar cómo se había adoptado la tecnología de aprendizaje automático (ML) para la predicción de fallas en el mantenimiento industrial y sintetizar los resultados obtenidos. Se había llevado a

cabo una búsqueda sistemática de estudios experimentales en revistas revisadas por pares publicadas entre 2012 y 2020. Se habían examinado un total de 1,024 artículos, de los cuales 34 habían cumplido con los criterios de inclusión. La revisión se había centrado en comprender los conjuntos de datos analizados, el preprocesamiento para la generación de características, y el entrenamiento y evaluación de los modelos de predicción. Los resultados habían mostrado una amplia variedad de sistemas y dominios tratados, la adopción de enfoques actualizados para el preprocesamiento y el entrenamiento, una cierta falta de evaluación del desempeño que había limitado el problema del sobreajuste y una notable heterogeneidad en la presentación de los diseños y resultados experimentales. Se habían identificado oportunidades para futuras investigaciones y se habían sugerido formas de facilitar la comparación e integración de la evidencia obtenida de estudios individuales.

En su artículo, Torim et al. (2022) examinaron la transición de varios sectores industriales hacia la Revolución Industrial 4.0, subrayando la importancia del Internet de las Cosas y el mantenimiento predictivo como pilares fundamentales. El mantenimiento predictivo, se basó en la monitorización continua para anticipar fallos en el equipo y permitir intervenciones preventivas. El objetivo del estudio fue presentar un sistema inteligente de monitoreo y diagnóstico que pueda alertar a los operadores antes de que ocurran fallos, previniendo así daños materiales y ambientales. La innovación principal radicó en una arquitectura flexible para el mantenimiento predictivo, basada en patrones de software que combinan conocimientos expertos sobre fallos y técnicas de detección de anomalías como el Matrix Profile. Esta arquitectura dividió el sistema en una capa de reglas/sistema experto y una capa de indicadores/aprendizaje automático, facilitando la evolución independiente y flexible de ambos componentes. También se proponen tecnologías específicas para implementar esta arquitectura, como la plataforma de observabilidad Grafana y el marco Stumpy de Python para el Matrix Profile. La aplicación de patrones de reglas y observación en esta arquitectura de mantenimiento predictivo representaron una contribución novedosa.

En su investigación, Li et al (2024) analizan los diferentes métodos de Deep learning. El propósito de este estudio es sintetizar las estructuras de aprendizaje profundo utilizadas en el mantenimiento predictivo y ofrecer recomendaciones para que tanto investigadores como profesionales elijan la arquitectura más apropiada de

acuerdo a su entorno. Se llevó a cabo un análisis de 106 artículos primarios utilizando la base de datos Scopus y una búsqueda específica de artículos publicados entre el 1 de enero de 2017 y el 16 de abril de 2023. Como resultado, se seleccionaron finalmente aquellos artículos que abordan las ventajas y limitaciones de los métodos de aprendizaje profundo. Se clasificaron cinco arquitecturas principales en el estudio, a saber, redes neuronales profundas (DNNs), redes neuronales convolucionales (CNNs), codificadores auto-organizados apilados (SAEs), redes neuronales bayesianas profundas (DBNs) y redes neuronales recurrentes profundas (DRNNs). Estas arquitecturas fueron evaluadas utilizando métricas de rendimiento tales como el error cuadrático medio (MSE), la precisión y la pérdida de reconstrucción. Los resultados evidencian que cada estructura arquitectónica presenta ventajas y limitaciones particulares, lo que ofrece pautas para elegir las en función del tipo de datos y la tarea a realizar. Se concluye que las estrategias de mantenimiento predictivo fundamentadas en el aprendizaje profundo tienen la capacidad de aumentar la confiabilidad de los equipos industriales y prolongar su durabilidad, lo que supone un avance eficaz con respecto a los métodos convencionales de análisis de señales.

Shandookh et al. (2024) estudian la mejora de la fiabilidad de los sistemas mecánicos a través del mantenimiento predictivo, el cual tiene como objetivo prevenir la fatiga y el fallo en dichos sistemas. En el presente contexto, el estudio analizó información de vibración en el dominio temporal con el fin de optimizar la detección de anomalías en el ámbito del mantenimiento predictivo. Emplearon un sistema de transmisión por correa que presentaba velocidades de rotación operativas que variaban entre 500 y 2000 RPM, pretensiones de correa de 70 y 150 N, y se consideraron tres situaciones operativas distintas (saludable, defectuosa y desequilibrada), lo que resultó en un total de 12 casos analizados. Implementando dos acelerómetros piezoeléctricos uniaxiales con el fin de registrar las señales de vibración en las proximidades del motor y la polea. Durante el procesamiento de señales, calcularon cinco estadísticas avanzadas: la varianza, la desviación media absoluta (MAD), la tasa de cruce por cero (ZCR), el coeficiente de autocorrelación y la energía de la señal. Con la metodología Taguchi evaluaron las cinco características elegidas a través del análisis de la relación Señal/Ruido (S/N). En la clasificación, emplearon un sistema experto basado en inteligencia artificial. Entrenaron un modelo

de Bosque Aleatorio (RF) con parámetros no convencionales con el fin de mejorar la precisión. Los resultados obtenidos revelaron una precisión del 0.990 y un área bajo la curva (AUC) de 0.999, lo cual evidencia la elevada confiabilidad del modelo de bosques aleatorios (RF). La metodología resaltó el potencial de las características al ser integradas en sistemas expertos, lo que impulsa el avance de las estrategias de mantenimiento predictivo para sistemas de transmisión por correa.

De manera consiguiente, se aborda la teoría base para la investigación, el Machine Learning, una disciplina perteneciente al campo de la inteligencia artificial posibilita que las computadoras adquieran conocimiento a partir de datos y experiencias, prescindiendo de la programación explícita (Carbonell et al., 1983; Shaveta, 2023). Shaveta (2023) menciona la inclusión de enfoques como el aprendizaje supervisado, no supervisado y por refuerzo. Los algoritmos de Machine Learning se perfeccionan de forma adaptativa a medida que aumenta la cantidad de datos procesados. Estos algoritmos resultan especialmente beneficiosos en contextos donde la creación de algoritmos explícitos presenta dificultades (Raj, 2019; Shaveta, 2023). El área en cuestión se fundamenta en disciplinas como la informática, la estadística y la ciencia cognitiva (Raj, 2019). Las aplicaciones de este enfoque son diversas, incluyendo la adquisición de conocimientos declarativos, el desarrollo de habilidades motoras y cognitivas, la organización de conocimientos en representaciones efectivas, y el descubrimiento de nuevos hechos y teorías a través de la observación y la experimentación (Carbonell et al., 1983). Este proceso, como un campo en constante evolución, continúa generando un considerable interés en la investigación y se ha consolidado como un elemento fundamental en los sistemas contemporáneos de inteligencia artificial (Lewis & Denning, 2018).

El mantenimiento predictivo es una estrategia de gestión de activos que, a través de monitoreo en tiempo real y análisis de datos avanzados, predice y previene fallos en equipos. Utilizando sensores y software especializado, evalúa continuamente el estado de los activos para identificar anomalías y patrones que requieren intervención. Esta metodología ofrece beneficios como reducción de costos, prolongación de la vida útil de los equipos, mayor seguridad operativa y optimización de recursos. De este modo, se logra una gestión más eficiente y proactiva de los activos industriales (Arafat et al, 2023; chen et al, 2023; Siraskar et al, 2023;

Vanderschueren et al, 2023; Hurtado et al, 2023; Pagano y Davide 2023; Tasci et al, 2023; Zhuang et al, 2023; Orosnjak et al, 2023).

En la presente investigación se toma como variable dependiente al compresor industrial, donde se dimensiona por su estado operativo para determinar la presencia de falla o de no falla. El compresor industrial es un dispositivo mecánico utilizado para aumentar la presión del aire o gas mediante la reducción de su volumen, con aplicaciones en diversas industrias (Fernández, s.f.).

Como variable independiente están las condiciones operativas del compresor industrial, que es el conjunto de parámetros o variables ambientales y mecánicas que afectan el funcionamiento del compresor industrial, como temperatura, presión, vibración, humedad y tiempos de inactividad (Pascual, 2019).

La temperatura, una magnitud física fundamental, se define dentro del marco de la termodinámica y la mecánica estadística. Su medición se basa en una variedad de fenómenos físicos (De Boer, 1965). Esto abarca desde la clasificación de los medios por McLuhan como "calientes" o "fríos" hasta los análisis de la transmisión de información inspirados en la termodinámica (Starosielski, 2014).

Según Sommeria (2020), la presión se refiere a la cantidad de resistencia que enfrenta el flujo de aire producido por un compresor. Este parámetro se expresa en PSI (libras por pulgada cuadrada), unidad que representa la fuerza ejercida de una libra sobre un área de una pulgada cuadrada. En términos técnicos, la presión es el resultado de aplicar una libra de fuerza sobre una superficie con estas dimensiones, lo que permite cuantificar la intensidad con la que el aire es comprimido y distribuido dentro del sistema. Este valor es esencial para evaluar el rendimiento y la eficiencia de los equipos neumáticos en diversas aplicaciones industriales.

La vibración, según Manakova et al. (2019) y Dumitrescu y Dumitrescu (2010), se caracteriza como el movimiento oscilatorio de un cuerpo sólido bajo la influencia de fuerzas. Dependiendo de la fuerza impulsora y la estructura, el fenómeno puede manifestarse como continuo o intermitente, periódico o no periódico

Gaffney (1978) define la humedad como la presencia de vapor de agua en la atmósfera, que puede medirse en términos de cantidad absoluta o cantidad relativa.

El concepto de humedad relativa (HR) se aplica comúnmente en diversas disciplinas, aunque no tiene una definición estandarizada. Feistel y Lovell (2017) han propuesto una definición novedosa y completa basada en la termodinámica de actividades para explicar las desviaciones del comportamiento de gases ideales.

Los tiempos de inactividad o tiempos muertos según Garcés y Castrillón (2017) se definen como períodos en los que una línea de producción o máquina permanece inactiva debido a fallos, lo que ocasiona pérdidas de producción y altos costos de mantenimiento. Estos tiempos inactivos surgen de componentes críticos y fallos, y su identificación es crucial para mejorar la eficiencia operativa. Para abordarlos, se utilizan herramientas como el análisis de Pareto, la teoría de colas, y la simulación de eventos discretos, además de técnicas avanzadas como la minería de datos.

II. METODOLOGÍA

El presente estudio tuvo como metodología de tipo aplicada, la investigación aplicada tiene como objetivo generar conocimiento que pueda ser utilizado de manera directa para resolver problemas específicos en la sociedad o en el ámbito productivo. Se apoya principalmente en los descubrimientos tecnológicos provenientes de la investigación básica, enfocándose en conectar la teoría con la creación de productos o soluciones concretas (Lozada, 2014).

La presente investigación es cuantitativa, según Sánchez (2019) este enfoque se refiere al estudio en fenómenos medibles numéricamente, utilizando técnicas estadísticas para analizar datos con el objetivo de describir, explicar, predecir y controlar estos fenómenos. Se basa en la cuantificación rigurosa a lo largo de todo el proceso investigativo.

El diseño de investigación adoptado fue no experimental y correlacional. Este diseño se basó en el uso de datos históricos y operativos de los compresores, sin manipular las variables independientes, lo cual permitió observar y analizar las relaciones entre variables en un entorno real.

En cuanto a la variable y su operacionalización se describe en el siguiente orden:

Variable 1: Compresor industrial, es un dispositivo mecánico utilizado para aumentar la presión del aire o gas mediante la reducción de su volumen, con aplicaciones en diversas industrias (Fernández, s.f.).

El compresor industrial para su concepto operacional considera en estado operativo si está en funcionamiento y realizando la compresión de aire o gas según lo diseñado. El estado se clasifica en "falla" (1) si el compresor presenta un problema que afecta su funcionamiento, y "no falla" (0) si opera normalmente.

Variable 2: Condiciones operativas del compresor industrial, es el conjunto de parámetros o variables ambientales y mecánicas que afectan el funcionamiento del compresor industrial, como temperatura, presión, vibración, humedad y tiempos de inactividad (Pascual, 2019).

Las condiciones operativas del compresor industrial se operacionalizan en medición y monitoreo continuo de las variables de temperatura (en °C), presión (en PSI), vibración (en mm/s), humedad (en %) y tiempos de inactividad (en horas). Estos valores se registran en intervalos constantes para determinar el estado operativo del compresor.

Para determinar la población, muestra, técnicas de muestreo y la unidad de análisis, se siguió un esquema metodológico detallado que se presenta a continuación:

La población objeto de estudio en esta investigación estuvo compuesta por los compresores industriales utilizados en las operaciones de la empresa en el sector industrial de estudio. Estos compresores, son esenciales para el proceso de pilado de arroz. Para asegurar la eficacia de la investigación, se seleccionaron compresores que representaran adecuadamente la población objetivo y cumplieran con los criterios necesarios para un análisis detallado.

La selección de la muestra fue mediante el muestreo no probabilístico intencional, del cual se eligió a 1 compresor de aire industrial de 350 Lts – 7.5 HP de 2 etapas. según los criterios de inclusión y exclusión rigurosos. Los criterios de inclusión consideraron que el compresor debía tener datos históricos operativos de al menos 3 años y contar con la infraestructura necesaria para la instalación de sensores avanzados que permitieran la recolección de datos en tiempo real entre enero y febrero del 2024. Se excluyeron aquellos compresores que no cumplían con estos requisitos, como la falta de registros históricos suficientes o la imposibilidad de instalar los sensores necesarios.

La recopilación de datos fue proporcionada por el responsable del área de mantenimiento de la empresa privada del presente estudio, la base de datos se recopiló mediante la técnica de observación instrumentada, para Useche et al. (2019), se empleó una combinación de equipos mecánicos y electrónicos para capturar y analizar el fenómeno en cuestión. Estos dispositivos, que van desde instrumentos manuales para mediciones físicas hasta sensores digitales y sistemas de monitoreo, permiten una recolección precisa y detallada de datos. En el presente estudio la recopilación de datos se llevó a cabo mediante la instalación de sensores en el compresor industrial, tales como el sensor de presión (transductor de presión no

incendiario M3200), sensor de vibración (acelerómetro de 3 cables TO-5 modelo 805M1), sensor de temperatura (RTD – NB-PTCO-035) y sensor de humedad (FC-28: Sensor de humedad del suelo), los cuales registraron continuamente variables operativas claves como humedad, temperatura, vibración y presión. Estos sensores, diseñados para operar en entornos industriales exigentes, proporcionaron datos de alta fidelidad necesarios para el desarrollo de modelos predictivos.

La validación y análisis de consistencia interna de los instrumentos de recolección de datos se llevaron a cabo mediante calibraciones periódicas. Los resultados de estas calibraciones demostraron una alta consistencia interna de los datos recolectados, validando así la fiabilidad de los sensores utilizados.

En cuanto a los métodos para el análisis de datos, se describen las metodologías empleadas para desarrollar modelos de Machine Learning (ver anexos) con el objetivo de detectar fallas en sistemas mecánicos, utilizando las variables de condiciones operativas como predictores (Soares, et al. 2022; San Miguel, 2021). Estas variables incluyen temperatura, humedad, vibración y presión, las cuales son monitoreadas continuamente para evaluar el estado de los equipos. El enfoque metodológico adoptado se divide en dos fases principales.

En la primera fase, se realiza una prueba de diferencia de medias con el propósito de determinar si los niveles de las variables de condiciones operativas varían de manera significativa en presencia de fallas. Este análisis preliminar es fundamental, ya que permite identificar patrones en los datos que puedan estar asociados con eventos de falla, lo cual es crucial para la construcción de modelos predictivos robustos.

