



UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DEPARTAMENTO INGENIERÍA MECÁNICA



**DISEÑO E IMPLEMENTACIÓN DE UN ALGORITMO BASADO EN
APRENDIZAJE AUTOMÁTICO Y SEÑALES VIBRATORIAS PARA LA
DETECCIÓN Y PRONÓSTICO DE FALLAS EN MOTORES DE
RETROEXCAVADORAS**

POR

Rodrigo Antonio Veloso Huircapán

Memoria de Título presentada a la Facultad de Ingeniería de la Universidad de Concepción para
optar al título profesional de Ingeniero Civil Mecánico

Profesores Guía:

PhD. Bernardo Andrés Hernández Vicente
M. Sc. Luciano Ignacio Radrigan Figueroa

Marzo 2023

Concepción (Chile)

© 2023 Rodrigo Antonio Veloso Huircapán

© 2023 Rodrigo Antonio Veloso Huircapán

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento

Agradecimientos

Agradecer a mis padres por el amor y dedicación con la que me criaron, por siempre incentivar me a mejorar personalmente y confiar siempre en mí.

A mis hermanos, Esteban y Paula, por enseñarme lo lindo que es la vida.

A mis tatas por enseñarme el valor del esfuerzo y la perseverancia, por siempre recibirme con los brazos abiertos y mucho cariño.

A mi abuela Ana, por brindarme el cobijo en el inicio de esta aventura universitaria.

A Natalia, quién me ha sido mi compañera en los últimos años y ha sido fundamental, gracias por el amor, motivación y felicidad que me brindas.

A los profesores Bernardo y Luciano por guiarme en la última etapa de este periplo universitario, por el tiempo, paciencia y motivación que han tenido en esta memoria.

A mis compañeros y amigos Chris, Sebastián y Vicente por su fidelidad y amistad, por siempre tener el espíritu de superación, ganas de aprender más y alentarnos a ser mejores estudiantes y futuros profesionales. Por siempre estar en los buenos y malos momentos que no ha deparado la vida.

Esta memoria está dedicada a Alberto

Resumen

En la Industria 4.0 el monitoreo basado en condiciones es clave para aumentar la disponibilidad y la confiabilidad de los equipos, lo que también se ve reflejado en el factor económico. Así, aplicando Inteligencia Artificial y Big Data en el mantenimiento predictivo se busca mejorar la gestión del mantenimiento.

La memoria de título se sustenta en una base de datos obtenida en una investigación anterior, en donde se midieron las vibraciones del motor de dos retroexcavadoras, las cuales fueron monitoreadas en un periodo aproximado de 60 días. Una de estas retroexcavadoras presenta una falla en el motor al final del tiempo de monitoreo. La falla se atribuye a la falta de lubricación del motor, lo que ocasionó que un pistón fuera eyectado del motor, con lo cual este queda fuera de servicio, debiendo ser cambiado. Mediante el análisis de las vibraciones, se distinguen 4 estados de salud del motor de la retroexcavadora, con los cuáles mediante algoritmos de clasificación se busca predecir la vida útil remanente (RUL) del motor, utilizando los datos obtenidos a lo largo del monitoreo.

En la presente investigación se genera un modelo con algoritmos de Inteligencia Artificial para realizar la estimación del RUL, pero a diferencia de la investigación realizada anteriormente, en este caso la estimación se realiza en tiempo real. Para ello se crea un modelo, que tiene como propósito predecir el comportamiento vibracional a futuro utilizando los datos medidos, y estos datos generados por el regresor, clasificarlos en sus respectivos estados de salud, con tal de calcular el RUL.

Es necesario destacar, que los datos con los que se trabajan fueron preprocesados, ya que fueron sometidos a una Transformada de Fourier, luego se calculó el promedio por cada minuto de medición, para finalmente, entregar el valor más alto de esos promedios cada 15 minutos. Debido a esto, la investigación se verá limitada por los datos sensorizados, lo que permitirá analizar si estos datos son válidos para modelos de inteligencia artificial, con tal de establecer un precedente para futuros estudios.

Los resultados obtenidos demuestran que con los datos recopilados en una primera etapa de sensorización, realizada por terceros, no es posible realizar el pronóstico del tiempo de vida útil restante, pues, el algoritmo de clasificación del modelo alcanza un 26,5% de precisión, con lo cual es imposible pronosticar un RUL confiable para ser implementado en el monitoreo del equipo. Lo anterior es principalmente consecuencia de que no hay un parámetro que revele las perturbaciones externas que sufre la máquina en su funcionamiento, lo que entorpece el entrenamiento de los modelos utilizados.

Palabras clave: Monitoreo basado en condiciones, Pronóstico, Machine Learning.

Abstract

In Industry 4.0, condition-based monitoring is key to increase the availability and reliability of equipment, which is also reflected in the economic factor. Thus, applying Artificial Intelligence and Big Data in predictive maintenance seeks to improve maintenance management.

The title memory is based on a database obtained in previous research, where the engine vibrations of two backhoe loaders were measured, which were monitored over a period of approximately 60 days. One of these backhoes presented an engine failure at the end of the monitoring time. The failure was attributed to a lack of engine lubrication, which caused a piston to be ejected from the engine, putting it out of service and requiring replacement. By means of vibration analysis, 4 states of health of the backhoe loader engine are distinguished, with which classification algorithms are used to predict the remaining useful life (RUL) of the engine, using the data obtained throughout the monitoring.

In the present research, a model is generated with Artificial Intelligence algorithms to estimate the RUL, but unlike the previous research, in this case the estimation is done in real time. For this purpose, a model is created, whose purpose is to predict the vibrational behavior in the future using the measured data, and these data generated by the regressor, classify them in their respective health states, in order to calculate the RUL.

It is necessary to emphasize that the data used were pre-processed, since they were subjected to a Fourier Transform, then the average was calculated for each minute of measurement, to finally deliver the highest value of these averages every 15 minutes. Due to this, the research will be limited by the sensorized data, which will allow to analyze if these data are valid for artificial intelligence models, to establish a precedent for future studies.

The results obtained show that with the data collected in a first stage of sensorization, performed by third parties, it is not possible to forecast the remaining useful lifetime, since the classification algorithm of the model reaches 26.5% accuracy, making it impossible to forecast a reliable RUL to be implemented in the monitoring of the equipment. This is mainly due to the fact that there is no parameter that reveals the external disturbances suffered by the machine in its operation, which hinders the training of the models used.

Keywords: Condition-based Management, Prognostics, Machine Learning

Tabla de Contenidos

Tabla de Contenidos.....	5
Lista de Tablas	6
Lista de Figuras	7
Glosario	8
1 CAPÍTULO 1: Introducción.....	9
1.1 Contexto.....	9
1.2 Hipótesis	10
1.3 Objetivos.....	11
1.4 Metodología.....	11
1.5 Alcances y Limitaciones.....	12
2 CAPÍTULO 2: Marco Teórico	13
2.1 Mantenimiento.....	13
2.2 Mantenimiento Predictivo	13
2.3 Vibraciones	14
2.4 Monitoreo Basado en Condiciones.....	16
2.5 Machine Learning.....	19
2.6 Conceptos de Machine Learning	22
2.7 Tipos de entrenamiento	25
3 CAPÍTULO 3: Estado del Arte	26
3.1 Pronósticos.....	26
3.2 Fallas mecánicas y monitoreo en motor de retroexcavadora.....	29
4 CAPÍTULO 4: Adquisición de Datos y Modelo de predicción de RUL propuesto.....	34
4.1 Muestra de datos	35
4.2 Modelo Propuesto.....	36
4.3 Entrenamiento.....	38
4.4 Clasificadores	38
4.5 Regresores	42
5 CAPÍTULO 5: Resultados	46
5.1 Clasificadores	46
5.2 Regresores	47
5.3 Modelo.....	51
6 CAPÍTULO 6: Conclusiones	53

Referencias	55
-------------------	----

Lista de Tablas

Tabla 4.1: Análisis Estadístico.....	35
Tabla 4.2: Amplitudes [g] de cada frecuencia en su respectivo estado.....	36
Tabla 4.3: Cantidad de datos por Etiqueta.	38
Tabla 5.1: Resultados del entrenamiento de los clasificadores.....	46
Tabla 5.2: Resultados del entrenamiento de los clasificadores utilizando RMS acumulado.....	46
Tabla 5.3: R2 obtenido utilizando RMS.....	47
Tabla 5.4: R2 obtenido utilizando Amplitud.....	47
Tabla 5.5: R2 de RMS obtenido utilizando RMS y Amplitud como entradas.....	48
Tabla 5.6: R2 de Amplitud obtenido utilizando RMS y Amplitud como entradas.....	49
Tabla 5.7: R2 de RMS obtenido algoritmo Bagguing.....	50
Tabla 5.8: R2 de Amplitud obtenido algoritmo Bagguing.....	50

Lista de Figuras

Figura 1.1: Diagrama metodológico de la consecución de los objetivos específicos.	12
Figura 2.1: Porcentaje de OEE alcanzado por los tipos de mantenimiento [10].....	14
Figura 2.2: Funcionamiento del CBM [15].....	17
Figura 2.3: Funcionamiento del KNN.....	21
Figura 2.4: Aplicación de método de conjunto Bagging.....	22
Figura 2.5: Aplicación de método de conjunto Boosting.....	22
Figura 3.1: Paso a paso del Diagnóstico y Pronóstico [24].....	28
Figura 3.2: Proceso del CBM [25]	29
Figura 3.3: Cuadro Esquemático de las principales fallas en los cigüeñales [28].	32
Figura 4.1: Comparación de las retroexcavadoras que falla y que no falla.	35
Figura 4.2: RMS de retroexcavadora que no falla.	36
Figura 4.3: RMS de retroexcavadora que falla.	36
Figura 4.4: Amplitud de retroexcavadora que no falla.....	36
Figura 4.5: Amplitud de retroexcavadora que falla.....	36
Figura 4.6: Cuadro esquemático de modelo propuesto	37
Figura 4.7: Matriz de Confusión [38].....	42
Figura 5.1: <i>Bagging Regressor</i>	48
Figura 5.2: Filtro de Media Móvil con $N = 4$	49
Figura 5.3: Filtro de Media Móvil con $N = 6$	49
Figura 5.4: Filtro de Media Móvil con $N = 10$	50
Figura 5.5: Comparación de Filtros.....	50
Figura 5.6: <i>Bagging Regressor</i>	51
Figura 5.7: Modelo realizado.	52

Glosario

CbM	:	<i>Condition-Based Monitoring</i>
PdM	:	<i>Predictive Maintenance</i>
IoT	:	<i>Internet of Things</i>
ML	:	<i>Machine Learnings</i>
OEE	:	<i>Overall Equipment Effectiveness</i>
RMS	:	<i>Root Mean Square</i>
FFT	:	<i>Fourier Fast Transform</i>
DFT	:	<i>Discret Fourier Transform</i>
MTTR	:	<i>Mean Time To Repair</i>
RUL	:	<i>Remaining Useful Life</i>
PSD	:	<i>Power Spectral Density</i>
AI	:	<i>Artificial Intelligence</i>
SVM	:	<i>Support Vector Machine</i>
IMU	:	<i>Inertial Measurement Unit</i>

CAPÍTULO 1: Introducción

1.1 Contexto

La industria se encuentra en la búsqueda constante de mejoras para los procesos que desarrollan. Entre los principales objetivos destacan el establecer un ambiente seguro de trabajo, la reducción de las emisiones de gases contaminantes y el aumento de la productividad, los que se llevan a cabo con ideas innovadoras, logrando optimizar los procesos industriales. La confiabilidad de los equipos es un factor importante que se debe considerar si se quieren lograr los objetivos mencionados, pues se asegura que el equipo funcione correctamente, permitiendo que el proceso se lleve a cabo de la manera deseada. El monitoreo de condiciones operacionales de los equipos es fundamental para asegurar la confiabilidad y disponibilidad, dos conceptos importantes en el área del mantenimiento. Con este método se conoce de manera inmediata, o en un periodo de tiempo breve dependiendo de la aplicación del monitoreo, el estado de salud del equipo analizado, de esta manera se realiza el plan de mantenimiento conforme a la información adquirida [1]. Lo anterior permite reducir los costos de mantenimiento, disminuir los tiempos de detención, significando un aumento de la producción.

La retroexcavadora es un equipo móvil que tiene como principales funciones el desplazamiento de material y realizar excavaciones, teniendo un rol fundamental en las industrias de la construcción y minería, en donde se desempeña principalmente en la construcción de caminos los que permiten la extracción del material. Las operaciones de desplazamiento de material son unas de las más importantes en los rubros mencionados, requieren de una alta presencia de máquinas, significando altos costos para las empresas. Por lo que el manejo adecuado de las flotas de maquinarias puede significar un gran beneficio costo-efectivo para la industria [2].

Actualmente, la mayor concentración de fallas que afectan a la retroexcavadora, cerca del 50%, se presentan en sellos, rodamientos y en el árbol de levas, siendo la principal causa la falta de mantenimiento [3]. A esto se suma que por más leve que se considere una falla, esta puede derivar en grandes e incluso catastróficas daños para el equipo [4]. Es por esto por lo que se requiere de la elaboración de un correcto plan de mantención, en donde se realicen las detenciones a tiempo para asegurar el buen estado del equipo. Aun así, dependiendo del ambiente, tipo de trabajo y operación al que sea sometido la retroexcavadora esta puede fallar, siendo necesario el monitoreo en tiempo real para pronosticar el tiempo de vida útil remanente y poder actuar a tiempo.

A partir de lo anterior, durante la realización de esta Memoria de Título se implementará algoritmos de aprendizaje automático, más precisamente aplicando métodos de ensamble, con lo cual se pretende entrenar un modelo para la detección y pronóstico de fallas presentes en el motor diésel de una retroexcavadora. El objetivo de esto es calcular la vida útil remanente o RUL, por sus siglas en inglés, lo que otorga el margen de tiempo en el que se debe elaborar el plan a seguir ante una falla inesperada del equipo [5].

Anteriormente, se han desarrollado sistemas de monitoreo de retroexcavadoras capaces de anticipar la falla del equipo 50 horas antes en comparación a un sistema de detección manual [6], basado en la mantención preventiva. En esta investigación se propone ampliar este rango de tiempo de detección de fallas, con tal de poder disminuir lo más posible el tiempo de mantención del equipo y así aumentar la disponibilidad de este, además de innovar en la manera de realizar el pronóstico de fallas en equipos no estacionarios.

Para desarrollar este algoritmo se utilizarán mediciones realizadas en el motor de dos retroexcavadoras JCB 3CX ECO15FT, que tienen como particularidad que una se encuentra en normal funcionamiento, mientras que la otra presenta fallas en el motor relacionadas con la lubricación inadecuada de este. De esta manera, se permite diferenciar el comportamiento vibracional del equipo. La lubricación de componentes o equipos es uno de los grandes problemas en la mantención de maquinaria, ya que, si no se realiza de forma adecuada o en el tiempo establecido genera desgaste en las piezas del motor, lo que ocasiona graves problemas [7]. Estos datos serán utilizados por distintos algoritmos de *Machine Learning*, los que permitirán realizar la detección temprana de fallas.

Este informe está organizado de la siguiente manera. En el Capítulo 1 se establece el contexto en el que se encuentran las retroexcavadoras, se detalla la metodología, objetivos, alcances y limitaciones que tendrá este trabajo. En el Capítulo 2, se establecerá el marco teórico de la investigación, mientras que en el Capítulo 3 se realizará una búsqueda bibliográfica para conocer las investigaciones más destacadas en el área del CbM y del Machine Learning. En el Capítulo 4 se dan a conocer el proceso de sensorización de las máquinas y se muestran los datos a grandes rasgos. Además, se explica el modelo propuesto y como se lleva a cabo, explicando el trabajo realizado con los datos previo al entrenamiento de los modelos y a continuación, el Capítulo 5 expone los resultados más importantes de la investigación. Finalmente, en el Capítulo 6 se presentan las principales ideas a modo de conclusión de la memoria de título.

