



Universidad Católica de Santa María

**Facultad de Ciencias e Ingenierías Físicas y Formales
Escuela Profesional de Ingeniería Industrial**

**Propuesta para la optimización del proceso de mantenimiento
industrial mediante el análisis de datos con herramientas de
inteligencia artificial**

Tesis presentada por:

Canevaro Guevara, Joaquin Sebastian

ORCID: 0009-0006-3716-2099

para optar el Título Profesional de Ingeniero Industrial

Asesor:

Mag. Urday Luna, Ferly Elmer

ORCID: 0000-0002-9340-4017

Arequipa - Perú

2026

UNIVERSIDAD CATÓLICA DE SANTA MARÍA

INGENIERIA INDUSTRIAL

TITULACIÓN CON TESIS

DICTAMEN APROBACIÓN DE BORRADOR

Arequipa, 19 de Julio del 2026

Dictamen: 018105-C-EPII-2026

Visto el borrador del expediente 018105, presentado por:

2019176201 - CANEVARO GUEVARA JOAQUIN SEBASTIAN

Titulado:

**PROPUESTA PARA LA OPTIMIZACIÓN DEL PROCESO DE MANTENIMIENTO INDUSTRIAL
MEDIANTE EL ANÁLISIS DE DATOS CON HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL**

Nuestro dictamen es:

APROBADO

Titulo Profesional/Titulo de Segunda Especialidad/Grado Académico a optar:

INGENIERO INDUSTRIAL

**29267647 - VALENCIA BECERRA ROLARDI MARIO
DICTAMINADOR**



**29434502 - MURILLO QUISPE EFRAIN RAFAEL
DICTAMINADOR**



**29653773 - CARRASCO BOCANGEL JULIO CESAR
DICTAMINADOR**



Propuesta para la optimización del proceso de mantenimiento industrial mediante el análisis de datos con herramientas de inteligencia artificial

INFORME DE ORIGINALIDAD

13%

INDICE DE SIMILITUD

11%

FUENTES DE INTERNET

5%

PUBLICACIONES

7%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	repositorio.ucsm.edu.pe	1%
	Fuente de Internet	
2	dgsa.uaeh.edu.mx:8080	1%
	Fuente de Internet	
3	Submitted to Universidad Internacional del Ecuador	1%
	Trabajo del estudiante	
4	Submitted to Universidad Católica de Santa María	<1%
	Trabajo del estudiante	
5	oa.upm.es	<1%
	Fuente de Internet	
6	Submitted to Universidad de Guayaquil	<1%
	Trabajo del estudiante	
7	repositorio.usm.cl	<1%
	Fuente de Internet	
8	Submitted to Universidad San Marcos	<1%
	Trabajo del estudiante	
9	Submitted to Uniminuto Virtual	<1%
	Trabajo del estudiante	
10	editorial.edulearn.ec	<1%
	Fuente de Internet	
11	repositorio.unap.edu.pe	<1%
	Fuente de Internet	
12	Submitted to Universidad Nacional Tecnológica De Lima Sur	<1%
	Trabajo del estudiante	
13	repositorio.comillas.edu	<1%
	Fuente de Internet	
14	Submitted to FUNIBER	<1%
	Trabajo del estudiante	

Submitted to ITESM: Instituto Tecnológico y de Estudios Superiores de Monterrey

DEDICATORIA

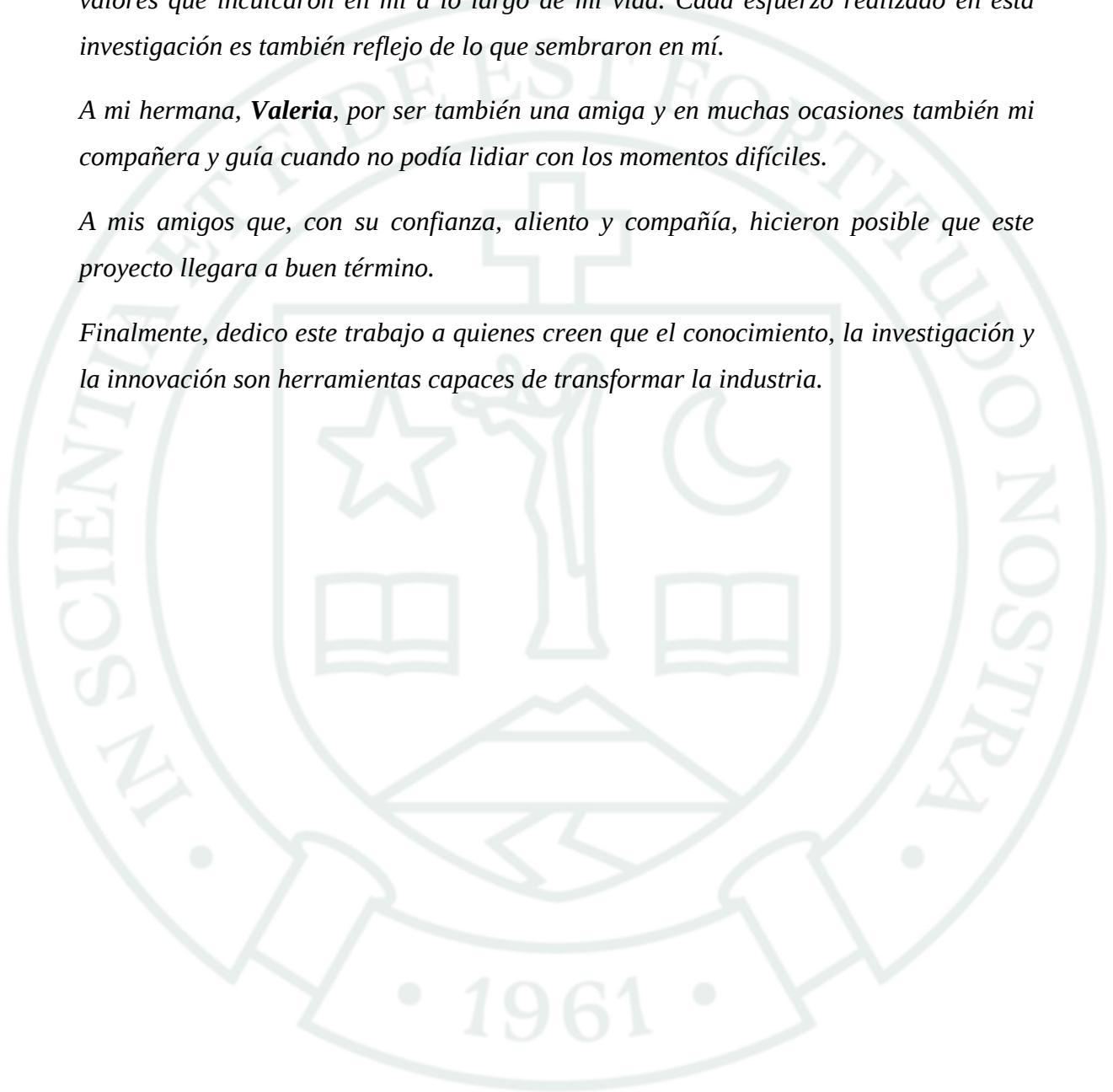
A **Dios** por ser mi guía y fortaleza en cada etapa de mi vida, brindándome la sabiduría, la perseverancia y las oportunidades necesarias para culminar esta meta.

A mis padres, **Cesar y Karim**, por su amor incondicional, sus enseñanzas, el apoyo y los valores que inculcaron en mí a lo largo de mi vida. Cada esfuerzo realizado en esta investigación es también reflejo de lo que sembraron en mí.

A mi hermana, **Valeria**, por ser también una amiga y en muchas ocasiones también mi compañera y guía cuando no podía lidiar con los momentos difíciles.

A mis amigos que, con su confianza, aliento y compañía, hicieron posible que este proyecto llegara a buen término.

Finalmente, dedico este trabajo a quienes creen que el conocimiento, la investigación y la innovación son herramientas capaces de transformar la industria.



AGRADECIMIENTOS

A Dios por concederme la fortaleza, la sabiduría y la perseverancia necesarias para culminar esta etapa de formación profesional. Su guía ha sido fundamental en cada desafío y logro alcanzado.

A mis padres y a mi familia, por su amor incondicional, comprensión y apoyo constantes. Gracias por ser mi principal fuente de motivación, por confiar siempre en mí y por impulsarme a seguir adelante, incluso en los momentos difíciles.

Mi sincero agradecimiento a mi asesor de tesis, el Mag. Ferly Elmer Urday Luna, por su orientación, conocimientos y valiosas observaciones, las cuales contribuyeron significativamente al desarrollo y fortalecimiento de esta investigación. Asimismo, agradezco a los docentes que formaron parte de mi preparación académica y que, con su experiencia y compromiso, enriquecieron mi formación profesional; en especial al Dr. Julio Cesar Carrasco Bocangel, Dr. Efraín Rafael Murillo Quispe y al Dr. Rolardi Mario Valencia Becerra por su valiosa asesoría, dedicación y acompañamiento durante el desarrollo de este trabajo.

Finalmente, agradezco a todas las personas que, de una forma u otra, contribuyeron a la realización de este trabajo. Cada consejo, palabra de aliento y muestra de confianza representó un impulso para alcanzar esta meta.

RESUMEN

La presente investigación tuvo como objetivo presentar una propuesta para la optimización del proceso de mantenimiento industrial mediante el análisis de datos con herramientas de Inteligencia Artificial, orientada a fortalecer la toma de decisiones y mejorar la asignación de recursos en las actividades de mantenimiento. La investigación se desarrolló bajo un enfoque cuantitativo, de tipo aplicada y con un diseño no experimental basado en una validación retrospectiva, utilizando una base de datos histórica recopilada a partir de registros de mantenimiento predictivo obtenidos de diversas fuentes disponibles en Internet.

La propuesta se estructuró en tres niveles. En el primero, se desarrolló un modelo predictivo basado en técnicas de Machine Learning para estimar la condición de los equipos. En el segundo, se implementó un modelo de optimización mediante programación dinámica, formulado como un problema de mochila 0-1, con el propósito de priorizar las intervenciones de mantenimiento bajo restricciones de tiempo. Finalmente, se realizó una evaluación comparativa entre un procedimiento tradicional de priorización y la propuesta desarrollada mediante una validación retrospectiva.

Los resultados evidenciaron que el algoritmo Random Forest presentó el mejor desempeño entre los modelos evaluados, permitiendo generar probabilidades confiables para la clasificación de la condición de los equipos. Asimismo, el modelo de optimización permitió identificar las intervenciones con mayor beneficio esperado bajo las restricciones establecidas, mientras que la comparación retrospectiva mostró una mejora respecto al procedimiento tradicional de priorización.

En conclusión, la investigación evidencia la factibilidad técnica de integrar herramientas de análisis de datos, Machine Learning y programación dinámica para apoyar la gestión del mantenimiento industrial. La propuesta constituye una alternativa para priorizar intervenciones y optimizar la utilización de los recursos disponibles, contribuyendo a la incorporación de tecnologías asociadas con la Industria 4.0.

Palabras clave: mantenimiento industrial, inteligencia artificial, machine learning

ABSTRACT

The objective of this research was to present a proposal for optimizing the industrial maintenance process through data analysis using Artificial Intelligence tools, aimed at strengthening decision-making and improving resource allocation in maintenance activities. The study employed a quantitative, applied approach with a non-experimental design based on retrospective validation, utilizing a historical database compiled from predictive maintenance records obtained from various online sources.

The proposal was structured across three levels. At the first level, a predictive model based on machine learning techniques was developed to estimate equipment condition. At the second level, an optimization model using dynamic programming—formulated as a 0-1 knapsack problem—was implemented to prioritize maintenance interventions under time constraints. Finally, a comparative evaluation was conducted between a traditional prioritization procedure and the developed proposal through a retrospective validation.

The results demonstrated that the Random Forest algorithm exhibited the best performance among the evaluated models, enabling the generation of reliable probabilities for classifying equipment condition. Furthermore, the optimization model identified the interventions offering the greatest expected benefit within the established constraints, while the retrospective comparison showed an improvement over the traditional prioritization procedure.

In conclusion, the research demonstrates the technical feasibility of integrating data analysis tools, machine learning, and dynamic programming to support industrial maintenance management. The proposal offers an alternative for prioritizing interventions and optimizing the use of available resources, contributing to the adoption of technologies associated with Industry 4.0.

Keywords: industrial maintenance, artificial intelligence, machine learning

ÍNDICE

DEDICATORIA

AGRADECIMIENTOS

RESUMEN

ABSTRACT

INTRODUCCIÓN 1

CAPÍTULO I..... 3

1. Planteamiento del Problema..... 4

1.1. Identificación del Problema 4

1.2. Descripción del problema..... 4

1.3. Síntesis 6

1.4. Formulación del Problema 7

1.4.1. Problema General:..... 7

1.4.2. Problemas Específicos: 7

1.5. Objetivos de la Investigación 7

1.5.1. Objetivo General 7

1.5.2. Objetivo Específicos 7

1.6. Hipótesis..... 8

1.6.1. Variables..... 8

1.6.2. Tabla de operacionalización de variables..... 11

1.7. Justificación de la Investigación 13

1.7.1. Justificación Económica:..... 13

1.7.2. Justificación Ambiental:..... 13

1.7.3. Justificación Técnica: 13

1.7.4. Justificación Académica.....	13
1.8. Metodología	13
1.8.1. Enfoque de la investigación	13
1.8.2. Tipo de investigación	14
1.8.3. Diseño de la investigación.....	15
1.8.4. Definición y cálculo de indicadores	15
1.8.5. Población.....	17
1.8.6. Muestra.....	17
1.8.7. Técnicas de recolección de datos	18
1.8.8. Instrumentos de Investigación.....	19
1.8.9. Técnicas de análisis de datos.....	20
1.9. Alcance.....	20
1.10. Desafíos y Limitaciones	21
CAPÍTULO II	23
2. Marco Teórico	24
2.1. Antecedentes de la Investigación	24
2.1.1. Antecedentes Internacionales	24
2.1.2. Antecedentes Nacionales.....	26
2.1.3. Antecedentes locales	31
2.2. Marco Teórico.....	33
2.2.1. Mantenimiento Industrial	33
2.2.2. Gestión del Mantenimiento Industrial.....	33
2.2.3. Tipos de Mantenimiento.....	34
2.2.4. Indicadores de Desempeño del Mantenimiento	35
2.2.5. Industria 4.0 y Mantenimiento Inteligente	35
2.2.6. Análisis de datos en el Mantenimiento Industrial	36

2.2.7. Inteligencia Artificial aplicada al Mantenimiento Industrial.....	37
2.2.8. Machine Learning en el Mantenimiento Predictivo	37
2.2.9. Regresión Logística.....	41
2.2.10. Máquinas de Vectores de Soporte (SVM).....	41
2.2.11. Deep Learning y Análisis de Series Temporales.....	42
2.2.12. Integración de la Inteligencia Artificial con la Gestión del Mantenimiento	42
2.2.13. Internet Industrial de las Cosas (IIoT).....	43
2.2.14. Optimización del Proceso de Mantenimiento.....	43
2.2.15. Programación Dinámica.....	44
2.2.16. Vacíos Teóricos Identificados y Tendencias Emergentes	47
CAPÍTULO III	48
3. Diagnóstico Situacional.....	49
3.1. Enfoque del diagnóstico.....	49
3.2. Contexto del sector y análisis PESTEL.....	50
3.2.1. Priorización de factores PESTEL mediante matriz de enfrentamiento	50
3.2.2. Factor Tecnológico.....	52
3.2.3. Factor Económico	55
3.2.4. Factor Legal.....	58
3.2.5. Factor Político	60
3.2.6. Factor Social.....	63
3.2.7. Factor Ecológico	68
3.3. Descripción del sistema actual de mantenimiento	70
3.4. Identificación de problemáticas en el mantenimiento industrial.....	72
3.5. Análisis de la situación actual del mantenimiento	76
3.6. Análisis de causas del problema.....	79
3.6.1. Máquina.....	82

3.6.2. Método	83
3.6.3. Mano de Obra.....	84
3.6.4. Tecnología.....	85
3.6.5. Información	87
3.6.6. Gestión	88
3.6.7. Síntesis del análisis causal.....	89
3.7. Evaluación de alternativas de solución	89
3.8. Justificación de la propuesta	91
CAPÍTULO IV	93
4. Propuesta del modelo	94
4.1. Base de datos utilizada para el entrenamiento del modelo.....	94
4.2. Selección del modelo de Machine Learning	99
4.3. Modelo Predictivo de Fallas.....	102
4.3.1. Introducción del Modelo	102
4.3.2. Fase 1: Limpieza y preprocesamiento de datos	104
4.3.3. Fase 2: Construcción de la variable objetivo.....	107
4.3.4. Fase 3: Ingeniería de características (Feature Engineering).....	107
4.3.5. Fase 4: Codificación de variables.....	110
4.3.6. Fase 5: División del conjunto de datos.....	110
4.3.7. Fase 6: Entrenamiento de modelos de Machine Learning.....	111
4.3.8. Fase 7: Evaluación del modelo.....	118
4.3.9. Fase 8: Interpretación de resultados	120
4.3.10. Fase 9: Generación de probabilidades de clasificación.....	120
4.4. Modelo de Priorización y Optimización	123
4.5. Evaluación Comparativa y validación de la propuesta	124
CAPÍTULO V	127

5. Resultados y Discusión	128
5.1. Resultados de la variable independiente	128
5.2. Análisis económico de la propuesta	129
5.2.1. Estructura económica de la propuesta	129
5.2.2. Modelo de optimización económica.....	131
5.2.3. Resultado económico de la optimización.....	132
5.2.4. Análisis del costo de oportunidad de la capacidad disponible	135
5.2.5. Análisis de sensibilidad económica.....	137
5.2.6. Evaluación de VAN, TIR, ROI y periodo de recuperación.....	138
5.2.7. Modelo financiero propuesto para una futura implementación.....	140
5.2.8. Interpretación integral del análisis económico	141
5.2.9. Alcance del análisis económico	142
5.3. Discusión de resultados.....	143
5.4. Validación de la hipótesis	144
CONCLUSIONES	145
RECOMENDACIONES	147
REFERENCIAS BIBLIOGRÁFICAS	149
ANEXOS.....	158

ÍNDICE DE TABLAS

Tabla 1. Variable independiente y dependiente de la investigación.....	11
Tabla 2. Resultados de Parámetros Mejorados.....	28
Tabla 3. Matriz de enfrentamiento de factores PESTEL.....	50
Tabla 4. Priorización de factores PESTEL mediante análisis de impacto.....	51
Tabla 5. Evolución de los enfoques de mantenimiento.....	54
Tabla 6. Estructura empresarial peruana según tamaño de empresa	56
Tabla 7. Impacto económico de las estrategias de mantenimiento.....	57
Tabla 8. Relación entre mantenimiento y cumplimiento normativo	59
Tabla 9. Contribución del mantenimiento predictivo al cumplimiento normativo	60
Tabla 10. Principales políticas nacionales relacionadas con la transformación digital y competitividad.....	61
Tabla 11. Impacto de las políticas públicas sobre la gestión del mantenimiento industrial	62
Tabla 12. Relación entre aspectos sociales y gestión del mantenimiento	64
Tabla 13. Impacto de la transformación digital sobre los recursos humanos.....	65
Tabla 14. Relación entre mantenimiento y sostenibilidad ambiental.....	69
Tabla 15. Contribución del modelo propuesto a la sostenibilidad	70
Tabla 16. Tipo de mantenimiento predominante.....	77
Tabla 17. Frecuencia de fallas no planificadas.....	77
Tabla 18. Criterios de priorización de intervenciones.....	78
Tabla 19. Información registrada en el proceso de mantenimiento.....	78
Tabla 20. Uso de información para análisis y toma de decisiones	79
Tabla 21. Factores limitantes para la mejora del mantenimiento	79
Tabla 22. Matriz de enfrentamiento de criterios	90
Tabla 23. Pesos normalizados en la evaluación	90

Tabla 24. Criterios de evaluación de la matriz de decisión.....	90
Tabla 25. Matriz de evaluación de alternativas.....	91
Tabla 26. Resultados Ponderados.....	91
Tabla 27. Diccionario de datos de las variables utilizadas en el modelo	96
Tabla 28. Disponibilidad temporal de las variables utilizadas en el modelo	97
Tabla 29. Matriz de enfrentamiento de criterios para la selección de modelos de Machine Learning	99
Tabla 30. Pesos de los criterios para selección de modelos de Machine Learning normalizados	99
Tabla 31. Criterios de Evaluación	100
Tabla 32. Matriz de Evaluación de Modelos de Machine Learning.....	100
Tabla 33. Resultado ponderado de Evaluación de Modelos.....	101
Tabla 34. Cuadro Comparativo: Método Tradicional vs Modelo Propuesto	125
Tabla 35. Evaluación del Modelo de Inteligencia Artificial	128
Tabla 36. Comparación económica del procedimiento tradicional y el modelo propuesto.....	133
Tabla 37. Información disponible para los indicadores de VAN-TIR-ROI-Payback	139
Tabla 38. Construcción del flujo financiero	141

ÍNDICE DE FIGURAS

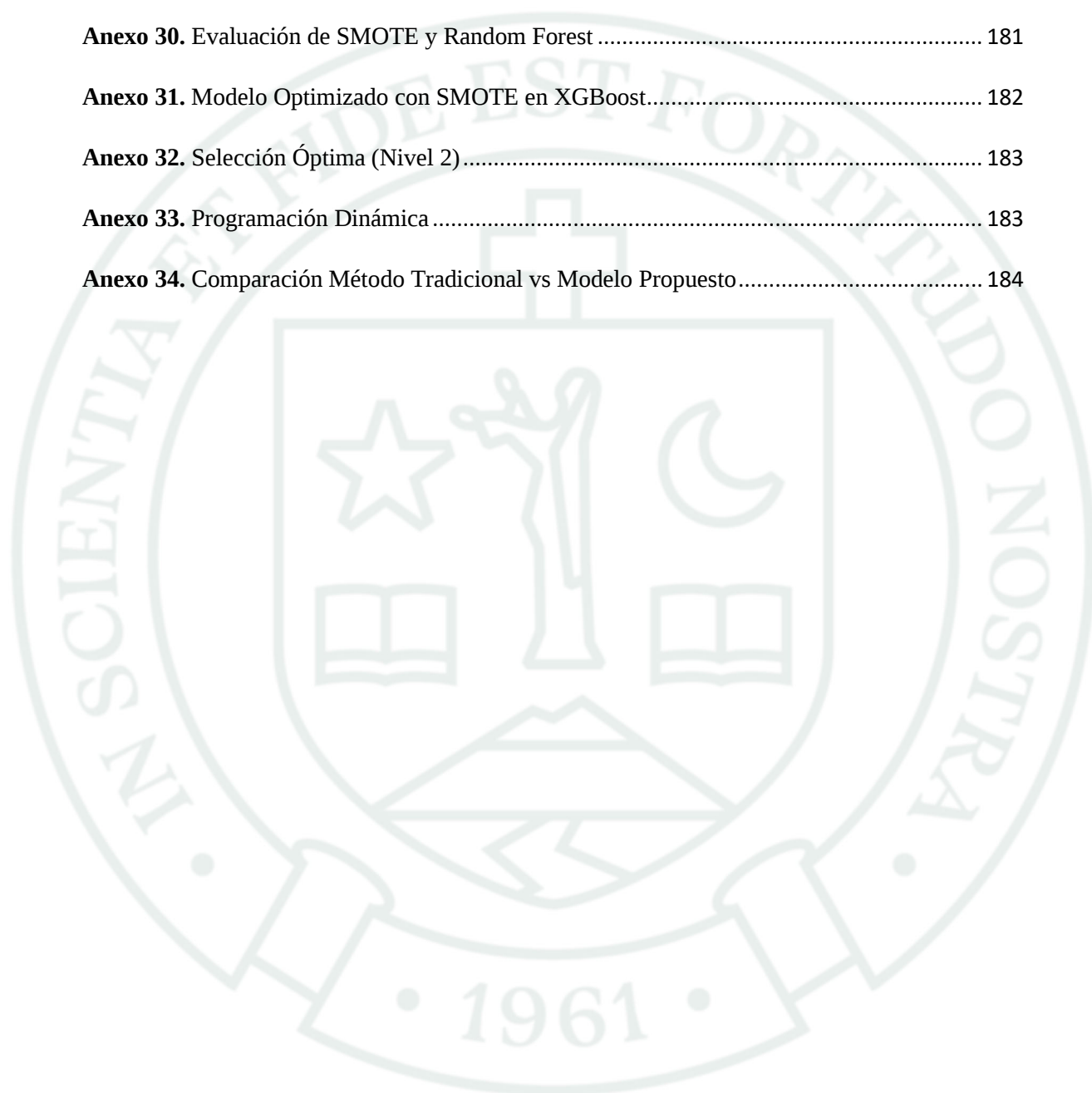
Figura 1. Descripción general de los Tipos de Mantenimiento.....	24
Figura 2. Pasos principales para el Diseño de un Modelo de Aprendizaje Automático	25
Figura 3. Identificación de patrones mediante diferentes técnicas de monitoreo.....	25
Figura 4. Flujo de Machine Learning y Deep Learning	26
Figura 5. Modelo de Parámetros obtenidos.....	27
Figura 6. Arquitectura para la implementación de Machine Learning, XGBoost.....	29
Figura 7. Comparación de la Precisión de los Modelos de Machine Learning.....	30
Figura 8. Resultados Modelo Predictivo.....	32
Figura 9. Esquema PESTEL aplicado a la investigación	52
Figura 10. Análisis PESTEL del sector transporte y logística	52
Figura 11. Línea de tiempo de la Evolución de las Políticas Públicas a la Transformación Digital en el Perú.....	63
Figura 12. Impacto social de una gestión eficiente del mantenimiento (BPMN simplificado) .	66
Figura 13. Impacto del mantenimiento sobre la seguridad laboral (BPMN simplificado)	67
Figura 14. Esquema secuencial del proceso actual de mantenimiento.....	71
Figura 15. Tipo de mantenimiento predominante en las empresas encuestadas	72
Figura 16. Principal problemática del mantenimiento en las empresas encuestadas	72
Figura 17. Frecuencia de ocurrencia de fallas no planificadas (1: Muy baja - 5: Muy alta).....	73
Figura 18. Criterios utilizados para la priorización de intervenciones.....	74
Figura 19. Percepción sobre la eficiencia de la priorización de intervenciones.....	74
Figura 20. Frecuencia de problemas relacionados con la priorización de intervenciones	75
Figura 21. Factores que limitan la mejora del mantenimiento	76
Figura 22. Diagrama de Ishikawa	81
Figura 23. Muestra representativa de la base de datos utilizada	98
Figura 24. Metodología general del Modelo	103

Figura 25. Flujo de limpieza de preprocesamiento de datos	106
Figura 26. Proceso de Ingeniería de Características	109
Figura 27. Proceso de entrenamiento de Modelos de Machine Learning	114
Figura 28. Aplicación del Modelo Random Forest	115
Figura 29. Aplicación del Modelo XGBoost	116
Figura 30. Resultados del modelo aplicando Random Forest y XGBoost	117
Figura 31. Resultados del modelo de Random Forest Optimizado con SMOTE.....	117
Figura 32. Resultados del modelo de XGBoost Optimizado con SMOTE	118
Figura 33. Evaluación del Desempeño Predictivo	119
Figura 34. Generación de Probabilidades	122
Figura 35. Comparación del Beneficio Total.....	134
Figura 36. Beneficio por Hora	135
Figura 37. Utilización del Tiempo Disponible.....	135

ÍNDICE DE ANEXOS

Anexo 1. Ficha de Validación de la Encuesta aplicada	159
Anexo 2. Evaluación del Instrumento.....	160
Anexo 3. Firma del Evaluador Experto.....	160
Anexo 4. Código completo de Google Colab #1	161
Anexo 5. Código completo de Google Colab #2	162
Anexo 6. Código completo de Google Colab #3	163
Anexo 7. Código completo de Google Colab #4	164
Anexo 8. Código completo de Google Colab #5	165
Anexo 9. Código completo de Google Colab #6	166
Anexo 10. Código completo de Google Colab #7	167
Anexo 11. Código completo de Google Colab #8	168
Anexo 12. Código completo de Google Colab #9	169
Anexo 13. Código completo de Google Colab #10	170
Anexo 14. Código completo de Google Colab #11	171
Anexo 15. Código completo de Google Colab #12	172
Anexo 16. Código completo de Google Colab #13	173
Anexo 17. Código completo de Google Colab #14	174
Anexo 18. Fase 1: Limpieza y Preprocesamiento.....	175
Anexo 19. Fase 2: Variable Objetivo.....	175
Anexo 20. Fase 3: Feature Engineering.....	176
Anexo 21. Fase 4: Codificación.....	176
Anexo 22. Fase 5: División Temporal	177
Anexo 23. Entrenamiento	177
Anexo 24. Fase 7: Evaluación.....	178
Anexo 25. Fase 8: Importancia de variables	179

Anexo 26. Gráfica de la Importancia de Variables	179
Anexo 27. Fase 9: Probabilidades.....	180
Anexo 28. Balanceo de Técnica SMOTE	180
Anexo 29. Modelo Optimizado con SMOTE en Random Forest	181
Anexo 30. Evaluación de SMOTE y Random Forest	181
Anexo 31. Modelo Optimizado con SMOTE en XGBoost.....	182
Anexo 32. Selección Óptima (Nivel 2).....	183
Anexo 33. Programación Dinámica	183
Anexo 34. Comparación Método Tradicional vs Modelo Propuesto.....	184



INTRODUCCIÓN

En la actualidad, las organizaciones industriales operan en entornos altamente competitivos, caracterizados por una creciente automatización, mayores exigencias de calidad, presión por la reducción de costos y la necesidad de garantizar la continuidad operativa de sus procesos productivos. En este contexto, la gestión eficiente del mantenimiento industrial se ha convertido en un factor estratégico para asegurar la disponibilidad, confiabilidad y vida útil de los activos físicos, así como sostener la productividad y rentabilidad de las empresas.

Tradicionalmente, el mantenimiento industrial ha sido abordado mediante enfoques correctivos y preventivos, los cuales, si bien han permitido mantener operativos los equipos, presentan limitaciones significativas frente a sistemas productivos cada vez más complejos y altamente instrumentados. La ocurrencia de fallas no planificadas, los tiempos de inactividad prolongados y el uso ineficiente de recursos técnicos continúan generando impactos negativos tanto en los costos operativos como en la competitividad organizacional.

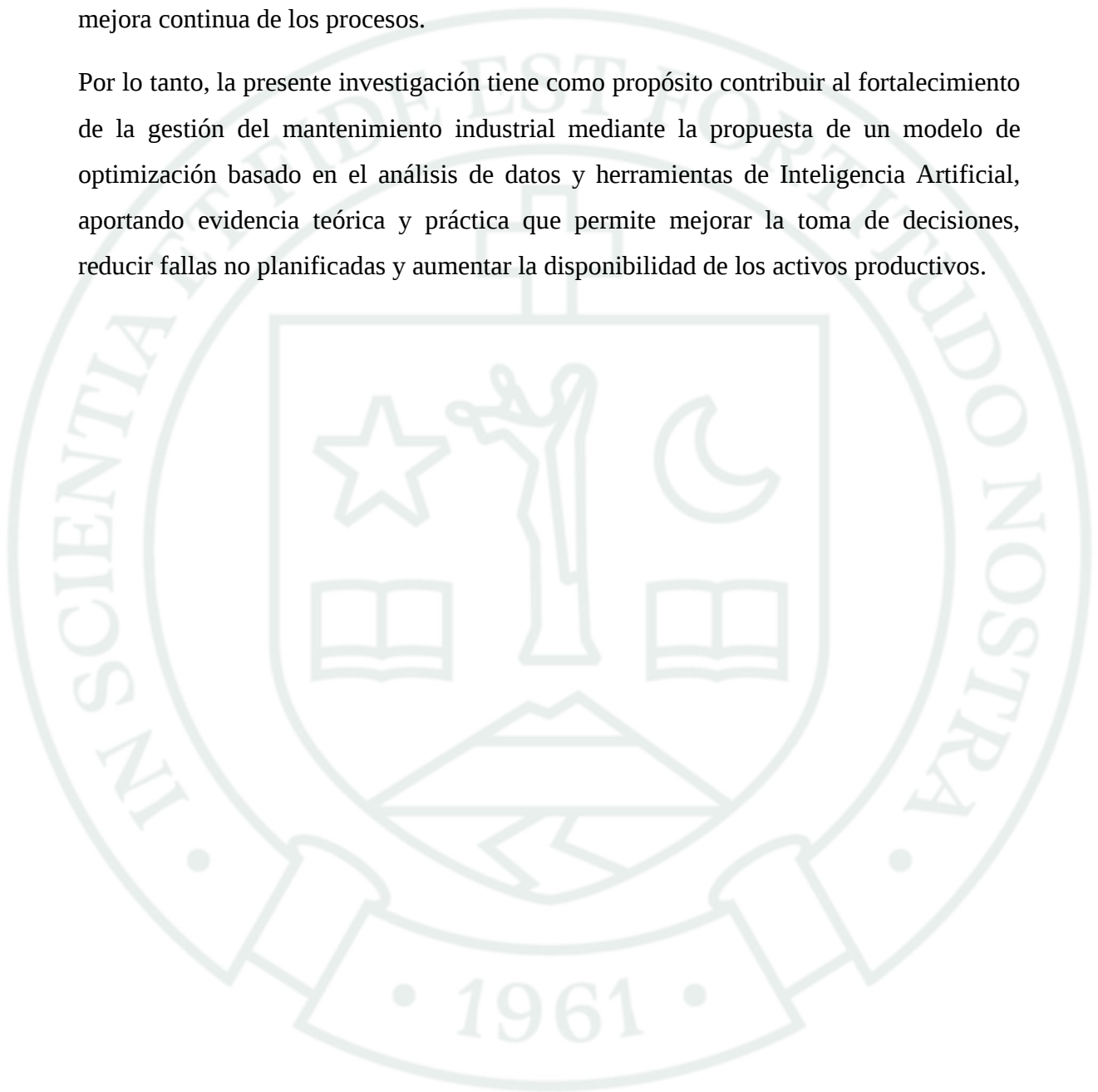
Paralelamente, el avance de la industria 4.0 ha impulsado la incorporación de tecnologías digitales como el Internet Industrial de las Cosas (IoT), el análisis de grandes volúmenes de datos (Big Data) y la Inteligencia Artificial (IA), las cuales han abierto nuevas oportunidades para transformar la gestión del mantenimiento.

Estas tecnologías permiten recopilar, procesar y analizar datos operativos en tiempo real, facilitando la detección temprana de fallas, la predicción del comportamiento de los equipos y la toma de decisiones basada en información objetiva.

En particular, la aplicación de herramientas de Inteligencia Artificial, tales como técnicas de Machine Learning y Deep Learning, ha demostrado un alto potencial para optimizar el proceso de mantenimiento industrial, al identificar patrones complejos de degradación que no pueden ser detectados mediante métodos tradicionales. No obstante, a pesar de los avances científicos y tecnológicos reportados en la literatura, muchas organizaciones aún enfrentan dificultades para integrar estas herramientas de manera efectiva en sus procesos de mantenimiento, especialmente en contextos donde existen limitaciones tecnológicas, organizacionales y de gestión de datos.

En ese escenario, surge la necesidad de desarrollar investigaciones aplicadas que analicen y evalúen el impacto del uso de análisis de datos e Inteligencia Artificial en la optimización del mantenimiento industrial, desde un enfoque propio de la ingeniería industrial. Este enfoque permite no solo considerar el desempeño técnico de los equipos, sino también la eficiencia operativa, la asignación de recursos, los costos y la mejora continua de los procesos.

Por lo tanto, la presente investigación tiene como propósito contribuir al fortalecimiento de la gestión del mantenimiento industrial mediante la propuesta de un modelo de optimización basado en el análisis de datos y herramientas de Inteligencia Artificial, aportando evidencia teórica y práctica que permite mejorar la toma de decisiones, reducir fallas no planificadas y aumentar la disponibilidad de los activos productivos.





CAPÍTULO I

1. Planteamiento del Problema

1.1. Identificación del Problema

En el contexto industrial actual, las organizaciones enfrentan crecientes desafíos para garantizar la continuidad operativa de sus activos productivos. Los procesos de mantenimiento, tradicionalmente basados en rutinas programadas o acciones correctivas reactivas, suelen generar paradas no planificadas, sobrecostos operativos, baja disponibilidad de equipos y un uso ineficiente de los recursos técnicos.

A nivel nacional y global, se observa que muchas plantas industriales aún no utilizan plenamente datos operacionales ni herramientas de inteligencia artificial (IA) para optimizar la toma de decisiones en mantenimiento, a pesar del aumento del volumen de datos provenientes de sensores, sistemas SCADA, PLCs y plataformas IoT.

1.2. Descripción del problema

El proceso de mantenimiento industrial presenta tres problemas centrales:

- **Falta de predicción de fallas:**
La mayoría de industrias aún se basan en modelos preventivos rígidos que no consideran patrones reales de degradación, generando mantenimiento excesivo o insuficiente.
- **Subutilización de datos operativos:**
Los sistemas industriales recolectan datos de vibración, aceite, temperatura, presión o consumo energético, pero no son analizados con modelos avanzados.
- **Limitada capacidad de decisión en tiempo real:**
La ausencia de modelo de inteligencia artificial limita la capacidad para anticipar fallas, priorizar intervenciones o asignar recursos de manera óptima.

Como resultado, las empresas experimentan mayores costos por mantenimiento correctivo, disminución de vida útil de equipos y menor confiabilidad del sistema productivo.

Asimismo, en los últimos años la gestión del mantenimiento industrial ha cobrado una importancia estratégica debido al incremento de la automatización, la complejidad de los sistemas productivos y la necesidad de garantizar altos niveles de disponibilidad y confiabilidad de los activos físicos. Diversos estudios recientes coinciden en que las

fallas no planificadas continúan siendo una de las principales causas de pérdidas económicas en el sector industrial, afectando la continuidad operativa, los costos de producción y la competitividad empresarial. (Carvalho, 2019)

A partir de 2019, la literatura científica evidencia un crecimiento significativo en investigaciones orientadas al mantenimiento predictivo basado en análisis de datos e Inteligencia Artificial, impulsado por la consolidación de la Industria 4.0 y el uso masivo de sensores industriales, sistemas IoT y plataformas de Big Data. Estas investigaciones destacan que los enfoques tradicionales de mantenimiento preventivo resultan insuficientes para entornos industriales dinámicos, ya que no consideran el estado real de los equipos ni los patrones complejos de degradación. (Zonta et al., 2020)

Estudios recientes demuestran que la aplicación de técnicas de Machine Learning, tales como Random Forest, Support Vector Machines y Gradient Boosting, permite mejorar significativamente la precisión en la predicción de fallas en maquinaria industrial, superando a los métodos estadísticos convencionales. Asimismo, el uso de Deep Learning, especialmente redes neuronales recurrentes (LSTM) y autoencoders, ha mostrado un alto desempeño en el análisis de series temporales provenientes de sensores industriales. (Zhao,2019; Wen,2020)

Entre los principales beneficios reportados por la literatura se encuentran la reducción de tiempos de inactividad no planificados, el incremento del MTBF, la disminución del MTTR y una optimización global de los costos de mantenimiento. Investigaciones aplicadas realizadas entre 2020 y 2022 indican que la implementación de sistemas de mantenimiento predictivo basados en Inteligencia Artificial puede reducir hasta en un 25%-35% los costos de mantenimiento y mejorar la disponibilidad operativa de los equipos en más de un 20%. (Selcuk, 2017; Zhang, 2021)

No obstante, a pesar de estos avances, diversos autores señalan que aún existen limitaciones significativas para la adopción efectiva de la Inteligencia Artificial en el mantenimiento industrial, especialmente en contextos reales de operación. Entre los principales desafíos identificados se encuentran la baja calidad y heterogeneidad de los datos industriales, la falta de integración entre los modelos de IA y los sistemas de gestión de mantenimiento (CMMS), así como la escasa capacitación del personal técnico en análisis de datos. (Bokrantz, 2020)

Asimismo, investigaciones recientes resaltan la necesidad de evaluar no solo el desempeño técnico de los modelos predictivos, sino también su impacto en la gestión del mantenimiento desde una perspectiva de Ingeniería Industrial, considerando variables como la planificación de recursos, la toma de decisiones, los costos operativos y la sostenibilidad del sistema productivo. (Franciosi, 2020)

En el contexto latinoamericano, los estudios empíricos sobre la implementación de Inteligencia Artificial en procesos de mantenimiento industrial aún son limitados y, en muchos casos, se centran en simulaciones o pruebas piloto, sin un análisis integral de su impacto en indicadores reales de desempeño. Esta situación evidencia un vacío de investigación aplicada, particularmente en lo relacionado con la optimización del proceso de mantenimiento mediante el uso de análisis de datos e Inteligencia Artificial en entornos industriales reales. (Rodríguez, 2022)

En consecuencia, se identifica la necesidad de desarrollar investigaciones que integren de manera sistemática el análisis de datos y herramientas de Inteligencia Artificial al proceso de mantenimiento industrial, evaluando su contribución a la optimización operativa, económica y estratégica de la gestión de activos, lo cual justifica plenamente el desarrollo del presente estudio.

1.3. Síntesis

En el contexto actual de la industria, la gestión del mantenimiento continúa enfrentando importantes limitaciones debido al uso predominante de enfoques tradicionales como el mantenimiento correctivo y preventivo, los cuales no permiten anticipar fallas de manera oportuna. Esta situación genera paradas no planificadas, incremento de costos operativas y una utilización ineficiente de los recursos. A pesar de que las organizaciones disponen de grandes volúmenes de datos operativos provenientes de sensores, análisis de condición, registros históricos, estos no son aprovechados para la predicción de fallas y la mejora de la toma de decisiones.

Asimismo, incluso cuando se aplican modelos de inteligencia artificial para la predicción de fallas, estos suelen limitarse únicamente al análisis predictivo, sin integrarse a procesos de optimización que permitan priorizar intervenciones y asignar recursos de manera eficiente. En este sentido, el problema radica en la falta de un enfoque integral que articule la predicción, la optimización y la evaluación del impacto frente a métodos tradicionales. Por ello, la presente investigación propone el desarrollo

de un modelo estructurado en tres niveles que permita mejorar de manera significativa el proceso de mantenimiento industrial mediante el uso de análisis de datos e inteligencia artificial.

1.4. Formulación del Problema

1.4.1. Problema General:

¿De qué manera el análisis de datos y las herramientas basadas en Inteligencia Artificial pueden optimizar el proceso de mantenimiento industrial en una empresa?

1.4.2. Problemas Específicos:

- ¿Qué variables operativas son determinantes para predecir fallas o degradación de activos industriales?
- ¿Qué modelos de Inteligencia Artificial ofrecen mayor precisión para detectar patrones de falla en los equipos críticos?
- ¿Cómo se puede diseñar una propuesta de optimización del proceso de mantenimiento basada en información generada por Inteligencia Artificial?
- ¿Cuál es el impacto operativo esperado al migrar de un mantenimiento reactivo/preventivo a uno predictivo?

1.5. Objetivos de la Investigación

1.5.1. Objetivo General

Presentar una propuesta para la optimización del proceso de mantenimiento industrial mediante el análisis de datos con la aplicación de herramientas de Inteligencia Artificial.

1.5.2. Objetivo Específicos

- Realizar un diagnóstico situacional del área de mantenimiento de una empresa dedicada al sector de transporte internacional, identificando las prácticas actuales de gestión y toma de decisiones en el mantenimiento de equipos.
- Identificar las variables operativas relevantes a partir de datos históricos y desarrollar un modelo integral que combine técnicas de Machine Learning para predecir el estado futuro de los equipos, con un modelo de optimización que utilice dichas probabilidades para seleccionar las intervenciones que maximicen el beneficio esperado bajos restricciones de recursos.

- Evaluar el impacto del modelo propuesto mediante la comparación con el método tradicional de mantenimiento, analizando mejoras en la eficiencia en la asignación de recursos y en el beneficio obtenido.

1.6. Hipótesis

La aplicación de técnicas de análisis de datos y herramientas de inteligencia artificial permitirá optimizar el proceso de mantenimiento industrial, mediante la mejora en la asignación de recursos y la maximización del beneficio asociado a las intervenciones, en comparación con los métodos tradicionales de mantenimiento.

1.6.1. Variables

- Variable independiente

El análisis de datos con herramientas de Inteligencia Artificial

La variable independiente corresponde al uso e implementación de técnicas de análisis de datos y modelos de inteligencia artificial aplicados al proceso de mantenimiento industrial, orientados a la predicción del estado de los equipos y al apoyo en la toma de decisiones para la gestión del mantenimiento.

Esta variable se sustenta en el aprovechamiento de datos operativos históricos, los cuales son procesados mediante herramientas analíticas para generar información predictiva. En particular, se emplean modelos de Machine Learning que permiten estimar probabilidad asociadas a diferentes condiciones de los equipos, tales como estado normal, seguimiento y crítico.

Estas probabilidades constituyen la base para la construcción de indicadores de criticidad, los cuales permiten priorizar intervenciones y optimizar el uso de recursos disponibles en el proceso de mantenimiento.

Incluye los siguientes componentes:

En primer lugar, los modelos de inteligencia artificial, mediante la aplicación de algoritmos de Machine Learning como Random Forest y XGBoost, utilizados para la predicción del estado futuro de los equipos.

En segundo lugar, el análisis de datos operativos, que comprende procesos de limpieza, transformación, codificación de variables y selección de características relevantes a partir de datos históricos.

En tercer lugar, la generación de información predictiva, expresada en probabilidades asociadas a cada estado del equipo, las cuales son utilizadas para la toma de decisiones en el mantenimiento.

Finalmente, la evaluación del desempeño del modelo, mediante métricas como precisión, recall y F1-score, con énfasis en la capacidad de identificar condiciones críticas.

- Variable dependiente

Optimizar el Proceso de mantenimiento Industrial

La variable dependiente representa el nivel de mejora en la gestión del mantenimiento industrial como resultado de la aplicación de técnicas de análisis de datos e inteligencia artificial, integradas con modelos de optimización para la toma de decisiones.

Esta optimización se entiende como la capacidad de asignar de manera eficiente los recursos disponibles, priorizando las intervenciones que generan un mayor impacto en la reducción del riesgo operativo y en el aprovechamiento del tiempo disponible para mantenimiento.

En este contexto, la variable dependiente se mide a través de indicadores cuantitativos relacionados con la eficiencia en la toma de decisiones y el uso de recursos, derivados de la comparación entre el método tradicional y el modelo propuesto.

Los principales indicadores considerados son los siguientes:

En primer lugar, el beneficio total de las intervenciones, entendido como la reducción del costo esperado de fallas mediante la intervención oportuna de los equipos.

En segundo lugar, el beneficio por unidad de tiempo, el cual permite evaluar la eficiencia en el uso de los recursos disponibles, particularmente el tiempo destinado a mantenimiento.

En tercer lugar, el nivel de utilización de los recursos, expresado como el grado de aprovechamiento del tiempo disponible para intervenciones bajo restricciones operativas.

En cuarto lugar, la eficiencia en la selección de equipos, medida a partir de la capacidad del modelo para priorizar aquellas intervenciones que generan mayor impacto en la reducción del riesgo operativo.

Finalmente, se considera la comparación del desempeño entre el método tradicional de mantenimiento y el modelo propuesto, evidenciando las mejoras en la asignación de recursos y en la toma de decisiones.