En la segunda fase, se procede al desarrollo de modelos de machine learning, diseñados para predecir la ocurrencia de fallas basándose en los datos históricos de las condiciones operativas. Estos modelos serán entrenados y evaluados utilizando un conjunto de datos representativo, con el fin de asegurar su precisión y eficacia en la predicción de fallas en escenarios reales. La metodología aquí descrita no solo busca mejorar la eficiencia del mantenimiento predictivo, sino también reducir los costos asociados a las fallas inesperadas y optimizar la disponibilidad operativa de los equipos.

La prueba de diferencia de medias es una herramienta estadística utilizada para determinar si existe una diferencia significativa entre las medias de dos grupos de datos. En el contexto de este estudio, se emplea para comparar los niveles de las variables de condiciones operativas (temperatura, humedad, vibración y presión) durante los periodos de operación normal y los periodos en los que se presentan fallas. Esta prueba es fundamental para establecer si las condiciones operativas muestran variaciones significativas cuando ocurre una falla, lo que a su vez justifica su uso como predictores en los modelos de machine learning.

La prueba de diferencia de medias se basa en la formulación de dos hipótesis:

Hipótesis nula (H_0): No existe una diferencia significativa entre las medias de las variables de condiciones operativas durante los periodos de operación normal y los periodos con fallas. Matemáticamente, esto se expresa como:

$$H_0: \mu_1 = \mu_2$$

Donde μ_1 y μ_2 representan las medias de las variables de condiciones operativas en los periodos de operación normal y con fallas, respectivamente.

Hipótesis alternativa (H_1): Existe una diferencia significativa entre las medias de las variables de condiciones operativas durante los periodos de operación normal y los periodos con fallas. Esto se expresa como:

$$H_1: \mu_1 \neq \mu_2$$

Para evaluar la hipótesis planteada, se utiliza el estadístico t, que se calcula bajo el supuesto de que las varianzas de los dos grupos son iguales. El estadístico t se define como:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Donde \bar{X}_1 y \bar{X}_2 son las medias muestrales de los períodos de operación normal y con fallas, respectivamente, n_1 y n_2 son los tamaños de las muestras de los dos grupos, y S_p es la desviación estándar combinada de las dos muestras, calculada como:

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Donde S_1^2 y S_2^2 son las varianzas muestrales de los dos grupos.

Una vez calculado el valor de t , se compara con el valor crítico de t correspondiente al nivel de significancia deseado (α) y los grados de libertad $df = n_1 + n_2 - 2$. Si el valor absoluto de t calculado es mayor que el valor crítico, se rechaza la hipótesis nula, lo que indica que existe una diferencia significativa entre las medias de las variables de condiciones operativas en los periodos de operación normal y con fallas.

Un modelo de clasificación es un tipo de algoritmo de machine learning diseñado para asignar etiquetas a datos nuevos basándose en las características aprendidas a partir de un conjunto de datos de entrenamiento. En el contexto del mantenimiento predictivo, un modelo de clasificación se utiliza para predecir si un equipo o sistema está en riesgo de fallar, basándose en las variables de condiciones operativas como temperatura, humedad, vibración y presión. Estas predicciones permiten a los ingenieros tomar decisiones informadas sobre cuándo realizar mantenimientos preventivos o correctivos, mejorando así la eficiencia y reduciendo el tiempo de inactividad.

La matriz de confusión es una herramienta fundamental en la evaluación de un modelo de clasificación, ya que muestra las predicciones realizadas por el modelo comparadas con los resultados reales. La matriz de confusión se organiza en una tabla cuadrada donde las filas representan las clases reales y las columnas representan las clases predichas por el modelo. A continuación, se muestra la matriz de confusión:

	Predicción Normal	Predicción Falla
Real Normal	TN	FN
Real Falla	FP	TD

La estructura general de una matriz de confusión binaria, en la que las clases posibles son "falla" (positiva) y "no falla" (negativa), se desglosa en cuatro categorías principales: los Verdaderos Positivos (TP) representan el número de casos en los que el modelo predice correctamente una falla; los Falsos Positivos (FP) indican los casos en los que el modelo predice una falla que en realidad no ocurre; los Falsos Negativos (FN) corresponden a los casos en los que el modelo no predice una falla que sí sucede; y los Verdaderos Negativos (TN) son los casos en los que el modelo predice correctamente la ausencia de falla.

Respecto a Métricas de Precisión, la eficacia de un modelo de clasificación se evalúa utilizando varias métricas de precisión derivadas de la matriz de confusión:

Precisión (Precision): La precisión es la proporción de predicciones correctas de fallas entre todas las predicciones positivas realizadas por el modelo. Se calcula como:

$$Precisión = \frac{TP}{TP + FP}$$

Exhaustividad (Recall): También conocida como sensibilidad o tasa de verdaderos positivos, la exhaustividad mide la capacidad del modelo para identificar correctamente todas las fallas reales. Se calcula como:

$$Exhaustividad = \frac{TP}{TP + FN}$$

Exactitud (Accuracy): La exactitud es la proporción de todas las predicciones correctas (tanto positivas como negativas) respecto al total de casos. Se calcula como:

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN}$$

La Curva ROC (Receiver Operating Characteristic) es una representación gráfica que evalúa el rendimiento de un modelo de clasificación para diferentes umbrales de decisión. En esta curva, se grafica la tasa de verdaderos positivos (también conocida como sensibilidad o recall) en función de la tasa de falsos positivos (1 - especificidad) a medida que se varía el umbral utilizado por el modelo para determinar las predicciones.

La ROC permite observar la relación entre la sensibilidad del modelo y la probabilidad de emitir falsas alarmas, lo que es crucial en contextos donde se debe equilibrar la detección de fallas reales con la minimización de alarmas incorrectas. Un modelo que se acerque al punto (0, 1) en la gráfica ROC es ideal, ya que este punto representa una sensibilidad perfecta y una tasa de falsos positivos nula.

El área bajo la curva ROC, conocida como AUC (Area Under the Curve), proporciona una medida cuantitativa del rendimiento general del modelo. Un AUC cercano a 1 indica que el modelo tiene una excelente capacidad discriminativa, siendo capaz de distinguir entre clases positivas (fallas) y negativas (no fallas) con alta precisión. Un AUC de 0.5, por otro lado, sugiere que el modelo no tiene capacidad discriminativa y su rendimiento es equivalente a una clasificación aleatoria.

La curva ROC y el AUC son especialmente útiles en problemas de clasificación desequilibrada, donde las clases no están igualmente representadas en los datos. Estas herramientas permiten evaluar el modelo de manera más holística, ayudando a los ingenieros a seleccionar el umbral óptimo que maximiza la identificación de fallas reales mientras minimiza las falsas alarmas.

El sobreajuste (overfitting) es un problema común en el desarrollo de modelos de machine learning, donde el modelo se ajusta demasiado a los datos de entrenamiento, capturando el ruido o las anomalías específicas de ese conjunto. Esto puede resultar en un excelente rendimiento en el conjunto de entrenamiento, pero un desempeño deficiente en datos nuevos o no vistos, ya que el modelo pierde capacidad de generalización.

Para mitigar el riesgo de sobreajuste, se utiliza una técnica fundamental: la partición de datos en conjuntos de entrenamiento y prueba. Esta técnica asegura que el modelo sea evaluado de manera justa y pueda generalizar bien a datos no vistos.

El proceso de partición de datos comienza con la división en conjuntos, donde el conjunto de entrenamiento se utiliza para entrenar el modelo, permitiendo que el algoritmo aprenda las relaciones entre características y etiquetas, mientras que el conjunto de prueba se reserva para evaluar el rendimiento del modelo tras el entrenamiento, proporcionando una medida de su desempeño con datos nuevos. En la partición simple, los datos se dividen aleatoriamente en conjuntos de entrenamiento y prueba, comúnmente en proporciones como 70%/30% o 80%/20%, lo que permite

una evaluación justa del modelo en datos que no ha visto durante el entrenamiento y ofrece una evaluación precisa de su capacidad de generalización. Durante la evaluación del rendimiento, se calculan métricas como precisión, recall, exactitud y AUC utilizando el conjunto de prueba para determinar la efectividad del modelo en datos no vistos, y se realizan ajustes en el modelo basado en su desempeño para mejorar su capacidad de generalización y evitar el sobreajuste. Implementar una partición adecuada de los datos es crucial para identificar el sobreajuste y proporcionar una evaluación precisa del rendimiento del modelo en condiciones prácticas.

Los algoritmos de clasificación son el núcleo de las técnicas de machine learning supervisado, diseñados para asignar una clase o categoría a cada observación dentro de un conjunto de datos. Estos algoritmos aprenden de datos etiquetados, es decir, datos en los que las clases correctas ya están definidas, y utilizan esta información para hacer predicciones sobre nuevas observaciones.

En el contexto del mantenimiento predictivo, los algoritmos de clasificación juegan un papel crucial al permitir la detección anticipada de fallas en sistemas o equipos basándose en las condiciones operativas observadas. La selección del algoritmo adecuado depende de múltiples factores, como la naturaleza de los datos, la complejidad del problema, y la necesidad de interpretabilidad del modelo.

A continuación, se explorarán algunos de los algoritmos de clasificación más utilizados en la industria, cada uno con sus propias características, ventajas y limitaciones. Entre ellos se incluyen el modelo Logit, que es ampliamente utilizado en la regresión logística para problemas binarios; el Árbol de Clasificación, que ofrece una estructura jerárquica fácil de interpretar para la toma de decisiones; y el Random Forest, que combina múltiples árboles de decisión para mejorar la precisión y la robustez del modelo.

El Modelo Logit, también conocido como regresión logística, es un método estadístico utilizado para modelar la probabilidad de ocurrencia de un evento binario (como falla/no falla) en función de una o más variables independientes. Este modelo es particularmente útil en contextos donde el resultado que se desea predecir es categórico y las relaciones entre las variables independientes y el resultado no son lineales.

En su forma más general, el Modelo Logit puede expresarse de la siguiente manera:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k$$

En el modelo, p representa la probabilidad de que ocurra el evento de interés, como una falla, mientras que $x_1, x_2 \dots x_k$ son las variables independientes que pueden incluir temperatura, humedad, vibración y presión. El término β_0 denota el intercepto del modelo, y los coeficientes $\beta_0, \beta_1, \dots, \beta_k$ indican la magnitud y la dirección del efecto de cada variable independiente sobre la probabilidad del evento.

La expresión matemática anterior es la transformación logit, que convierte la relación lineal entre las variables independientes y la probabilidad de que ocurra el evento en un valor que puede ser interpretado en términos de probabilidad. La transformación logit se define como el logaritmo del odds ratio, donde el odds ratio es la razón entre la probabilidad de que el evento ocurra y la probabilidad de que no ocurra.

Para comprender mejor el Modelo Logit, es útil considerar el caso de una sola variable independiente:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1$$

Este modelo con una sola variable puede ser visualizado mediante la función sigmoide, que es la función que convierte la combinación lineal de las variables independientes en una probabilidad entre 0 y 1:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1x_1)}}$$

La función sigmoide tiene una forma característica en "S", lo que refleja cómo pequeños cambios en la variable independiente cerca del punto medio (donde $p = 0.5$) pueden llevar a cambios significativos en la probabilidad predicha. Esta no linealidad es clave para capturar la relación compleja entre las condiciones operativas y la probabilidad de falla.

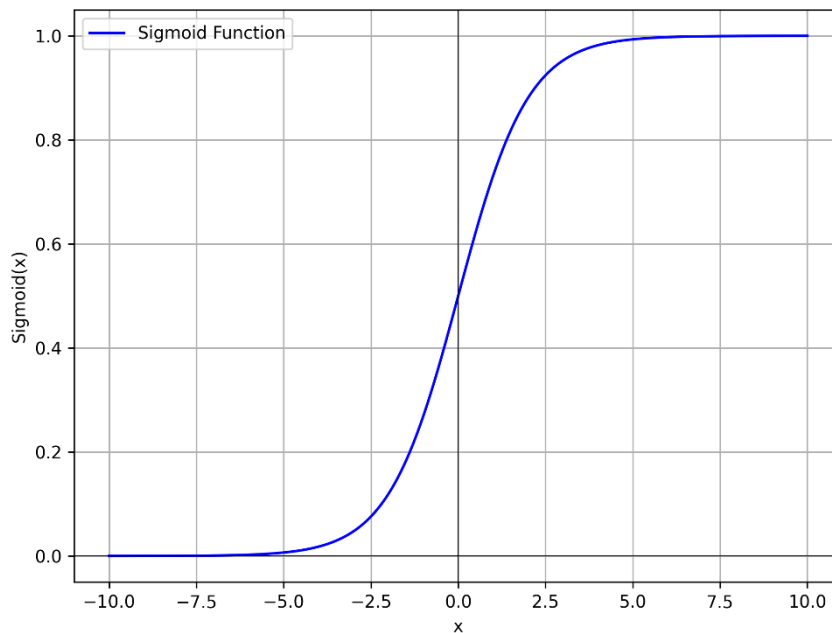


Figura 1. Función sigmoide

La interpretación de los coeficientes del Modelo Logit difiere de la regresión lineal. En lugar de representar un cambio absoluto en la variable dependiente, los coeficientes del Modelo Logit (β_i) representan el cambio en el logaritmo del odds ratio por cada unidad adicional en la variable independiente. En otras palabras, un coeficiente positivo indica que un aumento en la variable independiente está asociado con un aumento en la probabilidad de que ocurra el evento de interés, mientras que un coeficiente negativo sugiere una disminución en dicha probabilidad.

El Modelo de Árbol de Clasificación es una técnica comúnmente utilizada en el ámbito del machine learning supervisado para realizar predicciones categóricas, especialmente útil para tareas de clasificación binaria. En un árbol de clasificación, las decisiones se estructuran de manera jerárquica: cada nodo interno corresponde a una prueba sobre una variable predictora, y cada rama representa un resultado de dicha prueba. Las hojas finales del árbol indican la clase a la que se asigna la observación, en este caso, una de dos posibles clases, como "Falla" o "No Falla".

En la construcción de un árbol de clasificación binaria, el objetivo principal es dividir el conjunto de datos de manera que se maximice la pureza de las ramas, es decir, que cada rama contenga el máximo posible de ejemplos de una sola clase. Para lograr esto, se utilizan criterios de división como el índice Gini o la ganancia de información, los cuales miden la impureza o la incertidumbre dentro de un nodo.

Respecto al Índice Gini, en una clasificación binaria, el índice Gini mide la probabilidad de que una observación seleccionada al azar sea clasificada incorrectamente si se asigna a una clase en función de la distribución de la muestra en el nodo. Para un nodo que contiene dos clases, "Falla" y "No Falla", se calcula como:

$$\text{Índice Gini}(S) = 1 - (p_1)^2 - (p_2)^2$$

Donde p_1 es la proporción de observaciones de la clase "Falla" y p_2 es la proporción de observaciones de la clase "No Falla".

Ganancia de Información:

La ganancia de información mide la reducción en la incertidumbre sobre la clase de una observación cuando se realiza una partición en función de una variable predictora. Para una clasificación binaria, se calcula a partir de la entropía, la cual es máxima cuando las clases están equilibradas y mínima cuando una de las clases domina:

$$\text{Ganancia de Información}(S, A) = \text{Entropía}(S) - \sum_{v \in \text{Valores}(A)} \frac{|s_v|}{|S|} \text{Entropía}(s_v)$$

Donde A es la variable predictora utilizada para dividir el nodo, y s_v es el subconjunto de datos resultante de esa división.