1.2 Hipótesis

Implementando un algoritmo de aprendizaje automático se aumentará el tiempo de la ventana de detección de fallas ocurridas en el motor de la retroexcavadora, lo que permite realizar un pronóstico oportuno, con el objetivo de disminuir el tiempo de inactividad del equipo.

1.3 Objetivos

1.3.1 Objetivo General.

Diseñar e implementar algoritmos utilizando aprendizaje automático y señales vibratorias no estacionarias para la detección y pronóstico de fallas mecánicas en motor de retroexcavadoras. Con lo cual se busca aumentar la ventana de 50 horas de detección en comparación a la clasificación manual alcanzada con anterioridad.

1.3.2 Objetivos Específicos.

1. Analizar y comprender los datos obtenidos en la sensorización de las retroexcavadoras, con tal de entender los alcances y limitaciones de la investigación.
2. Desarrollar e implementar algoritmos de aprendizaje automático que clasifiquen los datos generados en los distintos estados de salud de máquina.
3. Desarrollar e implementar algoritmos de aprendizaje automático capaces de predecir datos de la vibración del motor de la retroexcavadora.
4. Generar un modelo capaz de predecir el RUL del motor de la retroexcavadora, el cual se busca determinar en una ventana superior a la ya realizada.
5. Realizar distintas iteraciones con el modelo generado, probando la validez de este.

1.4 Metodología

Primero que todo, se confeccionará el marco teórico de la investigación. Posteriormente, se realizará una revisión del estado del arte desde los aspectos generales a lo más particular, así, se establecen las bases de la investigación y se comprenden las contribuciones al desarrollo de la industria en donde se encuentran inmersas las retroexcavadoras.

Luego, se analizarán los datos, que ya fueron recopilados en una investigación anterior, utilizando estadística descriptiva, con tal de conocer el comportamiento de los datos recopilados.

Posteriormente, se establecerá el porcentaje de datos que serán utilizados para entrenamiento y validación de los algoritmos, para luego, poder someter estos datos a los algoritmos de aprendizaje automático seleccionados.

Finalmente, se realizará un análisis de resultados el que permitirá validar los algoritmos diseñados a través de métricas de validación, con lo cual se obtendrá la precisión de los algoritmos desarrollados.

Se presenta un cuadro esquemático, Figura 1.1 en donde se resume la metodología aplicada en el desarrollo de la investigación y como esta satisface los objetivos específicos propuestos con anterioridad.

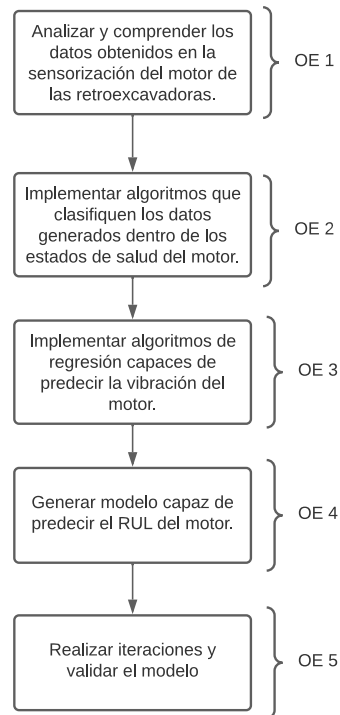


Figura 1.1: Diagrama metodológico de la consecución de los objetivos específicos.

1.5 Alcances y Limitaciones

Los datos recopilados pueden presentar errores de medición, que puede afectar en el entrenamiento de datos. Además, se restringe sólo para el modelo de retroexcavadora utilizado, por lo que la predicción no se puede generalizar a más modelos, sin embargo, al entrenar el algoritmo con datos provenientes de otros modelos, la predicción será válida.

Los algoritmos se entrenarán con datos ya recopilados y preprocesados, por lo que puede que no se adapten de buena forma al modelo que se busca realizar.

Desde el punto de vista computacional, no se presentan limitaciones, ya que los programas utilizados no requieren de grandes rendimientos de memoria para este caso.

CAPÍTULO 2: Marco Teórico

2.1 Mantenimiento

Se define como mantenimiento las acciones que ayudan a preservar y recrear el estado de una maquinaria o equipo.

Como se ha mencionado anteriormente, la mantención de equipos, y en especial de los equipos críticos, es de vital importancia para el desarrollo productivo de una empresa, es por eso por lo que con el paso de los años cada vez se le brinda mayor atención y dedicación al área de mantención. Se estima que en el mundo se gastan anualmente más de 7000 billones de euros en conceptos de mantención y se generan alrededor de 150 millones de puestos de trabajo indirectamente [8].

El origen del mantenimiento se remonta a la primera revolución industrial, en ese entonces consistía en reparar las máquinas una vez que estas detenían su funcionamiento, lo que se conoce como Mantenimiento Correctivo. Pero el gran desarrollo de la mantención fue en la Segunda Guerra Mundial, con lo cual se ha ido desarrollando con el paso de los años hasta la actualidad en pleno desarrollo de la Industria 4.0 y el internet de las cosas o IoT, por sus siglas en inglés. A continuación, se explican brevemente los tipos de mantenimiento que han surgido con los años [9].

Mantenimiento Correctivo: es el más simple ya que ocurre cuando el equipo ya ha fallado, por lo que no requiere de monitoreo previo. Tiene las desventajas de que conlleva un gran tiempo de reparación y puede generar otras fallas más graves.

Mantenimiento Preventivo: consta de realizar la calendarización de mantenciones para anticipar la ocurrencia de fallas. La principal desventaja es que aumenta los costos de mantención y disminuye la producción, al ser muchas veces innecesarias las detenciones.

Mantenimiento Predictivo o PdM, por sus siglas en inglés: requiere de una sensorización de los equipos, los datos recopilados son analizados mediante *Machine Learning*, lo que permite obtener un pronóstico del estado de salud del equipo. De esta manera se agenda la mantención.

Mantenimiento Basado en Condiciones o CBM, por sus siglas en inglés: Consta del continuo monitoreo del estado de salud del equipo, por lo que las tareas de mantención se realizan únicamente cuando sea estrictamente necesario.

2.2 Mantenimiento Predictivo

El mantenimiento predictivo también se conoce como mantenimiento basado en estadísticas, ya que las detenciones de mantenimiento son agendadas solo cuando es altamente necesario realizarlas. Este tipo de monitoreo se basa en la adquisición de mediciones que reflejan el estado de salud del equipo, de esta manera se establece el comportamiento histórico permitiendo establecer los puntos

de inflexión en el tiempo del estado de salud del equipo y cuando es apropiado reparar las fallas generadas. Esto es posible utilizando herramientas de ML y de estadística [10].

A diferencia de las otras técnicas de monitoreo, el Mantenimiento Predictivo tiene la ventaja de poder establecer la mantención en el momento preciso que permita optimizar la producción y el costo de reparación. Ya que por ejemplo en el caso del mantenimiento correctivo, la reparación no se realiza hasta una vez que falla el equipo, por lo que el rendimiento de ese activo decae en el tiempo, afectando de manera considerable la producción. A su vez, el mantenimiento preventivo si bien no es necesario esperar la ocurrencia de la falla, las reparaciones se realizan cuando aún queda vida útil a las piezas, de esta manera se elevan los costos de reparación y mantención. Para establecer un punto de comparación entre rendimiento de los equipos y tiempos de reparación, se utiliza un indicador conocido como *Overall Equipment Effectiveness* (OEE), el que indica la optimización del proceso productivo. En la Figura 2.1, se observa los porcentajes alcanzados por los tipos de mantenimientos utilizados en la industria. Se observa que el mantenimiento predictivo tiene una efectividad superior al 90%, lo que refleja el por qué es uno de los tipos de mantenimiento más utilizados en la actualidad.

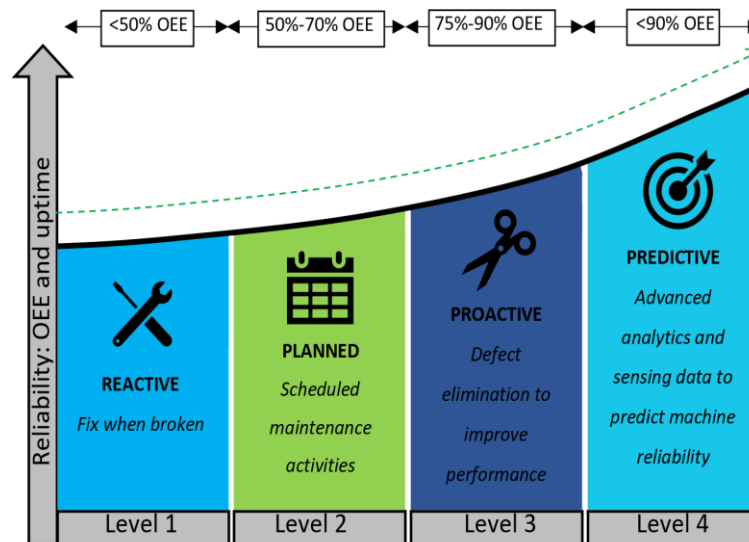


Figura 2.1: Porcentaje de OEE alcanzado por los tipos de mantenimiento [10].

2.3 Vibraciones

La vibración se define como el movimiento de un equipo o máquina alrededor de una posición de reposo [11] o como un movimiento oscilatorio de pequeña amplitud [12]. El análisis de estas vibraciones es de vital importancia en el área del mantenimiento de equipos industriales, ya que permite la detección de fallas de manera preventiva, es decir antes de que el equipo presente mayores complicaciones en su funcionamiento. Las vibraciones se presentan como periódicas o no periódicas:

- **Vibraciones Periódicas:** se entiende como vibración periódica cuando la forma de onda de repite con el tiempo. Las vibraciones periódicas pueden ser sinusoidales o de cualquier forma.
- **Vibraciones No Periódicas:** estas vibraciones no son constantes en el tiempo, por lo que presentan discontinuidades y saltos repentinos en el tiempo. Dentro de esta clasificación se encuentran las vibraciones transientes y vibraciones aleatorias. Las vibraciones transientes duran un periodo corto de tiempo y desaparecen, mientras que las vibraciones aleatorias son de carácter no determinista, por lo que no se puede predecir su comportamiento a futuro.

Dentro de los parámetros que definen una vibración periódica se tienen los siguientes:

- **Valor pico:** es el valor máximo alcanzado por la vibración durante el periodo de tiempo medido, se utiliza especialmente para analizar vibraciones transientes.
- **Valor pico a pico:** es la máxima diferencia entre los valores extremos de la vibración, valor máximo y mínimo, en el tiempo de monitoreo.
- **Valor RMS (*Root Mean Square*):** se define como la raíz del promedio de los valores instantáneos de medición durante un periodo de tiempo, como se observa en la Ec. 1, en donde N es la cantidad de valores utilizados para el cálculo. De esta manera, se consideran todas las vibraciones, por lo que se utiliza para analizar la severidad de la vibración.

$$RMS = \sqrt{\frac{\sum_{i=1}^n (V_i)^2}{N}} \quad \text{Ec (1)}$$

Las vibraciones registradas presentan ruido de por medio, por lo que se dificulta la interpretación de las mediciones, para facilitar esta tarea se obtienen el espectro vibracional. El espectro es calculado con la Transformada Rápida de Fourier (FFT).

2.3.1 Transformada de Fourier

Las vibraciones pueden ser medidas en el dominio del tiempo o en el dominio de la frecuencia. Las vibraciones en el dominio del tiempo reflejan el comportamiento de la máquina o componente monitoreado, pero la forma en que se encuentra la información del comportamiento es compleja, principalmente por el ruido de la medición, lo que supone un problema para realizar las tareas de diagnóstico. Por lo anterior, se recurre al análisis en el dominio de la frecuencia, el cuál presenta un espectro que distingue las frecuencias y amplitud de cada perturbación a la que es sometida la máquina.

La FFT permite pasar del dominio tiempo al dominio de frecuencia una señal vibratoria, lo cual se realiza según la Ec 2:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad \text{Ec (2)}$$

La complejidad es esta ecuación, es que para aplicarla se debe conocer la ecuación de la señal, lo que aplicado en la realidad es imposible ya que la señal presenta ruido y señales provenientes de otros equipos. Por ello, es necesario discretizar las mediciones, aplicando la Transformada Discreta de Fourier (DFT). La DFT transforma N puntos de la señal vibratoria en $N/2$ puntos discretos del espectro, esto se realiza mediante la ecuación 3:

$$X(k) = \frac{2}{N} \sum_{n=1}^N x(n) e^{-\frac{j2\pi kn}{N}} \quad k=1, 2, \dots, \frac{N}{2} \quad \text{Ec (3)}$$

En donde:

- $x(n)$: son los valores discretos de la señal vibratoria medida, $n=1, 2, \dots, N$.
- N : número total de puntos a considerar en la discretización de la señal.
- Δt : intervalo de discretización de la señal.
- $T = N\Delta t$: tiempo de muestreo de la señal vibratoria para realizar el espectro
- $f_s = \frac{1}{\Delta t}$: frecuencia demuestreo

Mientras que, si se analiza el resultado de la ecuación 3 se pueden obtener la siguiente información que modela el gráfico Frecuencia vs Amplitud:

- $\Delta f = \frac{1}{N}$: resolución de frecuencias
- $f_{max} = \frac{N}{2} \Delta t$: frecuencia máxima del espectro
- $X(k)$: valor del componente del espectro que corresponde a la frecuencia $f_k = k\Delta f$
- $|X(k)|$: amplitud de la componente que corresponde a f_k

2.4 Monitoreo Basado en Condiciones

El Monitoreo de Condiciones (CBM), es fundamental en la Industria 4.0, en el cual la aplicación del IoT permite optimizar procesos, ahorrar recursos y mejorar la toma de decisiones, lo que al aplicarlo sobre equipos críticos de la industria logra un mejor desempeño de estos[13]. Si bien el mantenimiento preventivo busca asegurar la disponibilidad de los equipos, siempre existe el riesgo de las fallas imprevistas, por distintas razones, ya sea por problemas de fabricación del equipo o por la mala utilización del activo. Las fallas imprevistas pueden llegar a ser catastróficas, lo que significa que afecta los componentes más importantes del equipo, aumentando considerablemente *Mean Time to Repair* (MTTR), desde el punto de vista económico, además de los costos de reparación, el principal costo es el de tener el equipo detenido, lo que dependiendo de la criticidad de este puede significar grandes pérdidas.

El ambiente de trabajo en donde se desempeña la máquina puede haber muchos factores que alteren su normal funcionamiento, es por eso que los planes de mantenimiento realizados por el fabricante o por el equipo de mantenimiento, es sujeto a cambios. Como se menciona anteriormente, una falla del equipo puede generar grandes pérdidas, en comparación al costo de implementación del CBM,

sobre todo si es dentro del rubro minero, en donde las faenas se encuentran alejadas de los centros urbanos y el tiempo requerido para reparar el equipo, es mucho mayor.

Para realizar un CBM adecuado lo primero es identificar las fallas críticas que puede tener o ha presentado el equipo, así es necesario realizar un análisis causa raíz, para establecer los mecanismos de falla, el origen y los efectos que estos tienen. También, es necesario establecer las cargas o procesos que pueden afectar el estado de salud del equipo y sus componentes. De esta manera, se realiza la sensorización de los equipos, monitoreando el o los parámetros más importantes del equipo, los cuales son registrados y analizados, permitiendo realizar ajustes al plan de mantenimiento, obteniendo una mayor disponibilidad del equipo. La sensorización de los equipos es uno de los aspectos más complejos del monitoreo, partiendo por la base de que es necesario que el sensor utilizado sea capaz de entregar información confiable, para esto se deben considerar los alcances y limitaciones del equipo, para seleccionar el adecuado, además en este caso los equipos monitoreados son equipos móviles, por lo que la conexión debe ser inalámbrica y se debe asegurar el traspaso de información a la base de datos, pero en este aspecto surge otro problema, ya que los lugares de trabajo se encuentran en zonas remotas, donde existe la presencia de ruido y las condiciones ambientales son extremas, por todo lo anterior, es fundamental realizar un vasto estudio de las condiciones operacionales del equipo crítico el que se requiere monitorear, pues así se garantiza la fiabilidad del sistema. Analizados los datos obtenidos, se calculan el RUL y MTTR [14].