1.6.2. Tabla de operacionalización de variables

Tabla 1. Variable independiente y dependiente de la investigación

Variables	Definición Conceptual	Definición Operativa	Indicadores	Unidad	Instrumento
El análisis de datos con herramientas de Inteligencia Artificial	Es el uso de técnicas de análisis de datos y modelos de inteligencia artificial para procesar información operativa, identificar patrones, predecir el estado de los equipos y apoyar la toma de decisiones en la gestión del mantenimiento industrial.	Se operacionaliza mediante la aplicación de modelos de Machine Learning (Random Forest y XGBoost) sobre datos históricos, incluyendo procesos de limpieza, transformación y selección de variables, con el fin de generar probabilidad asociadas a los estados de los equipos (normal, seguimiento y crítico).	• Precisión • Recall • F1-score • Probabilidades por clase	% % % Probabilidad (0-1)	Python (Random Forest, XGBoost) Google Colab Dataset histórico
Optimización del proceso de	Es la mejora en la eficiencia de la gestión del	Se operacionaliza mediante la integración de los resultados del	• Beneficio total • Beneficio por	\$ \$/hora	Python (modelo de optimización)

mantenimiento industrial	mantenimiento mediante la asignación óptima de recursos, priorizando intervenciones que maximicen el beneficio y reduzcan el riesgo operativo.	modelo predictivo con un modelo de optimización basado en programación dinámica, el cual selecciona los equipos a intervenir maximizando el beneficio bajo restricciones de tiempo.	hora	• Número de equipos intervenidos • Utilización del tiempo • Mejora vs Método tradicional	Cantidad % %	Excel Solver/algoritmo mochila
--------------------------	--	---	------	--	----------------------------	---------------------------------------

Nota: Elaboración propia

1.7. Justificación de la Investigación

1.7.1. Justificación Económica:

La adopción de Inteligencia Artificial en mantenimiento permite reducir costos por fallas imprevistas, optimizar inventarios de repuestos, mejorar el uso de mano de obra técnica y minimizar pérdidas por tiempos muertos. Esto genera mayor rentabilidad y uso eficiente del capital.

1.7.2. Justificación Ambiental:

Un mantenimiento optimizado y adecuado reduce el consumo energético, evita fugas o mal funcionamiento de equipos contaminantes y prolonga la vida útil de la maquinaria, contribuyendo a la sostenibilidad industrial.

1.7.3. Justificación Técnica:

El análisis de datos y la inteligencia Artificial permiten detectar patrones ocultos, generar alertas tempranas y realizar diagnósticos más precisos que los métodos tradicionales. Esto impulsa la transformación digital del mantenimiento y mejora la confiabilidad de los activos.

1.7.4. Justificación Académica

La investigación contribuye al cuerpo teórico sobre Mantenimiento Predictivo 4.0, generando evidencia sobre el desempeño de modelos de Inteligencia Artificial en contextos industriales. Además, reduce varios existentes respecto a la integración de datos reales, técnicas estadísticas y algoritmos de aprendizaje automático.

1.8. Metodología

1.8.1. Enfoque de la investigación

La presente investigación adopta un enfoque cuantitativo, ya que se basa en el análisis de datos operativos y en la aplicación de modelos matemáticos y de inteligencia artificial para optimizar el proceso de mantenimiento industrial.

Este enfoque permite desarrollar y evaluar de manera objetiva un modelo predictivo y un modelo de optimización, orientados a mejorar la toma de decisiones en la gestión del mantenimiento. En este sentido, se emplean técnicas de Machine Learning para la

predicción del estado futuro de los equipos, generando probabilidades asociadas a diferentes condiciones operativas, tales como normal, seguimiento y crítico.

Asimismo, el enfoque cuantitativo se utiliza para la construcción de un modelo de optimización que, a partir de la información generada por el modelo predictivo, permite seleccionar el conjunto de equipos a intervenir que maximiza el beneficio esperado bajo restricciones de recursos, particularmente el tiempo disponible para mantenimiento.

La evaluación del modelo propuesto se realiza mediante indicadores cuantificables relacionados con la eficiencia en la asignación de recursos y la reducción del riesgo operativo. En particular, se consideran métricas como el beneficio total obtenido, el beneficio por unidad de tiempo y la comparación del desempeño entre el método tradicional de mantenimiento y el modelo basado en la inteligencia artificial y optimización.

De esta manera el enfoque cuantitativo permite abordar integralmente las tres etapas de la investigación: el desarrollo del modelo predictivo de fallas, el modelo de priorización y optimización y la evaluación comparativa y validación de la propuesta.

1.8.2. Tipo de investigación

La investigación es de tipo aplicada, debido a que busca generar una solución práctica y viable para optimizar el proceso de mantenimiento industrial mediante el uso de análisis de datos e Inteligencia Artificial, contribuyendo directamente a la mejora de la gestión de activos en una organización real.

Asimismo, presenta un alcance propositivo y de validación analítica, debido a que desarrolla un modelo basado en técnicas de Machine Learning y optimización matemática, cuya eficacia se evalúa mediante una validación retrospectiva utilizando datos históricos del proceso de mantenimiento. En este sentido, la investigación busca determinar el desempeño del modelo propuesto y compararlo con el procedimiento tradicional empleado por la organización, sin establecer relaciones causales derivadas de una implementación experimental.

1.8.3. Diseño de la investigación

El diseño de la investigación es no experimental, debido a que no existe manipulación deliberada de las variables de estudio y el análisis se realiza a partir de registros históricos generados por el proceso de mantenimiento de la organización.

Asimismo, corresponde a un estudio retrospectivo basado en datos históricos, en el cual se desarrolla y valida analíticamente un modelo de apoyo a la decisión utilizando información previamente registrada. La evaluación del modelo se realiza mediante la comparación de los resultados obtenidos con el procedimiento tradicional aplicado sobre el mismo conjunto de datos, sin intervenir directamente en la operación real.

1.8.4. Definición y cálculo de indicadores

Para la evaluación del modelo, se establecen indicadores cuantitativos que permiten medir tanto el desempeño del modelo predictivo (variable independiente) como el nivel de optimización del proceso de mantenimiento (variable dependiente). Estos indicadores constituyen la base para el análisis comparativo entre el método tradicional y el modelo propuesto.

a) Indicadores de la variable independiente

El desempeño del modelo predictivo se evalúa mediante métricas de clasificación que permiten medir la capacidad del modelo para identificar correctamente las condiciones de los equipos

- Precisión: La precisión se define como la proporción de predicciones positivas correctas respecto al total de predicciones positivas realizadas por el modelo.

$$Precisión = \frac{TP}{TP + FP}$$

Donde:

- TP: verdaderos positivos
- FP: falsos positivos
- Recall: El recall mide la capacidad del modelo para identificar correctamente los casos positivos reales.

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

Donde:

- FN: falsos negativos
- F1-score: Es una medida que combina precisión y recall proporcionando un equilibrio entre ambas métricas.

$$F1 = 2 \cdot \frac{Precisión \cdot Recall}{Precisión + Recall}$$

Adicionalmente, el modelo generará probabilidades por clase, las cuales representan la estimación de pertenencia de cada equipo a las categorías definidas (normal, seguimiento y crítico). Estas probabilidades constituyen la salida principal del modelo predictivo y sirven como insumo para el proceso de optimización.

b) Indicadores de la variable dependiente

La optimización del mantenimiento se evalúa mediante indicadores que permiten medir la eficiencia en la asignación de recursos y el impacto de las decisiones tomadas.

- Beneficio total: El beneficio total se define como la suma de los beneficios asociados a las intervenciones seleccionadas.

$$Beneficio\ Total = \sum_{i=1}^n B_i \cdot x_i$$

Donde:

- B_i : Beneficio asociado al equipo i.
- x_i : variable binaria (1 si el equipo es intervenido, 0 en caso contrario).
- Beneficio por hora: El beneficio por hora mide la eficiencia en el uso del recurso tiempo.

$$Beneficio\ por\ hora = \frac{Beneficio\ total}{Tiempo\ Total}$$

- Nivel de utilización del tiempo: El nivel de utilización del tiempo permite evaluar el grado de aprovechamiento de los recursos disponibles.

$$Utilización = \frac{Tiempo\ utilizado}{Tiempo\ disponible} \times 100$$

- Mejora respecto al método tradicional: La mejora respecto al método tradicional se define como la variable relativa del beneficio obtenido por el modelo propuesto en comparación con el método convencional.

$$Mejora = \frac{Beneficio_{modelo} - Beneficio_{tradicional}}{Beneficio_{tradicional}} \times 100$$

Finalmente, la eficiencia en la selección de equipos se interpreta como la capacidad del modelo para elegir el conjunto de intervenciones que maximiza el beneficio bajo restricciones, en comparación con la selección realizada mediante criterios tradicionales.

1.8.5. Población

La población del estudio está conformada por el conjunto total de registros históricos de inspecciones y reportes de condición de equipos industriales generados por el área de mantenimiento predictivo de la empresa donde se desarrolló la investigación. Estos registros corresponden a evaluaciones técnicas realizadas mediante diferentes técnicas de monitoreo de condición, tales como análisis de aceite, análisis de vibraciones, inspección por fuga de flujo magnético (MFL), inspección visual e inspección por ultrasonido (UT).

Cada registro contiene información relevante relacionada con el estado operativo de los equipos, incluyendo variables como la técnica de inspección utilizada, el sistema al que pertenece el equipo, la fecha de observación, el equipo inspeccionado, el componente evaluado, la condición diagnosticada (normal, seguimiento o crítico), el diagnóstico técnico emitido por el personal especializado, las recomendaciones de mantenimiento, los comentarios asociados a la inspección, el estatus de atención y el aviso generado para la programación del mantenimiento correspondiente.

1.8.6. Muestra

La muestra utilizada en la presente investigación corresponde a un muestreo no probabilístico de tipo censal, dado que se trabajará con la totalidad de los registros disponibles en la base de datos histórica de mantenimiento predictivo. En este sentido, no se realizará un proceso de selección de una parte representativa de la población, sino que se analizarán todos los datos disponibles con el objetivo de maximizar la cantidad de información utilizada para el desarrollo del modelo de análisis de datos.

De esta manera, la muestra está constituida por los 2123 registros de inspecciones técnicas, los cuales contienen reportes asociados a diferentes sistemas industriales, entre

los que se incluyen sistemas de minerales, sistemas de granos y equipos motorizados, evaluados mediante distintas técnicas de monitoreo de condición.

La utilización de la totalidad de los registros disponibles permite mejorar la calidad del análisis estadístico y del entrenamiento de los modelos de Inteligencia Artificial propuestos en la investigación, ya que se incrementa la cantidad de observaciones utilizadas para identificar patrones en el comportamiento de los equipos y en la evolución de su condición operativa. Asimismo, el uso de una muestra censal contribuye a reducir posibles sesgos asociados a la selección de datos, permitiendo obtener resultados más representativos del comportamiento real del proceso de mantenimiento industrial.

1.8.7. Técnicas de recolección de datos

En la presente investigación se emplearon técnicas de recolección de datos de tipo documental y analítico, debido a que la información utilizada proviene de registros históricos generados por el área de mantenimiento predictivo de la empresa objeto de estudio.

La principal técnica utilizada fue el análisis documental, el cual consistió en la revisión y recopilación de información contenida en los registros técnicos de inspección y monitoreo de condición de equipos industriales. Estos registros incluyen reportes generados a partir de diversas técnicas de diagnóstico, tales como análisis de aceite, análisis de vibraciones, inspección por fuga de flujo magnético (MFL), inspección visual e inspección por ultrasonido (UT). Cada uno de estos reportes contiene información relacionada con el estado operativo de los equipos, el diagnóstico técnico realizado por los especialistas y las recomendaciones de mantenimiento correspondientes.

Asimismo, se utilizó la técnica de extracción y consolidación de datos, mediante la cual se recopiló la información contenida en los registros históricos y se organizó en una base de datos estructurada en formato digital. Este proceso permitió integrar variables relevantes para el análisis, tales como la técnica de inspección utilizada, el sistema al que pertenece el equipo, la fecha de observación, el equipo inspeccionado, el componente evaluado, la condición diagnosticada, el diagnóstico técnico, las recomendaciones de mantenimiento, el estatus de atención y los avisos generados para la intervención correspondiente.

El uso de estas técnicas permitió disponer de una base de datos confiable que refleja el comportamiento histórico del proceso de mantenimiento predictivo dentro de la organización, constituyendo el insumo principal para el análisis de datos y el desarrollo del modelo basado en herramientas de Inteligencia Artificial propuesto en la investigación.

1.8.8. Instrumentos de Investigación

Para la recolección y organización de la información utilizada en la investigación se emplearon instrumentos de carácter técnico y digital que permitieron registrar, estructurar y procesar los datos históricos del proceso de mantenimiento.

El principal instrumento utilizado fue una base de datos digital en formato de hoja de cálculo, en la cual se consolidaron los registros históricos de inspección generados por el área de mantenimiento predictivo. Esta base de datos contiene variables relacionadas con la identificación del equipo, el sistema al que pertenece, la técnica de inspección aplicada, la fecha de observación, el componente evaluado, la condición diagnosticada, el diagnóstico técnico emitido por los especialistas, las recomendaciones de mantenimiento, los comentarios asociados al reporte, el estatus de atención y el aviso correspondiente.

Adicionalmente, se emplearon herramientas de análisis de datos y procesamiento estadístico, las cuales permitieron realizar el tratamiento, limpieza y transformación de la información recopilada. Estas herramientas facilitan la preparación de los datos para su posterior análisis mediante algoritmos de Inteligencia Artificial y técnicas de aprendizaje automático.

Asimismo, se utilizaron herramientas computacionales para el análisis de datos y el desarrollo de modelos de aprendizaje automático, las cuales permitieron ejecutar procesos de preprocesamiento de datos, entrenamiento de modelos de clasificación y evaluación del desempeño de los algoritmos utilizados en la investigación.

El uso de estos instrumentos permitió garantizar una adecuada organización y procesamiento de la información, facilitando la aplicación de métodos analíticos orientados a la optimización del proceso de mantenimiento industrial.

1.8.9. Técnicas de análisis de datos

Para el desarrollo de la investigación se emplearon técnicas de análisis de datos de tipo descriptivo, estadístico y predictivo, con el objetivo de identificar patrones en la información histórica del proceso de mantenimiento y desarrollar un modelo de análisis basado en herramientas de Inteligencia Artificial.

En una primera etapa se realizó un análisis estadístico descriptivo, mediante el cual se examinaron las principales características de la base de datos recopilada. Este análisis permitió identificar la distribución de las condiciones de los equipos (normal, seguimiento y crítico), la frecuencia de utilización de las diferentes técnicas de inspección, la cantidad de reportes asociados a cada sistema industrial y la recurrencia de condiciones críticas en los equipos evaluados.

Posteriormente se llevó a cabo un proceso de preprocesamiento y preparación de los datos, el cual incluyó actividades como la limpieza de registros, la estandarización de variables, la codificación de variables categóricas y la organización de la información en un formato adecuado para el entrenamiento de modelos de aprendizaje automático.

En una tercera etapa se aplicaron técnicas de análisis predictivo mediante algoritmos de aprendizaje automático, con el propósito de desarrollar un modelo de clasificación capaz de identificar patrones en la condición operativa de los equipos industriales. Para ello se utilizaron algoritmos de clasificación comúnmente empleados en el análisis de datos, tales como Random Forest y XGBoost, los cuales permiten analizar relaciones complejas entre variables y mejorar la capacidad de predicción del modelo.

Finalmente, se realizó la evaluación del desempeño de los modelos desarrollados, utilizando métricas de validación ampliamente empleadas en el análisis de modelos de clasificación, tales como exactitud (accuracy), precisión (precision), sensibilidad (recall) y F1-score. Estas métricas permiten medir la capacidad del modelo para identificar correctamente las diferentes condiciones operativas de los equipos y evaluar su utilidad como herramienta de apoyo para la toma de decisiones en el proceso de mantenimiento industrial.

1.9. Alcance

La presente investigación tiene como alcance el desarrollo y la validación analítica de una propuesta para apoyar la optimización del proceso de mantenimiento industrial

mediante el uso de técnicas de Machine Learning y programación dinámica. La validación se realiza de manera retrospectiva utilizando una base de datos histórica de mantenimiento, lo que permite evaluar el desempeño del modelo predictivo y del modelo de optimización bajo ciertas condiciones.

En este contexto, la comparación entre el procedimiento tradicional y la propuesta desarrollada corresponde a una evaluación analítica realizada sobre información histórica, con el propósito de estimar el comportamiento esperado del modelo frente al proceso convencional de priorización de intervenciones. En consecuencia, los resultados obtenidos no constituyen evidencia de una implementación industrial en tiempo real, sino una validación del desempeño potencial de la propuesta bajo el escenario analizado.

Asimismo, la investigación se limita al desarrollo del modelo predictivo, la optimización de la priorización de intervenciones y la comparación de ambos enfoques sobre la base de datos disponible. Aspectos relacionados con la implementación operativa, integración de sistemas empresariales, evaluación económica de largo plazo y medición del impacto organizacional quedan fuera del alcance del presente estudio y se plantean como líneas de trabajo futuro.

1.10. Desafíos y Limitaciones

La presente investigación evidencia la factibilidad técnica de emplear herramientas de inteligencia artificial para apoyar la optimización del proceso de mantenimiento industrial mediante una validación analítica basada en datos históricos. No obstante, los resultados obtenidos deben interpretarse dentro del alcance metodológico del estudio, ya que la propuesta no fue implementada en un entorno industrial real. En consecuencia, existen diversos desafíos y limitaciones que deben ser considerados para una adecuada interpretación de los resultados y para una eventual implementación futura.

En primer lugar, la disponibilidad y calidad de los datos constituyen un factor crítico. El desempeño del modelo predictivo depende directamente de la existencia de registros históricos consistentes y de variables operativas confiables. En este sentido, la presencia de datos incompletos, ruido en las mediciones o inconsistencias en el registro puede afectar la capacidad del modelo para identificar patrones de falla de manera precisa.

Asimismo, se identifica como desafío la resistencia al cambio organizacional. La incorporación de herramientas de inteligencia artificial en procesos de mantenimiento implica no solo la adopción de nuevas tecnologías, sino también un cambio en la cultura organizacional, requiriendo capacitación del personal, adaptación de procesos y aceptación de la toma de decisiones basada en datos.

Por otro lado, la infraestructura tecnológica representa una limitación importante. No todas las organizaciones cuentan con sistemas de adquisición de datos en tiempo real, plataformas de almacenamiento estructurado o arquitecturas basadas en IoT, lo cual restringe la implementación directa del modelo propuesto en contextos industriales reales.

Adicionalmente, existen consideraciones éticas y de seguridad relacionadas con el manejo de datos industriales. El uso de información operativa debe cumplir con las políticas de confidencialidad, protección de datos y ciberseguridad, especialmente en sectores estratégicos como el transporte internacional.

Finalmente, es importante señalar una limitación inherente del propio modelo de Machine Learning a desarrollar. El rendimiento del modelo predictivo se encuentra condicionado por la cantidad y calidad de los datos disponibles, existiendo un límite natural en su capacidad de mejora. A medida que el modelo alcanza un nivel de desempeño estable, la incorporación de nuevos datos que no aporten información significativamente diferente o relevante puede no generar mejoras sustanciales o incluso introducir ruido, afectando la generalización del modelo. En este sentido, la optimización del mismo, no depende únicamente del aumento en el volumen de datos, sino también de la calidad, representatividad y pertinencia de los mismos.



CAPÍTULO II

2. Marco Teórico

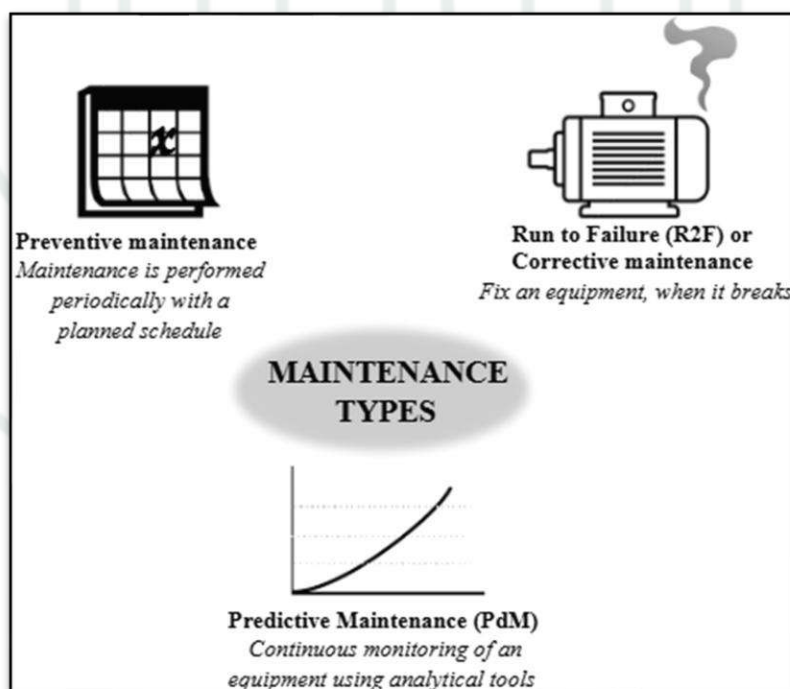
2.1. Antecedentes de la Investigación

2.1.1. Antecedentes Internacionales

En los últimos años, la aplicación de técnicas de análisis de datos y herramientas de inteligencia artificial en el mantenimiento industrial ha cobrado gran relevancia dentro del contexto de la industria 4.0. Diversas investigaciones han demostrado que el uso de algoritmos de aprendizaje automático permite mejorar la predicción de fallas y optimizar la gestión del mantenimiento.

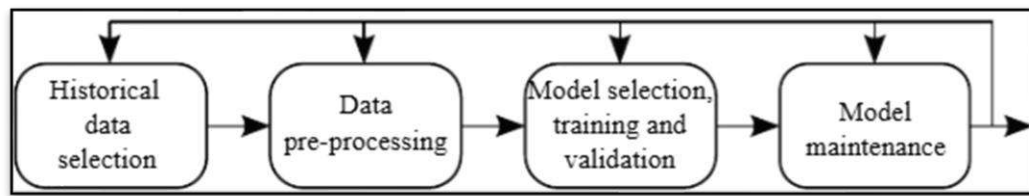
En este sentido, Carvalho, Soares, Francisco & Alcalá (2019), en su estudio titulado “Predictive maintenance: A review”, analizaron diferentes enfoques de mantenimiento predictivo basados en datos, figura 1, concluyendo que los modelos de Machine Learning permiten mejorar la confiabilidad de los equipos mediante la identificación temprana de fallas y la optimización de las intervenciones de mantenimiento. Asimismo, como se muestra en la figura 2, destacan que el uso de datos históricos es fundamental para el desarrollo de modelos robustos.

Figura 1. Descripción general de los Tipos de Mantenimiento



Nota: Adaptado de Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019)

Figura 2. Pasos principales para el Diseño de un Modelo de Aprendizaje Automático

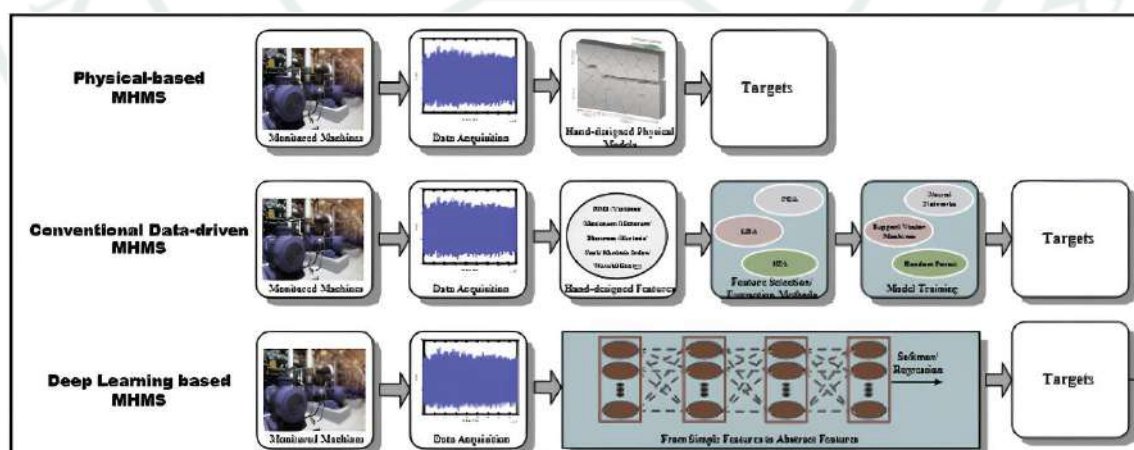


Nota: Adaptado de Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019)

Por otro lado, Zonta et al., (2020) en la investigación “Predictive maintenance in the Industry 4.0: A systematic literatura review”, realizaron una revisión sistemática sobre la aplicación de inteligencia artificial en el mantenimiento industrial, concluyendo que los algoritmos como Random Forest y Support Vector Machine son ampliamente utilizados para la clasificación de condiciones operativas de equipos, permitiendo mejorar la toma de decisiones en entornos industrial.

Asimismo, Zhao, Yan, Chen, Mao, Wang & Gao (2019), en su estudio “Deep learning and it’s applications to machine health monitoring”, demostraron que los modelos basados en aprendizaje profundo, figura 3, pueden identificar patrones complejos en los datos de monitoreo de condición, lo que contribuye a una detección más precisa de fallas en equipos industriales.

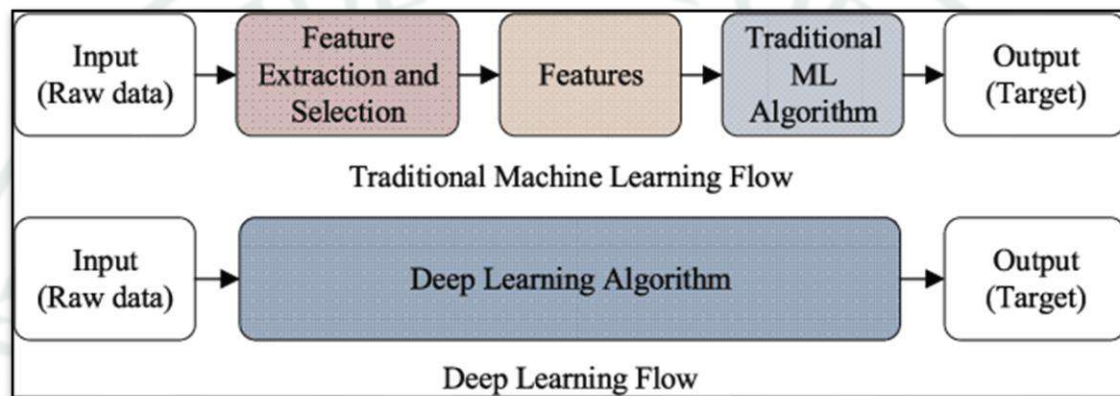
Figura 3. Identificación de patrones mediante diferentes técnicas de monitoreo



Nota: Adaptado de Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019)

Finalmente, Zhang, Yang, & Wang (2019), en su estudio “Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey”, concluyen que la implementación de herramientas de inteligencia artificial en el mantenimiento industrial permite no solo predecir fallas, como se muestra en la figura 4, sino que también permite optimizar la planificación de intervenciones, reducir costos y mejorar la eficiencia operativa.

Figura 4. Flujo de Machine Learning y Deep Learning



Nota: Adaptado de Zhang, W., Yang, D., & Wang, H. (2019)

En conjunto, estos estudios evidencian que el uso del análisis de datos e inteligencia artificial en el mantenimiento industrial constituye una tendencia creciente a nivel internacional, aunque aún existen desafíos relacionados con la calidad de los datos y la integración de estos modelos en entornos reales, lo que justifica el desarrollo de investigaciones aplicadas.

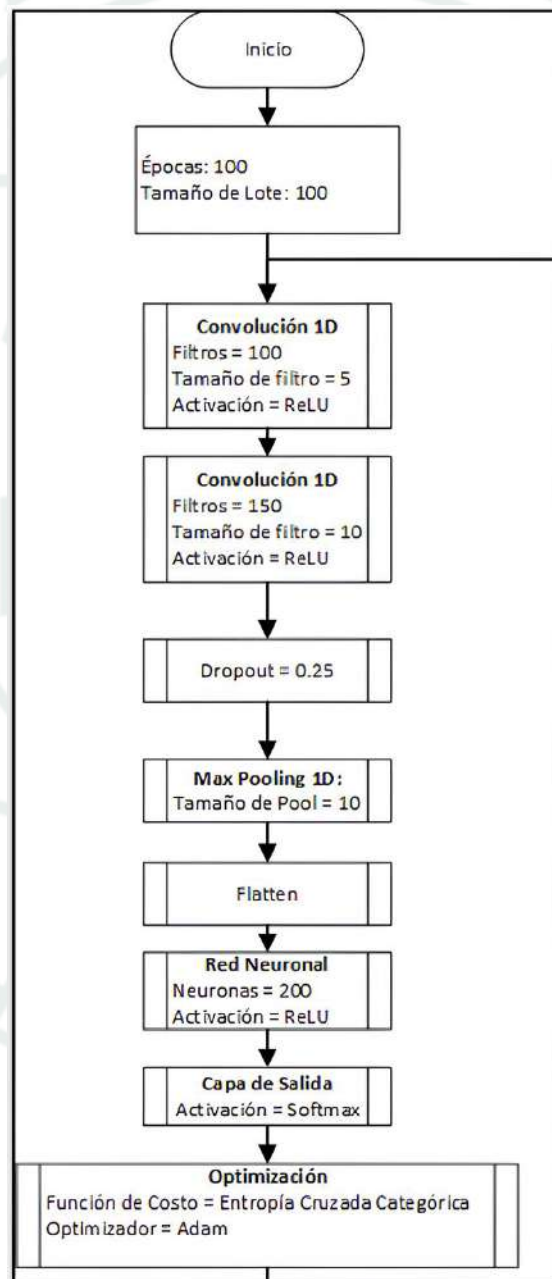
2.1.2. Antecedentes Nacionales

En el contexto nacional, diversas investigaciones han abordado la aplicación de técnicas de inteligencia artificial y machine learning en la mejora de procesos industriales, particularmente en el ámbito del mantenimiento predictivo y la optimización de la toma de decisiones.

En primer lugar, la investigación desarrollada por Inafuku Yoshida (2020) plantea el diseño e implementación de un sistema de diagnóstico de fallas en equipos industriales mediante el uso de algoritmos de inteligencia artificial integrados en un robot móvil. Este estudio se centra en la detección de fallas en motores y bombas a partir del análisis de vibraciones, empleando técnicas avanzadas como la Transformada de Hilbert-Huang

y redes neuronales convolucionales, como se observa en la figura 5. Los resultados evidencian que el uso de inteligencia artificial permite identificar patrones complejos en los datos operativos, mejorando significativamente la precisión en el diagnóstico de fallas y reduciendo riesgos asociados a inspecciones manuales en entornos peligrosos (Ver tabla 2). Este aporte resulta relevante, ya que demuestra la viabilidad del uso de modelos predictivos basados en datos reales para anticipar fallas en sistemas industriales.

Figura 5. Modelo de Parámetros obtenidos



Nota: Adaptado de Inafuku Yoshida, A.H. (2020)

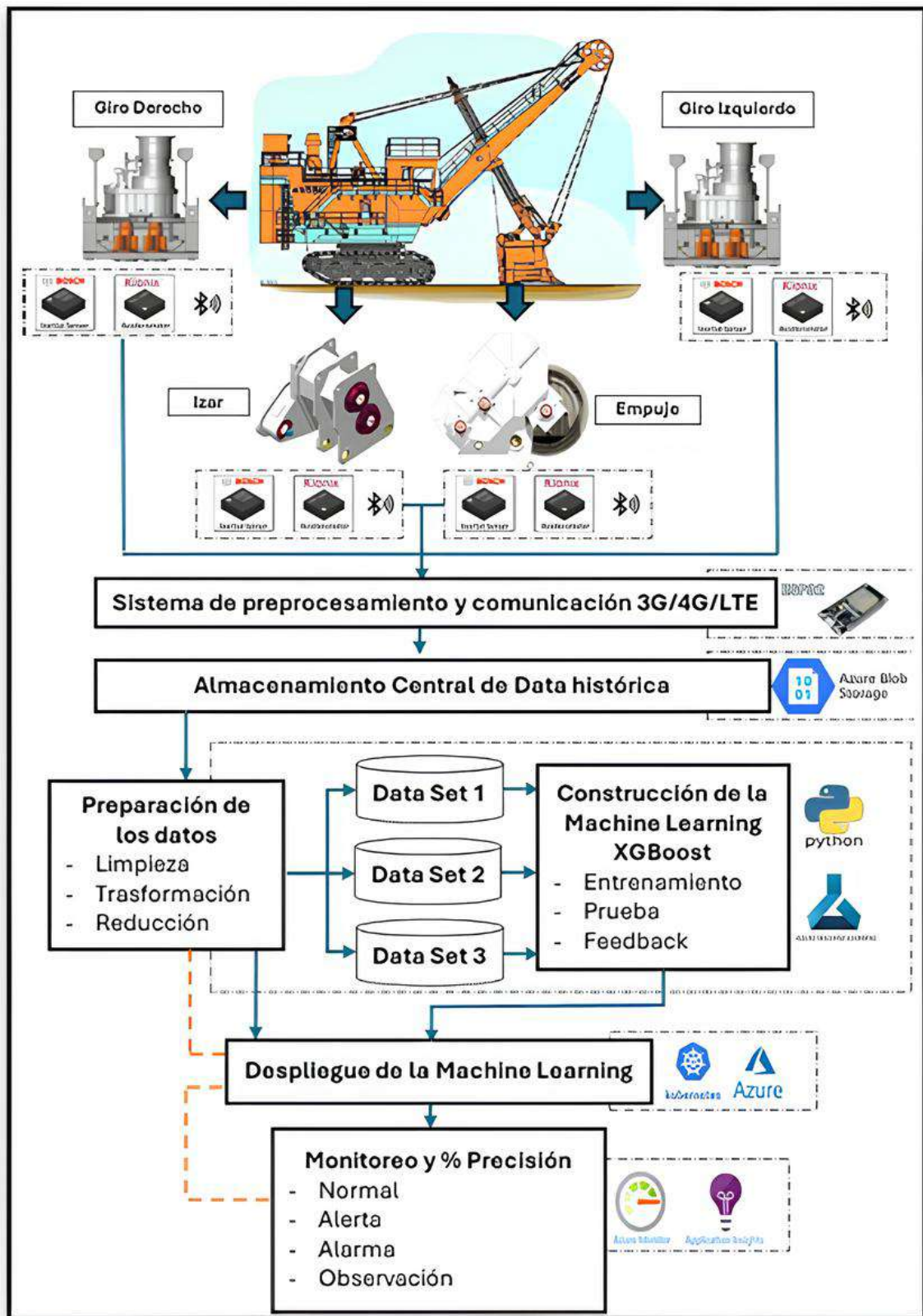
Tabla 2. Resultados de Parámetros Mejorados

Condiciones	Porcentaje de efectividad en la etapa de Entrenamiento (%)	Error	Porcentaje de efectividad en Validación (%)
Iniciales	100	0.0000024778	80
Mejoradas	100	0.00019136	90

Nota: Adaptado de Inafuku Yoshida, A.H. (2020)

Por su parte, Pucho Medina (2024) desarrolló una investigación orientada a la aplicación de modelos de machine learning en el mantenimiento predictivo de transmisiones de una pala eléctrica en el sector minero como se muestra en la figura 6. En este estudio, se evaluaron múltiples algoritmos, destacando el modelo XGBoost por su alta precisión (99.76%) en la predicción de fallas. La implementación de modelo permitió mejorar indicadores clave como la disponibilidad, confiabilidad y mantenibilidad, generando un incremento significativo en la efectividad operativa del sistema. Este trabajo evidencia que el impacto directo del machine learning en la optimización del mantenimiento industrial, alineándose con el enfoque de la presente investigación al integrar modelos predictivos con indicadores operativos.

Figura 6. Arquitectura para la implementación de Machine Learning, XGBoost



Nota: Adaptado de Pucho Medina, J.J. (2024)

Asimismo, Benavente (2024) abordó el mantenimiento predictivo en bombas de agua mediante el uso de algoritmos de machine learning, específicamente regresión lineal y Random Forest. La investigación concluyó procesos de análisis exploratorio de datos, tratamiento de valores atípicos y evaluación de modelos predictivos, concluyendo que estas técnicas permiten anticipar el comportamiento de equipos y mejorar la planificación del mantenimiento. Este estudio aporta una base metodológica importante en cuanto al preprocesamiento de datos y selección de algoritmos, aspectos fundamentales para el desarrollo de modelos predictivos robustos.

Finalmente, Farfán Arias (2025) desarrolló un modelo predictivo basado en machine learning aplicado a la estimación del índice de Rugosidad Internacional (IRI) en pavimentos. Aunque el enfoque se orienta al ámbito de la ingeniería civil, el estudio demuestra la capacidad de los modelos predictivos para analizar variables (tráfico, temperatura, edad del pavimento) y generar estimaciones con alto nivel de precisión ($R^2 > 0.82$) como se muestra en la figura 7. Este trabajo evidencia la versatilidad del machine learning en la predicción de fenómenos complejos y en la optimización de la toma de decisiones, lo cual es extrapolable al mantenimiento industrial.

Figura 7. Comparación de la Precisión de los Modelos de Machine Learning

IRI Evaluado	Modelo	RMSE	MAE	R ²
IRI Creciente (m/Km)	Decision Tree (DT)	0.321023	0.251702	0.820337
IRI Creciente (m/Km)	XGBoost (XGB)	0.321024	0.251703	0.820336
IRI Creciente (m/Km)	Random Forest (RF)	0.321152	0.251861	0.820112
IRI Creciente (m/Km)	K-Nearest Neighbor (KNN)	0.336958	0.262824	0.791884
IRI Creciente (m/Km)	Support Vector Machine (SVM)	0.359568	0.261868	0.749148
IRI Decreciente (m/Km)	Decision Tree (DT)	0.321023	0.251702	0.820337
IRI Decreciente (m/Km)	XGBoost (XGB)	0.321024	0.251703	0.820336
IRI Decreciente (m/Km)	Random Forest (RF)	0.321152	0.251861	0.820112
IRI Decreciente (m/Km)	K-Nearest Neighbor (KNN)	0.336958	0.262824	0.791884
IRI Decreciente (m/Km)	Support Vector Machine (SVM)	0.359568	0.261868	0.749148

Nota: Adaptado de Farfán Arias, P. C. (2025)

En conjunto, los antecedentes nacionales revisados coinciden en señalar que la aplicación de modelos de inteligencia artificial y machine learning permite mejorar significativamente la predicción de fallas, optimizar indicadores operativos y fortalecer la toma de decisiones basada en datos. Sin embargo, se identifica como oportunidad de investigación la integración de estos modelos con técnicas de optimización para la

priorización de intervenciones, aspecto que es abordado en la presente tesis mediante un enfoque multinivel.

2.1.3. Antecedentes locales

En el ámbito local, particularmente en la región sur del Perú y en la Universidad Nacional de San Agustín de Arequipa, se han desarrollado diversas investigaciones orientadas al uso de inteligencia artificial y machine learning en procesos industriales, evidenciando el creciente interés por la digitalización y optimización de operaciones.

En primer lugar, Palma Villanueva (2021) desarrolló una investigación centrada en la aplicación de técnicas de inteligencia artificial y machine learning para la implementación de soft sensors en la predicción del tamaño de partícula P80 en un circuito de molienda SAG en la empresa minera Las Bambas, figura 8. El estudio empleó múltiples algoritmos de aprendizaje supervisado, incluyendo CatBoost y otros modelos avanzados, logrando inferir variables críticas del proceso a partir de datos históricos provenientes de sistemas industriales como PI System. Los resultados demostraron que es posible reemplazar mediciones físicas complejas mediante modelos predictivos con un alto nivel de precisión, mejorando el control del proceso y la toma de decisiones operativas. Este trabajo guarda una relación directa con la presente investigación, ya que valida el uso de inteligencia artificial en la predicción de variables críticas del mantenimiento y operación industrial. Asimismo, en la misma investigación se analiza el uso de metodologías cuantitativas y herramientas tecnológicas para la optimización de procesos en entornos industriales, resaltando la importancia de la modelación de datos y el uso de indicadores de desempeño para la mejora continua. Este estudio refuerza la necesidad de incorporar técnicas de análisis de datos en la gestión industrial, permitiendo identificar patrones, reducir incertidumbre y mejorar la eficiencia operativa.

Figura 8. Resultados Modelo Predictivo

RESULTADOS EVALUACIÓN ESTADÍSTICA MODELOS P80 NIDO N°1						
Modelos	RMSE	MSE	MAE	R2	R2 adj	MAPE
Catboost Opt	9.77	95.52	7.06	0.86	0.86	2.36
Random Forest	11.14	124.17	7.96	0.82	0.82	2.66
XGBoost	11.43	130.66	8.24	0.81	0.81	2.75
Bagging	12.02	144.59	8.61	0.79	0.79	2.88
Catboost	13.24	175.26	10.07	0.75	0.75	3.36
LightGBM	15.54	241.34	12.02	0.65	0.65	4.03
Decision tree	17.20	295.86	11.26	0.57	0.57	3.76
Gradient Boosting	18.45	340.53	14.52	0.51	0.51	4.87
Extra Trees	19.81	392.53	15.61	0.43	0.43	5.23
Regresion Lineal	20.57	423.04	16.32	0.39	0.39	5.46
AdaBoost	21.36	456.14	17.06	0.34	0.34	5.72
SVM	23.37	545.94	18.52	0.21	0.21	6.25
Neural Networks	25.10	630.17	20.06	0.09	0.09	6.51

Nota: Adaptado de Palma Villanueva, R. M. (2021)

Por otro lado, la investigación desarrollada por Marin & Pineda (2019) aborda la mejora de procesos industriales mediante el uso de herramientas analíticas y modelos computacionales aplicados a la gestión operativa. En este estudio se evidencia que la incorporación de modelos de análisis de datos permite optimizar la eficiencia de los procesos, reducir tiempos de respuesta y mejorar la planificación de actividades operativas. Asimismo, se destaca la importancia de la integración de datos provenientes de diferentes fuentes para la toma de decisiones basada en evidencia, lo cual constituye un fundamento clave para el desarrollo de modelos predictivos en mantenimiento industrial.

En conjunto, los antecedentes locales evidencian que la aplicación de inteligencia artificial y técnicas de análisis de datos en la industria, especialmente en el contexto minero e industrial del sur del Perú, ha demostrado ser una herramienta eficaz para la predicción de variables críticas, la optimización de procesos y la mejora en la toma de decisiones. No obstante, se identifica que la mayoría de investigaciones se enfocan en la predicción o monitoreo de variables específicas, existiendo una oportunidad de integrar estos modelos con técnicas de optimización y priorización en la gestión del mantenimiento, enfoque que es desarrollado en la presente tesis mediante un modelo estructurado de tres niveles.

2.2. Marco Teórico

2.2.1. Mantenimiento Industrial

El mantenimiento Industrial es el conjunto de actividades técnicas, administrativas y de gestión destinadas a conservar o restablecer un activo físico a un estado en el cual pueda cumplir su función requerida. Su correcta gestión es fundamental para garantizar la continuidad operativa, la seguridad industrial y la competitividad empresarial.

Según la norma ISO 55000, el mantenimiento forma parte esencial de la gestión de activos físicos, integrando procesos técnicos financieros y organizacionales al ciclo de vida de los equipos (ISO,2014).

Tradicionalmente, el mantenimiento se ha clasificado en:

- Mantenimiento Correctivo
- Mantenimiento Preventivo
- Mantenimiento Predictivo

Sin embargo, los modelos tradicionales presentan limitaciones frente a sistemas industriales complejos y altamente instrumentados.

2.2.2. Gestión del Mantenimiento Industrial

La gestión del mantenimiento industrial constituye un conjunto de actividades técnicas, administrativas y estratégicas orientadas a garantizar la disponibilidad, confiabilidad y seguridad de los activos físicos durante su ciclo de vida. En el contexto actual, el mantenimiento ha dejado de ser una función meramente operativa para convertirse en un factor clave de competitividad, impactando directamente en los costos, la productividad y la calidad del proceso productivo. (Carvalho et al., 2019)

Estudios recientes señalan que una gestión inadecuada del mantenimiento genera incrementos significativos en los costos operativos, especialmente debido a paradas no planificadas y fallas críticas en equipos. Por ello, la ingeniería industrial enfatiza la necesidad de adoptar enfoques sistemáticos que integren planificación, control y mejora continua del mantenimiento, alineados con los objetivos estratégicos de la organización. (Franciosi, 2020)

2.2.3. Tipos de Mantenimiento

2.2.3.1. Mantenimiento Correctivo

El mantenimiento correctivo se realiza después de que ocurre una falla, con el objetivo de restablecer el funcionamiento del equipo. Aunque este enfoque sigue siendo utilizado en muchas industrias, la literatura reciente lo identifica como el más costoso e ineficiente, debido a su carácter reactivo y a la generación de tiempos muertos no programados. (Selcuk, 2017)

2.2.3.2. Mantenimiento Preventivo

El mantenimiento preventivo se basa en intervenciones programadas según intervalos de tiempo o uso del equipo. Si bien este enfoque reduce la probabilidad de fallas, investigaciones recientes indican que puede generar sobre mantenimiento, incrementando costos innecesarios al no considerar el estado real de los activos. (Zonta et al., 2020)

2.2.3.3. Mantenimiento Predictivo

El mantenimiento predictivo se fundamenta en el monitoreo continuo de la condición de los equipos mediante datos obtenidos de sensores y sistemas de información industrial. A partir de estos datos, se emplean técnicas analíticas para predecir fallas antes de que ocurran, optimizando la toma de decisiones y la asignación de recursos. (Zhang et al., 2021)

2.2.3.4. Mantenimiento Prescriptivo

El mantenimiento prescriptivo representa la evolución natural del mantenimiento predictivo dentro del paradigma de la Industria 4.0. Mientras que el mantenimiento predictivo busca anticipar la ocurrencia de fallas mediante el análisis de datos históricos y condiciones operativas, el mantenimiento prescriptivo va un paso más allá al recomendar las acciones óptimas que deben ejecutarse para minimizar riesgos y maximizar el desempeño de los activos.

Este enfoque combina técnicas de análisis predictivo, inteligencia artificial y optimización matemática para determinar no solo qué equipos presentan mayor probabilidad de falla, sino también cuáles deben ser intervenidos primero considerando restricciones de recursos, tiempo y costos. Según Carvalho et al. (2019), los sistemas

prescriptivos permiten transformar información predictiva en recomendaciones accionables para la toma de decisiones.

En el contexto industrial, el mantenimiento prescriptivo resulta especialmente útil cuando existen múltiples equipos que requieren atención y los recursos disponibles son limitados. En estos casos, la organización debe decidir qué intervenciones generan el mayor beneficio operativo posible. Por esta razón, el mantenimiento prescriptivo integra modelos de predicción con técnicas de optimización para priorizar actividades de mantenimiento de manera objetiva y fundamentada.

La presente investigación se enmarca dentro de este enfoque, ya que combina un modelo predictivo basado en Machine Learning para estimar el estado futuro de los equipos con un modelo de optimización basado en programación dinámica para seleccionar las intervenciones más convenientes bajo restricciones operativas. (Bokrants et al. 2020)

2.2.4. Indicadores de Desempeño del Mantenimiento

La evaluación del desempeño del mantenimiento se realiza mediante indicadores clave que permiten medir la efectividad de las estrategias implementadas. Entre los más utilizados se encuentra:

- MTBF (Mean Time Between Failures): mide la confiabilidad de los equipos.
- MTTR (Mean Time To Repair): evalúa la eficiencia en la reparación.
- Disponibilidad operativa: relaciona el tiempo efectivo de operación con el tiempo total.
- Costo de Mantenimiento: refleja el impacto económico del mantenimiento en la organización.