En el contexto del mantenimiento predictivo, el Modelo de Random Forest se utiliza para mejorar la precisión en la detección de fallas en equipos mediante la combinación de múltiples árboles de decisión. Esta técnica es eficaz para manejar la complejidad y la variabilidad de las condiciones operativas que afectan el rendimiento de los equipos.

El proceso de construcción del modelo Random Forest se basa en tres etapas clave: Primero, se aplica el método de Bootstrap Aggregating (Bagging) para crear múltiples subconjuntos de datos mediante muestreo con reemplazo, utilizando cada subconjunto para entrenar un árbol de decisión independiente; este enfoque es fundamental en el mantenimiento predictivo, ya que permite al modelo captar diversos aspectos del comportamiento del equipo y minimizar el riesgo de sobreajuste a condiciones específicas. En la segunda etapa, se realiza una selección aleatoria de

variables operativas, como temperatura, humedad, vibración y presión, durante la construcción de cada árbol, lo que asegura que cada árbol examine diferentes aspectos de las condiciones del equipo, ofreciendo una visión más completa y robusta para predecir fallas. Finalmente, para la predicción de la probabilidad de falla, el modelo Random Forest agrega las predicciones de todos los árboles en el bosque mediante una votación mayoritaria en clasificación binaria (falla/no falla), donde la clase final se determina por la mayoría de los árboles, mejorando así la precisión de la detección de fallas y reduciendo los errores de predicción.

Aspectos éticos: La investigación sobre la implementación de un sistema de mantenimiento predictivo basado en algoritmos de Machine Learning en una empresa en el sector industrial, se llevó a cabo siguiendo estrictamente los principios de integridad científica. La transparencia, honestidad y responsabilidad fueron pilares fundamentales en todas las etapas del estudio. Para asegurar la integridad de los datos, se implementaron procedimientos rigurosos de recolección y almacenamiento, garantizando que los datos fueran precisos y no estuvieran sujetos a manipulación. Además, la interpretación de los resultados se realizó de manera imparcial, basada únicamente en la evidencia obtenida, sin sesgos ni influencias externas, siguiendo las mejores prácticas establecidas por la comunidad científica.

III. RESULTADOS

El periodo de recolección de datos se estableció desde el 1 de enero de 2024 a las 00:00 horas hasta el 10 de febrero de 2024 a las 15:00 horas, con intervalos de recolección de una hora. Durante este intervalo, se registraron variables operativas críticas, tales como temperatura, presión, vibración y humedad, las cuales son esenciales para el análisis del rendimiento de los compresores industriales.

Asimismo, los fallos fueron documentados en un formato binario, donde un valor de 1 indica la ocurrencia de una falla y un valor de 0 denota la ausencia de fallas. Este método de registro permite una identificación precisa de los eventos de fallo, facilitando el análisis posterior y la implementación de estrategias de mantenimiento predictivo. La distribución de los fallos en el conjunto de datos proporciona una visión inicial sobre la frecuencia y la naturaleza de los problemas operativos, lo que resulta fundamental para la optimización de los procesos de mantenimiento.

Objetivo específico 1

Determinar la influencia de variables como temperatura, presión, vibración y humedad en la ocurrencia de fallos, mediante análisis estadísticos y visualizaciones gráficas para solucionar la inactividad productiva causada por reparaciones de maquinaria en el sector industrial.

Los estadísticos descriptivos de las variables temperatura, presión, vibración y humedad, presentados en la Tabla 1, muestran que la temperatura promedio fue de 75,1 °C con una desviación estándar de 4,9 °C, reflejando una variabilidad moderada. La presión tuvo un promedio de 30,14 psi (libras por pulgada cuadrada) con una desviación estándar de 1,99 psi, indicando una menor dispersión en comparación con la temperatura. La vibración presentó un promedio de 0,5 mm/s (milímetros por segundo) y una desviación estándar de 0,1 mm/s, lo que sugiere baja variabilidad. La humedad, con un promedio de 49,79% y una desviación estándar de 10,27%, mostró una mayor dispersión. Estos resultados resaltan la importancia del monitoreo constante para mantener el control sobre las condiciones operativas y prevenir posibles fallos.

Tabla 1. Estadísticos descriptivos de las variables temperatura, presión, vibración y humedad.

Estadística	Temperatura	Presión	Vibración	Humedad
Promedio	75,1	30,14	0,5	49,79
Std	4,9	1,99	0,1	10,27
Min	58,79	24,12	0,2	20,71
25%	71,76	28,79	0,44	42,64
50%	75,13	30,13	0,5	49,99
75%	78,25	31,47	0,57	56,65
Max	94,26	36,39	0,89	82,43

Fuente: Elaboración propia

Los datos preliminares sobre la frecuencia de fallas, como se muestra en la Tabla 2, indican que el 87% de las observaciones corresponden a casos en los que no se registran fallas, mientras que el 13% restante presenta fallas. Estos resultados proporcionan una visión inicial de la distribución de las fallas en el conjunto de datos y servirán como base para el análisis posterior del modelo de clasificación, permitiendo evaluar su efectividad en la identificación de fallas.

Tabla 2. Frecuencia de fallas.

Categoría	Frecuencia	%
No falla	872	87%
falla	128	13%

Fuente: Elaboración propia

En el análisis de la variable temperatura, se identificaron valores atípicos en ambos extremos de la distribución, como se muestra en la Figura 2. La gráfica de cajas horizontal superior presenta estos valores atípicos, que se extienden más allá de los límites de los bigotes, indicando valores que se desvían significativamente de la tendencia central.

En respuesta a estos valores atípicos, se aplicó un proceso de truncamiento para eliminar estos extremos y ajustar el rango de la variable. La gráfica de cajas inferior, que muestra la distribución después del truncamiento, revela una distribución más centrada y sin los valores extremos. Este ajuste es fundamental para mejorar la

precisión del análisis posterior y asegurar que los modelos estadísticos y de clasificación no se vean sesgados por estos valores atípicos.

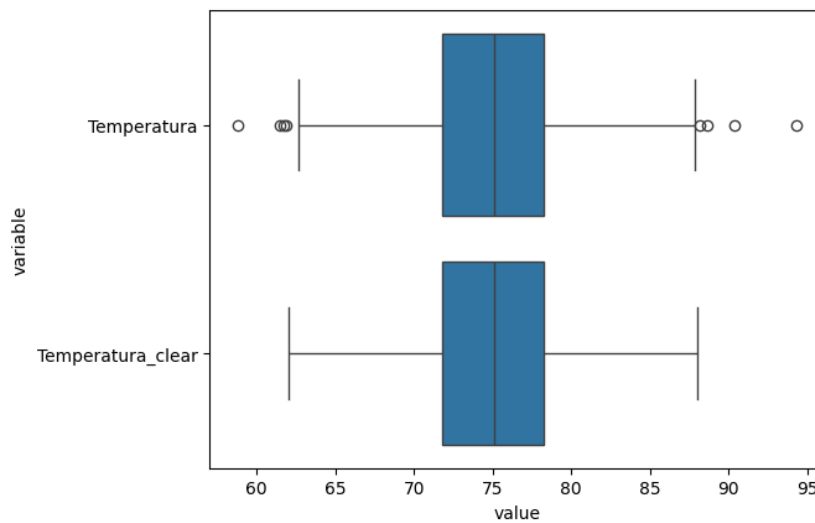


Figura 2. Gráfica de cajas de la temperatura: antes y después del truncamiento.

El análisis de la variable presión revela valores atípicos en ambos extremos, como se ilustra en la Figura 3. La gráfica de cajas superior muestra estos valores extremos, mientras que la gráfica inferior presenta la distribución tras truncar los valores atípicos. Este ajuste mejora la visualización de la distribución de la presión y minimiza el impacto de los valores extremos en el análisis posterior.

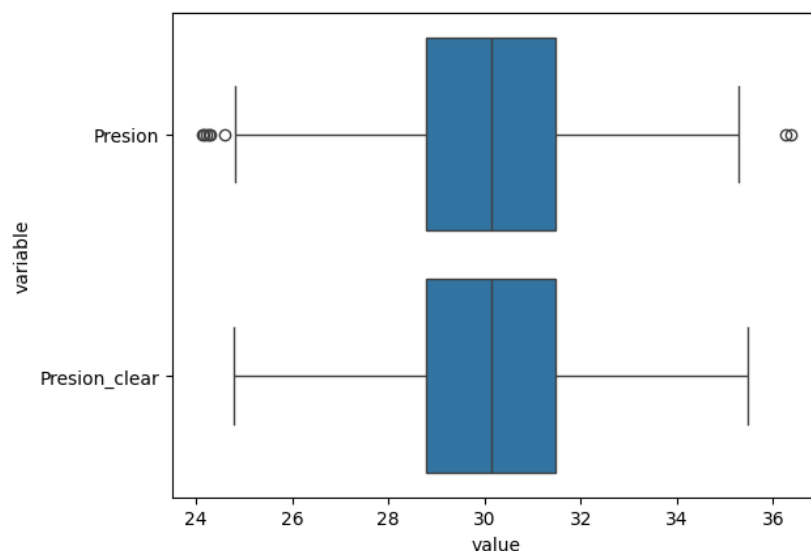


Figura 3. Gráfica de cajas de la presión: antes y después del truncamiento.

La variable vibración también muestra valores atípicos en ambos extremos, como se observa en la Figura 4. La gráfica de cajas superior destaca estos valores

extremos, mientras que la gráfica inferior muestra la distribución tras su eliminación. Este proceso de truncamiento facilita una evaluación más clara de la distribución de la vibración, sin la influencia de los valores atípicos.

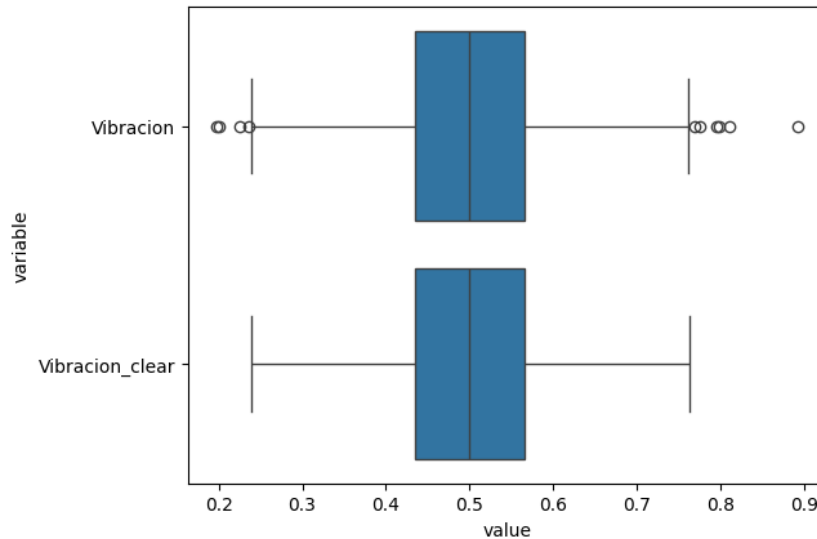


Figura 4. Gráfica de cajas de la vibración: antes y después del truncamiento.

En la variable humedad, se identificaron valores atípicos en ambos extremos, como se presenta en la Figura 5. La gráfica de cajas superior muestra estos valores extremos, mientras que la gráfica inferior ilustra la distribución después de su truncamiento. Este ajuste mejora la representación de la distribución de la humedad y reduce el impacto de los valores atípicos en el análisis.

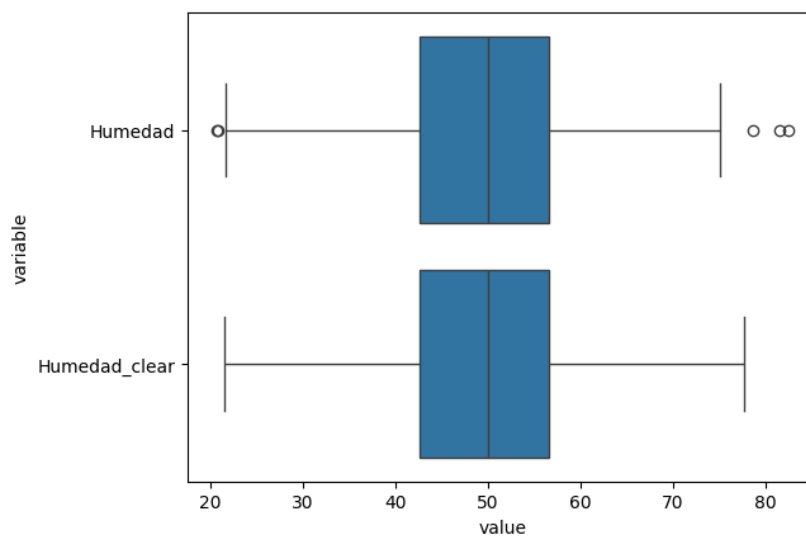


Figura 5. Gráfica de cajas de la humedad: antes y después del truncamiento.

La estandarización de los datos fue aplicada para asegurar que todas las variables tengan una escala comparable y normalizar su distribución. La Figura 6 muestra los gráficos de cajas para las variables estandarizadas. Como se observa, todas las variables presentan distribuciones simétricas y sin valores atípicos significativos, lo que indica que el proceso de estandarización ha sido efectivo en la normalización de los datos.

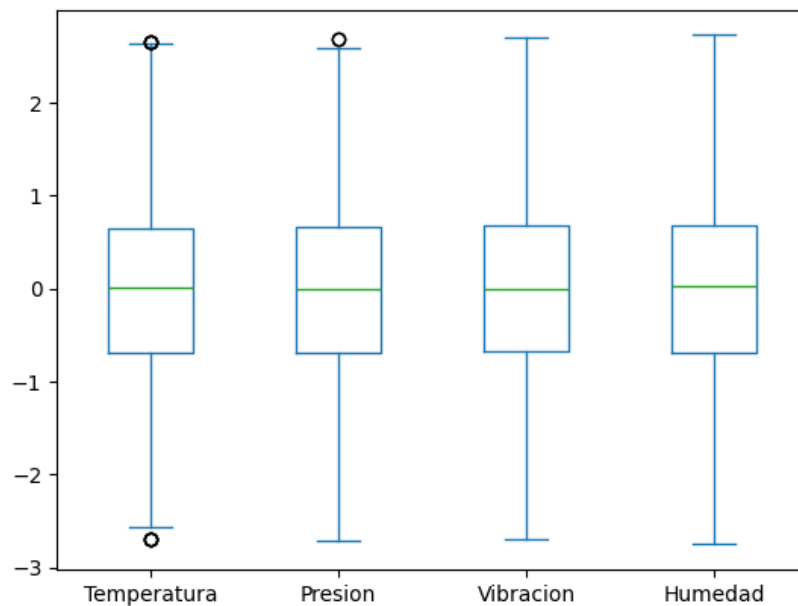


Figura 6. Gráfica de cajas de las variables estandarizadas.

La Tabla 3 presenta la cantidad de valores faltantes para cada una de las variables en el conjunto de datos. La variable de humedad tiene el mayor número de valores faltantes (6), seguida por temperatura (4), presión (3) y vibración (2). Estos valores faltantes serán gestionados mediante el proceso de imputación basado en el método KNN (K-Nearest Neighbors). Posteriormente, se incluirán gráficos de densidad para cada variable, mostrando su distribución antes y después de aplicar el proceso de imputación.

Tabla 3. Cantidad de valores faltantes por variable.

Variable	Valores faltantes
Temperatura	4
Presión	3
Vibración	2

Humedad	6
---------	---

Fuente: Elaboración propia

La Figura 7 muestra la comparación de las densidades de la variable temperatura antes y después de aplicar el proceso de imputación de valores faltantes mediante el método KNN. A pesar de que se imputaron algunos valores, la densidad de la temperatura se mantiene muy similar en ambas representaciones. Esto sugiere que el proceso de imputación no ha alterado significativamente la distribución original de la temperatura, manteniendo la integridad de la variable para el análisis posterior.

