



UNIVERSIDAD CÉSAR VALLEJO

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

**Modelo *Machine Learning* con Transformación Logarítmica y
Validación Cruzada para estimar la confiabilidad en un Sistema
de Molienda de Cemento**

**TESIS PARA OBTENER EL GRADO ACADÉMICO DE:
Ingeniero Mecánico Electricista**

AUTORES:

Angulo Gutierrez, Vicente Manuel (orcid.org/0000-0002-4077-0371)

Jara Flores, Jose Andres (orcid.org/0000-0003-1444-4887)

ASESORA:

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

LÍNEA DE INVESTIGACIÓN:

Sistema y Planes de Mantenimiento

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo sostenible y adaptación al cambio climático

TRUJILLO – PERÚ

2023

Dedicatoria

A mis queridos padres, vuestra fe inquebrantable y apoyo constante han guiado mi camino académico. Esta tesis es un tributo a vuestro amor y sacrificio.

A mi esposa e hijos, ustedes son mi mayor motivación y espero que esta tesis inspire vuestro amor por el aprendizaje y la perseverancia.

Vicente Angulo

Dedicado a la memoria de mi amado padre, cuya inspiración y orientación perduran desde el cielo. A mi querida madre, símbolo de apoyo inquebrantable y ejemplo de fortaleza y determinación. Esta tesis es un tributo a ustedes.

A mis queridos hermanos, gracias por estar presentes en cada paso de este camino. Su constante confianza en mí y su apoyo incondicional han sido el pilar fundamental de mi éxito académico.

José Jara

Agradecimiento

Queremos expresar nuestro sincero agradecimiento a la Universidad César Vallejo por ser la plataforma que hizo posible nuestro crecimiento académico y profesional. Esta institución no solo me proporcionó una educación de calidad, sino que también nos brindó oportunidades invaluable para desarrollar habilidades cruciales para nuestro futuro profesional. Agradecemos a nuestra asesora Mg. Miriam Serrepe por su guía invaluable, apoyo continuo y dedicación durante el desarrollo de esta tesis. Su conocimiento, orientación y compromiso han sido fundamentales en cada paso de este proceso académico. Agradecemos a todos los docentes y personal administrativo que nos brindaron apoyo en este camino, así como a nuestros asesores que enriquecieron nuestra experiencia. Estamos orgullosos de ser parte de esta comunidad académica y esperamos desarrollar un excelente camino profesional gracias a los conocimientos adquiridos y puestos en práctica.

Declaratoria de autenticidad del asesor



UNIVERSIDAD CÉSAR VALLEJO

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

Declaratoria de Autenticidad del Asesor

Yo, SERREPE RANNO MIRIAM MARCELA, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA MECÁNICA ELÉCTRICA de la UNIVERSIDAD CÉSAR VALLEJO SAC - TRUJILLO, asesor de Tesis titulada: "Modelo de Machine Learning con Transformación Logarítmica y Validación Cruzada para Estimar la Confiabilidad en un Sistema de Molienda de Cemento", cuyos autores son JARA FLORES JOSE ANDRES, ANGULO GUTIERREZ VICENTE MANUEL, constato que la investigación tiene un índice de similitud de 9.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

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

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

TRUJILLO, 12 de Diciembre del 2023

| Apellidos y Nombres del Asesor: | Firma |
|---|--|
| SERREPE RANNO MIRIAM MARCELA DNI: 06437594 ORCID: 0000-0001-9342-1717 | Firmado electrónicamente por: SRANNOMM el 14- 12-2023 09:40:55 |

Código documento Trilce: TRI - 0694053



Declaratoria de originalidad de los autores



UNIVERSIDAD CÉSAR VALLEJO

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

Declaratoria de Originalidad de los Autores

Nosotros, JARA FLORES JOSE ANDRES, ANGULO GUTIERREZ VICENTE MANUEL estudiantes de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA MECÁNICA ELÉCTRICA de la UNIVERSIDAD CÉSAR VALLEJO SAC - TRUJILLO, declaramos bajo juramento que todos los datos e información que acompañan la Tesis titulada: "Modelo de Machine Learning con Transformación Logarítmica y Validación Cruzada para Estimar la Confiabilidad en un Sistema de Molienda de Cemento", es de nuestra autoría, por lo tanto, declaramos que la Tesis:

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

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

| Nombres y Apellidos | Firma |
|--|--|
| JOSE ANDRES JARA FLORES DNI: 76560080 ORCID: 0000-0003-1444-4887 | Firmado electrónicamente por: JJARAFLO el 12-12-2023 09:12:03 |
| VICENTE MANUEL ANGULO GUTIERREZ DNI: 71104402 ORCID: 0000-0002-4077-0371 | Firmado electrónicamente por: VANGULOGU el 12-12-2023 10:08:04 |

Código documento Trilce: TRI - 0694054



Índice de contenidos

| | |
|--|-----|
| Carátula..... | i |
| Dedicatoria | ii |
| Agradecimiento | iii |
| Declaratoria de autenticidad del asesor | iv |
| Declaratoria de originalidad del autor/autores | v |
| Declaratoria de originalidad del autor/autores | vi |
| Índice de contenidos | 1 |
| Índice de tablas | 2 |
| Índice de figuras | 3 |
| Resumen | 4 |
| Abstract | 5 |
| I. INTRODUCCIÓN | 6 |
| II. MARCO TEÓRICO | 10 |
| III. METODOLOGÍA | 21 |
| 3.1. Tipo y diseño de investigación | 21 |
| 3.2. Variables y operacionalización | 23 |
| 3.3. Población, muestra y muestreo | 23 |
| 3.4. Técnicas e instrumentos de recolección de datos | 25 |
| 3.5. Procedimientos | 26 |
| 3.6. Método de análisis de datos | 26 |
| 3.7. Aspectos éticos | 27 |
| IV. RESULTADOS | 28 |
| 4.1. Análisis de Pareto para identificar equipos críticos en base a tiempo entre fallas. | 28 |
| 4.2. Análisis de Confiabilidad por métodos tradicionales | 30 |
| 4.3. Análisis de Confiabilidad por métodos de ML | 40 |
| 4.4. Comparación de métodos de análisis de distribuciones convencionales y de ML | 45 |
| V. DISCUSIÓN | 49 |
| VI. CONCLUSIONES | 53 |
| VII. RECOMENDACIONES | 55 |
| BIBLIOGRAFÍA | 58 |
| ANEXOS | 64 |

Índice de tablas

| | |
|---|----|
| Tabla 1 Equipos críticos a través de Análisis de Pareto | 28 |
| Tabla 2 Mejor ajuste para los datos de TBF a nivel de equipo por método MLE . | 34 |
| Tabla 3 Confiabilidad del sistema de molienda en diferentes intervalos de tiempo | 36 |
| Tabla 4 Vida útil confiable de Molino y equipos principales | 38 |
| Tabla 5 Mejor ajuste para los datos de TBF a nivel de equipo por métodos de ML | 42 |
| Tabla 6 Confiabilidad basada en el tiempo, basada en el modelo Híbrido 2: Crow_AMSAA + técnica de validación cruzada..... | 43 |
| Tabla 7 Vida útil confiable de Molino y equipos principales Híbrido 2: Crow_AMSAA y técnica de validación cruzada | 44 |
| Tabla 8 Evaluación MSE y R ² para los modelos convencionales y de ML | 47 |
| Tabla 9 Matriz de operacionalización de variables..... | 65 |
| Tabla 10 Instrumento ficha de registro..... | 66 |

Índice de figuras

| | |
|--|----|
| Figura 1 - Esquema de metodología | 22 |
| Figura 2 - Diagrama de Pareto completo..... | 29 |
| Figura 3 - Diagrama de Pareto de equipos críticos | 30 |
| Figura 4 - Diagrama de flujo para análisis por modelo K-S | 31 |
| Figura 5 - Diagrama de flujo para análisis por modelo MLE..... | 32 |
| Figura 6 - Distribución lognormal para el molino unidad total..... | 39 |
| Figura 7 - Diagrama de flujo para análisis por modelo Crow_AMSAA + Validación cruzada Mayor K fold..... | 41 |

Resumen

En el presente trabajo de tesis, se desarrolla y propone un modelo innovador de estimación de la confiabilidad para equipos industriales aplicado a un caso de estudio dentro de un sistema de equipos de molienda de cemento. Este modelo incorpora de aprendizaje automático situándose en la vanguardia de los estudios contemporáneos en el ámbito de la ingeniería de confiabilidad. A través de este enfoque, se busca mejorar la precisión en la estimación de la confiabilidad empleando una metodología que realiza la comparación entre métodos estadísticos convencionales, como los modelos de distribución normal, exponencial, lognormal y Weibull, y la integración del modelo innovador híbrido Crow-AMSAA incluyendo métodos de *machine learning* al ofrecer herramientas analíticas más robustas para la gestión de mantenimiento y prolongación de vida útil de sistemas productivos en la industria del cemento. La métrica de evaluación, precisión y eficacia de cada modelo de estimación de confiabilidad fue realizada por medio de las medidas de Error medio cuadrado (MSE), Error promedio absoluto (MAE), y coeficiente de determinación (R^2). Los resultados de la comparación obtenida mostraron que el modelo híbrido propuesto ofrece estimar la confiabilidad de equipos de forma más precisa. Esta mejora en la precisión asegura que las proyecciones de tiempos entre fallas para la toma de decisiones sean más exactas y estén estrechamente alineados con las condiciones operativas reales del sistema de manufactura industrial.

Palabras clave: Confiabilidad, *machine learning*, tiempo entre fallas, transformación logarítmica, validación cruzada

Abstract

In this thesis, an innovative reliability estimation model for industrial equipment is developed and proposed, focusing specifically on a case study of a cement grinding system. This model incorporates machine learning, positioning it at the forefront of contemporary studies in the field of reliability engineering. This approach aims to improve the accuracy of reliability estimation by using a methodology that compares conventional statistical methods, such as normal, exponential, lognormal and Weibull distribution models, with the integration of the innovative hybrid Crow-AMSAA model incorporating machine learning methods. This provides more robust analytical tools for maintenance management and extending the life of productive systems in the cement industry. The evaluation metrics, precision and effectiveness of each reliability estimation model were performed using Mean Squared Error (MSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2) measures. The results obtained confirm that the proposed hybrid model provides a more accurate reliability estimation. This improvement in precision ensures that the failure time predictions for decision making are more accurate and closer to the real operating conditions of the industrial manufacturing.

Keywords: Reliability, machine Learning, time between failures, logarithmic transformation, cross validation

I. INTRODUCCIÓN

La confiabilidad es un aspecto esencial en la ingeniería y el mantenimiento de los sistemas industriales, que requieren altos niveles de eficiencia (Ahmadi *et al.*, 2019). Para mejorar la confiabilidad, se han desarrollado nuevas técnicas que combinan métodos analíticos clásicos con tecnologías emergentes como el aprendizaje automático y el Internet de las Cosas (IoT¹) (Shojaeinasab *et al.*, 2022) permitiendo una evaluación más detallada y exacta de la fiabilidad de los sistemas (Choudhary *et al.*, 2019). Estos progresos han contribuido tanto al avance teórico del campo como a la aplicación práctica de soluciones innovadoras para la gestión efectiva y el mantenimiento preventivo, garantizando así la calidad y el funcionamiento óptimo de los equipos en diferentes sectores industriales (Xu *et al.*, 2021).

El análisis de confiabilidad se ha transformado con la incorporación de técnicas avanzadas de aprendizaje automático y la adaptación de modelos estadísticos probados como el Proceso No Homogéneo de Poisson (NHPP²) de Duane-Crow-AMSA (Miranda *et al.*, 2023; Peng *et al.*, 2017). Esta técnica, en combinación con el uso de la distribución Weibull, permite una modelización más precisa de la tasa de fallos en sistemas complejos (Attanayake *et al.*, 2023). El modelo de estimación puede ser enriquecido con la aplicación de métodos de *machine learning*, proporcionando un modelo robusto y generalizable que previene el sobreajuste y facilita la interpretación de datos sesgados. Este marco avanzado es esencial para la gestión de mantenimiento y mejora la confiabilidad y seguridad en la industria, marcando un progreso significativo para la aplicación práctica en ingeniería (V. G. Talkit *et al.*, 2022).

La estimación de la confiabilidad en sistemas de industriales se calcula analizando los datos históricos de tiempo entre fallas (TBF³) y tiempo para reparar (TTR⁴). Se emplean modelos estadísticos para ajustar las distribuciones de probabilidad a estos datos, los cuales pueden ser analizados usando software especializado en confiabilidad (Choudhary *et al.*, 2019). Al integrar estos modelos con técnicas de mantenimiento preventivo y predictivo,

¹ IoT: Internet of Things

² NHPP: *Non-Homogeneous Poisson Process*

³ TBF: *Time Between Failure*

⁴ TTR: *Time to Repair*

se mejora la disponibilidad y se optimiza la operación en la industria, conduciendo a una utilización más eficiente de los recursos y a una reducción de las paradas no planificadas (Ahmadi *et al.*, 2019).

El análisis de la confiabilidad se ha convertido en un aspecto crítico en los sistemas industriales modernos, los cuales exigen una mayor robustez, complejidad y disponibilidad (Bakon *et al.*, 2022). La evolución de la industria 4.0 y la gestión del mantenimiento inteligente (Jasiulewicz-Kaczmarek & Gola, 2019; Pinciroli *et al.*, 2023; Shojaeinasab *et al.*, 2022; Silvestri *et al.*, 2020) han propuesto la integración de información física y digital (Bag *et al.*, 2021; Ramesh *et al.*, 2020), mediante el uso de tecnologías respaldadas por esta tendencia, para abordar los desafíos actuales. Mediante la creación de gemelos digitales (Abbate *et al.*, 2022; D'Amico *et al.*, 2022) y la aplicación de la ciencia de datos en el mantenimiento, se busca diagnosticar, pronosticar y predecir de manera anticipada y precisa eventos de riesgo, como fallas y paradas no deseadas, que conllevan a maximizar el rendimiento de los activos y su confiabilidad operativa en las líneas de producción (Diez-Olivan *et al.*, 2019; Zermane *et al.*, 2022).

Las herramientas técnicas y tecnológicas basadas en la industria 4.0, aplicadas al análisis de confiabilidad, aprovechan las tecnologías existentes para anticiparse a las fallas y mejorar la eficiencia de los procesos (Silvestri *et al.*, 2020). Además, permiten adaptabilidad a entornos cambiantes, optimización de costos y proporcionan herramientas valiosas para la toma de decisiones tácticas y estratégicas (Culot *et al.*, 2020). El estudio de la confiabilidad, disponibilidad, mantenibilidad y seguridad (RAMS⁵) (Hoffmann Souza *et al.*, 2020; Kristjanpoller *et al.*, 2021) de los sistemas, subraya la importancia del análisis de confiabilidad mediante el uso de estadísticas y métricas basadas en datos de fallas. Al no aplicar estas técnicas, las industrias se exponen a mayores riesgos de fallas, averías, paradas no programadas, desperdicios y accidentes, que afectan negativamente su rentabilidad y resultados operativos (Guo *et al.*, 2021).

El desarrollo de la gestión de mantenimiento moderno y el análisis de

⁵ RAMS: *Reliability, Availability, Maintainability and Security*

indicadores clave de mantenimiento enfrentan desafíos y barreras (Jasiulewicz-Kaczmarek *et al.*, 2019), como lo es la falta de herramientas digitales para el análisis estadístico de confiabilidad en sistemas de activos de las empresas industriales. Dentro de las causas encontradas, se encuentra la dificultad para la recopilación de bases de datos de calidad (Martón *et al.*, 2023; Theissler *et al.*, 2021), la complejidad de los sistemas industriales (Hoffmann Souza *et al.*, 2020), la poca experiencia profesional del personal, el nivel de cultura de mantenimiento, y nivel organizacional de gestión por indicadores (Gallo *et al.*, 2021; Raj *et al.*, 2020).

El contexto operacional de producción y mantenimiento de las empresas regionales del país depende del grado de madurez y solidez económica, así como de las estrategias directivas que decidan abortar para una gestión efectiva. Así mismo las organizaciones tienen la potestad de centrar sus esfuerzos en aquellos indicadores de producción y mantenimiento, que les permitan tener una visión holística de su giro de negocio; tal es así que la disponibilidad es considerada como un parámetro de control determinante para la operación de plantas cementeras, como lo sostiene Rivera en su investigación respecto a la implementación de un equipo para mejorar la confiabilidad de un equipo dentro de una planta cementera (Rivera, 2019).

Dentro de la línea de producción de cemento, en una empresa en la localidad de Pacasmayo, se han suscitado resultados de baja disponibilidad y fallas recurrentes que comprometen el cumplimiento los objetivos de producción planteados por la organización. Una baja disponibilidad de los activos tiene como principal factor una elevada tasa de fallas y tiempo de fallas (Choudhary *et al.*, 2019) que se producen de manera no programada, las cuales afectan a los sistemas críticos de la planta de molienda. Los elevados tiempos de falla están asociadas a un bajo rendimiento de confiabilidad de los activos, desincronización de las operaciones (Guo *et al.*, 2021; Jasiulewicz-Kaczmarek & Gola, 2019), generando paradas no programadas, paradas con causas repetitivas, pérdidas de producción, altos costos de reparación y tiempos de entrega de equipos extensos (El-Naggar *et al.*, 2022; Fang *et al.*, 2021; Kovacs *et al.*, 2020).

La problemática expuesta plantea la necesidad de formular el siguiente

problema: ¿Cuál es el impacto de la implementación de un modelo *machine learning* con transformación logarítmica y validación cruzada para estimar la confiabilidad de un sistema de molienda de cemento?; y las siguientes preguntas específicas: ¿Cuál es el tiempo entre falla de los equipos del sistema de molienda de cemento en un determinado periodo de análisis?, ¿Cuál es la confiabilidad de los equipos del sistema de molienda de cemento en un determinado periodo de análisis?, ¿Cómo desarrollar un modelo para estimar la confiabilidad utilizando *machine learning* con transformación logarítmica y validación cruzada?, ¿Cuál sería la confiabilidad estimada del sistema de molienda de cemento, luego de implementar el análisis de confiabilidad mediante el uso del modelo *machine learning* con transformación logarítmica y validación cruzada?.