Investigaciones recientes resaltan que la implementación de mantenimiento predictivo basado en análisis de datos mejora significativamente estos indicadores, especialmente el MTBF y la disponibilidad operativa (Bokrantz et al., 2020).

2.2.5. Industria 4.0 y Mantenimiento Inteligente

La industria 4.0 se caracteriza por la integración de tecnologías digitales como el internet industrial de las Cosas (IIoT), Big Data y sistemas ciber físicos. En este contexto, el mantenimiento evoluciona hacia el concepto de mantenimiento inteligente,

donde los datos en tiempo real permiten una gestión proactiva y automatizada de los activos industriales. (Wen, 2020)

Estudios recientes indican que la digitalización del mantenimiento facilita la detección temprana de anomalías, reduce la incertidumbre en la toma de decisiones y mejora la eficiencia global del sistema productivo. (Zonta et al., 2020)

2.2.6. Análisis de datos en el Mantenimiento Industrial

El análisis de datos industriales permite transformar señales crudas en información útil para la toma de decisiones. Incluye procesos de:

- Limpieza de datos
- Normalización
- Extracción de características
- Reducción de dimensionalidad

Las fuentes de datos más comunes son:

- Sensores de vibración
- Temperatura
- Presión
- Corriente eléctrica
- Sistemas SCADA y PLC

El uso de Big Data Analytics ha demostrado mejoras significativas en la confiabilidad y eficiencia del mantenimiento.

2.2.6.1. Ingeniería de Características (Feature Engineering)

La ingeniería de características constituye una de las etapas más importantes dentro del proceso de construcción de modelos de Machine Learning. Consiste en transformar los datos originales en variables más representativas y útiles para el aprendizaje de los algoritmos, permitiendo mejorar la precisión y capacidad predictiva de los modelos.

Asimismo, es considerada uno de los factores de mayor impacto sobre el desempeño de los modelos predictivos, debido a que permite representar de forma más adecuada los patrones presentes en los datos (Kuhn & Jhonson, 2019; Zhang et al. 2021).

En aplicaciones de mantenimiento industrial, los datos suelen provenir de registros históricos de operación, sensores, sistemas de monitoreo y reportes de mantenimiento. Sin embargo, los datos originales no siempre contienen información suficiente para identificar patrones complejos asociados al deterioro de los equipos. Por ello, resulta necesario generar nuevas variables que representen de manera más adecuada el comportamiento de los activos.

Entre las técnicas más utilizadas se encuentran:

- Variables históricas: número de fallas previas, frecuencia de intervenciones y tiempo acumulado de operación.
- Variables temporales: antigüedad del equipo, tiempo desde la última intervención y tendencia de deterioro.
- Variables derivadas: relaciones entre indicadores operativos, costos y tiempo de intervención.
- Variables agregadas: promedios, máximos, mínimos y desviaciones calculadas sobre ventanas temporales específicas.

Diversos estudios han demostrado que la calidad de las características utilizadas tiene una influencia directa sobre el desempeño de los modelos predictivos, siendo en muchos casos más importante que la selección del algoritmo utilizado (Zhang et al., 2021). Por ello, la ingeniería de características constituye una etapa fundamental para el desarrollo de sistemas de mantenimiento predictivo basados en inteligencia artificial.

2.2.7. Inteligencia Artificial aplicada al Mantenimiento Industrial

La Inteligencia Artificial (IA) se define como el conjunto de técnicas computacionales capaces de imitar procesos cognitivos humanos, tales como el aprendizaje, la predicción y la toma de decisiones. En el ámbito del mantenimiento industrial, la Inteligencia Artificial permite analizar grandes volúmenes de datos operativos para identificar patrones de degradación y predecir fallas con alta precisión. (Zhao, 2019)

2.2.8. Machine Learning en el Mantenimiento Predictivo

El Machine Learning se basa en algoritmos que aprenden a partir de datos históricos sin ser explícitamente programados. Investigaciones entre 2019 y 2023 destacan el uso de algoritmos como Random Forest, Support Vector Machines y Gradient Boosting para la

predicción de fallas en equipos industriales, mostrando resultados superiores a los métodos tradicionales. (Zhang, 2021)

2.2.8.1. Aprendizaje Supervisado

El aprendizaje supervisado es una rama del Machine Learning en la cual los algoritmos aprenden a partir de ejemplos previamente etiquetados. Cada observación utilizada durante el entrenamiento contiene un conjunto de variables de entrada y una salida conocida que sirve como referencia para el aprendizaje del modelo.

Dentro del aprendizaje supervisado existen dos grandes categorías:

- Clasificación: utilizada cuando la variable objetivo corresponde a categorías discretas, como estados de funcionamiento o tipos de falla.
- Regresión: utilizada cuando la variable objetivo corresponde a valores continuos, como temperatura, vibración o tiempo restante de vida útil.

En el ámbito del mantenimiento industrial, los problemas de clasificación son ampliamente utilizados para identificar el estado de los equipos y determinar niveles de criticidad que faciliten la toma de decisiones (Goodfellow et al. 2016; Géron, 2022).

2.2.8.2. Clasificación multiclase

La clasificación multiclase corresponde a problemas donde la variable objetivo puede asumir más de dos categorías. A diferencia de la clasificación binaria, que distingue únicamente entre dos estados posibles, la clasificación multiclase permite representar situaciones más complejas y cercanas a los escenarios reales de operación industrial.

En la presente investigación se consideran tres estados operativos:

- Estado Normal: el equipo opera dentro de parámetros aceptables.
- Estado de Seguimiento: el equipo presenta señales de deterioro que requieren monitoreo.
- Estado Crítico: el equipo presenta una alta probabilidad de falla y requiere intervención prioritaria.

La utilización de clasificación multiclase permite generar una representación más detallada del estado de los activos y proporciona información más útil para la toma de decisiones de mantenimiento.

2.2.8.3. Datos Tabulares industriales

Los datos tabulares constituyen uno de los formatos más comunes en aplicaciones industriales. Se caracterizan por organizar la información en filas y columnas, donde cada fila representa una observación y cada columna una variable asociada al proceso analizado.

En mantenimiento industrial, los datos tabulares suelen incluir variables como:

- Horas de operación
- Número de fallas
- Costos de mantenimiento
- Tiempo de intervención
- Edad del equipo
- Indicadores operativos

Diversos estudios señalan que algoritmos como Random Forest y XGBoost presentan un desempeño sobresaliente sobre este tipo de datos debido a su capacidad para modelar relaciones complejas y manejar variables heterogéneas (Shwartz-Ziv & Armon, 2022; Grinsztajn et al., 2022).

2.2.8.4. Modelos Ensemble

Los modelos Ensemble corresponden a técnicas de aprendizaje automático que combinan múltiples modelos individuales con el propósito de mejorar la precisión y robustez de las predicciones. Su principio fundamental consiste en que la combinación de varios modelos suele producir mejores resultados que la utilización de un único modelo.

Los principales enfoques Ensemble son:

- i. Bagging (Bootstrap Aggregating)
Construye múltiples modelos de forma independiente y combina sus resultados para reducir la varianza
- ii. Boosting
Construye modelos de manera secuencial, donde cada nuevo modelo intenta corregir los errores cometidos por los anteriores.

iii. Random Forest

Random forest es un algoritmo basado en Bagging que construye múltiples árboles de decisión utilizando subconjuntos aleatorios de datos y variables. La predicción final se obtiene mediante votación entre los árboles generados.

Sus principales ventajas son:

- Alta robustez frente al ruido.
- Bajo riesgo de sobreajuste.
- Capacidad para manejar relaciones no lineales.
- Facilidad para calcular importancia de variables.

iv. XGBoost

XGBoost (Extreme Gradient Boosting) es un algoritmo de aprendizaje supervisado basado en la técnica de Boosting, cuyo objetivo consiste en construir secuencialmente múltiples árboles de decisión de manera secuencial, donde cada nuevo árbol corrige los errores generados por los modelos anteriores, mejorando progresivamente la capacidad predictiva del modelo. (Chen & Guestrin, 2016)

Este algoritmo incorpora técnicas avanzadas de regularización, optimización del gradiente y procesamiento paralelo, permitiendo obtener modelos con alta capacidad de generalización y reduciendo el riesgo de sobreajuste en comparación con otros métodos tradicionales de aprendizaje automático (Chen & Guestrin, 2016).

Entre sus principales características destacan:

- Alta precisión predictiva en problemas de clasificación y regresión.
- Capacidad para modelar relaciones complejas y no lineales entre las variables analizadas.
- Incorporación de mecanismos de regularización que contribuyen a mejorar la capacidad de generalización del modelo.
- Manejo eficiente de grandes volúmenes de datos y reducción de los tiempos de entrenamiento mediante procesamiento paralelo.
- Capacidad para determinar la importancia relativa de las variables utilizadas durante el proceso de clasificación.

2.2.9. Regresión Logística

La regresión logística es un algoritmo de aprendizaje supervisado utilizado en problemas de clasificación, cuyo objetivo consiste en estimar la probabilidad de que una observación pertenezca a una determinada categoría a partir de un conjunto de variables predictoras (Becerra Correa & Leguizamón Páez, 2024).

A diferencia de los modelos de regresión tradicionales, la regresión logística transforma la combinación lineal de las variables de entrada mediante una función logística o sigmoide, generando valores comprendidos entre 0 y 1 que pueden interpretarse como probabilidades de clasificación. (Becerra Correa & Leguizamón Páez, 2024)

Entre sus principales características destacan:

- Facilidad de interpretación de los resultados obtenidos.
- Capacidad para estimar probabilidades asociadas a cada categoría de salida.
- Bajo costo computacional en comparación con algoritmos más complejos.
- Facilidad de implementación en problemas de clasificación supervisada.

En aplicaciones de mantenimiento predictivo, la regresión logística ha sido utilizada para la identificación de condiciones de falla y clasificación de estados operativos de equipos industriales, obteniendo resultados competitivos en escenarios donde las relaciones entre variables presentan una estructura relativamente simple. (Álvarez Campos, 2024)

2.2.10. Máquinas de Vectores de Soporte (SVM)

Las Máquinas de Vectores de Soporte (SVM) constituyen una técnica de aprendizaje supervisado orientada a problemas de clasificación y regresión, cuyo objetivo es encontrar una frontera de decisión óptima que permita separar diferentes categorías maximizando la distancia entre las observaciones y el hiperplano de clasificación. (Cortes & Vapnik, 1995)

El funcionamiento de SVM se basa en la identificación de los denominados vectores de soporte, los cuales corresponden a las observaciones más cercanas a la frontera de decisión y determinan la capacidad de generalización del modelo (Cortes & Vapnik, 1995).

Una de las principales ventajas de este algoritmo es su capacidad para trabajar con problemas lineales y no lineales mediante el uso de funciones kernel, permitiendo representar relaciones complejas entre variables y mejorar el desempeño de clasificación en conjuntos de datos de alta dimensionalidad. (Maheshwari et al., 2024)

Entre sus principales características destacan:

- Capacidad para modelar relaciones lineales y no lineales.
- Buen desempeño en conjuntos de datos de alta dimensionalidad.
- Alta capacidad de generalización frente a nuevos datos.
- Robustez frente al sobreajuste cuando sus parámetros son correctamente ajustados.

Diversas investigaciones recientes en mantenimiento predictivo reportan el uso exitoso de SVM para la clasificación de estados operativos, detección de anomalías y diagnóstico de fallas en equipos industriales, posicionándolo como uno de los algoritmos más empleados dentro de los sistemas de mantenimiento predictivo basado en Machine Learning. (Tapia et al., 2025)

Asimismo, estudios comparativos realizados en aplicaciones industriales incluyen a SVM dentro de los modelos más utilizados para la predicción de fallas, junto con Random Forest, Regresión Logística y Redes Neuronales, debido a su capacidad para identificar patrones complejos en datos operativos provenientes de sensores y sistemas de monitoreo de condición. (Maheshwari et al., 2024)

2.2.11. Deep Learning y Análisis de Series Temporales

El Deep Learning, particularmente mediante redes neuronales profundas como LSTM y autoencoders, ha demostrado un alto desempeño en el análisis de señales industriales complejas. Estos modelos son especialmente eficaces para detectar comportamientos anómalos en series temporales provenientes de sensores, permitiendo anticipar fallas con mayor precisión. (Wen, 2020)

2.2.12. Integración de la Inteligencia Artificial con la Gestión del Mantenimiento

Diversos autores coinciden en que el éxito del mantenimiento predictivo no depende únicamente de la precisión de los modelos de inteligencia artificial, sino también de su integración con los procesos de gestión del mantenimiento. Esto incluye la conexión

con sistemas CMMS, la capacitación del personal y la alineación los objetivos estratégicos de la organización. (Bokrantz, 2020)

Desde la perspectiva de la Ingeniería Industrial, esta integración permite optimizar la planificación de recursos, mejorar la toma de decisiones y reducir los costos operativos, fortaleciendo la sostenibilidad del sistema productivo (Franciosi, 2020).

2.2.13. Internet Industrial de las Cosas (IIoT)

El Internet Industrial de las Cosas (IIoT) conecta sensores, máquinas y sistemas de análisis para permitir mantenimiento en tiempo real. Su integración con inteligencia artificial es clave para el mantenimiento inteligente.

2.2.14. Optimización del Proceso de Mantenimiento

La optimización busca minimizar costos y maximizar disponibilidad mediante modelos matemáticos, simulación y algoritmos inteligentes.

Indicadores clave:

- Mean Time Between Failures (MTBF)
- Mean Time To Repair (MTTR)
- Overall Equipment Effectiveness (OEE)
- Downtime

2.2.14.1. Optimización Aplicada al Mantenimiento Industrial

La optimización aplicada al mantenimiento industrial tiene como objetivo determinar la mejor asignación posible de recursos para maximizar la disponibilidad de los activos y minimizar los costos asociados a fallas, reparaciones e interrupciones operativas. En entornos industriales modernos, las organizaciones deben gestionar simultáneamente múltiples equipos, recursos limitados y diferentes niveles de criticidad, lo que convierte la planificación del mantenimiento en un problema complejo de toma de decisiones. (Giacotto et al., 2025)

El avance de la Industria 4.0 y la disponibilidad creciente de datos operativos han impulsado el desarrollo de enfoques que combinan modelos predictivos con técnicas de optimización matemática. Estos enfoques permiten transformar la información obtenida a partir del monitoreo de activos en decisiones concretas sobre qué equipos intervenir,

cuándo realizar las intervenciones y cómo asignar eficientemente los recursos disponibles. (Zonta et al., 2020)

Diversos estudios recientes indican que la integración de modelos predictivos con algoritmos de optimización mejora significativamente la eficiencia de los programas de mantenimiento, reduciendo fallas inesperadas, mejorando la utilización de recursos y aumentando la disponibilidad operativa de los equipos (Giacotto et al., 2025).

Dentro de este contexto surge el mantenimiento prescriptivo, enfoque que combina capacidades predictivas y mecanismos de optimización para recomendar acciones específicas de mantenimiento. A diferencia del mantenimiento predictivo, que únicamente estima probabilidades de falla, el mantenimiento prescriptivo incorpora procesos de decisión que permiten determinar la mejor estrategia de intervención bajo restricciones operativas reales. (Bokrantz et al., 2020)

En la presente investigación, la optimización se utiliza como un mecanismo complementario al modelo predictivo, permitiendo seleccionar el conjunto de intervenciones que genera el mayor beneficio esperado considerando las restricciones de tiempo disponibles para mantenimiento. De esta manera, la predicción y la optimización se integran dentro de un mismo proceso de apoyo a la toma de decisiones. (Giacotto et al., 2025)

2.2.15. Programación Dinámica

La programación dinámica es una combinación de optimización matemática y enfoques de programación informática que simplifica un problema al dividirlo en subproblemas más pequeños, es una herramienta eficiente para optimizar y controlar sistemas debido a su capacidad para manejar muchos problemas de optimización con funciones objetivo y restricciones no lineales. (Behzadi, 2022)

2.2.15.1. Problema de la mochila 0-1

El problema de la mochila 0–1 (0–1 Knapsack Problem) es uno de los problemas clásicos de optimización combinatoria y consiste en seleccionar un conjunto de elementos que maximice un beneficio total sin exceder una capacidad limitada de recursos. Cada elemento puede ser seleccionado una única vez o descartado, representándose mediante una variable binaria que toma valores de 0 o 1. (Behzadi et al., 2022)

Matemáticamente, el problema puede formularse de la siguiente manera:

$$\max \sum_{i=1}^n b_i x_i$$

Sujeto a:

$$\sum_{i=1}^n t_i x_i \leq T$$

Donde (b_i) representa el beneficio asociado al elemento (i), (t_i) corresponde al recurso consumido por dicho elemento, (T) representa la capacidad disponible y (x_i) indica si el elemento es seleccionado o no (Behzadi et al., 2022).

El problema de la mochila ha sido ampliamente utilizado en escenarios de asignación eficiente de recursos debido a su capacidad para modelar decisiones donde existen restricciones de capacidad. Aplicaciones recientes incluyen planificación de inversiones, asignación presupuestal, programación de proyectos y selección de actividades de mantenimiento bajo recursos limitados. (Ghadi et al., 2023)

En el ámbito del mantenimiento industrial, cada elemento puede representar un equipo candidato a intervención, mientras que la capacidad disponible puede asociarse al tiempo total destinado para mantenimiento. Bajo esta perspectiva, el beneficio de intervenir un equipo puede relacionarse con la reducción del riesgo de falla, la criticidad del activo o el impacto operativo esperado de la intervención. (Abensur et al., 2023)

La principal ventaja de este modelo es que permite seleccionar objetivamente las intervenciones más convenientes cuando no es posible atender todos los requerimientos de mantenimiento existentes. Esto resulta particularmente útil en organizaciones donde los recursos son limitados y las decisiones deben orientarse a maximizar el beneficio global obtenido por las actividades de mantenimiento. (Ghadi et al., 2023)

En la presente investigación, el problema de la mochila se utiliza para seleccionar el conjunto óptimo de equipos que deben ser intervenidos, considerando simultáneamente las probabilidades generadas por los modelos de Machine Learning y las restricciones de tiempo disponibles para mantenimiento. (Abensur et al., 2023)

2.2.15.2. Programación dinámica aplicada al problema de la mochila

La programación dinámica es una técnica de optimización que permite resolver problemas complejos mediante la descomposición en subproblemas más pequeños e interdependientes. Este enfoque resulta especialmente eficiente cuando un problema presenta subestructuras óptimas y subproblemas repetitivos, permitiendo reutilizar soluciones previamente calculadas para reducir el costo computacional del proceso de optimización. (Behzadi et al., 2022)

En aplicaciones de mantenimiento industrial, la programación dinámica se emplea frecuentemente para resolver problemas relacionados con asignación de recursos, planificación de intervenciones, reemplazo de equipos y programación de actividades de mantenimiento. Su principal ventaja radica en la capacidad de encontrar soluciones óptimas sin necesidad de evaluar exhaustivamente todas las combinaciones posibles. (Hillier & Lieberman, 2021)

Cuando se aplica al problema de la mochila 0–1, la programación dinámica construye progresivamente una matriz de decisiones donde cada celda representa la mejor solución alcanzable para una determinada capacidad disponible y un conjunto específico de elementos considerados. Este procedimiento permite obtener la solución óptima de manera sistemática y eficiente. (Ghadi et al., 2023)

La relación recursiva utilizada para resolver el problema puede expresarse de la siguiente manera:

$$DP(i, w) = \max \begin{cases} DP(i - 1, w) \\ DP(i - 1, w - t_i) + b_i \end{cases}$$

donde $(DP(i, w))$ representa el beneficio máximo alcanzable considerando los primeros (i) elementos y una capacidad disponible (w) , mientras que (b_i) y (t_i) corresponden al beneficio y recurso asociado al elemento evaluado, respectivamente (Behzadi et al., 2022).

En el contexto de la presente investigación, cada elemento representa un equipo candidato a mantenimiento, mientras que la capacidad corresponde al tiempo total disponible para realizar intervenciones. El beneficio asociado a cada equipo se calcula a partir de las probabilidades generadas por los modelos predictivos, permitiendo integrar

la información obtenida mediante Machine Learning dentro del proceso de optimización. (Giacotto et al., 2025)

La combinación de programación dinámica y modelos predictivos constituye actualmente una de las estrategias más utilizadas dentro del mantenimiento prescriptivo, debido a su capacidad para transformar predicciones en decisiones operativas concretas. Esta integración permite mejorar la asignación de recursos, reducir riesgos operativos y aumentar la efectividad de los programas de mantenimiento industrial. (Bokrantz et al., 2020)

2.2.16. Vacíos Teóricos Identificados y Tendencias Emergentes

A pesar de los avances reportados en la literatura reciente, persisten vacíos teóricos y prácticos relacionados con la aplicación de Inteligencia Artificial en contextos industriales reales, especialmente en países en desarrollo. La mayoría de los estudios se enfocan en pruebas experimentales o simulaciones, sin evaluar integralmente el impacto de estas tecnologías en indicadores reales de desempeño del mantenimiento. (Rodríguez, 2022)

Este vacío justifica la necesidad de investigaciones aplicadas que analicen la optimización del proceso de mantenimiento mediante el uso de análisis de datos e Inteligencia Artificial, aportando evidencia empírica desde una perspectiva de Ingeniería Industrial.

Los vacíos identificados son los siguientes:

- Integración completa IA-CMMS-ERP
- Explicabilidad de modelos de IA
- Aplicación en PYMES industriales
- Falta de estudios en contextos latinoamericanos

Tendencias emergentes:

- Explainable AI (XAI)
- Digital Twins
- Edge Computing
- MLOps Industria



CAPÍTULO III

3. Diagnóstico Situacional

3.1. Enfoque del diagnóstico

El diagnóstico situacional de la presente investigación se desarrolló bajo un enfoque cuantitativo, orientado a identificar y analizar las principales problemáticas relacionadas con la gestión del mantenimiento industrial en empresas del sector transporte y logística. El propósito de este diagnóstico fue obtener evidencia objetiva sobre las prácticas actuales de mantenimiento, los mecanismos de priorización de intervenciones, el uso de información histórica y las limitaciones existentes en la toma de decisiones operativas.

Para el desarrollo del diagnóstico se emplearon dos tipos de fuentes de información. En primer lugar, se utilizó información secundaria proveniente de organismos oficiales y literatura especializada, tales como reportes del Instituto Nacional de Estadística e Informática (INEI), estudios sobre mantenimiento industrial y documentos relacionados con transformación digital e industria 4.0. Esta información permitió contextualizar la importancia del sector transporte y almacenamiento dentro de la estructura nacional peruana, así como identificar tendencias problemáticas generales asociadas al mantenimiento industrial.

En segundo lugar, se recopiló información primaria mediante la aplicación de una encuesta estructurada dirigida a los responsables y personal vinculado al área de mantenimiento de empresas del sector transporte y logística. El instrumento tuvo como finalidad identificar aspectos relacionados con el tipo de mantenimiento predominante, la frecuencia de fallas no planificadas, los métodos actuales de priorización de intervenciones, el uso de datos históricos y las principales limitaciones presentes en la gestión del mantenimiento. El aspecto general del mantenimiento en estas empresas se muestra en la figura 14.

La encuesta fue validada mediante juicio de expertos considerando criterios de pertinencia, claridad, coherencia y relevancia. Asimismo, la recolección de información se realizó utilizando un muestreo no probabilístico por conveniencia, debido a la accesibilidad de los participantes y a la naturaleza aplicada a la investigación.

3.2. Contexto del sector y análisis PESTEL

3.2.1. Priorización de factores PESTEL mediante matriz de enfrentamiento

Con el propósito de determinar la influencia relativa de los factores del entorno sobre la problemática de mantenimiento industrial identificada en la presente investigación, se realizó una priorización de los componentes del análisis PESTEL, ejemplo simplificado en figuras 9 y 10, mediante una matriz de enfrentamiento por pares, como se ve en la tabla 3. Esta técnica permite comparar sistemáticamente cada factor respecto a los demás, identificando aquellos que ejercen una mayor incidencia sobre la gestión del mantenimiento, la toma de decisiones operativas y la adopción de tecnologías orientadas a la optimización de recursos.

La comparación se efectuó considerando el contexto del sector transporte y logística, así como los objetivos de la investigación relacionados con la implementación de modelos predictivos y de optimización del mantenimiento. Para cada enfrentamiento se asignó un valor de uno (1) cuando el factor evaluado presentaba una mayor influencia sobre la problemática analizada, y cero (0) cuando la influencia era menor respecto al factor comparado.

Tabla 3. Matriz de enfrentamiento de factores PESTEL

Factor	Político	Económico	Social	Tecnológico	Ecológico	Legal	Total
Político	-	0	1	0	1	0	2
Económico	1	-	1	0	1	1	4
Social	0	0	-	0	1	0	1
Tecnológico	1	1	1	-	1	1	5
Ecológico	0	0	0	0	-	0	0
Legal	1	0	1	0	1	-	3

Nota: Elaboración Propia

Los resultados muestran que los factores tecnológico, económico y legal presenta la mayor influencia sobre la problemática estudiada. El factor tecnológico obtuvo la mayor puntuación debido a que la disponibilidad de herramientas asociadas a la industria 4.0, Inteligencia Artificial, Internet de las Cosas (IoT) y analítica de datos constituye el principal habilitador para la implementación de estrategias de mantenimiento predictivo y prescriptivo. Por su parte, el factor económico refleja el impacto directo que las fallas

no planificadas generan sobre los costos operativos, mientras que el factor legal esta relacionados con las exigencias normativas en materia de seguridad, continuidad operativa y gestión de activos.

Con el fin de evitar la exclusión total de factores que, si bien representan una menor influencia relativa, forman parte del entorno que afecta a las organizaciones, los resultados obtenidos fueron normalizados, como se muestra en la tabla 4, para determinar los pesos finales utilizados en el análisis PESTEL.

Tabla 4. *Priorización de factores PESTEL mediante análisis de impacto*

Factor	Impacto (%)	Acumulado (%)
Tecnológico	30.00	30.00
Económico	25.00	55.00
Legal	18.00	73.00
Político	12.00	85.00
Social	8.00	93.00
Ecológico	7.00	100.00

Nota: Elaboración propia

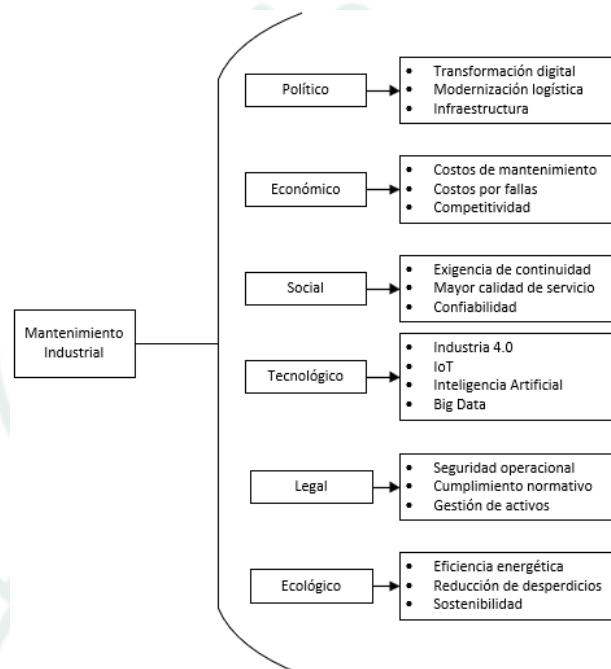
A partir de los pesos obtenidos se elaboró un análisis detallado de cada uno de los factores del entorno, permitiendo identificar las condiciones externas que influyen sobre la gestión del mantenimiento industrial que justifican la necesidad de implementar soluciones basadas en inteligencia artificial y optimización matemática para mejorar la toma de decisiones.

Figura 9. Esquema PESTEL aplicado a la investigación



Nota: Elaboración propia a partir de información del INEI, Gobierno del Perú y literatura especializada

Figura 10. Análisis PESTEL del sector transporte y logística



Nota: Elaboración propia a partir de información del INEI, Gobierno del Perú y literatura especializada

3.2.2. Factor Tecnológico

El factor tecnológico constituye el elemento de mayor influencia dentro del análisis PESTEL realizado para la presente investigación, alcanzando un peso relativo de 30% en la matriz de enfrentamiento. Este resultado refleja la importancia que tienen las tecnologías emergentes en la transformación de los procesos industriales y,

particularmente, en la evolución de los sistemas de mantenimiento hacia enfoques basados en datos y toma de decisiones inteligente.

Durante los últimos años, el Perú ha experimentado un crecimiento sostenido en los procesos de transformación digital impulsado tanto por iniciativas gubernamentales como por la necesidad de las organizaciones de incrementar su competitividad. La Política Nacional de Transformación Digital al 2030 establece como objetivo promover el uso estratégico de tecnologías digitales para mejorar la productividad, competitividad y eficiencia de los sectores económicos (Presidencia del Consejo de ministros [PCM], 2023). Asimismo, diversas organizaciones vienen incorporando herramientas digitales para optimizar procesos, automatizar operaciones y mejorar la gestión de recursos, consolidando la transformación digital como uno de los principales motores de modernización empresarial en país. (Banco Interamericano de Desarrollo [BID], 2022)

Esta transformación tecnológica se encuentra estrechamente relacionada con el concepto de industria 4.0, paradigma que promueve la integración de tecnologías digitales, sistemas inteligentes, Internet de las Cosas (IoT), análisis de datos, computación en la nube e Inteligencia Artificial dentro de los procesos productivos y operativos. De acuerdo a Kagermann et al. (2013), la industria 4.0 representa una nueva etapa de evolución industrial caracterizada por la interconexión de sistemas físicos y digitales. De manera similar, Xu et al. (2018) sostienen que este paradigma permite a las organizaciones capturar, procesar y analizar grandes volúmenes de información para optimizar la toma de decisiones y mejorar la eficiencia operativa.

En el ámbito del mantenimiento industrial, esta evolución tecnológica ha generado una transición progresiva desde modelos reactivos hacia enfoques predictivos y prescriptivos. Históricamente, las organizaciones dependían principalmente de estrategias correctivas, donde la intervención ocurría después de la falla. Posteriormente surgieron los programas de mantenimiento preventivo basados en cronogramas periódicos. Sin embargo, el incremento en la disponibilidad de datos operativos y las capacidades de procesamiento han permitido el desarrollo del mantenimiento predictivo, orientado a anticipar fallas mediante modelos analíticos y algoritmos de aprendizaje automático. (Mobley, 2002; Selcuk, 2017) (ver tabla 5)

Tabla 5. *Evolución de los enfoques de mantenimiento*

Característica	Correctivo	Preventivo	Predictivo	Prescriptivo
Momento de intervención	Después de la falla	Según cronograma	Antes de la falla	Antes de la falla y optimizando recursos
Uso de datos históricos	Bajo	Bajo	Alto	Muy alto
Aplicación de inteligencia artificial	No	No	Sí	Sí
Optimización de recursos	No	Parcial	Parcial	Alta
Nivel de automatización	Bajo	Medio	Alto	Muy alto

Nota: Adaptado de Mobley (2002), Selcuk (2017) y Carvalho et al. (2019).

La incorporación de tecnologías asociadas a la industria 4.0 resulta especialmente relevantes para las empresas del sector transporte y logística, donde la disponibilidad de los activos constituye un factor crítico para la continuidad operativa. Una falla no planificada puede generar retrasos, incumplimientos contractuales, incremento de costos y pérdida de competitividad. En consecuencia, las organizaciones requieren mecanismos capaces de anticipar comportamientos futuros y apoyar la toma de decisiones de mantenimiento. (World Economic Forum, 2023)

Diversos estudios coinciden en que el mantenimiento predictivo basado en Machine Learning representa una de las aplicaciones con mayor potencial dentro de la industria 4.0 debido a su capacidad para identificar patrones complejos de comportamiento, estimar probabilidades de falla y generar alertas tempranas antes de que ocurran eventos críticos. Carvalho et al. (2019) identificaron que los algoritmos de aprendizaje automático permiten mejorar significativamente las capacidades predictivas de los sistemas de mantenimiento. Asimismo, Zonta et al. (2020) concluyen que las técnicas de Machine Learning constituyen una de las principales herramientas para la evolución hacia modelos de mantenimiento inteligente.

El avance de tecnologías complementarias como Big Data, computación en la nube, sensores inteligentes e Internet de las Cosas permite que los sistemas industriales generen cantidades crecientes de información susceptible de ser analizada mediante algoritmos de Inteligencia Artificial. Según Lee et al. (2015), la integración de sistemas ciber-físicos y análisis avanzado de datos constituye uno de los pilares fundamentales de la Industria 4.0. De igual forma, Tao et al. (2018) destacan que la manufactura inteligente depende de la capacidad de transformar grandes volúmenes de datos en conocimiento útil para la toma de decisiones.

En este contexto, la presente investigación se alinea con las tendencias tecnológicas actuales del sector industrial al proponer un modelo basado en Machine Learning para la predicción del estado futuro de los equipos, complementado con un mecanismo de optimización orientado a la priorización de intervenciones de mantenimiento. La propuesta responde a las oportunidades identificadas dentro del entorno tecnológico y busca contribuir a la transición desde esquemas tradicionales de mantenimiento hacia estrategias predictivas y prescriptivas sustentadas en datos. (Carvalho et al., 2019; Zonta et al., 2020)

3.2.3. Factor Económico

El factor económico obtuvo una ponderación de 25% dentro del análisis PESTEL, posicionándose como el segundo elemento de mayor influencia sobre la problemática estudiada. Este resultado evidencia que las decisiones relacionadas con la gestión del mantenimiento industrial se encuentran estrechamente vinculadas al desempeño económico de las organizaciones, particularmente en sectores donde la disponibilidad de los activos constituye un factor determinante para la continuidad operativa.

En el contexto peruano, el sector transporte, almacenamiento y mensajería representa una actividad económica estratégico debido a su participación en las cadenas de suministro, el comercio y la movilidad de bienes y servicios. Según el Instituto Nacional de Estadística e Informática (INEI, 2024), este sector registró un crecimiento de 7.9% en el último año, impulsado por la recuperación económica y el incremento de la demanda de servicios logísticos. Este comportamiento evidencia la necesidad de garantizar elevados niveles de disponibilidad operativa en los activos utilizados para el transporte y la gestión logística.

Asimismo, la estructura empresarial peruana se caracteriza por una elevada presencia de micro, pequeñas y medianas empresas, las cuales enfrentan importantes restricciones presupuestarias para la gestión de activos y mantenimiento. De acuerdo con la Estructura Empresarial del INEI (2024), más del 99% de las empresas registradas en el país corresponden a micro, pequeñas y medianas empresas, ver tabla 6, situación que limita la capacidad de realizar inversiones significativas en infraestructura tecnológica y estrategias avanzadas de mantenimiento.

Tabla 6. Estructura empresarial peruana según tamaño de empresa

Tamaño de empresa	Participación (%)
Microempresa	96.4%
Pequeña empresa	3.0%
Grande y Mediana empresa	0.6%

Nota: Adaptado de INEI (2024)

La disponibilidad de recursos económicos influye directamente sobre las estrategias de mantenimiento implementadas por las organizaciones. Tradicionalmente, muchas empresas han priorizado enfoques correctivos debido a sus menores requerimientos iniciales de inversión. Sin embargo, diversos estudios demuestran que las fallas no planificadas generan costos significativamente superiores a los asociados con estrategias preventivas o predictivas, debido a la interrupción de operaciones, pérdida de productividad, consumo adicional de recursos y retrasos en la prestación de servicios. (Mobley, 2002; Selcuk, 2017)

Desde una perspectiva económica, el mantenimiento debe entenderse como una inversión orientada a preservar la disponibilidad, confiabilidad y vida útil de los activos. En ese sentido, las organizaciones buscan maximizar el retorno obtenido de cada intervención realizada, priorizando aquellas acciones capaces de generar el mayor beneficio operativo con la menor utilización posible de recursos. Esta necesidad resulta particularmente relevante en entornos donde el tiempo disponible para ejecutar actividades de mantenimiento constituye una restricción crítica.

La creciente competitividad empresarial también ha incrementado la presión por optimizar la utilización de recursos. Según el Ministerio de la Producción (PRODUCE, 2025), la incorporación de tecnologías digitales constituye uno de los principales mecanismos para mejorar la productividad y competitividad de las organizaciones

peruanas. Dentro de este contexto, las herramientas basadas en analítica de datos e Inteligencia Artificial permiten reducir la incertidumbre asociada a la toma de decisiones y mejorar la asignación de recursos de mantenimiento.

Diversos autores señalan que las estrategias de mantenimiento predictivo pueden generar reducciones significativas en los costos operativos al disminuir la ocurrencia de fallas inesperadas y optimizar la planificación de intervenciones. Carvalho et al. (2019) identificaron que los modelos de Machine Learning permiten anticipar condiciones de falla con suficiente anticipación para programar acciones correctivas oportunas. Por su parte, Zonta et al. (2020) destacan que la integración de técnicas predictivas y métodos de optimización contribuye a mejorar la eficiencia económica de los programas de mantenimiento industrial.

En el caso específico de la presente investigación, la dimensión económica se encuentra representada tanto en el proceso de predicción de fallas como en el modelo de optimización desarrollado. Mientras el componente predictivo busca reducir los costos asociados a interrupciones no planificadas, el componente de optimización permite seleccionar el conjunto de intervenciones que maximiza el beneficio esperado bajo restricciones de tiempo y recursos disponibles, como se ve en la tabla 7. De esta manera, la propuesta responde directamente a una de las principales necesidades identificadas dentro del entorno económico: mejorar la eficiencia en la asignación de recursos para mantenimiento sin incrementar significativamente los costos operativos.

Tabla 7. *Impacto económico de las estrategias de mantenimiento*

Aspecto	Correctivo	Preventivo	Predictivo
Costos por fallas inesperadas	Alto	Medio	Bajo
Tiempo de inactividad	Alto	Medio	Bajo
Uso eficiente de recursos	Bajo	Medio	Alto
Planificación de intervenciones	Baja	Media	Alta
Impacto sobre la productividad	Negativo	Moderado	Positivo

Nota: Adaptado de Mobley (2002), Selcuk (2017), Carvalho et al. (2019) y Zonta et al. (2020).

Finalmente, el análisis económico evidencia que las organizaciones enfrentan una necesidad creciente de maximizar la rentabilidad de sus operaciones mediante una gestión más eficiente de sus activos. En este escenario, las tecnologías predictivas y los

modelos de optimización constituyen herramientas estratégicas capaces de reducir costos, mejorar la productividad y fortalecer la competitividad empresarial, justificando la pertinencia de la solución propuesta en la presente investigación.

3.2.4. Factor Legal

El factor legal obtuvo una ponderación de 18% del análisis PESTEL, posicionándose como el tercer elemento de mayor influencia sobre la problemática estudiada. Este resultado evidencia que las actividades de mantenimiento industrial no solo responden a criterios técnicos y económicos, sino también al cumplimiento de disposiciones normativas orientadas a garantizar la seguridad, continuidad operativa y adecuada gestión de los activos organizacionales.

En el Perú, las organizaciones que operan activos industriales se encuentran sujetas a diversas regulaciones relacionadas con la seguridad ocupacional, gestión de riesgos y protección de la integridad de los trabajadores. La ley N° 29783, Ley de Seguridad y Salud en el Trabajo, establece la obligación de identificar peligros, evaluar riesgos e implementar medidas preventivas destinadas a reducir la ocurrencia de accidentes y enfermedades ocupacionales (Congreso de la República del Perú, 2011). Dentro de este contexto, el mantenimiento adecuado de equipos e infraestructura constituye una de las principales estrategias para minimizar riesgos operacionales.

La relación entre mantenimiento y seguridad resulta particularmente relevante debido a que una proporción importante de los incidentes industriales se encuentra asociada al deterioro de componentes, fallas mecánicas, errores operativos o deficiencias en los programas de inspección y mantenimiento. Por esta razón, las organizaciones deben garantizar que sus activos operen dentro de condiciones seguras y controladas, reduciendo la probabilidad de eventos que puedan afectar a los trabajadores, las instalaciones o la continuidad de las operaciones.

Asimismo, las exigencias regulatorias actuales han impulsado la adopción de enfoques más sistemáticos para la gestión de activos. La norma ISO 55000 establece principios orientados a maximizar el valor obtenido de los activos mediante una adecuada planificación, operación, mantenimiento y mejora continua, (ver tabla 8). Según esta norma, las organizaciones deben tomar decisiones fundamentadas en información confiable que permita optimizar el desempeño y minimizar los riesgos asociados al ciclo de vida de los activos. (ISO, 2014)

Tabla 8. *Relación entre mantenimiento y cumplimiento normativo*

Aspecto regulatorio	Relación con el mantenimiento
Seguridad y Salud en el Trabajo	Reducción de riesgos asociados a fallas de equipos.
Gestión de riesgos operacionales	Prevención de incidentes y paradas no planificadas.
Continuidad operativa	Garantiza la disponibilidad de activos críticos.
Gestión de activos	Optimiza el ciclo de vida de los equipos.
Cumplimiento normativo	Reduce sanciones y responsabilidades legales.

Nota: Elaboración propia basada en la Ley N° 29783 e ISO 55000

En sectores relaciones con transporte y logística, las exigencias legales adquieren una relevancia adicional debido a la necesidad de garantizar operaciones seguras y confiables. La indisponibilidad de equipos críticos puede generar incumplimientos contractuales, afectaciones al servicio e incluso responsabilidades administrativas derivadas de incidentes operacionales. En consecuencia, las organizaciones requieren mecanismos que permitan identificar oportunamente condiciones de riesgos y priorizar intervenciones de mantenimiento antes de que ocurran fallas significativas.

Por otro lado, la creciente digitalización de los procesos industriales también ha impulsado el desarrollo de marcos regulatorios relacionados con la gestión de información y la transformación digital. La Política Nacional de Transformación Digital al 2030 reconoce la importancia de utilizar tecnologías avanzadas para mejorar la productividad y eficiencia organizacional. (PCM, 2023). En este escenario, las herramientas basadas en Inteligencia Artificial constituyen una oportunidad para fortalecer la toma de decisiones y mejorar el cumplimiento de los objetivos operacionales y regulatorios.

Diversos estudios señalan que los sistemas de mantenimiento predictivo contribuyen significativamente a la reducción de riesgos operacionales al permitir la detección temprana de patrones asociados al deterioro de los equipos. Carvalho et al. (2019) identifican que los modelos de Machine Learning pueden utilizarse para anticipar condiciones de falla antes de que se produzcan eventos críticos. De manera similar, Zonta et al. (2020) concluyen que las técnicas predictivas permiten fortalecer la confiabilidad de los activos y mejorar la planificación de intervenciones, reduciendo la exposición a riesgos operativos.

En el marco de la presente investigación, el componente legal se encuentra directamente relacionado con la necesidad de mejorar la gestión de riesgos asociados a la operación de activos industriales. La utilización de modelos predictivos y mecanismos de optimización permite identificar equipos con mayor probabilidad de transición a estados críticos, facilitando la priorización de intervenciones y contribuyendo al cumplimiento de los principios de prevención establecidos por la normativa vigente, como se ve en la tabla 9.

Tabla 9. *Contribución del mantenimiento predictivo al cumplimiento normativo*

Requisito	Aporte del modelo propuesto
Prevención de riesgos	Identificación temprana de equipos críticos.
Seguridad operacional	Reducción de fallas inesperadas.
Continuidad operativa	Priorización eficiente de intervenciones.
Gestión de activos	Soporte a la toma de decisiones basada en datos.
Cumplimiento regulatorio	Disminución de riesgos asociados a incumplimientos.

Nota: Elaboración propia

Finalmente, el análisis legal evidencia que las organizaciones enfrentan una creciente necesidad de garantizar operaciones seguras, confiables y alineadas con las disposiciones regulatorias vigentes. En este contexto, las tecnologías de Inteligencia Artificial y los modelos de optimización representan herramientas capaces de fortalecer los programas de mantenimiento, reducir riesgos operacionales y contribuir al cumplimiento de los objetivos de seguridad, continuidad y gestión eficiente de activos.

3.2.5. Factor Político

El factor político obtuvo una ponderación de 12% dentro del análisis PESTEL realizado para la presente investigación. Aunque su influencia es menor en comparación con los factores tecnológico, económico y legal, continúa desempeñando un papel relevante debido a que las políticas públicas establecen el marco estratégico que orienta la modernización de los sectores productivos y promueve la adopción de tecnologías digitales para incrementar la competitividad empresarial.

Durante los últimos años, el estado peruano ha impulsado diversas iniciativas orientadas a fortalecer la transformación digital y la innovación tecnológica en las organizaciones, como se ve en la tabla 10 y en la figura 11. Uno de los principales instrumentos en este

ámbito es la Política Nacional de Transformación Digital al 2030, la cual tiene como objetivo impulsar la transformación digital en todos los sectores de la sociedad mediante el aprovechamiento de tecnologías emergentes, el fortalecimiento de capacidades digitales y la generación de valor a partir del uso de datos. (Presidencia del Consejo de ministros [PCM], 2023). Esta política reconoce que la digitalización constituye un factor fundamental para mejorar la productividad y competitividad del país.

De manera complementaria, el Plan Nacional de Competitividad y Productividad establece lineamientos orientados a promover la innovación, la transformación tecnológica y la modernización de los procesos productivos. Dentro de este marco, el uso de herramientas digitales, sistemas inteligentes y tecnologías de análisis de datos es considerado un elemento clave para incrementar la eficiencia operativa de las organizaciones y fortalecer su capacidad de adaptación frente a entornos cada vez más dinámicos. (Consejo Nacional de Competitividad y Formalización, 2024)

Tabla 10. Principales políticas nacionales relacionadas con la transformación digital y competitividad

Política o estrategia	Objetivo principal	Relación con la investigación
Política Nacional de Transformación Digital al 2030	Impulsar la transformación digital del país	Favorece la adopción de tecnologías de Inteligencia Artificial
Plan Nacional de Competitividad y Productividad	Incrementar la productividad empresarial	Promueve la modernización de procesos industriales
Agenda Digital Peruana	Fortalecer capacidades digitales y uso de tecnologías	Impulsa la innovación tecnológica
Estrategias de Gobierno Digital	Promover el uso de tecnologías de información	Fomenta la transformación organizacional

Nota: Elaboración propia basada en PCM (2023) y Consejo Nacional de Competitividad y Formalización (2024).

Asimismo, las políticas de desarrollo económico y modernización logística implementadas por el Estado buscan fortalecer la competitividad de sectores estratégicos como transporte, almacenamiento y distribución. El desarrollo de

infraestructura, la digitalización de procesos y la incorporación de tecnologías avanzadas forman parte de las estrategias nacionales orientadas a mejorar la eficiencia de las cadenas logísticas y facilitar el crecimiento económico sostenible. (Ministerio de Transportes y Comunicaciones [MTC], 2024)

En este contexto, la transformación digital promovida por las políticas públicas genera un entorno favorable para la implementación de sistemas de mantenimiento inteligentes. La disponibilidad de marcos normativos y estrategias nacionales orientadas a la innovación tecnológica facilita la incorporación de soluciones basadas en Inteligencia Artificial, análisis de datos y automatización dentro de los procesos industriales.

Por otro lado, el crecimiento de las iniciativas gubernamentales relacionadas con Gobierno Digital y economía digital ha contribuido a incrementar la disponibilidad de infraestructura tecnológica, conectividad y capacidades digitales dentro de las organizaciones. Según la PCM (2023), el aprovechamiento estratégico de los datos constituye uno de los pilares fundamentales para impulsar la innovación y la generación de valor en los distintos sectores económicos, ver tabla 11.