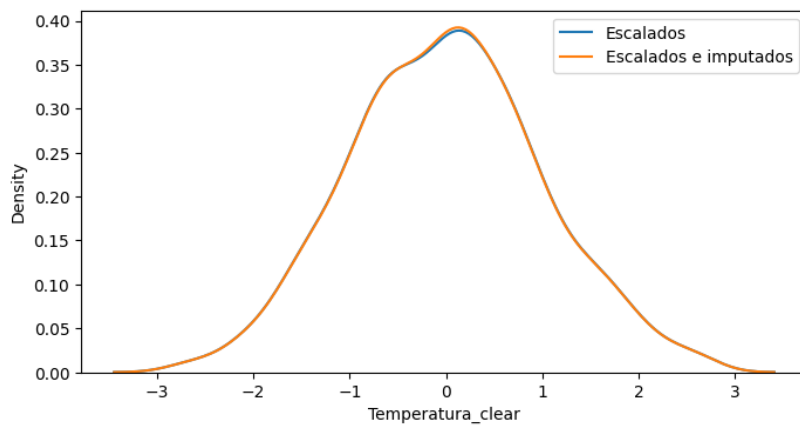


Figura 7. Comparación de las densidades de la temperatura antes y después de la imputación.

La Figura 8 presenta las densidades de la variable presión antes y después de la imputación de valores faltantes utilizando el método KNN. Los gráficos muestran que la distribución de la presión permanece prácticamente inalterada tras la imputación, lo que indica que el proceso ha sido efectivo sin modificar significativamente la forma original de la densidad.

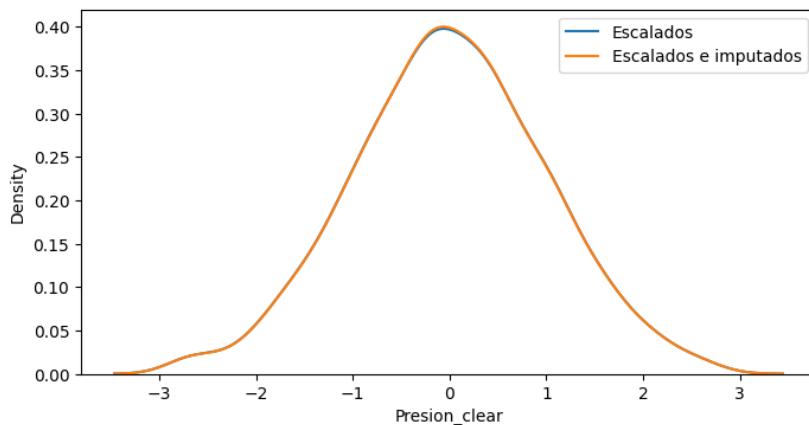


Figura 8. Comparación de las densidades de la presión antes y después de la imputación

La Figura 9 muestra las densidades de la variable vibración antes y después de la imputación de valores faltantes mediante el método KNN. Los gráficos revelan que la densidad de la vibración se mantiene muy similar en ambas situaciones, indicando que la imputación no ha afectado de manera significativa la distribución original de la variable.

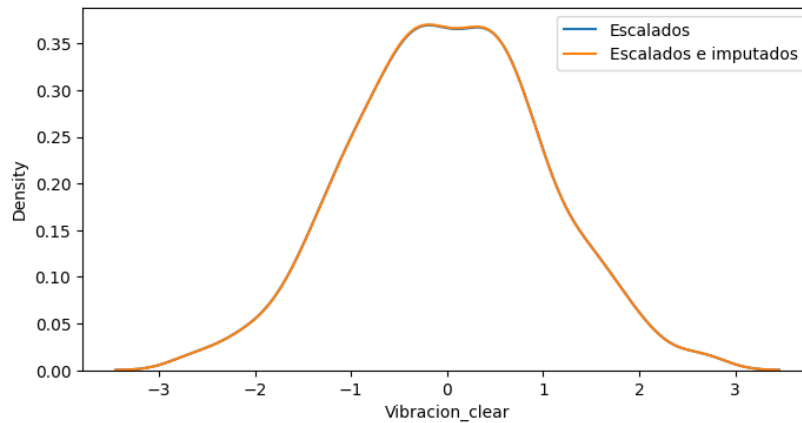


Figura 9. Comparación de las densidades de la vibración antes y después de la imputación.

La Figura 10 presenta las densidades de la variable humedad antes y después de aplicar el proceso de imputación de valores faltantes con el método KNN. La comparación muestra que la distribución de la humedad sigue siendo muy similar tras la imputación, lo que sugiere que el proceso ha mantenido la integridad de la densidad original de la variable.

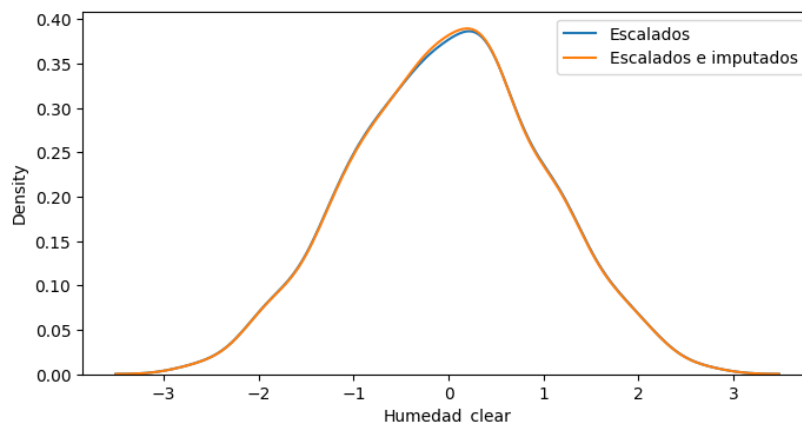


Figura 10. Comparación de las densidades de la humedad antes y después de la imputación.

Para determinar qué variables se relacionan mejor con la variable objetivo de fallas, se realizó un análisis comparativo utilizando gráficos de cajas. La Figura 11 muestra las comparaciones de cada variable (Temperatura, Presión, Vibración y

Humedad) en función de la presencia o ausencia de fallas. Este análisis visual ayuda a identificar diferencias significativas en la distribución de cada variable asociada con la ocurrencia de fallas.

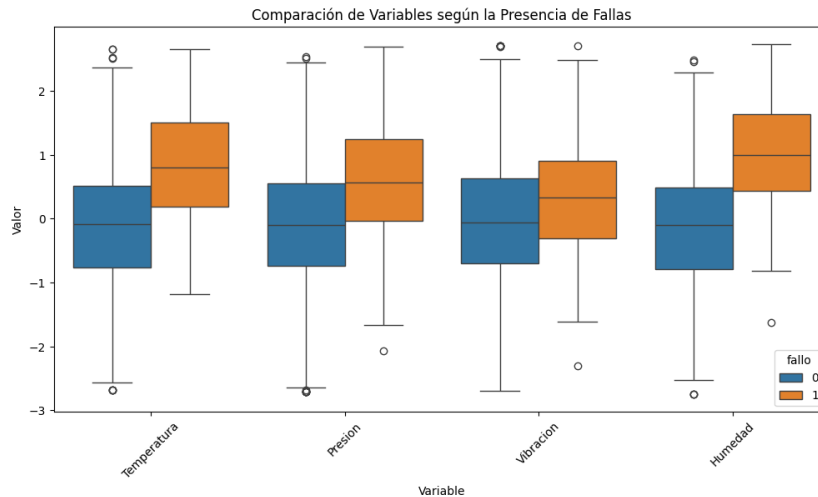


Figura 11. Gráficos de cajas comparativos de cada variable según la presencia de fallas.

La Tabla 4 presenta los promedios y desviaciones estándar de las variables operativas clave (temperatura, presión, vibración y humedad) para el compresor con y sin fallos. Esta comparación inicial permite observar posibles diferencias en las condiciones de operación que podrían estar relacionadas con la ocurrencia de fallos.

Los datos de la Tabla 4 se utilizarán para evaluar la hipótesis nula de igualdad de medias entre los grupos con y sin fallos, a través de una prueba t. Este análisis ayudará a determinar si las diferencias en las condiciones operativas son estadísticamente significativas, lo que podría ser relevante para el mantenimiento predictivo.

Tabla 4. Estadísticas descriptivas según falla o no falla

Variables	Falla		No falla	
	Promedio	Desviación estándar	Promedio	Desviación estándar
Temperatura	0,79	0,90	-0,12	0,96
Presión	0,61	0,97	-0,09	0,97
Vibración	0,31	0,89	-0,05	1,01
Humedad	0,95	0,86	-0,14	0,94

Fuente: Elaboración propia

A continuación, se presenta una tabla con los resultados de la prueba t para la diferencia de medias. Esta prueba estadística permite evaluar si las diferencias en las medias de cada variable, entre las instancias con fallas y sin fallas, son estadísticamente significativas.

Tabla 5. Resultados de la prueba t para la diferencia de medias por variable.

Variable	Estadístico T	P valor
Temperatura	-10,12	0,00
Presión	-7,61	0,00
Vibración	-3,78	0,00
Humedad	-12,37	0,00

Fuente: Elaboración propia

Respecto a la tabla 5, se plantea el siguiente análisis para temperatura: Hipótesis Nula (H_0): No existe una diferencia significativa en las medias de la temperatura entre las instancias con fallas y sin fallas ($\mu_1 = \mu_2$). Hipótesis Alternativa (H_1): Existe una diferencia significativa en las medias de la temperatura entre las instancias con fallas y sin fallas ($\mu_1 \neq \mu_2$). Decisión: El valor del estadístico t (-10.12) es muy grande en valor absoluto y el p-valor es 0.00, lo cual indica que se rechaza la hipótesis nula a un nivel de significancia común (como 0.05). Por lo tanto, la diferencia en la temperatura entre las instancias con fallas y sin fallas es estadísticamente significativa.

En la misma línea para presión: Hipótesis Nula (H_0): No existe una diferencia significativa en las medias de la presión entre las instancias con fallas y sin fallas ($\mu_1 = \mu_2$). Hipótesis Alternativa (H_1): Existe una diferencia significativa en las medias de la presión entre las instancias con fallas y sin fallas ($\mu_1 \neq \mu_2$). Decisión: El valor del estadístico t (-7.61) y el p-valor de 0.00 indican que se rechaza la hipótesis nula. Por lo tanto, la diferencia en la presión es estadísticamente significativa.

El siguiente análisis es para vibración: Hipótesis Nula (H_0): No existe una diferencia significativa en las medias de la vibración entre las instancias con fallas y sin fallas ($\mu_1 = \mu_2$). Hipótesis Alternativa (H_1): Existe una diferencia significativa en las medias de la vibración entre las instancias con fallas y sin fallas ($\mu_1 \neq \mu_2$). Decisión: El valor del estadístico t (-3.78) y el p-valor de 0.00 indican que se rechaza la hipótesis

nula. La vibración también presenta una diferencia estadísticamente significativa entre las categorías de fallas.

Finalmente, la Humedad se analiza como: Hipótesis Nula (H_0): No existe una diferencia significativa en las medias de la humedad entre las instancias con fallas y sin fallas ($\mu_1 = \mu_2$). Hipótesis Alternativa (H_1): Existe una diferencia significativa en las medias de la humedad entre las instancias con fallas y sin fallas ($\mu_1 \neq \mu_2$). Decisión: El valor del estadístico t (-12.37) y el p-valor de 0.00 indican que se rechaza la hipótesis nula. La humedad presenta una diferencia estadísticamente significativa entre las instancias con y sin fallas.

Los resultados sugieren que todas las variables evaluadas (Temperatura, Presión, Vibración y Humedad) tienen una relación significativa con la ocurrencia de fallas. Dado que los valores de t son negativos y los p-valores son extremadamente bajos (0.00), se concluye que hay diferencias significativas en las medias de estas variables entre las instancias con y sin fallas, apoyando la hipótesis alternativa para cada caso.

Se realizó la partición de los datos, reservando el 30% para el conjunto de prueba y el 70% para el conjunto de entrenamiento, los cuales conforman 300 y 700 casos respectivamente. La Figura 12 ilustra la distribución de la variable "fallo" en ambos conjuntos. La proporción de fallas en el conjunto de entrenamiento es del 12,3%, mientras que en el conjunto de prueba es del 14%. Las diferencias entre ambas proporciones son leves, indicando que la partición ha preservado de manera adecuada la representación de la variable objetivo en los conjuntos de entrenamiento y prueba.

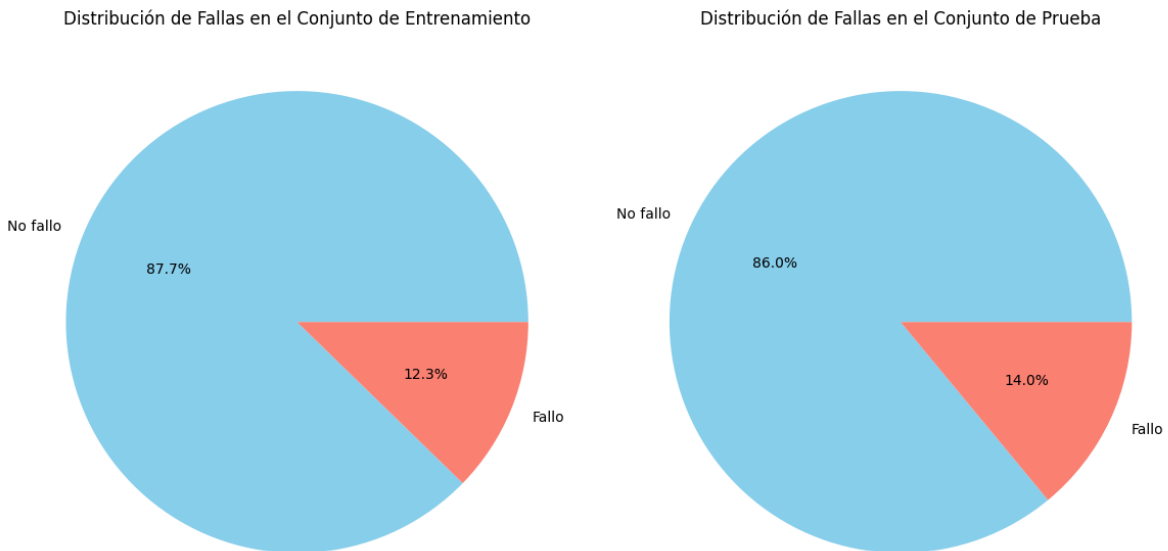


Figura 12. Distribución de la variable "fallo" en los conjuntos de entrenamiento y prueba.

Objetivo específico 2

Ajustar el punto de corte del modelo para maximizar la tasa de verdaderos positivos y minimizar la tasa de falsos positivos, mejorando así la efectividad del sistema de mantenimiento preventivo.

Para mejorar el rendimiento de los modelos y prevenir el sobreajuste, se realizó una optimización de parámetros. En el caso del árbol de decisión, se evaluaron diferentes niveles de profundidad (`max_depth`) y se seleccionó la que maximizaba el `accuracy` en el conjunto de prueba. Para el `random forest`, se optimizó tanto la profundidad máxima de los árboles (`max_depth`) como el número de árboles (`n_estimators`), buscando la combinación que ofreciera el mejor `accuracy` en el conjunto de prueba. La regresión logística no tiene parámetros para ajustar, solo debemos evaluar la significancia de las variables.