La Figura 2.2 demuestra el funcionamiento del monitoreo de condiciones y como se asocian con el diagnóstico y pronóstico. Como se menciona anteriormente, primero se obtiene una base de datos del equipo, la cual se necesita procesar. Una vez analizados estos datos se obtiene un estado de salud del equipo actual, en la cual se puede realizar un diagnóstico y un pronóstico a futuro del comportamiento del equipo.

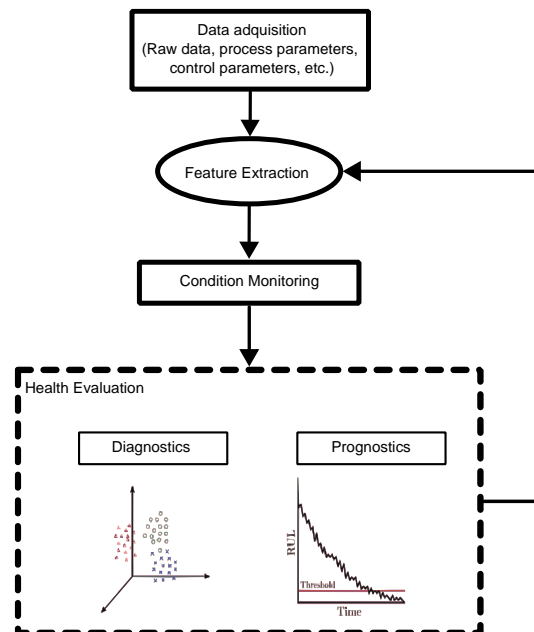


Figura 2.2: Funcionamiento del CBM [15].

El CBM se compone de 3 fases, detección de fallas, diagnóstico de fallas y pronóstico de fallas. La detección de fallas se realiza en base al análisis de los datos recopilados se detecta si hay alguna anomalía o resalto en el comportamiento de la máquina, lo que permite detectar la aparición de una falla. Pero anteriormente, para establecer el criterio de salud del equipo, es necesario conocer el comportamiento normal del equipo, ya que cada equipo presenta distinto comportamiento, por lo que es necesario la individualización y caracterización de cada equipo. El diagnóstico de fallas entrega una estimación del daño causado por la falla y el mecanismo de falla, lo cual es fundamental para la toma de decisiones de mantención. En el pronóstico de fallas se calcula el valor de RUL, para esto es importante conocer la incertidumbre, ya que en el caso de un CBM basado en datos tiene la base datos tiene que ser representativa [16].

En resumen, el CBM está constituido por las siguientes etapas:

- **Sensorización:** existen dos tipos principales de sensorización, la aplicación de sensores internos y sensores externos. Los sensores internos principalmente vienen incorporados en los equipos ya que tienen la función de medir variables significativas para el correcto funcionamiento del equipo, ya sea velocidad lineal, torque, corriente, etc. En cambio, los sensores externos son los más usados para el CBM, ya que permiten recolectar gran cantidad de datos, los cuales deben ser almacenados en algún dispositivo de almacenamiento o computador. Los sensores más utilizados son los acelerómetros, *Inertial Measurement Unit* (IMU), sensores acústicos y sensores de temperatura. 1
- **Procesamiento de señales:** una vez obtenidos los datos, es necesario procesarlos para poder interpretar de mejor manera la información. Para esto es común la utilización de filtros, remuestreo de datos y la clasificación de datos. Las principales técnicas de procesamiento de señales son: dominio de tiempo, dominio de frecuencia y dominio tiempo-frecuencia. En el primero se suelen realizar análisis estadísticos, como la amplitud *peak to peak*, el promedio, RMS, varianza y *kurtosis*, los que son determinantes para evaluar el estado de salud del equipo. El análisis del dominio de la frecuencia se realiza principalmente utilizando la transformada rápida de Fourier (FFT), lo que permite obtener el espectro de la frecuencia, facilitando el análisis de los datos, también es utilizada ampliamente utilizada la densidad espectral de frecuencia (PSD), la cual es la medición de la energía contenida a lo largo de la frecuencia. El análisis del dominio frecuencia-tiempo es utilizado principalmente para las señales no estacionarias
- **Aplicación de métodos de CBM:** existen distintos métodos con los cuales se realiza el CBM, los cuales son: basado en modelos y basado en datos. El primero pueden ser modelos deterministas o estocásticos, mientras que los modelos basados en datos suelen aplicar técnicas de *Machine Learning* para realizar regresiones y clasificaciones, los que permiten realizar predicciones de las fallas.

2.5 Machine Learning

Machine Learning se define como una subcategoría de la Inteligencia Artificial (IA), los cuales son algoritmos que tienen las características de aprender con o sin ayuda por parte del programador. Existen 3 tipos de aprendizajes: supervisados, no supervisados y semisupervisado:

- Aprendizaje Supervisado: el objetivo del aprendizaje supervisado es el de encontrar patrones en los datos que son ingresados, de esta manera entrega una clasificación de los datos en la salida. Los datos de entrada tienen que ser previamente etiquetados. Los algoritmos más utilizados son: SVM, *Naive Bayes*, *Nearest Neighbor*, *Decision Tree*, *Linear Regression*, *Ensemble Methods*.
- Aprendizaje No Supervisado: a diferencia del aprendizaje supervisado, los datos de entrada no se encuentran etiquetados, por lo que el objetivo de este tipo de aprendizaje es el de encontrar patrones para poder entrenar los datos, y así etiquetar a los datos que son ingresados.

2.5.1 Support Vector Machine (SVM)

Es un método de ML no paramétrico utilizado tanto para realizar una clasificación como una regresión. Los algoritmos de SVM entrenan un conjunto de datos etiquetado en n categorías para realizar un modelo que sea capaz de distinguir a que categoría pertenecen los nuevos datos ingresados, estableciendo un hiperplano N -dimensional.

Las ventajas de este método son:

- Resultados obtenidos son robustos y precisos con gran cantidad de datos de entrada.
- Buen rendimiento.
- Presenta una predicción eficaz para muestras pequeñas, por lo que es ampliamente utilizado para el diagnóstico de fallas en maquinaria.

Principales desventajas:

- Difícil de construir una serie de tiempo univariada, es aquella que toma datos de observaciones pasadas y predice los valores de la siguiente secuencia, por vida restante y tiempo de muestreo.
- Los parámetros deben ser ajustados específicamente para el problema y en ocasiones puede ser difícil.

2.5.2 Naive Bayes

Algoritmo de clasificación paramétrico que aplica el teorema de Bayes para clasificación probabilística. Para aplicar este método primero se requieren las probabilidades de que suceda cada clase existente. De esta manera, al ingresar nuevos datos se va recalculando la probabilidad obtenida inicialmente. El clasificador se realiza según la siguiente ecuación:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad \text{Ec (4)}$$

En donde:

- $P(c|x)$ es la probabilidad posterior de la clase c dada las características x .
- $P(c)$ es la probabilidad de la clase.
- $P(x|c)$ es la probabilidad de la clase dada el predictor.
- $P(x)$ es la probabilidad previa del predictor.

2.5.3 Decision Tree

Es un sistema de red, similar a un diagrama de flujo, que se compone de nodos y ramas, los nodos pueden ser nodos raíz o nodos intermedios. Los nodos intermedios representan una función, los nodos de hoja representan una etiqueta de clases. Los árboles de decisiones permiten dividir la base de datos en base a diferentes condiciones.

2.5.4 K-Nearest Neighbors (KNN)

Método de aprendizaje utilizado para la clasificación de datos, el cual genera *clusters* o grupos según los datos proporcionados para entrenamiento. Al ingresar nueva data, se consideran los k elementos más cercanos, a los que les asigna una clasificación según los k vecinos. En la Figura 2.3 se observa el comportamiento del algoritmo.

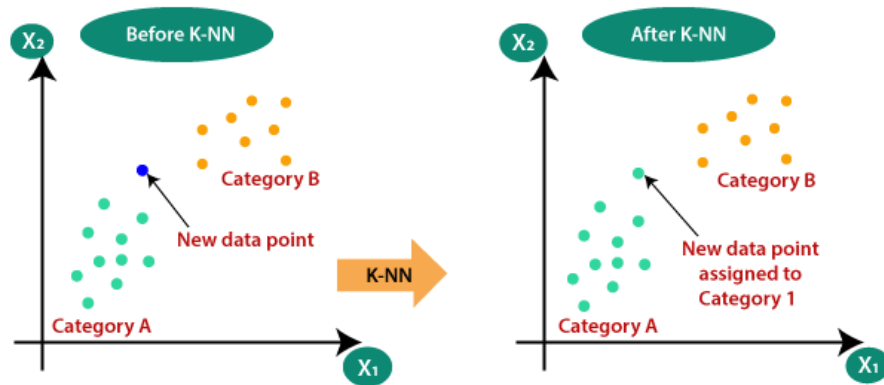


Figura 2.3: Funcionamiento del KNN.

2.5.5 Métodos de Ensamble

Los métodos de conjunto o de ensamble son algoritmos de aprendizaje automático que entrenan varios modelos, en este caso las distintas mediciones que se realizarán, logrando resultados mejores y más robustos que los obtenidos si se entrena cada modelo por separado.

El principio de los métodos de conjunto es tomar aprendices débiles y transformarlos en alumnos fuertes. En esta analogía el estudiante débil es un modelo base que tiene un alto sesgo o, por el contrario, tiene demasiada varianza. El alumno fuerte es un modelo de conjunto que reúne las mejores condiciones de los modelos bases, así dependiendo del caso reduce el sesgo o la varianza de los modelos de base, logrando mejores resultados.

Existe gran cantidad de combinaciones de modelos débiles, sin embargo, esta combinación debe ser de acuerdo con el objetivo que se busca lograr. En el caso de que los modelos bases se entrenen con el mismo algoritmo de aprendizaje, el modelo de conjunto obtenido se denomina homogéneo, mientras que si los modelos base se entrenan con respecto a distintos algoritmos de aprendizaje el modelo obtenido es del tipo heterogéneo. Los principales métodos de combinaciones son el *boosting*, *bagging* y *stacking*.

El *bagging* es un método que trabaja modelo homogéneo que entrena los modelos base en paralelo, como se observa en la Figura 2.4 promediando los resultados al final del proceso. La ventaja de entrenar los modelos en paralelo es aprovechar la independencia entre los algoritmos, además al promediar los resultados obtenidos por separado se reduce significativamente el error. Este método es ideal para reducir la varianza de los modelos débiles.

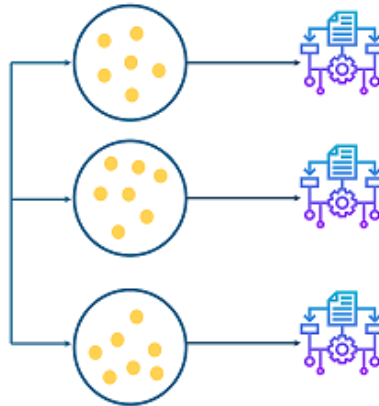


Figura 2.4: Aplicación de método de conjunto Bagging.

El *boosting* también es un método que trabaja con un modelo homogéneo, pero a diferencia del *bagging*, entrena los modelos en serie, o sea de manera secuencial, como se observa en la Figura 2.5, por lo que en este caso se resalta la dependencia entre los modelos base arrastrando los errores de cada modelo. El modelo final se obtiene promediando los resultados de los modelos débiles. Este modelo es ideal en el caso que se requiera reducir el sesgo de los modelos base.

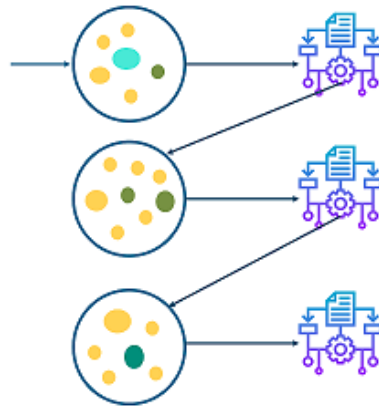


Figura 2.5: Aplicación de método de conjunto Boosting.

El *stacking* trabaja con modelos base heterogéneos, trabajando con ellos en paralelo, finalmente obtiene los resultados obtenidos con un metamodelo entrenado para generar predicciones basadas en datos.

2.6 Conceptos de Machine Learning

Para trabajar con datos en modelos de aprendizaje automático, es importante manejar nociones básicas de Machine Learning, las que son vistas a continuación.

2.6.1 División Train-Test

Al someter una base de datos a un modelo de Machine Learning, lo principal es la separación de los datos en dos, datos de entrenamiento y datos de testeo. Con esto, nos aseguramos de que el modelo sea capaz de reconocer las características principales de la base de datos y con ello extrapolar estas, hacia un conjunto de datos desconocidos. La división de conjuntos se realiza generalmente en 70% - 30% o 80% - 20% para entrenamiento y testeo, respectivamente. Realizar esta división de manera correcta es de suma importancia, ya que puede generar problemas en los modelos, debido al overfitting o underfitting [17].

2.6.2 Overfitting

El sobreajuste u overfitting, se manifiesta cuando el modelo entrenado se ha ajustado demasiado bien a los datos de entrenamiento. Ocurre generalmente cuando el modelo es complejo, se presentan mayores características que observaciones, lo que genera una generalización de datos, ya que asume el ruido como un comportamiento, en vez de distinguir las distintas variables de la base de datos. Es decir, el modelo pasará por alto los datos con los cuales no fue entrenado ya que se remite a los datos con los que fue entrenado [18]. Un ejemplo básico es el caso en que se entrene un modelo con distintas razas de perros, pero todos del mismo color, y al presentarle un perro de una raza con la que ya fue entrenada el modelo, pero de distinto color, no va a ser capaz de clasificar la raza del perro.

2.6.3 Underfitting

El underfitting, al contrario del overfitting, no es capaz de generalizar los datos de entrenamiento, ya que no se ajusta a la tendencia de datos con la que fue entrenado el modelo, ya que el modelo es muy simple, al no presentar variables suficientes para caracterizar los datos [18]. Como ejemplo al entrenar un modelo con solo una raza de perro, al presentar otra raza de perro, el modelo no va a ser capaz de atribuirlo dentro de una clasificación por falta de muestras.

Para evitar tener los problemas de ajuste de datos, es necesario considerar los siguientes factores:

- Encontrar una adecuada separación de datos de entrenamiento y testeo.
- Datos balanceados.
- Realizar un correcto ajuste de parámetros de los modelos.
- Contar con la cantidad necesaria de dimensiones.
- Ajustar el número de dimensiones del modelo.

2.6.4 Desbalanceo de Datos

El problema de los datos desbalanceados se presenta cuando una base de datos tiene clases muy disminuidas en comparación a otras, ocasionando una disparidad de las clases. El desbalanceo afecta directamente a los clasificadores de aprendizaje automático, ya que impacta considerablemente el entrenamiento de los modelos, pues el modelo no logra entender el comportamiento de la muestra [19].

Para trabajar con datos desbalanceados existen los siguientes métodos con lo cual se puede mejorar el rendimiento del clasificador:

- Trabajar con Matriz de Confusión como métrica de rendimiento.
- Entrenar con varios modelos de aprendizaje automático, con tal de encontrar el que tenga mejor rendimiento para estos casos.
- Realizar un remuestreo de datos, con tal de emparejar la cantidad de datos por clase.