Tabla 11. *Impacto de las políticas públicas sobre la gestión del mantenimiento industrial*

Política pública	Impacto esperado
Transformación Digital	Facilita la adopción de tecnologías predictivas
Competitividad y Productividad	Incrementa la eficiencia operativa
Gobierno Digital	Favorece la gestión basada en datos
Modernización logística	Mejora la continuidad operativa
Innovación tecnológica	Impulsa el uso de Inteligencia Artificial

Nota: Elaboración propia

Desde la perspectiva del mantenimiento industrial, estas políticas contribuyen a crear condiciones favorables para la transición desde enfoques tradicionales hacia modelos predictivos y prescriptivos. La promoción de tecnologías emergentes permite a las organizaciones acceder a herramientas capaces de mejorar la capacidad de anticipar fallas, optimizar recursos y fortalecer la toma de decisiones basada en evidencia.

En el marco de la presente investigación, el factor político se manifiesta a través del respaldo institucional que existe hacia la transformación digital y la innovación tecnológica. La propuesta desarrollada se encuentra alineada con los objetivos

nacionales de modernización productiva, al emplear técnicas de Machine Learning y optimización matemática para mejorar la gestión del mantenimiento industrial y contribuir a una utilización más eficiente de recursos disponibles.

Finalmente, el análisis político evidencia que las estrategias gubernamentales orientadas a la digitalización, competitividad e innovación tecnológica constituyen un elemento facilitador para la implementación de soluciones inteligentes en el ámbito industrial. Si bien estas políticas no determinan directamente la ocurrencia de fallas o la disponibilidad de activos, sí generan un entorno propicio para la adopción de herramientas avanzadas que permitan mejorar la eficiencia operativa y fortalecer la competitividad empresarial.

Figura 11. Línea de tiempo de la Evolución de las Políticas Públicas a la Transformación Digital en el Perú



Nota: Elaboración propia basada en el Consejo Nacional de Competitividad y Formalización. (2024), la Agenda Digital Peruana y la Política Nacional de Transformación Digital al 2030

3.2.6. Factor Social

El factor social obtuvo una ponderación de 8% dentro del análisis PESTEL realizado para la presente investigación. Aunque su influencia es menor respecto a los factores tecnológico, económico, legal y político, continúa desempeñando un papel importante debido a que las actividades de mantenimiento industrial impactan directamente sobre la

seguridad de los trabajadores, la calidad de los servicios prestados y la continuidad de las operaciones empresariales.

En el contexto actual, las organizaciones enfrentan una creciente demanda por parte de clientes, usuarios y grupos de interés para garantizar niveles elevados de confiabilidad y disponibilidad en sus operaciones. En sectores relacionados con transporte, almacenamiento y logística, la interrupción de servicios debido a fallas de equipos puede afectar la satisfacción de los usuarios, generar retrasos operativos e impactar negativamente sobre la imagen institucional de las empresas. Como consecuencia, la gestión eficiente del mantenimiento se ha convertido en un factor clave para asegurar la continuidad operativa y responder a las expectativas de los diferentes actores involucrados.

Asimismo, la seguridad de los trabajadores constituye uno de los principales aspectos sociales vinculados al mantenimiento industrial. La organización Internacional del Trabajo (ILO, 2023) señala que las condiciones seguras de trabajo dependen, entre otros factores, del adecuado funcionamiento de equipos, herramientas e infraestructura. En este sentido, las actividades de mantenimiento contribuyen a reducir riesgos operacionales, prevenir accidentes y proteger la integridad física de los trabajadores, como se ve en la tabla 12.

Tabla 12. *Relación entre aspectos sociales y gestión del mantenimiento*

Aspecto social	Relación con el mantenimiento
Seguridad de los trabajadores	Reduce riesgos asociados a fallas de equipos
Calidad del servicio	Mejora la disponibilidad operativa
Satisfacción del cliente	Disminuye interrupciones y retrasos
Continuidad operativa	Favorece la confiabilidad de los activos
Desarrollo de competencias	Impulsa la capacidad tecnológica

Nota: Elaboración propia

Otro aspecto relevante corresponde a la transformación de las competencias laborales requeridas dentro de los entornos industriales. La incorporación de tecnologías asociadas a Industria 4.0, Inteligencia Artificial y análisis de datos ha generado una creciente demanda de profesionales capaces de interpretar información, gestionar sistemas digitales y utilizar herramientas analíticas para la toma de decisiones. Según el Foro Económico Mundial (World Economic Forum, 2023), las habilidades relacionadas

con análisis de datos, automatización y tecnologías digitales se encuentran entre las competencias con mayor crecimiento proyectado para los próximos años.

En consecuencia, las organizaciones enfrentan el desafío de fortalecer las capacidades de su personal para aprovechar adecuadamente las nuevas tecnologías disponibles. La adopción de sistemas predictivos no depende únicamente de la infraestructura tecnológica, sino también de la capacidad de los trabajadores para interpretar resultados, comprender indicadores y utilizar información generada como apoyo a la toma de decisiones, como se ve en la tabla 13.

Por otra parte, la incorporación de herramientas de Inteligencia Artificial suele generar procesos de cambio organizacional que requieren aceptación por parte de los usuarios finales. Diversos estudios indican que la resistencia al cambio constituye una de las principales barreras para la implementación exitosa de tecnologías avanzadas dentro de las organizaciones. (Davenport & Ronanki, 2018). Por ello, resulta necesario desarrollar estrategias que faciliten la integración entre las capacidades humanas y las herramientas tecnológicas.

Tabla 13. *Impacto de la transformación digital sobre los recursos humanos*

Aspecto	Impacto esperado
Capacitación tecnológica	Incremento de competencias digitales
Toma de decisiones	Mayor soporte analítico
Productividad laboral	Mejora en la eficiencia operativa
Adaptación organizacional	Necesidad de gestión del cambio
Uso de Inteligencia Artificial	Complementa la experiencia técnica

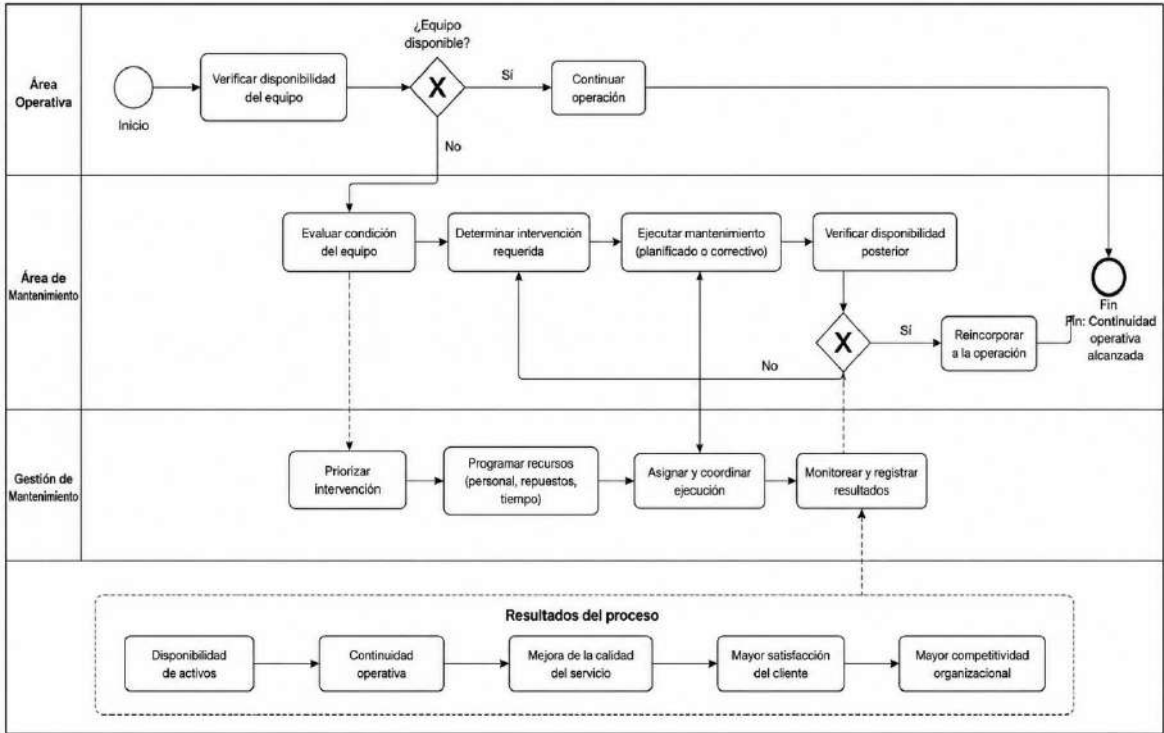
Nota: Elaboración propia

Desde la perspectiva de la presente investigación, el componente social se encuentra relacionado principalmente con la capacidad de los modelos inteligentes para apoyar la toma de decisiones del personal responsable del mantenimiento. La propuesta no busca reemplazar la experiencia de los especialistas, sino proporcionar información adicional que permita identificar riesgos potenciales, priorizar intervenciones y optimizar la asignación de recursos disponibles.

Finalmente, el análisis social evidencia que la gestión eficiente del mantenimiento contribuye tanto a la seguridad de los trabajadores como a la calidad de los servicios

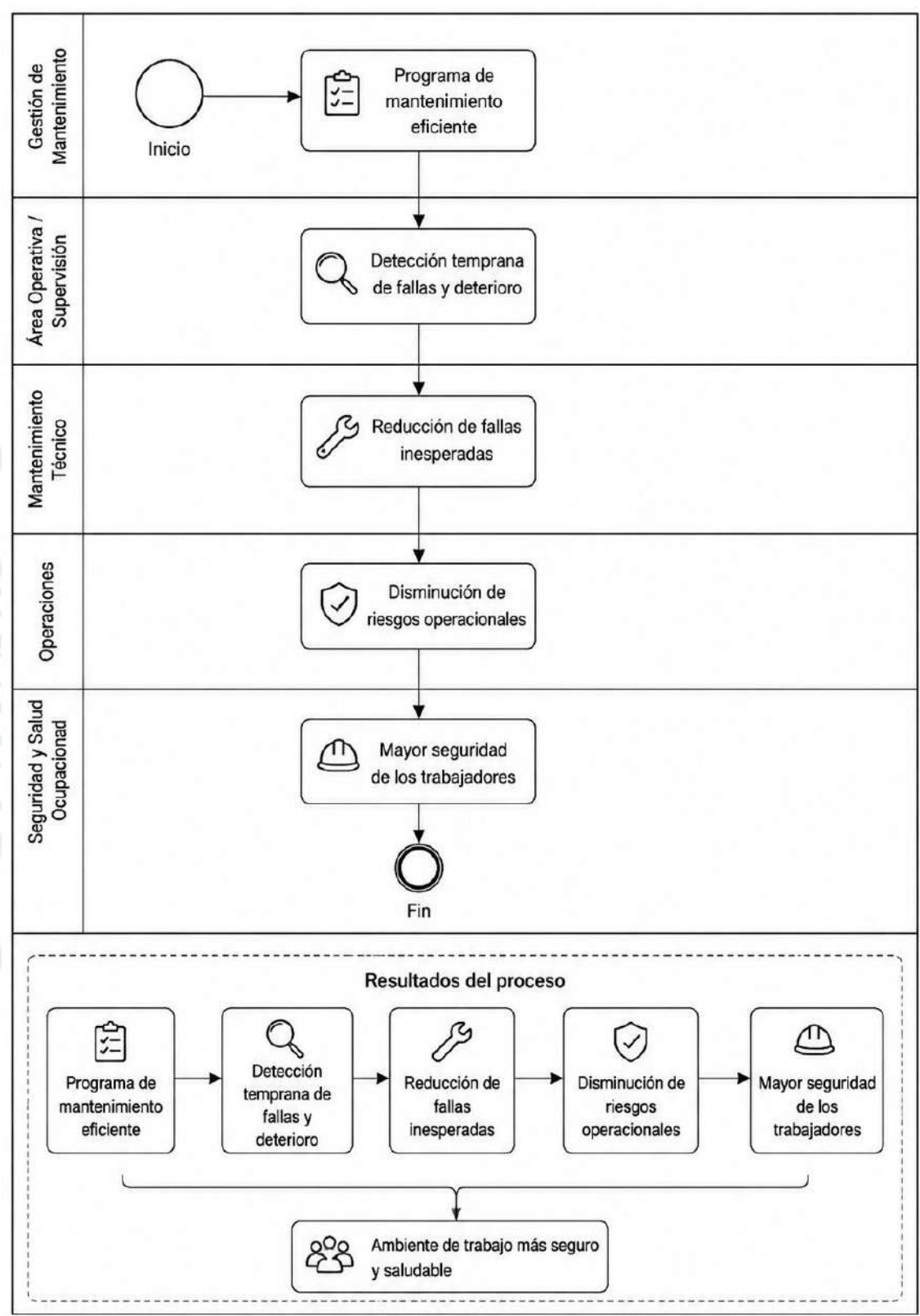
proporcionados por las organizaciones. Asimismo, la creciente digitalización de los procesos industriales exige el desarrollo de nuevas competencias y capacidades que permitan aprovechar el potencial de las tecnologías basadas en Inteligencia Artificial y analítica de datos, fortaleciendo la competitividad y sostenibilidad de las organizaciones.

Figura 12. Impacto social de una gestión eficiente del mantenimiento (BPMN simplificado)



Nota: Elaboración propia basada en International Labour Organization (2023), World Economic Forum (2023) y Davenport y Ronanki (2018)

Figura 13. Impacto del mantenimiento sobre la seguridad laboral (BPMN simplificado)



Nota: Elaboración propia basada en International Labour Organization (2023), World Economic Forum (2023) y Davenport y Ronanki (2018)

3.2.7. Factor Ecológico

El factor ecológico obtuvo una ponderación de 7% dentro del análisis PESTEL realizado para la presente investigación. Aunque represente el componente de menor influencia relativa respecto a los demás factores analizados, continúa siendo relevante debido a la creciente preocupación por la sostenibilidad, el uso eficiente de los recursos y la reducción de los impactos ambientales asociados a las actividades industriales.

Durante los últimos años, las organizaciones han incrementado su interés por incorporar criterios de sostenibilidad dentro de sus procesos operativos, motivadas tanto por exigencias regulatorias como por compromisos corporativos orientados a la protección del medio ambiente. En este contexto, la gestión eficiente de los activos industriales desempeña un papel importante, ya que influye directamente sobre el consumo energético, la generación de residuos y la utilización de recursos materiales a lo largo del ciclo de vida de los equipos. (United Nations, 2015)

Desde la perspectiva del mantenimiento industrial, una estrategia adecuada contribuye a prolongar la vida útil de los activos, reducir la frecuencia de reemplazo de componentes y minimizar la generación de residuos derivados de fallas prematuras. Según Tsang (2002), los programas de mantenimiento eficaces no solo mejoran la confiabilidad de los equipos, sino que también favorecen un uso más eficiente de los recursos disponibles.

Asimismo, la aparición de enfoques asociados a la economía circular ha impulsado una visión más amplia del mantenimiento, donde las actividades de conservación y reparación son consideradas mecanismo para extender el ciclo de vida de los activos y disminuir la necesidad de fabricar o adquirir nuevos equipos, como se ve en la tabla 14. De acuerdo con Kirchherr et al. (2017), la economía circular promueve estrategias orientadas a maximizar el aprovechamiento de los recursos mediante la reutilización, reparación y extensión de la vida útil de los productos.

Tabla 14. *Relación entre mantenimiento y sostenibilidad ambiental*

Aspecto ambiental	Contribución del mantenimiento
Consumo energético	Mejora la eficiencia operativa de los equipos
Generación de residuos	Reduce reemplazos innecesarios
Uso de recursos materiales	Prolonga la vida útil de los activos
Emisiones indirectas	Disminuye pérdidas asociadas a fallas
Sostenibilidad organizacional	Favorece la continuidad de las operaciones

Nota: Elaboración propia basada en Tsang (2002) y Kirchherr et al. (2017).

En sectores relacionados con transporte y logística, las fallas inesperadas pueden generar consumos adicionales de combustible, pérdidas de productividad y utilización ineficiente de recursos. Por ello, la capacidad de anticipar condiciones de deterioro y programar intervenciones oportunas permite reducir impactos operativos que indirectamente también afectan el desempeño ambiental de las organizaciones.

Por otra parte, la creciente digitalización de los procesos industriales ha permitido desarrollar herramientas capaces de optimizar la utilización de recursos mediante el análisis de datos y la identificación temprana de anomalías. Según Carvalho et al. (2019), los modelos de mantenimiento predictivo contribuyen a mejorar la eficiencia operativa al reducir intervenciones innecesarias y optimizar la planificación de actividades de mantenimiento.

La propuesta desarrollada en la presente investigación se encuentra alineada con estos principios, ya que busca identificar de manera anticipada aquellos equipos con mayor probabilidad de transición hacia estados críticos, permitiendo priorizar intervenciones únicamente cuando resulten necesarias, como se muestra en la tabla 15. Esta capacidad contribuye a evitar el uso excesivo de recursos, mejorar la eficiencia operativa y favorecer una gestión más sostenible de los activos industriales.

Tabla 15. Contribución del modelo propuesto a la sostenibilidad

Objetivo ambiental	Aporte del modelo propuesto
Uso eficiente de recursos	Priorización de intervenciones necesarias
Reducción de desperdicios	Menor reemplazo prematuro de componentes
Extensión de vida útil	Mantenimiento basado en condición
Optimización energética	Equipos operando en mejores condiciones
Sostenibilidad operativa	Mejor gestión del ciclo de vida de activos

Nota: Elaboración propia

Finalmente, el análisis ecológico evidencia que las prácticas modernas de mantenimiento contribuyen indirectamente al logro de objetivos de sostenibilidad mediante la optimización del uso de recursos y la extensión del ciclo de vida de los activos. Aunque su influencia sobre la problemática estudiada es menor en comparación con otros factores del entorno, continúa representando un elemento complementario que fortalece la pertinencia de soluciones basadas en mantenimiento predictivo y gestión inteligente de activos.

3.3. Descripción del sistema actual de mantenimiento

A partir de la información recopilada mediante la encuesta aplicada al área de mantenimiento y operaciones, se identificó que el sistema actual de mantenimiento en las empresas analizadas se encuentra orientado principalmente hacia estrategias preventivas. Como se observa en la figura 15, el 72.7% de los participantes indicó que el mantenimiento preventivo es el enfoque predominante en sus organizaciones, mientras que el mantenimiento predictivo aún presenta una adopción limitada (18.2%).

Asimismo, los resultados evidencian que las fallas no planificadas continúan representando una problemática relevante dentro del proceso de mantenimiento. El 72.7% de los encuestados identificó las fallas inesperadas como el principal problema presente en sus organizaciones, lo cual afecta directamente la continuidad operativa y la eficiencia en la utilización de recursos.

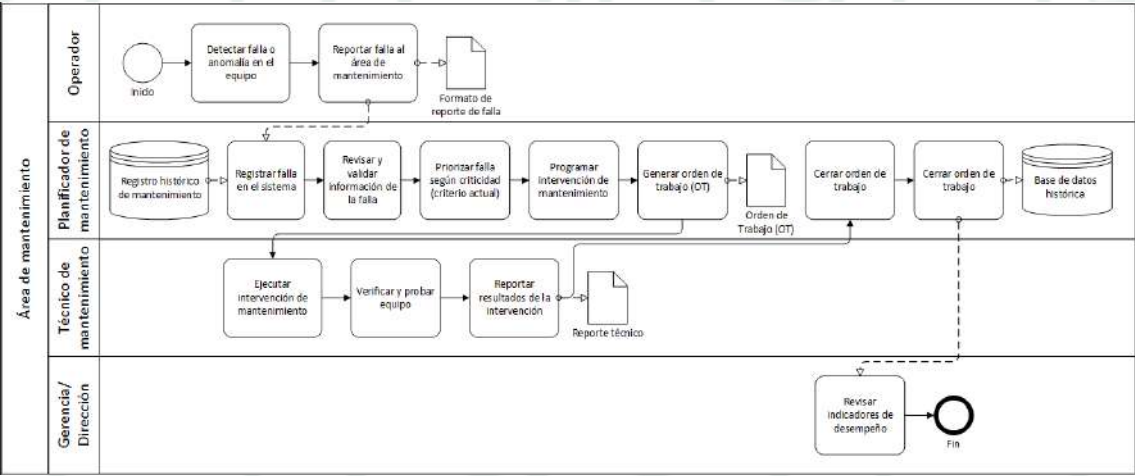
En relación a la toma de decisiones, se observó que la priorización de intervenciones se realiza principalmente mediante indicadores técnicos (45.5%) y criterios de criticidad percibida (45.5%). Esto demuestra que, aunque las empresas utilizan información técnica para apoyar sus decisiones, la priorización aún depende considerablemente de

evaluaciones convencionales y no de modelos predictivos basados en inteligencia artificial.

Por otro lado, el 81.8% de los participantes indicó que la información registrada sí es utilizada para el análisis o toma de decisiones. Del mismo modo, se identificó que las organizaciones registran principalmente información relacionada con fallas (81.8%), kilometraje u horas de uso (72.7%), tiempos de intervención (63.6%) y costos de mantenimiento (63.6%). Estos resultados evidencian la existencia de datos históricos relevantes que podrían ser aprovechados mediante técnicas de Machine Learning para mejorar la gestión del mantenimiento.

Finalmente, entre las principales limitaciones identificadas destacan la falta de tiempo para una adecuada gestión del mantenimiento (45.5%) y la resistencia al cambio (45.5%), factores que limitan la implementación de estrategias más avanzadas de análisis y optimización.

Figura 14. Esquema secuencial del proceso actual de mantenimiento



Nota: Elaboración propia a partir de la encuesta aplicada y el diagnóstico situacional

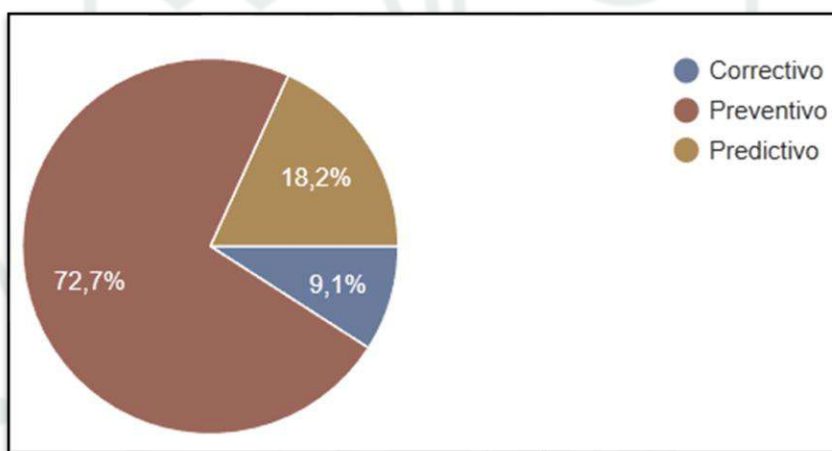
El proceso actual de mantenimiento identificado en las empresas encuestadas evidencia una gestión basada principalmente en evaluaciones técnicas convencionales y priorización operativa manual. Aunque las organizaciones registran información relacionada con fallas, tiempos de intervención y costos, las decisiones continúan dependiendo de criterios tradicionales y de la disponibilidad inmediata de recursos, lo cual puede contribuir a la ocurrencia de fallas no planificadas y a una asignación parcialmente eficiente del mantenimiento.

3.4. Identificación de problemáticas en el mantenimiento industrial

A partir de la información recopilada mediante la encuesta aplicada a personal vinculado a las áreas de mantenimiento y operaciones, se identificaron diversas problemáticas relacionadas con la gestión actual del mantenimiento en las empresas analizadas. Los resultados evidencian limitaciones asociadas a la ocurrencia de fallas no planificadas, la priorización de intervenciones y las restricciones operativas para la gestión eficiente de recursos.

Uno de los principales hallazgos corresponde al tipo de mantenimiento predominante en las organizaciones encuestadas. Como se observa en la figura 15, el 72.7% de los participantes indicó que el mantenimiento preventivo constituye la estrategia principal utilizada en sus empresas, mientras que únicamente el 18.2% señaló utilizar mantenimiento predictivo. Esto evidencia que, aunque existe una orientación preventiva en las organizaciones, la adopción de enfoques basados en análisis predictivo aún es limitada.

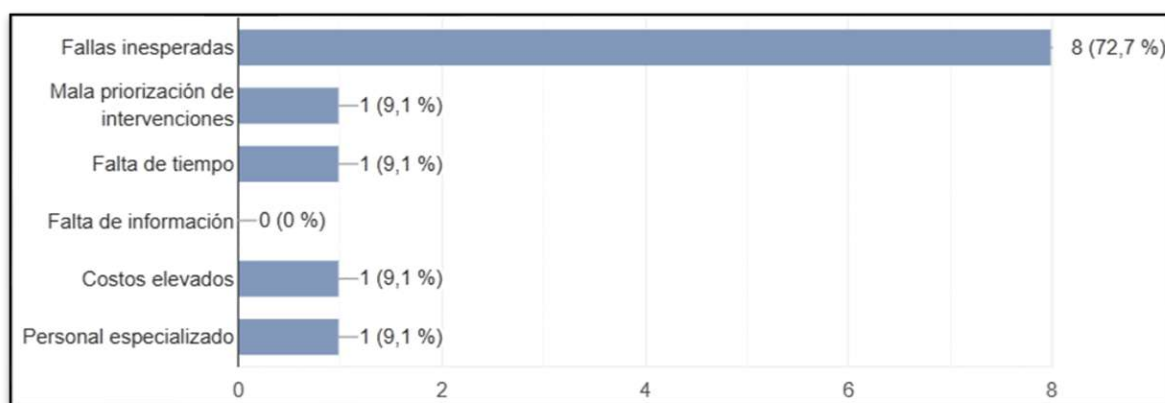
Figura 15. Tipo de mantenimiento predominante en las empresas encuestadas



Nota: Elaboración propia a partir de la encuesta aplicada

Asimismo, se identificó que las fallas inesperadas representan la principal problemática dentro del proceso de mantenimiento. El 72.7% de los encuestados indicó que las fallas no planificadas constituyen el principal problema operativo en sus organizaciones, superando otros factores como costos elevados, falta de tiempo o problemas de priorización, como se muestra en la figura 16.

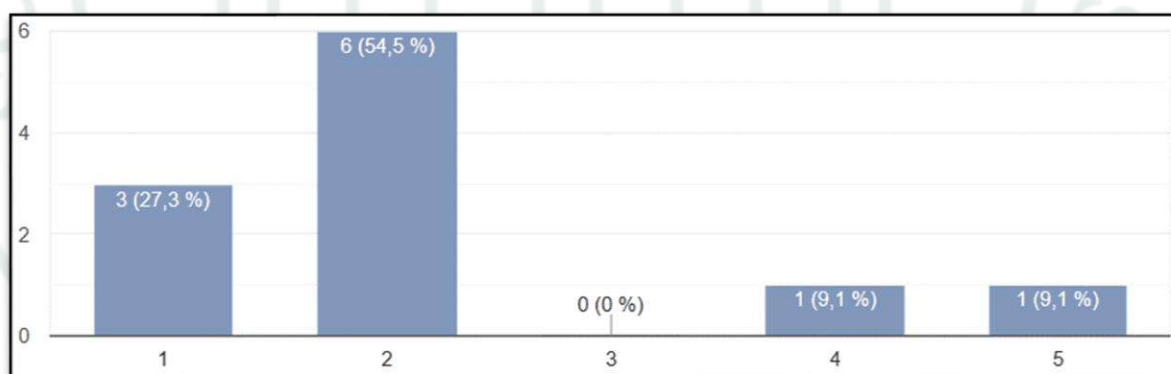
Figura 16. Principal problemática del mantenimiento en las empresas encuestadas



Nota: Elaboración propia a partir de la encuesta aplicada

En relación con la ocurrencia de fallas no planificadas, los resultados muestran que estas continúan presentándose con una frecuencia significativa dentro de las empresas analizadas, como se ve en la figura 17. Aunque parte de los participantes indicó una frecuencia baja o moderada, también se registraron respuestas asociadas a frecuencias altas, evidenciando que aún existen dificultades para anticipar oportunamente el deterioro de los equipos.

Figura 17. Frecuencia de ocurrencia de fallas no planificadas (1: Muy baja - 5: Muy alta)

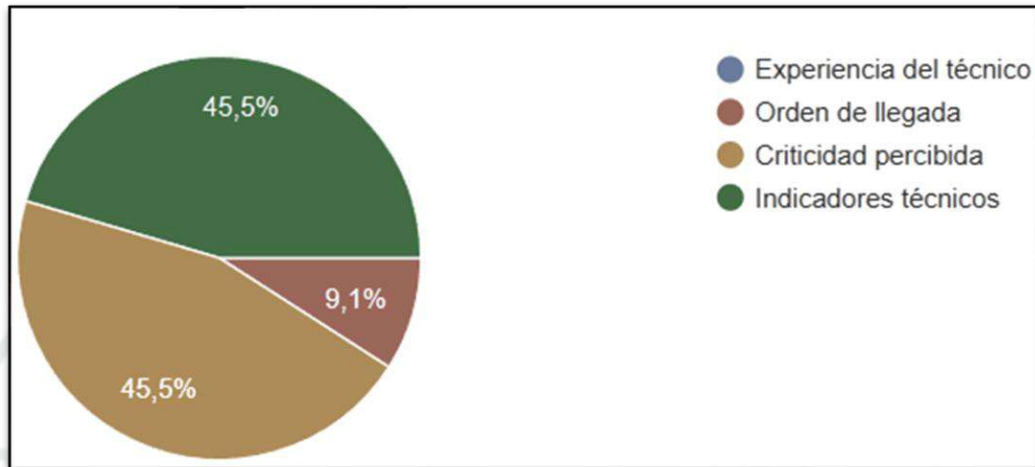


Nota: Elaboración propia a partir de la encuesta aplicada

Por otro lado, la priorización de intervenciones continúa dependiendo principalmente de criterios tradicionales. El 45.5% de los participantes indicó que las decisiones se toman mediante indicadores técnicos, mientras que otro 45.5% señaló que la criticidad percibida constituye el principal criterio de priorización, así como se ve en la figura 18. Esto demuestra que, aunque existe utilización de información operativa, la toma de

decisiones aún no incorpora modelos inteligentes de análisis predictivo ni mecanismos de optimización para la asignación eficiente de recursos.

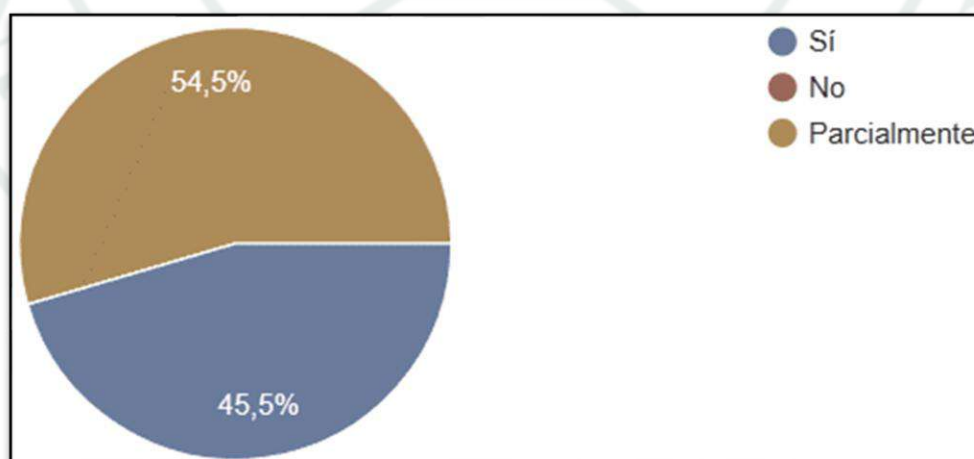
Figura 18. Criterios utilizados para la priorización de intervenciones



Nota: Elaboración propia a partir de la encuesta aplicada

Del mismo modo, los resultados evidencian limitaciones relacionadas con la eficiencia de la priorización actual, como se ve en la figura 19. El 54.5% de los encuestados considera que la priorización de intervenciones es solo parcialmente eficiente, lo cual refleja oportunidades de mejora en la planificación y asignación de recursos de mantenimiento.

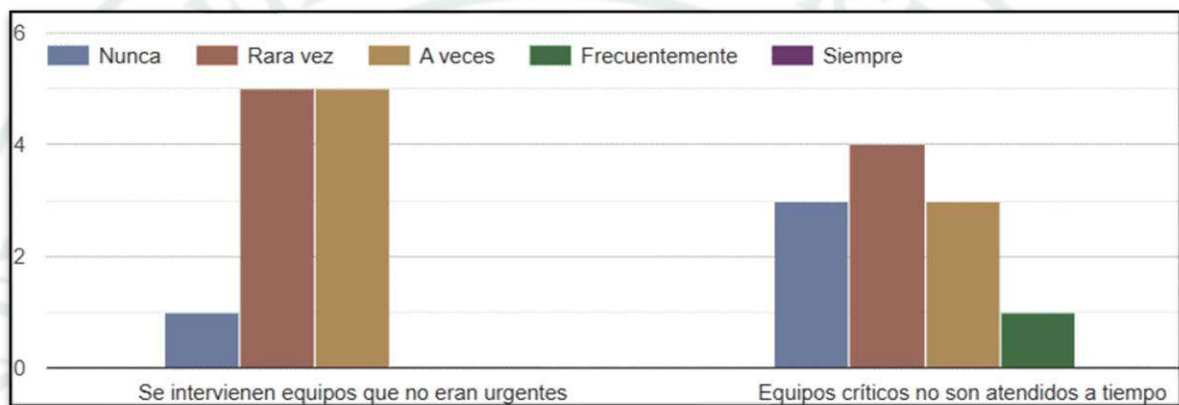
Figura 19. Percepción sobre la eficiencia de la priorización de intervenciones



Nota: Elaboración propia a partir de la encuesta aplicada

Otro aspecto relevante identificado corresponde a la atención de equipos críticos y a la asignación de intervenciones. Los resultados, vistos en la figura 20, muestran que, en determinados casos, se realizan intervenciones sobre equipos no urgentes mientras algunos equipos críticos no son atendidos oportunamente. Esta situación evidencia posibles deficiencias en los mecanismos actuales de priorización y programación de mantenimiento.

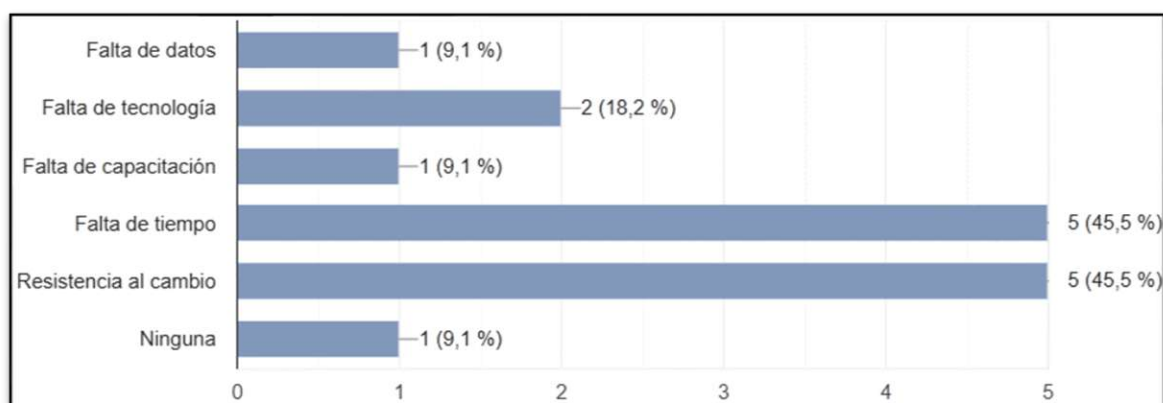
Figura 20. Frecuencia de problemas relacionados con la priorización de intervenciones



Nota: Elaboración propia a partir de la encuesta aplicada

Finalmente, se identificó que las principales limitaciones para mejorar el mantenimiento corresponden a la falta de tiempo operativo y a la resistencia al cambio, ambas con un 45.5% de incidencia entre los participantes, como se ve en la figura 21. Estos factores representan barreras importantes para la incorporación de tecnologías avanzadas y metodologías de mantenimiento basadas en análisis de datos e inteligencia artificial.

Figura 21. Factores que limitan la mejora del mantenimiento



Nota: Elaboración propia a partir de la encuesta aplicada

En conjunto, los resultados obtenidos evidencian que las empresas analizadas representan condiciones favorables para la incorporación de modelos inteligentes de mantenimiento, debido a que ya existe registro de información operativa e interés en mejorar la eficiencia del proceso. Sin embargo, persisten problemáticas relacionadas con la ocurrencia de fallas inesperadas, la priorización parcial de intervenciones y las limitaciones operativas, justificando así la necesidad de implementar herramientas orientadas a mejorar la toma de decisiones en mantenimiento industrial.

3.5. Análisis de la situación actual del mantenimiento

Con el propósito de evaluar el estado actual de la gestión del mantenimiento en las empresas analizadas, se realizó un análisis de los principales indicadores identificados mediante la encuesta aplicada al personal vinculado a las áreas de mantenimiento y operaciones. Estos indicadores permitieron obtener evidencia cuantitativa sobre las prácticas actuales de mantenimiento, el uso de información operativa y la eficiencia percibida en la toma de decisiones.

Uno de los principales indicadores analizados corresponde al tipo de mantenimiento predominante en las organizaciones encuestadas. Los resultados muestran que el 72.7% de las empresas utiliza principalmente mantenimiento preventivo, mientras que el mantenimiento predictivo aún presenta una adopción reducida del 18.2% (Ver tabla 16). Este resultado evidencia que, aunque existe una orientación preventiva, la utilización de herramientas avanzadas para anticipar fallas aún es limitada.

Tabla 16. *Tipo de mantenimiento predominante*

Tipo de mantenimiento	Frecuencia	Porcentaje
Preventivo	8	72.7%
Predictivo	2	18.2%
Correctivo	1	9.1%
Total	11	100.0%

Nota: Elaboración propia a partir de la encuesta aplicada

Otro indicador relevante corresponde a la frecuencia de ocurrencia de fallas no planificadas. Los resultados muestran que las fallas inesperadas continúan siendo una problemática presente dentro de las empresas analizadas, lo cual afecta la continuidad operativa y genera necesidades de intervención no previstas (Ver tabla 17).

Tabla 17. *Frecuencia de fallas no planificadas*

Nivel	Frecuencia	Porcentaje
Muy baja	3	27.3%
Baja	6	54.5%
Alta	1	9.1%
Muy alta	1	9.1%
Total	11	100.0%

Nota: Elaboración propia a partir de la encuesta aplicada

En relación con los mecanismos utilizados para la toma de decisiones, se identificó que la priorización de intervenciones se realiza principalmente mediante indicadores técnicos y criticidad percibida, ambos con una participación de 45.5% (Ver tabla 18). Esto demuestra que las organizaciones utilizan información operativa para apoyar la gestión del mantenimiento, sin embargo, aún no se incorporan modelos predictivos ni algoritmos de optimización para la asignación eficiente de recursos.

Tabla 18. *Criterios de priorización de intervenciones*

Criterio utilizado	Frecuencia	Porcentaje
Indicadores técnicos	5	45.5%
Criticidad percibida	5	45.5%
Orden de llegada	1	9.1%
Total	11	100.0%

Nota: Elaboración propia a partir de la encuesta aplicada

Asimismo, se analizó la disponibilidad y utilización de información operativa dentro del procesos de mantenimiento (Ver tabla 19). Los resultados muestran que las empresas registran principalmente información relacionada con fallas (81.8%), kilometraje y horas de uso (72.7%), tiempos de intervención (63.6%) y costos de mantenimiento (63.6%). Esto evidencia la existencia de datos históricos relevantes que pueden ser aprovechados mediante técnicas de análisis predictivo.

Tabla 19. *Información registrada en el proceso de mantenimiento*

Tipo de información registrada	Frecuencia	Porcentaje
Fallas	9	81.8%
Kilometraje/Horas de uso	8	72.7%
Tiempo de intervención	7	63.6%
Costos de mantenimiento	7	63.6%
Datos de sensores	6	54.5%

Nota: Elaboración propia a partir de la encuesta aplicada

Del mismo modo, el 81.8% de los participantes indicó que la información registrada sí es utilizada para análisis de toma de decisiones (Ver tabla 20). Este resultado evidencia que las organizaciones cuentan con una base mínima de información operativa susceptible de ser utilizada en modelos de inteligencia artificial y mantenimiento predictivo u optimización.

Tabla 20. *Uso de información para análisis y toma de decisiones*

Respuesta	Frecuencia	Porcentaje
Si	9	81.8%
Parcialmente	2	18.2%
No	0	0.0%
Total	11	100.0%

Nota: Elaboración propia a partir de la encuesta aplicada

Finalmente, se identificó que las principales limitaciones para mejorar el mantenimiento corresponden a la falta de tiempo de intervención y a la resistencia al cambio, ambas con una incidencia del 45.5% (Ver tabla 21). Estas restricciones representan factores críticos que dificultan la implementación de estrategias avanzadas de mantenimiento y limitan la eficiencia en la asignación de recursos.

Tabla 21. *Factores limitantes para la mejora del mantenimiento*

Factor limitante	Frecuencia	Porcentaje
Falta de tiempo	5	45.5%
Resistencia al cambio	5	45.5%
Falta de tecnología	2	18.2%
Falta de datos	1	9.1%
Falta de capacitación	1	9.1%

Nota: Elaboración propia a partir de la encuesta aplicada

En conjunto, los indicadores analizados evidencian que las empresas encuestadas cuentan con procesos de mantenimiento parcialmente estructurados y con disponibilidad de información operativa relevante; sin embargo, persisten limitaciones relacionadas con la ocurrencia de fallas inesperadas, la priorización parcial de intervenciones y la ausencia de modelos inteligentes de análisis y optimización, justificando así la necesidad de implementar herramientas para mejorar la toma de decisiones en mantenimiento industrial.

3.6. Análisis de causas del problema

Con el propósito de identificar las principales causas asociadas a las problemáticas detectadas en el proceso de mantenimiento industrial, se realizó un análisis causa-efecto en la información obtenida mediante la encuesta aplicada y el diagnóstico situacional

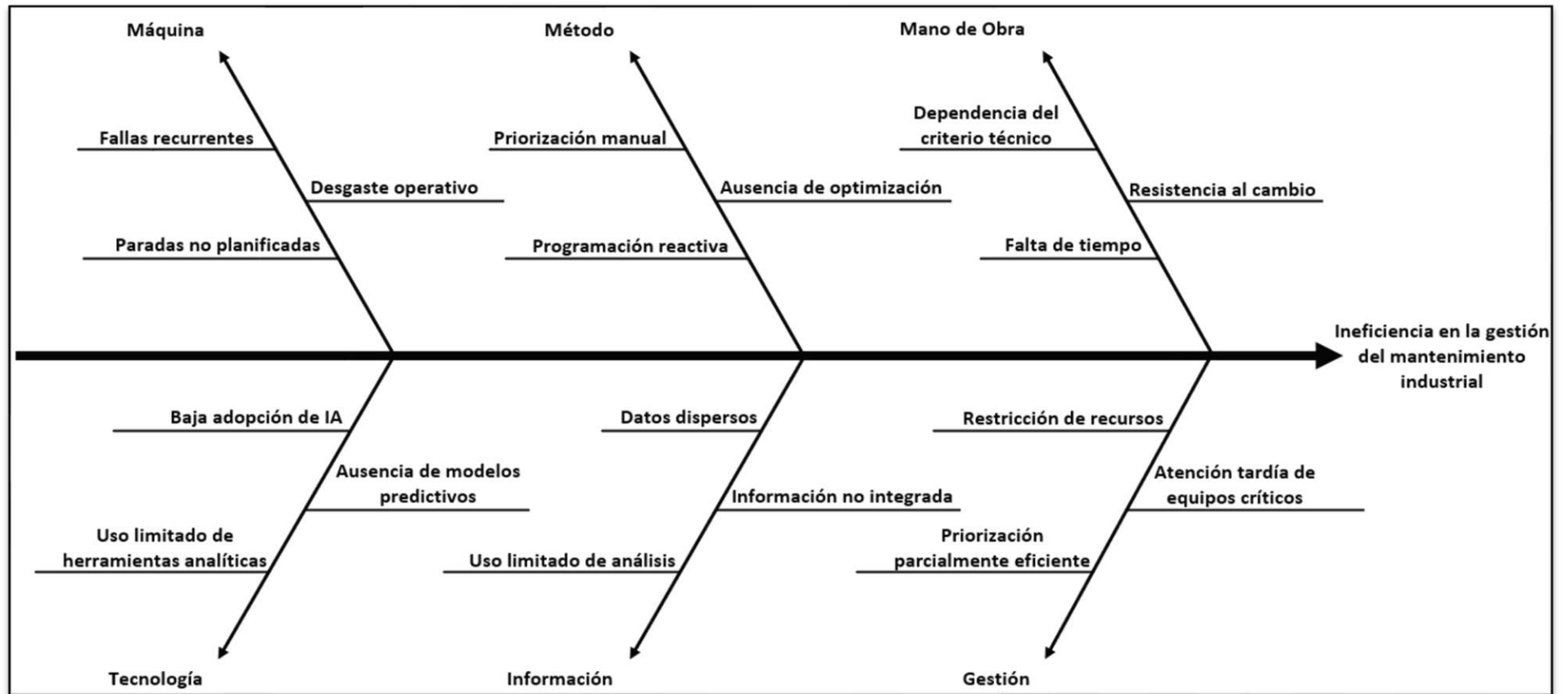
desarrollado en las secciones anteriores. Este análisis permitió reconocer los factores que contribuyen a la ocurrencia de fallas no planificadas, la priorización parcialmente eficiente de intervenciones y las limitaciones operativas presentes en las empresas analizadas.

A partir de los resultados obtenidos, se identificó que las problemáticas actuales del mantenimiento no dependen de un único factor, sino de la interacción de elementos técnicos, operativos y organizacionales. Entre las principales causas identificadas destacan la dependencia de criterios convencionales para la toma de decisiones, el uso limitado de herramientas predictivas, ocurrencia de fallas inesperadas y las restricciones relacionadas con el tiempo y adaptación tecnológica.

Asimismo, aunque las empresas registran información relevante como fallas, tiempos de intervención, costos y datos operativos, se evidencio que dicha información aún no es aprovechada mediante modelos inteligentes capaces de anticipar comportamientos futuros u optimizar la asignación de recursos. Como consecuencia, las decisiones continúan dependiendo principalmente de evaluaciones técnicas manuales y criterios de criticidad percibida.

Con base en estos hallazgos, se elaboró el siguiente diagrama de Ishikawa, como se ve en la figura 22, el cual resume las principales causas asociadas a las deficiencias identificadas en el proceso de mantenimiento.

Figura 22. Diagrama de Ishikawa



Nota: Elaboración propia

El diagrama de Ishikawa evidencia que las problemáticas identificadas en el proceso de mantenimiento industrial están asociadas a factores técnicos, operativos y organizacionales, destacando la dependencia de criterios manuales de priorización, la limitada incorporación de herramientas predictivas y las restricciones de tiempo y recursos para la gestión eficiente del mantenimiento.

3.6.1. Máquina

3.6.1.1. Fallas recurrentes

Las fallas recurrentes representan una de las causas más visibles de la ineficiencia en la gestión del mantenimiento industrial, debido a que evidencian la presencia de problemas repetitivos en los equipos o componentes. De acuerdo con los resultados de la encuesta aplicada, el 72.7% de los participantes identificó las fallas inesperadas como el principal problema del mantenimiento en sus organizaciones. Este resultado muestra que, aun cuando predomina el mantenimiento preventivo, todavía existen eventos de falla que no son anticipados oportunamente.