La investigación se justifica desde un enfoque económico, dado que la confiabilidad de los equipos de un sistema de molienda de cemento está relacionada directamente con los costos de operación y mantenimiento, dada la exigente demanda de volúmenes de producción constante, con bajos tiempos de parada. Desde un enfoque social, la capacitación y uso de herramientas avanzadas para el análisis de tiempos de falla y confiabilidad, abre las puertas del desarrollo profesional e incremento del compromiso activo de los profesionales del área de mantenimiento. Desde un enfoque ambiental, una gestión efectiva, centrada en confiabilidad, evitará fallas que atenten contra el medio ambiente, evitando así contaminación de los recursos naturales y medio donde se desarrollen las operaciones industriales.

Para dar respuesta a la problemática, se plantea el siguiente objetivo general: Estimar la confiabilidad empleando un modelo *machine learning* con transformación logarítmica y validación cruzada en un sistema de molienda de cemento. Para lograr el desarrollo del objetivo general, se proyectan los siguientes objetivos específicos: Obtener y analizar los datos históricos de paradas del sistema, empleando técnicas de manejo de datos y análisis de Pareto; desarrollar un modelo de *machine learning* con transformación logarítmica y validación cruzada para estimar la confiabilidad de los equipos seleccionados del sistema de molienda; estimar el TBF y confiabilidad de los equipos seleccionados del sistema de molienda de cemento empleando el

modelo propuesto; comparar el modelo propuesto con los convencionales, y utilizado las métricas de error cuadrático medio (MSE^6), coeficiente de determinación (R^2^7) y error absoluto medio (MAE^8).

Para dar respuesta a la formulación del problema, se plantea la siguiente hipótesis: La implementación de un modelo para estimar la confiabilidad empleando transformación logarítmica y validación cruzada estima con un mayor ajuste la confiabilidad de los equipos de un sistema de molienda de cemento.

II. MARCO TEÓRICO

Los estudios realizados por V. Talkit *et al.*, (2022), se centraron en la aplicación del análisis de confiabilidad en modificaciones de programación de mantenimiento en un contexto no especificado. El objetivo fue optimizar las prácticas de mantenimiento y reducir los tiempos de inactividad. Para lograrlo, se utilizaron datos históricos de fallas y se identificaron los componentes críticos. A través del software Reliasoft Weibull++⁹, se analizaron los datos de TBF y TTR. La integración del análisis de confiabilidad a la programación de mantenimiento permitió adoptar un enfoque proactivo en el mantenimiento, asignar recursos en función de la confiabilidad del equipo y mitigar riesgos potenciales. Dentro de los resultados destacaron los beneficios de utilizar el análisis de confiabilidad en la programación de mantenimiento, incluyendo estrategias de mantenimiento más rentables, mejor confiabilidad del equipo y mayor eficiencia operativa. En conclusión, este enfoque permitió tomar decisiones informadas y mejorar el rendimiento general de los sistemas de mantenimiento.

La investigación realizada por Peng, *et al.*, (2017), se enfoca en el ámbito del análisis de confiabilidad de sistemas reparables, proponiendo abordar la problemática de los datos de fiabilidad que a menudo se ven afectados por censuras en intervalos debido a fallos en instrumentos, omisiones del personal y estrategias de inspección imperfectas. La investigación, que se centró en un

⁶ MSE: *Mean Squared Error*

⁷ R^2 : Coeficiente de determinación

⁸ MAE: *Mean Absolute Error*

⁹ Reliasoft Weibull ++ : Software especializado para cálculo de confiabilidad

sistema compresor de gasoductos como caso de estudio, utilizó un método novedoso de imputación múltiple y de inferencia, basándose en la suposición de que la tendencia de crecimiento de la fiabilidad seguía un Proceso No Homogéneo de Poisson (NHPP). El enfoque incluyó un modelo estadístico de orden para transformar variables dependientes en independientes, simplificando así el procedimiento de imputación. Los parámetros desconocidos del modelo fueron inferidos iterativamente mediante el algoritmo Monte Carlo de Expectación-Maximización (MCEM¹⁰). Los principales resultados demostraron la eficacia del método propuesto en la mejora de la precisión en la estimación de parámetros y en la predicción de futuras fallas, superando métodos existentes como el de Crow-AMSAA. Este estudio subraya la relevancia de abordar adecuadamente los datos faltantes y censurados en la evaluación de la confiabilidad, ofreciendo un marco sólido para futuras investigaciones en este campo. Las conclusiones reafirmaron la necesidad de métodos que minimicen el sesgo en la estimación de parámetros y mejoren la precisión en la predicción de fallas en sistemas reparables.

El análisis de confiabilidad utilizando datos de falla del sistema de husillo de una rectificadora CNC¹¹ llevado a cabo por Fan *et al.*, (2021), se desarrolló tomando la información de industrias del maquinado de piezas y la población de muestreo consistió en 820 datos de fallas recopilados de 20 rectificadoras CNC a lo largo de un período de 4 años y 8 meses. Para realizar el análisis, se emplearon varios métodos de evaluación de confiabilidad, como el análisis Weibull, distribución normal, distribución logarítmica y distribución gamma. Los resultados indicaron que la función de distribución gamma ofreció el mejor ajuste de confiabilidad, con un valor R del 99.58% según la prueba de Kolmogorov-Smirnov y un error del 7.4%. Además, se calculó el valor del MTBF y se validó la precisión del ajuste utilizando el método de Montecarlo con un nivel de certeza del 97%. Durante el análisis, se identificaron los principales modos de fallo y sus causas, destacando el daño en las piezas mecánicas como el más frecuente, con un índice de fallo del 19.35%. Asimismo, se determinó que el diseño estructural y la selección de piezas/componentes representaron

¹⁰ MCEM: *Monte Carlo Expectation-Maximization*

¹¹ CNC: *Computer Numerical Control*

el 35.48% de los fallos. A partir de estos hallazgos, los investigadores propusieron medidas de mejora, brindando conocimientos valiosos para la evaluación de la confiabilidad de productos similares en la industria.

El estudio realizado por Hartoyo *et al.*, (2023), se centró en optimizar las frecuencias de inspección en tuberías de petróleo utilizando la distribución Weibull y el método Monte Carlo. Este estudio fue desarrollado en una empresa de la industria del Oil & Gas. La población de muestreo consistió en datos de inspección recopilados previamente. Los investigadores llevaron a cabo un análisis estadístico de los datos de inspección, ajustando los tiempos de falla y determinando el porcentaje de fallo mediante la simulación Monte Carlo, permitiendo reducir el riesgo asociado con las inspecciones y optimizar los costos, mejorando así la toma de decisiones tácticas con mayor certeza. Los resultados indicaron que el enfoque basado en la distribución Weibull fue más preciso y confiable que el uso de la distribución normal para el análisis de riesgo en los tiempos de falla e inspección.

La investigación llevada a cabo por M. Kumar *et al.*, (2019), se enfocó en la administración probabilística de la necesidad en sistemas de red inteligente. El caso se desarrolló en el campus de la BHU¹² y la población de muestreo consistió en datos de demanda de carga eléctrica variable del año 2014. Los autores utilizaron funciones de distribución de probabilidad como la logarítmica, Weibull y gamma para estimar la demanda de carga eléctrica. Para analizar verosimilitud de los datos a las funciones teóricas, se realizaron pruebas de bondad de ajuste como Kolmogorov-Smirnov, Anderson-Darling y Cramer-von Mises. El estudio se llevó a cabo utilizando el lenguaje R en el software RStudio para el análisis, desarrollo y visualización de los resultados. El objetivo principal fue desarrollar modelos estocásticos que abordaran la variabilidad distribuida de datos de energía y estimar la proyección de demanda para enfatizar el consumo de energía renovable, con el propósito de reducir el consumo de energía eléctrica proveniente de fuentes fósiles, volviendo el algoritmo de consumo cada vez más inteligente. Los resultados mostraron que la distribución logarítmica fue la mejor ajustada según las pruebas de bondad y

¹² BHU: *Banaras Hindu University*

verosimilitud. Este análisis proporciona información confiable para la toma de decisiones en inversión y para lograr un flujo óptimo de energía.

En el estudio realizado por Patil *et al.*, (2021), en un lugar no especificado, se propone un enfoque para el análisis de confiabilidad de un sistema de caldera de vapor. El objetivo principal fue mejorar las prácticas de gestión de la confiabilidad y el mantenimiento al evaluar de manera efectiva las tasas de falla del sistema y la identificación de los componentes críticos. Para lograr esto, se combinaron el método de juicio de expertos y el método del modelo de falla de mejor ajuste. El método de juicio de expertos proporcionó datos sobre el tiempo para la falla y el tiempo de reparación, y se realizó un análisis estadístico para evaluar la correlación de los datos y estimar las tasas de falla. La integración de estos enfoques permitió obtener un análisis integral que respalda la toma de decisiones informadas en las estrategias de mantenimiento y optimiza el rendimiento general del sistema de una caldera de vapor. Los resultados de la investigación proporcionaron información valiosa para mejorar la confiabilidad y la eficiencia del sistema, destacando la importancia de una gestión adecuada y una identificación precisa de los componentes críticos.

En la investigación realizada por Choudhary *et al.* (2019), abordaron el análisis RAM¹³ para identificar los factores que afectan la eficiencia operativa de una planta industrial de cemento. La población de muestreo consistió en datos de TBF y TTR, utilizados para realizar análisis estadísticos y determinar los subsistemas críticos de la planta. Se emplearon métodos estadísticos para el análisis de confiabilidad. Además, mediante el análisis RAM se establecieron programas de mantenimiento preventivo para los subsistemas críticos, lo que sirvió como referencia para mejorar las estrategias de mantenimiento y optimizó la utilización de la capacidad de la planta. Los principales resultados indicados fueron los factores que afectan la eficiencia operativa y permitieron tomar decisiones informadas para optimizar el rendimiento de la planta. Las conclusiones vinculadas resaltaron la importancia del análisis RAM y el mantenimiento preventivo en la mejora de la eficiencia y la confiabilidad de las

¹³ RAM: *Reliability, Availability and Maintainability*

operaciones industriales.

En el antecedente de Benchekroun *et al.*, (2023), realizado en una planta de cemento, se aborda la aplicación de análisis predictivos de aprendizaje automático para mejorar la seguridad y el rendimiento de los hornos de cemento. El objetivo principal fue integrar técnicas de aprendizaje automático (ML¹⁴) para mejorar la confiabilidad, la gestión del mantenimiento y el rendimiento general del horno de Clinker. Para ello, se utilizó el análisis de datos de sensores históricos y en tiempo real para desarrollar modelos predictivos que permitieran identificar riesgos de seguridad, predecir fallas en los equipos y optimizar los programas de mantenimiento. Los algoritmos de aprendizaje automático fueron entrenados para reconocer patrones en los datos y realizar predicciones precisas, lo que facilitó la toma de acciones de mantenimiento oportunas. Los resultados obtenidos destacaron la eficacia del análisis predictivo de aprendizaje automático en la mejora de la seguridad y el rendimiento de los hornos de cemento. Mediante el aprovechamiento del análisis de datos y el modelado predictivo, se pueden implementar medidas preventivas, optimizar los programas de mantenimiento y mejorar la confiabilidad del horno. Estos hallazgos resaltan la importancia de la aplicación de técnicas de aprendizaje automático en la industria del cemento para lograr una mayor eficiencia y seguridad en los procesos.

En la investigación realizada por Fang *et al.*, (2021), en un laboratorio de pruebas de tecnología aeronáutica, se centraron en aplicar una versión mejorada de la distribución Weibull para predecir cómo evolucionará la confiabilidad a largo plazo de un mecanismo de cierre de puertas en aeronaves, utilizando modelos de distribución Weibull de 2 modos de falla y realizaron una evaluación de verosimilitud para optimizar el número de pruebas y reducir los costos asociados. Se emplearon múltiples parámetros para mejorar la flexibilidad de la distribución. Los datos utilizados provinieron de pruebas en seis sistemas de cierre de puerta, con cada mecanismo sometido a 18,000 pruebas y registrando datos cada 500 pruebas. El estudio resaltó la importancia de equilibrar el número de parámetros y la precisión de los resultados del

¹⁴ ML: *Machine Learning*

cálculo en la estimación de parámetros. Para el desarrollo, utilizaron software como MATLAB o propusieron una alternativa como RStudio para obtener soluciones, especialmente en la estimación de máxima verosimilitud. El estudio señaló que una distribución estadística Weibull de cuatro o cinco parámetros es adecuada para una curva de bañera con tres etapas, proporcionando mayor certeza y flexibilidad en la predicción de datos.

En el estudio llevado a cabo por Kirbas (2019), tuvo como objetivo determinar la función de distribución más adecuada para el análisis de datos de velocidad del viento. El estudio comparó el desempeño de diferentes funciones estadísticas de distribución, como gamma, logística, logarítmica y Weibull, utilizando lenguaje de programación y el análisis estadístico. Los resultados revelaron que la distribución Weibull obtuvo mejores resultados en comparación con las distribuciones gamma y logarítmica, especialmente en velocidades de viento más altas. Sin embargo, para velocidades de viento bajas, las distribuciones logarítmica y gamma ofrecieron mejores resultados que la ampliamente utilizada distribución Weibull. El estudio también destacó el uso de un software de código abierto, para realizar cálculos estadísticos y análisis gráfico, resaltando su rentabilidad y su interfaz de fácil uso en el campo de la investigación. Estos hallazgos brindaron información valiosa para la toma de decisiones en inversiones relacionadas con la energía eólica, al considerar diferentes modelos de distribución y realizar un análisis de eficiencia basado en los datos de velocidad del viento.

En la investigación sobre la predicción de la vida útil de una caja de engranajes de turbina eólica marina realizada por Bhardwaj *et al.*, (2019), se centró específicamente en el análisis de la tasa de fallas y la evaluación de la confiabilidad. El objetivo principal fue estimar la tasa de fallas y predecir el rendimiento de confiabilidad de la caja de engranajes en condiciones operativas. Para lograrlo, se analizaron datos históricos de fallas y se utilizaron métodos de predicción probabilísticos, como el análisis de Weibull y los diagramas de bloques de confiabilidad (RBD¹⁵). Los resultados obtenidos permitieron una estimación precisa de la tasa de fallas y brindaron información

¹⁵ RBD: Reliability block diagram

sobre el rendimiento de confiabilidad de la caja de engranajes. Esto facilita la toma de decisiones en cuanto a estrategias de mantenimiento, gestión de repuestos y mejoras generales en la confiabilidad del sistema.

Vanderschueren *et al.*, (2023) presentan un enfoque para optimizar la frecuencia de mantenimiento preventivo mediante el uso de datos y causalidad. El objetivo fue encontrar la frecuencia óptima de mantenimiento que minimice los costos asociados a fallos y revisiones. Los autores utilizaron aprendizaje automático (ML) para predecir los resultados de diferentes frecuencias de mantenimiento utilizando datos observacionales. Con base en estas predicciones, optimizaron la frecuencia de mantenimiento de cada activo para minimizar los costos totales. El enfoque se valida utilizando datos reales de más de 4000 contratos de mantenimiento. Los resultados demostraron que este enfoque personalizado y causal ofrece horarios de mantenimiento más precisos y rentables en comparación con enfoques que no consideran el sesgo de selección o las características específicas de cada activo. Los resultados mostraron una mejora significativa en la precisión de las predicciones y en la reducción de costos mediante el enfoque causal y personalizado propuesto. Sin embargo, se destaca que el enfoque tiene limitaciones, como la necesidad de datos observacionales y el cumplimiento de ciertas suposiciones que pueden ser difíciles de verificar en la práctica.

La confiabilidad será medida a partir de número de fallas y duración (tiempo según escala) de las mismas (Mora, 2009). La confiabilidad refiere a la probabilidad de que un sistema o conjunto de equipos realice su función de diseño predicha de forma adecuada en un período específico de tiempo bajo las condiciones operacionales establecidas. Por lo general, se expresa como un porcentaje y se mide por el tiempo medio entre fallas MTBF (Gulati & Smith, 2013). En tal sentido un indicador de MTBF más alto generalmente indica equipos más confiables.

Por definición, según Gulati & Smith (2013) y la SMRP (2020), indican que el MTBF es una medida de la confiabilidad de los equipos. Se calcula dividiendo el tiempo operativo total del activo por el número de fallas durante un período de tiempo, como se muestra en la Ecuación 1:

$$MTBF = \frac{\text{Tiempo de operación}}{\text{Número de fallas}} \quad (1)$$

Donde:

MTBF: Tiempo promedio entre fallas (Horas)

Tiempo de operación: Tiempo de operación sin fallas de un equipo o sistema en un periodo de tiempo determinado (Horas)

Número de fallas: Cantidad de fallas producidas en un equipo o sistema en un periodo de tiempo determinado (Número)

Por definición, según Gulati & Smith (2013) y la SMRP (2020), el tiempo promedio de reparación (MTTR) es el tiempo promedio necesario para restaurar un equipo a su condición operativa luego de un evento de parada por falla MTTR es el tiempo para reparar el equipo o sistema y es una medida importante de mantenibilidad, como se muestra en la Ecuación 2:

$$MTTR = \frac{\text{Tiempo de reparación (horas)}}{\text{Número de eventos de reparación}} \quad (2)$$

Donde:

MTTR: Tiempo promedio de reparación (Horas)

Tiempo de reparación: Tiempo empleado para la reparación de un equipo a su estado operativo en un periodo de tiempo determinado (Horas)

Número de eventos de reparación: Cantidad de eventos de reparación de un equipo en un periodo de tiempo determinado (Número)

De acuerdo con la investigación realizada por S. Kumar & Singh, (2020), se define la confiabilidad como la medida de la capacidad de un sistema para operar de manera continua durante un periodo de tiempo específico.