2.6.5 Predictores

Las variables predictoras o atributos son los datos de entrada con los cuales trabaja el modelo para obtener los resultados o la variable objetivo. Los predictores generalmente tienen una relación empírica con la variable objetivo [20].

2.6.6 Parámetros e Hiperparámetros

Los hiperparámetros son propiedades de los modelos que regulan el proceso de aprendizaje y determinan los valores del parámetro que aprende el modelo. Los hiperparámetros son asignados antes del entrenamiento del modelo y son utilizados durante el entrenamiento del modelo. Algunos ejemplos de hiperparámetros son: tamaño de Kernel, número de iteraciones, número de capas, etc. Por lo cual, asignar los hiperparámetros correctos es de suma importancia para el éxito del modelo de aprendizaje automático.

Por su parte el parámetro tiene relación con los valores que el modelo de aprendizaje automático puede modificar durante el entrenamiento, utilizando algoritmos de optimización. Así, se puede regular el gradiente, coeficientes, sesgos, etc. [20].

2.7 Tipos de entrenamiento

Existen dos formas de realizar pronóstico, las cuales se diferencian en los datos que se entregan al algoritmo de Machine Learning.

2.7.1 Modelo de lazo abierto

El modelo de lazo abierto utiliza solo datos reales para realizar pronósticos. Por ejemplo, para predecir en el tiempo t se es necesario contar con los valores 1 al $t - 1$ registrados previamente, en este caso en la sensorización de la máquina. En otro caso, si lo que se busca es predecir para el tiempo $t + 1$, los valores utilizados serán los datos reales desde el 1 al t , no tomando en cuenta el valor t predicho con anterioridad [21].

2.7.2 Modelo de lazo cerrado

Este modelo es contrario al anterior, ya que, para predecir, se le entrega cierta cantidad de datos reales, los que son utilizados para predecir el valor inmediatamente posterior, el cual se agrega a los valores reales. Por ejemplo, si se requieren predecir valor desde el tiempo t hasta $t + k$, se utilizan solo los datos reales desde 1 hasta $t - 1$, desde ahí en adelante para predecir un valor en el tiempo i se utiliza hasta el valor predicho $i - 1$ [21].

CAPÍTULO 3: Estado del Arte

3.1 Pronósticos

El pronóstico ha tenido un gran desarrollo en los últimos años, debido a todos los beneficios que otorga a la industria. Así, Kan M. et al. [22] realizan una revisión de técnicas de pronóstico para maquinas rotatorias no estacionarias y no lineales. Las técnicas más comunes de pronósticos son para sistemas lineales. Primero que todo, los autores detallan la caracterización de los modelos de pronóstico, estos pueden ser basados en modelos, en datos o la combinación de estas dos. Los modelos se guían por leyes físicas y matemáticas que describan el funcionamiento de una máquina, pero para esto es imprescindible el total conocimiento de la máquina y su interacción con el exterior. En cambio, los modelos de datos utilizan la data adquirida mediante sensorización de la máquina, para a través del análisis de estos, predecir el comportamiento en el futuro. Esto está sujeto a la disponibilidad de los datos históricos de la máquina, ya que, si en el futuro se presentan situaciones, que no fueron sensorizadas, y por ende el modelo no es entrenado con esos datos, la predicción será absolutamente errónea.

La primera técnica de pronóstico es la matriz de coincidencia, la cual se enfoca en la predicción a largo plazo y es ampliamente utilizada para sistemas dinámicos y sistemas no estacionarios. Esta técnica se basa en dos matrices, una con vectores de características P que describe el ciclo de mantenimiento actual y una matriz con vectores Q que describe el ciclo anterior, una matriz S muestra la similitud entre cada valor de las matrices P y Q . La matriz S sirve como Q para la próxima comparación realizando un modelo de media móvil autorregresiva (ARMA). Esta técnica se complica para grandes cantidades de datos, debido al tiempo computacional que requiere su ejecución.

Los métodos secuenciales de Monte Carlo es otra técnica para el pronóstico, los cuales se basan en la inferencia bayesiana. Esta técnica es de carácter estadístico, trabajando con la distribución de los datos, para realizar la predicción en el tiempo. Mientras que el filtrado de partículas de Kalman es ampliamente utilizado en sistemas dinámicos no lineales, como por ejemplo en el monitoreo de turbinas eólicas, etc. Otra técnica estadística es el modelo oculto de Markov, que se trata principalmente de la secuencia de estados que define el comportamiento de la máquina, en el que cada estado depende únicamente del estado anterior. La dificultad de esta técnica es el gran número de estados y la cantidad de parámetros libres que se involucran con la máquina. Es utilizado para la predicción del RUL de herramientas de corte.

Uno de los algoritmos más ampliamente utilizado es el *Support Vector Machine* (SVM), el cual optimiza la curva limite al maximizar la distancia al punto más cercano, mediante un hiperplano de separación con la distancia máxima entre el plano y los datos más cercanos.

Los autores señalan que las técnicas mencionadas con anterioridad requieren de mejoras para ser aplicarse con efectividad en situaciones reales en los sistemas no lineales y no estacionarios, sobre

todo para tener un pronóstico confiable. Además, señalan que técnicas como la matriz de coincidencia requiere gran tiempo de proceso computacionalmente, por lo que no es aplicable en equipos que requieran de un pronóstico rápido el filtrado de partículas y el modelo de Markov necesitan de grandes cantidades de datos históricos, por lo que se limita el uso en la industria actual, ya que, al incluso tener una gran cantidad de datos, no se garantiza un pronóstico confiable.

Otro aspecto que destacan los investigadores es que en la industria se tienen pocos datos de equipos sensorizados son los que fallan, ya que constantemente se realizan mantenciones preventivas, por ende, no se tiene un real conocimiento del comportamiento de la máquina o equipo en situaciones de fallas críticas. Por lo demás, faltan técnicas de validación universal de pronósticos, ya que cada técnica cuenta con su propio sistema de validación, haciendo difícil la comparación entre ellas, ya que cuentan con enfoques y ponderaciones diferentes.

Peng Y. et al. [23] destacan los principales pasos a realizar en el CBM, en donde se vuelve a mencionar temas como el comportamiento físico de los componentes de la máquina, el de un modelo que describa el comportamiento del sistema y la importancia de la data histórica. Así mismo, mencionan las principales mejoras que se necesitan para ejecutar esta técnica, como lo es la velocidad y precisión del pronóstico, la validación y el desarrollo de software,

Los autores concluyen que la aplicación de un modelo por sí solo tiene grandes complicaciones para alcanzar predicción aceptable, por lo que se es necesario de combinar dos o más técnicas de predicción para potenciar las ventajas que otorga cada técnica por si sola, aminorando las desventajas de estas mismas, formando un modelo más fuerte y capaz de realizar una predicción más precisa y confiable. En la aplicación de esta técnica combinada está el desafío, ya que existe un sin número de combinaciones posibles, por lo que en encontrar la correcta para cada caso de aplicación está el éxito de la predicción. De igual manera, se mencionan los problemas que afectan a las técnicas de monitoreo, como lo son la necesidad de un método de validación general, lo costoso técnica y económicamente de la recolección de la data histórica que requieren algunos modelos y lo difícil de obtener datos de máquinas que estén fallando. También, los autores mencionan el caso de la aplicación en tiempo real de los pronósticos, ya que los modelos son entrenados con los datos en tiempo real, pero sin que el modelo de pronóstico se vaya actualizando, por lo que se necesitan modelos rápidos que sean capaces de actualizar el modelo en línea.

Sikorska J. et al. [24] realizan una investigación acerca de la selección de modelos de pronósticos de vida útil restante y dejan claro que el éxito de estos en la industria se ha visto limitado. Además de destacar las ventajas y desventajas de las distintas técnicas. Se indica que la correcta elección del método es imprescindible para alcanzar el éxito del modelo, ya que además de necesitar un modelo validado, es necesario conocer el tipo de implementación se quiere realizar, sin embargo, siempre está el riesgo de que el modelo no cumpla con las expectativas.

Los autores presentan diversas definiciones de pronóstico, y determinan que la descripción más acabada es la que otorga la ISO 13381-1, la que entiende como pronóstico como ‘estimación del tiempo hasta la falla y el riesgo de uno o más modos de fallas existentes y futuros’, por lo que infieren que además en el cálculo de la vida útil, el pronóstico este asociado a determinar cómo

pueden iniciarse otros métodos de fallas. Así, la definición de pronóstico queda determinada por: modos de fallas existentes y las tasas de deterioro; criterios de iniciación de futuros modos de fallas, relación entre modos de falla y tasas de deterioro; sensibilidad de las técnicas de monitoreo y las condiciones y suposiciones que se realizan al implementar un modelo. En la figura 3.1 se observa un cuadro elaborado por los autores acerca de su percepción de la relación entre el diagnóstico y el pronóstico y las preguntas que se atribuyen a estos.

Los autores detallan el paso a paso para realizar un pronóstico, el que se detalla a continuación:

- Preprocesamiento de datos: diagnósticos, definiciones de falla, identificación de posibles fallas futuras y selección de un modelo de pronóstico adecuado.
- Pronóstico del modo de falla existente: en este paso se obtiene la identificación del tiempo hasta la falla, cálculo del RUL e iterar el proceso.
- Pronóstico de modo de falla futuro: se analizan los modos de fallas futuras más probables y se realiza el cálculo del RUL para las potenciales fallas futuras.
- Pronóstico posterior a la acción: de detectan acciones que podrían afectar los modos de fallas críticos.

Así con la clasificación en 3 niveles la capacidad de pronóstico, en donde cada uno es más complejo que el otro, entregando resultados confiables y útiles. En donde el nivel 3 es el más real, ya que consideran las acciones de mantenimiento que se realizan en el equipo.

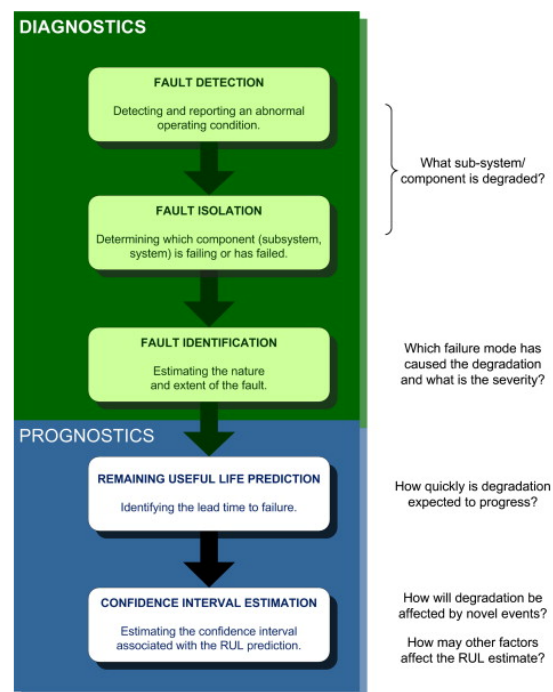


Figura 3.1: Paso a paso del Diagnóstico y Pronóstico [24]

Benkedjough T. et al. [25] realizan una investigación sobre el cálculo del RUL en rodamientos, componentes que tienen un comportamiento no lineal. Los rodamientos son unos de los componentes que presentan más fallas en máquina eléctricas, por lo que aumentar la confiabilidad de este componente impacta de gran forma a la máquina. Las técnicas de pronóstico utilizadas en los rodamientos por lo general son basadas en datos, utilizando el análisis de vibraciones.

El pronóstico aplicado en rodamientos, de igual manera cuenta con el problema de la alta dimensionalidad de los datos, al buscar caracterizar el comportamiento de la máquina con los datos obtenidos. Es por esto que los autores buscan generar un indicador de salud que sirva para evaluar la degradación de los rodamientos, realizando un preprocesamiento de datos y técnicas de reducción de datos, las cuales no mencionaremos ya que no son de interés para esta investigación. Lo que resalta de esta investigación es la utilización de SVR, *Support Vector Regression*, los que entregan los indicadores de salud. La Figura 3.2 demuestra la estructura que siguen los autores para realizar el CBM sobre los rodamientos, en ella se aprecia también el cómo se llevan a cabo estos pasos.

Support Vector Regresor junto al *Support Vector Classification* componen el SVM. El SVR tiene por objetivo estimar la relación entre las variables de entrada y de salida, con el supuesto de que la distribución conjunta de ambas variables es desconocida. Para realizar la parte experimental de la investigación, los rodamientos son puestos a pruebas en una plataforma experimental que permite controlar todas las cargas sobre estos. Los rodamientos son puestos a pruebas hasta que fallan y los datos obtenidos son preprocesados, para caracterizar el estado de salud del componente. Entrenan un modelo SVR, entre 16000 y 25000 datos, los que entrega valores de RMSE, de entre 0,021 a 0,023, por lo que es se consideran buenas predicciones.

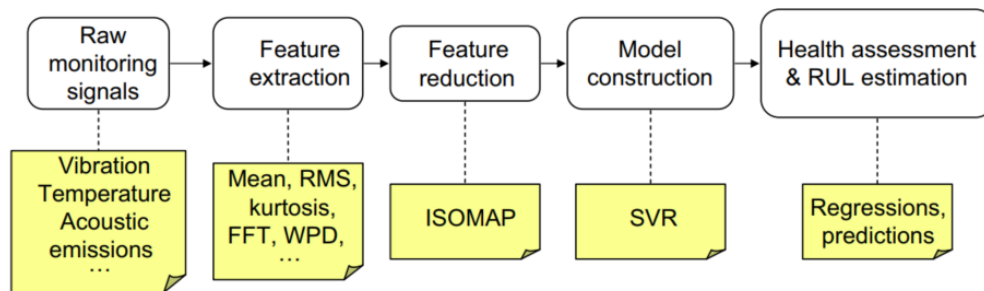


Figura 3.2: Proceso del CBM [25]

3.2 Fallas mecánicas y monitoreo en motor de retroexcavadora.

Los autores [3] estudian el comportamiento de una retroexcavadora Komatsu modelo WB93R-5, además establece la fiabilidad de cada sistema de componentes del equipo. En el estudio se analizaron 180 retroexcavadoras, con un promedio de 2120 y 177 horas de operación, anualmente y mensualmente, respectivamente. Para establecer los estados de salud del motor de la

retroexcavadora, se recolectaron información de 39 retroexcavadoras en donde se estableció el comportamiento de los siguientes aspectos: origen, causa de ocurrencia, complejidad de la reparación, elementos involucrados, consecuencias de las fallas, naturaleza de estas y forma de eliminación. Los resultados arrojaron que la mayor cantidad de fallas se debieron a la fractura de componentes, como, por ejemplo: rodamientos del árbol de levas, sellos, bomba de alimentación de combustible.

Existe una gran cantidad de autores que se enfocaron en el cigüeñal, señalando que los principales orígenes de las fallas recaen en vibraciones torsionales, mal ajuste del motor o mala lubricación.

Los autores [26] señalan que el origen de las grietas en los cigüeñales se localiza principalmente en el área cercana a los filetes del eje y en los filetes del muñón principal, debido a que son concentradores de esfuerzo y cualquier mala terminación en el proceso de rectificado puede provocar esto. Los cigüeñales están expuestos a flexión rotatoria y flexión inversa, por lo que sufre una fatiga multiaxial. Los autores analizan un cigüeñal de un motor mono cilíndrico que se desempeña en el área de la agricultura, el cual falló a los dos años catastróficamente. Para dar con la falla se le practica metalografía a la muñequilla y realizan un análisis de elementos finitos para dar con las causas de la falla. Los resultados arrojaron que las grietas iniciaron en el filete del eje y en el orificio de aceite lubricante, originados por el desalineamiento de muñones, desequilibrio de contrapesos y mala lubricación, esta última produce mayor deflexión en las muñequillas. Para prevenir esta falla se considera poner énfasis en proceso de producción para tener mejor calidad de mecanizado y realizar un mejor tratamiento térmico. A su vez, durante la operación es fundamental el monitoreo de vibraciones, para detectar a tiempo las desalineaciones y realizar una buena lubricación.