La recurrencia de fallas puede estar asociada a intervenciones que corrigen el síntoma inmediato, pero no necesariamente la causa raíz del problema. En este sentido, cuando no existe un análisis histórico del comportamiento del equipo, las mismas condiciones de deterioro pueden repetirse en el tiempo, generando nuevas intervenciones, mayores costos y reducción de la disponibilidad operativa. Esta situación justifica la necesidad de incorporar modelos predictivos capaces de identificar patrones previos a la falla.

3.6.1.2. Desgaste operativo

El desgaste operativo corresponde al deterioro progresivo de los equipos como consecuencia del uso continuo, las condiciones de trabajo, la carga operativa y el envejecimiento de los componentes. En sectores como transporte, logística e industria, los activos están sometidos a condiciones exigentes que pueden acelerar su degradación si no se monitorean adecuadamente.

Este desgaste no siempre se manifiesta de manera inmediata, sino que puede evolucionar gradualmente hasta convertirse en una falla crítica. Por ello, depender únicamente de inspecciones manuales o de cronogramas fijos puede ser insuficiente para detectar señales tempranas en deterioro. En este punto, el uso de datos históricos y

modelos de Machine Learning permite analizar la evolución de la condición de los equipos y anticipar posibles transiciones hacia estado de seguimiento o crítico.

3.6.1.3. Paradas no planificadas

Las paradas no planificadas son interrupciones inesperadas de la operación ocasionadas por fallas no previstas. Estas paradas afectan directamente la continuidad del servicio, la productividad y la programación de actividades. Aunque en la encuesta la mayoría de participantes reportó frecuencias bajas o moderadas de fallas no planificadas, el hecho de que las fallas inesperadas hayan sido señaladas como el principal problema evidencia que su impacto operativo es significativo.

Desde el punto de vista de la gestión del mantenimiento, una parada no planificada genera urgencias, reasignación de recursos, retrasos y posibles sobrecostos. Además, reduce la capacidad de planificar intervenciones de manera ordenada. Por ello, una estrategia basada en predicción y priorización permite disminuir la incertidumbre, anticipar escenarios críticos y reducir la dependencia de respuestas reactivas.

3.6.2. Método

3.6.2.1. Priorización manual

La priorización manual se refiere a la selección de equipos a intervenir mediante criterios definidos principalmente por el personal responsable, sin el apoyo de modelos analíticos formales. En la encuesta, el 45.5% de los participantes indicó que las decisiones se toman mediante indicadores técnicos, mientras que otro 45.5% señaló que se utiliza la criticidad percibida. Esto demuestra que existen criterios técnicos, pero también un componente importante de interpretación humana.

Aunque la experiencia del personal es valiosa, la priorización manual puede generar variabilidad en las decisiones, especialmente cuando existen varios equipos con necesidades simultáneas de mantenimiento. En estos casos, resulta difícil evaluar de manera conjunta la probabilidad de falla, el tiempo requerido, el costo de intervención y el beneficio esperado. Esta limitación justifica la incorporación de un modelo de optimización que permita seleccionar objetivamente las intervenciones con mayor impacto.

3.6.2.2. Programación reactiva

La ausencia de optimización implica que la programación de intervenciones no considera formalmente restricciones y objetivos de manera simultánea. En la práctica, las empresas pueden priorizar equipos por urgencia, disponibilidad del personal o percepción de criticidad, pero sin evaluar cuál combinación de intervenciones maximiza el beneficio total bajo restricciones de tiempo o recursos.

Esta causa se relaciona directamente con el segundo nivel de la propuesta, donde se plantea un modelo de optimización basado en programación dinámica. Dicho modelo permite seleccionar el conjunto de equipos a intervenir considerando el beneficio esperado y el tiempo disponible. De esta manera, la decisión deja de ser únicamente secuencial o empírica y pasa a estar respaldada por un procedimiento cuantitativo.

3.6.2.3. Ausencia de optimización

La programación reactiva ocurre cuando las intervenciones se ejecutan principalmente como respuesta a fallas, alertas o problemas ya ocurridos. Aunque las empresas encuestadas reportan predominio del mantenimiento preventivo, la presencia de fallas inesperadas como principal problemática indica que todavía existen componentes reactivos dentro del sistema actual.

Este tipo de programación limita la capacidad de anticipación y obliga a reorganizar recursos de manera imprevista. Además, puede provocar que algunas actividades programadas sean postergadas para atender emergencias. La propuesta de investigación busca reducir esta limitación mediante la predicción del estado futuro de los equipos y la priorización optimizada de intervenciones antes de que se presenten condiciones críticas.

3.6.3. Mano de Obra

3.6.3.1. Dependencia del criterio técnico

La dependencia del criterio técnico hace referencia al peso que tiene la experiencia del personal en la toma de decisiones de mantenimiento. Esta experiencia es fundamental para interpretar condiciones operativas, identificar señales de deterioro y ejecutar intervenciones adecuadas. Sin embargo, cuando no se complementa con herramientas analíticas, puede generar decisiones heterogéneas entre diferentes técnicos o áreas.

En la encuesta se evidenció que la toma de decisiones se apoya principalmente en indicadores técnicos y criticidad percibida. Esto muestra que el conocimiento humano sigue siendo central, pero también evidencia la oportunidad de fortalecerlo con modelos predictivos. La propuesta no busca reemplazar al personal técnico, sino proporcionarle información adicional basada en probabilidades de falla y priorización objetiva.

3.6.3.2. Falta de tiempo

La resistencia al cambio fue identificada como una de las principales limitaciones para mejorar el mantenimiento, con una incidencia del 45.5% en los resultados de la encuesta. Esta resistencia puede presentarse cuando se introducen nuevas tecnologías, metodologías o herramientas que modifican la forma tradicional de trabajo.

En contextos industriales, la adopción de inteligencia artificial puede generar dudas relacionadas con la confiabilidad del modelo, la pérdida de control sobre las decisiones o la necesidad de capacitación adicional. Por ello, cualquier propuesta tecnológica debe ser entendida como un sistema de apoyo a la decisión y no como una sustitución del criterio profesional. Esta consideración es importante para facilitar la aceptación del modelo dentro de las organizaciones.

3.6.3.3. Resistencia al cambio

La falta de tiempo también fue identificada como una limitación relevante, con 45.5% de incidencia entre los participantes. Esta causa afecta directamente la capacidad del personal para analizar datos, revisar historiales, comparar alternativas y planificar intervenciones con mayor detalle.

Cuando el tiempo disponible es limitado, las decisiones tienden a enfocarse en resolver problemas inmediatos, reduciendo el espacio para el análisis preventivo o predictivo. Esta situación refuerza la importancia del modelo de optimización, ya que permite seleccionar intervenciones considerando una restricción explícita de tiempo, como se desarrolló posteriormente en la investigación.

3.6.4. Tecnología

3.6.4.1. Baja adopción de Inteligencia Artificial

La baja adopción de inteligencia artificial constituye una causa tecnológica relevante dentro del sistema actual de mantenimiento. Aunque las empresas encuestadas registran

información operativa y en muchos casos la utilizan para análisis o toma de decisiones, esto no significa necesariamente que empleen modelos predictivos avanzados o algoritmos de optimización.

La inteligencia artificial permite identificar patrones complejos que no siempre son visibles mediante análisis manual o revisión de registros. Por ello, su baja adopción limita la capacidad de anticipar fallas y generar recomendaciones basadas en datos. Esta causa se conecta directamente con la propuesta de la tesis, que plantea el uso de Random Forest y XGBoost para predecir el estado futuro de los equipos.

3.6.4.2. Ausencia de modelos predictivos

La ausencia de modelos predictivos implica que las organizaciones no cuentan con herramientas capaces de estimar probabilidades futuras de falla o deterioro. En consecuencia, las decisiones se basan principalmente en el estado actual del equipo, el historial observado o la percepción técnica del responsable.

Esta limitación es importante porque el mantenimiento moderno no solo requiere saber qué ocurrió, sino anticipar qué podría ocurrir. El modelo predictivo desarrollado en la investigación responde precisamente a esta necesidad, al transformar registros históricos en probabilidades asociadas a estados futuros: normal, seguimiento y crítico.

3.6.4.3. Uso limitado de herramientas analíticas

El uso limitado de herramientas analíticas no significa necesariamente ausencia de datos. De hecho, la encuesta mostró que las empresas registran información sobre fallas, tiempos de intervención, costos, kilometraje u horas de uso y datos de sensores. Sin embargo, el desafío está en convertir esa información en conocimiento útil para la toma de decisiones.

Cuando los datos se utilizan solo para registro o consulta, se pierde la posibilidad de identificar patrones, tendencias o relaciones entre variables. El uso de herramientas analíticas avanzadas permite aprovechar mejor los datos disponibles, mejorar la precisión de las decisiones y construir modelos de mantenimiento predictivo y prescriptivo.

3.6.5. Información

3.6.5.1. Datos dispersos

Los datos dispersos constituyen una causa frecuente en la gestión del mantenimiento, especialmente cuando la información se encuentra distribuida en diferentes archivos, reportes, sistemas o áreas responsables. Esta dispersión dificulta el análisis integral de la condición de los equipos y limita la capacidad de construir una visión completa del comportamiento histórico de los activos.

Aunque las empresas encuestadas registran información relevante, esta puede encontrarse fragmentada según el tipo de dato o el área que la genera. Para aplicar modelos de Machine Learning, resulta necesario consolidar dicha información en una base estructurada, limpia y coherente. Por ello, la fase de preprocesamiento de datos fue un componente fundamental del modelo desarrollado.

3.6.5.2. Información no integrada

La información no integrada se refiere a la falta de conexión entre distintos tipos de registros, como fallas, costos, tiempos de intervención, condiciones de operación y datos de sensores. Cuando estos elementos se analizan de manera aislada, es difícil determinar la relación entre el estado del equipo, el impacto económico y la necesidad de intervención.

La integración de información permite construir variables más completas y útiles para el análisis predictivo. En la investigación, esta necesidad se abordó mediante la ingeniería de características, generando variables temporales, históricas y de estado previo que permitieron representar mejor la evolución de los equipos.

3.6.5.3. Uso limitado de análisis

El uso limitado de análisis se refiere a la utilización parcial de la información disponible. La encuesta mostró que el 81.8% de los participantes indicó que la información registrada sí se utiliza para análisis o toma de decisiones; sin embargo, esto no implica necesariamente el uso de técnicas avanzadas como Machine Learning, optimización o análisis probabilístico.

En este sentido, el problema no es únicamente la existencia de datos, sino el nivel de profundidad con que estos son analizados. La propuesta busca elevar ese nivel de análisis, pasando de una revisión descriptiva o técnica a un enfoque predictivo y

prescriptivo, capaz de generar probabilidades de falla y seleccionar intervenciones óptimas.

3.6.6. Gestión

3.6.6.1. Restricción de recursos

La restricción de recursos representa una causa crítica en la gestión del mantenimiento industrial. Las organizaciones no siempre disponen de tiempo, personal, presupuesto o repuestos suficientes para atender simultáneamente todos los requerimientos de mantenimiento. Esta limitación obliga a seleccionar qué equipos serán intervenidos primero y cuáles deberán esperar.

En la encuesta, la falta de tiempo fue identificada como una de las principales barreras para mejorar el mantenimiento. Esta condición justifica la necesidad de modelos de optimización, ya que permiten tomar decisiones bajo restricciones reales. En la investigación, esta problemática se abordó mediante un modelo tipo mochila 0–1, resuelto con programación dinámica, donde el tiempo disponible actúa como restricción principal.

3.6.6.2. Priorización parcialmente eficiente

La priorización parcialmente eficiente fue evidenciada en la encuesta, donde el 54.5% de los participantes señaló que la priorización actual es solo parcialmente eficiente. Este resultado indica que, aunque existen criterios de decisión, todavía hay oportunidades de mejora en la forma en que se seleccionan y programan las intervenciones.

Esta causa sintetiza varias limitaciones del sistema actual: dependencia del criterio técnico, ausencia de optimización, restricciones de tiempo y uso limitado de análisis predictivo. Por ello, constituye uno de los principales puntos de conexión con la propuesta de la investigación, orientada a mejorar la priorización mediante la integración de Machine Learning y programación dinámica.

3.6.6.3. Atención tardía de equipos críticos

La atención tardía de equipos críticos ocurre cuando un activo con alto riesgo o impacto operativo no es intervenido oportunamente. Esta situación puede producirse por falta de recursos, errores de priorización, ausencia de información integrada o dependencia excesiva de criterios manuales.

El problema es especialmente relevante porque los equipos críticos pueden generar mayores consecuencias en caso de falla, afectando la continuidad operativa y elevando los costos de mantenimiento. La propuesta planteada contribuye a reducir esta causa al identificar probabilidades de transición hacia estados críticos y utilizar dichas probabilidades como insumo para la priorización optimizada.

3.6.7. Síntesis del análisis causal

El análisis de causas realizado mediante el diagrama de Ishikawa evidencia que la ineficiencia en la gestión del mantenimiento industrial no se origina en un único factor, sino en la interacción de causas técnicas, metodológicas, humanas, tecnológicas, informativas y de gestión. Entre las causas más relevantes destacan las fallas inesperadas, la ausencia de modelos predictivos, la priorización manual, la falta de tiempo y la restricción de recursos.

Estos hallazgos justifican la necesidad de una propuesta que no se limite únicamente a predecir fallas, sino que también permita optimizar la toma de decisiones. En ese sentido, el modelo planteado en la investigación integra un primer nivel predictivo basado en Machine Learning y un segundo nivel de optimización basado en programación dinámica, con el propósito de anticipar condiciones críticas y seleccionar intervenciones que maximicen el beneficio esperado bajo restricciones operativas.

3.7. Evaluación de alternativas de solución

Con el propósito de determinar la alternativa más adecuada para abordar las problemáticas identificadas en el proceso de mantenimiento industrial, se realizó una evaluación comparativa de diferentes enfoques de mejora aplicables al contexto analizado. Las alternativas consideradas incluyeron estrategias tradicionales de mejora de procesos, metodologías de optimización operativa y enfoques basados en análisis inteligente de datos.

Entre las alternativas evaluadas, se consideró la reingeniería del proceso de mantenimiento, la implementación de la metodología SMED para la reducción de tiempos operativo, la utilización de software convencional de gestión de mantenimiento y un modelo inteligente basado en Machine Learning y optimización matemática (Ver tabla 25). Para realizar la comparación se definieron criterios relacionados con la

capacidad de anticipar fallas, optimización de recursos, utilización de datos históricos, escalabilidad e impacto operativo esperado.

Tabla 22. *Matriz de enfrentamiento de criterios*

Criterio	1	2	3	4	5	Total
1	-	1	1	1	1	4
2	0	-	1	1	1	3
3	0	0	-	1	1	2
4	0	0	0	-	1	1
5	0	0	0	0	-	0

Nota: Elaboración propia

Tabla 23. *Pesos normalizados en la evaluación*

	Criterio	Puntos	Peso	Peso normalizado
1	Anticipación de fallas	4	40%	30%
2	Optimización de recursos	3	30%	25%
3	Impacto operativo esperado	2	20%	20%
4	Uso de datos históricos	1	10%	15%
5	Escalabilidad	0	0%	10%

Nota: Elaboración propia

Tabla 24. *Criterios de evaluación de la matriz de decisión*

Criterio	Peso	Acumulado
Anticipación de fallas	30%	30%
Optimización de recursos	25%	55%
Impacto operativo esperado	20%	75%
Uso de datos históricos	15%	90%
Escalabilidad	10%	100%

Nota: Elaboración propia

Con el propósito de justificar los pesos utilizados en la matriz de evaluación de alternativas, se realizó la priorización de criterios basada en el principio de Pareto (Ver

tabla 24). Los resultados evidencian que la capacidad de anticipar fallas, la optimización de recursos y el impacto operativo esperado concentran aproximadamente el 75% del impacto sobre la problemática identificada, razón por la cual dichos criterios recibieron una mayor ponderación dentro del proceso de evaluación.

Tabla 25. *Matriz de evaluación de alternativas*

Criterio	Peso	Reingeniería	SMED	Software	IA + optimización
Anticipación de fallas	0.30	2	1	3	5
Optimización de recursos	0.25	3	4	3	4
Impacto operativo	0.20	2	1	4	5
Uso de datos históricos	0.15	3	3	4	4
Escalabilidad	0.10	4	4	4	4

Nota: Elaboración propia

Tabla 26. *Resultados Ponderados*

Alternativa	Puntaje final
Reingeniería	2.60
SMED	2.35
Software convencional	3.45
IA + Optimización	4.50

Nota: Elaboración propia

Los resultados obtenidos en la tabla 26 muestran que la alternativa basada en inteligencia artificial y optimización matemática alcanzó el mayor puntaje global, debido principalmente a su capacidad para integrar análisis predictivo, utilización de información histórica y optimización de recursos bajos restricciones operativas. En consecuencia, esta alternativa fue seleccionada como base para el desarrollo de la propuesta planteada en la presente investigación.

3.8. Justificación de la propuesta

El diagnóstico situacional realizado evidenció que las empresas analizadas presentan problemáticas relacionadas con fallas no planificadas, priorización parcialmente eficiente de intervenciones y limitaciones operativas para la gestión adecuada de

recursos. Asimismo, aunque las organizaciones registran información relevante sobre fallas, tiempos de intervención, costos y condiciones operativas, la toma de decisiones continúa dependiendo principalmente de criterios tradicionales y evaluaciones manuales.

En este contexto, la propuesta planteada busca aprovechar los datos históricos disponibles mediante técnicas de Machine Learning para predecir el estado futuro de los equipos y estimar probabilidades de falla. Complementariamente, se incorpora un modelo de optimización basado en programación dinámica, orientado a seleccionar el conjunto óptimo de intervenciones bajo restricciones de tiempo operativo, maximizando el beneficio esperado del mantenimiento.

La integración de ambos enfoques permite evolucionar hacia una estrategia de mantenimiento predictivo y prescriptivo alienada con los principios de industria 4.0, contribuyendo a mejorar la toma de decisiones, optimizar recursos y reducir la ocurrencia de fallas inesperadas. Además, la evaluación comparativa de alternativas realizada previamente evidenció que la propuesta basada en inteligencia artificial y optimización matemática representa la alternativa con mayor potencial para abordar integralmente las problemáticas identificadas.



CAPÍTULO IV

4. Propuesta del modelo

El diagnóstico situacional desarrollado en el capítulo anterior permitió identificar las principales limitaciones del proceso de mantenimiento industrial relacionadas con la toma de decisiones, el aprovechamiento de la información histórica y la asignación eficiente de recursos, concluyendo que todas están asociadas a la utilización de enfoques tradicionales.

Asimismo, el análisis de causas realizado permitió evidenciar que gran parte de los problemas relacionados con la disponibilidad de los equipos, la eficiencia en la asignación de recursos y la gestión del mantenimiento se encuentran vinculados a la falta de mecanismos que transformen la información operativa en conocimiento útil para la toma de decisiones. En este sentido, la evaluación de alternativas realizada determinó que la aplicación de técnicas de análisis de datos e inteligencia artificial constituye la alternativa con mayor potencial para abordar las causas identificadas y contribuir a la optimización del proceso de mantenimiento.

A partir de estos hallazgos, el presente capítulo desarrolla la propuesta de solución planteada en la investigación, la cual se estructura en tres niveles complementarios. En primer lugar, se presenta el desarrollo de un modelo predictivo basado en técnicas de Machine Learning, orientado a estimar el estado futuro de los equipos industriales a partir de información histórica de mantenimiento. En segundo lugar, se desarrolla un modelo de priorización y optimización que utiliza las probabilidades generadas por el modelo predictivo para seleccionar las intervenciones que maximizan el beneficio esperado bajo restricciones operativas. Finalmente, se establece el esquema de evaluación comparativa que permitirá analizar el desempeño de la propuesta frente al método tradicional de mantenimiento.

De esta manera, el capítulo constituye la materialización de la solución propuesta a partir de los problemas identificados en el diagnóstico, integrando herramientas de inteligencia artificial y optimización matemática con el propósito de fortalecer la toma de decisiones y mejorar la eficiencia del proceso de mantenimiento industrial.

4.1. Base de datos utilizada para el entrenamiento del modelo

La construcción de modelos basados en Inteligencia Artificial requiere disponer de información histórica confiable que permita identificar patrones y relaciones asociadas

al comportamiento de los equipos industriales. En este sentido, la calidad de los resultados obtenidos por los algoritmos de Machine Learning depende directamente de la calidad, consistencia y representatividad de los datos utilizados durante el proceso de entrenamiento.

Para el desarrollo de la presente investigación se empleó una base de datos histórica, ver figura 23. Dicha base de datos está conformada por 2123 registros correspondientes a inspecciones técnicas, las cuales fueron obtenidas mediante diferentes técnicas de monitoreo de condición aplicadas a equipos industriales pertenecientes a diversos sistemas operativos,

Los registros recopilados contienen información relacionada con la técnica de inspección utilizada, el sistema al que pertenece, el equipo evaluado, el componente inspeccionado, la condición diagnosticada, así como información complementaria asociada al diagnóstico técnico, recomendaciones, comentarios y estado de atención de cada observación.

Sin embargo, debido a los objetivos planteados en la investigación, no todas las variables disponibles en la base de datos fueron consideradas para la construcción del modelo predictivo. Las variables relacionadas con diagnósticos técnicos, recomendaciones de mantenimiento, comentarios y otros campos de carácter descriptivo fueron excluidas del proceso de modelado, debido a que no forman parte de los atributos utilizados para la clasificación automática de la condición de los equipos.

En consecuencia, se seleccionaron únicamente aquellas variables que permiten caracterizar operativamente cada registro y que presentan relación directa con la variable objetivo de la investigación. La variable de interés corresponde a la condición del equipo, la cual fue clasificada en tres categorías: normal, seguimiento y crítico. Estas categorías representan distintos niveles de riesgo operativo y constituyen la base para el entrenamiento del modelo de clasificación desarrollado en la presente propuesta.

Tabla 27. *Diccionario de datos de las variables utilizadas en el modelo*

Variable	Tipo de dato	Descripción	Función en el modelo
Técnica	Categórica	Técnica de monitoreo de condición aplicada durante la inspección.	Variable predictora
Sistema	Categórica	Sistema industrial al que pertenece el equipo inspeccionado.	Variable predictora
Equipo	Categórica	Equipo evaluado durante la inspección.	Variable predictora
Componente	Categórica	Componente específico del equipo sometido a evaluación.	Variable predictora
Condición	Categórica	Estado operativo diagnosticado durante la inspección. Categorías: Normal, Seguimiento y Crítico.	Variable predictora

Nota: Elaboración propia

Tabla 28. Disponibilidad temporal de las variables utilizadas en el modelo

Variable	Disp. Antes de determinar la condición	Utilizada en el modelo
Técnica	Si	Si
Sistema	Si	Si
Equipo	Si	Si
Componente	Si	Si
Condición	No (variable objetivo)	Si (etiqueta)
Diagnóstico	No	No
Recomendaciones	No	No
Comentarios	No	No
Estatus	No	No

Nota: Elaboración propia

Con el fin de evitar la fuga de información, únicamente se utilizaron como variables predictoras aquellos atributos disponibles antes de la determinación de la condición del equipo, ver figura 28. Variables generadas como resultado de la inspección, tales como el diagnóstico técnico, las recomendaciones del mantenimiento, los comentarios y el estado de atención, fueron excluidas del proceso de entrenamiento debido a que contienen información posterior a la variable objetivo y su utilización habría producido una sobreestimación artificial del desempeño del modelo.

Como puede observarse en la tabla 27, el conjunto de variables seleccionado se encuentra orientado a representar las características operativas del proceso de inspección y la condición de los equipos. Esta información constituye el insumo principal para las etapas posteriores de limpieza, transformación de datos, entrenamiento y evaluación de los modelos de Machine Learning que serán desarrollados más adelante.

Figura 23. Muestra representativa de la base de datos utilizada

RELATIVO	TÉCNICA	SISTEMA	FECHA OBSERVACION	EQUIPO	COMPONENTE	CONDICION	DIAGNÓSTICO	RECOMENDACIONES	COMENTARIOS	ESTATUS
1	ACEITE	Minerales C	6/06/2024	Faja 03 - Huidbay	Reductor	SEGUIMIENTO	Cantidades no típicas de calcio y zinc, elementos	Verificar limpieza de los materiales utilizados para el	No se tiene respuesta	NO ATENDIDO
2	ACEITE	Minerales C	6/06/2024	Faja 03 - Huidbay	Reductor	SEGUIMIENTO	Presencia de calcio. Podría provenir del	Investigar posibles orígenes Identificar-eliminar posibles	No se tiene respuesta	NO ATENDIDO
3	ACEITE	Minerales C	6/06/2024	Faja 03 - Huidbay	Reductor	SEGUIMIENTO	Cantidades no típicas de calcio y zinc, elementos	Verificar limpieza de los materiales utilizados para el	No se tiene respuesta	NO ATENDIDO
4	ACEITE	Minerales F	6/06/2024	Faja 04	Reductor	SEGUIMIENTO	Cantidad no típica de zinc, elemento de aditivo,	Verificar limpieza de los materiales utilizados para el	No se tiene respuesta	NO ATENDIDO
5	ACEITE	Minerales F	6/06/2024	REDUCTOR FAJA 08-1005	REDUCTOR	CRÍTICO	Cantidad no típica de zinc, elemento de aditivo,	Identificar-eliminar posibles accesos de ingreso de calcio./	No se tiene respuesta	ATENDIDO
6	ACEITE	Minerales F	6/06/2024	Faja 07	Reductor	SEGUIMIENTO	Cantidad no típica de zinc, elemento de aditivo,	Verificar limpieza de los materiales utilizados para el	No se tiene respuesta	NO ATENDIDO
7	ACEITE	Minerales F	6/06/2024	Faja 08	Reductor	SEGUIMIENTO	Cantidad no típica de zinc, elemento de aditivo,	Verificar limpieza de los materiales utilizados para el	No se tiene respuesta	ATENDIDO
8	ACEITE	Minerales C	6/06/2024	FAJA 8 - C2	REDUCTOR	SEGUIMIENTO	Presencia de partículas no ferrosas en la	Utilizar técnicas de remoción de contaminantes sólidos	No se tiene respuesta	ATENDIDO
9	ACEITE	Minerales F	12/06/2024	Feeder 03	Reductor Conico	SEGUIMIENTO	Desgaste de hierro en pretensión. Investigar	Verificar limpieza de los materiales utilizados para el	No se tiene respuesta	NO ATENDIDO
10	ACEITE	Minerales F	12/06/2024	Feeder 03	REDUCTOR PLANETARIO	SEGUIMIENTO	Se puede observar la presencia de una partícula	Utilizar técnicas de remoción de contaminantes sólidos	No se tiene respuesta	NO ATENDIDO
11	ACEITE	Minerales F	12/06/2024	Feeder 04	Reductor Conico	SEGUIMIENTO	Se puede observar la presencia de una partícula	Utilizar técnicas de remoción de contaminantes sólidos	No se tiene respuesta	NO ATENDIDO
12	ACEITE	Minerales F	12/06/2024	Feeder 04	REDUCTOR PLANETARIO	SEGUIMIENTO	Se puede observar la presencia de una partícula	Utilizar técnicas de remoción de contaminantes sólidos	No se tiene respuesta	NO ATENDIDO
13	ACEITE	Minerales C	21/06/2024	Faja 01-Huidbay	Reductor	SEGUIMIENTO	Presencia de partículas no ferrosas en la	Utilizar técnicas de remoción de contaminantes sólidos	No se tiene respuesta	NO ATENDIDO
14	ACEITE	Minerales C	21/06/2024	Faja 02 -Huidbay	Reductor	SEGUIMIENTO	Cantidades no típicas de zinc y boro, elementos	Verificar limpieza de los materiales utilizados para el	No se tiene respuesta	NO ATENDIDO
15	ACEITE	Minerales C	21/06/2024	Faja 03 - C2	Reductor	CRÍTICO	Viscosidades muy elevadas, crítico. Se asemeja	Verificar información del nombre y grado del lubricante./	No se tiene respuesta	ATENDIDO
16	ACEITE	Minerales C	21/06/2024	FAJA 10 - C2	REDUCTOR	SEGUIMIENTO	Verificar limpieza de los materiales utilizados	Continuar con envío de muestra para monitoreo	No se tiene respuesta	ATENDIDO
17	ACEITE	Eq Monovizado	3/07/2024	CARGADOR FRONTAL 911	DIFERENCIAL DELANTERO	SEGUIMIENTO	Viscosidad dentro del rango de servicio Ligera	Investigar posibles orígenes Continuar con el envío de	No se tiene respuesta	ATENDIDO
18	ACEITE	Eq Monovizado	3/07/2024	CARGADOR FRONTAL 911	DIFERENCIAL POSTERIOR	SEGUIMIENTO	Viscosidad dentro del rango de servicio Ligera	Investigar posibles orígenes Continuar con el envío de	No se tiene respuesta	ATENDIDO

Nota: Elaboración propia

4.2. Selección del modelo de Machine Learning

Para identificar los modelos de Machine Learning más adecuados para el desarrollo del sistema predictivo propuesto, se realizó una evaluación comparativa de diferentes algoritmos de clasificación comúnmente utilizados en problemas de mantenimiento predictivo y análisis de datos industriales. La selección consideró criterios relacionados con el desempeño predictivo, capacidad para trabajar con datos tabulares, interpretación de resultados, robustez frente a ruido y capacidad para modelar relaciones no lineales entre variables operativas.

Las alternativas evaluadas incluyeron Regresión logística, Support Vector Machine (SVM), Random Forest y XGBoost (Ver tabla 31). La comparación se realizó mediante una matriz de decisión ponderada, considerando las características del conjunto de datos y los requerimientos del problema planteado.

Tabla 29. Matriz de enfrentamiento de criterios para la selección de modelos de Machine Learning

Criterio	1	2	3	4	5	Total
1	-	1	1	1	1	4
2	0	-	1	1	1	3
3	0	0	-	1	1	2
4	0	0	0	-	1	1
5	0	0	0	0	-	0

Nota: Elaboración propia

Tabla 30. Pesos de los criterios para selección de modelos de Machine Learning normalizados

	Criterio	Puntos	Peso	Peso normalizado
1	Desempeño en clasificación multiclase	4	40%	32%
2	Robustez frente a ruido y datos variables	3	30%	24%
3	Manejo de relaciones no lineales	2	20%	18%
4	Adaptabilidad a datos tabulares industriales	1	10%	16%
5	Interpretabilidad y análisis de variables	0	0%	10%

Nota: Elaboración propia

Tabla 31. Criterios de Evaluación

Criterio	Peso	Acumulado
Desempeño en clasificación multiclase	32%	32%
Robustez frente a ruido y datos variables	24%	56%
Manejo de relaciones no lineales	18%	74%
Adaptabilidad a datos tabulares industriales	16%	90%
Interpretabilidad y análisis de variables	10%	100%

Nota: Elaboración propia

Los resultados muestran que el desempeño en clasificación multiclase, la capacidad para modelar relaciones no lineales y la robustez frente a ruido y variabilidad concentran aproximadamente el 75% del impacto sobre el problema de mantenimiento predictivo analizado (Ver tabla 30).

Debido a ello, dichos criterios recibieron una mayor ponderación dentro del proceso de evaluación, ya que representan los factores con mayor influencia sobre la capacidad predictiva y la estabilidad de los modelos en entornos industriales con datos históricos categóricos y variables operativas heterogéneas.

Tabla 32. Matriz de Evaluación de Modelos de Machine Learning

Criterio	Peso	Reg. Logística	SVM	Random Forest	XGBoost
Desempeño en clasificación multiclase	0.32	3	4	5	5
Robustez frente a ruido y datos variables	0.24	2	4	5	5
Manejo de relaciones no lineales	0.18	3	4	5	4
Adaptabilidad a datos tabulares industriales	0.16	4	3	4	4
Interpretabilidad y análisis variables	0.10	3	4	5	5

Nota: Elaboración Propia

Tabla 33. Resultado ponderado de Evaluación de Modelos

Modelo	Puntaje Final
Regresión Logística	2.92
SVM	3.84
Random Forest	4.84
XGBoost	4.66

Nota: Elaboración propia

Los resultados obtenidos en la tabla 32 evidencian que cada uno de los algoritmos analizados presentan fortalezas y limitaciones en función de los criterios considerados. La Regresión Logística destaca por su simplicidad, facilidad de interpretación y bajo costo computacional; sin embargo, presenta limitaciones para modelar relaciones complejas entre las variables. Por su parte, las Máquinas de Vectores de Soporte (SVM) ofrecen una adecuada capacidad de clasificación y generalización, aunque requieren un mayor ajuste de parámetros y presentan una menor interpretabilidad en comparación con otros métodos.

En contraste, los modelos basados en técnicas Ensemble, particularmente Random Forest y XGBoost, obtienen las mejores valoraciones globales debido a su capacidad para manejar relaciones no lineales, procesar grandes volúmenes de información y generar modelos robustos frente a la variabilidad de los datos. Estas características resultan especialmente relevantes en aplicaciones de mantenimiento industrial, donde los patrones asociados al comportamiento de los equipos suelen representar un alto nivel de complejidad.

Si bien ambos algoritmos muestran un desempeño competitivo en los criterios evaluados, se observa que los modelos basados en árboles de decisión ofrecen ventajas adicionales relacionadas con la estabilidad de las predicciones, la capacidad de interpretación de los resultados y la generación de probabilidades de clasificación útiles para la toma de decisiones. Por esta razón, constituyen alternativas especialmente adecuadas para el desarrollo de sistemas de apoyo a la gestión del mantenimiento basados en inteligencia artificial.

En consecuencia, considerando los objetivos de la investigación y los resultados obtenidos en la evaluación comparativa, los modelos Ensemble representan las

alternativas más prometedoras para el problema de clasificación planteado, justificando su análisis más detallado en etapas posteriores.

4.3. Modelo Predictivo de Fallas

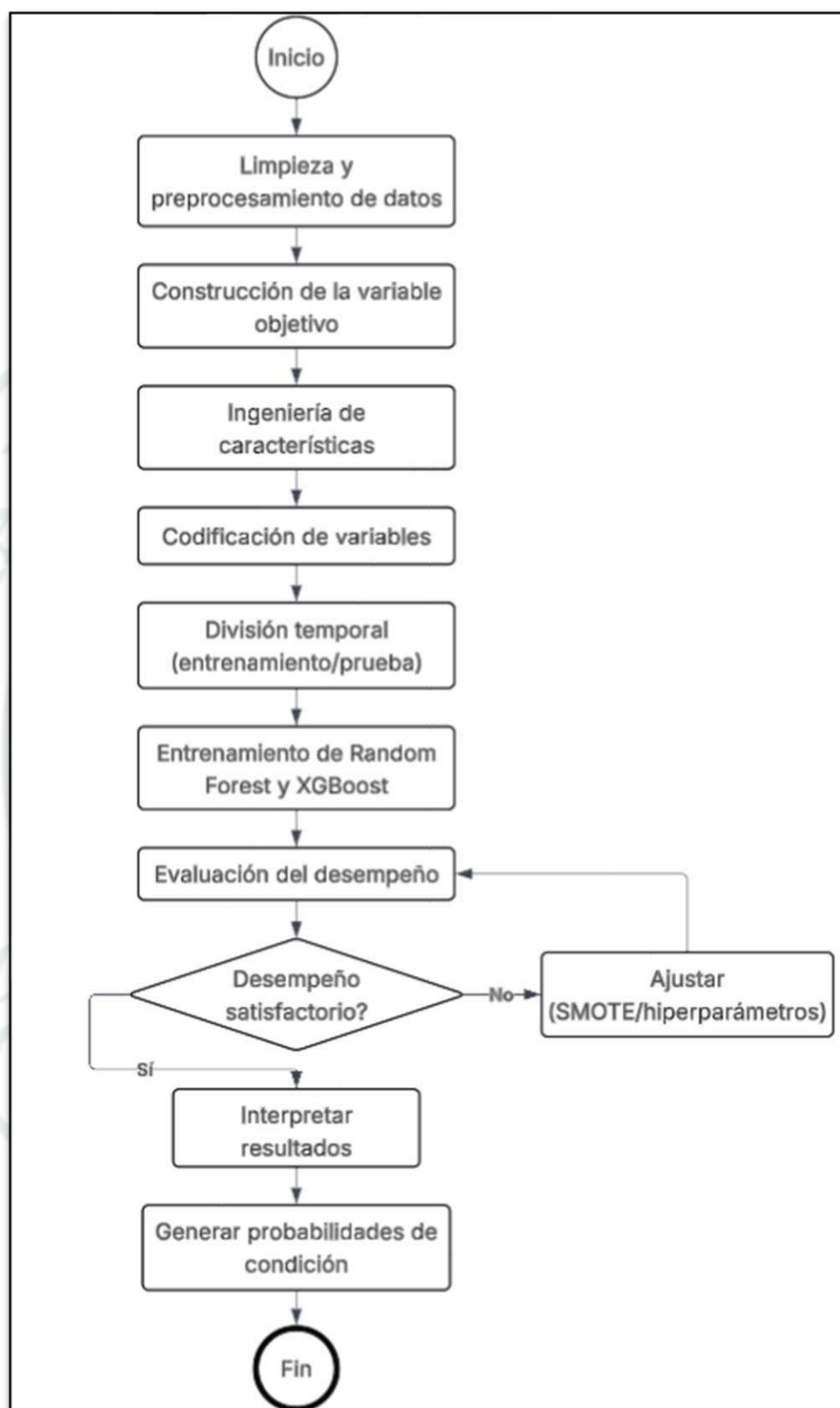
4.3.1. Introducción del Modelo

Con el propósito de optimizar el proceso de mantenimiento industrial mediante el uso de herramientas de Inteligencia artificial, se propone el desarrollo de un modelo de Machine Learning orientado a la predicción de la condición futura de los equipos industriales, a partir del análisis de datos históricos de predicción.

El desarrollo del modelo se estructura en una secuencia metodológica que permite transformar datos operativos en información predictiva útil para la toma de decisiones en mantenimiento. Esta secuencia comprende desde la preparación inicial de los datos hasta la obtención de resultados interpretables que sirvan de base para la optimización del proceso.

Bajo ese concepto, la metodología se organiza de la siguiente manera en la figura 24:

Figura 24. Metodología general del Modelo



Nota: Elaboración propia

La figura 23 presenta la metodología general empleada para el desarrollo del modelo propuesto. Como parte de este proceso, la división del conjunto de datos en entrenamiento y prueba precede al entrenamiento de los algoritmos de Machine Learning. Posteriormente, cuando se identificó desbalance entre las clases del conjunto de entrenamiento, se aplicó la técnica SMOTE exclusivamente sobre dicho conjunto, seguida del ajuste de hiperparámetros y la reevaluación de los modelos. Este procedimiento permitió preservar la independencia del conjunto de prueba y obtener una estimación objetiva del desempeño de los modelos.

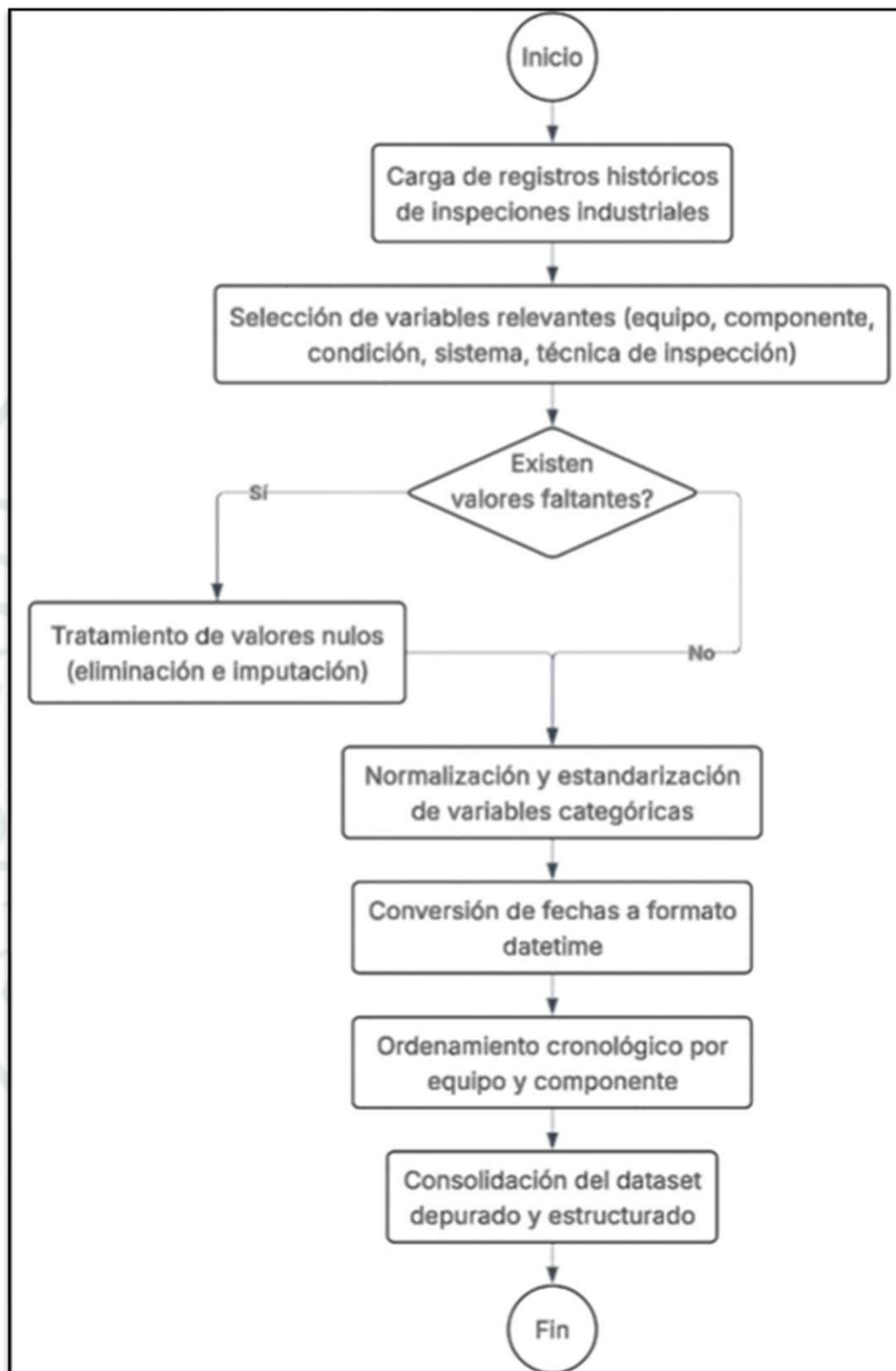
4.3.2. Fase 1: Limpieza y preprocesamiento de datos

El desarrollo del modelo, resumido en la figura 25, inició con la preparación del conjunto de la base de datos, la cual consolida registros históricos de inspecciones industriales, incluyendo variables categóricas, temporales y descriptivas sobre el estado operativo de los equipos. Debido a que la calidad de los modelos de Machine Learning depende de la consistencia de los insumos, se ejecutó un proceso de limpieza orientado a eliminar el ruido o asegurar la idoneidad de la información. En primer lugar, se realizó una selección de variables con potencial valor predictivo mediante un enfoque de selección de características basado en dominio; bajo este criterio, se conservaron dimensiones operativas como la técnica de inspección, el sistema, el equipo, el componente y la condición registrada, mientras que se excluyeron identificadores y campos de texto no estructurado – como diagnósticos y recomendaciones – por su limitada contribución analítica en esta etapa.

Posteriormente, se abordó el tratamiento de valores faltantes mediante estrategias diferenciadas según la relevancia de cada variable. Para la variable objetivo, denominada “Condición”, se aplicó la técnica de eliminación de casos incompletos (listwise deletion) con el fin de garantizar la integridad de la variable dependiente y evitar sesgos en el entrenamiento. En contraste, para las variables categóricas restantes, los nulos se imputaron bajo la categoría “Desconocido”, lo que permitió preservar la representatividad del conjunto sin introducir supuestos artificiales. A la par de este tratamiento, se procedió con la estandarización de las variables categóricas mediante la normalización de texto, convirtiendo los valores a mayúsculas y unificando etiquetas semánticamente idénticas para reducir la redundancia y fragmentación de los datos.

Finalmente, la variable “Fecha de observación” se transformó al tipo de dato datetime, permitiendo una correcta validación cronológica y el análisis de la evolución temporal de los activos. Con esta base tipificada, los datos fueron estructurados siguiendo técnicas de series temporales, ordenando los registros de manera ascendente por equipo, componente y fecha. Este procedimiento de ordenamiento cronológico resultó fundamentalmente para reconstruir la trayectoria histórica de cada componente y facilitar la posterior identificación de patrones dinámicos. Como resultado de esta fase, se consolidó un conjunto de datos depurado y estructurado, estableciendo una base sólida para la construcción de la variable objetivo y la generación de características en las etapas subsecuentes de la investigación.

Figura 25. Flujo de limpieza de preprocesamiento de datos



Nota: Elaboración propia

4.3.3. Fase 2: Construcción de la variable objetivo

Tras la limpieza y estructuración de los datos, se procedió a la definición de la variable objetivo, adoptando un enfoque de predicción futura en lugar de una clasificación estática tradicional. En este contexto, la variable se define como la condición del equipo en la inspección subsiguiente, categorizada en tres niveles: Normal, Seguimiento y Crítico. Matemáticamente, el modelo busca estimar la probabilidad de que un activo se encuentre en un estado determinado en el tiempo $t + 1$, dadas sus características en el tiempo t . Este planteamiento permite capturar la dinámica de cambio y la evolución operativa de los componentes industriales entre observaciones sucesivas.

Para operativizar este enfoque, se aplicó una técnica de desplazamiento temporal (time shifting) basada en el análisis de series temporales discretas. El proceso consistió en agrupar los registros por la combinación equipo-componente y generar una nueva variable denominada “condición siguiente”, la cual vincula cada observación con el estado registrado en la inspección inmediatamente posterior. De este modo, el sistema aprende los patrones de transición entre estados, transformando el historial de inspecciones en un problema de aprendizaje supervisado orientados a la anticipación de escenarios futuros.

Como consecuencia natural del desplazamiento temporal, en los últimos registros de cada serie cronológica (es decir, la inspección más reciente de cada equipo) no cuentan con una observación posterior que actúe como etiqueta. Por tanto, se procedió a la depuración de estos registros para garantizar la coherencia del conjunto de entrenamiento y asegurar que todas las instancias posean una variable objetivo válida. Esta estrategia metodológica no solo dota al modelo de una capacidad de predicción dinámica sin requerir variables físicas continuas, sino que también se alinea con el objetivo de proporcionar información valiosa para la toma de decisiones preventivas en el mantenimiento industrial. El resultado de esta fase es un conjunto de datos robusto donde cada registro asocia el estado actual (t) con su desenlace futuro ($t+1$), permitiendo abordar el modelado como una tarea de clasificación supervisada multiclase.