La confiabilidad (Ahmadi *et al.*, 2019), se refiere a la habilidad de un sistema para cumplir una tarea dentro de condiciones y plazos establecidos,

aumentando su nivel de confiabilidad a medida que pueda mantener su rendimiento en el tiempo.

En la Ecuación 3 (Dhillon, 2002), se define la confiabilidad como la probabilidad de que un sistema funcione sin fallos hasta un determinado tiempo, considerando una tasa de fallas y un límite máximo de fallas aceptables.

$$R(t) = e^{(-\lambda*t)} \quad (3)$$

Donde:

R(t): Confiabilidad en una función de un tiempo determinado (%).

λ : Tasa de fallas (fallas/horas).

t: Periodo de tiempo transcurrido en horas (horas).

Para la evaluación de la confiabilidad se emplean modelos estadísticos, especialmente en el contexto de la selección de modelos y la verificación de ajustes. Entre los modelos estadísticos empleados para esta función encontramos:

Log-Likelihood (Logaritmo de la verosimilitud): Es una medida de la probabilidad de observar los datos dados un modelo estadístico. Cuanto mayor es el logaritmo de la verosimilitud, mejor se ajusta el modelo a los datos. Para una muestra de datos x_1, x_2, \dots, x_n que se asume que siguen una distribución probabilística con parámetros θ , la log-verosimilitud $L(\theta|x)$ se expresa como (Lehmann, *et al.*, 1998):

$$L(\theta | x) = \log f(x_1, x_2, \dots, x_n | \theta) \quad (4)$$

Donde:

f es la función de densidad de probabilidad (para datos continuos) o la función de masa de probabilidad (para datos discretos) que describe la distribución de los datos bajo el modelo con parámetros θ .

El objetivo suele ser encontrar los valores de θ que maximizan esta log-

verosimilitud, lo que se conoce como estimación de máxima verosimilitud (MLE¹⁶).

La ecuación de la log-verosimilitud depende del modelo estadístico específico que se esté utilizando para analizar los datos. Por ejemplo, en el caso de una distribución normal, la log-verosimilitud estaría basada en la fórmula de la función de densidad de probabilidad de la distribución normal. Para otras distribuciones o modelos estadísticos, la ecuación variará en función de la fórmula de su función de densidad o masa de probabilidad.

Criterio de Información de Akaike corregido (AICc¹⁷): Burnham, *et al.*, (2002), la definen como una versión del AIC¹⁸ (Criterio de Información de Akaike) ajustada para pequeños tamaños de muestra. Este método se utiliza para comparar modelos, penalizando la complejidad para evitar el sobreajuste y se obtiene mediante la Ecuación 5:

El AICc se calcula utilizando la fórmula:

$$AICc = -2 \times \log - \text{verosimilitud máxima} + 2k + \frac{2k(k + 1)}{n - k - 1} \quad (5)$$

Donde:

log-verosimilitud máxima es el logaritmo de la verosimilitud máxima alcanzada por el modelo (Ver Ecuación 4).

k es el número de parámetros en el modelo.

n es el número de observaciones.

Criterio de Información Bayesiano (BIC¹⁹): Según Gelman, *et al.*, (1995), es similar al AIC, el BIC también penaliza la complejidad del modelo pero tiene un enfoque más bayesiano y da una penalización mayor a los modelos con más parámetros y se puede calcular con la Ecuación 6:

$$BIC = -2 \times \log - \text{verosimilitud máxima} + k \times \log(n) \quad (6)$$

¹⁶ MLE: Maximum Likelihood Estimation

¹⁷ AICc: Akaike Information Criterion corrected

¹⁸ AIC: Akaike Information Criterion

¹⁹ BIC: Bayesian Information Criterion

Donde:

log-verosimilitud máxima es el logaritmo de la verosimilitud máxima alcanzada por el modelo (Ver Ecuación 4).

k es el número de parámetros en el modelo.

n es el número de observaciones.

Test de Anderson-Darling (AD²⁰): D'Agostino, *et al.*, (1986) lo definen como una prueba de bondad de ajuste que evalúa si una muestra proviene de una población específica con una distribución dada. Es más sensible a las colas de la distribución que otras pruebas. Este test se basa en la comparación de los valores ordenados de los datos observados con los valores esperados bajo la distribución teórica considerada. Calcula una estadística de prueba que se utiliza para obtener un p-valor, el cual indica si hay suficiente evidencia para rechazar la hipótesis nula de que los datos siguen la distribución teórica asumida. El resultado del test de Anderson-Darling incluye un p-valor. Si el p-valor es menor que un umbral predefinido (normalmente 0.05), se rechaza la hipótesis nula y se concluye que los datos no siguen la distribución teórica considerada. En caso contrario, no se dispone de suficiente evidencia para rechazar la hipótesis nula y se asume que los datos podrían provenir de esa distribución teórica.

La fórmula general se basa en las diferencias entre las funciones de distribución acumulada observada y teórica. Dada una muestra ordenada $x_1 \leq x_2 \leq \dots \leq x_n$ y la función de distribución acumulada empírica $F_n(x)$ basada en estos datos, y una función de distribución acumulada teórica $F(x)$, el estadístico de prueba Anderson-Darling se calcula mediante la Ecuación 7:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i - 1) \cdot (\ln F(x_i) + \ln(1 - F(x_n + 1 - i)))] \quad (7)$$

Donde:

n es el tamaño de la muestra.

x_i son los valores de la muestra ordenada.

²⁰ AD: Anderson-Darling test

$F(x_i)$ es la función de distribución acumulada teórica evaluada en x_i .

Test de Kolmogorov-Smirnov (K-S²¹): Según Hollander, *et al.*, (1973), es otra prueba de bondad de ajuste utilizada para comparar una muestra con una distribución de referencia o para comparar dos muestras. La idea básica detrás de este test es comparar la función de distribución acumulada empírica (ECDF²²) de la muestra con la función de distribución acumulada teórica (CDF²³) que se espera que sigan los datos, como la CDF de una distribución normal, uniforme, exponencial, etc.

El estadístico de prueba D se calcula como la máxima diferencia vertical absoluta entre las dos funciones de distribución acumulada. Para una muestra ordenada X_1, X_2, \dots, X_n con una función de distribución acumulada empírica $F_n(x)$ y una función de distribución acumulada teórica $F(x)$, el estadístico D se calcula siguiendo la Ecuación 8:

$$D = \max(\sup |F_n(x) - F(x)|) \quad (8)$$

Donde:

\sup representa el supremo, es decir, la mayor diferencia absoluta.

$F_n(x)$ es la función de distribución acumulada empírica.

$F(x)$ es la función de distribución acumulada teórica.

Estas herramientas son fundamentales en el análisis estadístico y son ampliamente utilizadas para la selección de modelos, comparando la calidad de ajuste de diferentes modelos estadísticos y verificando si los datos se ajustan a una distribución específica. Cada una tiene sus propias ventajas y limitaciones y se eligen según el contexto específico del análisis.

III. METODOLOGÍA

3.1. Tipo y diseño de investigación

²¹ K-S: Kolmogorov-Smirnov test

²² ECDF: Empirical Cumulative Distribution Function

²³ CDF: Cumulative Distribution Function

Tipo de investigación

El presente estudio seguirá un enfoque de investigación aplicada, en concordancia con el objetivo general y la problemática planteada. Se utilizó los modelos Crow_AMSAA y validación cruzada de *machine learning* con el fin de estimar la confiabilidad del sistema y de los equipos de molienda de cemento. Esta elección metodológica se sustenta en las contribuciones previas de Arias Gonzales (2020); Gabriel-Ortega (2017); Huairé Inacio *et al.* (2022).

Diseño de investigación

El diseño de investigación adoptado para este estudio se fundamenta en un enfoque cuantitativo, mediante la aplicación de un modelo de distribución estadística combinado de *machine learning* con transformación logarítmica y validación cruzada, a la variable confiabilidad. Se puede clasificar como un diseño de investigación de alcance correlacional no-causal, tal como lo menciona Huairé Inacio *et al.* (2022), visto que nuestra investigación radica en estimar bajo un modelo propuesto la confiabilidad de un sistema y compararlo con la confiabilidad estimada mediante modelos convencionales.

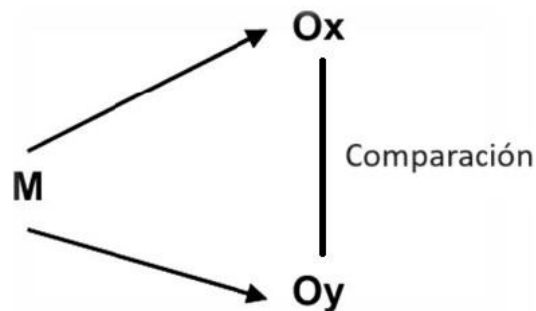


Figura 1 - Esquema de metodología

Donde:

M: Muestra de investigación

Ox y Oy: Variables observadas - Variable Ox: Confiabilidad estimada bajo modelo de *machine learning* con transformación logarítmica y validación cruzada. Variable Oy: Confiabilidad estimada bajo modelos convencionales

c: Comparación de variables

3.2. Variables y operacionalización

Variable: Modelo de *machine learning* con transformación logarítmica y validación cruzada.

Modarres, *et al.*, (1999), definen el método de transformación logarítmica, combinado con la validación cruzada, como una estrategia estadística empleada para estimar los parámetros de modelos de fiabilidad y garantizar su robustez, aplicado frecuentemente en la evaluación de datos de TBF en sistemas de ingeniería y en análisis de supervivencia mediante el ajuste lineal de conjuntos de datos con distribuciones no lineales, como la Weibull, habitual en el estudio de la confiabilidad. La validación cruzada se utiliza para evaluar la generalización de modelos estadísticos, especialmente útil en contextos con limitaciones en la cantidad de datos (Arlot, *et al.*, 2009).

Variable 2: Confiabilidad

Según el estudio de Kumar y Singh (2020), la confiabilidad es la medida de la capacidad de un sistema para funcionar sin interrupciones durante un período de tiempo determinado. Según lo expuesto por Ahmadi *et al.*, (2019), refiriendo a Modarres *et al.*, (1999) y Gulati (2013), la confiabilidad es la capacidad de un sistema para cumplir una tarea en condiciones y plazos establecidos, volviéndose más confiable a medida que pueda sostener su condición en el tiempo.

En el Anexo 1 se detalla la Tabla 9, de operacionalización de variables.

3.3. Población, muestra y muestreo

Población

El objeto de estudio estará conformado por la línea de equipos de la planta de molienda (Molino de rodillos verticales) de cemento, en una empresa productora de cemento en la localidad de Pacasmayo. El universo de equipos para la planta de molienda es de 69 equipos, a los cuales se les estimará la confiabilidad en el periodo de evaluación, con la finalidad de estimar los tiempos entre falla.

Muestra

La muestra seleccionada para este estudio consiste en 9 equipos específicos, elegidos mediante el método de Pareto y cantidad de datos de eventos de falla, provenientes de una población total de 69 equipos. Este proceso de selección se llevó a cabo basándose en criterios clave de priorización que identificaron aquellos elementos que mostraban la mayor influencia o impacto en determinados aspectos del sistema estudiado y representan una muestra representativa y relevante dentro de la población total.

Muestreo

Para esta tesis, se ha realizado un muestreo selectivo de 9 equipos principales, utilizando el método de Pareto, a partir de una población compuesta por 69 equipos en total. Este enfoque de selección se basó en la metodología de identificar y priorizar aquellos equipos que ejercen una influencia significativa en el funcionamiento general del sistema. La utilización del método de Pareto en la selección de estos equipos asegura que la investigación se centre en los elementos que tienen un peso considerable en el funcionamiento y desempeño del sistema.

Criterios de inclusión

En el marco de esta investigación, se considera un enfoque objetivo al abordar todos los equipos de la línea del molino. Para obtener datos precisos, se utilizará la base de datos de registro de paradas de planta, el cual forma parte del sistema de gestión de operaciones y paradas (SGOP²⁴) implementado en la organización. Mediante el uso del programa Excel, se ha analizará el detalle de las paradas y los tiempos asociados, centrándose en las fallas eléctricas, mecánicas y electrónicas de los equipos considerados. El análisis se realizará en un periodo que abarca desde enero del 2021 hasta julio del 2023, con el objetivo de recopilar información actualizada y representativa. Este enfoque garantiza la solidez y

²⁴ SGOP: Sistema de gestión de operaciones y paradas

confiabilidad de la base de datos.

Criterios de exclusión

Dentro del presente estudio, se considerarán únicamente los valores registrados por fallas durante el periodo especificado, es decir, a partir del año 2021, no serán tomados en cuenta de la base de datos los registros de años anteriores. Asimismo, se excluirán aquellos valores del registro de paradas de equipos que no estén relacionados con trabajos de mantenimiento correctivo mecánico, eléctrico y electrónico, dado que el enfoque de estudio será sobre el impacto de la tasa de fallas sobre los resultados de confiabilidad de los sistemas de la planta de molienda de cemento. Se eliminarán los valores no pertinentes a las paradas investigadas, con el fin de garantizar la precisión y relevancia de los datos recopilados, respecto a los indicadores de mantenimiento. No se considerarán tiempos de parada de equipos por motivos relacionados a temas operativos, maniobras por cambio de producción o paradas programadas según cronograma de operación. No se considerará el primer evento de falla para cada equipo crítico debido a no tener información en registro de fallas del 2020.

Unidad de análisis

Se considera el registro de tiempos de paradas por motivo de fallas en equipos de tipo mecánica, eléctrica o electrónica dentro del periodo enero 2019 a agosto 2023 para todos los equipos de la línea de molienda de cemento, en una empresa en Pacasmayo.

3.4. Técnicas e instrumentos de recolección de datos

Técnicas de recolección de datos

En la investigación se emplea la técnica de análisis documental. Esta técnica se caracteriza por recopilar datos de fuentes primarias y secundarias utilizando una ficha de registro de datos, lo que permite obtener información relevante para el estudio (Huaire Inacio *et al.*, 2022).

Instrumentos de recolección de datos

Se empleará una ficha de registro de datos para almacenar la información recopilada a través del sistema SGOP de la planta de cemento ubicada en Pacasmayo. Este instrumento registrará las paradas de los equipos debido a fallas, brindando una fuente confiable de información para el presente estudio (Huaire Inacio *et al.*, 2022).

En el Anexo 2 se detalla el instrumento de ficha de registro empleada para control de datos de paradas de equipos en falla.

3.5. Procedimientos

Para llevar a cabo esta investigación, se obtuvo el permiso correspondiente de la organización para utilizar la información obtenida del SGOP referente a los datos de paradas de planta. Además, se empleó la ficha documental que se soporta en la información obtenida del SGOP y posterior clasificación de la información a través de un análisis de Pareto para identificar los equipos que presenten mayor tiempo de falla en el periodo analizado. Para el análisis de los datos, se desarrolló un modelo de machine learning con transformación logarítmica y validación cruzada como propuesta para la estimación de la confiabilidad. Se utilizó un software estadístico para estimar la confiabilidad de los equipos seleccionados del sistema de molienda mediante el uso de modelos convencionales y de machine learning propuesto. Los resultados obtenidos se compararon para determinar la precisión de los modelos empleados mediante el uso de las métricas MSE, R^2 y MAE en cada uno de ellos. Al obtener estos resultados, se realizó una discusión sobre los datos obtenidos en la aplicación del modelo de machine learning propuesto con los de los modelos convencionales, así como la elaboración de las conclusiones y recomendaciones de esta investigación.

3.6. Método de análisis de datos

En este proyecto de investigación, se realizó la recolección, clasificación, segmentación, ordenamiento y procesamiento con las herramientas de análisis - software *Microsoft Excel*. Se utilizó el método de Pareto para identificar aquellos equipos donde se valide la mayor cantidad de horas de

fallas, que impactan al sistema de molienda de cemento. Posteriormente se emplearon modelos de análisis de datos convencionales y de *machine learning* para estimar el tiempo entre fallas y confiabilidad de aquellos equipos que afecten con mayor impacto al sistema de molienda, empleando algoritmos de programación en lenguaje *Python*. Posteriormente, se compararon los resultados al aplicar el modelo propuesto de *machine learning* con transformación logarítmica y validación cruzada sobre los resultados de confiabilidad.

3.7. Aspectos éticos

El presente estudio de investigación se regirá por principios éticos que salvaguarden la autonomía y el respeto a los derechos de autor. Se dará especial consideración a las fuentes de información relevantes para la construcción del marco teórico y el diseño metodológico del proyecto, asegurando una adecuada atribución a cada autor de acuerdo con las directrices establecidas en la norma ISO 690. Se realizarán las correspondientes referencias bibliográficas y se garantizará la integridad intelectual de las ideas y conocimientos utilizados en este estudio.

Dentro del marco de esta investigación, se velará por salvaguardar el derecho de las personas a decidir libremente su participación en el desarrollo de este estudio. Se garantizará que el presente proyecto se enfoque en el bienestar y no tenga intenciones de provocar ningún tipo de daño moral a los individuos y entidades que decidan formar parte de este.

Los estudiantes de la Universidad César Vallejo se enfocaron en asegurar la conducta ética en su investigación, la cual se centró principalmente en objetivos académicos. Obtuvieron permiso de la empresa de producción de cemento para usar información, y tras su aprobación, confirmaron que los datos obtenidos son auténticos, transparentes y fiables. Garantizaron que no hubiera manipulaciones o comportamientos poco éticos durante el proceso. En los Anexos 3 y 4, se detallan los documentos empleados para solicitud y uso de información de empresa.

En el contexto de este proyecto de investigación, se dará especial atención al principio de justicia, promoviendo la inclusión y rechazando enérgicamente

cualquier forma de discriminación. Además, se asumirá el compromiso de garantizar la veracidad de los resultados y la integridad de la información obtenida en la investigación. Para ello, se llevará a cabo una exhaustiva evaluación y análisis de los resultados mediante el uso del programa anti-plagio Turnitin, asegurando la originalidad y la calidad de los hallazgos obtenidos.