Los autores de [27], realizan un análisis de confiabilidad de motores diésel de camiones, basado en datos recopilados en el servicio de post venta. Para obtener una mejor interpretación de los datos dividen el motor en seis subsistemas, biela-cigüeñal, sistema de distribución de gas, sistema de arranque, sistema de suministro de combustible, sistema de refrigeración y sistema de lubricación. De estos sistemas, el que presenta más fallas fue el, sistema de suministro de combustible seguido del sistema de arranque y del sistema de distribución de gas. Los resultados obtenidos demuestran que los componentes electrónicos presentan una gran cantidad de fallas dentro del motor, esto se debe a las condiciones de trabajo en que se desempeñan los vehículos pesados con motor diésel, como lo es las vibraciones, el polvo y la corrosión. De esta manera destacan que un mismo motor utilizado en distintos vehículos pesados va a tener un comportamiento distinto, ya que las condiciones de trabajo y las exigencias sufridas por cada equipo van a depender de su uso.

Siguiendo con el estudio de fallas de los cigüeñales, los autores [28] establecen las principales fallas que presentan los cigüeñales, realizan un cuadro esquemático en el cual desmenuzan las fallas. En la Figura 3.3 se observa el cuadro esquemático, en donde se puede distinguir las dos principales causas de falla: alta torsión vibracional y mala lubricación. En el caso de la vibración torsional, esta se presenta por una alta tensión debida al cambio de los pares, momentos de flexión, cargas axiales y velocidades. Los pares de torsión son acompañados de esfuerzos cortantes sobre el cigüeñal, como

se menciona anteriormente en especial sobre los filetes del eje y del muñón principal. La lubricación depende por sobre todo del estado de los elementos del sistema de lubricación, como la bomba de aceite, el filtro de aceite y el regulador de presión. El motor necesita funcionar con una presión de aceite y utilizar un índice de viscosidad según lo estimado por el fabricante, ya que sin ellos no se forma la película de aceite que impida el desgaste de los metales. Los autores realizaron mediciones experimentales con 3 cigüeñales distintos, que fueron rebajados, entre 1 a 5 mm, mediante mecanizado. Estos cigüeñales fueron utilizados a distintas velocidades, entre 20 y 240 [km/h], las cuales midieron utilizando un dinamómetro. Obtenida la velocidad lineal, la transforman a velocidad de rotación, la que introdujeron en la ecuación 5, que permite estimar la torsión que se aplica sobre el cigüeñal. En donde D es el diámetro del cigüeñal y N es la velocidad de rotación.

$$T = \frac{4\pi m}{3600} N^2 D \quad \text{Ec. (5)}$$

Los resultados obtenidos por los autores demostraron que a medida que se aumentaba el mecanizado en la pieza, aumenta la vibración torsional, lo que reduce la vida útil de la pieza. Ya que en todas las velocidades que fue puesta a prueba la pieza, la vibración era mayor en la pieza modificada que en la pieza estándar. Mientras que para los muñones su vida útil se conserva solo si se reduce la velocidad de operación, además de una adecuada lubricación. Mencionan, además, que, al aumentar las vibraciones, aumenta la temperatura de operación del motor, por lo que la viscosidad del lubricante ya no es la misma que la recomendada por el fabricante, así la película de aceite requerida para prevenir el desgaste no es capaz de cumplir su función.

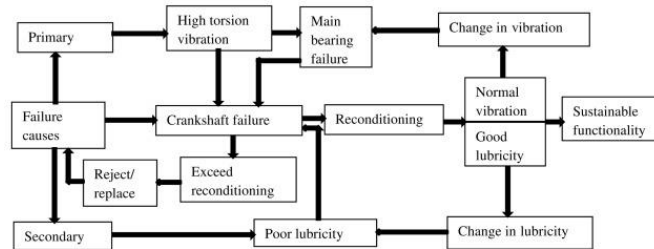
Por su parte [29], realiza una revisión de las investigaciones realizadas sobre el monitoreo condición en motores de combustión interna. Los autores señalan que el análisis de vibración del motor es una de las técnicas más eficientes al momento de encontrar la generación de una falla o un indicio de mal comportamiento entre la gran cantidad de componentes que interactúan en un motor de combustión interna. También, destacan la importancia del monitoreo de condición, ya que la detección de una falla incipiente puede prevenir de que está falla se prolongue a otros componentes, significando un ahorro en los costos de mantención y además aumenta la disponibilidad de la máquina. De esta manera, se mencionan las principales causas de vibración de la máquina, las que, dicho de paso, pueden ser torsionales o longitudinales:

Holgura: se presenta cuando las piezas sueltas son las causantes de las vibraciones.

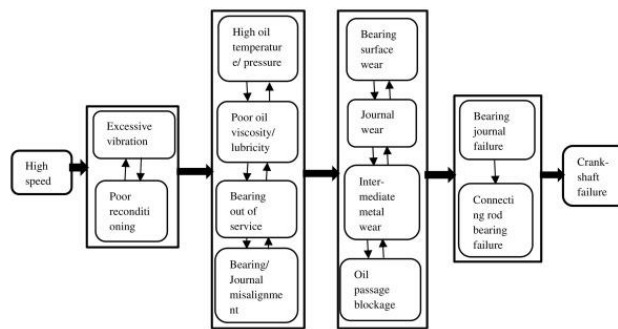
Resonancia: se presenta cuando la frecuencia de vibración es igual o próxima a la frecuencia natural de la máquina, provocando que se amplifiquen las vibraciones de la máquina.

Además, resaltan la importancia del monitoreo que recae en el conocimiento del estado de salud de la máquina, permitiendo identificar las fallas antes de que sean catastróficas, lo que genera demoras en el caso de que no haya el repuesto indicado. Además, mencionan que el rendimiento de la

máquina decae, ocasionando un mayor consumo de energía. Por lo que el monitoreo es fundamental si se quiere optimizar el rendimiento y disponibilidad del motor.



a. Crankshaft failure modelling and analysis concepts



b. Crankshafts' failure mechanisms

Figura 3.3: Cuadro Esquemático de las principales fallas en los cigüeñales [28].

En [30] los autores estudian las consecuencias en el rendimiento y vibración de un motor de combustión interna debido al desgaste del pistón. Para ello montan un banco de pruebas donde monitorean el comportamiento del motor durante una falla por desgaste del pistón. La falla por desgaste ocurrida en el mecanismo pistón-cilindro es debido a las altas tensiones de contacto sobre estas superficies, producidas por la alta fricción y temperatura, que se presentan por mala lubricación, condiciones de funcionamiento o el juego excesivo. Así, el desgaste genera una destrucción de las superficies, afectando directamente el rendimiento del motor y aumentando las emisiones de gases contaminantes.

De esta manera, los autores montan un banco de ensayos con un motor de cuatro cilindros en línea, el cuál fue equipado con cerca de 30 sensores, entre ellos acelerómetros, sensores de torque, posición, temperatura, presión, etc. En específico, el acelerómetro fue instalado en el bloque del motor, frente al cilindro n°3. El acelerómetro cuenta con un rango de frecuencias de 16 [kHz]. El motor fue sensorizado en 4 estados: encendido, velocidad de potencia nominal, inactivo y detenido. Finalmente, los resultados revelan una disminución de 5,5 [Nm] del par motriz y 3,5 [kW] de la potencia del motor. Por su parte, la tasa de flujo de escape aumenta a 128,7 [L/min] siendo que para un estado normal tenía un valor de 33 [L/min], este gran aumento se debe a la mayor holgura entre el pistón y el cilindro, generado por el desgaste, permitiendo un mayor flujo de gases. La señal vibratoria fue sometida a la transformada *wavelet* continua, la que permite una representación

tiempo-frecuencia de la señal. Así, se demuestra que la amplitud de las vibraciones del motor aumenta considerablemente, además de un aumento de un 475% en la media y un 101% en el RMS de la señal vibratoria, con lo cual, los autores demuestran que la señal vibratoria del motor permite identificar las fallas por desgaste del pistón.

CAPÍTULO 4: Adquisición de Datos y Modelo de predicción de RUL propuesto

Como se ha mencionado anteriormente, se monitorearon dos retroexcavadoras durante 60 días en los que operaron 12 horas aproximadamente por día, con la particularidad de que a una retroexcavadora sufrió una falla en el motor.

La falla del motor fue producida por la mala lubricación, según el monitoreo, no se cumple con las horas de trabajo establecidas por el fabricante, lo que afectó directamente a un pistón, el cuál salió eyectado, provocando un forado en el bloque del motor, quedando inutilizable, siendo necesario un nuevo motor. Este tipo de falla, como se describe anteriormente en el estado del arte, es altamente prevenible a través del monitoreo de la señal vibratoria, ya que como explica [30], la señal vibratoria se ve notablemente afectada por fallas en el pistón de un motor diésel, lo que de igual manera se puede apreciar en los datos utilizados en esta investigación, como se verá más adelante.

El monitoreo del motor fue llevado a cabo gracias a sensores inalámbricos que fueron instalados en puntos críticos del motor para obtener la aceleración que presenta en los 3 ejes triaxiales. Los sensores crearon una red inalámbrica que se comunica vía *Bluetooth*, a su vez, esta red envía los datos recopilados a un servidor web mediante señal 3G/4G/LTE. Los sensores fueron equipados con un sensor IMU (Bosh BMI160) y un acelerómetro (Rohm Semiconductor kx220), con los cuales se obtiene un rango de medición de frecuencia de 1 a 5000 [Hz]. Cabe mencionar que los sensores estaban equipados con batería LiPo de 3,7 [V] y 1400 [mAh], que eran cargadas inalámbricamente, por lo que contaban con autonomía de funcionamiento los sensores. El sensor es protegido por una carcasa elaborada con resina epoxi, con un tamaño de 90 x 40 x 29 [mm], la cual fue diseñada para tener resistencia al contacto, ser hermética al polvo y resistencia a la humedad [6]. En el interior de la carcasa, fue instalado un imán de alto campo, el que permite una instalación rápida y sencilla de los sensores en el motor.

El proceso de adquisición de datos consistía en que, una vez obtenida la aceleración del motor, es calculado el valor RMS y se realiza el análisis del espectro de frecuencia, mediante la Transformada Rápida de Fourier (FFT). La FFT actúa en una ventana de 1 minuto de datos recopilados, en donde se guardan las 5 frecuencias predominantes más altas, con sus respectivas amplitudes en los 3 ejes. Posteriormente, cada 15 minutos se promedian los datos guardados, obteniendo los datos con los que finalmente quedan almacenados para ser analizados por el experto en vibraciones.

Los datos fueron etiquetados por un experto según la norma ISO 2372, la que sirve para evaluar el estado de salud del equipo, mediante el análisis de la severidad vibratoria. Se consideran 4 estados, los cuales son: Bueno, Aceptable, Insatisfactorio e Inaceptable, estos estados son atribuidos según el valor de RMS. Es importante mencionar que la norma utilizada por el experto es para máquinas rotativas estacionarias que operan en el rango de 600 a 12000 rpm, la cual se encuentra obsoleta, ya que fue reemplazada por la norma ISO 10816 e ISO 4867, sin embargo, es la norma que más se ajusta al rango de frecuencias registrados, razón por la cuál es utilizada [31].

4.1 Muestra de datos

Realizando un análisis estadístico de RMS de ambas retroexcavadoras, se obtiene lo observado en la Tabla 4.1. En donde se aprecia que la retroexcavadora que no falla obtiene una menor media de RMS, esto principalmente a que no valores tan altos, como se observa las Figuras 4.2 y 4.3. De la misma manera, el coeficiente de variación demuestra una menor dispersión en los valores de la retroexcavadora que no falla. En la Figura 4.1 se realiza una comparación de las amplitudes de las 5 frecuencias predominantes en la hora 516 de medición, y en la Tabla 4.2 se presenta un resumen de las amplitudes con sus respectivas frecuencias para cada estado de salud, en donde se observa que la máquina que falla presenta una mayor amplitud para las mismas frecuencias y que a medida que empeora el estado de salud aumenta la amplitud para la misma frecuencia. En las Figuras 4.4 y 4.5 se observan el comportamiento de la amplitud relacionada a la frecuencia predominante en el eje Z, se observa que el comportamiento de la amplitud es incluso más notorio que el RMS en los cambios de estados. En definitiva, se tiene que todos los parámetros arrojan una notoria diferencia en el comportamiento de las máquinas, por lo que es posible generar una distinción en el comportamiento vibracional de ambas máquinas.

Tabla 4.1: Análisis Estadístico.

Retroexcavadora	Falla	No Falla
Media	0,0359	0,331
Desviación estándar	0,0096	0,0067
Coefficiente Variación	26,68%	20,38%

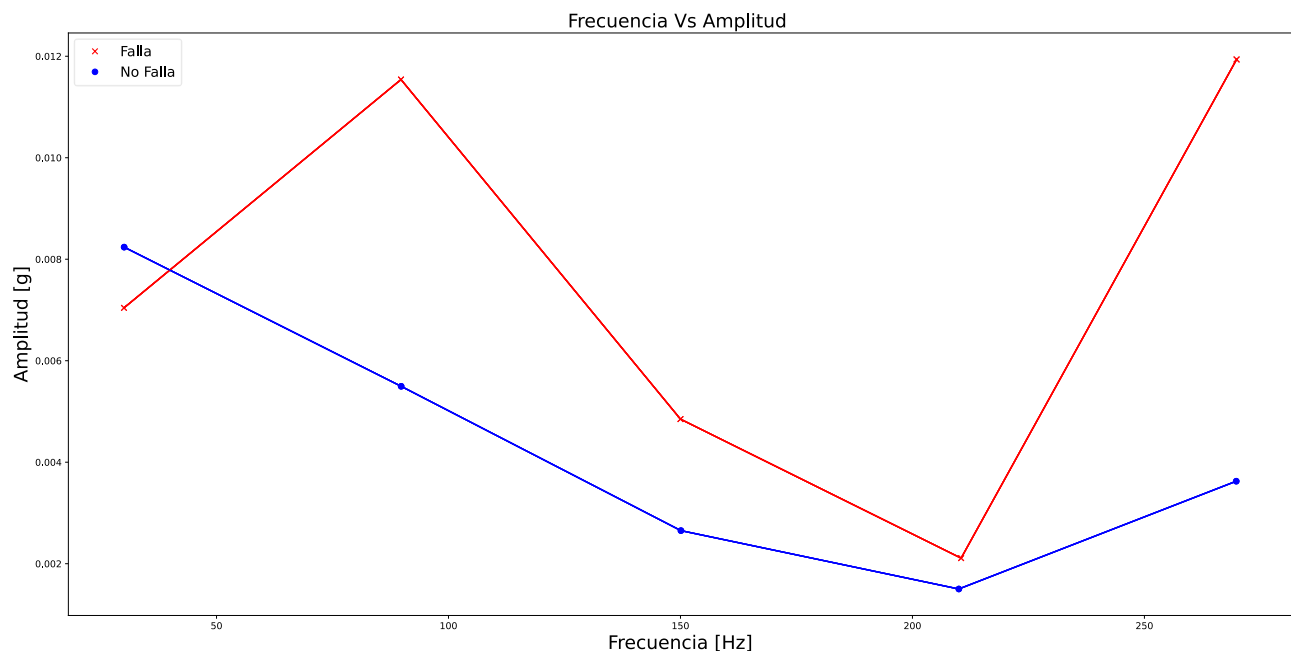


Figura 4.1: Comparación de las retroexcavadoras.

Tabla 4.2: Amplitudes [g] de cada frecuencia en su respectivo estado.