4.3.4. Fase 3: Ingeniería de características (Feature Engineering)

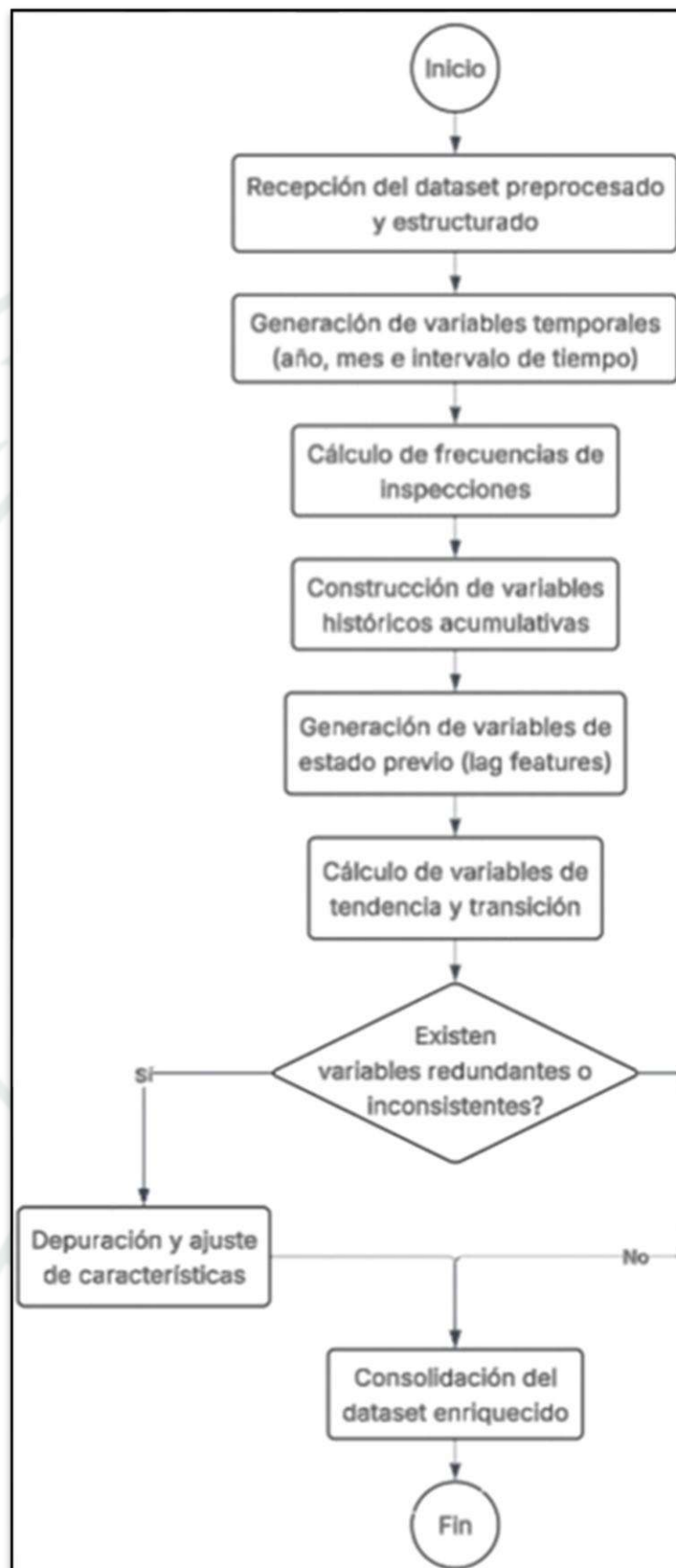
Tras la definición de la variable objetivo, se procedió a la ingeniería de característica, una etapa crucial para transformar los datos brutos en variables informativas que

incrementen la capacidad predictiva del modelo, ver resumen en figura 26. Dado que el mantenimiento industrial en este estudio se basa en registros categóricos y no en mediciones físicas continuas, fue imperativo construir indicadores que capturaran el comportamiento histórico y la evolución temporal de los activos. En primera instancia, se generaron variables temporales mediante la descomposición de la fecha de observación en componentes de año y mes, permitiendo identificar patrones estacionales o tendencias en el régimen de fallas. Asimismo, se calculó el intervalo en días entre inspecciones sucesivas para incorporar información sobre la frecuencia de monitoreo y la dinámica operativa del proceso.

Para dotar al modelo de una perspectiva longitudinal, se construyeron variables históricas acumulativas mediante técnicas de agregación secuencial. Para cada combinación de equipo y componente, se computó el número total de inspecciones previas y la frecuencia acumulada de cada estado (Normal, Seguimiento y Crítico) hasta el momento de la observación. Este procedimiento se realizó bajo un enfoque de agregación acumulativa (cumulative aggregation), asegurando que cada registro contenga información del pasado sin incurrir en la fuga de datos (data leakage) al omitir información futura. Complementariamente, integraron variables de estado previo (lag features), asignando a cada registro la condición observada en la inspección inmediata anterior. Esta inclusión es fundamental para que el modelo identifique trayectorias de transición, como la probabilidad de que un equipo en estado de “seguimiento” evolucione hacia una condición “Crítica”.

Finalmente, se desarrollaron variables de tendencia y frecuencia para capturar la dirección del cambio en el estado de los activos. Al asignar valores ordinales a las categorías de condición, se calculó la diferencia entre el estado actual y el previo, facilitando la detección de patrones de deterioro progresivo o estabilidad. Adicionalmente, se obtuvieron indicadores del tiempo promedio entre inspecciones, permitiendo diferenciar entre equipos con monitoreo intensivo y aquellos con intervalos prolongados. El resultado de esta fase fue un conjunto de datos enriquecido con dimensiones temporales, históricas y de comportamiento, proporcionando una base robusta para que el algoritmo de Machine Learning identifique patrones complejos y mejore la precisión en la determinación de la condición futura de los equipos industriales.

Figura 26. *Proceso de Ingeniería de Características*



Nota: Elaboración propia

4.3.5. Fase 4: Codificación de variables

Una vez consolidada las características del modelo, se procedió a la transformación de las variables categóricas en representaciones numéricas, paso indispensable para que los algoritmos de Machine Learning puedan ejecutar operaciones matemáticas y establecer una alta densidad de información cualitativa – como la técnica de inspección, el sistema, el equipo, el componente y el estatus – la cual requirió un tratamiento específico para preservar su valor semántico sin introducir sesgos jerárquicos artificiales.

Para las variables de naturaleza nominal, se empleó la técnica de codificación one-hot (one-hot encoding), mediante la cual cada categoría se transforma en una variable binaria independiente. Este enfoque se seleccionó para evitar que el modelo interpretara erróneamente un orden o una distancia numérica entre entidades que no poseen una jerarquía natural, como es el caso de lo nombre de los equipos o sistemas. Si bien esta técnica incrementa la dimensionalidad del conjunto de datos, la capacidad computacional de los algoritmos seleccionados (basados principalmente en árboles de decisión y métodos de boosting) permite gestionar este volumen de columnas sin comprometer el rendimiento. por su parte, la variable objetivo fue sometida a una codificación ordinal controlada, asignando valores enteros a las categorías “Normal”, “Seguimiento” y “Crítico”, facilitando así el procesamiento eficiente de la clasificación supervisada.

Finalmente, es importante señalar que, debido a la predominancia de variables categóricas y conteos discretos, no se requirieron técnicas de escalamiento o normalización (como Min-Max Scaling), dado que los modelos de ensamble basados en árboles son invariantes a la escala de los predictores. Como resultado de esta fase, se obtuvo un conjunto de datos íntegramente numérico y estructurado, listo para la fase de entrenamiento. Esta transformación asegura que la información original del sistema y las variables derivadas sean interpretadas correctivamente por los algoritmos garantizando la robustez del proceso de identificación de patrones complejos en el comportamiento de los activos.

4.3.6. Fase 5: División del conjunto de datos

Una vez preprocesada y codificada la información, se procedió a la partición del conjunto de datos en subgrupos de entrenamiento y prueba. Esta etapa es fundamental para estimar con objetividad la capacidad de generalización del modelo ante datos no

observados previamente, asegurando que los patrones aprendidos en el historial técnico sean aplicables a nuevos escenarios de mantenimiento industrial. A diferencia de los enfoques de validación cruzada aleatoria, en esta investigación se adoptó una estrategia basada en el tiempo (time-based split), la cual es más coherente con la naturaleza secuencial de los registros de inspección.

Bajo este enfoque cronológico, se destinó el 70% de los registros más antiguos al conjunto de entrenamiento, utilizado para el ajuste de los parámetros del modelo, mientras que el 30% restante, correspondiente a las observaciones más recientes, se reservó como conjunto de prueba para la evaluación del desempeño. Esta metodología permite simular con precisión un entorno operativo real, donde el sistema se entrena con el histórico acumulado y se valida proyectando predicciones hacia eventos futuros. La elección de una partición temporal, en lugar de una aleatoria, es una medida crítica para mitigar la fuga de información (data leakage), evitando que el modelo tenga acceso implícito a eventos futuros durante su fase de aprendizaje y previniendo así una sobreestimación artificial de su precisión.

Finalmente, durante el proceso de división se verificó la distribución de la variable objetivo para garantizar que las clases “Normal”, “Seguimiento” y “Crítico” estuvieran representadas de manera proporcional en ambos subconjuntos. Asimismo, se aseguró la integridad de las series temporales por equipo, evitando fragmentaciones inconsistentes que pudieran alterar la lógica secuencias de las observaciones. Como resultado, se consolidaron dos subconjuntos claramente para el entrenamiento y la validación de los modelos predictivos en las fases subsecuentes.

4.3.7. Fase 6: Entrenamiento de modelos de Machine Learning

Tras la segmentación de los datos, se procedió el entrenamiento de los modelos de Machine Learning con el objetivo de identificar patrones complejos que permitan predecir la condición futura de los activos, ver el resumen en la figura 27. El problema se abordó como una tarea de clasificación supervisada multiclase, donde el sistema debe asignar cada instancia a una de las tres categorías predefinidas: Normal, Seguimiento o Crítico. Para ello, se seleccionaron algoritmos basados en ensambles de árboles de decisión, dada su robustez para manejar variables categóricas, capturar relaciones no lineales y ofrecer un nivel superior de interpretabilidad en entornos industriales.

Específicamente, se implementaron dos arquitecturas principales: Random Forest y XGBoost. El primero se fundamenta en la construcción de múltiples árboles independientes aparte de subconjuntos aleatorios, combinando sus resultados para reducir la varianza y el sobreajuste, ver figura 28. Por su parte, XGBoost emplea un enfoque de potenciación de gradiente (gradient boosting), entrenando modelos de forma secuencial para corregir los errores de las iteraciones previas, lo que optimiza el rendimiento predictivo en conjuntos de datos estructurados, ver figura 29. Durante este proceso, los algoritmos ajustaron sus parámetros internos sobre el conjunto de entrenamiento para minimizar el error de clasificación, estableciendo inicialmente configuraciones estándar para definir una línea base de desempeño (baseline).

Desde el punto de vista metodológico, se mantuvo un control estricto para evitar el sobreajuste (overfitting), asegurando que los modelos conserven su capacidad de generalización ante datos no observados. Asimismo, se garantizó la integridad del flujo de información utilizando exclusivamente predictores disponibles el tiempo t , evitando cualquier sesgo por incorporación de datos futuros. Como resultado de esta fase, se obtuvieron modelos entrenados capaces de generar tanto etiquetas de clasificación como probabilidades asociadas a cada estado, proporcionando la base necesaria para la evaluación de desempeño y la selección final del modelo con mayor capacidad de anticipación de fallas, como se ve en la figura 30.

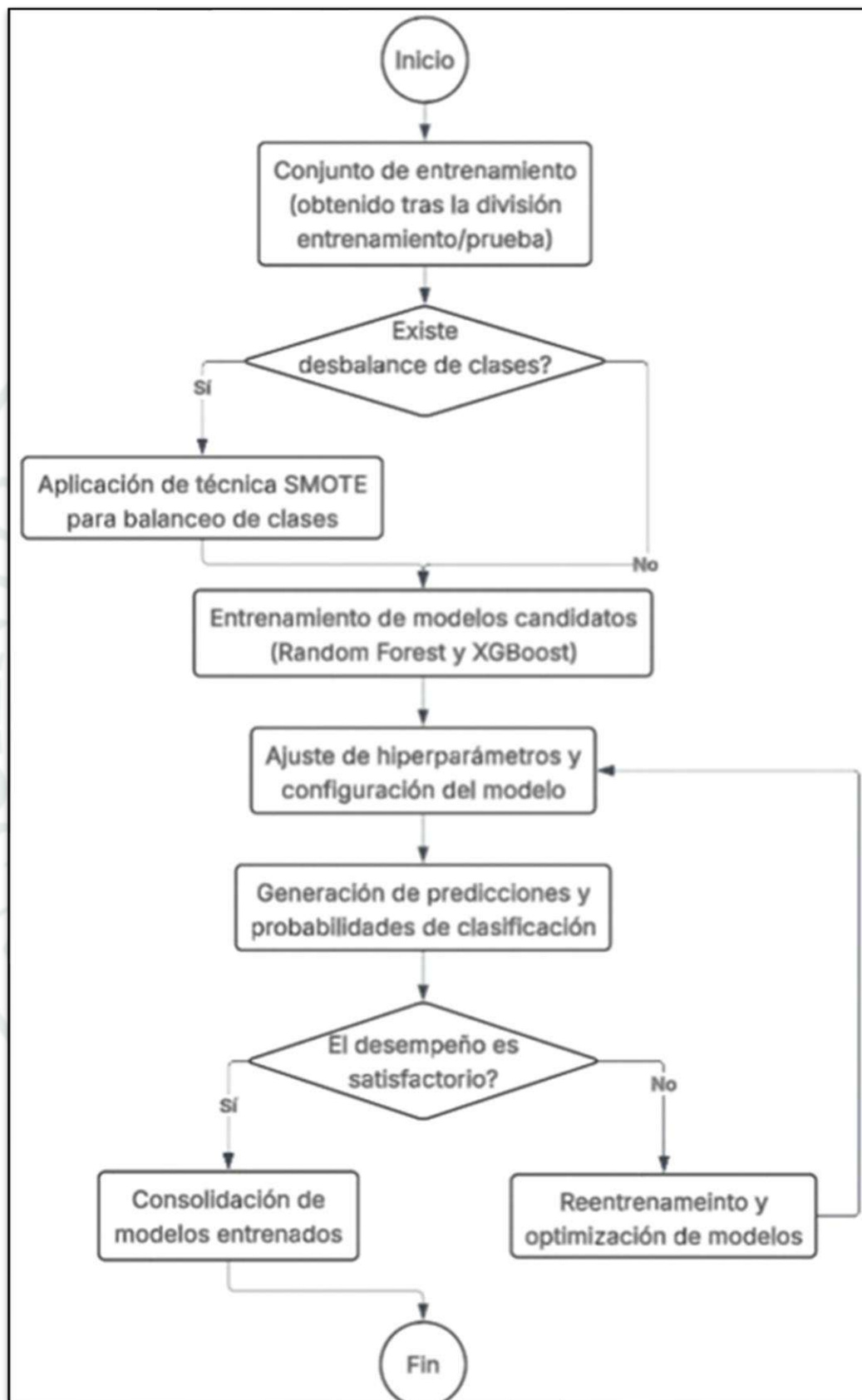
Tras el entrenamiento inicial, se identificó un desbalance de clases significativo en la variable objetivo, donde la predominancia de la condición “Normal” generaba un sesgo en el aprendizaje, dificultando la identificación de estados críticos. Para mitigar este efecto, se implementó la técnica de sobre muestreo sintético SMOTE (Synthetic Minority Over-sampling Technique) aplicada exclusivamente al conjunto de entrenamiento. Esta estrategia permitió preservar la independencia del conjunto de prueba, evitando que las muestras sintéticas influyeran en la evaluación final del modelo y garantizando una estimación más confiable de su capacidad de generalización, ver figura 31 y 32. Este procedimiento se complementó con un ajuste de hiperparámetros en el algoritmo de Random Forest – optimizando la profundidad de los árboles y el número de estimadores – e incorporando un esquema de ponderación de clases (class weights) que penaliza con mayor rigor los errores en la detección de condiciones críticas.

Los resultados de esta optimización reflejaron una mejora sustancial en la métrica de sensibilidad (recall) para la clase “Crítico”, lo cual es prioritario en el mantenimiento

industrial para garantizar la detección temprana de fallas. Si bien el F1-score macro mostró un equilibrio superior entre categorías, se observó una ligera disminución en la precisión de la clase mayoritaria, un trade-off esperado al desestimar el sesgo hacia la clase dominante. No obstante, el análisis de la matriz de confusión reveló persistencia en la dificultad para diferenciar estados de “Seguimiento” y “Crítico”, comportamiento atribuido a la naturaleza categórica de los datos y a la ausencia de variables físicas continuas (como vibración o temperatura) que permitan una delimitación más nítida de la degradación.

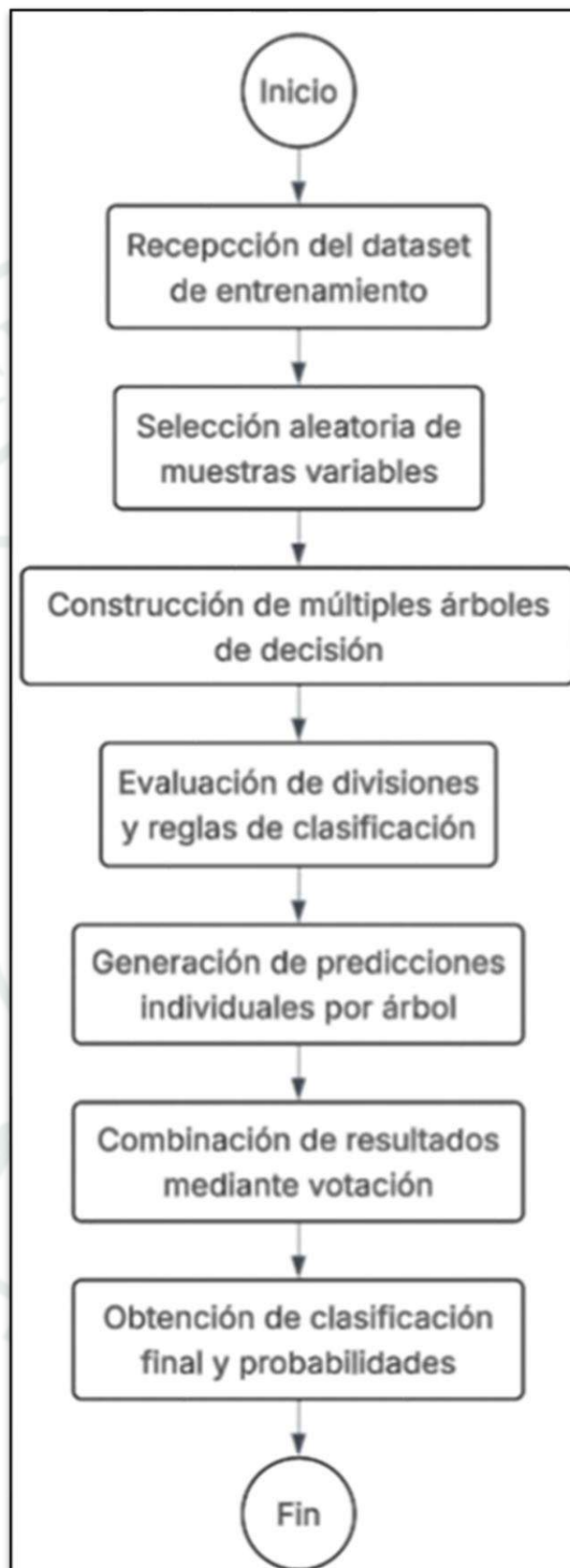
A pesar de estas limitaciones, inherentes a la subjetividad de las inspecciones técnicas y a la resolución de los datos disponibles, se determinó que el desempeño alcanzado es adecuado para los fines de esta investigación. Se descartó una optimización exhaustiva adicional, dado que el objetivo primordial no es la maximización aislada de métricas de precisión, sino la obtención de un insumo predictivo estable e interpretable. De este modo, el modelo cumple su función crítica: generar probabilidades de riesgo que alimenten el posterior modelo de optimización de mantenimiento. En esta etapa final, la probabilidad de falla se integra con estructuras de costos e impacto operativo, transformando una salida algorítmica en una herramienta estratégica de ingeniería industrial para la priorización eficiente de recursos.

Figura 27. *Proceso de entrenamiento de Modelos de Machine Learning*



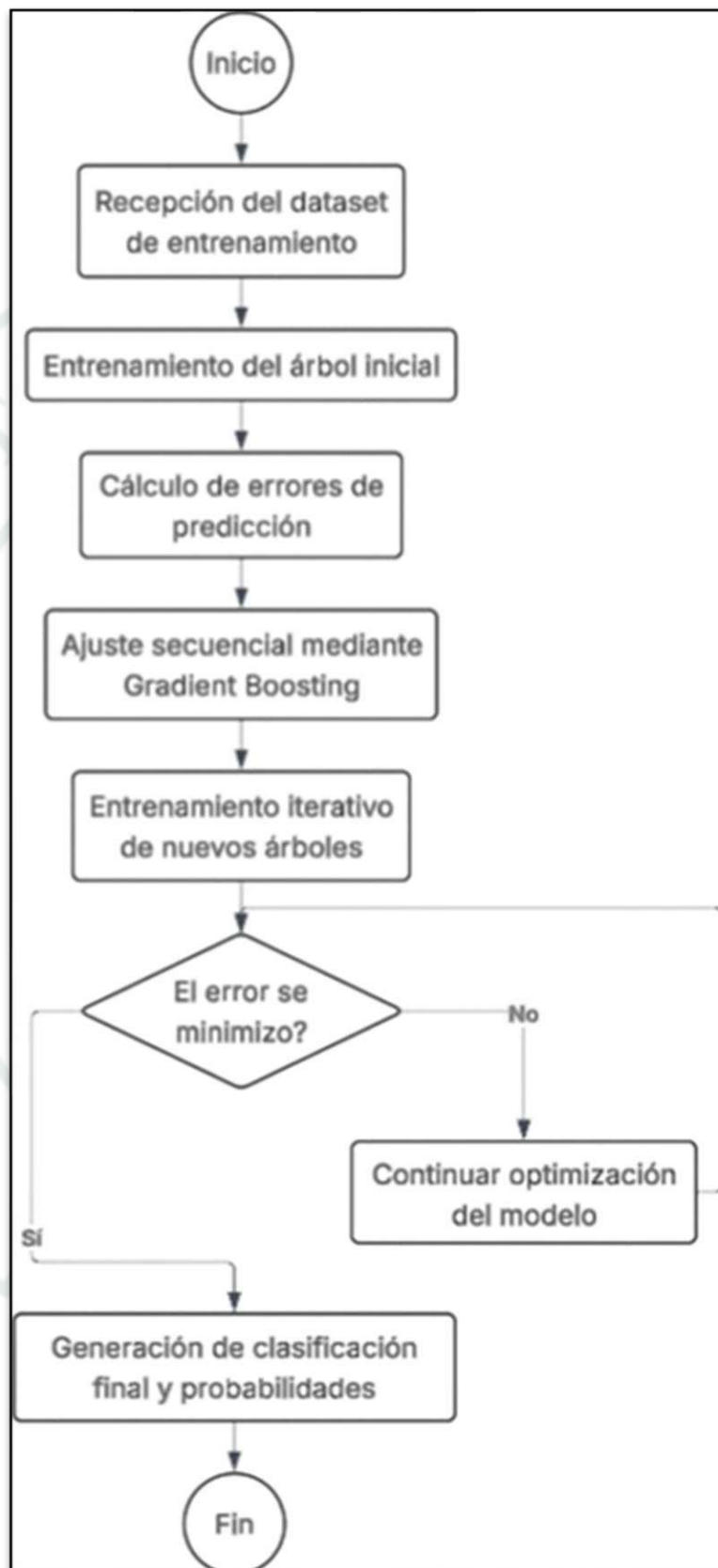
Nota: Elaboración propia

Figura 28. *Aplicación del Modelo Random Forest*



Nota: Elaboración propia

Figura 29. Aplicación del Modelo XGBoost



Nota: Elaboración propia

Figura 30. Resultados del modelo aplicando Random Forest y XGBoost

Random Forest					
	precision	recall	f1-score	support	
0	0.39	0.47	0.43	19	
1	0.79	0.77	0.78	195	
2	0.35	0.36	0.36	64	
accuracy			0.66	278	
macro avg	0.51	0.54	0.52	278	
weighted avg	0.67	0.66	0.66	278	
XGBoost					
	precision	recall	f1-score	support	
0	0.28	0.42	0.33	19	
1	0.77	0.88	0.82	195	
2	0.42	0.17	0.24	64	
accuracy			0.69	278	
macro avg	0.49	0.49	0.47	278	
weighted avg	0.66	0.69	0.66	278	
Matriz de confusión					
[[9 4 6]					
[8 151 36]					
[6 35 23]]					

Nota: Elaboración propia, a partir de la implementación en Python (Random Forest y XGBoost)

Figura 31. Resultados del modelo de Random Forest Optimizado con SMOTE

Random Forest Optimizado					
	precision	recall	f1-score	support	
0	0.32	0.53	0.40	19	
1	0.87	0.71	0.78	195	
2	0.41	0.56	0.48	64	
accuracy			0.67	278	
macro avg	0.54	0.60	0.55	278	
weighted avg	0.73	0.67	0.69	278	
Matriz de confusión					
[[10 3 6]					
[11 139 45]					
[10 18 36]]					

Nota: Elaboración propia, a partir de la implementación en Python optimizado por la técnica SMOTE

Figura 32. Resultados del modelo de XGBoost Optimizado con SMOTE

	precision	recall	f1-score	support
0	0.26	0.37	0.30	19
1	0.78	0.84	0.81	195
2	0.36	0.23	0.28	64
accuracy			0.67	278
macro avg	0.47	0.48	0.47	278
weighted avg	0.65	0.67	0.66	278

Nota: Elaboración propia, a partir de la implementación en Python optimizado por la técnica SMOTE

4.3.8. Fase 7: Evaluación del modelo

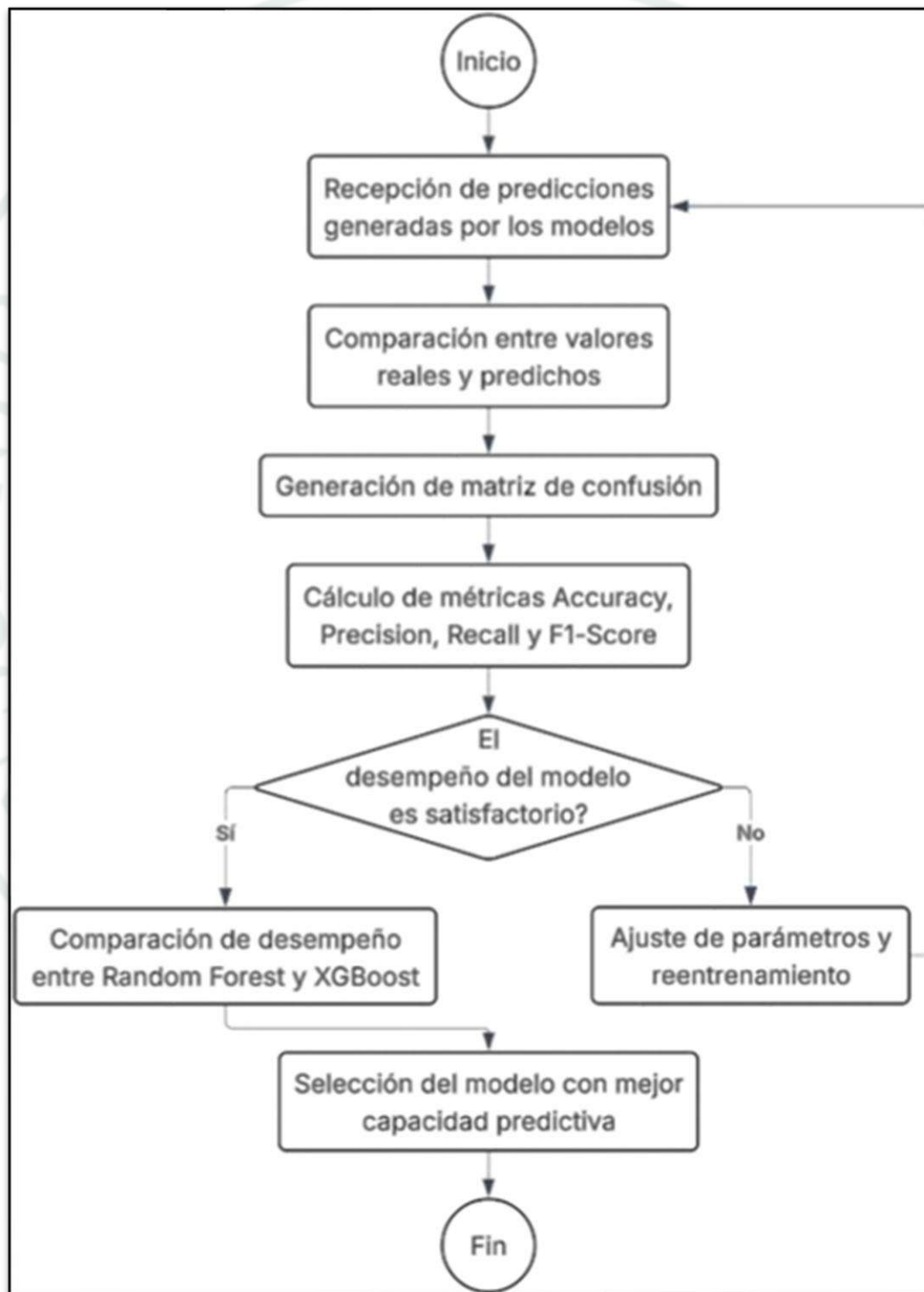
La fase final de la metodología consistió en la evaluación del desempeño de los modelos entrenados, con el objetivo de cuantificar su capacidad para predecir la condición futura de los equipos sobre el conjunto de datos de prueba, como se puede ver en la figura 33. Dado que el problema se define como una clasificación multiclase, se empleó un marco de evaluación integral para validar la utilidad del modelo en un entorno de mantenimiento real. Más allá de la precisión global (accuracy), se incorporaron métricas robustas como precisión, sensibilidad (recall) y puntuación F1 (F1-score) para cada categoría. En particular, se priorizó la optimización del recall en la clase “crítico”, ya que en el contexto industrial es imperativo minimizar los fallos negativos para evitar fallas catastróficas no detectadas por el sistema.

Para un análisis detallado de los errores de clasificación, se utilizó la matriz de confusión como herramienta diagnóstica. Esta permitió visualizar las interacciones entre las etiquetas reales y las predicciones, facilitando la identificación de posibles confusiones entre estados adyacentes (como "Normal" y "Seguimiento"). Este nivel de detalle es fundamental para comprender el comportamiento del modelo ante el desbalance de clases intrínseco a los registros de inspección, donde las condiciones críticas suelen ser menos frecuentes, pero de mayor impacto operativo.

Finalmente, se realizó una evaluación comparativa entre las arquitecturas de Random Forest y XGBoost bajo las mismas condiciones y métricas. Este procedimiento permitió una selección objetiva basada en el equilibrio entre la precisión general y la eficacia en

la detección de estado de alto riesgo. Como resultado de esta fase, se determinó el modelo con la mejor capacidad de generalización y robustez predictiva, consolidando la solución tecnológica que servirá de base para las conclusiones y aplicaciones prácticas de la presente investigación.

Figura 33. *Evaluación del Desempeño Predictivo*



Nota: Elaboración propia

4.3.9. Fase 8: Interpretación de resultados

Más allá de la capacidad predictiva, la utilidad de un modelo en el mantenimiento industrial depende de su interpretabilidad, factor que permite a los tomadores de decisiones comprender las causas subyacentes a una predicción de falla. Por ello, esta fase se centró en desglosar la “caja negra” del algoritmo para identificar los factores críticos que influyen en la condición futura de los activos. Para cuantificar este impacto, se empleó la técnica de importancia de las características (feature importance), la cual mide la contribución relativa de cada variable en el proceso de clasificación basándose en la reducción de impureza o la ganancia de información dentro de las estructuras de los árboles de decisión.

El análisis de los resultados permitió jerarquizar las variables con mayor peso en el modelo, destacando la relevancia de las características históricas acumulativas y las tendencias temporales frente a los datos estáticos. Esta jerarquización no solo proporciona una validación técnica, sino que también ofrece una validación de dominio, al contrastar si los factores identificados por el algoritmo son coherentes con el conocimiento experto y las leyes físicas del deterioro industrial. La convergencia entre los hallazgos del modelo y la experiencia en campo refuerza la fiabilidad del sistema propuesto.

Finalmente, esta etapa de interpretación transformó las métricas abstractas en información estratégica. Al identificar las variables clave que disparan las alertas de condición “Crítica”, el modelo deja de ser una herramienta meramente descriptiva para convertirse en un soporte de decisión que permite formular estrategias de mantenimiento preventivo más precisas. Como resultado, se consolidó un conjunto de hallazgos que explican el comportamiento del modelo, cerrando el ciclo metodológico con una solución que es, a la vez, robusta en su ejecución y transparente en su lógica.

4.3.10. Fase 9: Generación de probabilidades de clasificación

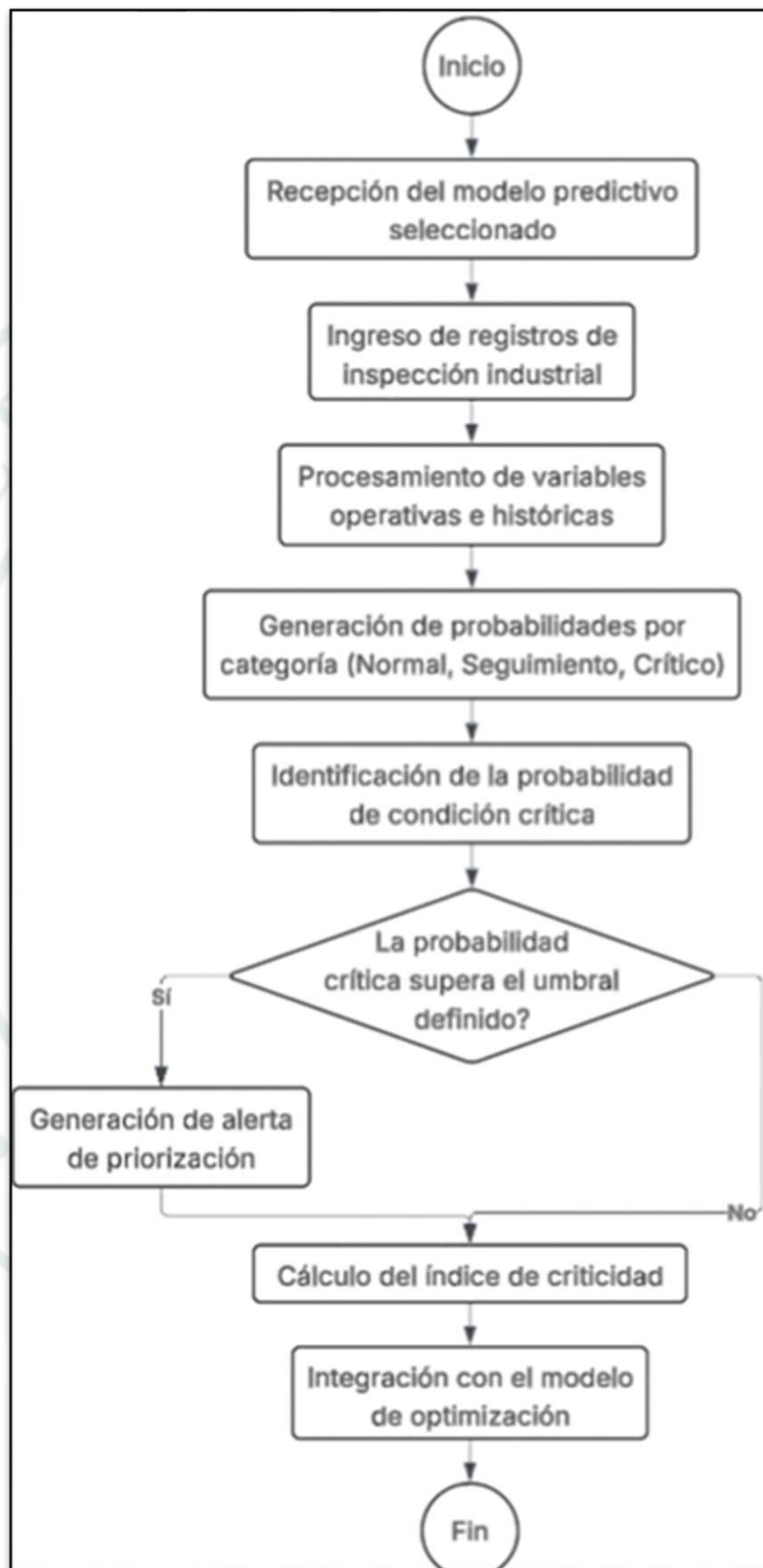
Una vez seleccionado el modelo con el desempeño predictivo óptimo, la fase final de la metodología se centró en la extracción de las probabilidades de la clasificación asociadas a cada categoría de la variable objetivo, se puede ver el resumen en la figura 34. A diferencia de una clasificación discreta, este enfoque permite cuantificar el nivel de confianza del algoritmo para cada posible estado futuro. Para cada registro de inspección, el modelo genera un vector de probabilidad que distribuye la certeza entre

las condiciones “Normal”, “Seguimiento” y “Crítico”, proporcionando así una visión estocástica de la salud del activo en el tiempo $t+1$.

La utilidad técnica de estas probabilidades radica en su capacidad para servir como un indicador de riesgo dinámico. En particular, la probabilidad asignada a la condición “Crítica” se convierte en una métrica de alerta temprana, permitiendo a la gestión de mantenimiento priorizar intervenciones en aquellos equipos donde la incertidumbre de fallas es mayor, incluso si el estado actual parece estable. Este matiz probabilístico es fundamental para transitar de un mantenimiento reactivo a uno basado en el riesgo, optimizando la asignación de recursos técnicos y económicos.

Finalmente, los resultados obtenidos en esta fase constituyen el insumo para el posterior modelo de optimización de mantenimiento. En dicha etapa, estas probabilidades se integran con variables de impacto operativo y estructura de costos para definir la estrategia de intervención más costo-efectiva. Con la culminación de esta fase, el proceso de transformación de datos históricos en información estratégica queda completo, entregando un modelo predictivo robusto que no solo estima estados futuros, sino que fundamenta matemáticamente la toma de decisiones en el entorno industrial.

Figura 34. *Generación de Probabilidades*



Nota: Elaboración Propia

4.4. Modelo de Priorización y Optimización

Como continuación natural del componente predictivo, el segundo nivel de la investigación plantea un modelo de optimización orientado a la selección eficiente de activos para intervención. El propósito fundamental es transformar las probabilidades de falla generadas por el Machine Learning en decisiones concretas de mantenimiento, permitiendo maximizar el impacto operativo bajo restricciones de recursos limitados. Para ello, se adoptó un enfoque de programación dinámica, modelando el problema como una variante del problema clásico de la mochila (Knapsack Problem), en el cual cada equipo es evaluado simultáneamente por su probabilidad de transición a un estado crítico, el beneficio esperado de su reparación y el tiempo requerido para la ejecución de las tareas.

La formulación del problema se estructuró mediante una variable de decisión binaria ($x_i \in \{0,1\}$), que determina si un equipo es seleccionado para mantenimiento dentro de un periodo dado. La función objetivo del modelo busca maximizar el beneficio total esperado, definido como la reducción del riesgo operativo al intervenir oportunamente un componente antes de que ocurra una falla crítica. Esta meta se encuentra sujeta a una restricción de capacidad operativa de 40 horas hombre, reflejando una limitación realista de personal y tiempo en el entorno industrial. En este contexto, el problema se modela como un problema de mochila 0-1, donde cada equipo representa un elemento con un “peso” (tiempo de intervención) y un “valor” (beneficio), siendo resuelto mediante programación dinámica a través de la construcción de una tabla de estados que permite identificar la combinación óptima de equipos.

La fortaleza de este enfoque radica en la integración sistemática de variables heterogéneas. Por un lado, se capturan las salidas probabilísticas del modelo predictivo (probabilidad de falla) y, por otro, se incorporan parámetros de terministas como tiempos de intervención y costos asociados. Esta integración permite construir una métrica de beneficio de intervención, lo cual cuantifica el valor esperado de actuar sobre un activo. Al resolver el modelo mediante un método exacto basado en programación dinámica, se garantiza la obtención de una solución óptima global dentro del espacio de soluciones factibles, asegurando que el subconjunto de equipos seleccionados aproveche al máximo las 40 horas disponibles y maximice el beneficio total.

La solución resultante representa una priorización cuantitativa que trasciende los criterios subjetivos tradicionales. El modelo es capaz de identificar, por ejemplo, que un equipo con alta probabilidad de falla, pero con un tiempo de intervención elevado, puede ser desplazado por un conjunto de equipos con menor criticidad individual, pero mayor beneficio acumulado en el mismo horizonte temporal. Esta capacidad de balancear múltiples factores posiciona a la propuesta como una herramienta de mantenimiento prescriptivo alineada con los principios de la industria 4.0.

En última instancia, este modelo cierra el flujo de valor de la investigación, transformando datos históricos en acciones estratégicas que optimizan el uso de recursos y contribuyen a la mejora de la continuidad operativa.

4.5. Evaluación Comparativa y validación de la propuesta

Con el objetivo de validar la efectividad del sistema desarrollado, se realizó una evaluación comparativa entre el enfoque tradicional de mantenimiento y el modelo propuesto basado en inteligencia artificial y optimización. El experimento se llevó cabo bajo condiciones operativas equivalentes, considerando una restricción estricta de 40 horas disponibles para la ejecución de intervenciones. Este análisis busca evidenciar las diferencias en la asignación de recursos y el impacto generado, determinando si la transición hacia técnicas analíticas avanzadas representa una mejora frente a los métodos convencionales.

El método tradicional se modeló siguiendo la heurística habitual en el entorno industrial, donde la toma de decisiones se basa en la clasificación subjetiva del estado del equipo o en la percepción de criticidad del personal técnico. Bajo este esquema, la priorización es secuencial: se atienden primero los equipos clasificados como “Críticos” y posteriormente los de “Seguimiento” hasta agotar la disponibilidad temporal. En contraste, el modelo propuesto integra las salidas probabilísticas del modelo predictivo con un modelo de optimización basado en programación dinámica, formulado como un problema de mochila 0-1. Este enfoque evalúa simultáneamente la probabilidad de fallas, el costo asociado y el tiempo de ejecución, identificando la combinación de equipos que maximiza el beneficio total esperado bajo la restricción de tiempo disponible.

Los resultados obtenidos demuestran una mejora cuantitativa del modelo optimizado. Mientras que el enfoque tradicional alcanzó un beneficio por hora de \$1108.41, el

modelo propuesto logró elevar esta métrica a \$1191,41. Esta diferencia evidencia que el modelo es capaz de identificar configuraciones de intervención más eficientes, que no son fácilmente detectables mediante criterios empíricos. El análisis muestra que el método tradicional tiende a concentrar recursos en intervenciones individuales de mayor duración, sin considerar el costo de oportunidad, mientras que el modelo propuesto prioriza combinaciones de intervenciones que, en conjunto, generan un mayor beneficio acumulado en el mismo horizonte temporal.

Con el fin de sintetizar las diferencias entre ambos enfoques, se presenta el siguiente cuadro comparativo (Ver tabla 33).

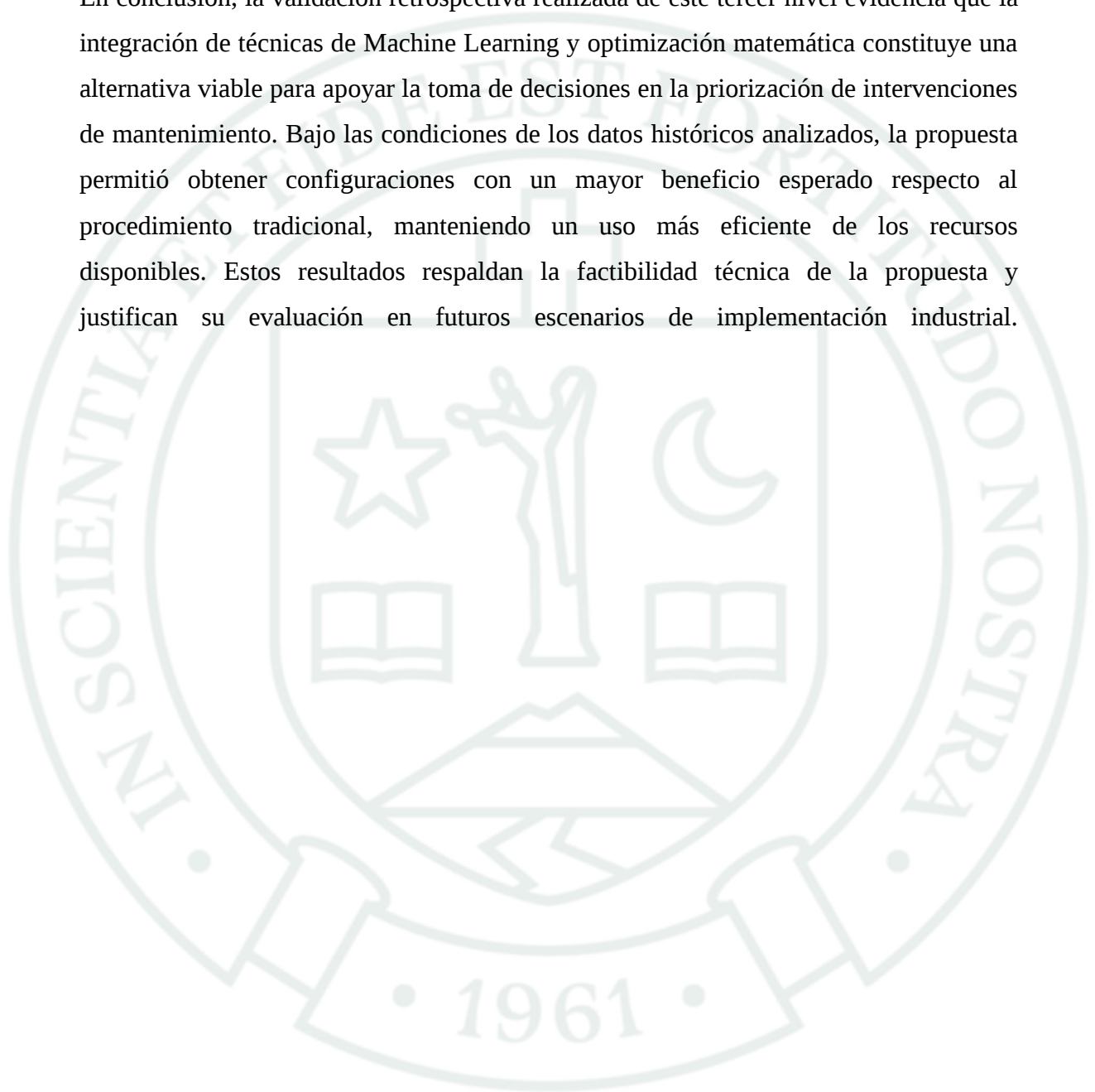
Tabla 34. Cuadro Comparativo: Método Tradicional vs Modelo Propuesto

Criterio	Método Tradicional	Modelo Propuesto (IA + Optimización)
Enfoque de decisión	Basado en la experiencia y reglas empíricas	Basado en datos, modelo predictivo y optimización
Uso de información	Clasificación del estado actual	Probabilidades de estado futuro
Consideración del riesgo	Parcial o implícita	Cuantificada mediante índice de criticidad
Criterio de priorización	Orden de criticidad (discreto)	Maximización del beneficio esperado
Uso de recursos	No optimizado	Optimizado bajo restricciones de tiempo
Capacidad de análisis	Unidimensional	Multivariable (probabilidad, costo, tiempo)
Adaptabilidad	Baja	Alta (modelo ajustable y escalable)
Dependencia del factor humano	Alta	Reducida (soporte analítico)
Beneficio por hora (\$)	1108.41	1191.41
Eficiencia operativa	Media	Alta
Toma de decisiones	Reactiva o preventiva	Predictiva y optimizada

Impacto en la operación	Variable	Mayor reducción del riesgo operativo
-------------------------	----------	--------------------------------------

Nota: Elaboración propia, a partir de la implementación de programación dinámica aplicada al problema de la mochila en Python

En conclusión, la validación retrospectiva realizada de este tercer nivel evidencia que la integración de técnicas de Machine Learning y optimización matemática constituye una alternativa viable para apoyar la toma de decisiones en la priorización de intervenciones de mantenimiento. Bajo las condiciones de los datos históricos analizados, la propuesta permitió obtener configuraciones con un mayor beneficio esperado respecto al procedimiento tradicional, manteniendo un uso más eficiente de los recursos disponibles. Estos resultados respaldan la factibilidad técnica de la propuesta y justifican su evaluación en futuros escenarios de implementación industrial.