IV. RESULTADOS

4.1. Análisis de Pareto para identificar equipos críticos en base a tiempo entre fallas.

En este estudio, se emplea el diagrama de Pareto para analizar las fallas de cada elemento dentro del sistema de molienda de cemento. Este diagrama se utiliza con el fin de distinguir los componentes que experimentan fallas con mayor frecuencia de aquellos que presentan una ocurrencia menos recurrente.

A continuación, se muestran los resultados del análisis de Pareto aplicado a los 69 equipos que comprende el sistema de molienda de cemento, teniendo como resultado un total de 15 equipos que comprenden el 80.69% del total de horas de parada registrados en el periodo de enero 2019 a agosto 2023. La Tabla 1 muestra el tiempo de parada por equipo y el tiempo de parada acumulado del periodo analizado en el sistema de molienda de cemento, con su respectivo porcentaje acumulado.

Tabla 1

Equipos críticos a través de Análisis de Pareto

| Equipo | Tiempo de falla (Horas) | Tiempo H % | Acumulado Tiempo de falla (Horas) | Acumulado Tiempo H % |
|--|-------------------------|------------|-----------------------------------|----------------------|
| REDUCTOR PRINCIPAL 4000KW - 1190/21.9RPM | 331.79 | 16.44% | 331.79 | 16.44% |
| CLASIFICADOR LSKS-76S | 246.8 | 12.23% | 578.59 | 28.66% |
| SUBESTACIÓN ELÉCTRICA | 215.87 | 10.69% | 794.46 | 39.36% |
| ELEVADOR CANGILONES CT6 A SILOS CTO | 139.42 | 6.91% | 933.88 | 46.26% |
| GENERADOR DE GASES CALIENTES CT6 | 107.91 | 5.35% | 1041.79 | 51.61% |
| MOLINO VERTICAL LOESCHE LM56.2+2C/S | 102.26 | 5.07% | 1144.05 | 56.68% |
| UNIDAD DOSIFICACION DE AGUA CT6 | 92.54 | 4.58% | 1236.59 | 61.26% |
| MOTOR ELECT. PRINCIPAL 4000 KW-1192 RPM | 82.95 | 4.11% | 1319.54 | 65.37% |

| | | | | |
|--|-------|-------|---------|--------|
| RODILLOS ESCLAVOS 1-2 | 69.65 | 3.45% | 1389.19 | 68.82% |
| BALANZA DOSIFICADORA CLINKER TIPO I CT6 | 46.86 | 2.32% | 1436.05 | 71.14% |
| SEPARADOR MAGNETICO STEINERT | 46.19 | 2.29% | 1482.24 | 73.43% |
| FAJA INCLINADA ALIMENTACION TOLVAS CT6/7 | 44.8 | 2.22% | 1527.04 | 75.65% |
| DETECTOR DE METALES | 37.49 | 1.86% | 1564.53 | 77.51% |
| VENTILADOR EXTRACTOR FILTRO PPAL | 33.56 | 1.66% | 1598.09 | 79.17% |
| FAJA ALIMENTACION A MOLINO CT6 | 30.67 | 1.52% | 1628.76 | 80.69% |

Fuente: Elaboración propia

Con los datos del total de horas de falla por equipo y total de horas de falla acumulado, se elaboró el gráfico presentado en la Figura 2 sobre la distribución de horas de parada del total de equipos y en la Figura 3, se presenta la distribución de horas y porcentaje acumulado de los equipos que representan el 80.69% del total del tiempo de falla acumulado.

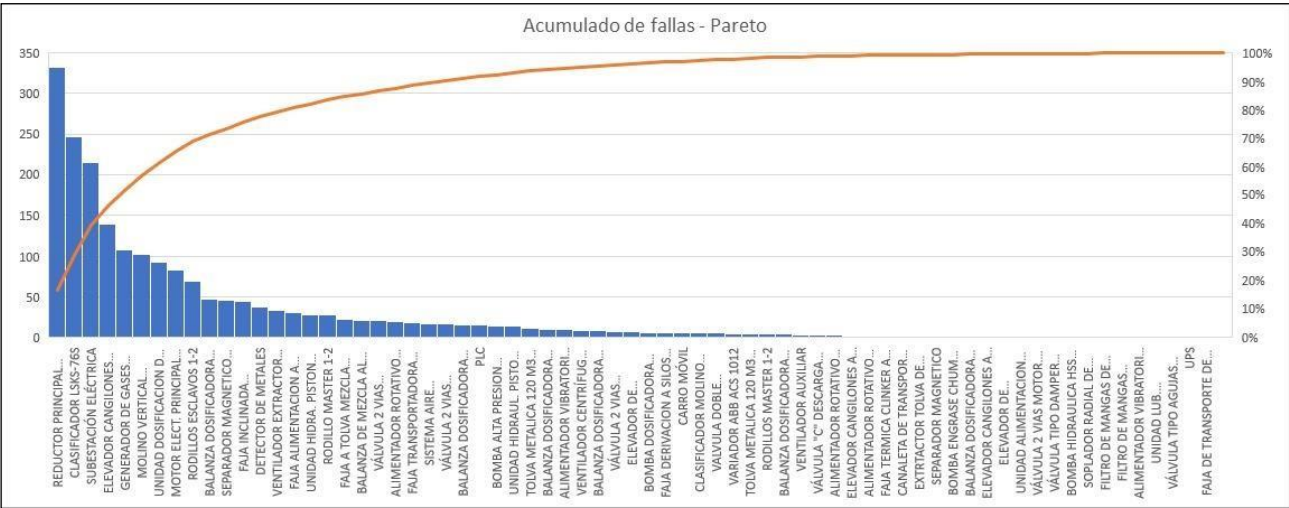


Figura 2 - Diagrama de Pareto completo

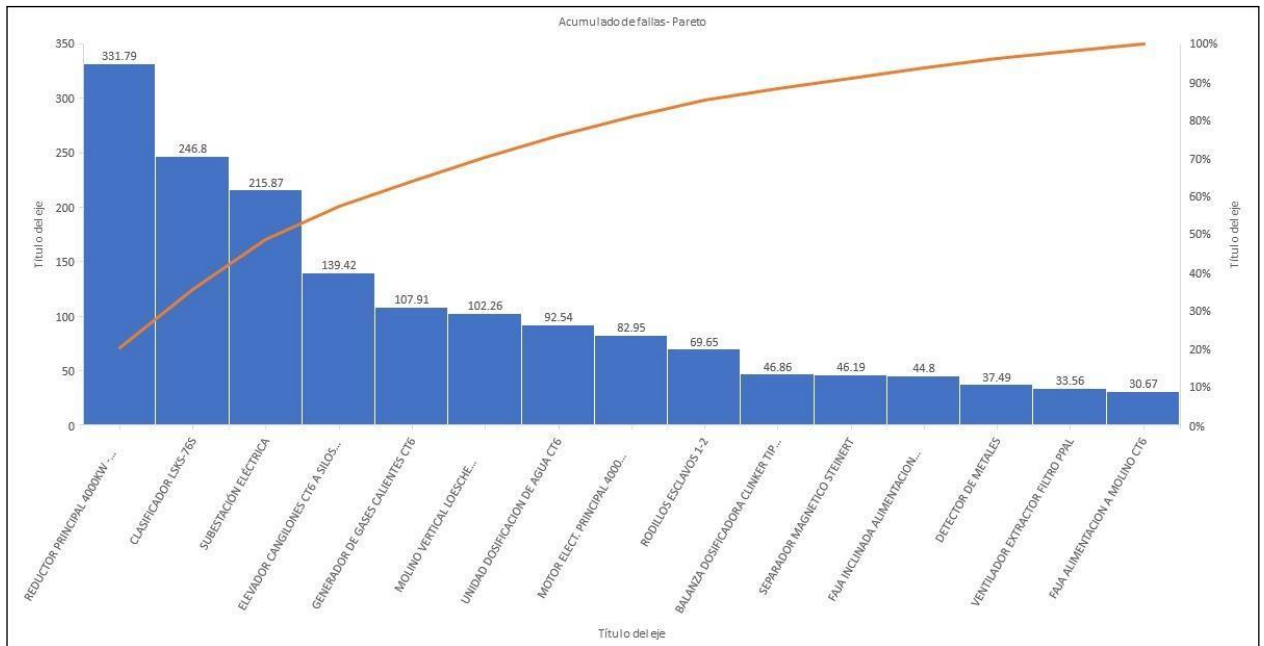


Figura 3 - Diagrama de Pareto de equipos críticos

A partir de la Figura 3 se hace referencia a la identificación y selección de se seleccionaron 8 de los 15 equipos para un análisis detallado, basándose en que estos tienen un registro más significativo de eventos., dado que cuentan con el mayor registro de eventos para el análisis de distribuciones. La recomendación de utilizar un valor de eventos superior a 10 se basa en la necesidad de tener suficiente información para que los análisis estadísticos sean válidos y confiables. En estadísticas, una mayor cantidad de datos puede ayudar a mejorar la precisión de las estimaciones y a hacer inferencias más sólidas sobre las poblaciones o procesos subyacentes. (Krejcie *et al.*, 1970; Greenwood *et al.*, 1996).

4.2. Análisis de Confiabilidad por métodos tradicionales

Para estimar la confiabilidad según métodos convencionales, se ha optado por emplear un enfoque estadístico. En el estudio se utilizó lenguaje de programación abierta mediante algoritmos en lenguaje *Phyton*, para identificar el modelo que tenga el mejor ajuste a los datos de TBF para una distribución teórica a nivel de equipo, en la Figura 4 podemos observar la secuencia de análisis para el modelo estadístico K-S a través de un diagrama de flujo, en donde se detallan los pasos que siguen los bloques de

programación para el análisis de los datos y procesamiento de resultados finales ((V. Talkit *et al.*, 2022; Fan *et al.*, 2021), de igual manera, en la Figura 05 nos muestra la secuencia para el modelo estadístico MLE (Mclinn, 2010).

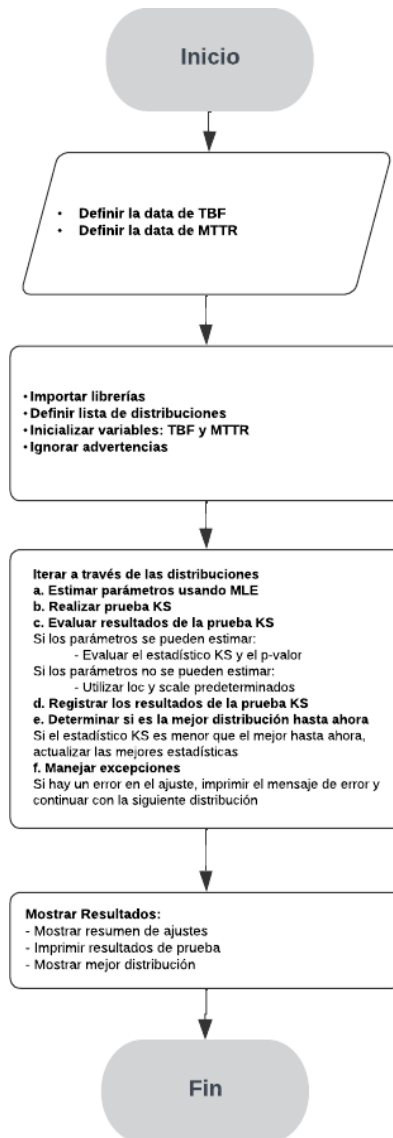


Figura 4 - Diagrama de flujo para análisis por modelo K-S

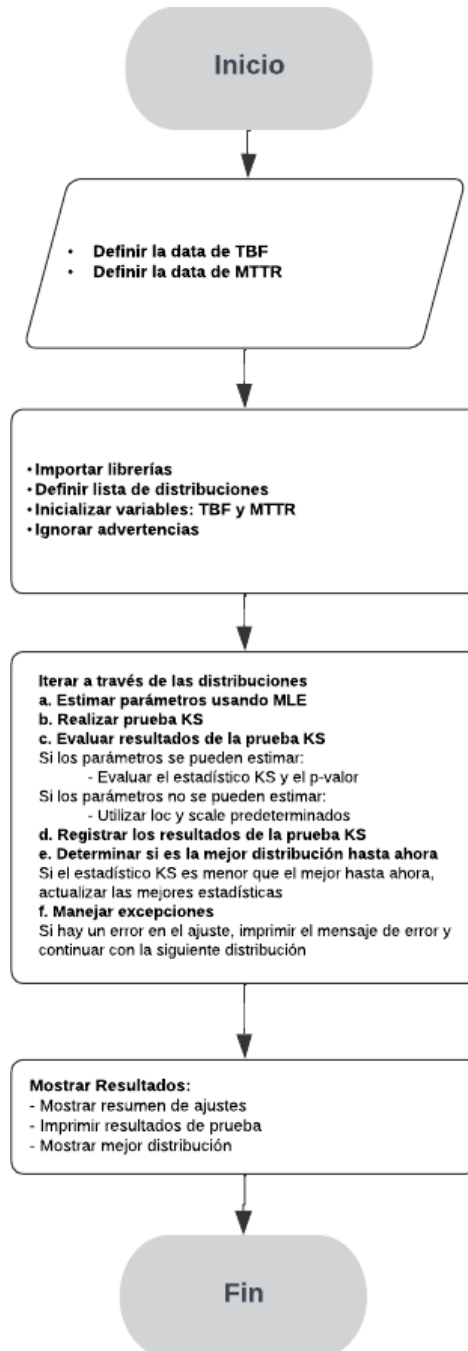


Figura 5 - Diagrama de flujo para análisis por modelo MLE

Se ha buscado la distribución más adecuada para todos los equipos en el sistema de molienda de cemento a través del modelo MLE, utilizando los datos de TBF. En la Tabla 2, se han considerado las distribuciones convencionales de normal, lognormal, 1p exponencial, 2p exponencial, 2p

Weibull y 3p Weibull (V. Talkit *et al.*, 2022) utilizadas en la estimación de confiabilidad para identificar la distribución de datos de fallas libres de tendencias. La distribución se utiliza para pronosticar el comportamiento de falla de los equipos. Los resultados presentados en la Tabla 2 cuentan con hasta siete decimales para resaltar la precisión en los cálculos del TBF para cada equipo, así como el valor i que representa el número de datos que se tuvo en cuenta para cada uno de ellos al momento de hacer los cálculos.

Tabla 2

Mejor ajuste para los datos de TBF a nivel de equipo por método MLE

| Mejor ajuste de data TBF | | Método de análisis MLE (Maximun long-likelihood), máxima verosimilitud. | | | | | | | |
|------------------------------------|---|---|--|---|--|---|---|--|---|
| Distribución | Molino unidad total | Reductor principal 4000kw - 1190/21.9rpm | Clasificador LSKS-76S | Generador de gases calientes ct6 | Molino vertical Loesche LM56.2+2c/s | Motor elect. Principal 4000 kW-1192 rpm | Rodillos esclavos 1-2 | Balanza dosificadora clinker tipo I CT6 290-140 | Ventilador extractor filtro ppal 290-665 |
| | i = 136 | i = 13 | i = 17 | i = 19 | i = 36 | i = 33 | i = 21 | i = 21 | i = 26 |
| MTBF Lineal -aritmético (estático) | 253.1985294 | 2324.6153846 | 2253.0000000 | 1981.8421053 | 1038.6666667 | 1183.9090909 | 1420.6190476 | 1073.7619048 | 1507.2692308 |
| Normal | 253.1985294 | 2324.6153846 | 2253.0000000 | 1981.8421053 | 1038.6666667 | 1183.9090909 | 1420.6190476 | 1073.7619048 | 1507.2692308 |
| Exponencial | 222.1985294 | 2187.6153846 | 2218.0000000 | 1910.8421053 | 962.6666667 | 1112.9090909 | 1340.6190476 | 1013.761905 | 1423.2692308 |
| Exponencial 2p | 253.1985294 | 2324.6153846 | 2253.0000000 | 1981.8421053 | 1038.6666667 | 1183.9090909 | 1420.6190476 | 1073.7619048 | 1507.2692308 |
| Lognormal | 244.19174755 | 2004.628976 | 2270.320186 | 1728.8536098 | 1014.0827693 | 1238.2837814 | 1260.9441101 | 1197.0809346 | 1537.4780960 |
| Weibull 2P = Weibul.min | 256.1923139 | 2296.514136 | 2258.251911 | 1971.8610600 | 1023.2002636 | 1181.5392249 | 1382.1403189 | 1073.1872739 | 1507.9751644 |
| Weibull 2P, MTBF T anual | 174.1833143 | 1177.080814 | 1179.161649 | 900.1058816 | 544.7085828 | 655.1167961 | 657.8641232 | 617.7210129 | 877.0751758 |
| Weibull 3P | 237.0315500 | 3861.487285 | 2837.268902 | 2774.9837769 | 1036.7129983 | 1118.4325016 | 5597.2804423 | 1283.2522018 | 1854.8832431 |
| Weibull 3P, MTBF T anual | 237.0315493 | 2580.257713 | 2450.651853 | 2014.8067055 | 1028.6629486 | 1111.2807507 | 2686.6316834 | 1274.6122185 | 1811.7756620 |
| Mejor ajuste de distribución | Lognormal | Weibull | Weibull | Weibull | Weibull | Exponencial | Weibull | Weibull | Weibull |
| | Forma (s) = 0.736793 Escala = 186.144213 | Forma (β) = 0.807832 Escala (n-eta) = 2040.956130 | Forma (β) = 0.9502652 Escala (n-eta) = 2206.90173 | Forma (β) = 0.711368 Escala, n-eta = 1581.481731 | Forma (β) = 0.829737 Escala, n-eta = 926.028456 | Escala = 1112.909091 Localización = 71.0 | Forma (β) = 0.710973 Escala, n-eta = 1107.941724 | Forma (β) = 1.0657046 Escala, n-eta = 1100.161920 | Forma (β) = 1.0197575 Escala, n-eta = 1520.19343 |

En el resumen se presentan los cálculos de TBF obtenidos mediante diversos modelos de distribuciones clásicas para cada equipo. Estos modelos se evaluaron utilizando criterios como la máxima verosimilitud, AICc, BIC, AD y K-S. Se enfatiza la identificación de los modelos más adecuados según los parámetros de cada evaluación. Estas herramientas son vitales en el análisis estadístico, ya que permiten comparar y seleccionar modelos estadísticos que se ajusten mejor a los datos reales, verificando su correspondencia con una distribución específica. Cada herramienta tiene sus ventajas y limitaciones, seleccionándose según el contexto particular del análisis.