Frecuencias [Hz] / Estado	0	1	2	3
30	0,007219	0,009017	0,012889	0,030321
90	0,005479	0,007306	0,011574	0,027764
150	0,002944	0,004913	0,008947	0,025581
210	0,002075	0,004048	0,008188	0,024852
270	0,004483	0,006318	0,010694	0,027171

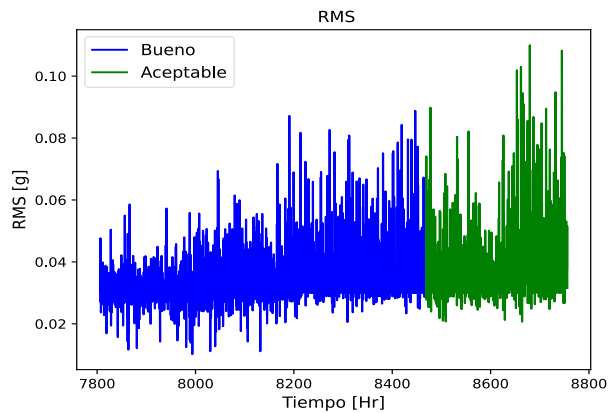


Figura 4.2: RMS de retroexcavadora que no falla.

4

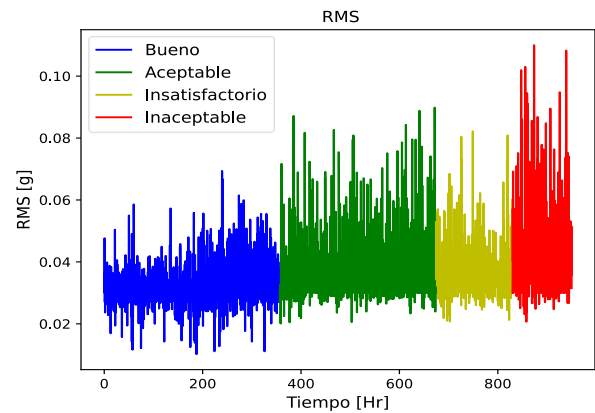


Figura 4.3: RMS de retroexcavadora que falla.

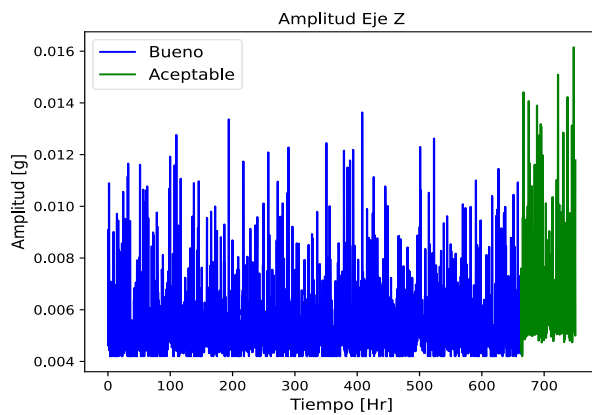


Figura 4.4: Amplitud de retroexcavadora que no falla.

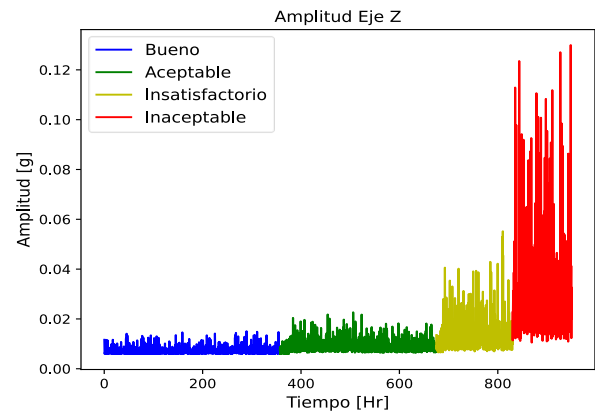


Figura 4.5: Amplitud de retroexcavadora que falla.

4.2 Modelo Propuesto

Como bien se ha mencionado, este trabajo se sustenta en mediciones realizadas en una investigación anterior [6]. En esta investigación trabajaron con clasificadores con el objetivo de poder determinar el RUL de la retroexcavadora en un periodo de tiempo anterior al predicho por el experto encargado de etiquetar los datos. El experto determinó que el motor de la retroexcavadora entraría en el estado

inaceptable 120 horas antes de la falla, mientras que los investigadores mediante técnicas de *Machine Learning* lograron aumentar el RUL a 169 horas antes de la falla, considerando que la falla se produce a las 950 horas de funcionamiento desde la última mantención, es decir lograron aumentar el pronóstico del RUL en casi 50 horas, las que en el ámbito de la mantención son significativamente favorables.

Ahora bien, en la investigación anterior se trabajó siempre desde la perspectiva de manejar la totalidad de los datos, en los que pasa por los 4 estados de salud del motor de la retroexcavadora, y en base a eso se enfocaron en calcular el RUL. Pero en la realidad, cada máquina es expuesta a distintas condiciones de trabajo y tiene diferentes cuidados por parte del operador, es por eso que se propone un modelo en el que se trabaje con los datos en tiempo real del trabajo de la retroexcavadora, es decir, que a medida que se vayan recopilado las mediciones, someterlas a los modelos de *Machine Learning* para obtener el RUL. De esta manera, al entrenar los algoritmos de regresión y clasificación con la totalidad de datos, el entrenamiento será siguiendo la técnica del método de lazo abierto, ya que como predictores se utilizarán sólo datos reales, provenientes de la sensorización, por el contrario, el modelo pondrá en práctica los algoritmos bajo el método de lazo cerrado, ya que los predictores serán datos obtenidos gracias al algoritmo de regresión. Así, se pretende crear un modelo que sea útil en la industria para predecir la vida útil del motor del equipo.

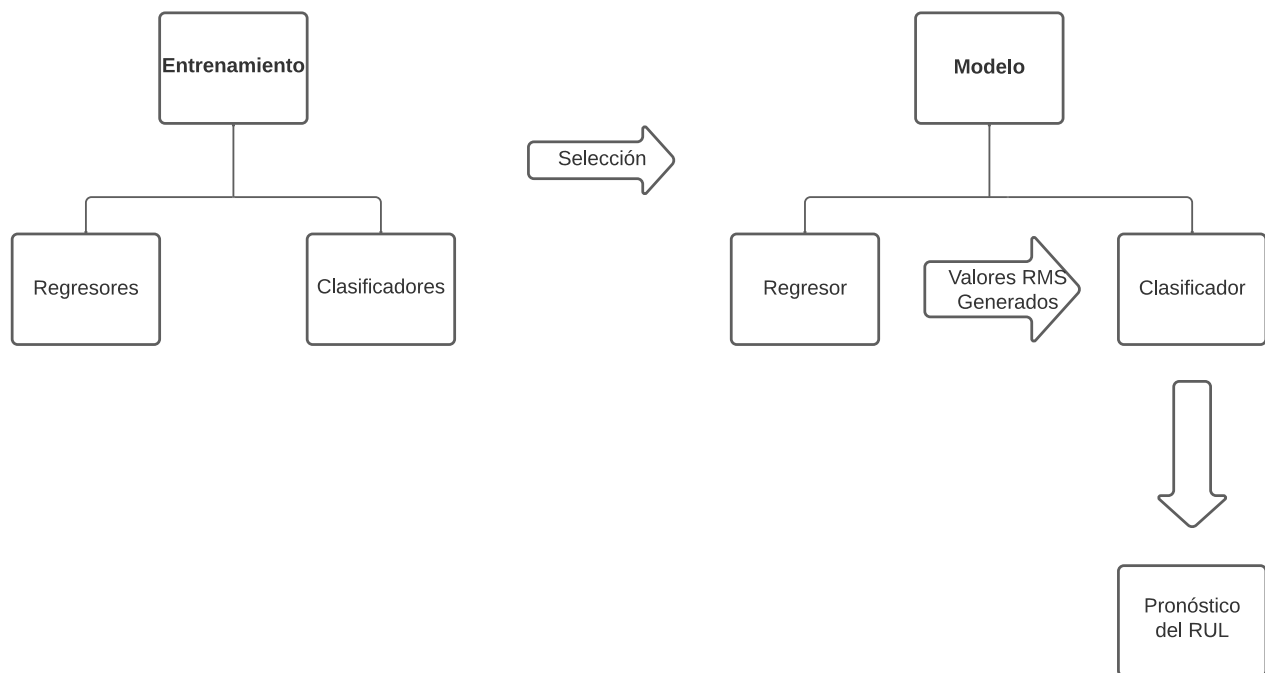


Figura 4.6: Cuadro esquemático de modelo propuesto

En la Figura 4.6 se observa el funcionamiento del modelo, el cual se explica de la siguiente manera. En primera instancia, los datos medidos son sometidos a un regresor, el cual se encarga de pronosticar el comportamiento de la máquina en un determinado periodo de tiempo futuro. A

continuación, estos datos son ingresados a un clasificador, que tiene el objetivo de asignar a cada uno de los 4 estados de salud los datos predichos. De esta manera, se busca calcular el RUL de la máquina a medida que está sume horas de trabajo.

4.3 Entrenamiento

Primero que todo, es necesario seleccionar los algoritmos de regresión y clasificación para ingresarlos al modelo. Estos se seleccionarán a partir de la comparación de los resultados de entrenamiento con los datos recopilados en ambas retroexcavadoras. En primera instancia, se explicará el método entrenamiento del clasificador, ya que, en el modelo propuesto, el clasificador es el encargado de atribuir un estado de salud a los datos, en consecuencia, se necesita conocer que datos de entrada necesita el clasificador para tener un buen rendimiento, y en base a esos datos, entrenar el regresor para que tenga como variable de salida la variable de entrada del clasificador.

4.4 Clasificadores

En primera instancia seleccionará el clasificador con el cuál se trabajará, así a continuación se presentarán los algoritmos que fueron entrenado, como también las métricas de validación pertinentes a los clasificadores. Es necesario destacar el problema de desbalanceo de datos que afectan directamente el rendimiento de los algoritmos, en la Figura 4.3 se logra apreciar que la cantidad de datos decae al avanzar por los estados. En la Tabla 4.3 se detalla la cantidad de datos por estado de salud, en donde a medida que se deteriora el estado de la máquina, disminuye la cantidad de datos registrados. Cabe mencionar, que los estados de salud fueron etiquetados con los números 0, 1, 2 y 3, correspondiendo a los estados Bueno, Aceptable, Insatisfactorio e Inaceptable, respectivamente.

Tabla 4.3: Cantidad de datos por Etiqueta.

Estado de Salud	Cantidad de Datos
0	1427
1	1272
2	620
3	481

Como se estableció en el marco teórico la asignación de los conjuntos de entrenamiento y testeo es de gran importancia para evitar problemas como el *overfitting* y el *underfitting*. Así, al entrenar el algoritmo obtiene una generalización del comportamiento de los datos, con lo cual puede extender este conocimiento a otros datos. El conjunto de entrenamiento permite conocer que tan bien está aplicando el conocimiento adquirido en el entrenamiento en nuevos datos. La cantidad de datos con

que se entrena el modelo es del 70% del total, dejando el restante 30% para el testeo del modelo, también se entrena el modelo distribuyendo 80 % -20 % los conjuntos, pero los resultados no sufrieron una gran variación. Como medida para enfrentar el desbalanceo de datos, los porcentajes de entrenamiento y testeo no se aplican directamente al total de datos, sino que se aplica por cada etiqueta, de esta manera se concatena el 70% de cada estado de salud para conformar el vector de entrenamiento, al igual que para el vector de testeo. Lo anterior se realiza con el fin de contar con datos provenientes de todas las etiquetas, pues en el supuesto de que se tomará el 70% del total, los datos de la etiqueta 2 sería mínimos y de la etiqueta 3 serían inexistentes en el vector de entrenamiento.

Además, como se verá más adelante en el capítulo de resultados, al trabajar con variables de entrada RMS o amplitud no se obtienen tan buenos resultados como al trabajar con el RMS acumulado, el cual contiene la información histórica del comportamiento de la máquina, por lo que es fundamental para los clasificadores.

4.4.1 Algoritmos de Clasificación y principales parámetros

A. *Random Forest*: algoritmo meta-estimador, es decir, toma como parámetro a otro estimador, como por ejemplo un árbol de decisiones. Se constituye de una serie de conjuntos de árbol de decisión, calculando el promedio de estos para mejorar la precisión y regular el sobreajuste [32]. Al igual que el algoritmo *Decision Tree*, es utilizado el índice de Gini, el cuál es el encargado de decidir cómo se ramifican los nodos del árbol de decisión. La fórmula se denota a continuación.

$$G(\text{nodo}) = \sum_{k=1}^c p_k(1 - p_k), p_k = \frac{\text{número de observaciones de clase } k}{\text{todas las observaciones del nodo}} \quad \text{Ec. (6)}$$

Los principales párametros son:

- *n_estimators*: determina el número de árboles en el bosque, el valor ingresado debe ser un valor entero y el valor por default es 100.
 - *Criterion*: sirve para seleccionar el criterio de ramificación de los nodos, en este caso el índice de Gini, las otras opciones *entropy* y *log_loss*.
 - *max_depth*: profundidad que alcanza el árbol, si no se señala un valor el árbol se expandirá hasta separar las clases puras.
- B. *SVM*: algoritmo de clasificación que minimiza el error cuadrático, a través de un hiperplano en un espacio n-dimensional, siendo n la cantidad de características [33]. En este algoritmo es utilizado la función *kernel*, que se encargan de asignar los datos a los espacios n-dimensionales diferentes, el *kernel* puede ser de tipo gaussiano, lineal, polinómico o *sigmoide*.

C. *Gaussian Bayes*: complementando a lo detallado en la sección 2.5.2, el algoritmo entrega medidas de incertidumbre a la clasificación, a través del cálculo de la probabilidad de distribución de datos [34].

El principal parámetro es *priors* el cuál le detalla las probabilidades previas de las clases detalladas en la Ec.4 en el capítulo 2.

D. *Decision Tree*: algoritmo que utiliza un conjunto de reglas para tomar decisiones, mediante preguntas de si/no que definan un conjunto de datos, de manera que se vaya parcializando las características, hasta aislar todos los datos que pertenecen a las determinadas clases, así se forma el árbol de decisión. Cada pregunta se denomina nodo, mientras que los últimos nodos del árbol se denominan nodos hoja.

Por cada división el algoritmo trata de obtener un subconjunto cada vez más pequeño, por lo que trata de minimizar la función de pérdida lo máximo posible, el que entrega por default el algoritmo es la impureza de Gini, Ec. (6), la cuál es una medida de varianza entre las diferentes clases [35].

Los parámetros principales son los mismos que para *Random Forest*.

E. *Boosting*: es un método de ensamble, es decir, utiliza múltiples algoritmos con tal de mejorar el rendimiento en comparación a ocupar cada uno por separado, lo que se realiza por qué cada modelo tratar de reducir los errores de modelo anterior [36]. Por lo es necesario minimizar la función de pérdida, la que se detalla a continuación.

$$L = - \sum_{i=1}^n y_i \log p + (1 - p) \log(1 - p) \quad \text{Ec. (6)}$$

En donde y_i es el valor observado y p es la función de probabilidad predicha para la clase.