La Figura 13 presenta un gráfico de líneas comparadas que muestra el `accuracy` del modelo de árbol de decisión en el conjunto de entrenamiento y en el de prueba a medida que se ajusta la profundidad máxima del árbol. Se observa que el `accuracy` en el conjunto de entrenamiento aumenta continuamente hasta alcanzar un valor cercano a 1, indicando un ajuste cada vez más preciso a los datos de entrenamiento. En contraste, el `accuracy` del conjunto de prueba sube hasta una profundidad máxima de 3 y luego comienza a decrecer. Este patrón sugiere que una

profundidad de 3 es óptima para el árbol de decisión, ya que proporciona un equilibrio adecuado entre la complejidad del modelo y su capacidad de generalización, evitando el sobreajuste.

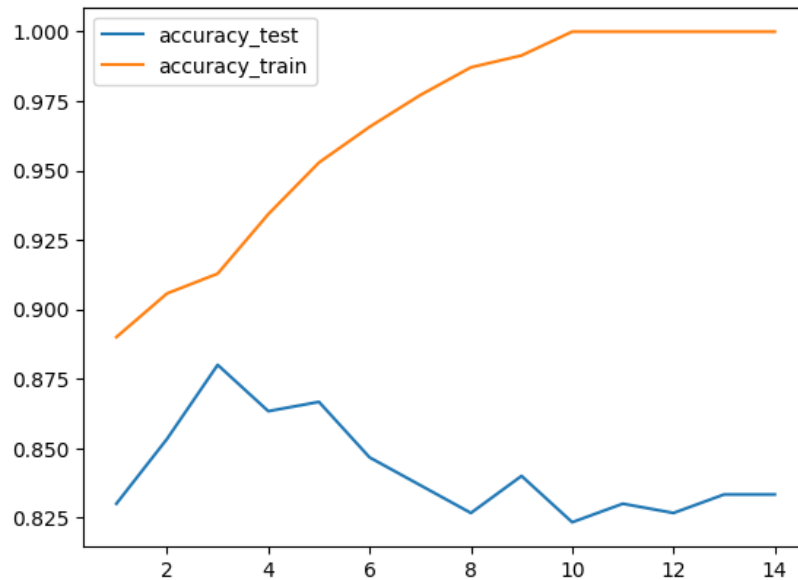


Figura 13. Accuracy del modelo de árbol de decisión en los conjuntos de entrenamiento y prueba a diferentes niveles de profundidad.

La Figura 14 muestra un gráfico de líneas comparadas del accuracy del modelo de random forest en el conjunto de entrenamiento y en el de prueba para bosques de 5, 10 y 15 árboles. En el gráfico, se observa que el accuracy en el conjunto de entrenamiento se aproxima a 1 para todos los tamaños de bosque, lo que indica un ajuste muy preciso a los datos de entrenamiento. El accuracy en el conjunto de prueba muestra un aumento hasta estabilizarse, destacando el modelo con 5 árboles como el más adecuado. La mejor configuración se identificó con una profundidad de 7 y un número de árboles de 5, proporcionando el equilibrio óptimo entre la capacidad de ajuste y la generalización del modelo.

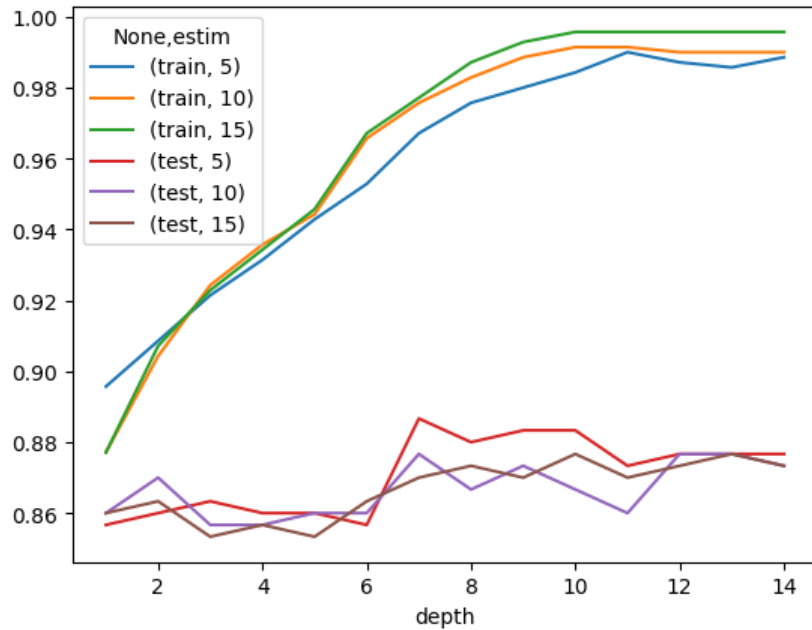


Figura 14. Accuracy del modelo de random forest en los conjuntos de entrenamiento y prueba para diferentes tamaños de bosque y profundidades.

La Tabla 6 muestra los resultados del modelo de regresión logística, donde todos los coeficientes son positivos y estadísticamente significativos con p valores de 0. El coeficiente para Temperatura es 1,75, indicando que un aumento en la temperatura incrementa la probabilidad de fallo. Presión tiene un coeficiente de 1,42, sugiriendo que mayores niveles de presión también están asociados con una mayor probabilidad de fallo. Vibración, con un coeficiente de 0,65, demuestra que un aumento en la vibración está vinculado con un incremento en la probabilidad de fallo, aunque su efecto es menor en comparación con temperatura y presión. Finalmente, Humedad presenta el coeficiente más alto de 2,60, indicando que tiene la mayor influencia en la probabilidad de fallo entre las variables evaluadas. Estos resultados destacan la importancia de todas las variables en la predicción de fallas en el sistema.

Tabla 6. Resultados del modelo de regresión logística.

Variable	Coficiente	z	p valor
const	-4,41	-10,95	0
Temperatura	1,75	7,60	0
Presión	1,42	6,79	0
Vibración	0,65	3,78	0
Humedad	2,60	8,77	0

Fuente: Elaboración propia

Para determinar el punto de corte óptimo para cada modelo, se comenzará con la presentación de una matriz de confusión inicial, que permitirá interpretar indicadores clave como la precisión, la sensibilidad, la especificidad y el valor predictivo positivo. A continuación, se generará la curva ROC (Receiver Operating Characteristic), que ayudará a identificar el punto de corte que maximiza el balance entre la tasa de verdaderos positivos y la tasa de falsos positivos. Con base en el análisis de la curva ROC, se seleccionará el punto de corte óptimo para cada modelo.

Finalmente, se elaborará una nueva matriz de confusión utilizando este punto de corte ajustado, y se proporcionará una interpretación detallada de los resultados, destacando las mejoras en la precisión y el desempeño general del modelo. Este proceso garantizará una evaluación más precisa de la capacidad predictiva de cada modelo y su eficacia en la clasificación de fallas.

La Tabla 7 presenta la matriz de confusión inicial para el modelo logit, que muestra el desempeño del modelo antes de ajustar el punto de corte. De los 300 casos evaluados, se identificaron 246 verdaderos negativos (VN), 12 falsos negativos (FN), 20 falsos positivos (FP) y 22 verdaderos positivos (VP). Esto resulta en una precisión general de 0,89 y una tasa de aciertos del 89%.

Tabla 7. *Matriz de confusión inicial del modelo logit.*

	Predicción: 0	Predicción: 1	Total
Real: 0	246	12	258
Real: 1	20	22	42
Total	266	34	300

Fuente: Elaboración propia

El reporte de clasificación, que también se detalla en la Tabla 8, revela que el modelo tiene una alta precisión para la clase 0 (0,92) y un buen recall (0,95), indicando que la mayoría de los casos negativos fueron correctamente identificados. Para la clase 1, la precisión es más baja (0,65) y el recall es del 0,52, lo que sugiere que el modelo tiene dificultades para identificar correctamente los casos positivos de fallo.

La puntuación F1 para la clase 1 es 0,58, indicando un equilibrio moderado entre precisión y recall para esta clase.

Tabla 8. Reporte de clasificación inicial del modelo logit.

Métrica	Clase 0 (No fallo)	Clase 1 (Fallo)	Macro Avg	Weighted Avg
Exactitud (Accuracy)	-	-	-	0.89
Precisión	0.92	0.65	0.79	0.89
Recall	0.95	0.52	0.74	0.89
F1-Score	0.94	0.58	0.76	0.89
Soporte (Support)	258	42	300	300

Fuente: Elaboración propia

Para mejorar la capacidad del modelo para identificar fallas y optimizar su desempeño, se procederá a ajustar el punto de corte utilizando la curva ROC. Esto permitirá seleccionar un punto de corte que maximice la relación entre verdaderos positivos y falsos positivos, lo que a su vez debería mejorar el balance entre precisión y recall para la clase 1.

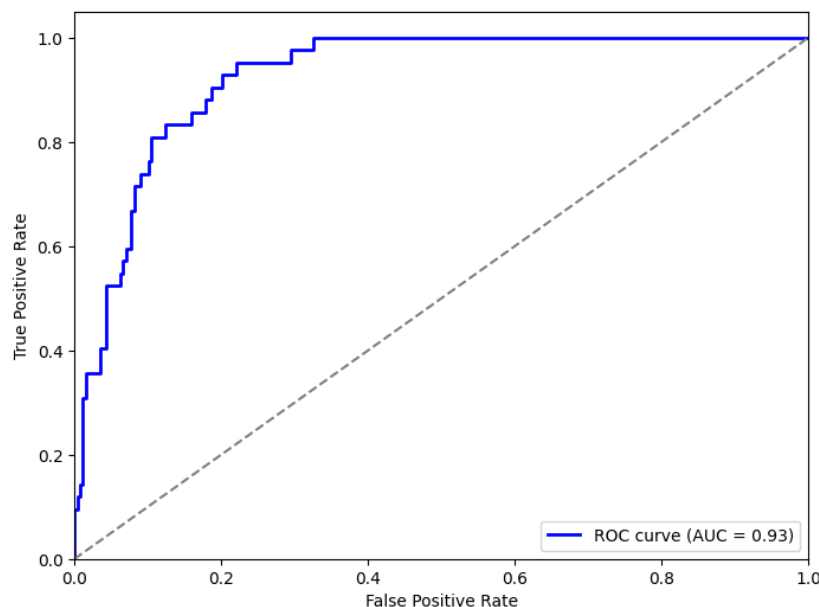


Figura 15. Curva ROC del modelo logit

Se ha generado la curva ROC (Receiver Operating Characteristic) para el modelo logit, como se muestra en la Figura 15. La curva ROC permite evaluar el rendimiento del modelo en diferentes umbrales de clasificación, comparando la tasa de verdaderos positivos frente a la tasa de falsos positivos. El área bajo la curva (AUC) es de 0,93, lo que indica una excelente capacidad del modelo para discriminar entre

las clases de fallo y no fallo. Se obtuvo un punto de corte óptimo de 0,0858, a continuación, veremos los nuevos resultados.

La nueva matriz de confusión para el modelo logit se presenta en la Tabla 9. Esta matriz muestra la distribución de las predicciones clasificadas con el nuevo umbral, comparado con los valores reales.

Tabla 9. Matriz de Confusión del Modelo Logit con el Punto de Corte Óptimo

	Predicción: 0	Predicción: 1	Total
Real: 0	201	57	258
Real: 1	3	39	42
Total	204	96	300

Fuente: Elaboración propia

El reporte de clasificación asociado con este nuevo umbral se detalla a continuación. La precisión y el recall para cada clase muestran un cambio notable. La precisión para la clase de no fallo (0) ha mejorado a 0,99, aunque el recall ha disminuido a 0,78. Esto indica que el modelo es ahora muy bueno en identificar correctamente los casos de no fallo, pero puede haber aumentado la tasa de falsos negativos para esta clase. Para la clase de fallo (1), la precisión ha disminuido a 0,41, mientras que el recall ha aumentado significativamente a 0,93. Esto refleja una mejora en la identificación de los casos de fallo, aunque con una menor precisión en las predicciones positivas.

Tabla 10. Reporte de Clasificación con el Nuevo Punto de Corte

Métrica	Clase 0 (No fallo)	Clase 1 (Fallo)	Macro Avg	Weighted Avg
Exactitud (Accuracy)	-	-	-	0.8
Precisión	0.99	0.41	0.7	0.9
Recall	0.78	0.93	0.85	8
F1-Score	0.87	0.57	0.72	0.83
Soporte (Support)	258	42	300	300

Fuente: Elaboración propia

La matriz de confusión inicial para el modelo de árbol, presentada en la Tabla 11, muestra la distribución de las predicciones comparadas con los valores reales.

Tabla 11. Matriz de Confusión del Modelo de Árbol Inicial

	Predicción: 0	Predicción: 1	Total
Real: 0	240	18	258
Real: 1	18	24	42
Total	258	42	300

Fuente: Elaboración propia

El reporte de clasificación asociado a esta matriz se muestra en la Tabla 12. La precisión para la clase de no fallo (0) es de 0,93, con un recall también de 0,93, indicando un buen desempeño en la identificación de los casos de no fallo. Para la clase de fallo (1), la precisión y el recall son ambos de 0,57, lo que refleja una capacidad moderada del modelo para identificar correctamente los fallos.

Tabla 12. Reporte de Clasificación del Modelo de Árbol Inicial

Métrica	Clase 0 (No fallo)	Clase 1 (Fallo)	Macro Avg	Weighted Avg
Exactitud (Accuracy)	-	-	-	0.88
Precisión	0.93	0.57	0.75	0.88
Recall	0.93	0.57	0.75	0.88
F1-Score	0.93	0.57	0.75	0.88
Soporte (Support)	258	42	300	300

Fuente: Elaboración propia

El modelo de árbol muestra una buena precisión general y un equilibrio aceptable en la identificación de casos de no fallo, pero necesita mejorar en la detección de fallos para optimizar su desempeño global.

La Figura 16 ilustra la curva ROC (Receiver Operating Characteristic) para el modelo de árbol. Esta curva compara la tasa de verdaderos positivos frente a la tasa de falsos positivos en distintos umbrales de clasificación. El área bajo la curva (AUC) es de 0.78, lo que indica una capacidad moderada del modelo para discriminar entre las clases de fallo y no fallo.

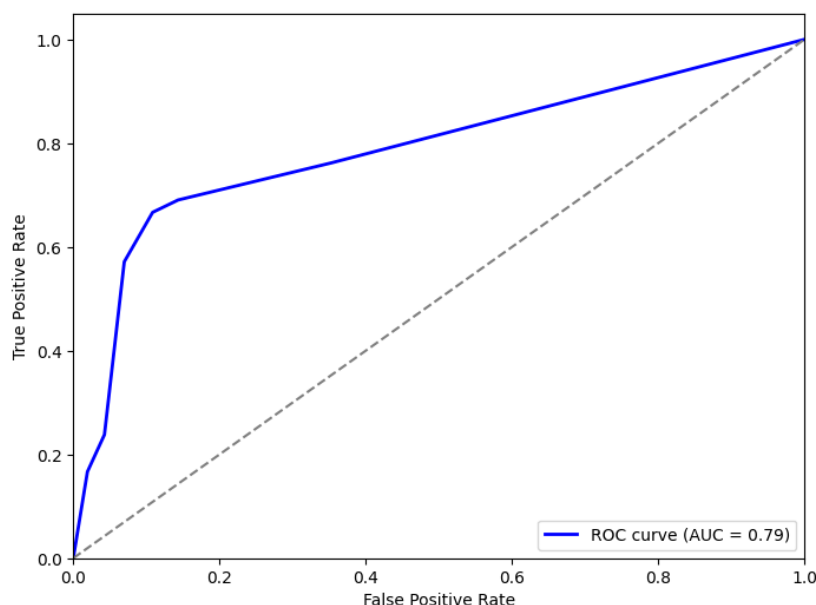


Figura 16. Curva ROC del modelo de árbol de clasificación

El punto de corte óptimo determinado para el modelo es 0,3. Este umbral maximiza el equilibrio entre la tasa de verdaderos positivos y la tasa de falsos positivos, ajustando el modelo para mejorar la clasificación de fallos sin comprometer significativamente la identificación de casos de no fallo.