En la Tabla 3, se observa cómo se desenvuelve la función confiabilidad basada en el tiempo para los componentes de un sistema de molino. Los datos analizados revelan tendencias significativas en la degradación de la confiabilidad a lo largo del tiempo, abarcando un rango desde 24 horas hasta tres años. La evaluación correlacional entre diversos componentes, tales como el molino unidad total, el reductor principal, el clasificador, y otros, muestra una disminución progresiva y consistente en la fiabilidad con el incremento del tiempo operativo. El incremento de la tasa de fallas entre los diferentes componentes sugiere diferencias en su robustez y diseño, evidenciando una diversidad en la resistencia al desgaste y en los patrones de falla.

El análisis de los modelos estadístico para cada componente indica una predominancia del modelo de Weibull, aunque se observan variaciones, como el uso del modelo exponencial para ciertos componentes. Esta elección del modelo refleja los patrones subyacentes de fallo y desgaste, proporcionando una base sólida para el desarrollo de estrategias de mantenimiento predictivo y preventivo.

Tabla 3
Confiabilidad del sistema de molienda en diferentes intervalos de tiempo

Confiabilidad basada en el tiempo

| Tiempo (horas) | Molino unidad total | Reductor principal 4000kw - 1190/21.9rpm | Clasificador LSKS-76S | Generador de gases calientes ct6 | Molino vertical Loesche LM56.2+2c/s | Motor elect. Principal 4000 kW-1192 rpm | Rodillos esclavos 1-2 | Balanza dosificadora clinker tipo I CT6 290-140 | Ventilador extractor filtro ppal 290-665 |
|--------------------------|-------------------------|--|-------------------------|----------------------------------|-------------------------------------|---|-------------------------|---|--|
| 24 [1 día] | 0.997284133130 8150 | 0.972760092134 2660 | 0.986665333087 5870 | 0.915429979092 4970 | 0.921784147924 3730 | 0.978665760592 6490 | 0.959183873458 2270 | 0.983138730010 2300 | 0.975040680249 97300 |
| 168 [1 semana] | 0.555351870920 40000 | 0.875455791821 86400 | 0.918234572669 40200 | 0.702768772856 98600 | 0.664097335026 61600 | 0.859885768236 84400 | 0.846857299356 43900 | 0.873479744353 25300 | 0.832046069578 63600 |
| 336 [2 semanas] | 0.211401582429 81300 | 0.792275890417 75900 | 0.848043235356 80600 | 0.561277052749 51300 | 0.483104845169 08200 | 0.739403534416 26700 | 0.761784297452 94900 | 0.753409182783 20800 | 0.688799082173 20000 |
| 504 [3 semanas] | 0.088207142961 58650 | 0.723910475379 28400 | 0.784820901058 69400 | 0.462719408474 74700 | 0.361140966685 86200 | 0.635802576228 57000 | 0.695583515548 38500 | 0.646496683086 57800 | 0.569095916131 85800 |
| 720 [1 mes] | 0.033181104449 95410 | 0.649873271381 45400 | 0.711730217108 96300 | 0.370386027310 78600 | 0.254297041653 43300 | 0.523638817325 66500 | 0.626390569018 72000 | 0.528400848288 10200 | 0.444418024660 37100 |
| 1440 [2 meses] | 0.002745425847 51960 | 0.470264086158 21300 | 0.518369832645 35200 | 0.196667926501 09400 | 0.087715443396 51340 | 0.274197611010 22100 | 0.465001501730 10400 | 0.263093075782 07700 | 0.193139044198 23500 |
| 2160 [3 meses] | 0.000438865465 69830 | 0.351038037421 26100 | 0.380631630747 09100 | 0.114183093390 67100 | 0.033141289023 88200 | 0.143580512742 91500 | 0.360033525084 41700 | 0.127843729947 29900 | 0.083212645201 34990 |
| 4320 [6 meses] | 0.000009870313 20150 | 0.159995062432 32100 | 0.154682028560 98700 | 0.028638985234 18990 | 0.002345094544 58300 | 0.020615363639 51850 | 0.187834055614 32100 | 0.013492999567 16360 | 0.006465460805 82950 |
| 6480 [9 meses] | 0.000000724603 84550 | 0.078641010588 29200 | 0.064332811640 74700 | 0.008730917488 04450 | 0.000208152210 13980 | 0.002959964481 74370 | 0.107429872576 37300 | 0.001316542772 34000 | 0.000489198014 50960 |
| 8760 [1 año] | 0.000000086007 47130 | 0.039004442075 63500 | 0.025891042248 07620 | 0.002809157602 94370 | 0.000018707668 51200 | 0.000381552208 67870 | 0.063026097779 51620 | 0.000106659167 92120 | 0.000031472941 77920 |
| 13140 [1.5 años] | 0.00000003788 72640 | 0.011093903835 68560 | 0.004647758529 67530 | 0.000394037973 23260 | 0.000000240496 59990 | 0.000007452997 19580 | 0.025027163183 57530 | 0.000000760566 51640 | 0.000000155808 02050 |
| 17520 [2 años] | 0.00000000345 68690 | 0.003416912875 20420 | 0.000858703859 52340 | 0.000066436688 33150 | 0.000000003950 62700 | 0.000000145582 08790 | 0.010837414898 84660 | 0.000000004851 06910 | 0.00000000744 24540 |
| 21900 [2.5 años] | 0.000000000048 70800 | 0.001112880101 60600 | 0.000162033337 01510 | 0.000012698027 07170 | 0.000000000077 16680 | 0.000000002843 70750 | 0.004978255786 16330 | 0.000000000028 43160 | 0.000000000003 46120 |
| 26280 [3 años] | 0.000000000009 18900 | 0.000378238590 84480 | 0.000031087310 48500 | 0.000002664895 01860 | 0.000000000001 72120 | 0.000000000055 54720 | 0.002389723297 31530 | 0.000000000000 15560 | 0.000000000000 01580 |
| Mejor modelo estadístico | Lognormal | Weibull | Weibull | Weibull | Weibull | Exponencial | Weibull | Weibull | Weibull |

La confiabilidad tanto del sistema de molienda de cemento en su conjunto como de cada equipo individual se presenta en la Tabla 3 en intervalos de tiempo definidos. Por ejemplo, la confiabilidad del reductor principal es del 97.276 % después de 24 horas de funcionamiento, pero disminuye al 64.987 % después de 720 horas. Después de 8760 horas, la probabilidad de que el reductor principal opere sin fallas es del 3.900 %. En la Tabla 4, se detalla un resumen de las horas estimadas según valores predeterminados de confiabilidad en el tiempo.

Bajo los modelos estudiados, en la Tabla 4 se proyectaron los valores de confiabilidad en el tiempo para evaluar la función de supervivencia operativa del grupo de activos, tomando en cuenta intervalos de confiabilidad entre el 95% al 50% para estimar el tiempo de vida útil confiable en cada equipo.

Tabla 4*Vida útil confiable de Molino y equipos principales*

| Vida útil confiable de Molino y equipos principales | | | | | | | | | | |
|--|-----------------------------|-------|--------|--------|--------|--------|--------|--------|--------|--------|
| Equipo | Vida útil confiable (horas) | | | | | | | | | |
| | Mejor modelo estadístico | 95% | 90% | 85% | 80% | 75% | 70% | 65% | 60% | 50% |
| Molino unidad total | Lognormal | 55.40 | 72.41 | 86.74 | 100.13 | 113.25 | 126.49 | 140.14 | 154.45 | 186.14 |
| Reductor principal 4000kw - 1190/21.9rpm | Weibull | 26.65 | 64.97 | 111.10 | 164.49 | 225.28 | 293.96 | 371.34 | 458.56 | 669.08 |
| Clasificador LSKS-76S | Weibull | 50.66 | 108.05 | 170.49 | 238.01 | 310.95 | 389.89 | 475.58 | 568.99 | 784.51 |
| Generador de gases calientes ct6 | Weibull | 11.17 | 30.73 | 56.52 | 88.26 | 126.14 | 170.65 | 222.51 | 282.75 | 434.24 |
| Molino vertical Loesche LM56.2+2c/s | Weibull | 13.75 | 32.73 | 55.19 | 80.86 | 109.83 | 142.31 | 178.67 | 219.41 | 316.96 |
| Motor elect. Principal 4000 kW-1192 rpm | Exponencial | 57.08 | 117.26 | 180.87 | 248.34 | 320.16 | 396.95 | 479.42 | 568.50 | 771.41 |
| Rodillos esclavos 1-2 | Weibull | 8.10 | 22.30 | 41.02 | 64.06 | 91.58 | 123.91 | 161.59 | 205.36 | 315.47 |
| Balanza dosificadora clinker tipo I CT6 290-140 | Weibull | 39.01 | 76.65 | 115.11 | 155.00 | 196.72 | 240.69 | 287.33 | 337.16 | 448.97 |
| Ventilador extractor filtro ppal 290-665 | Weibull | 48.04 | 97.31 | 148.85 | 203.12 | 260.59 | 321.74 | 387.17 | 457.59 | 617.25 |
| | | Min | 22.30 | | | Min | 123.91 | | | |

En la Tabla 4, se denota que los principales subsistemas que comprometen la operación del molino y afectan a su disponibilidad son los rodillos esclavos y la unidad mecánica del molino quienes contribuyen a un rendimiento de 124 y 142 horas respectivamente al 70% de confiabilidad. Como marco comparativo, se conoce que los estándares de confiabilidad de clase mundial varían según la industria. Sin embargo, un buen referente proporcionado por la SMRP es del 70% para el sector de la manufactura industrial. Basado en los resultados operativos de la planta de molino y el KPI de 150 horas predeterminado como política de la empresa privada, se ha tomado como límite inferior de referencia un valor del 50% para la comparación de las estimaciones de valores de confiabilidad en los distintos subsistemas. Este enfoque permite establecer un marco de referencia claro para evaluar y mejorar continuamente la confiabilidad de los equipos en la planta

Tomando como ejemplo al molino como unidad total, en la Figura 6 muestra el gráfico de confiabilidad por aspersion utilizando Lognormal como la distribución de mejor ajuste, con un parámetro de forma (s) menor que 1, lo que indica una tendencia a fallas iniciales más rápidas seguidas por una disminución gradual de la tasa de falla y una estabilización con menor variabilidad en los tiempos hasta la falla a medida que los equipos envejecen.

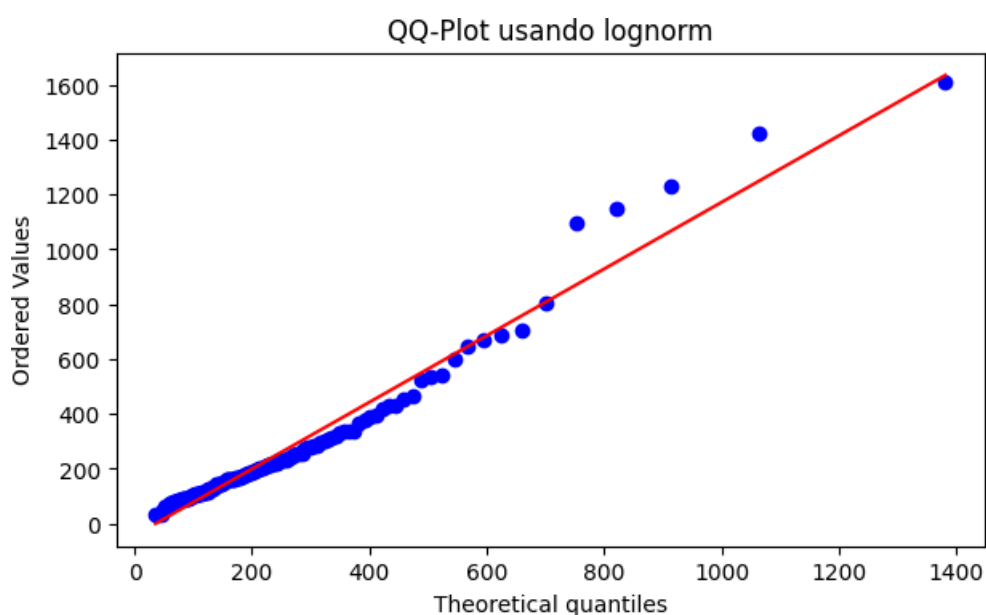


Figura 6 - Distribución lognormal para el molino unidad total

4.3. Análisis de Confiabilidad por métodos de ML

Para estimar la confiabilidad según métodos de ML, se ha optado por emplear un enfoque estadístico buscando la distribución más adecuada para los equipos en el sistema de molienda de cemento a través de los modelos Crow-AMSAA, Crow-AMSAA + ML Feature engineering y Crow_AMSAA + Validación cruzada Mayor K fold, utilizando los datos de TBF.

Se utilizaron modelos estadísticos de ML para estudio de confiabilidad mediante lenguaje de programación abierta, para este caso de estudio se utilizaron algoritmos en lenguaje *Python*, para identificar el mejor modelo que se ajuste a los datos de TBF para una distribución teórica a nivel de equipo, como se muestra en la Tabla 5. Los resultados presentados en la Tabla 5 cuentan con hasta siete decimales para resaltar la precisión en los cálculos del TBF para cada equipo, así como el valor i que representa el número de datos que se tuvo en cuenta para cada uno de ellos al momento de hacer los cálculos. En la Figura 7 podemos observar la secuencia de análisis para el modelo estadístico Crow_AMSAA + Validación cruzada Mayor K fold a través de un diagrama de flujo, en donde se detallan los pasos que siguen los bloques de programación para el análisis de los datos y procesamiento de resultados finales utilizados para la elaboración de la Tabla 5.

En la Tabla 5 se presenta un resumen de los cálculos de Tiempo Entre Fallas (TBF) obtenidos mediante los modelos de distribuciones de ML indicados. Se enfatiza la identificación de los modelos más adecuados según los parámetros de cada evaluación. Estas herramientas son críticas en el análisis estadístico, pues facilitan la comparación y elección de modelos estadísticos más adecuados para los datos reales, asegurando su concordancia con una distribución específica. Cada herramienta presenta ventajas y limitaciones particulares, siendo seleccionadas en función del contexto específico del análisis.

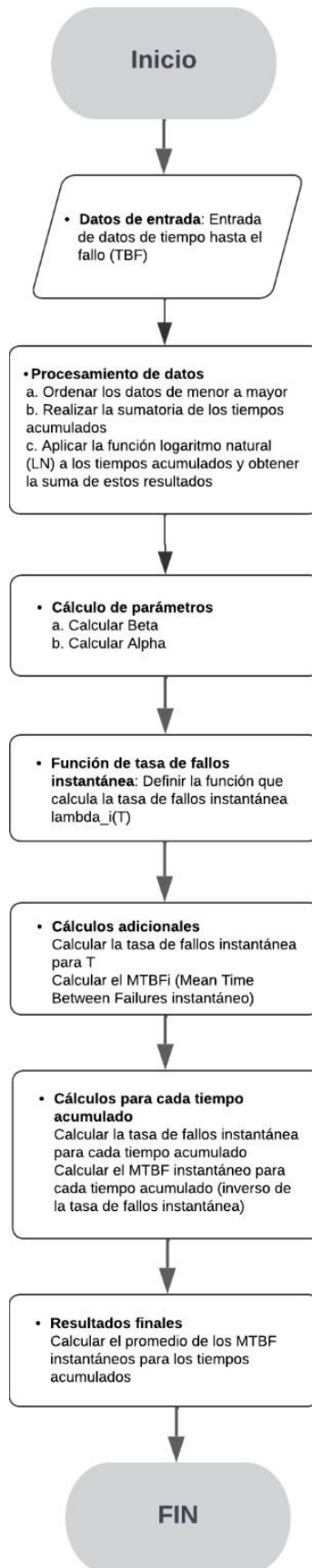


Figura 7 - Diagrama de flujo para análisis por modelo Crow_AMSAA + Validación cruzada Mayor K fold

Tabla 5*Mejor ajuste para los datos de TBF a nivel de equipo por métodos de ML*

| Modelo de estudio | Molino unidad total | Reductor principal 4000kw - 1190/21.9rpm | Clasificador LSKS-76S | Generador de gases calientes ct6 | Molino vertical Loesche LM56.2+2c/s | Motor elect. Principal 4000 kW- 1192 rpm | Rodillos esclavos 1-2 | Balanza dosificadora clinker tipo I CT6 290-140 | Ventilador extractor filtro ppal 290-665 |
|---|---------------------|--|--------------------------|--|--|---|--------------------------|--|--|
| | i = 136 | i = 13 | i = 17 | i = 19 | i = 36 | i = 33 | i = 21 | i = 21 | i = 26 |
| Crow_AMSAA, con optimizador | 269.8080333 | 2542.716302 | 2770.546377 | 2516.106897 | 1065.871847 | 1310.342363 | 12543.57704 | 1000.768472 | 1852.358753 |
| Híbrido 1: Crow_AMSAA + ML Feature engineering | 190.9351733 | 1029.248704 | 1370.117791 | 819.3618612 | 442.0399466 | 680.1040769 | 501.7749562 | 515.0542978 | 757.0018047 |
| Híbrido 2: Crow_AMSAA + Validación cruzada Mayor K fold | 240.1704627 | 2058.043452 | 2155.010405 | 1704.427896 | 878.4336215 | 1073.692407 | 1155.657443 | 1022.723301 | 1388.146421 |