Los principales párametros son:

- *loss*: es la función de pérdida a optimizar, el valor por defecto para la clasificación es el de *log_loss*, el cuál corresponde a la ec. 6. Las otras opciones son *deviance* y *exponential*.
 - *n_estimators*: indica el número de árboles con los que contará el modelo.
 - *Learning_rate*: indica la contribución que cada árbol aporta al resultado final, toma valores entre 0 en infinito.
- F. *Bagging*: al igual que el algoritmo anterior es un algoritmo de ensamble, pero a diferencia, este pondera a través de una ponderación o votación los resultados que entrega cada algoritmo por separado, consiguiendo una predicción final [37]. En el caso de la clasificación las predicciones se obtienen a través del voto mayoritario de los algoritmos, que consiste en elegir la clase que contenga la mayor suma de votos o la mayor probabilidad sumada de los algoritmos, como se observa en la ec. 7.

$$\widehat{f_{bagg}} = \widehat{f}_1(X) + \widehat{f}_2(X) + \dots + \widehat{f}_b(X) \quad \text{Ec. (7)}$$

Los principales parámetros son:

- *estimators*: es el modelo que ajusta los datos, el modelo por default es *Decision Tree*.
- *n_estimators*: es el número de estimadores bases en el conjunto.

4.4.2 Métricas de Validación

Para validar el entrenamiento de los clasificadores es necesario utilizar métricas que sirvan de comparación entre los distintos algoritmos utilizados, en busca de seleccionar el algoritmo que mejor rendimiento alcance con los datos proporcionados. Es por eso por lo que se utiliza la Matriz de Confusión, la cual se observa en la Figura 4.7. En ella se dan a conocer los términos TP, FP, FN y TN, los que significan verdadero positivo, falso positivo, falso negativo y verdadero negativo, respectivamente. En palabras simples, la matriz de confusión demuestra los valores de la clasificación correctos e incorrectos. Así, a partir de la matriz se obtienen las siguientes métricas:

- *Recall*: hace alusión a la cantidad de variables positivas que fueron clasificadas correctamente.

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

- *Precisión*: discrimina la cantidad de valores positivos verdaderos dentro de la totalidad de valores positivos clasificados.

$$Precision = \frac{TP}{TP + FP}$$

- *Exactitud*: Demuestra la totalidad de clases que fueron clasificadas correctamente.

$$F - measure = \frac{2 * Recall * Precision}{Recall + Precision}$$

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figura 4.7: Matriz de Confusión [38]

En relación con las métricas, hay cuatro casos que entregan información relevante, los que son:

- Alta precisión y alto *recall*: el algoritmo maneja perfectamente la clase.
- Alta precisión y bajo *recall*: el algoritmo no se adapta bien a la clase, pero en el caso de que si lo haga es altamente confiable.
- Baja precisión y alto *recall*: el algoritmo detecta bien la clase, pero de igual manera incluye datos de otras clases.
- Baja precisión y bajo *recall*: el algoritmo no es capaz de clasificar acertadamente la clase.

4.5 Regresores

La regresión se define como una relación empírica entre las variables de entrada o predictores y las variables de salida u objetivo, sin embargo, con la sensorización realizada y las características extraídas de ella se hace imposible esta realización. Siendo más preciso, las características extraídas del motor son las siguientes: RMS, las 5 frecuencias predominantes con sus respectivas amplitudes de onda para cada eje, tiempo de trabajo durante el monitoreo, tiempo de trabajo desde la última mantención, tiempo de trabajo desde el último cambio de aceite de motor y el tiempo desde la última revisión de válvulas del motor. Si se decide enfocarse en los tiempos monitoreados, lo único que se puede tomar como base son las horas recomendadas por el fabricante para realizar las mantenciones, el cambio de aceite y revisar las válvulas del motor, así que desde este punto de vista sólo se puede determinar si se cumple o no con el tiempo entre las mantenciones. Ahora, desde el punto de vista de la aceleración, de la cual se calcula el RMS, frecuencia y amplitudes, se puede dar cuenta de que no existe una relación empírica entre alguna de estas variables, lo que no es óptimo para ser utilizadas como variables de entrada o salida del algoritmo. Por ejemplo, al monitorear la velocidad de un automóvil, es posible sensorizar el torque que es entregado a la rueda del y con ello calcular la velocidad del automóvil, o al calcular la fuerza realizada por un equipo o componente, al sensorizar la aceleración y conociendo la masa es posible conocer la fuerza. Lo que se pretende

demostrar con estos ejemplos es que para las características extraídas de la base de datos son insuficientes para entrenar de buena manera el algoritmo.

Con tal de proponer una posible solución, en esta investigación se trabajará con valores consecutivos de RMS y amplitud, excluyendo la frecuencia, ya que las 5 frecuencias predominantes no sufren mayor variación en relación a los estados de salud de la retroexcavadora, en cambio, como se observa en las Figuras 4.4 y 4.6, el RMS se incrementa a medida que empeora el estado de la máquina, al igual que la amplitud, la cual incluso tiene variaciones más determinantes, la amplitud utilizada es la que corresponde a la primera frecuencia en el eje z. Así, el trabajar con mayor cantidad de predictores tiene por objetivo mejorar el rendimiento de los regresores. Por su parte, la variable objetivo corresponderá al RMS, ya que es la variable con la que el clasificador obtiene mejores resultados y además es el que evalúa de mejor manera la severidad de las vibraciones, por lo que es óptimo para el pronóstico del RUL. De esta manera, se conforma un proceso iterativo, en donde el valor RMS pronosticado se agrega al vector de entrenamiento, para posteriormente ser utilizado como predictor, así, se logra pronosticar el estado de la máquina en un cierto periodo de tiempo.

Por consiguiente, se obtiene la siguiente matriz de entrenamiento, en donde los predictores son valores consecutivos de amplitud y RMS, y el variable objetivo es el valor RMS inmediatamente posterior a los valores de entrada. De esta manera la matriz de entrenamiento consta de 6 columnas, 3 con datos de amplitud y 3 con datos de RMS, las cuales toman distintos valores de inicio y termino. Como se aprecia en la matriz que está a continuación, en donde m es la cantidad total de valores de entrenamiento. Es necesario mencionar, que el entrenamiento del regresor se realizará siguiendo la técnica del modelo de lazo abierto, es decir, considerando sólo valores reales (medidos) para entrenar el algoritmo, mientras que la aplicación del modelo será siguiendo la técnica de modelo de lazo cerrado, mencionada en la sección 2.7.2, que utiliza como predictores datos generados por el mismo algoritmo de regresión.

Ya que el fin del entrenamiento es pronosticar valores en cierta ventana de tiempo, es necesario configurar la matriz de entrenamiento para que el pronóstico sea enfocado en el valor de RMS siguiente a los valores de entrenamiento. De esta manera la matriz de entrenamiento consta de 6 columnas, 3 con datos de amplitud y 3 con datos de RMS, las cuales toman distintos valores de inicio y termino. Como se aprecia en la matriz que está a continuación, en donde m es la cantidad total de valores de entrenamiento.

$$\begin{bmatrix} amp_n & amp_{n+1} & amp_{n+2} & rms_n & rms_{n+1} & rms_{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ amp_{m-2} & amp_{m-1} & amp_m & rms_{m-2} & rms_{m-1} & rms_m \end{bmatrix} \Rightarrow rms_{n+3}$$

4.5.1 Algoritmos de Regresión

El entrenamiento de los algoritmos se realizará con diversos algoritmos, los cuales unos son más complejos que otros, la aplicación será desde el modelo más sencillo hasta los más complejos. Los modelos cuentan con diversos hiperparámetros, los cuales permiten realizar un sinnúmero de configuraciones, los que permiten adaptar y potenciar su rendimiento dependiendo el caso de aplicación. A continuación, se mencionan los principales hiperparámetros de los algoritmos utilizados:

- A. *Linear Regression*: método de regresión más básico, consiste en hallar la relación lineal entre la variable independiente x y la variable dependiente y , como se observa en la ec. 8. Los términos w y b representan las estimaciones de parámetros lineales que deben ser calculadas a través de la función de costo, el cual minimiza el error entre el valor predicho y el valor real, lo que se observa en la ec.9, la que se conoce como función de error cuadrático medio [39].

$$Y = wX + b \quad \text{Ec. (8)}$$

$$\min \left(\frac{1}{n} \sum_{i=1}^n (\text{pred}_i - y_i)^2 \right) \quad \text{Ec. (9)}$$

- B. *>Linear Regression Ridge*: algoritmo utilizado para reducir la complejidad del modelo y evitar el ajuste excesivo que realiza la regresión lineal. Lo anterior lo realiza minimizando la ec. 10, llamada función de regularización [40].

$$\text{Min}(\|y - Xw\|^2 + \alpha * \|w\|^2) \quad \text{Ec. (10)}$$

El principal parámetro es *Alpha*, constante que multiplica al término L2, que es la función de regularización. Puede tomar valores positivos, en el caso de que sea cero se considera una regresión lineal simple.

- C. *SVR*: es similar al SVM, pero solo que es para regresión, cuenta con los mismos parámetros. Tiene por objetivo minimizar el error entre el valor predicho y el valor real, como se observa en la ec. 11, en donde C es el parámetro de regularización [41].

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n |\xi_i| \quad \text{Ec. (11)}$$

- D. *Bagging*: se diferencia del clasificador principalmente por que el regresor calcula el resultado final en base a la ponderación en vez de una votación de los resultados obtenidos por cada modelo por separado [37].

- E. *Boosting*: se diferencia del clasificador por la función de pérdida utilizada, ya que en el caso de la regresión se utiliza el error cuadrático medio, la que se observa en la ec. 9 [42].
- F. *Random Forest*: al igual que el *Bagging*, en la regresión en comparación a la clasificación, el resultado final se obtiene gracias a la ponderación del resultado de cada árbol [43].
- G. *Decision Tree*: al igual que los otros algoritmos de regresión, es necesario calcular el error cuadrático medio, salvo esta diferencia, los parámetros son los mismo que para la clasificación [35].
- H. *XGB Regressor: Extreme Gradient Boosting*, como su nombre lo da a entender es muy parecido al *Boosting*, por lo que sus parámetros son prácticamente los mismos [44].

4.5.2 Métricas de validación

R^2 : conocido como coeficiente de determinación, mide como se ajusta un modelo de regresión a los datos reales. Toma un valor de 0 a 1, en donde 0 corresponde al peor de los casos y 1 a un ajuste perfecto de los datos. También puede resultar valores negativos, que indican un pésimo ajuste del modelo.

MSE: calcula el error cuadrático medio entre los valores predichos y los valores reales. En el mejor de los casos toma un valor 0, a medida que incrementa el resultado es peor el rendimiento del algoritmo.

CAPÍTULO 5: Resultados

5.1 Clasificadores

Los clasificadores que se entrenarán son los siguientes: *Random Forest*, *Decision Tree*, *Boosting*, *Bagging*, *SVM* y *Gaussian Bayes*.

En primera instancia se entrena el clasificador con datos provenientes de los regresores, es decir, las salidas de los modelos. Sin embargo, no dan buenos resultados, ya que rondan el 50% de precisión, como se observa en la Tabla 5.1. Por lo que se procede a entrenar los clasificadores con RMS y RMS acumulado, con lo cual mejoran considerablemente los resultados bordeando el 99% de precisión. Lo anterior, se observa en la Tabla 5.2, en donde se exponen los resultados de la matriz de confusión.

Tabla 5.1: Resultados del entrenamiento de los clasificadores.

Modelo	Accuracy	Precision	Recall	F1
<i>Random Forest</i>	47,93%	39,46%	39,43%	39,36%
<i>Decision Tree</i>	48,49%	39,24%	39,26%	39,15%
<i>Boosting</i>	53,53%	40,83%	40,23%	38,47%
<i>Bagging</i>	47,25%	38,99%	38,93%	38,88%
<i>SVM</i>	39,31%	30,37%	26,09%	15,98%
<i>Gaussian Bayes</i>	54,28%	27,08%	37,79%	31,53%

Tabla 5.2: Resultados del entrenamiento de los clasificadores utilizando RMS acumulado

Modelo	Accuracy	Precision	Recall	F1
<i>Random Forest</i>	99,73%	99,66%	99,53%	99,59%
<i>Decision Tree</i>	99,88%	99,89%	99,83%	99,86%
<i>Boosting</i>	99,88%	99,89%	99,83%	99,86%
<i>Bagging</i>	99,88%	99,89%	99,83%	99,86%
<i>SVM</i>	95,83%	86,98%	38,44%	49,87%
<i>Gaussian Bayes</i>	97,66%	98,01%	97,69%	97,85%

Según los resultados de la Tabla 5.2, todos los algoritmos presentan buenos resultados, ahora bien, si también consideramos los resultados de la Tabla 5.1, el clasificador seleccionado es *Boosting*, ya que tiene resultados levemente mejores que el resto. Cabe mencionar, que, para la utilización de RMS acumulado en los algoritmos de clasificación de estado de salud de la retroexcavadora, es estrictamente necesario tener en cuenta de que es necesario manejar la data histórica de RMS desde un punto base, como, por ejemplo, la última mantención del motor.

5.2 Regresores

Primero que todo, se utilizan como predictores los valores consecutivos de RMS, resultados que se observan en la Tabla 5.3, en ella se observan los valores de R^2 , observando el pésimo rendimiento de los algoritmos de regresión, al igual que para los resultados de la Tabla 5.4, la cual presenta los resultados utilizando los datos de amplitud como predictor.

Como se menciona anteriormente, el vector de entrenamiento cuenta con valores de RMS y amplitud, es por eso por lo que para más adelante para que funcione el modelo, es necesario que el modelo haya sido entrenado para ambas variables. En ambos casos destacan un par de algoritmos, *Bagging* y *XGB Regressor*, pero solo para el entrenamiento, pues el rendimiento del conjunto de testeo es bajo comparado al rendimiento del entrenamiento.

Tabla 5.3: R^2 obtenido utilizando RMS.

Modelo	Train	Test
<i>Linear Regression</i>	0.0544	-0.0173
<i>Linear Regression Ridge</i>	0.0227	-0.0094
<i>SVR</i>	-6.805	-5.5862
<i>Bagging</i>	0.6319	-0.0608
<i>Boosting</i>	0.9209	-0.2047
<i>Random Forest</i>	0.7986	-0.2059
<i>Decision Tree</i>	0.1634	-0.1209
<i>XGB</i>	0.9161	-0.2517

Tabla 5.4: R^2 obtenido utilizando Amplitud.

Modelo	Train	Test
<i>Linear Regression</i>	0.3152	0.9759
<i>Linear Regression Ridge</i>	0.1929	0.1995
<i>SVR</i>	-22.84	-25.84
<i>Bagging</i>	0.7728	0.3241
<i>Boosting</i>	0.9937	0.1167
<i>Random Forest</i>	0.8778	0.1985
<i>Decision Tree</i>	0.5149	0.2727
<i>XGB</i>	0.9759	0.2022

Debido a los pobres rendimientos alcanzados, se intenta el predictor visto en el capítulo anterior, el que cuenta con los valores consecutivos de RMS y amplitud para predecir el RMS inmediatamente posterior a los predictores. Así, se obtienen los resultados de las Tablas 5.5 y 5.6. En ellas se observan mejores resultados para amplitud que para RMS, en el caso de RMS, *XGB Regressor* sigue teniendo un buen desempeño para el entrenamiento, pero con los valores de test el rendimiento cae rotundamente, lo que se evidencia en la Figura 5.1 en donde se observa que el rendimiento del *Bagging Regressor*.

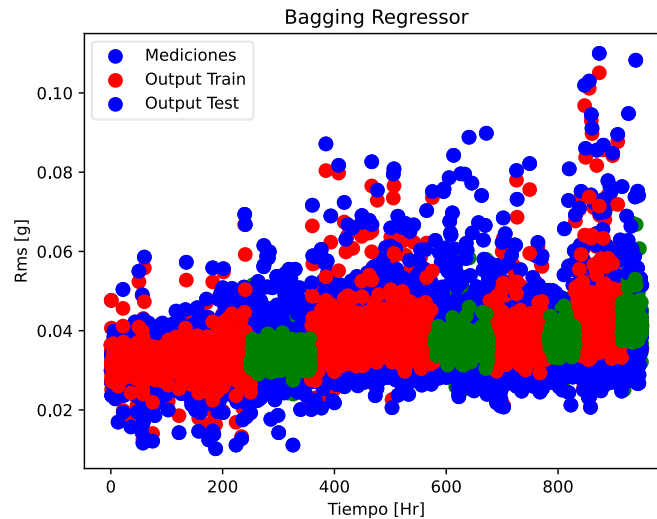


Figura 5.1: *Bagging Regressor*.