Aplicando el punto de corte óptimo de 0,3 al modelo de árbol, se obtuvo la siguiente matriz de confusión en la tabla 12:

Tabla 13. Matriz de confusión con punto de corte óptimo para el modelo de árbol

	Predicción: 0	Predicción: 1	Total
Real: 0	240	18	258
Real: 1	18	24	42
Total	258	42	300

Fuente: Elaboración propia

Tabla 14. Reporte de clasificación con punto de corte óptimo para el modelo de árbol

Métrica	Clase 0 (No fallo)	Clase 1 (Fallo)	Macro Avg	Weighted Avg
Exactitud (Accuracy)	-	-	-	0.88
Precisión	0.93	0.57	0.75	0.88

Recall	0.93	0.57	0.75	0.88
F1-Score	0.93	0.57	0.75	0.88
Soporte (Support)	258	42	300	300

Fuente: Elaboración propia

Los resultados obtenidos con la tabla 13 y la tabla 14 muestran una notable similitud con los resultados iniciales. Esta consistencia se debe a que el árbol de clasificación generó una cantidad limitada de valores únicos de probabilidad. En otras palabras, el modelo asignó probabilidades muy similares a las observaciones, lo que hace que el ajuste del punto de corte tenga un impacto mínimo en las métricas de rendimiento. Esto implica que, a pesar del cambio en el punto de corte, el árbol de clasificación no varía significativamente en su capacidad para distinguir entre las clases, ya que operaba cerca de un equilibrio en la diferenciación entre las categorías.

Objetivo general

Plantear un Sistema de Mantenimiento Predictivo utilizando Algoritmos de Machine Learning para solucionar la inactividad productiva causada por reparaciones de maquinaria en el sector industrial.

Aunque el modelo de Random Forest está compuesto por múltiples árboles de decisión, sus predicciones tienden a ser más variadas en comparación con los modelos individuales, como el árbol de decisión simple. A continuación, se presentan la matriz de confusión y el reporte de indicadores para el modelo de Random Forest.

Tabla 15. *Matriz de Confusión Inicial para el Modelo Random Forest*

	Predicción: 0	Predicción: 1	Total
Real: 0	241	17	258
Real: 1	20	22	42
Total	261	39	300

Fuente: Elaboración propia

La tabla 15 muestra que el modelo clasifica correctamente 241 de los 258 casos de no fallo y 22 de los 42 casos de fallo. Sin embargo, presenta una cierta dificultad para identificar los casos de fallo, con 20 falsos negativos. Esto se traduce en una precisión del 0,56 para la clase de fallo, indicando que el modelo tiene un rendimiento moderado en identificar las fallas.

Tabla 16. Reporte de Clasificación Inicial para el Modelo Random Forest

Métrica	Clase 0 (No fallo)	Clase 1 (Fallo)	Macro Avg	Weighted Avg
Exactitud (Accuracy)	-	-	-	0.88
Precisión	0.92	0.56	0.74	0.87
Recall	0.93	0.52	0.73	0.88
F1-Score	0.93	0.54	0.74	0.87
Soporte (Support)	258	42	300	300

Fuente: Elaboración propia

La tabla 16 revela una precisión de 0,92 y un recall de 0,93 para la clase de no fallo, lo que indica un buen desempeño en la identificación de casos no fallidos. No obstante, el modelo muestra un desempeño inferior en la clase de fallo, con una precisión de 0,56 y un recall de 0,52, sugiriendo que el modelo tiene dificultades para identificar correctamente los casos de fallo. La precisión general del modelo es del 0,88, reflejando una correcta clasificación mayoritaria, pero con margen de mejora en la detección de fallas.

Se ha generado la curva ROC para el modelo de Random Forest, como se muestra en la Figura 17. Esta curva ilustra la capacidad del modelo para distinguir entre las clases de fallo y no fallo en diferentes umbrales de clasificación. El área bajo la curva (AUC) es de 0,834, indicando una buena capacidad discriminativa del modelo. El punto de corte óptimo determinado es 0,0476, el cual maximiza el balance entre la tasa de verdaderos positivos y la tasa de falsos positivos, proporcionando así un umbral adecuado para la clasificación.

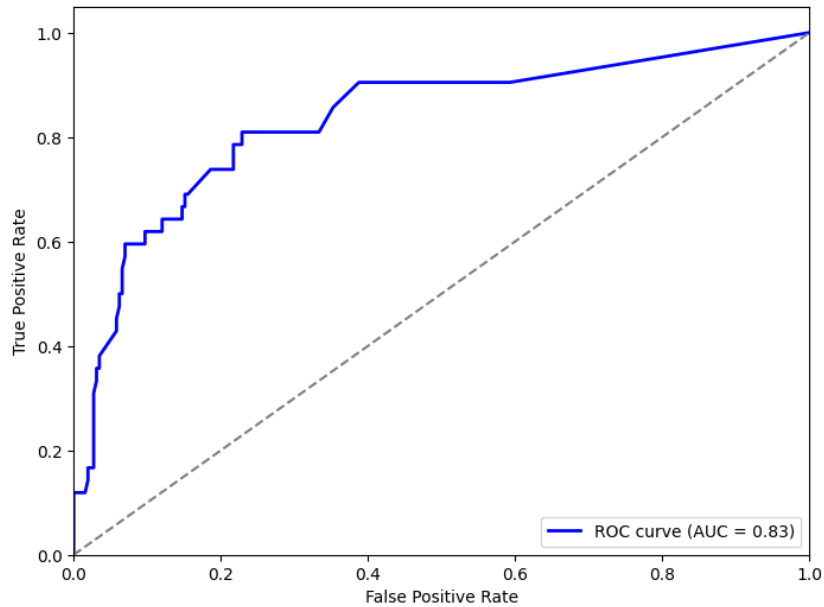


Figura 17. Curva ROC para el modelo de Random Forest

A continuación, se presentan los resultados del modelo Random Forest tras ajustar el punto de corte a 0,0476. Comparando con la matriz de confusión (tabla 16) y los indicadores (tabla 17), se observa un aumento significativo en el **recall** para la clase 1 (fallo), que pasa de 0,52 a 0,79. Esto sugiere que el modelo ahora identifica con mayor precisión los casos de fallo, lo que es crucial en contextos donde es importante no pasar por alto fallos críticos.

Tabla 17. Matriz de confusión con punto de corte óptimo para el modelo de random forest

	Predicción: 0	Predicción: 1	Total
Real: 0	199	59	258
Real: 1	9	33	42
Total	208	92	300

Fuente: Elaboración propia

Sin embargo, este ajuste ha reducido la precisión de la clase 1 de 0,56 a 0,36, lo que implica un incremento en las clasificaciones incorrectas de "fallo" cuando en realidad no lo son. En contraste, la precisión para la clase 0 (no fallo) se mantiene alta en 0,96, aunque su recall se reduce a 0,77, indicando que un número mayor de instancias sin fallos se clasifican incorrectamente como fallos.

Tabla 18. Reporte de clasificación con punto de corte óptimo para el modelo de *random forest*

Métrica	Clase 0 (No fallo)	Clase 1 (Fallo)	Macro Avg	Weighted Avg
Exactitud (Accuracy)	-	-	-	0.77
Precisión	0.96	0.36	0.66	0.87
Recall	0.77	0.79	0.78	0.77
F1-Score	0.85	0.49	0.67	0.8
Soporte (Support)	258	42	300	300

Fuente: Elaboración propia

El ajuste del punto de corte ha llevado a un modelo que es más sensible a detectar fallos (incremento en el recall de la clase 1), a expensas de generar más falsos positivos, como lo refleja la precisión reducida. Esta estrategia puede ser adecuada en escenarios donde el costo de no detectar un fallo es alto, incluso si se deben asumir más falsas alarmas.

IV. DISCUSIÓN

Con respecto al objetivo general, se buscó plantear un Sistema de Mantenimiento Predictivo utilizando Algoritmos de Machine Learning para solucionar la inactividad productiva causada por reparaciones de maquinaria en el sector industrial. El modelo de Random Forest, a pesar de su capacidad para manejar múltiples árboles de decisión, muestra una precisión general del 0,88 pero enfrenta desafíos en la identificación de fallos, con una precisión inicial de 0,56 y un recall de 0,52 para la clase de fallo. La curva ROC muestra un AUC de 0,834, lo que indica una buena capacidad discriminativa. Al ajustar el punto de corte a 0,0476, el recall para la clase de fallo aumenta a 0,79, mejorando la detección de fallos críticos, aunque la precisión de esta clase disminuye a 0,36 y la precisión para la clase de no fallo se mantiene alta en 0,96, pero con un recall reducido a 0,77. Este ajuste mejora la sensibilidad del modelo a los fallos, pero incrementa los falsos positivos, lo que puede ser beneficioso en contextos donde es crucial detectar todos los fallos, aunque implique aceptar más falsas alarmas. Estos resultados tienen semejanza con los expresados por Castro et al. (2022), quienes también a través de un algoritmo, bagged en su caso, demostraron una exactitud de 98,56% en la predicción de fallos en maquinaria pesada como excavadoras. Además de optimizar la capacidad de reacción y tiempo de respuesta en el mantenimiento. Tal fue el caso para Guerroum et al. (2022), quienes también lograron beneficios financieros y productivos al manejar los problemas ocasionados por escalabilidad e interoperabilidad.

En lo que respecta al objetivo específico uno, se buscó determinar la influencia de variables como temperatura, presión, vibración y humedad en la ocurrencia de fallos, mediante análisis estadísticos y visualizaciones gráficas. Para optimizar el rendimiento de los modelos y evitar el sobreajuste, se ajustaron los parámetros de los modelos. Para el árbol de decisión, se determinó que una profundidad máxima de 3 es óptima, ya que maximiza el accuracy en el conjunto de prueba sin sobreajustar el modelo. En el caso del random forest, el accuracy en el conjunto de prueba es más alto con 5 árboles y una profundidad máxima de 7, lo que ofrece el mejor equilibrio entre ajuste y generalización. La regresión logística, al no tener parámetros ajustables, mostró coeficientes estadísticamente significativos para todas las variables, siendo la humedad la más influyente con un coeficiente de 2,60, seguida por temperatura (1,75),

presión (1,42) y vibración (0,65). Anomalías que se analizan como señales de posibles futuros fallos, los cuales pueden generar daños materiales y ambientales, tal como también señala Torim et al. (2022).

En cuanto al objetivo específico dos, se pretendió analizar ajustar el punto de corte del modelo, el desempeño del modelo de regresión logística y del modelo de árbol de decisión en la identificación de fallos, maximizar la tasa de verdaderos positivos y minimizar la tasa de falsos positivos, utilizando métricas como precisión, recall y F1-score. El análisis comparativo de los modelos de clasificación utilizados para la detección de fallos en compresores industriales revela que el ajuste del punto de corte ha tenido un impacto significativo en el rendimiento de cada modelo. El modelo logit, tras ajustar el umbral, mostró una mejora considerable en la identificación de fallos (recall de 0,93), aunque con una reducción en la precisión para estos casos. Por otro lado, el modelo de árbol de clasificación mantuvo un rendimiento similar antes y después del ajuste, lo que sugiere que su capacidad para distinguir entre fallos y no fallos es limitada. El modelo Random Forest, al igual que el logit, incrementó su capacidad para detectar fallos (recall de 0,79) tras el ajuste, aunque a costa de un aumento en los falsos positivos. En general, estos ajustes mejoraron la sensibilidad de los modelos para detectar fallos, lo cual es crucial en aplicaciones de mantenimiento predictivo donde la anticipación de fallos es prioritaria, aunque se genere un mayor número de falsas alarmas. Coincidiendo con los resultados de Ayvaz y Alpay (2021) y Shandookh et al. (2024), el modelo Random Forest en comparación con otros, ofrece un mayor rendimiento en la predicción de fallos. Esta pronta captación de errores estructurales permite prolongar la durabilidad de los equipos industriales tal como se demuestra en la investigación y también por parte de Li et al. (2024).

Castro et al. (2022), desarrolló un modelo predictivo basado en algoritmos de aprendizaje automático para identificar fallas en excavadoras hidráulicas. Se destaca la alta precisión (98.56%) lograda por su modelo basado en "bagged trees". Comparado con los resultados de esta tesis, donde se implementó un modelo de "Random Forest", se observa una similitud en la eficacia de los algoritmos predictivos al manejar datos industriales. Sin embargo, a diferencia de Castro et al., que se enfocaron en la disponibilidad de repuestos, esta investigación prioriza la detección temprana de fallas críticas para optimizar la operación de compresores industriales.

Respecto al estudio de Guerroum et al. (2022), utilizó un enfoque integral que combina la Industria 4.0 y la ingeniería de sistemas basada en modelos (SysML) para el mantenimiento predictivo en la industria minera. En comparación, aunque ambos estudios emplean la tecnología para mejorar la operación industrial, el presente trabajo se concentra en modelos específicos de machine learning aplicados en compresores industriales. La integración tecnológica y la personalización de sistemas en el trabajo de Guerroum et al. superan la implementación tecnológica de esta investigación, la cual se limita a parámetros específicos y modelado predictivo.

En la investigación de Useche et al. (2019), aborda la técnica de observación instrumentada con dispositivos avanzados, alineándose con la metodología empleada en este trabajo para la recolección de datos mediante sensores. Sin embargo, la diferencia radica en que el estudio de Useche et al. se enfoca en la mejora de procesos productivos generales, mientras que el presente estudio se centra en la predicción específica de fallas en un contexto industrial limitado a compresores.

Respecto a Navarro (2022), En su estudio sobre la gestión de mantenimiento en una unidad minera, Navarro subraya la importancia de técnicas de minería de datos para identificar patrones de comportamiento y prevenir fallas. Este enfoque es similar al empleado en esta investigación, aunque en este trabajo se aplican modelos más avanzados de machine learning. Además, la metodología de Navarro es más genérica, mientras que esta tesis ofrece un análisis detallado de las variables operativas críticas para predecir fallas en compresores industriales.

En la pesquisa Soares et al. (2022), sobre el análisis predictivo en sistemas mecánicos en Brasil también empleó variables como temperatura y vibración para predecir fallos. La diferencia radica en que este trabajo ha logrado ajustar de manera más precisa los modelos de machine learning para reducir falsos positivos y mejorar el recall en fallas críticas, mientras que Soares et al. se concentraron en la optimización general de procesos sin un enfoque específico en compresores industriales.

San Miguel (2021), en su aplicación de técnicas de machine learning en compresores de aire, San Miguel destaca la importancia de modelos de regresión para la predicción de fallas. Sin embargo, en esta tesis, el modelo de "Random Forest" ha demostrado mayor efectividad, lo que sugiere que los modelos basados en

múltiples árboles pueden superar a las regresiones tradicionales en situaciones donde se requiere manejar múltiples variables interrelacionadas.

El antecedente de Ayvaz y Alpay (2021) muestra cómo los modelos de "Random Forest" ofrecen un rendimiento superior en la predicción de fallas. Este hallazgo se confirma en la presente investigación, donde "Random Forest" no solo mejora la sensibilidad del modelo, sino que también permite ajustar el punto de corte para maximizar la detección de fallas críticas, lo que es crucial en entornos industriales de alta exigencia.