Tabla 6

Confiabilidad basada en el tiempo, basada en el modelo Híbrido 2: Crow_AMSAA + técnica de validación cruzada

| Tiempo (horas) | Molino unidad total | Reductor principal 4000kw - 1190/21.9rpm | Clasificador LSKS-76S | Generador de gases calientes ct6 | Molino vertical Loesche LM56.2+2c/s | Motor elect. Principal 4000 kW- 1192 rpm | Rodillos esclavos 1-2 | Balanza dosificadora clinker tipo I CT6 290-140 | Ventilador extractor filtro ppal 290-665 |
|------------------|-------------------------|--|-------------------------|----------------------------------|-------------------------------------|--|-------------------------|---|--|
| 24 [1 día] | 0.7547994454856 6100 | 0.86153703648156 300 | 0.86788306216387 000 | 0.81452216908166 900 | 0.79545776794016 300 | 0.82447227571875 400 | 0.79136872515681 300 | 0.84241985204088 400 | 0.85863115162629 100 |
| 168 [1 semana] | 0.4398338784271 9900 | 0.70998183588491 700 | 0.71975913885336 300 | 0.65523450091448 100 | 0.60200233376652 900 | 0.63886006668556 200 | 0.61550561007030 200 | 0.65186942101332 400 | 0.68682158096894 500 |
| 336 [2 semanas] | 0.3002663741584 7000 | 0.63084806691533 300 | 0.64157313575339 500 | 0.57870967179266 600 | 0.50967821713271 700 | 0.54616771367774 800 | 0.53295207919802 100 | 0.55284349693630 300 | 0.59567727950085 100 |
| 504 [3 semanas] | 0.2222284901963 8200 | 0.57815617862438 200 | 0.58923253916058 900 | 0.52946397729080 200 | 0.45128997892671 000 | 0.48633799764438 000 | 0.48063944208841 900 | 0.48817095974474 400 | 0.53515891176006 800 |
| 720 [1 mes] | 0.1603401874372 0100 | 0.52824743026386 900 | 0.53945912358070 200 | 0.48383799254572 800 | 0.39823715076284 000 | 0.43120221163343 000 | 0.43281109744814 700 | 0.42829342665508 400 | 0.47825631485751 800 |
| 1440 [2 meses] | 0.0684833448384 6470 | 0.42384899146435 700 | 0.43473600937785 400 | 0.39090946030071 000 | 0.29442154050996 600 | 0.32127164511066 400 | 0.33757680577794 000 | 0.30898796491879 100 | 0.36162505132346 000 |
| 2160 [3 meses] | 0.0350182798630 7660 | 0.36027342490822 000 | 0.37055906063526 600 | 0.33554209616205 100 | 0.23610506428132 800 | 0.25839105238551 900 | 0.28245352366635 700 | 0.24147679791264 200 | 0.29302540637675 300 |
| 4320 [6 meses] | 0.0073748675379 6399 | 0.25331571049098 900 | 0.26187181838061 700 | 0.24346053769684 200 | 0.14704644709632 100 | 0.16093850825304 200 | 0.19409732130964 800 | 0.13971673973153 400 | 0.18402337961530 400 |
| 6480 [9 meses] | 0.0021594912949 6900 | 0.19532569266501 400 | 0.20253990362110 900 | 0.19349098190603 000 | 0.10403015537886 200 | 0.11336431177695 200 | 0.14829821579338 000 | 0.09243318043437 030 | 0.12967221334281 500 |
| 8760 [1 año] | 0.0007128176015 5305 | 0.15601613395694 100 | 0.16214220291694 300 | 0.15925917272586 600 | 0.07728244286748 260 | 0.08369395102502 250 | 0.11802065783724 800 | 0.06433141251939 300 | 0.09544899870199 590 |
| 13140 [1.5 años] | 0.0001163373734 6436 | 0.10975261156237 100 | 0.11439265529212 300 | 0.11813134917265 500 | 0.04868011669596 460 | 0.05200418312932 790 | 0.08310062103481 670 | 0.03616593267757 240 | 0.05871993660885 290 |
| 17520 [2 años] | 0.0000245695204 4234 | 0.08218603565858 050 | 0.08582230459172 310 | 0.09283494809054 660 | 0.03336717027317 340 | 0.03513933264609 290 | 0.06259914152651 880 | 0.02236661856868 110 | 0.03918547030018 250 |
| 21900 [2.5 años] | 0.0000061336895 5577 | 0.06399475014074 270 | 0.06691403946993 090 | 0.07559121901819 210 | 0.02410445009238 460 | 0.02502329009184 480 | 0.04915466840132 250 | 0.01469370586079 430 | 0.02752895572520 310 |
| 26280 [3 años] | 0.0000017268633 0212 | 0.05120828094374 680 | 0.05359497589381 200 | 0.06307532242894 860 | 0.01806099096244 340 | 0.01848579847434 650 | 0.03971958968614 780 | 0.01007315736344 530 | 0.02005302301667 210 |
| | k = 20 folds | k = 12 folds | k = 15 folds | k = 17 folds | k = 33 folds | k = 30 folds | k = 18 folds | k = 20 folds | k = 24 folds |

Tabla 7

Vida útil confiable de Molino y equipos principales Híbrido 2: Crow_AMSAA y técnica de validación cruzada

| Equipo | Modelo | Vida útil confiable (horas) | | | | | | | | |
|---|------------------------|-----------------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| | | 95% | 90% | 85% | 80% | 75% | 70% | 65% | 60% | 50% |
| Molino unidad total | Mayor R2, k = 20 folds | 1 | 4 | 9 | 16 | 25 | 37 | 52 | 71 | 123 |
| Reductor principal 4000kw - 1190/21.9rpm | Mayor R2, k = 12 folds | 2 | 11 | 29 | 62 | 112 | 185 | 287 | 428 | 873 |
| Clasificador LSKS-76S | Mayor R2, k = 15 folds | 2 | 12 | 33 | 69 | 123 | 203 | 314 | 465 | 942 |
| Generador de gases calientes ct6 | Mayor R2, k = 17 folds | 1 | 4 | 13 | 30 | 60 | 106 | 177 | 280 | 636 |
| Molino vertical loesche LM56.2+2c/s | Mayor R2, k = 33 folds | 1 | 4 | 10 | 23 | 42 | 71 | 113 | 171 | 360 |
| Motor elect. Principal 4000 kW-1192 rpm | Mayor R2, k = 30 folds | 1 | 6 | 16 | 34 | 60 | 99 | 153 | 227 | 460 |
| Rodillos esclavos 1-2 | Mayor R2, k = 18 folds | 0 | 3 | 9 | 21 | 42 | 74 | 122 | 193 | 435 |
| Balanza dosificadora clinker tipo I CT6 290-140 | Mayor R2, k = 18 folds | 2 | 9 | 21 | 42 | 72 | 114 | 170 | 245 | 469 |
| Ventilador extractor filtro ppal 290-665 | Mayor R2, k = 24 folds | 2 | 11 | 28 | 55 | 94 | 150 | 226 | 326 | 630 |

La Tabla 6 exhibe la confiabilidad tanto del conjunto del sistema de molienda de cemento como de cada equipo individual, destacando variaciones en diferentes lapsos de tiempo. Por ejemplo, el reductor principal muestra una confiabilidad del 86.154% tras 24 horas de operación, disminuyendo al 52.824% después de 720 horas. Tras 8760 horas, la probabilidad de funcionamiento sin fallos del reductor principal se reduce al 15.601%.

Esta tabla detalla la función confiabilidad en función del tiempo para los componentes del sistema de molino. Los datos revelan patrones notables en la disminución de la confiabilidad a lo largo del tiempo, abarcando desde 24 horas hasta tres años.

La tabla 7 ofrece un resumen de las horas estimadas según valores predeterminados de confiabilidad en distintos periodos. Proyectamos los valores de confiabilidad para evaluar la eficiencia operativa del conjunto de activos, considerando un estándar de 150 horas como KPI de la empresa. Hemos establecido un umbral del 50% como referencia comparativa entre las estimaciones de confiabilidad de los diversos subsistemas, permitiendo una evaluación constante y mejora de la confiabilidad de los equipos. Es notable que los rodillos esclavos y la unidad mecánica del molino, como subsistemas clave, contribuyen con 435 y 360 horas respectivamente al 50% de confiabilidad, comprometiendo la operatividad del molino.

4.4. Comparación de métodos de análisis de distribuciones convencionales y de ML

La evaluación comparativa entre los métodos estadísticos convencionales y el modelo de *machine learning* con transformación logarítmica y validación cruzada, considerando parámetros como el Error Cuadrático Medio (MSE), Coeficiente de Determinación (R²) y Error Absoluto Medio (MAE), revela sus diferencias y similitudes en el análisis predictivo. Los métodos convencionales, se basan en supuestos específicos sobre la relación entre variables y distribuciones paramétricas, siendo útiles cuando se cumplen tales condiciones. Sin embargo, estos métodos pueden subestimar la complejidad de los datos al asumir linealidad, lo que limita su capacidad

predictiva en escenarios no lineales.

Por otro lado, los enfoques de *machine learning*, ofrecen flexibilidad y capacidad para capturar patrones complejos en datos no lineales. Estos modelos se adaptan mejor a la variabilidad inherente a muchos conjuntos de datos reales, lo que les permite generar predicciones más precisas y robustas en comparación con los métodos convencionales.

El MSE, R^2 y MAE son métricas clave para evaluar la precisión y ajuste de los modelos. Mientras que el MSE y MAE miden la discrepancia entre las predicciones y los valores reales, donde un menor valor indica una mejor precisión, el R^2 cuantifica la proporción de la variabilidad de los datos que es explicada por el modelo, con valores cercanos a 1 que denotan un mejor ajuste. En la comparación, los modelos de *machine learning* tienden a mostrar un mejor rendimiento en términos de MSE, R^2 y MAE, especialmente en conjuntos de datos complejos y no lineales, debido a su capacidad para manejar relaciones más intrincadas entre las variables.

En la Tabla 8 se muestran los valores de MSE, R^2 y MAE para cada distribución utilizada en la presente investigación, resaltando los métodos más precisos en el modelo convencional y ML, así como los valores de R^2 que indicará que modelo tiene mejor ajuste a los datos.

Tabla 8
Evaluación MSE y R² para los modelos convencionales y de ML

| Modelo de estudio | Molino unidad total | Reductor principal 4000kw - 1190/21.9rpm | Clasificador LSKS-76S | Generador de gases calientes ct6 | Molino vertical Loesche LM56.2+2c/s | Motor elect. Principal 4000 kW- 1192 rpm | Rodillos esclavos 1-2 | Balanza dosificadora clinker tipo I CT6 290-140 | Ventilador extractor filtro ppal 290-665 |
|-------------------|--|--|---|--|--|---|---|---|--|
| | i = 136 | i = 13 | i = 17 | i = 19 | i = 36 | i = 33 | i = 21 | i = 21 | i = 26 |
| Normal | MSE: 0.018679887362856605 R2: 0.77406798866547745 MAE: 0.11742386392127159 | MSE: 0.02661770511021642 R2: 0.6786862740266733 MAE: 0.12941859658696345 | MSE: 0.010215765080729087 R2: 0.8769851621528872 MAE: 0.08339475871852974 | MSE: 0.0182360519031011 R2: 0.7746073103125912 MAE: 0.10512839168162819 | MSE: 0.0189185357721835 R2: 0.7728006393615586 MAE: 0.12226622500452547 | MSE: 0.010932730467917049 R2: 0.8686936475749392 MAE: 0.09389235608671627 | MSE: 0.02310995276265704 R2: 0.722050295409134 MAE: 0.12753194917615437 | MSE: 0.008108533736355566 R2: 0.90247645333456 MAE: 0.07204338663655516 | MSE: 0.012733902724422642 R2: 0.8469667868140497 MAE: 0.09061229559694867 |
| Exponencial | MSE: 0.008050844768908898 R2: 0.9026255601847856 MAE: 0.07453043982465826 | MSE: 0.017657397325808337 R2: 0.7868499894241708 MAE: 0.11336260402745756 | MSE: 0.004626151244409607 R2: 0.9442934287652343 MAE: 0.06161729404194715 | MSE: 0.02764234057876313 R2: 0.6583481158417119 MAE: 0.12585892882377037 | MSE: 0.008405844596612163 R2: 0.8990512510601153 MAE: 0.07436878458302824 | MSE: 0.0023750331018297043 R2: 0.9714749271094532 MAE: 0.039633502996993704 | MSE: 0.0283973711855818 R2: 0.6584570720134117 MAE: 0.13268849308947817 | MSE: 0.0016937551625465647 R2: 0.9796287447268264 MAE: 0.034546869917624914 | MSE: 0.002230608518827119 R2: 0.9731930425115176 MAE: 0.03616371809048601 |
| Exponencial 2p | MSE: 129183.45182724654 R2: - 0.9851611538427325 MAE: 253.1966609387274 | MSE: 13882293.137024827 R2: - 0.6373608286552574 MAE: 2324.615140853431 | MSE: 8870771.480054418 R2: -1.337635213994226 MAE: 2252.999778231343 | MSE: 9858738.236680103 R2: - 0.6622274492753011 MAE: 1981.8418129659058 | MSE: 3274221.946655908 R2: - 0.4914052008485732 MAE: 1038.6661331110136 | MSE: 3195954.2770364177 R2: -0.76000447233153 MAE: 1174.764268567534 | MSE: 6304883.981616246 R2: - 0.47079244044629465 MAE: 1420.6186266340806 | MSE: 2049072.2136312467 R2: - 1.2866351116125827 MAE: 1073.761448959621 | MSE: 4561673.47216223 R2: - 0.9921594560274538 MAE: 1507.2688947453523 |
| Lognormal | MSE: 0.0011942817385058031 R2: 0.985552406478328 MAE: 0.029667352501949112 | MSE: 0.0034836383494817854 R2: 0.9579475084955413 MAE: 0.04891486737666525 | MSE: 0.003752013854495081 R2: 0.9548194998354551 MAE: 0.05114263960518611 | MSE: 0.008209279420666177 R2: 0.8985355175094236 MAE: 0.07429157775238562 | MSE: 0.003079609904855359 R2: 0.9630158797792524 MAE: 0.04555425791797784 | MSE: 0.0032795269252810268 R2: 0.9606116038896118 MAE: 0.04816927753353385 | MSE: 0.005495910253841209 R2: 0.9338991884924371 MAE: 0.06169040458834132 | MSE: 0.003755679342561722 R2: 0.9548294202708258 MAE: 0.05037572212398417 | MSE: 0.0013127596200271283 R2: 0.984223546610874 MAE: 0.031682713368311084 |

| | | | | | | | | | |
|------------|---|---|---|--|--|---|---|--|---|
| Weibull 2P | MSE: 0.005710940857421086 R2: 0.9309265446333324 MAE: 0.06473546138641575 | MSE: 0.006209676171597538 R2: 0.9250403376428583, MAE: 0.06560503746305803 | MSE: 0.00369777579263495 R2: 0.9554726164970209 MAE: 0.054826884837485934 | MSE: 0.008666624400746024 R2: 0.892882856740362 MAE: 0.06639435847473293 | MSE: 0.00285158585235209 R2: 0.9657543009532157 MAE: 0.04410839566818792 | MSE: 0.0017993506254271856 R2: 0.9783890979429213 MAE: 0.034594636538099785 | MSE: 0.007075220065016975 R2: 0.9149043986725686 MAE: 0.06789789776378867 | MSE: 0.002163898539466224 R2: 0.9739742021116926 MAE: 0.037995074359524816 | MSE: 0.002553185026657837 R2: 0.9693163897240765 MAE: 0.03866409815075441 |
|------------|---|---|---|--|--|---|---|--|---|

| | | | | | | | | | |
|------------|---|---|--|--|---|---|---|---|---|
| Weibull 3P | MSE: 0.0002820312492770739 R2: 0.9965888505247553 MAE: 0.013462782261425782 | MSE: 0.0008863106338112055 R2: 0.9893009644918505 MAE: 0.024494479138144086 | MSE: 0.0020633011445532795 R2: 0.9751544153843376 MAE: 0.03691796328965671 | MSE: 0.002589026046283508 R2: 0.968000335415619 MAE: 0.04264043308811108 | MSE: 0.0010774975509981514 R2: 0.9870599523332977 MAE: 0.028438048500560927 | MSE: 0.0014441726905211423 R2: 0.9826549233221564 MAE: 0.031225461168296928 | MSE: 0.0013156428516533244 R2: 0.9841764046114787 MAE: 0.027968034719191365 | MSE: 0.0012947404499040133 R2: 0.9844278034979727 MAE: 0.029679947414174537 | MSE: 0.0010943203848495621 R2: 0.9868487007971857 MAE: 0.025731450860293388 |
|------------|---|---|--|--|---|---|---|---|---|

| | | | | | | | | | |
|------------|---|--|---|--|--|---|--|---|--|
| Crow_AMSAA | MSE: 22.865276509951485 R2: 0.9851644596853518 MAE: 4.103057847696444 | MSE: 0.6767416418476242 R2: 0.9516613112965983 MAE: 0.6851719990763119 | MSE: 1.564959245069611 R2: 0.9347933647887662 MAE: 1.0975830641646838 | MSE: 4.627127128137524 R2: 0.8457624290620825 MAE: 1.7528415464372202 | MSE: 3.4020515652448924 R2: 0.9684751978510125 MAE: 1.4653381559190306 | MSE: 4.737196482205526 R2: 0.9507823742108517 MAE: 1.8340329253577181 | MSE: 1.9611650188350058 R2: 0.9465136813044999 MAE: 1.1058763864696541 | MSE: 1.483425611578671 R2: 0.9595429378660363 MAE: 0.9372300143977326 | MSE: 2.4959252033632575 R2: 0.9556279963846532 MAE: 1.3691707403260147 |
|------------|---|--|---|--|--|---|--|---|--|

| | | | | | | | | | |
|---|---|--|---|---|---|--|---|--|--|
| Híbrido 1: Crow_AMSAA + ML Validación cruzada | MSE: 23.579867911382173 R2: 0.9837243082841123 MAE: 4.165148630494467 | MSE: 1.1163260253689917 R2: 0.37846427277452566 MAE: 0.9029641891528473 | MSE: 2.554746372953472 R2: 0.6550308697404732 MAE: 1.4414322524145926 | MSE: 5.889387165960654 R2: 0.7571927928007163 MAE: 1.979529546208639 | MSE: 3.640647517179965 R2: 0.9565203548753869 MAE: 1.4978716058278394 | MSE: 5.296744825249913 R2: 0.9305073420443402 MAE: 1.901752662956035 | MSE: 3.395165867068063 R2: 0.9183171413180254 MAE: 1.2761537507113931 | MSE: 6.206297106004179 R2: - 3.2457199969037704 MAE: 1.5312243897480928 | MSE: 2.922343155103744 R2: 0.9186631934824335 MAE: 1.478647417745464 |
|---|---|--|---|---|---|--|---|--|--|

| | | | | | | | | | |
|---|---|--|---|--|---|---|---|--|---|
| Híbrido 2: Crow_AMSAA + ML Feature engineering | MSE: 0.23936193367432224 R2: 0.9998446962311927 MAE: 0.4074393406365722 | MSE: 0.01486591108292538 R2: 0.9989381492083624 MAE: 0.08988984871744911 | MSE: 0.05457561830503812 R2: 0.9977260159039567 MAE: 0.1933390225618248 | MSE: 0.1400061148887915 R2: 0.995333129503707 MAE: 0.311028197709754 | MSE: 0.07555709624558173 R2: 0.999299857023207 MAE: 0.23183814622241927 | MSE: 0.13036355085001883 R2: 0.9986455734976621 MAE: 0.3022069945252552 | MSE: 0.05120391893708696 R2: 0.998603529483534 MAE: 0.17732956424086307 | MSE: 0.043855754938149916 R2: 0.9988039339562322 MAE: 0.1508460158696117 | MSE: 0.09096831525275119 R2: 0.9983827855066177 MAE: 0.2515102442811446 |
|---|---|--|---|--|---|---|---|--|---|

Los resultados de la Tabla 8 proporcionan una visión variada de cómo diferentes modelos estadísticos se desempeñan en la predicción de la confiabilidad de un sistema. Los modelos exponencial y lognormal muestran un ajuste significativamente mejor, exhibiendo valores altos de R^2 y MSE considerablemente más bajos en comparación con otros modelos. Esto sugiere que estos modelos podrían ser más apropiados para explicar la variabilidad en los datos y predecir la confiabilidad del sistema de manera más precisa.