CAPÍTULO V

5. Resultados y Discusión

En el presente capítulo se exponen los resultados obtenidos a partir de la aplicación del modelo propuesto, estructurado en tres niveles: un modelo predictivo basado en Machine Learning, un modelo de optimización del mantenimiento y una evaluación comparativa frente al método tradicional. Los resultados se presentan en función de los indicadores definidos en la operacionalización de las variables de estudio, con el propósito de analizar el desempeño de la propuesta basada en herramientas de inteligencia artificial y su potencial para apoyar la optimización del proceso de mantenimiento industrial mediante una validación retrospectiva basada en datos históricos.

Dado que la investigación se contextualiza en el sector de transporte internacional y que los datos económicos utilizados para el costo de mantenimiento, costo de falla y beneficio fueron trabajados en dólares estadounidenses, todos los resultados monetarios de este capítulo se expresan en dicha moneda.

5.1. Resultados de la variable independiente

La variable independiente fue evaluada a partir de los indicadores de desempeño del modelo predictivo, específicamente precisión, recall, F1-score y la generación de probabilidades por clase. El modelo de Machine Learning fue diseñado para clasificar el estado futuro de los equipos en tres categorías: normal, seguimiento y crítico, generando además probabilidades asociadas a cada clase.

Tabla 35. *Evaluación del Modelo de Inteligencia Artificial*

Indicador	Resultado
Precisión (Weighted avg)	0.73
Recall (Weighted avg)	0.67
F1-score (Weighted avg)	0.69
Probabilidades por clase	Generadas en rango (0-1)

Nota: Elaboración propia, a partir del modelo Random Forest optimizado en Python

Los resultados en la tabla 34 evidencian que el modelo presenta un desempeño adecuado para la clasificación multiclase del estado de los equipos. No obstante, el valor principal del modelo no reside únicamente en la clasificación discreta, sino en la generación de probabilidades por clase, las cuales permiten cuantificar el riesgo

asociado a cada equipo. Estas probabilidades constituyen el insumo fundamental para el cálculo del índice de criticidad y la posterior toma de decisiones en el proceso de optimización del mantenimiento.

5.2. Análisis económico de la propuesta

El análisis económico de la propuesta tiene como finalidad determinar si la incorporación del modelo predictivo basado en Machine Learning y del modelo de optimización mediante programación dinámica permite mejorar la asignación económica de los recursos destinados al mantenimiento industrial. Para ello, se evalúa el beneficio económico esperado asociado a las intervenciones seleccionadas y se compara el desempeño del modelo propuesto con el procedimiento tradicional bajo una misma restricción de disponibilidad de recursos.

Debe precisarse que la investigación corresponde a una validación retrospectiva basada en datos históricos y no a la implementación productiva de la propuesta en una empresa determinada. En consecuencia, no se dispone de información financiera empresarial completa relacionada con inversión inicial, costos reales de implementación, costos de parada por equipo, costos de falla, flujos de caja incrementales, horizonte de evaluación, tasa de descuento o valor residual.

Esta característica determina el alcance del análisis económico. Por tanto, se diferencia entre dos niveles de evaluación: la evaluación económica-operativa de la solución desarrollada, que si puede efectuarse con los resultados obtenidos en la investigación; y la evaluación financiera de una futura implementación empresarial, que requeriría información financiera que no forma parte de la base de datos disponibles.

Esta distinción resulta necesaria para evitar atribuir a la propuesta indicadores de rentabilidad financiera que no pueden ser estimados de manera válida sin una implementación real o, como mínimo, sin un escenario de inversión empresarial debidamente sustentado.

5.2.1. Estructura económica de la propuesta

La propuesta integra el resultado del modelo predictivo con a la decisión de mantenimiento. Para cada observación i , Random Forest proporciona las probabilidades de pertenencia a las tres condiciones consideradas:

$$P_i(Normal), P_i(Seguimiento), P_i(Crítico)$$

cumpléndose:

$$P_i(Normal) + P_i(Seguimiento) + P_i(Crítico) = 1$$

Estas probabilidades permiten superar una decisión basada únicamente en una clasificación discreta, debido a que incorporan información acerca del grado de incertidumbre asociado al estado estimado del equipo.

Para utilizar esta información dentro del modelo de decisión se plantea un índice de criticidad probabilístico:

$$IC_i = P_i(Crítico) + \alpha P_i(Seguimiento)$$

donde α representa la importancia relativa asignada a una condición de seguimiento respecto de una condición crítica.

Para el escenario de evaluación se considera:

$$\alpha = 0.5$$

obteniéndose:

$$IC_i = P_i(Crítico) + 0.5P_i(Seguimiento)$$

con:

$$0 \leq IC_i \leq 1$$

La ponderación de 0.5 constituye un parámetro del modelo de decisión y no un resultado aprendido por Random Forest. Su finalidad es representar una situación intermedia entre las condiciones Normal y Crítico, pudiendo ser modificada de acuerdo con las políticas de riesgos de una organización.

A partir del índice de criticidad puede formalizarse el costo esperado asociado a una eventual falla:

$$CEF_i = IC_i \times CF_i$$

donde:

- CEF_i : costo esperado de falla del equipo i ;
- CF_i : consecuencia económica estimada de la falla del equipo i .

En una implementación empresarial, CF_i podría estar constituido por:

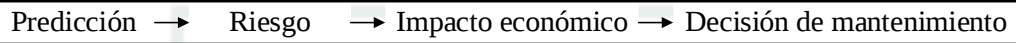
$$CF_i = C_{rep,i} + C_{mo,i} + C_{par,i} + C_{otros,i}$$

siendo C_{rep} el costo de reparación o repuestos, C_{mo} el costo de mano de obra, C_{par} el costo asociado a indisponibilidad o parada y C_{otros} otros costos atribuibles a la falla.

Por otra parte, ejecutar una intervención también genera un costo CM_i . En consecuencia, el beneficio económico esperado de intervenir el equipo puede expresarse como:

$$BE_i = (IC_i \times CF_i) - CM_i$$

Esta expresión permite vincular formalmente los tres componentes de la propuesta:



Por tanto, el modelo predictivo no sustituye la evaluación económica, sino que genera información probabilística que puede incorporarse a ella.

5.2.2. Modelo de optimización económica

Una vez determinado el beneficio esperado de cada alternativa de intervención, el problema consiste en seleccionar el conjunto de actividades que proporcione el mayor beneficio total posible dentro de la capacidad disponible.

Para cada intervención i , se define:

$$x_i = \begin{cases} 1, & \text{si la intervención es seleccionado} \\ 0, & \text{en caso contrario} \end{cases}$$

La función objetivo se expresa como:

$$\max Z = \sum_{i=1}^n BE_i x_i$$

sujeta a la restricción:

$$\sum_{i=1}^n t_i x_i \leq H$$

Donde t_i corresponde al tiempo requerido por la intervención y H representa las horas disponibles.

Para el escenario evaluado:

$$H = 40 \text{ horas}$$

y:

$$x_i \in \{0,1\}$$

El problema corresponde a una formulación de mochila 0-1 (0-1 Knapsack Problem) y fue resuelto mediante programación dinámica.

La finalidad económica del modelo no consiste simplemente en seleccionar los equipos con mayor criticidad o las intervenciones de menor duración, sino en encontrar la combinación que maximice el beneficio económico esperado agregado respetando la capacidad disponible.

Puede calcularse adicionalmente la relación:

$$RE_i = \frac{BE_i}{t_i}$$

como indicador de beneficio esperado por hora de intervención. Sin embargo, esta razón no constituye por sí sola el criterio de optimización, debido a que una selección voraz basada únicamente en BE_i/t_i no garantiza necesariamente la combinación óptima en un problema de mochila 0-1.

5.2.3. Resultado económico de la optimización

La evaluación realizada consideró una capacidad máxima de 40 horas tanto para el procedimiento tradicional como para el modelo propuesto, permitiendo realizar la comparación bajo igualdad de recursos.

El procedimiento tradicional alcanzó un beneficio esperado de:

$$Z_T = USD 44\,336.29$$

mientras que el modelo propuesto alcanzó:

$$Z_P = USD 47\,656.36$$

La diferencia absoluta corresponde a:

$$\Delta Z = Z_P - Z_T$$

$$\Delta Z = 47\,656.36 - 44\,336.29$$

$$\Delta Z = \text{USD } 3\,320.07$$

La mejora relativa se determina mediante:

$$MR = \frac{Z_P - Z_T}{Z_T} \times 100$$

obteniéndose:

$$MR = 7.49\%$$

Por otra parte, considerando que ambos procedimientos utilizan 40 horas, el beneficio esperado por hora del procedimiento tradicional fue:

$$\frac{44\,336.29}{40} = \text{USD } 1\,108.41/h$$

mientras que para el modelo propuesto fue:

$$\frac{47\,656.36}{40} = \text{USD } 1\,191.41/h$$

obteniéndose una diferencia de:

$$\text{USD } \frac{83.00}{h}$$

Tabla 36. Comparación económica del procedimiento tradicional y el modelo propuesto

Indicador	Tradicional	Propuesto	Diferencia
Horas disponibles	40 horas	40 horas	0
Beneficio económico esperado	USD 44,336.29	USD 47,656.36	USD 3,320.07
Beneficio esperado/hora	USD 1,108.41	USD 1,191.41	USD 83.00/h
Mejora relativa	-	7.49%	7.49%

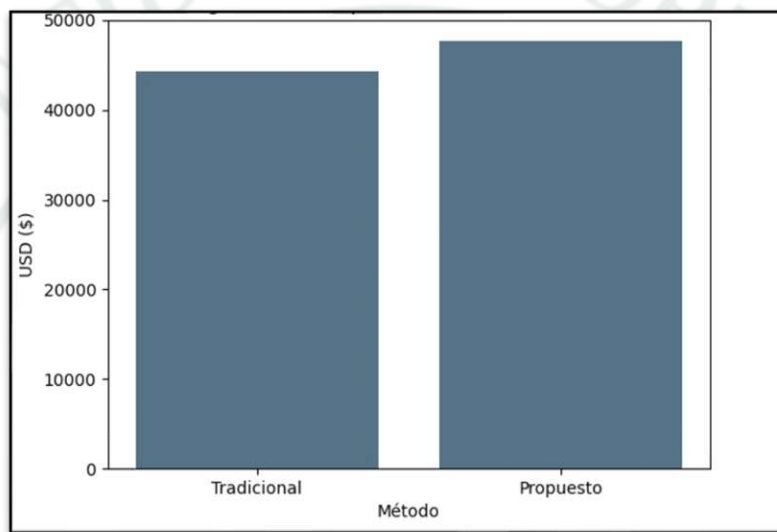
Nota: Elaboración propia

El resultado muestra que la ventaja económica obtenida no se explica por un incremento en los recursos disponibles, debido a que ambos escenarios utilizan la misma capacidad de 40 horas. La diferencia corresponde a una mejora asignación de dicha capacidad entre las alternativas de mantenimiento consideradas.

En consecuencia, dentro del escenario retrospectivo evaluado, el modelo permitió obtener un beneficio económico esperado adicional de USD 3,320.07, equivalente a una mejora de 7.49% respecto al procedimiento tradicional.

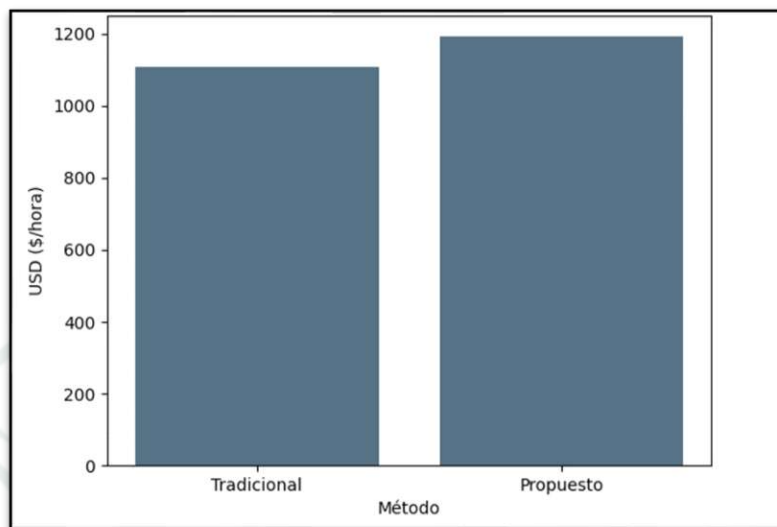
Debe enfatizarse que esta cantidad constituye un beneficio esperado derivado del escenario de evaluación, y no un ahorro contable efectivamente registrado por una empresa.

Figura 35. Comparación del Beneficio Total



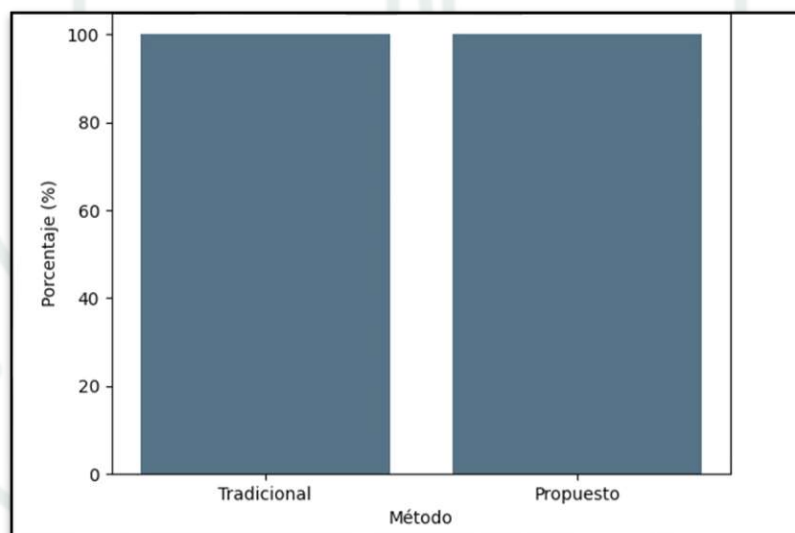
Nota: Elaboración propia a partir de los resultados del modelo de optimización basado en programación dinámica

Figura 36. Beneficio por Hora



Nota: Elaboración propia a partir de los resultados del modelo de optimización basado en programación dinámica

Figura 37. Utilización del Tiempo Disponible



Nota: Elaboración propia a partir de la restricción de tiempo del modelado (40 horas)

5.2.4. Análisis del costo de oportunidad de la capacidad disponible

La restricción de 40 horas genera una consecuencia económica adicional: no todas las intervenciones con beneficio esperado positivo pueden necesariamente ser ejecutadas.

Sea S^* el conjunto de intervenciones seleccionadas por el modelo:

$$S^* = \{i: x_i = 1\}$$

El beneficio capturado mediante la solución óptima es:

$$BC = \sum_{i \in S^*} BE_i$$

mientras que el beneficio potencial correspondiente a alternativas económicamente favorables que no fueron seleccionadas puede expresarse como:

$$BNS = \sum_{i \notin S^*} \max(BE_i, 0)$$

El término BNS no debe interpretarse como una pérdida contable real, sino como beneficio potencial no seleccionado debido a la restricción de capacidad.

Este concepto permite extender la utilidad del modelo hacia decisiones de planificación de recursos. Si la organización dispone de información sobre el costo de incrementar la capacidad de mantenimiento, puede compararse dicho costo con el beneficio esperado adicional generado al aumentar H .

Para ello se define:

$$Z^*(H)$$

como el beneficio óptimo obtenido con H horas disponibles.

El valor marginal de capacidad puede aproximarse mediante:

$$VM_H = \frac{Z^*(H + \Delta H) - Z^*(H)}{\Delta H}$$

Este indicador representa el incremento promedio del beneficio esperado asociado a una hora adicional de capacidad en el intervalo considerado.

De esta manera, la propuesta no solamente permite responder qué intervenciones deben ejecutarse, sino que proporciona una estructura para analizar posteriormente si económicamente resultaría conveniente incrementar la capacidad de mantenimiento.

5.2.5. Análisis de sensibilidad económica

La evaluación económica no debe depender exclusivamente de un único conjunto de parámetros, especialmente cuando los costos industriales pueden variar según el tipo de equipo, organización, disponibilidad de repuestos y consecuencias de una parada.

Por ello, la estructura económica propuesta permite desarrollar un análisis de sensibilidad respecto de los principales parámetros:

$$Z^* = f(CF, CM, H, \alpha)$$

donde:

- CF : costo asociado a la falla;
- CM : costo de mantenimiento;
- H : capacidad disponible;
- α : ponderación asignada a la condición Seguimiento.

Para un parámetro q , puede analizarse:

$$q' = q(1 + \delta)$$

considerando variaciones tales como:

$$\delta = \{-25\%, -10\%, 0, +10\%, +25\%\}$$

Esto permite estudiar si la superioridad económica del modelo se mantiene ante cambios razonables de las condiciones consideradas.

Particularmente, la sensibilidad respecto de la capacidad puede evaluarse mediante:

$$H = \{24, 32, 40, 48, 56\} \text{ horas}$$

resolviendo nuevamente el problema de optimización para cada escenario.

Asimismo, la ponderación de la condición Seguimiento puede evaluarse mediante:

$$\alpha = \{0.25, 0.50, 0.75\}$$

De esta manera se puede determinar si la selección de intervenciones presenta estabilidad ante diferentes niveles de aversión al riesgo.

En una implementación empresarial, el mismo procedimiento permitiría analizar simultáneamente escenarios conservadores, esperados y adversos utilizando los costos reales proporcionados por la organización.

En la presente investigación, este análisis se formula de manera paramétrica debido a que no se dispone de información financiera empresarial suficiente para asignar valores monetarios alternativos sin introducir supuestos no verificables.

Esta limitación evita presentar como resultados empíricos valores que no proceden de la información disponible.

5.2.6. Evaluación de VAN, TIR, ROI y periodo de recuperación

Como parte de una evaluación económica integral de proyectos de inversión, es habitual utilizar indicadores financieros como el Valor Actual Neto (VAN), la Tasa Interna de Retorno (TIR), el Retorno sobre la Inversión (ROI) y el periodo de recuperación o Payback. Sin embargo, su cálculo requiere información que debe corresponder a flujos monetarios identificables y a un horizonte temporal definido.

5.2.6.1. Valor actual Neto (VAN)

El VAN se determina mediante:

$$VAN = -I_0 + \sum_{t=1}^n \frac{FC_t}{(1+k)^t}$$

donde:

- I_0 : inversión inicial;
- FC_t : flujo de caja incremental en el periodo t ;
- k : tasa de descuento;
- n : horizonte de evaluación.

Para calcularlo sería necesario conocer, entre otros elementos, la inversión requerida para implementar la solución, costos de integración, infraestructura, capacitación, mantenimiento del sistema y los ahorros o beneficios monetarios efectivamente atribuibles a la propuesta en cada periodo.

5.2.6.2. Tasa Interna de Retorno (TIR)

La TIR corresponde a la tasa r que satisface:

$$TIR = -I_0 + \sum_{t=1}^n \frac{FC_t}{(1+r)^t}$$

Por consiguiente, la TIR tampoco puede determinarse sin una inversión inicial y una serie temporal de flujos de caja.

5.2.6.3. Retorno sobre la inversión (ROI)

El ROI puede expresarse como:

$$ROI = \frac{\text{Beneficio acumulado} - \text{Inversión}}{\text{Inversión}} \times 100$$

Por tanto, también requiere una inversión efectivamente cuantificada y beneficios acumulados durante un periodo determinado.

5.2.6.4. Periodo de recuperación

El periodo de recuperación requiere determinar el momento en que:

$$\sum_{t=1}^T FC_t \geq I_0$$

Lo que nuevamente exige contar con inversión inicial y flujos de caja periódicos.

5.2.6.5. Aplicabilidad a la presente investigación

La información disponible en esta investigación no contiene las variables financieras necesarias para calcular estos indicadores de manera empírica. La base histórica utilizada corresponde fundamentalmente a información técnica y operativa de mantenimiento y la propuesta no fue implementada productivamente en una organización durante un horizonte temporal que permita observar flujos de caja antes y después de su aplicación.

Tabla 37. Información disponible para los indicadores de VAN-TIR-ROI-Payback

Información requerida	Disponibilidad
Inversión inicial real I_0	No
Costos reales de integración e implementación	No

Flujos de caja incrementales por periodo FC_t	No
Horizonte financiero de evaluación	No
Tasa de descuento empresarial k	No
Ahorros realizados después de la implementación	No
Valor residual	No
Beneficio económico esperado del escenario retrospectivo	Si
Comparación tradicional vs propuesto	Si

Nota: Elaboración propia

En consecuencia, asignar valores arbitrarios a estas variables únicamente para obtener VAN, TIR, ROI o Payback produciría indicadores matemáticamente calculables, pero sin validez empírica para la investigación.

Por esta razón, no se considera metodológicamente apropiado reportar un VAN, TIR, ROI o periodo de recuperación como resultado de la tesis.

Esto no constituye una ausencia del análisis económico, sino una delimitación entre evaluación económica-operativa y evaluación financiera de una inversión. La presente investigación realizar la primera; la segunda podrá efectuarse cuando la propuesta sea implementada en una organización específica y se disponga de los datos financieros correspondientes.

5.2.7. Modelo financiero propuesto para una futura implementación

Aunque los indicadores financieros anteriores no pueden calcularse con rigor a partir de la información disponibles, la propuesta deja definida la estructura necesaria para hacerlo en una implementación empresarial posterior.

La inversión inicial podría expresarse como:

$$I_0 = C_{adaptación} + C_{integración} + C_{infraestructura} + C_{capacitación} + C_{puesta\ en\ marcha}$$

Los beneficios incrementales del periodo t podrían determinarse mediante:

$$BI_t = Ahorro_{fallas,t} + Ahorro_{paradas,t} + Ahorro_{MO,t} + Ahorro_{repuestos,t}$$

mientras que los costos recurrentes serían:

$$CR_t = C_{soporte,t} + C_{actualización,t} + C_{reentrenamiento,t} + C_{infraestructura,t}$$

El flujo de caja incremental correspondiente sería:

$$FC_t = BI_t - CR_t$$

Con estos valores sería posible construir el flujo financiero

Tabla 38. Construcción del flujo financiero

Periodo	Inversión	Beneficio incremental	Costo recurrente	Flujo neto
0	$-I_0$	0	0	$-I_0$
1	-	BI_1	CR_1	FC_1
2	-	BI_2	CR_2	FC_2
...	...	...	...	...
n	-	BI_n	CR_n	FC_n

Nota: Elaboración propia

y posteriormente determinar: VAN, TIR, ROI, Payback

La incorporación de esta estructura tiene dos finalidades. Primero, demostrar que la propuesta puede ser sometida a una evaluación financiera completa cuando se disponga de información empresarial. Segundo, establecer claramente cuáles son los datos adicionales que deberán recopilarse durante una futura implementación.

Por tanto, una organización que adopte la propuesta deberá registrar como mínimo los costos de implementación, mantenimiento del sistema, fallas evitadas, tiempos de parada, intervenciones ejecutadas, repuestos, mano de obra y beneficios obtenidos durante un horizonte temporal definido.

5.2.8. Interpretación integral del análisis económico

El análisis desarrollado permite diferenciar tres conceptos que no deben confundirse:

Beneficio esperado $\neq$ Ahorro realizado $\neq$ Flujo de caja

El beneficio esperado corresponde al valor utilizado por el modelo para comparar y priorizar alternativas.

El ahorro realizado solo puede determinarse después de una implementación, comparando costos efectivamente observados frente a una línea base.

El flujo de caja requiere además establecer cuando se producen los ingresos, ahorros, inversiones y costos durante el horizonte temporal.

Esta diferencia explica por qué el resultado USD 47 656.36 no debe utilizarse directamente como flujo de caja para calcular VAN o TIR.

Hacerlo supondría asumir, sin evidencia, que dicho beneficio se obtiene periódicamente, que se materializa íntegramente como ahorro monetario y que puede repetirse durante un determinado número de años. Ninguna de estas condiciones fue evaluada en la investigación.

El resultado que sí puede sostenerse empíricamente es que, bajo la misma restricción de 40 horas y dentro del escenario retrospectivo analizado, la propuesta alcanzó un beneficio económico esperado de USD 47,656.36 frente a USD 44,336.29 del procedimiento tradicional, representando una diferencia esperada de USD 3,320.07 y una mejora relativa de 7.49% en la asignación de los recursos disponibles.

Por tanto, la evidencia obtenida permite sostener la conveniencia económica-operativa del mecanismo de priorización dentro del escenario evaluado, pero no afirmar todavía la rentabilidad financiera de una implementación empresarial específica.

5.2.9. Alcance del análisis económico

A partir de los resultados obtenidos, el análisis económico permite establecer que la integración de las probabilidades generadas por Machine Learning con un modelo de optimización mediante programación dinámica proporciona un mecanismo cuantitativo para asignar recursos limitados de mantenimiento buscando maximizar el beneficio esperado.

La evaluación retrospectiva mostró una mejora de 7.49 % respecto del procedimiento tradicional manteniendo constante la capacidad de 40 horas, lo cual evidencia que la ventaja obtenida proviene de la selección de las intervenciones y no de una mayor disponibilidad de recursos.

No obstante, este resultado debe circunscribirse al escenario analizado. La investigación no demuestra que cualquier empresa que implemente la propuesta obtendrá necesariamente una mejora de 7.49 %, debido a que el resultado dependerá de sus

equipos, estructura de costos, frecuencia y consecuencias de las fallas, capacidad de mantenimiento y demás condiciones operativas.

Asimismo, la ausencia de información empresarial sobre inversión inicial y flujos de caja impide calcular de manera válida VAN, TIR, ROI y Payback. En lugar de sustituir dichos datos mediante supuestos arbitrarios, se ha definido la estructura matemática necesaria para su determinación futura.

5.3. Discusión de resultados

Los resultados obtenidos durante la validación retrospectiva evidencian que la propuesta desarrollada constituye una alternativa técnicamente viable para apoyar la toma de decisiones en el proceso de mantenimiento industrial. La integración de técnicas de Machine Learning y programación dinámica permitió clasificar la condición de los equipos y priorizar las intervenciones utilizando información histórica, generando un mayor beneficio esperado respecto al procedimiento tradicional bajo las condiciones evaluadas.

En primer lugar, el modelo predictivo mostró un desempeño adecuado en la clasificación de los estados de los equipos, generando probabilidades por clase que permitieron cuantificar el riesgo asociado a cada activo y servir como insumo para el modelo de optimización. Estos resultados respaldan la utilidad del modelo como herramienta de apoyo para la priorización de intervenciones, sin implicar por sí mismos una validación de su impacto operativo en un entorno industrial real.

En segundo lugar, el modelo de optimización, basado en programación dinámica y formulado como un problema de mochila 0-1, permitió seleccionar de manera óptima los equipos a intervenir. El hecho de que el modelo propuesto alcance un beneficio total de \$ 47,656.36 frente a los \$ 44,336.29 del método tradicional demuestra, como se ve en la figura 35, que la integración entre predicción y optimización genera mejores resultados que la asignación manual de recursos.

Asimismo, el incremento del beneficio por hora de \$ 1,108.41/h a \$ 1,191.41/h evidencia una mejora en la eficiencia del uso del tiempo disponible. Este resultado es especialmente relevante en contextos industriales donde los recursos son limitados y la correcta priorización de intervenciones es clave para garantizar la continuidad operativa.

Finalmente, la mejora del 7.49% respecto al método tradicional valida la utilidad del modelo propuesto como herramienta de apoyo a la toma de decisiones. El uso de programación dinámica permite evaluar múltiples combinaciones de intervención, superando las limitaciones de los enfoques heurísticos tradicionales y contribuyendo a una gestión de mantenimiento más eficiente y basada en datos.

5.4. Validación de la hipótesis

Los resultados obtenidos permiten aceptar la hipótesis general de la investigación, dado que la aplicación de análisis de datos y herramientas de inteligencia artificial, integradas con un modelo de optimización basado en programación dinámica, permitió mejorar la toma de decisiones en el proceso de mantenimiento industrial.

En particular, los resultados obtenidos evidenciaron que el modelo propuesto supera al método tradicional en términos de beneficio total, beneficio por hora y eficiencia en la selección de equipos, lo cual demuestra que la propuesta desarrollada constituye una alternativa más efectiva para la gestión del mantenimiento en el sector de transporte internacional.

CONCLUSIONES

En la presente investigación se desarrolló un modelo integral basado en inteligencia artificial y optimización matemática, orientado a mejorar la toma de decisiones en el proceso de mantenimiento industrial. A partir de los resultados obtenidos y de los objetivos redactados, se establecen las siguientes conclusiones.

Primero. Los resultados permiten concluir que se desarrolló una propuesta técnicamente viable para apoyar la optimización del proceso de mantenimiento industrial mediante el análisis de datos y herramientas de Inteligencia Artificial. La integración del diagnóstico, el modelo predictivo y el modelo de optimización permitió transformar los registros históricos de mantenimiento en información útil para anticipar la condición futura de los equipos y apoyar la selección de intervenciones bajo recursos limitados. La validación retrospectiva realizada mostró un mejor desempeño económico y operativo respecto al método tradicional dentro del escenario evaluado, evidenciando el potencial de la propuesta como herramienta de apoyo para la toma de decisiones en mantenimiento industrial.

Segundo. El diagnóstico situacional permitió identificar las principales características y limitaciones de las prácticas actuales de gestión y toma de decisiones en el mantenimiento. Los resultados evidenciaron una predominancia del mantenimiento preventivo, la persistencia de fallas inesperadas y una priorización de las intervenciones basada principalmente en indicadores técnicos y en la criticidad percibida por el personal. Asimismo, se identificaron limitaciones relacionadas con la disponibilidad de tiempo y la adaptación tecnológica. Estos hallazgos evidencian la necesidad de complementar los criterios tradicionales de mantenimiento con herramientas analíticas que permitan aprovechar los datos históricos para anticipar el estado de los equipos y apoyar la asignación de los recursos disponibles.

Tercero. A partir de las variables operativas obtenidas de los datos históricos, se desarrolló un modelo integral que articula la predicción del estado futuro de los equipos con la optimización de las intervenciones de mantenimiento. El modelo predictivo basado en Random Forest optimizado alcanzó un F1-score ponderado de 0.69 y permitió generar probabilidades para los estados normal, seguimiento y crítico. Estas probabilidades constituyeron el insumo para cuantificar el riesgo y determinar el

beneficio esperado de las intervenciones. Posteriormente, mediante programación dinámica y la formulación de un problema de mochila 0-1, se seleccionó la combinación de seis equipos que maximiza el beneficio esperado bajo una restricción de 40 horas, alcanzando un beneficio total de USD 47,656.36.

Cuarto. La evaluación comparativa del modelo propuesto frente a método tradicional evidenció una mejora en la eficiencia de la asignación de recursos dentro del escenario analizado. Bajo la misma disponibilidad de 40 horas, el método tradicional alcanzó un beneficio total esperado de USD 44,336.29 y un beneficio de USD 1,108.41 por hora, mientras que el modelo propuesto alcanzó USD 47,656.36 y USD 1,191.41 por hora, respectivamente. Esto representa un incremento del 7.49 % en el beneficio esperado, mostrando que la combinación de predicción y optimización permite identificar alternativas de intervención con un mayor aprovechamiento de los recursos disponibles que la priorización tradicional.

RECOMENDACIONES

A partir de los resultados obtenidos, las limitaciones identificadas en el desarrollo de la investigación y los objetivos, general y específicos, se plantean las siguientes recomendaciones.

PRIMERA: A partir de las limitaciones identificadas en el diagnóstico situacional, se recomienda a las organizaciones fortalecer la gestión y sistematización de sus registros históricos de mantenimiento, procurando mantener información estructurada, actualizada y consistente sobre las condiciones de los equipos, intervenciones realizadas y resultados obtenidos. Asimismo, se recomienda complementar los criterios técnicos y la experiencia del personal con herramientas de análisis de datos que proporcionen información objetiva para apoyar la priorización de las intervenciones.

SEGUNDA: Respecto al modelo integral desarrollado, se recomienda ampliar progresivamente la cantidad y diversidad de los datos utilizados para su entrenamiento, incorporando registros de diferentes equipos, condiciones operativas y periodos de tiempo, así como evaluar la incorporación de información proveniente de sensores IoT. También se recomienda continuar evaluando técnicas de Machine Learning que permitan mejorar especialmente el desempeño en las clases con menor representación, manteniendo la generación de probabilidades como elemento fundamental para su integración con el modelo de optimización.

TERCERA: Para fortalecer la aplicación del modelo de optimización, se recomienda incorporar progresivamente restricciones adicionales propias de un entorno industrial real, tales como disponibilidad de personal, costos diferenciados de intervención, disponibilidad de repuestos, prioridades operativas y ventanas de mantenimiento. Asimismo, los parámetros económicos y operativos utilizados por el modelo deberán ser ajustados a las características particulares de cada organización antes de su implementación.

CUARTA: Finalmente, se recomienda realizar una implementación piloto del modelo propuesto en un entorno industrial real, con el propósito de validar prospectivamente los resultados obtenidos en la presente investigación y evaluar su impacto sobre indicadores como disponibilidad de equipos, ocurrencia de fallas, costos de mantenimiento, beneficio obtenido y eficiencia en la utilización de los recursos. Esta implementación permitiría contrastar los resultados de la validación retrospectiva con el desempeño real

del modelo y determinar las condiciones necesarias para su posterior integración con plataformas digitales de gestión del mantenimiento.



REFERENCIAS BIBLIOGRÁFICAS

- Abensur, E. O., Santos, B. P., & Bandeira, A. A. (2023). *Optimization models as applied to equipment replacement problems: review and trends*. *Gestão & Produção*, 30, e4022. <https://doi.org/10.1590/1806-9649-2023v30e4022>
- Ahmad, R., & Kamaruddin, S. (2012). *An overview of time-based and condition-based maintenance in industrial application*. *Computers & Industrial Engineering*, 63(1), 135–149. <https://doi.org/10.1016/J.CIE.2012.02.002>
- Alvarez Campos, D. A. (2024). *Sistema de mantenimiento predictivo utilizando algoritmos de Machine Learning en un compresor de aire industrial en el sector industrial*. Universidad César Vallejo.
- Banco Interamericano de Desarrollo. (2022). *Índice de madurez Digital: Cómo medir el avance de la transformación digital en las administraciones tributarias*. BID. Obtenido de: <https://www.iadb.org/es/blog/modernizacion-del-estado/gestion-fiscal/indice-de-madurez-digital-como-medir-el-avance-de-la-transformacion-digital-en-las-administraciones>
- Becerra Correa, N., & Leguizamón Páez, M. A. (2024). *Regresión logística técnica de Machine Learning para predicciones académicas*. Xikua Boletín Científico de la Escuela Superior de Tlahuelilpan, 12(Especial). <https://doi.org/10.29057/xikua.v12iEspecial.12746>
- Behzadi, A., Holmberg, S., Duwig, C., Haghighat, F., Ooka, R., & Sadrizadeh, S. (2022). *Smart design and control of thermal energy storage in low-temperature heating and high-temperature cooling systems: A comprehensive review*. *Renewable and Sustainable Energy Reviews*, 166, 112625. <https://doi.org/10.1016/J.RSER.2022.112625>
- Benavente, A. R. (2024). *Mantenimiento predictivo usando algoritmos de Machine Learning aplicado a bombas de agua* (Tesis de título profesional). Universidad Nacional del Altiplano.
- Bokrantz, J., Skoogh, A., Berlin, C., & Stahre, J. (2020). *Maintenance in digitalised manufacturing: Delphi-based scenarios for 2030*. <https://doi.org/10.1016/j.ijpe.2017.06.010>

- Bousdekis, A., Magoutas, B., Apostolou, D., & Mentzas, G. (2021). *Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance*. J Intell Manuf 29, 1303–1316 (2018).
<https://doi.org/10.1007/s10845-015-1179-5>
- Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). *A systematic literature review of machine learning methods applied to predictive maintenance*. Computers & Industrial Engineering, 137, 106024. <https://doi.org/10.1016/J.CIE.2019.106024>
- Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, New York, NY, USA, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Congreso de la República del Perú. (2022). *Ley N.º 29783, Ley de Seguridad y Salud en el Trabajo*. Diario Oficial El Peruano
- Consejo Nacional de Competitividad y Formalización. (2024). *Plan Nacional de Competitividad y Productividad*. Ministerio de Economía y Finanzas.
<https://www.gob.pe/institucion/mef/campa%C3%B1as/229-plan-nacional-de-competitividad-y-productividad>
- Cortes, C., & Vapnik, V. (1995). *Support-vector networks*. Machine Learning, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Davenport, T. H., & Ronanki, R. (2018). *Artificial intelligence for the real world*. Harvard Business Review, 96(1), 108–116
- Ersöz, O. Ö., İnal, A. F., Aktepe, A., Türker, A. K., & Ersöz, S. (2022). *A Systematic Literature Review of the Predictive Maintenance from Transportation Systems Aspect*. In Sustainability (Switzerland) (Vol. 14, Issue 21). MDPI.
<https://doi.org/10.3390/su142114536>
- Farfán Arias, P. C. (2025). *Desarrollo de un modelo predictivo con Machine Learning y su influencia en la estimación del IRI en pavimentos flexibles de la Av. Antigua Panamericana Sur, Chilca* (Tesis de título profesional). Universidad Tecnológica del Perú.

- Franciosi, C., Iung, B., Miranda, S., & Riemma, S. (2020). *Maintenance for sustainability in the Industry 4.0 context: A scoping literature review*. <https://doi.org/10.1016/j.ifacol.2018.08.459>
- García, A., & Canibano, E. (2024). *iDIGIT4L. Nuevos ecosistemas de digitalización y aprendizaje hombre-máquina para sistemas de fabricación industrial heredados*. ArXiv, abs/2405.16676
- Ghadi, Y. Y., AlShoul, T., Nezami, Z. I., Ali, H., Asif, M., Aljuaid, H., & Ahmad, S. (2023). *An efficient optimizer for the 0/1 knapsack problem using group counseling*. PeerJ. Computer science, 9, e1315. <https://doi.org/10.7717/peerj-cs.1315>
- Ghobakhloo, M. (2020). *Industry 4.0, digitization, and opportunities for sustainability*. Journal of Cleaner Production, 252, 119869. <https://doi.org/10.1016/j.jclepro.2019.119869>
- Giacotto, A., Costa Marques, H., & Martinetti, A. (2025). *Prescriptive maintenance: A comprehensive review of current research and future directions*. Journal of Quality in Maintenance Engineering, 31(1), 129–173.
- Gobierno del Perú. (2025). *Transformación digital en el Perú*. <https://www.gob.pe/transformaciondigital>
- Goodfellow I., Bengio Y., Courville A. (2016). *Deep Learning*. MIT Press
- Grinsztajn L, Oyallon E, Varoquaux G. (2022). *Why do tree-based models still outperform deep learning on tabular data?* https://openreview.net/forum?id=Fp7__phQszn
- Grooss, O. F. (2024). *Digitalization of maintenance activities in small and medium-sized enterprises: A conceptual framework*. Computers in Industry, 154, Artículo 104039 <https://doi.org/10.1016/j.compind.2023.104039>
- Géron A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow*. 2nd ed. O'Reilly
- Haleel, A. J., & Dawood, L. M. (2023). *Maintenance scheduling optimization using artificial intelligence techniques: A review*. 5th International Congress on

- Human-Computer Interaction, Optimization and Robotic Applications. (pp. 1-6). IEEE. <https://ieeexplore.ieee.org/document/10156713>
- Hamaide, V., Joassin, D., Castin, L., & Glineur, F. (2022). *A two-level machine learning framework for predictive maintenance: comparison of learning formulations*. arXiv. <http://arxiv.org/abs/2204.10083>
- Hillier, F. S., & Lieberman, G. J. (2021). *Introduction to Operations Research* (11th ed.). McGraw-Hill.
- IBM. (2023). *Predictive maintenance*. IBM Corporation. <https://www.ibm.com/topics/predictive-maintenance>
- Inafuku Yoshida, A.H. (2020). *Diseño e implementación de un sistema de diagnóstico de fallas para la inspección y detección de fallas en componentes de procesos industriales utilizando un robot móvil y algoritmos de inteligencia artificial* (Tesis de Posgrado). Pontificia Universidad Católica del Perú.
- Instituto Nacional de Estadística e Informática. (2024, abril). *Producción Nacional abril 2024 (Nota de prensa N° 056)*. <https://m.inei.gob.pe/prensa/noticias/produccion-nacional-se-incremento-285-en-febrero-de-2024-15114/>
- Instituto Nacional de Estadística e Informática. (2024, junio). *Producción Nacional junio 2024 (Nota de prensa N° 090)*. <https://m.inei.gob.pe/prensa/noticias/produccion-nacional-aumento-en-528-durante-abril-de-2024-15214/>
- Instituto Nacional de Estadística e Informática. (2025). *Demografía empresarial en el Perú*. INEI.
- Instituto Nacional de Estadística e Informática. (2024). *Estructura empresarial 2024*. INEI.
- International Labour Organization. (2023). *Global Strategy on Occupational Safety and Health 2024–30 and plan of action for its implementation*. ILO

- International Organization for Standardization. (2014). *ISO 55000: Asset management — Overview, principles and terminology*. ISO.
<https://www.iso.org/standard/55088.html>
- Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). *A review on machinery diagnostics and prognostics implementing condition-based maintenance*. *Mechanical Systems and Signal Processing*, 20(7), 1483–1510.
<https://doi.org/10.1016/J.YMSSP.2005.09.012>
- Kagermann, H., Wahlster, W., & Helbig, J. (2013). *Recommendations for implementing the strategic initiative industrie 4.0*. Acatech
- Kane, A., Kore, A., Khandale, A., Nigade, S., & Joshi, P. P. (2022). *Predictive Maintenance using Machine Learning*. arXiv.
<https://doi.org/10.48550/arXiv.2205.09402>
- Kirchherr, J., Reike, D., & Hekkert, M. (2017). *Conceptualizing the circular economy: An analysis of 114 definitions*. *Resources, Conservation and Recycling*, 127, 221–232. <https://doi.org/10.1016/J.RESCONREC.2017.09.005>
- Kuhn, M., & Johnson, K. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models (1st ed.)*. Chapman and Hall/CRC.
<https://doi.org/10.1201/9781315108230>
- Lee, J., Bagheri, B., & Kao, H. A. (2015). *A cyber-physical systems architecture for Industry 4.0-based manufacturing systems*. *Manufacturing Letters*, 3, 18–23.
<https://doi.org/10.1016/j.mfglet.2014.12.001>
- Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). *Industrial Artificial Intelligence for industry 4.0-based manufacturing systems*. *Manufacturing Letters*, 18, 20–23. <https://doi.org/10.1016/J.MFGLET.2018.09.002>
- Liao, Y., Deschamps, F., Loures, E. F. R., & Ramos, L. F. P. (2019). *Past, present and future of Industry 4.0: A systematic literature review and research agenda proposal*. *International Journal of Production Research*, 55(12), 3609–3629.
<https://doi.org/10.1080/00207543.2017.1308576>

- Maheshwari, S., Tiwari, S., Rai, S., & Daman Pratap Singh, S. V. (2024). *Comprehensive Study of Predictive Maintenance in Industries Using Classification Models and LSTM Model*. arXiv:2403.10259
- Mallioris, P., Aivazidou, E., & Bechtsis, D. (2024). *Predictive maintenance in Industry 4.0: A systematic multi-sector mapping*. CIRP Journal of Manufacturing Science and Technology, 50, 80-103. <https://doi.org/10.1016/j.cirpj.2024.02.003>
- Ministerio de la Producción. (2025). *Nueva iniciativa busca impulsar la transformación digital de las Mype en el país*. PRODUCE.
- Ministerio de Transporte y Comunicaciones. (2025). *Informe Multianual de inversiones en Asociaciones Público y Privadas*.
https://www.mef.gob.pe/es/?option=com_content&language=es-ES&Itemid=101579&lang=es-ES&view=article&id=6049
- Mobley, R. K. (2020). *Maintenance fundamentals (4th ed.)*. Elsevier
- Mobley, R. K. (2002). *An introduction to predictive maintenance (2nd ed.)*. Butterworth-Heinemann
- Natekin, A., & Knoll, A. (2013). *Gradient boosting machines, a tutorial*. Front. Neurorobot. 7:21. doi: 10.3389/fnbot.2013.00021
- Palma Villanueva, R. M. (2021). *Inteligencia Artificial y Machine Learning para el desarrollo e Implementación de Softsensors en la Predicción del P80 para la Molienda SAG (SABC-A) – Minera Las Bambas* (Tesis de Posgrado). Universidad Nacional de San Agustín de Arequipa.
- Panuera Moreno, Y., Coronado Diaz, M. A., Castillo Romani, J., Canchanya Molina, F., & Gonzales Muñoz, N. (2025). *Estructura Empresarial 2024*. Instituto Nacional de Estadística e Informática.
- Patil, C. R., Jadhav, S. K., Bardiya, A. L., Davande, A. P., & Raverkar, M. P. (2023). *Machine Learning-Based Predictive Maintenance of Industrial Machines*. International Journal of Computer Trends and Technology, 71(3), 50-56. <https://doi.org/10.14445/22312803/%20IJCTT-V71I3P108>

- Presidencia del Consejo de ministros. (2023). *Política Nacional de Transformación Digital al 2030*. Gobierno del Perú. <https://www.gob.pe/institucion/pcm/normas-legales/4471543-085-2023-pcm>
- Presidencia del Consejo de ministros. (2025). *Estrategia Nacional de Gobierno de Datos 2026-2030*. Gobierno del Perú. <https://www.gob.pe/99097-estrategia-nacional-de-gobierno-de-datos-2026-2030>
- Pucho Medina, J.J. (2024). *Propuesta de un modelo de Machine Learning para el mantenimiento predictivo de transmisiones de una pala eléctrica Bucyrus 495 BII* (Tesis de título profesional). Universidad Tecnológica del Perú.
- Rai, A. (2020). *Explainable AI: From black box to glass box*. MIS Quarterly, 44(4), 1377–1380. <https://link.springer.com/article/10.1007/s11747-019-00710-5>
- Samatas, G. G., Moumgiakmas, S. S., & Papakostas, G. A. (2021). *Predictive Maintenance-Bridging Artificial Intelligence and IoT*. 2021 IEEE World AI IoT Congress (AIIoT), 0413-0419. <https://doi.org/10.1109/AIIoT52608.2021.9454173>
- Selcuk, S. (2017). *Predictive maintenance, its implementation and latest trends*. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 231(9), 1670–1679. <https://journals.sagepub.com/doi/10.1177/0954405415601640>
- Shwartz-Ziv, R. and Armon, A. (2022) *Tabular Data: Deep Learning Is Not All You Need*. Information Fusion, 81, 84-90. <https://doi.org/10.1016/j.inffus.2021.11.011>
- Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). *Machine Learning for Predictive Maintenance: A Multiple Classifier Approach*. IEEE Transactions on Industrial Informatics, 11(3), 812-820. <https://ieeexplore.ieee.org/document/6879441>
- Tao, F., Zhang, M., Liu, Y., & Nee, A. Y. C. (2019). *Digital twin in industry: State-of-the-art*. IEEE Transactions on Industrial Informatics, 15(4), 2405–2415. <https://doi.org/10.1109/TII.2018.2873186>

- Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). *Data-driven smart manufacturing*. *Journal of Manufacturing Systems*, 48, 157–169.
<https://doi.org/10.1016/J.JMSY.2018.01.006>
- Tapia, S., Aguilera, G., Rojas, L., & García, J. (2025). *Mantenimiento predictivo basado en Machine Learning: una revisión sistemática de la literatura y perspectivas en la Industria 4.0*. *ACI Avances en Ciencias e Ingenierías*, 15(4), 63–94.
- Taşcı, B., Omar, A., & Ayvaz, S. (2023). *Remaining useful lifetime prediction for predictive maintenance in manufacturing*. *Computers & Industrial Engineering*, 184, 109566. <https://doi.org/10.1016/J.CIE.2023.109566>
- Tortorella, G. L., de Castro Fettermann, D., Frank, A., & Marodin, G. (2018). *Lean manufacturing implementation: Leadership styles and contextual variables*. *International Journal of Operations & Production Management*, 38(5), 1205–1227. <https://doi.org/10.1108/IJOPM-08-2016-0453>
- Tsallis, C., Papageorgas, P., Piromalis, D., & Munteanu, R. A. (2025). *Application-Wise Review of Machine Learning-Based Predictive Maintenance: Trends, Challenges, and Future Directions*. In *Applied Sciences (Switzerland)* (Vol. 15, Issue 9). Multidisciplinary Digital Publishing Institute (MDPI).
<https://doi.org/10.3390/app15094898>
- Tsang, A. H. C. (2002). *Strategic dimensions of maintenance management*. *Journal of Quality in Maintenance Engineering*, 8(1), 7–39.
<https://doi.org/10.1108/13552510210420577>
- United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. United Nations. <https://sdgs.un.org/2030agenda>
- Wen, L., Li, X., Gao, L., & Zhang, Y. (2020). *A new convolutional neural network-based data-driven fault diagnosis method*. *IEEE Transactions on Industrial Electronics*, 67(7), 5857–5867. <https://ieeexplore.ieee.org/document/8114247>
- Wen, Y., Fashiar Rahman, M., Xu, H., & Tseng, T. L. B. (2022). *Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective*. *Measurement*, 187, 110276.
<https://doi.org/10.1016/J.MEASUREMENT.2021.110276>

- World Economic Forum. (2023). *The future of jobs report 2023*. World Economic Forum. <https://www.weforum.org/publications/the-future-of-jobs-report-2023/>
- Wu, F., Wu, Q., Tan, Y., & Xu, X. (2024). *Remaining Useful Life Prediction Based on Deep Learning: A Survey*. *Sensors*, 24(11), 3454. <https://doi.org/10.3390/s24113454>
- Xu, L. D., Xu, E. L., & Li, L. (2018). *Industry 4.0: state of the art and future trends*. *International Journal of Production Research*, 56(8), 2941–2962. <https://doi.org/10.1080/00207543.2018.1444806>
- Zhang, H., He, X., Yan, W., Jiang, Z., & Zhu, S. (2022). *A machine learning-based approach for product maintenance prediction with reliability information conversion*. *Autonomous Intelligent Systems*, 2(1). <https://doi.org/10.1007/s43684-022-00033-3>
- Zhang, W., Yang, D., & Wang, H. (2019). *Data-driven methods for predictive maintenance of industrial equipment: A survey*. *IEEE Systems Journal*, 15(3), 4341–4351. <https://ieeexplore.ieee.org/document/8707108>
- Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). *Deep learning and its applications to machine health monitoring*. *Mechanical Systems and Signal Processing*, 115, 213–237. <https://doi.org/10.1016/J.YMSSP.2018.05.050>
- Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). *Predictive maintenance in the Industry 4.0: A systematic literature review*. *Computers & Industrial Engineering*, 150, 106889. <https://doi.org/10.1016/J.CIE.2020.106889>



Anexo 1. Ficha de Validación de la Encuesta aplicada

FICHA DE VALIDACIÓN DEL INSTRUMENTO

Estimado experto:

Se solicita evaluar cada uno de los ítems del instrumento de investigación considerando los siguientes criterios:

- Pertinencia: El ítem guarda relación con el objetivo de la investigación.
- Claridad: El ítem está redactado de forma comprensible.
- Coherencia: El ítem mantiene relación lógica con la dimensión evaluada.
- Relevancia: El ítem es importante para obtener información significativa.