Shandookh et al. (2024) en su investigación reciente sobre mantenimiento predictivo, los autores muestran cómo la anticipación de fallas mediante machine learning puede incrementar la durabilidad de los equipos industriales. Este estudio se alinea estrechamente con los objetivos y resultados de la presente investigación, donde se han logrado optimizar las estrategias de mantenimiento para prolongar la vida útil de los compresores industriales, demostrando similitudes en los enfoques y beneficios.

V. CONCLUSIONES

Primera:

A pesar de tener una precisión general del 0,88, el modelo de Random Forest enfrenta desafíos en la identificación de fallos, con una precisión inicial baja, aunque el ajuste del punto de corte mejora significativamente el recall para la detección de fallos críticos, aumentando la sensibilidad del modelo a los fallos a costa de incrementar los falsos positivos.

Segunda:

En la evaluación de las variables que influyen en la ocurrencia de fallos, la regresión logística mostró que la humedad tiene el mayor impacto, seguida de la temperatura, presión y vibración. Estos resultados coinciden con investigaciones previas que destacan la importancia de estas variables en la predicción de fallos industriales.

Tercera:

El análisis de los modelos de clasificación revela que el ajuste del punto de corte mejoró la capacidad del modelo logit y Random Forest para detectar fallos, destacando la importancia de estos ajustes en aplicaciones de mantenimiento predictivo, donde la anticipación de fallos es esencial, aunque implique un aumento en los falsos positivos.

VI. RECOMENDACIONES

Primera:

Se recomienda ajustar el punto de corte en modelos como Random Forest y logit para mejorar la detección de fallos críticos. Esta estrategia es especialmente valiosa en entornos industriales donde la detección temprana de todos los posibles fallos es prioritaria, incluso si esto resulta en un aumento de falsos positivos, lo que permitirá tomar medidas preventivas más efectivas.

Segunda:

Se sugiere establecer un monitoreo continuo y detallado de variables clave como humedad, temperatura, presión y vibración, dado que han demostrado ser influyentes en la predicción de fallos. Este enfoque permitirá la detección temprana de anomalías, minimizando los riesgos operativos y optimizando las estrategias de mantenimiento predictivo.

Tercera:

Es recomendable realizar evaluaciones periódicas de los modelos de clasificación utilizados para la predicción de fallos, como el árbol de decisión, Random Forest y regresión logística. Comparar su rendimiento y ajustar sus parámetros regularmente asegurará que el sistema de mantenimiento predictivo mantenga su efectividad, adaptándose a cambios en las condiciones operativas y minimizando tanto los fallos no detectados como los falsos positivos.

REFERENCIAS

ABBASI, A. & KHAN, M. S. Recurrent Neural Networks for Predictive Maintenance in Oil and Gas Industry. *Journal of Petroleum Technology*, 2019, 45 (3), 23-34 pp. <https://doi.org/10.1016/j.jpetrol.2019.07.015>.

ALAVEDRA, C., [et al.]. Gestión de mantenimiento preventivo y su relación con la disponibilidad de la flota de camiones 730e Komatsu-2013. *Ingeniería industrial*, 2016, (034), 11-26 pp. <https://doi.org/10.26439/ING.IND2016.N034.529>

ALVES, F., [et al.]. Deployment of a Smart and Predictive Maintenance System in an Industrial Case Study. 2020 IEEE 29th International Symposium on Industrial Electronics (ISIE), 2020, 493-498pp. <https://doi.org/10.1109/ISIE45063.2020.9152441>

ANDRIOAIA, D., & GAITAN, V. A specialty literature review of the predictive maintenance systems. *Journal of Engineering Studies and Research*, 2024, 29(4), 17-23 pp. <https://doi.org/10.29081/jesr.v29i4.002>

ARAFAT, MY; HOSSAIN, MJ; ALAM, Md Morshed. Alcances del aprendizaje automático en el mantenimiento predictivo de microrredes: marcos potenciales, desafíos y perspectivas. *Renewable and Sustainable Energy Reviews*, 2024, vol. 190, 114088. DOI: <https://doi.org/10.1016/j.rser.2023.114088>

AYVAZ, S., & ALPAY, K. Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time. *Expert Syst. Appl.*, 2021, 173. ISSN: 114-598. <https://doi.org/10.1016/j.eswa.2021.114598>

BASTOS, P., LOPES, I., & PIRES, L. A maintenance prediction system using data mining techniques. *World Congress on Engineering. Proceedings of the World Congress on Engineering*, Julio 2012, 3. ISBN: 9789881925220. https://www.iaeng.org/publication/WCE2012/WCE2012_pp1448-1453.pdf

CACHADA, A., [et al.]. Maintenance 4.0: Intelligent and Predictive Maintenance System Architecture. 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), 2018, 1, 139-146 pp. <https://doi.org/10.1109/ETFA.2018.8502489>

CARBONELL, J., MICHALSKI, R., & MITCHELL, T. An overview of machine learning. *Machine learning*, 1983, 3-23 pp. https://doi.org/10.1007/978-3-662-12405-5_1

CASTRO, J., [et al.]. Automatic learning algorithm for troubleshooting in hydraulic machinery. *Indonesian Journal of Electrical Engineering and Computer Science*, 2023, 29(1), 535-544 pp. ISSN: 2502-4752. <https://doi.org/10.11591/ijeecs.v29.i1.pp535-544>

CHEN, Chong, et al. El avance del gemelo digital para el mantenimiento predictivo: el papel y la función del aprendizaje automático. *Journal of Manufacturing Systems*, 2023, vol. 71, 581-594 pp. DOI: <https://doi.org/10.1016/j.jmsy.2023.10.010>

DE BOER, J. Temperature as a basic physical quantity. *Metrologia*, 1965, 1(4), 158-169 pp. <https://pure.tue.nl/ws/files/4400672/606639.pdf>

DUMITRESCU, G., & DUMITRESCU, F. A Practical Approach to Vibration Detection and Measurement. *Universitatii Maritime Constanta. Analele*, 2010, 11(13), 74-81 pp. <https://www.proquest.com/openview/83d45b3870eccef2917cea0f47690ac6/1?pq-origsite=gscholar&cbl=60411>

FEISTEL, R., & LOVELL, J. Defining relative humidity in terms of water activity. Part 1: definition. *Metrologia*, 2017, 54(4), 566 pp. <https://doi.org/10.1088/1681-7575%2Faa7083>

FERNÁNDEZ DÍEZ, Pedro. Compresores, Departamento de ingeniería eléctrica y energética, Universidad de Cantabria. (s.f.). https://www.ing.una.py/pdf_material_apoyo/compresores-y-ventiladores.pdf

FLORES Q., Ludwing R.; BLANCO P., Joseph O. Determinación de costos operativos y su incidencia en la rentabilidad económica y financiera de las empresas de transportes urbano de pasajeros de la ciudad de Puno – Perú. *Actualidad Contable FACES*, [S.I.], 43: 76-92 pp, noviembre 2021. ISSN 2244-8772. Disponible en: <http://erevistas.saber.ula.ve/index.php/actualidadcontable/article/view/17551>

GAFFNEY, J. Humidity: Basic principles and measurement techniques. *HortScience*, 1978, 13(5), 551-555 pp. <https://doi.org/10.21273/hortsci.13.5.551>

GARCES, Diego A. y CASTRILLON, Omar D. Diseño de una Técnica Inteligente para Identificar y Reducir los Tiempos Muertos en un Sistema de Producción. Inf. tecnol. [online]. 2017, 8 (3): 157-170 pp. ISSN 0718-0764. <http://dx.doi.org/10.4067/S0718-07642017000300017>.

GUERROUM, M., [et al.]. Integration of MBSE into Mining Industry: Predictive Maintenance System. Int. J. Emerg. Technol. Adv. Eng, 2022, 12, 21. ISSN: 2250-2459. https://doi.org/10.46338/ijetae0422_21

HURTADO, Julio, et al. Aprendizaje continuo para mantenimiento predictivo: visión general y desafíos. Sistemas Inteligentes con Aplicaciones, 2023, vol. 19, 200-251 pp. DOI: <https://doi.org/10.1016/j.cirpj.2022.11.004>

IFTIKHAR, A., HUSSAIN, S. & AHMED, K. Machine Learning Algorithms for Fault Prediction in Industrial Compressors. IEEE Transactions on Industrial Informatics, 2022, 58(2), 456-467 pp. <https://doi.org/10.1109/TII.2022.3146758>.

Joël, S. Pressure, temperature and heat. 2020. https://www.encyclopedie-environnement.org/app/pdf/?idpost=6898&idauthor=A-9&urlimg=https%3A%2F%2Fwww.encyclopedie-environnement.org%2Fapp%2Fuploads%2F2016%2F10%2Fpression-temperature_couverture.jpg

KÜFNER, T., [et al.]. Predictive maintenance: Using recurrent neural networks for wear prognosis in current signatures of production plants. International Journal of Mechanical Engineering and Robotics Research, 2021, 10(11), 583-591 pp. <https://doi.org/10.18178/ijmerr.10.11.583-591>

LEUKEL, J., GONZÁLEZ, J., & RIEKERT, M. Adopción de tecnología de aprendizaje automático para la predicción de fallas en mantenimiento industrial: una revisión sistemática. Journal of Manufacturing Systems, 2021, 61, 87-96 pp. <https://doi.org/10.1016/j.jmsy.2021.08.012>

LEWIS, T., & DENNING, P. Learning machine learning. Communications of the ACM, 2018, 61, 24 – 27 pp. <https://doi.org/10.1145/3286868>

LI, Z., HE, Q., & LI, J. A survey of deep learning-driven architecture for predictive maintenance. *Engineering Applications of Artificial Intelligence*, 2024, 133, ISSN: 108285. <https://doi.org/10.1016/j.engappai.2024.108285>

LOZADA, José. Investigación Aplicada: Definición, Propiedad Intelectual e Industria. *CIENCIAMÉRICA*, 2014, 3, 34-39 pp. <https://dialnet.unirioja.es/descarga/articulo/6163749.pdf>

MANAKOVA, O., MAYOROV, M., & SIDOROV, A. Vibrations in engineering systems. In *IOP Conference Series: Materials Science and Engineering*, 2019, 666(1), ISSN: 012044. <https://doi.org/10.1088/1757-899X%2F666%2F1%2F012044>

NAVARRO JIMENEZ, Claudio Cesar. Propuesta de mejora de la gestión de mantenimiento para optimizar el desempeño en una unidad minera del sur del país - Arequipa 2021. https://repositorio.continental.edu.pe/bitstream/20.500.12394/11458/2/IV_FIN_108_T_E_Navarro_Jimenez_2022.pdf

OECD/EUROSTAT, Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris, 2018. <https://doi.org/10.1787/9789264304604-en>

ORGANIZACIÓN DE LAS NACIONES UNIDAS (ONU). Los Objetivos de Desarrollo Sostenible en Perú. 2024. <https://peru.un.org/es/sdgs>

OROŠNJAK, Marko, et al. Del paradigma de mantenimiento predictivo al paradigma de mantenimiento basado en energía: lograr una producción más limpia a través de la productividad funcional. *Journal of Cleaner Production*, 2023, vol. 408, 137-177 pp. DOI: <https://doi.org/10.1016/j.jclepro.2023.137177>

PAGANO, Davide. Un modelo de mantenimiento predictivo utilizando redes neuronales de memoria de corto plazo y largo plazo e inferencia bayesiana. *Decision Analytics Journal*, 2023, vol. 6, 100-174 pp. DOI: <https://doi.org/10.1016/j.dajour.2023.100174>

PASCUAL JIMÉNEZ, Rubén. Obtención de los parámetros característicos y del diagrama del indicador de un compresor alternativo montado en un banco de ensayos. 2019. https://oa.upm.es/54098/1/TFG_RUBEN_PASCUAL_JIMENEZ.pdf

RAJ, A. A Review on Machine Learning Algorithms. International Journal for Research in Applied Science and Engineering Technology. 2019, 7(6), 792-769 pp. ISSN: 2321-9653. <https://doi.org/10.22214/ijraset.2019.6138>

RAMOS, M., LOZANO, B., & SUÁREZ, H. Implementación de mantenimiento preventivo y predictivo a los equipos del proceso de producción en la empresa EQUIACEROS SAS. Ingenio Libre, 2020. 8(18), 70-77 pp. <https://doi.org/10.18041/2322-8415%2FINGELIBRE.2020.V8N18.7012>

SAN MIGUEL, Marcos. Aplicación de técnicas de ML para la optimización de compresores de aire. 2021, 22 pp. https://repositorio.unican.es/xmlui/bitstream/handle/10902/24927/TFM_MARCOS_SAN_MIGUEL.pdf?sequence=1&isAllowed=y

SANCHEZ FLORES, Fabio Anselmo. Fundamentos epistémicos de la investigación cualitativa y cuantitativa: consensos y disensos. Rev. Digit. Invest. Docencia Univ. [online]. 2019, 13 (1): 102-122 pp. ISSN 2223-2516. <http://dx.doi.org/10.19083/ridu.2019.644>.

SHANDOOKH, A., OGAILI, A., & AL-HADDAD, L. Failure analysis in predictive maintenance: Belt drive diagnostics with expert systems and Taguchi method for unconventional vibration features. Heliyon. 2024. <https://doi.org/10.1016/j.heliyon.2024.e34202>

SHAVETA. A review on machine learning. International Journal of Science and Research Archive, 2023. 09(01), 281–285 pp. <https://doi.org/10.30574/ijrsra.2023.9.1.0410>

SIRASKAR, Rajesh, et al. Reinforcement learning for predictive maintenance: A systematic technical review. Artificial Intelligence Review, 2023, 56, 11, 12885-12947 pp. DOI: <https://doi.org/10.1108/JQME-10-2020-0107>

SOARES RODRIGUES, Daniel, [et al.]. Detección De Fallas Mecánicas Mediante "Machine Learning", Utilizando El Clasificador "Random Forest". XV Congreso

Iberoamericano de Ingeniería Mecánica, 2022, 8 pp.
<https://apidspace.linhd.uned.es/server/api/core/bitstreams/2025b8c3-faed-4bc0-8462-16eacb20d4/content>

STAROSIELSKI, N. The materiality of media heat. *International Journal of Communication*, 2024, 8, 2504-2508.
<https://ijoc.org/index.php/ijoc/article/download/3298/1268>

TAŞCI, Bernar; OMAR, Ammar; AYVAZ, Serkan. Predicción de la vida útil restante para el mantenimiento predictivo en la fabricación. *Computers & Industrial Engineering*, 2023, 184, 109 -566 pp. DOI: <https://doi.org/10.1016/j.cie.2023.109566>

TORIM, A., [et al.]. Pattern Based Software Architecture for Predictive Maintenance. In Symposium of the Norwegian AI Society. Springer International Publishing, 2022, 26-38 pp. https://doi.org/10.1007/978-3-031-17030-0_3

USECHE M., [et al.]. Técnicas e instrumentos de recolección de datos cuali-cuantitativos. Universidad de La Guajira. Primera edición. 2019. ISBN 978-956-6037-04-0.
<https://repositoryinst.uniguajira.edu.co/bitstream/handle/uniguajira/467/88.%20Tecnicas%20e%20instrumentos%20recolecci%C3%B3n%20de%20datos.pdf?sequence=1>

VANDERSCHUEREN, Toon, [et al.]. Optimización de la frecuencia de mantenimiento preventivo con aprendizaje automático causal. *International Journal of Production Economics*, 2023, vol. 258, 108798 pp. DOI: <https://doi.org/10.1016/j.ijpe.2023.108798>

XU, G, [et al.]. Data-Driven Fault Diagnostics and Prognostics for Predictive Maintenance: A Brief Overview*. 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), 2019, 103-108 pp.
<https://doi.org/10.1109/COASE.2019.8843068>

ZHUANG, Liangliang; XU, Ancha; WANG, Xiao-Lin. Un marco de mantenimiento predictivo impulsado por pronósticos basado en aprendizaje profundo bayesiano. *Reliability Engineering & System Safety*, 2023, vol. 234, 109-181 pp. DOI: <https://doi.org/10.1016/j.ress.2023.109181>

ANEXOS

Anexo 1. Tabla de Operacionalización de Variables

VARIABLE	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	ESCALA DE MEDICIÓN
Compresor industrial	Un dispositivo mecánico utilizado para aumentar la presión del aire o gas mediante la reducción de su volumen, con aplicaciones en diversas industrias (Fernández, s.f.).	Se considera en estado operativo si está en funcionamiento y realizando la compresión de aire o gas según lo diseñado. El estado se clasifica en "falla" (1) si el compresor presenta un problema que afecta su funcionamiento, y "no falla" (0) si opera normalmente.	Estado operativo	Falla	1
				No falla	0
Condiciones operativas del compresor industrial	Conjunto de parámetros o variables ambientales y mecánicas que afectan el funcionamiento del compresor industrial, como temperatura, presión, vibración, humedad y tiempos de inactividad (Pascual, 2019).	Medición y monitoreo continuo de las variables de temperatura (en °C), presión (en PSI), vibración (en mm/s), humedad (en %) y tiempos de inactividad (en horas). Estos valores se registran en intervalos constantes para determinar el estado operativo del compresor.	Temperatura	Temperatura en °C	Intervalo
			Presión	Presión en PSI	Intervalo
			Vibración	Vibración en mm/s	Intervalo
			Humedad	Humedad relativa en %	Intervalo
			Tiempos de inactividad	Tiempo de inactividad en horas	Razón

Anexo 2. Instrumento de Recolección de Datos

Ficha de recolección de datos operativos

Fecha	Hora	Temperatura (°C)	Presión (PSI)	Vibración (mm/s)	Humedad (%)	Fallo (1=si 0=no)

Notas Adicionales:

- Asegúrese de registrar los datos en la tabla de forma precisa y completa.
- Revise regularmente los instrumentos para asegurar su correcto funcionamiento.
- Realice las mediciones en condiciones estables para obtener resultados consistentes.