Por otro lado, los modelos Weibull 2P y 3P también presentan resultados con valores de R^2 cercanos a 1 y MSE bajos, lo que indica una buena capacidad para ajustarse a los datos.

Sin embargo, el modelo exponencial 2p muestra un rendimiento muy pobre, evidenciado por un MSE extremadamente alto y un R^2 negativo, lo que sugiere una discrepancia significativa entre las predicciones del modelo y los datos reales.

El modelo híbrido 2 (Crow_AMSAA + ML Feature engineering) destaca por su excelente ajuste, presentando un MSE muy bajo y un R^2 cercano a 1, lo que sugiere una capacidad muy sólida para predecir la confiabilidad del sistema.

V. DISCUSIÓN

El proceso de análisis inicial de las fallas en la planta de molienda de cemento se basó en una revisión detallada de la matriz de datos de paradas. Según la metodología propuesta, se utilizó el principio de Pareto para discernir cuáles equipos presentaban mayor frecuencia de fallos y requerían tiempos extensos de reparación. En concordancia con el enfoque propuesto por Armstrong (2023). Este método permitió identificar los equipos con mayor impacto, sentando las bases para el análisis de confiabilidad y el cálculo de los MTBF.

Tras establecer qué equipos eran críticos, se procedió con la estimación de los Tiempos Medios Entre Fallos (MTBF) para cada equipo identificado bajo distribuciones estadística convencionales (V. G. Talkit et al., 2022). La importancia de estos cálculos reside en su capacidad para

proporcionar una medida objetiva del rendimiento y la fiabilidad de los equipos en cuestión. Esta metodología se apoya en la investigación de Govindrao, et al. (2023) y V. Talkit et al. (2022), quienes enfatizaron la importancia del análisis de confiabilidad en la modificación de programas de mantenimiento. Este estudio proporciona un marco teórico sólido, vinculando el análisis de Pareto, como se describe en el trabajo de Armstrong. Por lo tanto, podemos afirmar que nuestra metodología para identificar equipos críticos está basada en principios teóricos definidos según lo establecido.

Durante la investigación, se recurrió a Microsoft Excel para la organización y manejo preliminar de los datos considerando las fichas de recolección y análisis de datos, así como análisis de Pareto. Posteriormente, para realizar análisis más complejos, se optó por el lenguaje de programación Python, específicamente mediante la interfaz de Jupyter Notebook. Se aprovecharon librerías como SciPy, especialmente su módulo stats para ajustar curvas y modelar distribuciones estadísticas. La plataforma de código abierto Python proporcionó flexibilidad al desarrollo de modelos, permitiendo no solo realizar cálculos complejos sino también organizar diversos escenarios con una alta precisión en los resultados. Este enfoque computacional aportó un dinamismo significativo al estudio, decisivo para la obtención de conclusiones confiables.

Se aplicó una metodología estadística comprensiva para determinar la distribución de probabilidad más adecuada para modelar los tiempos entre fallas. Utilizando técnicas como la inferencia Bayesiana y pruebas de bondad de ajuste como Kolmogorov-Smirnov, Anderson-Darling y criterios de información como Akaike (AIC) y Bayesiano (BIC), se analizaron distintas distribuciones estadísticas. Los procedimientos estadísticos se implementaron eficazmente en Python, lo que permitió evaluar la idoneidad de cada distribución para cada equipo y para la unidad en su conjunto. Los resultados revelaron una heterogeneidad en los patrones de fallas entre los equipos, con variaciones significativas en las distribuciones de los tiempos entre fallas. Esta diversidad en los perfiles de fallas subraya la importancia de un enfoque de mantenimiento diferenciado y adaptativo, que considere las características específicas de cada equipo para optimizar las estrategias

de mantenimiento preventivo y predictivo.

En el análisis de confiabilidad de los equipos, se emplearon modelos estadísticos convencionales según lo propuesto por (V. Talkit et al., 2022), incluyendo modelos de distribución normal, exponencial, lognormal y de Weibull. Estos modelos facilitaron el cálculo del Tiempo Medio Entre Fallos (TBF) y proporcionaron una perspectiva detallada sobre el rendimiento y la fiabilidad de cada equipo. Las distintas distribuciones, al estimar la tasa de fallos (λ), y luego integrar la función de supervivencia, fue determinante para predecir la operatividad sin fallos de los equipos. Este enfoque diversificado permitió estimar valores de confiabilidad diversos según los modelos. La inclusión de modelos para diferentes equipos permitió abordar una gama más amplia de patrones de fallos, desde fallos aleatorios hasta aquellos relacionados con el envejecimiento u horas de uso. Este análisis integral subraya la importancia de estrategias de mantenimiento adaptativas y orientadas a datos, fundamentales en la mejora continua y la eficiencia de los procesos productivos.

Este enfoque incluyó también la inversión del input de datos, asignando valores fijos de confiabilidad para posteriormente determinar los correspondientes valores de tiempo. Esta estrategia permitió correlacionar de manera más efectiva los parámetros de tiempo y confiabilidad. Así, se facilitó la estimación de aquellos componentes o equipos que más influyen en las paradas del sistema debido a su baja confiabilidad, ofreciendo insights valiosos para la optimización del mantenimiento y la operatividad del sistema.

La metodología incluyó una inversión en el enfoque de entrada de datos, asignando valores fijos de confiabilidad para determinar los correspondientes tiempos de operación sin fallos. Esta estrategia, innovadora en su aplicación, facilitó la correlación entre los parámetros de tiempo y confiabilidad, permitiendo identificar los componentes o equipos que, debido a su baja confiabilidad, influyen significativamente en las paradas del sistema. Los resultados de este análisis proveen insights esenciales para la optimización del mantenimiento y la mejora de la operatividad del sistema.

La adaptación del modelo de Weibull en particular demostró ser eficaz para este propósito. La estimación de parámetros como la forma (c) y la escala (n -eta o λ según diversas bibliografías), junto con la integración de estos en cálculos de la tasa de fallos y la función de supervivencia, brindó un marco robusto para evaluar la confiabilidad a lo largo del tiempo. Esta aproximación proporcionó una comprensión más profunda de los patrones de fallo y la durabilidad de los equipos, destacando la importancia de un enfoque basado en datos para la toma de decisiones en la gestión de la fiabilidad y el mantenimiento de equipos en entornos de producción complejos.

En el análisis de la vida útil confiable de los componentes clave de un sistema de molino, se identificó una correlación significativa entre los modelos estadísticos adoptados y la proyección de la confiabilidad en el tiempo. La mayoría de los equipos, siguiendo los modelos de lognormal y Weibull con la excepción del Motor eléctrico Principal que se ajusta a un modelo exponencial, exhiben una disminución en la confiabilidad a medida que aumentan las horas de operación. Esta variabilidad en los patrones de fallo, reflejada en los diferentes modelos estadísticos, subraya la importancia de una estrategia de mantenimiento diferenciada y específica para cada componente. Por ejemplo, mientras el molino unidad total mantiene una confiabilidad del 70% hasta aproximadamente 126.49 horas, el reductor principal muestra una vida útil confiable de 293.96 horas, significativamente mayor al mismo nivel de confiabilidad.

En el estudio realizado en la planta de molienda de cemento, se puso especial énfasis en el análisis de la vida útil confiable de los equipos al umbral del 50% de confiabilidad, un nivel establecido como referencia óptima según los datos históricos de confiabilidad de 2019 a 2023. Este análisis reveló variaciones notables en la duración de la confiabilidad de los equipos a este nivel. Por ejemplo, se encontró que el Molino unidad total tenía una vida útil confiable de 186.14 horas al 50% de confiabilidad, en contraste con el Ventilador extractor filtro principal, que alcanzaba las 617.25 horas. Esta notable diferencia subraya la importancia de un enfoque de mantenimiento personalizado y basado en datos. La capacidad de identificar y responder a

las variaciones específicas en la vida útil confiable de cada equipo permite intervenciones de mantenimiento más precisas y oportunas. Este enfoque no solo maximiza la eficiencia operativa, sino que también contribuye a prolongar la vida útil de los equipos dentro del contexto específico de la planta de molienda. Este tipo de análisis detallado y orientado a datos es fundamental para la gestión estratégica de activos y la optimización de procesos en entornos industriales complejos, donde cada equipo puede presentar patrones de desgaste y fallo únicos.

El estudio destacó la relevancia de integrar métodos de Machine Learning, como la ingeniería de características y la validación cruzada, en el análisis de confiabilidad. Estas técnicas, aunque no específicamente contrastadas en la literatura para este fin, ofrecen una base robusta para el desarrollo de modelos predictivos híbridos. La combinación del método Chrome AMSA con estas técnicas de Machine Learning permitió evaluar y comparar métricas clave como MSE, R^2 y MAE, para determinar el modelo más preciso y con menor error en la predicción del MTBF. Este enfoque híbrido es fundamental para mejorar las estimaciones de confiabilidad, proponiendo un camino innovador en la modelización de la fiabilidad operativa de los equipos.

VI. CONCLUSIONES

- 4.1 La recopilación y análisis detallado de los datos históricos de paradas del sistema revelaron patrones críticos de fallas. Empleando técnicas avanzadas de manejo de datos y análisis de Pareto, se identificaron los equipos y componentes que contribuyen mayormente a los tiempos de paradas, lo que permitió una focalización más eficiente de los esfuerzos. Este análisis mostró que un número reducido de equipos responsables de la mayoría de las interrupciones, para los cuales se centró posteriormente el estudio de confiabilidad.
- 4.2 Se realizó un análisis de los tiempos entre falla mediante la utilización de transformación logarítmica y validación cruzada, logrando una evaluación integral de la confiabilidad en el sistema de molienda de cemento en Pacasmayo. La aplicación de la transformación logarítmica

y validación cruzada ha demostrado ser una herramienta esencial para comprender la distribución de los tiempos entre falla, permitiendo modelar de manera más efectiva la relación entre los datos y la confiabilidad del sistema, y evaluar la precisión de los modelos empleados, asegurando una estimación robusta y confiable de la confiabilidad del sistema de molienda de cemento.

- 4.3 El cálculo del TBF en el sistema de molienda de cemento a través de métodos convencionales proporciona una métrica fundamental para evaluar la confiabilidad de los equipos. La precisión en la estimación del tiempo medio entre fallas ofrece una visión clara sobre la vida útil de los componentes, seleccionando la distribución que mejor se ajuste a la serie de datos que se tienen en cada equipo y sistema.
- 4.4 La implementación de técnicas estadísticas tradicionales en la valoración de la confiabilidad operativa de los sistemas de molienda de cemento ha establecido un marco referencial para la interpretación de la dinámica de fallos y eficiencia operacional. Estos métodos, fundamentados en principios de estadística aplicada, han proporcionado una visión preliminar del comportamiento de fallos y la resiliencia operativa.
- 4.5 La integración de algoritmos avanzados de Machine Learning para el cálculo del TBF ha revelado divergencias notables en precisión y capacidad analítica. La selección del algoritmo óptimo se basó en criterios de exactitud analítica, manejo eficiente de grandes conjuntos de datos, y rendimiento computacional superior.
- 4.6 La aplicación de TBF derivado de técnicas de Machine Learning ha proporcionado una mejora sustancial en la precisión de la estimación de la confiabilidad de los componentes del sistema de molienda. Estas metodologías avanzadas han facilitado una comprensión más matizada y detallada de la fiabilidad operativa en comparación con los enfoques estadísticos tradicionales.
- 4.7 La propuesta de comparación ha demostrado una superioridad manifiesta de los modelos de *machine learning*, particularmente del modelo híbrido Crown AMSAA combinado con técnicas de validación

cruzada de 24 *folds*, sobre los métodos estadísticos tradicionales. Esto se evidencia especialmente en la unidad equipo molino, identificada como un componente crítico, alcanzando un 50% de confiabilidad con un TBF de 123 horas. Las estimaciones de TBF obtenidas a través de métodos tradicionales resultaron ser inferiores, subrayando la eficacia y precisión de los modelos de Machine Learning en la predicción y análisis de la confiabilidad.

- 4.8 Adicionalmente, se identificó que otros componentes, como el molino vertical y los rodillos esclavos, presentan índices de confiabilidad bajos, con un 50% y TBF de 360 y 435 horas respectivamente. Estos resultados, inferiores a las estimaciones obtenidas por métodos estadísticos convencionales, destacan la capacidad de los modelos de Machine Learning para identificar con precisión los puntos críticos y vulnerabilidades en el sistema de molienda de cemento.
- 4.9 Es fundamental señalar que, aunque algunos modelos de regresión mostraron resultados prometedores, con un coeficiente R^2 aproximándose a 1, estos no lograron replicar la precisión al predecir valores de MTBF y, por ende, estimar la confiabilidad a lo largo del tiempo. Estos hallazgos sugieren una discrepancia entre los valores predichos y los datos históricos reales, incluyendo registros de los últimos cinco años. Esta situación indica que, a pesar de una evaluación favorable en términos de MSE, MAE, o R^2 , la calidad y relevancia del *input* de datos son determinantes críticos para la precisión de las predicciones de MTBF y la estimación de la confiabilidad. La eficacia predictiva de un modelo de regresión no solo depende de su robustez analítica, sino también de la adecuación y exactitud del conjunto de datos utilizado. Por tanto, una consideración meticulosa de la calidad y pertinencia de los datos es esencial para garantizar resultados confiables y representativos en el análisis de confiabilidad.

VII. RECOMENDACIONES

- La aplicación del análisis de Pareto en estudios de confiabilidad, con

especial atención en la censura de datos que se encuentran fuera de frecuencia o enfoque. Esta censura es crucial porque los modelos de Machine Learning y sus métodos de evaluación de regresiones, tales como MSE, R^2 o MAE, ofrecen perspectivas diferentes en comparación con los ajustes de distribuciones estadísticas. La inclusión de datos no censurados puede comprometer la precisión de los modelos. Por lo tanto, es esencial asegurarse de que los datos utilizados en los modelos de Machine Learning hayan sido adecuadamente revisados y filtrados para evitar distorsiones en los resultados y conclusiones.

- La importancia de verificar la cantidad y calidad de los datos disponibles para el desarrollo de un modelo de regresión y ajuste probabilístico eficaces. Se ha observado que, aunque ciertos tiempos de falla pueden ser relevantes para el criterio del 80-20 de Pareto, en ocasiones la frecuencia de estas fallas es demasiado baja. Esto resulta en una cantidad insuficiente de eventos de falla con duración significativa, impidiendo así una predicción precisa. En tales casos, tanto los ajustes estadísticos como los modelos de Machine Learning pueden distorsionar y exagerar las predicciones debido a la escasez de datos. Esta situación nos llevó a la necesidad de censurar algunos equipos por falta de datos suficientes, los cuales no estaban acumulados de manera adecuada hasta la fecha del análisis.
- Una recomendación técnica esencial, derivada de nuestro análisis, se relaciona con el ajuste del parámetro de ubicación en distribuciones estadísticas, especialmente al programar en Python. Denominado 'flock', este parámetro debe iniciarse en cero para asegurar la precisión del modelo. Esta práctica se fundamenta en la teoría de la distribución gaussiana, donde se considera que el punto de inicio de cualquier evento se sitúa a la derecha del cero. Es decir, los eventos de fallo se representan como valores positivos, no negativos. Esta especificación es vital para garantizar que el área bajo la curva refleje adecuadamente la naturaleza de los eventos de fallo y para mantener la integridad del análisis estadístico.
- En el ámbito del análisis estadístico avanzado, una recomendación crucial es la selección rigurosa y uniforme de las librerías de *Python*, capitalizando su eficacia como lenguaje de programación de código abierto en términos

de celeridad computacional y flexibilidad operativa, en comparación con plataformas tradicionales como *Microsoft Excel*. La heterogeneidad en la implementación de algoritmos estadísticos entre las distintas librerías de *Python* impone una consideración crítica: no todas estas herramientas computacionales ofrecen idénticos patrones en el ajuste de curvas para el análisis probabilístico. Diferencias en la estimación de parámetros clave, tales como 'mu' y 'lambda', en distribuciones Gaussianas, Exponenciales, Weibull o Logarítmicas, son atribuibles a la variabilidad inherente a cada librería. Por lo tanto, es imperativo adherirse a una única librería a lo largo de un análisis para garantizar la coherencia metodológica. La mezcla de librerías puede inducir variaciones significativas en los modelos matemáticos subyacentes, comprometiendo la precisión y fiabilidad estadística del estudio. Se aconseja optar por una librería que demuestre suficiente robustez y abarque integralmente las exigencias analíticas del proyecto.