Marque con el puntaje correspondiente según la siguiente escala:

Valor	Nivel
1	Deficiente
2	Regular
3	Bueno
4	Muy Bueno
5	Excelente

Matriz de Validación del Instrumento

N.º	Ítem de la encuesta	Pertinencia	Claridad	Coherencia	Relevancia
1	¿Cuál es su cargo?	5	5	5	5
2	¿A qué tipo de empresa pertenece?	4	4	4	4
3	¿Qué tipo de mantenimiento predomina en su empresa?	5	4	5	5
4	¿Con qué frecuencia ocurren fallas no planificadas?	5	5	5	4
5	Cuando ocurre una falla, ¿cómo se decide la intervención?	4	4	4	4
6	¿Se utilizan datos históricos para tomar decisiones de mantenimiento?	5	5	5	4

Nota: Elaboración propia

Anexo 2. Evaluación del Instrumento

Aspecto	Valoración
Pertinencia global del instrumento	5
Claridad global del instrumento	4
Coherencia global del instrumento	4
Relevancia global del instrumento	4

Nota: Elaboración propia

Anexo 3. Firma del Evaluador Experto

Nombre y firma: _____

Wilbert Felipe Zevallos Gonzales

06/05/2026

Nota: Elaboración propia

Anexo 4. Código completo de Google Colab #1

```

FASE 1: LIMPIEZA Y PREPROCESAMIENTO

import pandas as pd
import numpy as np

# Cargar datos
df = pd.read_excel("DataHistorica.xlsx", sheet_name="alertas_tempranas_PD")

# Renombrar columnas (opcional para facilidad)
df.columns = df.columns.str.strip().str.upper()

# Selección de variables relevantes
df = df[[
    "TÉCNICA", "SISTEMA", "FECHA OBSERVACION",
    "EQUIPO", "COMPONENTE", "CONDICION", "ESTATUS"
]]

# Limpieza de valores nulos
df = df.dropna(subset=["CONDICION"])

# Rellenar categóricas
df = df.fillna("DESCONOCIDO")

# Estandarización de texto
for col in df.columns:
    if df[col].dtype == "object":
        df[col] = df[col].str.upper().str.strip()

# Corrección específica
df["CONDICION"] = df["CONDICION"].str.replace("CRÍTICO", "CRITICO")

# Conversión de fecha
df["FECHA OBSERVACION"] = pd.to_datetime(df["FECHA OBSERVACION"], errors="coerce")

# Eliminar fechas inválidas
df = df.dropna(subset=["FECHA OBSERVACION"])

# Ordenamiento
df = df.sort_values(by=["EQUIPO", "COMPONENTE", "FECHA OBSERVACION"])

df

```

	TÉCNICA	SISTEMA	FECHA OBSERVACION	EQUIPO	COMPONENTE	CONDICION	ESTATUS
1777	VIBRACIONES	MINERALES F	2025-08-30	ALIMENTADOR 02	MOTOR	SEGUIMIENTO	ATENDIDO
1780	VIBRACIONES	MINERALES F	2025-08-30	ALIMENTADOR 02	POLEA MOTRIZ	NORMAL	ATENDIDO
1775	VIBRACIONES	MINERALES F	2025-08-30	ALIMENTADOR 02	REDUCTOR CÓNICO	SEGUIMIENTO	NO ATENDIDO
1779	VIBRACIONES	MINERALES F	2025-08-30	ALIMENTADOR 02	REDUCTOR PLANETARIO	NORMAL	ATENDIDO
146	ACEITE	MINERALES F	2024-09-24	ALIMENTADOR 11	REDUCTOR PLANETARIO	CRÍTICO	ATENDIDO
---	---	---	---	---	---	---	---
1086	VIBRACIONES	MINERALES C	2025-05-30	WINCHE SHIPLOADER	REDUCTOR	NORMAL	ATENDIDO
1245	VIBRACIONES	MINERALES C	2025-08-30	WINCHE SHIPLOADER	REDUCTOR	NORMAL	ATENDIDO
1743	VIBRACIONES	MINERALES C	2025-08-12	WINCHE SHIPLOADER	REDUCTOR	NORMAL	ATENDIDO
1918	VIBRACIONES	MINERALES C	2025-09-25	WINCHE SHIPLOADER	REDUCTOR	NORMAL	ATENDIDO
2008	VIBRACIONES	MINERALES C	2025-10-28	WINCHE SHIPLOADER	REDUCTOR	NORMAL	ATENDIDO

2123 rows x 7 columns

Nota: Elaboración propia en Google Colab

Anexo 5. Código completo de Google Colab #2

FASE 2: VARIABLE OBJETIVO (PREDICCIÓN FUTURA)									
<pre># Crear condición futura df["CONDICION_SIGUIENTE"] = df.groupby(["EQUIPO", "COMPONENTE"])[["CONDICION"]].shift(-1) # Eliminar últimos registros (sin futuro) df = df.dropna(subset=["CONDICION_SIGUIENTE"])</pre>									
df									
	TÉCNICA	SISTEMA	FECHA	OBSERVACION	EQUIPO	COMPONENTE	CONDICION	ESTATUS	CONDICION_SIGUIENTE
737	ACEITE	EQ MOTORIZADO	2025-02-27		BARREDORA B 181	MOTOR	SEGUIMIENTO	NO ATENDIDO	NORMAL
805	ACEITE	EQ MOTORIZADO	2025-03-18		CAMION BARANDA 8TN V7W-778	MOTOR	CRÍTICO	ATENDIDO	SEGUIMIENTO
1041	ACEITE	EQ MOTORIZADO	2025-05-29		CAMION BARANDA 8TN V7W-778	MOTOR	SEGUIMIENTO	ATENDIDO	CRÍTICO
1109	ACEITE	EQ MOTORIZADO	2025-08-05		CAMION BARANDA VCG902	MOTOR	NORMAL	ATENDIDO	NORMAL
527	ACEITE	EQ MOTORIZADO	2024-12-13		CAMIÓN CIST 31.320 DC AUP-773	MOTOR	SEGUIMIENTO	ATENDIDO	SEGUIMIENTO
...	...	...	...	...	...	...	...	...	...
962	VIBRACIONES	MINERALES C	2025-04-30		WINCHE SHIPLOADER	REDUCTOR	NORMAL	ATENDIDO	NORMAL
1086	VIBRACIONES	MINERALES C	2025-05-30		WINCHE SHIPLOADER	REDUCTOR	NORMAL	ATENDIDO	NORMAL
1245	VIBRACIONES	MINERALES C	2025-08-30		WINCHE SHIPLOADER	REDUCTOR	NORMAL	ATENDIDO	NORMAL
1749	VIBRACIONES	MINERALES C	2025-08-12		WINCHE SHIPLOADER	REDUCTOR	NORMAL	ATENDIDO	NORMAL
1918	VIBRACIONES	MINERALES C	2025-09-25		WINCHE SHIPLOADER	REDUCTOR	NORMAL	ATENDIDO	NORMAL
926 rows x 10 columns									

Nota: Elaboración propia en Google Colab

Anexo 6. Código completo de Google Colab #3

FASE 3: FEATURE ENGINEERING

```
# Variables temporales
df["AÑO"] = df["FECHA OBSERVACION"].dt.year
df["MES"] = df["FECHA OBSERVACION"].dt.month

# Tiempo entre inspecciones
df["DIAS_DESDE_ULTIMA"] = df.groupby(["EQUIPO", "COMPONENTE"])["FECHA OBSERVACION"].diff().dt.days
df["DIAS_DESDE_ULTIMA"] = df["DIAS_DESDE_ULTIMA"].fillna(0)

# Condición previa
df["CONDICION_PREVIA"] = df.groupby(["EQUIPO", "COMPONENTE"])["CONDICION"].shift(1)
df["CONDICION_PREVIA"] = df["CONDICION_PREVIA"].fillna("SIN_HISTORIAL")

# Conteos acumulados
df["NUM_INSPECCIONES"] = df.groupby(["EQUIPO", "COMPONENTE"]).cumcount()

# Conteo acumulado por condición
df["NUM_CRITICO"] = (df["CONDICION"] == "CRÍTICO").groupby([df["EQUIPO"], df["COMPONENTE"]]).cumsum()
df["NUM_SEGUIMIENTO"] = (df["CONDICION"] == "SEGUIMIENTO").groupby([df["EQUIPO"], df["COMPONENTE"]]).cumsum()
df["NUM_NORMAL"] = (df["CONDICION"] == "NORMAL").groupby([df["EQUIPO"], df["COMPONENTE"]]).cumsum()
```

Nota: Elaboración propia en Google Colab

Anexo 7. Código completo de Google Colab #4

FASE 4: CODIFICACIÓN

```
from sklearn.preprocessing import LabelEncoder

# Variable objetivo
le = LabelEncoder()
df["TARGET"] = le.fit_transform(df["CONDICION_SIGUIENTE"])

# One-hot encoding
df_model = pd.get_dummies(df.drop(columns=["CONDICION_SIGUIENTE", "FECHA OBSERVACION"]), drop_first=True)
```

FASE 5: DIVISIÓN TEMPORAL

```
# Ordenar por fecha (ya está, pero reforzamos)
df_model = df_model.sort_values(by="AÑO")

# División 70/30
split_index = int(len(df_model) * 0.7)

train = df_model.iloc[:split_index]
test = df_model.iloc[split_index:]

X_train = train.drop(columns=["TARGET"])
y_train = train["TARGET"]

X_test = test.drop(columns=["TARGET"])
y_test = test["TARGET"]
```

Fuente: Elaboración propia en Google Colab

Anexo 8. Código completo de Google Colab #5

FASE 6: ENTRENAMIENTO

```
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier

# Modelo 1: Random Forest
rf = RandomForestClassifier(n_estimators=100, random_state=42)
rf.fit(X_train, y_train)

# Modelo 2: XGBoost
xgb = XGBClassifier(use_label_encoder=False, eval_metric='mlogloss')
xgb.fit(X_train, y_train)
```

```
*** /usr/local/lib/python3.12/dist-packages/xgboost/training.py:200: UserWarning: [22:43:17] WARNING: /_w/xgboost/xgboost/src/learner.cc:782:
Parameters: { "use label encoder" } are not used.
```

```
bst.update(dtrain, iteration=i, fobj=obj)
```

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
               colsample_bylevel=None, colsample_bynode=None,
               colsample_bytree=None, device=None, early_stopping_rounds=None,
               enable_categorical=False, eval_metric='mlogloss',
               feature_types=None, feature_weights=None, gamma=None,
               grow_policy=None, importance_type=None,
               interaction_constraints=None, learning_rate=None, max_bin=None,
               max_cat_threshold=None, max_cat_to_onehot=None,
               max_delta_step=None, max_depth=None, max_leaves=None,
               min_child_weight=None, missing=nan, monotone_constraints=None,
               multi_strategy=None, n_estimators=None, n_jobs=None,
```

Nota: Elaboración propia en Google Colab

Anexo 9. Código completo de Google Colab #6

```
FASE 7: EVALUACIÓN

from sklearn.metrics import classification_report, confusion_matrix

# Predicciones
y_pred_rf = rf.predict(X_test)
y_pred_xgb = xgb.predict(X_test)

print("Random Forest")
print(classification_report(y_test, y_pred_rf))

print("XGBoost")
print(classification_report(y_test, y_pred_xgb))

# Matriz de confusión
print(confusion_matrix(y_test, y_pred_rf))
```

Random Forest					
		precision	recall	f1-score	support
	0	0.39	0.47	0.43	19
	1	0.79	0.77	0.78	195
	2	0.35	0.36	0.36	64
	accuracy			0.66	278
	macro avg	0.51	0.54	0.52	278
	weighted avg	0.67	0.66	0.66	278

XGBoost					
		precision	recall	f1-score	support
	0	0.28	0.42	0.33	19
	1	0.77	0.88	0.82	195
	2	0.42	0.17	0.24	64
	accuracy			0.69	278
	macro avg	0.49	0.49	0.47	278
	weighted avg	0.66	0.69	0.66	278

```
[[ 9  4  6]
 [ 8 151 36]
 [ 6  35 23]]
```

Nota: Elaboración propia en Google Colab

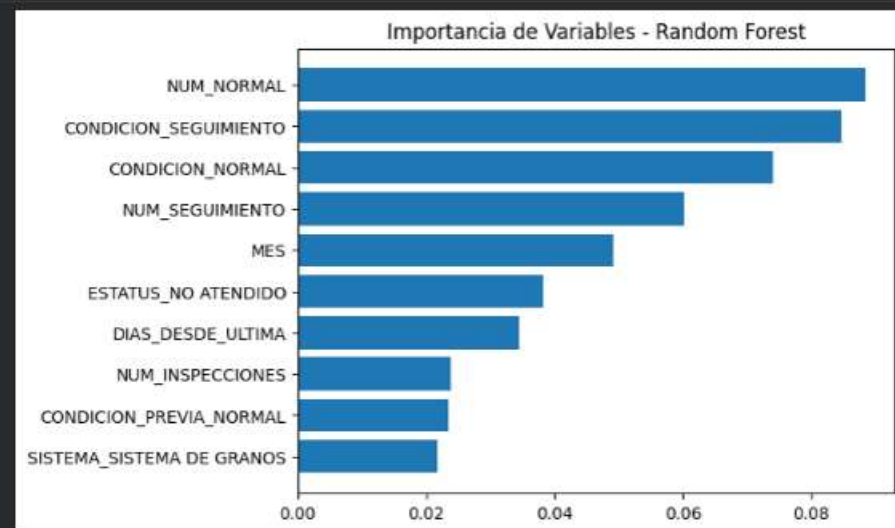
Anexo 10. Código completo de Google Colab #7

FASE 8: IMPORTANCIA DE VARIABLES

```
import matplotlib.pyplot as plt

# Random Forest
importances = rf.feature_importances_
indices = np.argsort(importances)[-10:]

plt.figure()
plt.title("Importancia de Variables - Random Forest")
plt.barh(range(len(indices)), importances[indices])
plt.yticks(range(len(indices)), [X_train.columns[i] for i in indices])
plt.show()
```



FASE 9: PROBABILIDADES

```
# Probabilidades
probs = rf.predict_proba(X_test)

# Convertir a DataFrame
probs_df = pd.DataFrame(probs, columns=le.classes_)

print(probs_df.head())
```

	CRÍTICO	NORMAL	SEGUIMIENTO
0	0.0	0.83	0.17
1	0.0	0.72	0.28
2	0.0	0.56	0.44
3	0.0	0.55	0.45
4	0.0	0.49	0.51

Nota: Elaboración propia en Google Colab

Anexo 11. Código completo de Google Colab #8

```
MODELO OPTIMIZADO -  
  
1 Balanceo de datos: Resolver desbalance (tu principal problema)  
2 Ajuste de hiperparámetros: Mejorar capacidad del modelo  
3 Evaluación enfocada en clase crítica: Alinear con mantenimiento industrial  
  
PASO 1: BALANCEO CON SMOTE  
  
# instalación imbalanced-learn  
!pip install imbalanced-learn  
  
Requirement already satisfied: imbalanced-learn in /usr/local/lib/python3.12/dist-packages (0.14.1)  
Requirement already satisfied: numpy<3,>=1.25.2 in /usr/local/lib/python3.12/dist-packages (from imbalanced-learn) (2.0.2)  
Requirement already satisfied: scipy<2,>=1.11.4 in /usr/local/lib/python3.12/dist-packages (from imbalanced-learn) (1.16.3)  
Requirement already satisfied: scikit-learn<2,>=1.4.2 in /usr/local/lib/python3.12/dist-packages (from imbalanced-learn) (1.6.1)  
Requirement already satisfied: sklearn-compat<0.2,>=0.1.5 in /usr/local/lib/python3.12/dist-packages (from imbalanced-learn) (0.1.5)  
Requirement already satisfied: joblib<2,>=1.2.0 in /usr/local/lib/python3.12/dist-packages (from imbalanced-learn) (1.5.3)  
Requirement already satisfied: threadpoolctl<4,>=2.0.0 in /usr/local/lib/python3.12/dist-packages (from imbalanced-learn) (3.6.0)  
  
from imblearn.over_sampling import SMOTE  
  
# Aplicar SOLO en entrenamiento  
smote = SMOTE(random_state=42)  
  
X_train_bal, y_train_bal = smote.fit_resample(X_train, y_train)  
  
print("Antes:", y_train.value_counts())  
print("Después:", pd.Series(y_train_bal).value_counts())  
  
Antes: TARGET  
1    416  
2    209  
0     23  
Name: count, dtype: int64  
Después: TARGET  
1    416  
2    416  
0    416  
Name: count, dtype: int64
```

Nota: Elaboración propia en Google Colab

Anexo 12. Código completo de Google Colab #9

```
PASO 2: MODELO OPTIMIZADO
Mejorar Random Forest

from sklearn.ensemble import RandomForestClassifier

rf_opt = RandomForestClassifier(
    n_estimators=300,      # más árboles
    max_depth=15,         # controla complejidad
    min_samples_split=5,
    min_samples_leaf=2,
    class_weight="balanced", # clave
    random_state=42,
    n_jobs=-1
)

rf_opt.fit(X_train_bal, y_train_bal)
```

RandomForestClassifier
RandomForestClassifier(class_weight='balanced', max_depth=15, min_samples_leaf=2, min_samples_split=5, n_estimators=300, n_jobs=-1, random_state=42)

```
PASO 3: EVALUACIÓN NUEVA

from sklearn.metrics import classification_report, confusion_matrix

y_pred_opt = rf_opt.predict(X_test)

print("Random Forest Optimizado")
print(classification_report(y_test, y_pred_opt))

print("Matriz de confusión")
print(confusion_matrix(y_test, y_pred_opt))
```

Random Forest Optimizado	precision	recall	f1-score	support
0	0.32	0.53	0.40	19
1	0.87	0.71	0.78	195
2	0.41	0.56	0.48	64
accuracy			0.67	278
macro avg	0.54	0.60	0.55	278
weighted avg	0.73	0.67	0.69	278

Matriz de confusión

```
[[ 10  3  6]
 [ 11 139 45]
 [ 10 18 36]]
```

Nota: Elaboración propia en Google Colab

Anexo 13. Código completo de Google Colab #10

MODELO OPTIMIZADO XgBoost

```
from xgboost import XGBClassifier

xgb_opt = XGBClassifier(
    n_estimators=400,
    max_depth=6,
    learning_rate=0.05,
    subsample=0.8,
    colsample_bytree=0.8,
    objective='multi:softprob',
    eval_metric='mlogloss',
    random_state=42
)

xgb_opt.fit(X_train_bal, y_train_bal)

# Evaluación
y_pred_xgb = xgb_opt.predict(X_test)

from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred_xgb))
```

	precision	recall	f1-score	support
0	0.26	0.37	0.30	19
1	0.78	0.84	0.81	195
2	0.36	0.23	0.28	64
accuracy			0.67	278
macro avg	0.47	0.48	0.47	278
weighted avg	0.65	0.67	0.66	278

Fuente: Elaboración propia en Google Colab

Anexo 14. Código completo de Google Colab #11

```
RESULTADOS DEL MODELO DE ML (NIVEL 1)

Paso 1: Crear DataFrame completo

# Probabilidades
probs = rf_opt.predict_proba(X_test)

# DataFrame de probabilidades
probs_df = pd.DataFrame(probs, columns=le.classes_)

# Predicción
predicciones = rf_opt.predict(X_test)

# Convertir a etiquetas originales
predicciones_labels = le.inverse_transform(predicciones)

# DataFrame final
resultados = X_test.copy()
resultados["PREDICCION"] = predicciones_labels

# Agregar probabilidades
for i, col in enumerate(le.classes_):
    resultados[f"PROB_{col}"] = probs[:, i]

Paso 2: Guardar en archivo

resultados["EQUIPO"] = df.loc[X_test.index, "EQUIPO"]
resultados["COMPONENTE"] = df.loc[X_test.index, "COMPONENTE"]
resultados["FECHA"] = df.loc[X_test.index, "FECHA OBSERVACION"]

# Archivo excel
resultados.to_excel("resultados_modelo.xlsx", index=False)

# Archivo csv
resultados.to_csv("resultados_modelo.csv", index=False)
```

Nota: Elaboración propia en Google Colab

Anexo 15. Código completo de Google Colab #12

```
Paso 2: Guardar en archivo

resultados["EQUIPO"] = df.loc[X_test.index, "EQUIPO"]
resultados["COMPONENTE"] = df.loc[X_test.index, "COMPONENTE"]
resultados["FECHA"] = df.loc[X_test.index, "FECHA OBSERVACION"]

# Archivo excel
resultados.to_excel("resultados_modelo.xlsx", index=False)

# Archivo csv
resultados.to_csv("resultados_modelo.csv", index=False)

Selección Óptima (nivel 2)

import pandas as pd

# Load file again
df = pd.read_excel('/mnt/data/resultados_modelo_nivel2.xlsx')

# Constraint
max_time = 40

# Calculate ratio
df['RATIO'] = df['BENEFICIO'] / df['TIEMPO_INTERVENCION']

df_sorted = df.sort_values(by='RATIO', ascending=False)

selected_rows = []
total_time = 0

for _, row in df_sorted.iterrows():
    if total_time + row['TIEMPO_INTERVENCION'] <= max_time:
        selected_rows.append(row)
        total_time += row['TIEMPO_INTERVENCION']

selected_df = pd.DataFrame(selected_rows)

# Save file
output_path = '/mnt/data/seleccion_optima.xlsx'
selected_df.to_excel(output_path, index=False)

output_path
```

Nota: Elaboración propia en Google Colab

Anexo 16. Código completo de Google Colab #13

Optimización con Programación Dinámica

```
import pandas as pd
import numpy as np
from pathlib import Path

input_path = Path("/mnt/data/resultados_modelo_nivel2.xlsx")
df = pd.read_excel(input_path)

# Normalize column names for robust matching
original_cols = list(df.columns)
norm_map = {c: c.strip().upper().replace(" ", "_") for c in df.columns}
df = df.rename(columns=norm_map)

# Required columns
time_col = None
benefit_col = None
for c in df.columns:
    if c in {"TIEMPO_INTERVENCION", "TIEMPO_DE_INTERVENCION"}:
        time_col = c
    if c == "BENEFICIO":
        benefit_col = c

if time_col is None or benefit_col is None:
    raise ValueError(f"No se encontraron las columnas requeridas. Columnas disponibles: {list(df.columns)}")

# Clean numeric fields
df[time_col] = pd.to_numeric(df[time_col], errors="coerce")
df[benefit_col] = pd.to_numeric(df[benefit_col], errors="coerce")
df = df.dropna(subset=[time_col, benefit_col]).copy()

# 0/1 Knapsack exact optimization with time constraint = 40 hours
capacity = 40
times = df[time_col].astype(int).to_list()
benefits = df[benefit_col].astype(float).to_list()
n = len(df)

# Dynamic programming table
dp = np.zeros((n + 1, capacity + 1), dtype=float)
```

```
for i in range(1, n + 1):
    wt = times[i - 1]
    val = benefits[i - 1]
    for w in range(capacity + 1):
        dp[i, w] = dp[i - 1, w]
        if wt <= w:
            cand = dp[i - 1, w - wt] + val
            if cand > dp[i, w]:
                dp[i, w] = cand

# Reconstruct chosen items
selected = [0] * n
w = capacity
for i in range(n, 0, -1):
    if dp[i, w] != dp[i - 1, w]:
        selected[i - 1] = 1
        w -= times[i - 1]

df["SELECCION_PL"] = selected
df["RATIO_BENEFICIO_TIEMPO"] = df[benefit_col] / df[time_col]

selected_df = df[df["SELECCION_PL"] == 1].copy()
selected_df = selected_df.sort_values(by=[benefit_col, "RATIO_BENEFICIO_TIEMPO"], ascending=[False, False])

summary = pd.DataFrame({
    "Indicador": [
        "Capacidad de tiempo (horas)",
        "Tiempo total utilizado (horas)",
        "Tiempo restante (horas)",
        "Beneficio total maximizado",
        "Número de equipos seleccionados"
    ],
    "Valor": [
        capacity,
        float(selected_df[time_col].sum()),
        float(capacity - selected_df[time_col].sum()),
        float(selected_df[benefit_col].sum()),
        int(selected_df.shape[0])
    ]
})

# Save workbook with 3 sheets
output_path = Path("/mnt/data/optimizacion_programacion_lineal.xlsx")
with pd.ExcelWriter(output_path, engine="openpyxl") as writer:
    summary.to_excel(writer, sheet_name="Resumen", index=False)
    selected_df.to_excel(writer, sheet_name="Seleccionados", index=False)
    df.to_excel(writer, sheet_name="Base_con_seleccion", index=False)

print(str(output_path))
print(summary)
```

Nota: Elaboración propia en Google Colab

Anexo 17. Código completo de Google Colab #14

Comparación del método tradicional vs modelo propuesto (nivel 3)

```
import pandas as pd
import numpy as np

# Cargar archivo con nivel 2
df = pd.read_excel("resultados_modelo_nivel2.xlsx")

# Normalizar nombres de columnas
df.columns = [c.strip().upper().replace(" ", "_") for c in df.columns]

# Verificar columnas mínimas
required = ["BENEFICIO", "TIEMPO_INTERVENCIÓN"]
missing = [c for c in required if c not in df.columns]
if missing:
    raise ValueError(f"Faltan columnas requeridas: {missing}")

# Elegir criterio para simular el método tradicional
manual_priority_col = None
for cand in ["PREDICCIÓN", "PREDICCIÓN", "CONDICIÓN SIGUIENTE", "CONDICIÓN SIGUIENTE", "ÍNDICE CRITICIDAD", "ÍNDICE CRITICIDAD"]:
    cand_norm = cand.strip().upper().replace(" ", "_")
    if cand_norm in df.columns:
        manual_priority_col = cand_norm
        break

manual_df = df.copy()

# Método tradicional: prioriza por clasificación o criticidad
if manual_priority_col in ["PREDICCIÓN", "PREDICCIÓN", "CONDICIÓN SIGUIENTE", "CONDICIÓN SIGUIENTE"]:
    order_map = {"CRÍTICO": 3, "CRÍTICO": 3, "SEGUIMIENTO": 2, "NORMAL": 1}
    manual_df["_PRIORIDAD_MANUAL"] = (
        manual_df[manual_priority_col].astype(str).str.upper().str.strip().map(order_map).fillna(0)
    )
    manual_df = manual_df.sort_values(by=["_PRIORIDAD_MANUAL", "BENEFICIO"], ascending=[False, False])
elif manual_priority_col in ["ÍNDICE CRITICIDAD", "ÍNDICE CRITICIDAD"]:
    manual_df = manual_df.sort_values(by=[manual_priority_col, "BENEFICIO"], ascending=[False, False])
else:
    manual_df = manual_df.sort_values(by=["BENEFICIO"], ascending=[False])

# Restricción de tiempo
capacity = 40

# Selección manual
selected_manual = []
time_acc = 0

for _, row in manual_df.iterrows():
    t = float(row["TIEMPO_INTERVENCIÓN"])
    if time_acc + t <= capacity:
        selected_manual.append(row)
        time_acc += t

manual_selected_df = pd.DataFrame(selected_manual).copy()
if not manual_selected_df.empty:
    manual_selected_df["SELECCIÓN_MANUAL"] = 1

# Optimización exacta (mochila 0-1)
opt_df = df.copy()
times = opt_df["TIEMPO_INTERVENCIÓN"].astype(int).tolist()
benefits = opt_df["BENEFICIO"].astype(float).tolist()
n = len(opt_df)

dp = np.zeros((n + 1, capacity + 1), dtype=float)

for i in range(1, n + 1):
    wt = times[i - 1]
    val = benefits[i - 1]
    for w in range(capacity + 1):
        dp[i, w] = dp[i - 1, w]
        if wt <= w:
            cand = dp[i - 1, w - wt] + val
            if cand > dp[i, w]:
                dp[i, w] = cand

selected = [0] * n
w = capacity
for i in range(n, 0, -1):
    if dp[i, w] != dp[i - 1, w]:
        selected[i - 1] = 1
        w -= times[i - 1]

opt_df["SELECCIÓN_OPTIMA"] = selected
opt_selected_df = opt_df[opt_df["SELECCIÓN_OPTIMA"] == 1].copy()

# Resumen comparativo
benef_manual = float(manual_selected_df["BENEFICIO"].sum()) if not manual_selected_df.empty else 0.0
time_manual = float(manual_selected_df["TIEMPO_INTERVENCIÓN"].sum()) if not manual_selected_df.empty else 0.0
benef_opt = float(opt_selected_df["BENEFICIO"].sum()) if not opt_selected_df.empty else 0.0
time_opt = float(opt_selected_df["TIEMPO_INTERVENCIÓN"].sum()) if not opt_selected_df.empty else 0.0

comparison = pd.DataFrame({
    "Método": ["Tradicional/manual", "Propuesto optimizado"],
    "Equipos seleccionados": [int(len(manual_selected_df)), int(len(opt_selected_df))],
    "Tiempo total horas": [time_manual, time_opt],
    "Beneficio total": [benef_manual, benef_opt],
    "Beneficio por hora": [
        round(benef_manual / time_manual, 4) if time_manual else 0.0,
        round(benef_opt / time_opt, 4) if time_opt else 0.0
    ]
})

# Guardar resultados
with pd.ExcelWriter("comparacion_nivel3.xlsx", engine="openpyxl") as writer:
    comparison.to_excel(writer, sheet_name="Resumen comparacion", index=False)
    manual_selected_df.to_excel(writer, sheet_name="Metodo tradicional", index=False)
    opt_selected_df.to_excel(writer, sheet_name="Metodo propuesto", index=False)

print(comparison)
```

Nota: Elaboración propia en Google Colab

Anexo 18. Fase 1: Limpieza y Preprocesamiento

```
import pandas as pd
import numpy as np

# Cargar datos
df = pd.read_excel("DataHistorica.xlsx", sheet_name="alertas tempranas_PD")

# Renombrar columnas (opcional para facilidad)
df.columns = df.columns.str.strip().str.upper()

# Selección de variables relevantes
df = df[[
    "TÉCNICA", "SISTEMA", "FECHA OBSERVACION",
    "EQUIPO", "COMPONENTE", "CONDICION", "ESTATUS"
]]

# Limpieza de valores nulos
df = df.dropna(subset=["CONDICION"])

# Rellenar categóricas
df = df.fillna("DESCONOCIDO")

# Estandarización de texto
for col in df.columns:
    if df[col].dtype == "object":
        df[col] = df[col].str.upper().str.strip()

# Corrección específica
df["CONDICION"] = df["CONDICION"].str.replace("CRITICO", "CRÍTICO")

# Conversión de fecha
df["FECHA OBSERVACION"] = pd.to_datetime(df["FECHA OBSERVACION"], errors="coerce")

# Eliminar fechas inválidas
df = df.dropna(subset=["FECHA OBSERVACION"])

# Ordenamiento
df = df.sort_values(by=["EQUIPO", "COMPONENTE", "FECHA OBSERVACION"])
```

Nota: Elaboración propia en Google Colab

Anexo 19. Fase 2: Variable Objetivo

```
# Crear condición futura
df["CONDICION_SIGUIENTE"] = df.groupby(["EQUIPO", "COMPONENTE"])["CONDICION"].shift(-1)

# Eliminar últimos registros (sin futuro)
df = df.dropna(subset=["CONDICION_SIGUIENTE"])
```

Nota: Elaboración propia en Google Colab

Anexo 20. Fase 3: Feature Engineering

```
# Variables temporales
df["AÑO"] = df["FECHA OBSERVACION"].dt.year
df["MES"] = df["FECHA OBSERVACION"].dt.month

# Tiempo entre inspecciones
df["DIAS_DESDE_ULTIMA"] = df.groupby(["EQUIPO", "COMPONENTE"])[["FECHA OBSERVACION"]].diff().dt.days
df["DIAS_DESDE_ULTIMA"] = df["DIAS_DESDE_ULTIMA"].fillna(0)

# Condición previa
df["CONDICION_PREVIA"] = df.groupby(["EQUIPO", "COMPONENTE"])[["CONDICION"]].shift(1)
df["CONDICION_PREVIA"] = df["CONDICION_PREVIA"].fillna("SIN_HISTORIAL")

# Conteos acumulados
df["NUM_INSPECCIONES"] = df.groupby(["EQUIPO", "COMPONENTE"]).cumcount()

# Conteo acumulado por condición
df["NUM_CRITICO"] = (df["CONDICION"] == "CRÍTICO").groupby([df["EQUIPO"], df["COMPONENTE"]]).cumsum()
df["NUM_SEGUIMIENTO"] = (df["CONDICION"] == "SEGUIMIENTO").groupby([df["EQUIPO"], df["COMPONENTE"]]).cumsum()
df["NUM_NORMAL"] = (df["CONDICION"] == "NORMAL").groupby([df["EQUIPO"], df["COMPONENTE"]]).cumsum()
```

Nota: Elaboración propia en Google Colab

Anexo 21. Fase 4: Codificación

```
from sklearn.preprocessing import LabelEncoder

# Variable objetivo
le = LabelEncoder()
df["TARGET"] = le.fit_transform(df["CONDICION_SIGUIENTE"])

# One-hot encoding
df_model = pd.get_dummies(df.drop(columns=["CONDICION_SIGUIENTE", "FECHA OBSERVACION"]), drop_first=True)
```

Nota: Elaboración propia en Google Colab

Anexo 22. Fase 5: División Temporal

```
# Ordenar por fecha (ya está, pero reforzamos)
df_model = df_model.sort_values(by="AÑO")

# División 70/30
split_index = int(len(df_model) * 0.7)

train = df_model.iloc[:split_index]
test = df_model.iloc[split_index:]

X_train = train.drop(columns=["TARGET"])
y_train = train["TARGET"]

X_test = test.drop(columns=["TARGET"])
y_test = test["TARGET"]
```

Nota: Elaboración propia en Google Colab

Anexo 23. Entrenamiento

```
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier

# Modelo 1: Random Forest
rf = RandomForestClassifier(n_estimators=100, random_state=42)
rf.fit(X_train, y_train)

# Modelo 2: XGBoost
xgb = XGBClassifier(use_label_encoder=False, eval_metric='mlogloss')
xgb.fit(X_train, y_train)
```

Nota: Elaboración propia en Google Colab

Anexo 24. Fase 7: Evaluación

```
from sklearn.metrics import classification_report, confusion_matrix

# Predicciones
y_pred_rf = rf.predict(X_test)
y_pred_xgb = xgb.predict(X_test)

print("Random Forest")
print(classification_report(y_test, y_pred_rf))

print("XGBoost")
print(classification_report(y_test, y_pred_xgb))

# Matriz de confusión
print(confusion_matrix(y_test, y_pred_rf))
```

Random Forest					
	precision	recall	f1-score	support	
0	0.39	0.47	0.43	19	
1	0.79	0.77	0.78	195	
2	0.35	0.36	0.36	64	
accuracy			0.66	278	
macro avg	0.51	0.54	0.52	278	
weighted avg	0.67	0.66	0.66	278	

XGBoost					
	precision	recall	f1-score	support	
0	0.28	0.42	0.33	19	
1	0.77	0.88	0.82	195	
2	0.42	0.17	0.24	64	
accuracy			0.69	278	
macro avg	0.49	0.49	0.47	278	
weighted avg	0.66	0.69	0.66	278	

```
[[ 9  4  6]
 [ 8 151 36]
 [ 6  35 23]]
```

Nota: Elaboración propia en Google Colab

Anexo 25. Fase 8: Importancia de variables

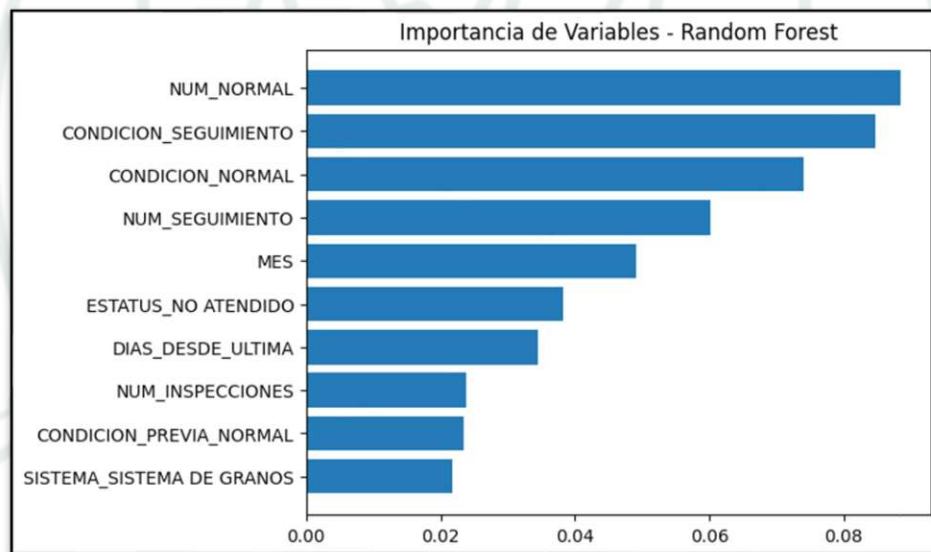
```
import matplotlib.pyplot as plt

# Random Forest
importances = rf.feature_importances_
indices = np.argsort(importances)[-10:]

plt.figure()
plt.title("Importancia de Variables - Random Forest")
plt.barh(range(len(indices)), importances[indices])
plt.yticks(range(len(indices)), [X_train.columns[i] for i in indices])
plt.show()
```

Nota: Elaboración propia en Google Colab

Anexo 26. Gráfica de la Importancia de Variables



Nota: Elaboración propia en Google Colab

Anexo 27. Fase 9: Probabilidades

```
# Probabilidades
probs = rf.predict_proba(X_test)

# Convertir a DataFrame
probs_df = pd.DataFrame(probs, columns=le.classes_)

print(probs_df.head())
```

Nota: Elaboración propia en Google Colab

Anexo 28. Balanceo de Técnica SMOTE

```
from imblearn.over_sampling import SMOTE

# Aplicar SOLO en entrenamiento
smote = SMOTE(random_state=42)

X_train_bal, y_train_bal = smote.fit_resample(X_train, y_train)

print("Antes:", y_train.value_counts())
print("Después:", pd.Series(y_train_bal).value_counts())
```

Antes: TARGET
1 416
2 209
0 23
Name: count, dtype: int64

Después: TARGET
1 416
2 416
0 416
Name: count, dtype: int64

Nota: Elaboración propia en Google Colab

Anexo 29. Modelo Optimizado con SMOTE en Random Forest

```
from sklearn.ensemble import RandomForestClassifier

rf_opt = RandomForestClassifier(
    n_estimators=300,      # más árboles
    max_depth=15,         # controla complejidad
    min_samples_split=5,
    min_samples_leaf=2,
    class_weight="balanced", # clave
    random_state=42,
    n_jobs=-1
)

rf_opt.fit(X_train_bal, y_train_bal)
```

RandomForestClassifier

```
RandomForestClassifier(class_weight='balanced', max_depth=15,
                        min_samples_leaf=2, min_samples_split=5,
                        n_estimators=300, n_jobs=-1, random_state=42)
```

Nota: Elaboración propia en Google Colab

Anexo 30. Evaluación de SMOTE y Random Forest

```
from sklearn.metrics import classification_report, confusion_matrix

y_pred_opt = rf_opt.predict(X_test)

print("Random Forest Optimizado")
print(classification_report(y_test, y_pred_opt))

print("Matriz de confusión")
print(confusion_matrix(y_test, y_pred_opt))
```

Random Forest Optimizado	precision	recall	f1-score	support
0	0.32	0.53	0.40	19
1	0.87	0.71	0.78	195
2	0.41	0.56	0.48	64
accuracy			0.67	278
macro avg	0.54	0.60	0.55	278
weighted avg	0.73	0.67	0.69	278

Matriz de confusión

```
[[ 10  3  6]
 [ 11 139 45]
 [ 10 18 36]]
```

Nota: Elaboración propia en Google Colab

Anexo 31. Modelo Optimizado con SMOTE en XGBoost

```
from xgboost import XGBClassifier

xgb_opt = XGBClassifier(
    n_estimators=400,
    max_depth=6,
    learning_rate=0.05,
    subsample=0.8,
    colsample_bytree=0.8,
    objective='multi:softprob',
    eval_metric='mlogloss',
    random_state=42
)

xgb_opt.fit(X_train_bal, y_train_bal)

# Evaluación
y_pred_xgb = xgb_opt.predict(X_test)

from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred_xgb))
```

	precision	recall	f1-score	support
0	0.26	0.37	0.30	19
1	0.78	0.84	0.81	195
2	0.36	0.23	0.28	64
accuracy			0.67	278
macro avg	0.47	0.48	0.47	278
weighted avg	0.65	0.67	0.66	278

Nota: Elaboración propia en Google Colab

Anexo 32. Selección Óptima (Nivel 2)

```
import pandas as pd

# Load file again
df = pd.read_excel('/mnt/data/resultados_modelo_nivel2.xlsx')

# Constraint
max_time = 40

# Calculate ratio
df['RATIO'] = df['BENEFICIO'] / df['TIEMPO_INTERVENCION']

df_sorted = df.sort_values(by='RATIO', ascending=False)

selected_rows = []
total_time = 0

for _, row in df_sorted.iterrows():
    if total_time + row['TIEMPO_INTERVENCION'] <= max_time:
        selected_rows.append(row)
        total_time += row['TIEMPO_INTERVENCION']

selected_df = pd.DataFrame(selected_rows)
```

Nota: Elaboración propia en Google Colab

Anexo 33. Programación Dinámica

```
# Dynamic programming table
dp = np.zeros((n + 1, capacity + 1), dtype=float)

for i in range(1, n + 1):
    wt = times[i - 1]
    val = benefits[i - 1]
    for w in range(capacity + 1):
        dp[i, w] = dp[i - 1, w]
        if wt <= w:
            cand = dp[i - 1, w - wt] + val
            if cand > dp[i, w]:
                dp[i, w] = cand
```

Nota: Elaboración propia en Google Colab

Anexo 34. Comparación Método Tradicional vs Modelo Propuesto

```
# Resumen comparativo
benef_manual = float(manual_selected_df["BENEFICIO"].sum()) if not manual_selected_df.empty else 0.0
time_manual = float(manual_selected_df["TIEMPO_INTERVENCION"].sum()) if not manual_selected_df.empty else 0.0
benef_opt = float(opt_selected_df["BENEFICIO"].sum()) if not opt_selected_df.empty else 0.0
time_opt = float(opt_selected_df["TIEMPO_INTERVENCION"].sum()) if not opt_selected_df.empty else 0.0

comparison = pd.DataFrame({
    "Metodo": ["Tradicional/manual", "Propuesto optimizado"],
    "Equipos_seleccionados": [int(len(manual_selected_df)), int(len(opt_selected_df))],
    "Tiempo_total_horas": [time_manual, time_opt],
    "Beneficio_total": [benef_manual, benef_opt],
    "Beneficio_por_hora": [
        round(benef_manual / time_manual, 4) if time_manual else 0.0,
        round(benef_opt / time_opt, 4) if time_opt else 0.0
    ]
})
```

Nota: Elaboración propia en Google Colab