Responsable de la Recolección: [Nombre del responsable]

Fecha de Elaboración: [Fecha]

Anexo. 3





Datos recolectados

	A	B	C	D	E	F
1	Fecha	Temperatura	Presion	Vibracion	Humedad	fallo
2	01/01/2024 0:00	7.748.357.076.505.610	32.798.710.873.172.000	0.4324821725025618	3.092.192.442.122.600	0
3	01/01/2024 1:00	7.430.867.849.414.400	31.849.267.365.825.500	0.48554813292844823	4.139.614.989.220.490	0
4	01/01/2024 2:00	7.823.844.269.050.340	3.011.926.073.984.030	0.42075800790000384	45.863.944.665.783.500	0
5	01/01/2024 3:00	8.261.514.928.204.010	28.706.126.444.588.800	0.46920384703609597	688.768.765.734.019	1
6	01/01/2024 4:00	7.382.923.312.638.330	3.139.644.662.722.710	0.3106385330462143	5.556.553.124.533.750	0
7	01/01/2024 5:00	738.293.152.152.541	307.869.707.708.435	0.5213293707372939	3.664.518.430.900.100	0
8	01/01/2024 6:00	8.289.606.407.753.690	31.790.386.440.055.400	0.5001205475362224	5.486.036.289.417.630	0
9	01/01/2024 7:00	7.883.717.364.576.450	3.127.034.360.336.390	0.4182911368931809	3.452.696.011.061.630	0
10	01/01/2024 8:00	7.265.262.807.032.520	3.209.910.543.063.860	0.5659245668432406	60.826.910.538.870.700	0
11	01/01/2024 9:00	7.771.280.021.792.980	28.929.529.576.878.800	0.5937570137626559	4.528.875.348.182.120	0
12	01/01/2024 10:00	7.268.291.153.593.760	3.263.478.813.126.860	0.3392440135631026	4.906.363.811.260.820	0
13	01/01/2024 11:00	7.267.135.123.214.870	303.951.992.093.848	0.42372752214337855	632.579.664.754.219	0
14	01/01/2024 12:00	7.620.981.135.783.010	3.415.052.174.525.050	0.42308576502127015	371.283.643.141.575	0
15	01/01/2024 13:00	6.543.359.877.671.100	28.621.624.363.820.800	0.4060096898891135	3.602.881.814.887.000	0
16	01/01/2024 14:00	6.637.541.083.743.480	33.471.927.606.330.400	0.5829474841931749	4.416.400.673.653.990	0
17	01/01/2024 15:00	7.218.856.235.379.510	30.395.821.566.925.200	0.48061738612271965	6.038.378.851.019.310	0
18	01/01/2024 16:00	6.993.584.439.832.780	28.697.163.992.771.100	0.4735485158689389	34.806.539.245.231.500	0
19	01/01/2024 17:00	7.657.123.666.297.630	29.032.228.331.891.300	0.2996137635593247	2.167.844.401.889.430	0
20	01/01/2024 18:00	7.045.987.962.239.390	2.935.930.538.361.130	0.5635418030209226	45.488.411.415.910.000	0
21	01/01/2024 19:00	6.793.848.149.332.350	30.848.331.892.803.800	0.37607416051465276	555.174.079.991.092	0
22	01/01/2024 20:00	8.232.824.384.460.770	31.045.670.976.070.900	0.5059932630266328	6.200.261.749.738.820	1
23	01/01/2024 21:00	7.387.111.849.756.730	2.885.259.999.212.280	0.5277376937958191	4.536.838.637.914.600	0
24	01/01/2024 22:00	7.532.764.102.242.060	20.051.200.815.478.000	0.6260658505328651	45.885.732.028.820.000	0

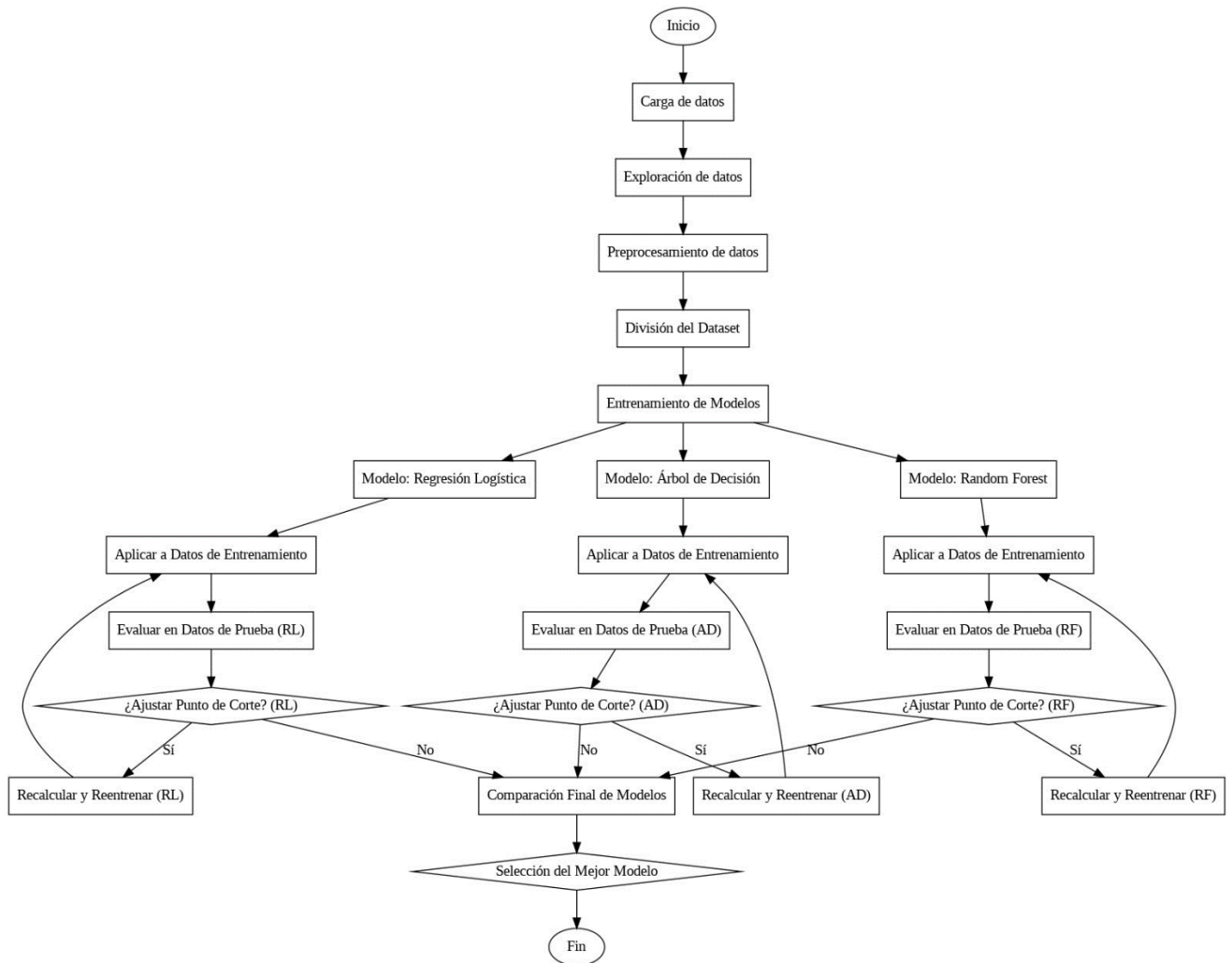
	A	B	C	D	E	F
979	10/02/2024 16:59	6.288.000.336.005.320	31.035.032.324.245.700	0.4133684022334204	6.447.303.039.391.800	0
980	10/02/2024 17:59	6.708.048.588.257.130	2.845.424.466.435.270	0.5543479378670093	55.681.032.051.286.400	0
981	10/02/2024 18:59	7.880.207.328.072.140	30.814.104.574.249.000	0.5199810433718713	3.950.344.501.090.050	0
982	10/02/2024 19:59	7.892.900.079.325.410	2.805.668.653.091.740	0.526402008133367	6.362.562.852.712.450	0
983	10/02/2024 20:59	7.712.728.780.892.480	2.724.076.367.042.210	0.6272353500106768	6.640.614.563.488.280	0
984	10/02/2024 21:59	7.016.511.928.435.390	2.874.656.555.082.000	0.5732492102779803	8.152.056.734.512.080	1
985	10/02/2024 22:59	747.614.432.192.901	3.172.478.694.476.550	0.5288723662825023	3.876.505.966.760.400	0
986	10/02/2024 23:59	7.498.198.730.454.710	31.906.250.090.752.100	0.3345112563771704	5.242.882.013.150.270	0
987	11/02/2024 0:59	6.920.817.655.403.720	31.026.170.840.194.500	0.40399537047205364	2.917.901.296.811.010	0
988	11/02/2024 1:59	8.251.699.150.883.570	31.450.191.573.695.900	0.48772910664042757	5.553.149.156.379.200	1
989	11/02/2024 2:59	7.938.681.145.287.830	31.032.356.574.229.900	0.5093372374276651	44.517.997.018.332.200	0
990	11/02/2024 3:59	7.389.517.913.085.160	28.717.036.809.554.000	0.3869796275326076	6.923.445.794.824.950	0
991	11/02/2024 4:59	7.513.442.919.497.260	3.086.384.508.934.780	0.7411676675578421	4.225.385.029.981.410	1
992	11/02/2024 5:59	7.604.191.403.973.770	31.600.819.050.455.300	0.6516394055253573	3.310.816.962.032.730	0
993	11/02/2024 6:59	6.479.132.565.787.850	31.508.582.659.627.100	0.5602118320392484	452.873.625.929.748	0
994	11/02/2024 7:59	7.376.411.308.739.700	3.237.782.674.294.920	0.5072036861283745	30.245.122.287.828.600	0
995	11/02/2024 8:59	7.159.007.876.001.100	31.416.607.694.402.100	0.47877910300988485	575.109.945.450.731	0
996	11/02/2024 9:59	6.999.189.995.052.540	30.702.896.415.083.100	0.4048081539977764	29.349.169.537.436.700	0
997	11/02/2024 10:59	7.359.449.853.557.020	3.214.030.047.663.280	0.5077480518518862	5.028.457.581.183.150	0
998	11/02/2024 11:59	8.398.843.263.424.760	2.994.695.748.150.180	0.525775254014275	2.922.188.180.143.800	0
999	11/02/2024 12:59	7.820.421.430.633.500	28.236.250.697.544.500	0.3758239424518376	4.679.702.195.463.350	0
1000	11/02/2024 13:59	7.214.410.505.108.600	2.967.386.607.211.240	0.5334176417244898	6.643.378.155.804.530	1
1001	11/02/2024 14:59	7.786.291.390.678.070	285.101.947.103.871	0.48447409511752515	5.360.647.891.435.790	0
1002						

Anexo. 4

Sensores de medición

Sensor de presión (transductor de presión no incendiario M3200)	sensor de vibración (acelerómetro de 3 cables TO-5 modelo 805M1)	sensor de temperatura (RTD – NB- PTCO-035)	sensor de humedad (FC-28: Sensor de humedad del suelo)
			

Algoritmo de Machine Learning



Turnitin Informe de Originalidad

Visualizador de documentos

Procesado el: 23-ago.-2024 8:20 p. m. -05
Identificador: 2437044663
Número de palabras: 13938
Entregado: 1

Tesis_Titulo Por DANIEL AUGUSTO ALVAREZ CAMPOS

Índice de similitud	Similitud según fuente
18%	Internet Sources: 14% Publicaciones: 4% Trabajos del estudiante: 9%

excluir citas	Excluir bibliografía	excluir las coincidencias menores	modo: ver informe en vista quickview (vista clásica)	imprimir	descargar
1% match (Internet desde 29-nov.-2016) http://www.theibr.com					
<1% match (trabajos de los estudiantes desde 19-jun.-2023) Submitted to Universidad Internacional de La Rioja on 2023-06-19					
<1% match (trabajos de los estudiantes desde 02-feb.-2024) Submitted to Universidad Internacional de La Rioja on 2024-02-02					
<1% match (trabajos de los estudiantes desde 19-may.-2024) Submitted to Universidad Internacional de La Rioja on 2024-05-19					
<1% match (trabajos de los estudiantes desde 20-sept.-2023) Submitted to Universidad Internacional de La Rioja on 2023-09-20					
<1% match (trabajos de los estudiantes desde 05-may.-2023) Submitted to Universidad Internacional de La Rioja on 2023-05-05					
<1% match (trabajos de los estudiantes desde 25-ene.-2024) Submitted to Universidad Internacional de La Rioja on 2024-01-25					
<1% match (trabajos de los estudiantes desde 03-jul.-2024) Submitted to Universidad Internacional de La Rioja on 2024-07-03					
<1% match (trabajos de los estudiantes desde 18-jul.-2023) Submitted to Universidad Internacional de La Rioja on 2023-07-18					
<1% match (trabajos de los estudiantes desde 19-jul.-2023) Submitted to Universidad Internacional de La Rioja on 2023-07-19					
<1% match (trabajos de los estudiantes desde 19-may.-2024) Submitted to Universidad Internacional de La Rioja on 2024-05-19					
<1% match () TALTEMA MODERNO: JFCUIC TALTEMA MODERNO: JFCUIC (Clases de identificación textual y análisis de textos: Un análisis comparativo) Universidad Autónoma del Estado de México					