BIBLIOGRAFÍA

ABBATE, R., CATERINO, M., FERA, M., & CAPUTO, F. (2022). Maintenance Digital Twin using vibration data. *Procedia Computer Science*, 200, 546-555. <https://doi.org/10.1016/j.procs.2022.01.252>

AHMADI, S., MOOSAZADEH, S., HAJIHASSANI, M., MOOMIVAND, H., & RAJAEI, M. M. (2019). Reliability, availability and maintainability analysis of the conveyor system in mechanized tunneling. *Measurement: Journal of the International Measurement Confederation*, 145, 756-764. <https://doi.org/10.1016/j.measurement.2019.06.009>

ARIAS GONZALES, J. Luis. (2020). Proyecto de tesis: guía para la elaboración.

Bag, S., Gupta, S., & Kumar, S. (2021). Industry 4.0 adoption and 10R advance manufacturing capabilities for sustainable development. *International Journal of Production Economics*, 231, 107844.

<https://doi.org/10.1016/j.ijpe.2020.107844>

BAKON, K., HOLCZINGER, T., SULE, Z., JASKO, S., & ABONYI, J. (2022). Scheduling Under Uncertainty for Industry 4.0 and 5.0. *IEEE Access*, 10, 74977-75017. <https://doi.org/10.1109/ACCESS.2022.3191426>

BENCHEKROUN, M. T., ZAKI, S., & ABOUSSALEH, M. (2023). Cement kiln safety and performance improvement based on machine learning predictive analytics. *International Journal of Advanced Manufacturing Technology*. <https://doi.org/10.1007/s00170-023-10813-7>

BHARDWAJ, U., TEIXEIRA, A. P., & SOARES, C. G. (2019). Reliability prediction of an offshore wind turbine gearbox. *Renewable Energy*, 141, 693-706. <https://doi.org/10.1016/j.renene.2019.03.136>

CHOUDHARY, D., TRIPATHI, M., & SHANKAR, R. (2019). Reliability, availability and maintainability analysis of a cement plant: a case study. *International Journal of Quality and Reliability Management*, 36(3), 298-313.

<https://doi.org/10.1108/IJQRM-10-2017-0215>

CULOT, G., NASSIMBENI, G., ORZES, G., & SARTOR, M. (2020). Behind the definition of Industry 4.0: Analysis and open questions. *International Journal of Production Economics*, 226.

<https://doi.org/10.1016/j.ijpe.2020.107617>

D'AMICO, R. D., ERKOYUNCU, J. A., ADDEPALLI, S., & PENVER, S. (2022). Cognitive digital twin: An approach to improve the maintenance management. In *CIRP Journal of Manufacturing Science and Technology* (Vol. 38, pp. 613-630). Elsevier Ltd. <https://doi.org/10.1016/j.cirpj.2022.06.004>

DHILLON, B. S. (2002). *ENGINEERING MAINTENANCE: A Modern Approach*.

DIEZ-OLIVAN, A., DEL SER, J., GALAR, D., & SIERRA, B. (2019). Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Information Fusion*, 50, 92-111.

<https://doi.org/10.1016/J.INFFUS.2018.10.005>

EL-NAGGAR, M. F., ABDELHAMID, A. S., ELSHAHED, M. A., & BEKHET, M. E. S. M. (2022). Dynamic Reliability and Availability Allocation of Wind Turbine Subassemblies Through Importance Measures. *IEEE Access*, 10, 99445-99459.

<https://doi.org/10.1109/ACCESS.2022.3203423>

ELSAYED, E. A. (2020). *Reliability Engineering, Third Edition*. Reliability Engineering, Third Edition, 1-512.

<https://doi.org/10.1002/9781119665946>

ELSEVIER. *Practical E-manufacturing and supply chain management*. (2004).

<https://doi.org/10.1016/B978-0-7506-6272-7.X5000-3>

ISBN 978-0-7506-6272-7

FAN, J., XUE, L., LIU, Y., & LI, W. (2021). Reliability analysis of spindle system of CNC grinder based on fault data. *The International Journal of Advanced Manufacturing Technology*.

<https://doi.org/10.1007/s00170-021-07552-y/Published>

FANG, J., KANG, R., & CHEN, Y. (2021). Reliability evaluation of non-repairable

systems with failure mechanism trigger effect. *Reliability Engineering and System Safety*, 210.

<https://doi.org/10.1016/j.res.2021.107454>

GABRIEL-ORTEGA, J. (2017). Cómo se genera una investigación científica que luego sea motivo de publicación. *Journal of the Selva Andina Research Society*, 8(2), 155-156.

GALLO, T., & SANTOLAMAZZA, A. (2021). Industry 4.0 and human factor: How is technology changing the role of the maintenance operator? *Procedia Computer Science*, 180, 388-393.

<https://doi.org/10.1016/j.procs.2021.01.364>

GULATI, Ramesh., & SMITH, R. (2013). *Maintenance and reliability best practices*. Industrial Press Inc.

GUO, D., LI, M., LYU, Z., KANG, K., WU, W., ZHONG, R. Y., & HUANG, G. Q. (2021). Synchronoperation in industry 4.0 manufacturing. *International Journal of Production Economics*, 238.

<https://doi.org/10.1016/j.ijpe.2021.108171>

HARTOYO, F., IRIANTI, G. P., FATRIANSYAH, J. F., OVELIA, H., MAS'UD, I. A., DIGITA, F. R., FAUZI, A., & ANIS, M. (2023). Weibull distribution optimization for piping risk calculation due to uniform corrosion using Monte Carlo method. *Materials Today: Proceedings*.

<https://doi.org/10.1016/j.matpr.2023.02.312>

HOFFMANN SOUZA, M. L., da COSTA, C. A., de OLIVEIRA RAMOS, G., & da ROSA RIGHI, R. (2020). A survey on decision-making based on system reliability in the context of Industry 4.0. In *Journal of Manufacturing Systems* (Vol. 56, pp. 133-156). Elsevier B.V.

<https://doi.org/10.1016/j.jmsy.2020.05.016>

HUAIRE INACIO, E. J., MARQUINA LUJÁN, R. J., HORNA CALDERÓN, V. E., LLANOS MIRANDA, K. N., HERRERA ÁLVAREZ, Á. M., RODRÍGUEZ SOSA, J., & VILLAMAR ROMERO, R. M. (2022). *Tesis fácil. El arte de dominar el método científico*. Analéctica.

JASIULEWICZ-KACZMAREK, M., & GOLA, A. (2019). Maintenance 4.0 Technologies for Sustainable Manufacturing - An Overview. IFAC-PapersOnLine, 52(10), 91-96. <https://doi.org/10.1016/j.ifacol.2019.10.005>

KIRBAS, I. (2019). Determination of Appropriate Distribution Functions for the Wind Speed Data Using the R Language. In Turkey) EJENS (Vol. 3, Issue 2).

KOVACS, K., ANSARI, F., & SIHN, W. (2020). A modified Weibull model for service life prediction and spare parts forecast in heat treatment industry. Procedia Manufacturing, 54, 172-177.

<https://doi.org/10.1016/j.promfg.2021.07.026>

KRISTJANPOLLER, F., CÁRDENAS-PANTOJA, N., VIVEROS, P., & MENA, R. (2021). Criticality analysis based on reliability and failure propagation effect for a complex wastewater treatment plant. Applied Sciences (Switzerland), 11(22). <https://doi.org/10.3390/app112210836>

KUMAR, M., & SAMUEL, C. (2019). Stochastic demand side management in smart grid system. International Journal of Networking and Virtual Organisations, 20(4), 319-339. <https://doi.org/10.1504/IJNVO.2019.100596>

KUMAR, S., & SINGH, R. (2020). Rank order clustering and imperialist competitive optimization based cost and RAM analysis on different industrial sectors. Journal of Manufacturing Systems, 56, 514-524.

<https://doi.org/10.1016/j.jmsy.2020.07.014>

MARTÓN, I., SÁNCHEZ, A. I., CARLOS, S., MULLOR, R., & MARTORELL, S. (2023). Prognosis of wear-out effect on of safety equipment reliability for nuclear power plants long-term safe operation. Reliability Engineering & System Safety, 233, 109121. <https://doi.org/10.1016/j.ress.2023.109121>

MORA, L. Alberto. (2009). Mantenimiento-planeación, ejecución y control. Alfaomega Grupo Editor.

PATIL, S. S., & BEWOOR, A. K. (2021). Reliability analysis of a steam boiler system by expert judgment method and best-fit failure model method: a new approach. International Journal of Quality and Reliability Management, 38(1), 389-409. <https://doi.org/10.1108/IJQRM-01-2020-0023>

PINCIROLI, L., BARALDI, P., & ZIO, E. (2023). Maintenance optimization in industry 4.0. *Reliability Engineering & System Safety*, 234, 109204.
<https://doi.org/10.1016/J.RESS.2023.109204>

RAJ, A., DWIVEDI, G., SHARMA, A., LOPES de Sousa Jabbour, A. B., & RAJAK, S. (2020). Barriers to the adoption of industry 4.0 technologies in the manufacturing sector: An inter-country comparative perspective. *International Journal of Production Economics*, 224.
<https://doi.org/10.1016/j.ijpe.2019.107546>

RAMESH, P. G., DUTTA, S. J., NEOG, S. S., BAISHYA, P., & BEZBARUAH, I. (2020). Implementation of Predictive Maintenance Systems in Remotely Located Process Plants under Industry 4.0 Scenario. *Springer Series in Reliability Engineering*, 293-326.
https://doi.org/10.1007/978-3-030-36518-9_12

RIVERA VITONERA, L. A. (2019). Diseño de válvula pendular para aumentar la confiabilidad del filtro de mangas de una cementera. Universidad de Piura.
<https://pirhua.udep.edu.pe/handle/11042/4030>

SHOJAEINASAB, A., CHARTER, T., JALAYER, M., KHADIVI, M., OGUNFOWORA, O., RAIYANI, N., YAGHOUBI, M., & NAJJARAN, H. (2022). Intelligent manufacturing execution systems: A systematic review. In *Journal of Manufacturing Systems* (Vol. 62, pp. 503-522). Elsevier B.V.
<https://doi.org/10.1016/j.jmsy.2022.01.004>

SILVESTRI, L., FORCINA, A., INTRONA, V., SANTOLAMAZZA, A., & CESAROTTI, V. (2020). Maintenance transformation through Industry 4.0 technologies: A systematic literature review. *Computers in Industry*, 123, 103335.
<https://doi.org/10.1016/J.COMPIND.2020.103335>

SMRP. (2020). *SMRP Best practice* (6th. Edition).

TALKIT, V. G., HONRAO, P., BHOSALE, S., BEWOOR, A., & PATIL, R. B. (2022). Application of the reliability analysis for modifications in maintenance scheduling. *Materials Today: Proceedings*.

<https://doi.org/10.1016/j.matpr.2022.10.075>

TALKIT, V., KININGE, R., KOKATE, P., NARKHEDE, D., MADAME, S., BEWOOR, A., & PATIL, R. B. (2022). Reliability analysis of repairable and replaceable system: Dairy product industry. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2022.10.074>

THEISLER, A., PEREZ-VELAZQUEZ, J., KETTELGERDES, M., & ELGER, G. (2021). Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. *Reliability Engineering and System Safety*, 215. <https://doi.org/10.1016/j.ress.2021.107864>

VANDERSCHUEREN, T., BOUTE, R., VERDONCK, T., BAESENS, B. and VERBEKE, W., 2023. Optimizing the preventive maintenance frequency with causal machine learning [en línea]. April 2023. S.I.: Elsevier BV. Disponible en: <http://dx.doi.org/10.1016/j.ijpe.2023.108798>.

ZERMANE, H., & DRARDJA, A. (2022). Development of an efficient cement production monitoring system based on the improved random forest algorithm. *International Journal of Advanced Manufacturing Technology*, 120(3-4), 1853-1866. <https://doi.org/10.1007/s00170-022-08884-z>

ANEXOS

Anexo 1 - Matriz de operacionalización de variables

Tabla 9

Matriz de operacionalización de variables

| Variable | Definición conceptual | Definición operacional | Dimensiones | Indicadores | Escala de medición |
|---|---|---|------------------------------|----------------------------------|--------------------|
| Modelo de machine learning con transformación logarítmica y validación cruzada. | Modarres, et al. (1999) define el método de transformación logarítmica, combinado con la validación cruzada, como una estrategia estadística empleada para estimar los parámetros de modelos de fiabilidad y garantizar su robustez. Esta técnica se aplica frecuentemente en la evaluación de datos de TBF en sistemas de ingeniería y en análisis de supervivencia. El método de transformación logarítmica busca linealizar conjuntos de datos con distribuciones no lineales, como la Weibull, habitual en el estudio de la confiabilidad. Mientras tanto, la validación cruzada se utiliza para evaluar la generalización de modelos estadísticos, especialmente útil en contextos con limitaciones en la cantidad de datos. | Montgomery, et al. (2008) indica que el método de transformación logarítmica reúne los datos de interés que siguen una distribución no lineal, como por ejemplo datos de TBF y aplica la siguiente fórmula para cada punto de datos x en el conjunto: Transformación logarítmica: $x = \log(x)$ | Número de fallas | Número de eventos de falla. | Razón |
| | | | Periodo de tiempo | Tiempo de funcionamiento (horas) | |
| Confiabilidad | Según el estudio de Kumar y Singh (2020), la confiabilidad es la medida de la capacidad de un sistema para funcionar sin interrupciones durante un período de tiempo determinado. Según lo expuesto por (Ahmadi et al., 2019) refiriendo a Modarres et al. (1999) y Gulati (2013), la confiabilidad es la capacidad de un sistema para cumplir una tarea en condiciones y plazos establecidos, volviéndose más confiable a medida que pueda sostener su condición en el tiempo. | Según (Dhillon, 2002), la confiabilidad es la probabilidad de que el sistema funcione sin fallas hasta el tiempo (t), con una tasa de fallas (λ) y un límite máximo de fallas tolerables (e). La fórmula para calcular la confiabilidad se define como: $R(t) = e^{(-\lambda t)}$ Donde: <ul style="list-style-type: none"> • R(t) representa la confiabilidad en un momento de tiempo t. • λ es la tasa de fallas, que representa el número promedio de fallas por unidad de tiempo. • t es el tiempo transcurrido. | Tasa de fallas. | Número de fallas | Razón |
| | | | | Tiempo de funcionamiento (horas) | |
| | | | Periodo de tiempo | Tiempo de funcionamiento (horas) | |
| | | Diferencia de errores | Error Cuadrático Medio (MSE) | | |

Anexo 3 - Documento de solicitud de uso de información de empresa

Trujillo, 16 de junio de 2023

Señor (a):
REYES MENDEZ, RONALD TEODORO
JEFATURA DE GESTIÓN DE MANTENIMIENTO
CEMENTOS PACASMAYO SAA
Presente. -

Es grato dirigirme a usted para saludarlo, y a la vez manifestarle que dentro de mi formación académica en la experiencia curricular de investigación del IX ciclo, se contempla la realización de una investigación con fines netamente académicos /de obtención de mi título profesional al finalizar mi carrera.

En tal sentido, considerando la relevancia de su organización, solicito su colaboración, para que pueda realizar mi investigación en su representada y obtener la información necesaria para poder desarrollar la investigación titulada: "Análisis de confiabilidad empleando la herramienta RStudio para mejorar la disponibilidad en una planta de molienda de cemento en Pacasmayo". En dicha investigación me comprometo a mantener en reserva el nombre o cualquier distintivo de la empresa, salvo que se crea a bien su socialización.

Se adjunta la carta de autorización de uso de información y publicación, en caso de que se considere la aceptación de esta solicitud para ser llenada por el representante de la empresa.

Agradeciéndole anticipadamente por vuestro apoyo en favor de mi formación profesional, hago propicia la oportunidad para expresar las muestras de mi especial consideración.

Atentamente,



Angulo Gutierrez, Vicente Manuel
DNI 71104402

Anexo 4 – Documento de autorización de uso de información de empresa

AUTORIZACIÓN DE USO DE INFORMACIÓN DE EMPRESA

Yo Ronald Teodoro Reyes Mendez, identificado con DNI 07525497, en calidad de jefe del área de gestión de mantenimiento de la empresa Cementos Pacasmayo SAA con R.U.C N°20419387658, ubicada en la ciudad de Pacasmayo.

OTORGO LA AUTORIZACIÓN,

A los señores Vicente Manuel Angulo Gutiérrez identificado con DNI N° 71104402 y José Andrés Jara Flores identificado con DNI N° 76560080, de la carrera profesional ingeniería mecánica eléctrica, para que utilice la siguiente información de la empresa: reporte de paradas de molino de cemento, con la finalidad de que pueda desarrollar su Tesis para optar el Título Profesional.

- Publique los resultados de la investigación en el repositorio institucional de la UCV.
- Mantener en reserva el nombre o cualquier distintivo de la empresa; o
- Mencionar el nombre de la empresa.



Ronald Teodoro Reyes Mendez
CIP 108928
DNI: 07525497

El Estudiante declara que los datos emitidos en esta carta y en el Trabajo de Investigación, en la Tesis son auténticos. En caso de comprobarse la falsedad de datos, el Estudiante será sometido al inicio del procedimiento disciplinario correspondiente; asimismo, asumirá toda la responsabilidad ante posibles acciones legales que la empresa, otorgante de información, pueda ejecutar.



Firma del Estudiante
DNI: 71104402



Firma del Estudiante
DNI: 76560080