Tabla 5.5: R^2 de RMS obtenido utilizando RMS y Amplitud como entradas.

Modelo	Train	Test
<i>Linear Regression</i>	0.0820	0.0160
<i>Linear Regression Ridge</i>	0.0511	0.0124
<i>SVR</i>	-6.805	-5.5862
<i>Bagging</i>	0.6647	-0.2052
<i>Boosting</i>	0.9695	-0.1288
<i>Random Forest</i>	0.8187	-0.1114
<i>Decision Tree</i>	0.1975	-0.0264
<i>XGB</i>	0.9689	0.2059

Tabla 5.6: R^2 de Amplitud obtenido utilizando RMS y Amplitud como entradas.

Modelo	Train	Test
<i>Linear Regression</i>	0.3230	0.3263
<i>Linear Regression Ridge</i>	0.2073	0.2080
<i>SVR</i>	-22.84	-25.91
<i>Bagging</i>	0.9832	0.2059
<i>Boosting</i>	0.9973	0.1942
<i>Random Forest</i>	0.8811	0.2763
<i>Decision Tree</i>	0.4958	0.3318
<i>XGB</i>	0.9832	0.2059

Con ello, nos vemos en la problemática de que ya utilizado los principales predictores con los que se contaban dada la sensorización realizada, los resultados de coeficiente de determinación son muy bajos para ocupar siquiera un algoritmo en el modelo propuesto. Así, se ve en la obligación de ocupar filtros de media móvil, los que sirven para suavizar señales vibratorias con el objetivo de facilitar la identificación de patrones, siendo ideal para la aplicación de los algoritmos. El objetivo de suavizar señales es resaltar las tendencias de la muestra, ya que elimina el ruido no deseado, el que en este caso proviene de sucesos propios de la operación de la máquina, como choques, impericia del operador o condiciones del terreno. El mencionado filtro, cada un N datos calcula la muestra de las N muestras consecutivas de la medición. En las Figuras 5.2, 5.3 y 5.4, se observa como quedan el gráfico de RMS vs Tiempo sometido a Filtros de $N=4$, 6 y 10 respectivamente, en comparación a la señal original. Es clara la tendencia de suavización de las señales a medida que aumenta el valor de N, por lo que los nuevos valores de la señal RMS serán sometidos a los algoritmos de regresión, resultados que se observan en las Tablas 5.7 y 5.8.

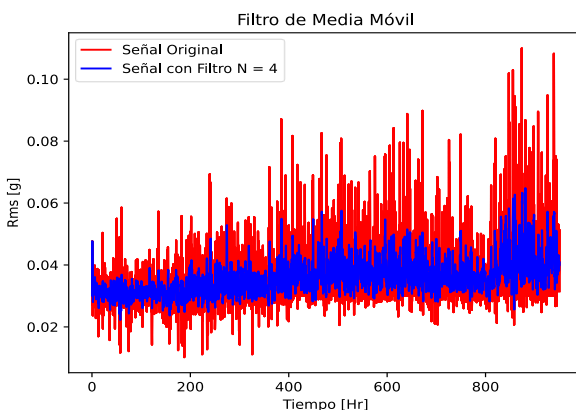


Figura 5.2: Filtro de Media Móvil con $N = 4$.

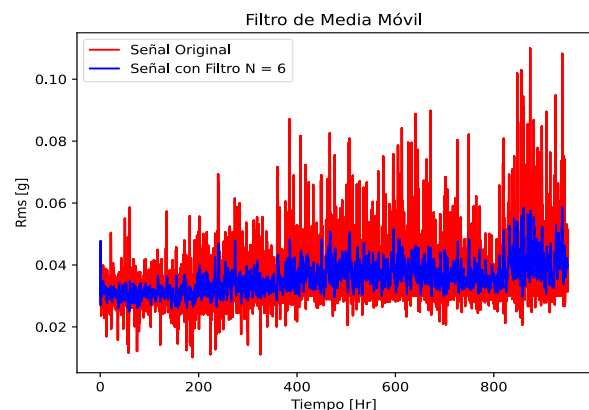


Figura 5.3: Filtro de Media Móvil con $N = 6$.

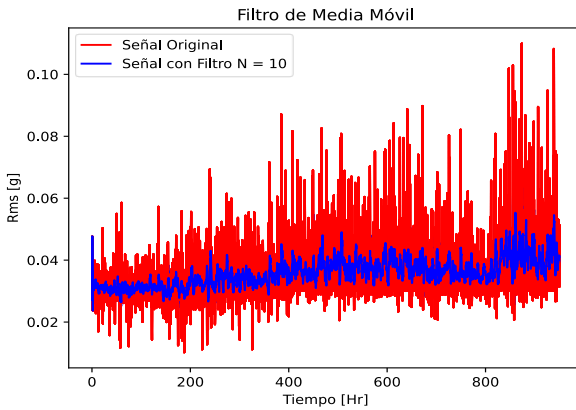


Figura 5.4: Filtro de Media Móvil con N = 10.

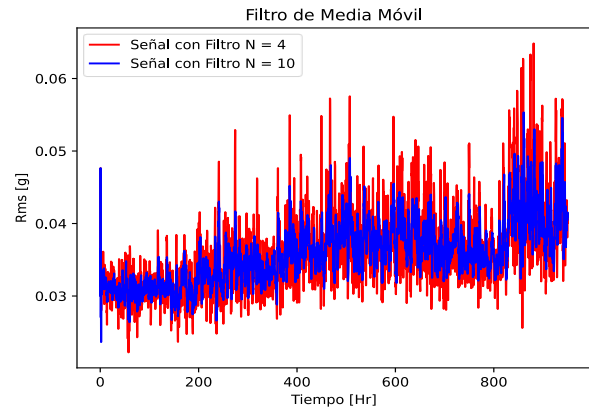


Figura 5.5: Comparación de Filtros.

Tabla 5.7: R^2 de RMS obtenido algoritmo Bagging.

	Filtro N =4	Filtro N =6	Filtro N =4
Train	0.9031	0.9378	0.9704
Test	0.6247	0.7696	0.9920

Tabla 5.8: R^2 de Amplitud obtenido algoritmo Bagging.

	Filtro N =4	Filtro N =4	Filtro N =4
Train	0.9551	0.9792	0.8795
Test	0.8956	0.9425	0.9769

En las tablas anteriores, se demuestra que a medida que el valor de N aumenta, los resultados del coeficiente de correlación de los algoritmos mejoran considerablemente. De esta manera, la selección del regresor será en base a los resultados de la señal sometidos al filtro de media móvil con N =10, en donde el mejor rendimiento lo tiene el algoritmo Bagging, por lo cuál es el seleccionado para ser utilizado en el modelo.

En la Figura 5.6 se observa el gráfico resultante del entrenamiento del modelo, en especial del *Bagging Regressor*. En él se logra identificar mínimas diferencias entre los datos obtenidos en la sensorización y los datos provenientes del entrenamiento del modelo.

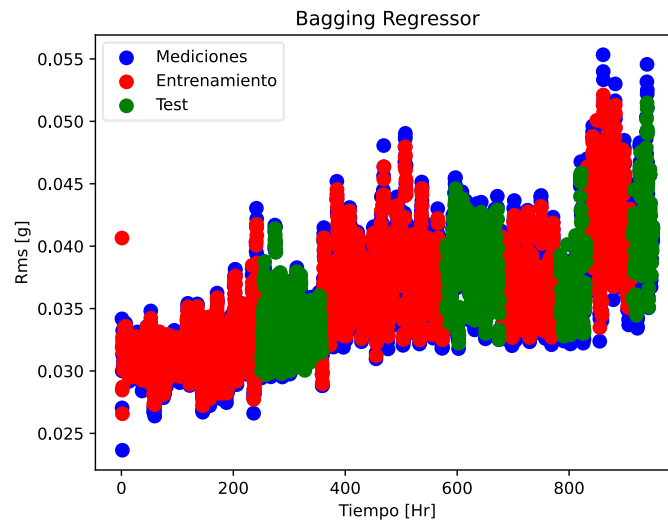


Figura 5.6: *Bagging Regressor.*

5.3 Modelo

Finalmente, el objetivo de esta investigación es la de aplicar un modelo de regresor y posteriormente clasificador, el cual tenga la tarea de pronosticar el comportamiento de la máquina y etiquetar los datos generados en un estado de salud. De esta forma, se permite establecer el RUL.

Para poder cumplir con lo anterior, es necesario seleccionar un modelo de regresor y otro de clasificador. Así en base a los resultados analizados, se determina que para el caso del regresor el modelo seleccionado será el de *Bagging*, gracias al desempeño que logra en los entrenamientos, en donde este modelo presenta los mejores resultados. Por su parte, para el clasificador se selecciona el modelo de *Decision Tree*, por las mismas razones que en el caso del regresor.

Dando paso a los resultados, en la Figura 5.7 se observa el resultado de la regresión, los cuales dista bastante de lo esperado. Para la realización del modelo, en primera instancia se le entregaron los primeros 2000 datos, los cuales se someten al modelo de lazo cerrado, es decir, las predicciones se sustentan en datos ya predichos, pues como se ha mencionado anteriormente, el modelo propuesto es para realizar predicciones en tiempo real, por lo que vamos a suponer que se han medido esa cantidad de datos en el momento en que entra en acción el modelo, prediciendo los datos futuros. De esta manera se observa en el gráfico que el modelo generado en ningún momento se asemeja a los datos reales, sin importar la cantidad de datos ingresados al predictor. Primero, se realizó el modelo utilizando el algoritmo *Bagging* como regresor, al no entregar resultados satisfactorios, se aplicaron la totalidad de regresores vistos con anterioridad, sin embargo, todos tuvieron bajos rendimientos en los resultados.

De igual manera, con los datos generados por el regresor, son aplicados en el algoritmo de *Decision Tree*, para ver el rendimiento del clasificador, así se obtiene una precisión del 26,5% y un recall de 24,6%, es decir, de acuerdo con lo visto en la sección 4.4.2, el clasificador no es capaz de clasificar

correctamente las clases, por lo que es demasiado impreciso para realizar una predicción del RUL confiable.

Algunas de las causas por las cuales se entiende que el modelo generado no cumple con las expectativas son las siguientes:

- En primera instancia el regresor es entrenado en lazo abierto, mientras que la aplicación del modelo es mediante lazo cerrado. De esta manera aumentan las probabilidades de falla, ya que lo ideal es que el entrenamiento y la aplicación sean realizadas de la misma manera.
- Los datos registrados no cuentan con una variable o parámetro que determine las variables o fuerzas externas a las que es sometida la retroexcavadora durante su operación. Las retroexcavadoras al ser una maquinaria móvil no estacionaria están afectadas por factores externos, ya sea el tipo de superficie del camino, tipo de trabajo realizado, material transportado, o incluso el nivel de experiencia del operador al realizar las labores. En palabras simples, el hecho de que una perturbación aumente considerablemente en un momento dado no se puede atribuir a si es por desgaste del funcionamiento del motor o si es porqué la maquina pasó por un bache de gran tamaño o si el operador realiza una operación indebida. A pesar de esto, en el momento de realizar las mediciones no se consideraron parámetro o elementos propios de la máquina que pudiesen explicar las mediciones registradas, ya que teniendo esos valores se podrían ocupar de punto de comparación o de base con las mediciones de aceleración, permitiendo tener mayor claridad del efecto del entorno operacional en las vibraciones. Debido a esto, no es posible entender de real manera el comportamiento de la máquina.

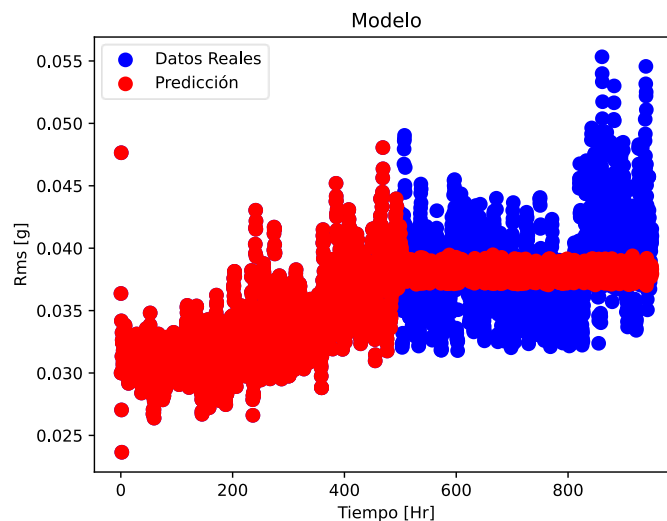


Figura 5.7: Modelo realizado.

CAPÍTULO 6: Conclusiones

En esta investigación, en donde se buscaba aumentar la ventana de pronóstico del RUL, para anticiparse a las fallas, se propuso un modelo basado en datos recopilados de la vibración del motor de dos retroexcavadoras en funcionamiento, con las cuales se caracterizó el estado de salud de la máquina en 4 estados.

El primer paso del modelo fue entrenar un regresor, el cuál es el encargado de pronosticar el comportamiento de la máquina en el futuro. Posteriormente, se entrenó un clasificador, el cual tiene como función de atribuir a uno de los 4 estados de salud cada dato generado por el regresor.

Finalmente, un modelo integrado por un regresor y un clasificador, los cuales fueron seleccionados mediante una comparación de las métricas de validación propias de estos, buscaron pronosticar el RUL de la máquina. A pesar del análisis, esto no fue posible, ya que el pronóstico generado por el regresor no es capaz de asimilar los datos generados a los datos reales, en parte por las diferencias entre los modelos de entrenamiento y aplicación, sumado a la limitación de los datos.

Así, tras realizada la presente investigación, se destacan las siguientes conclusiones:

- A pesar de que los datos con los que se trabaja reflejan diferencias en los comportamientos vibratorios de ambas máquinas, el alto procesamiento de los datos al momento de la medición limita la aplicación de nuevas técnicas y la obtención de una mejor predicción.
- Si bien los clasificadores tienen un alto rendimiento en el entrenamiento, este disminuye significativamente en el modelo. Esto se debe principalmente al bajo rendimiento del regresor con la predicción de datos. Además, para que el clasificador tenga un rendimiento aceptable es necesario contar con el RMS acumulado, lo que es difícil de obtener, ya que son requeridas las mediciones históricas de la máquina después del último ajuste del motor.
- Los regresores no son capaces de predecir el comportamiento, siendo la principal limitante del modelo. Aquí, se vuelve a mencionar el tema de la medición de datos, ya que los datos obtenidos fueron insuficientes para alcanzar un buen rendimiento del regresor. De manera que, se deja constancia para trabajos futuros que al aplicar este tipo de modelos se requiere medir una mayor cantidad de características vibratorias de la máquina.
- Si bien, la investigación se ve restringida a un modo de falla en particular, se cree que el modelo puede ser aplicado con el correcto entrenamiento de los algoritmos para otras máquinas que presenten otros modos de falla. Ya que, como se ve reflejado, el comportamiento de la máquina que falla dista bastante de la máquina que no falla, por lo que al ser aplicado en otro modo de falla los algoritmos debiesen ser capaces de distinguir los estados de salud. Para generalizar el funcionamiento del modelo, sería necesario sensorizar un mayor número de máquinas para observar el comportamiento vibracional, con sus respectivos modos de fallas y entrenar el modelo en base a ello.

En general, esta investigación sienta bases para trabajos futuros, ya que se demostró que con los datos registrados del motor de la retroexcavadora no es posible realizar un pronóstico en tiempo real utilizando algoritmos de *Machine Learning*, ya que no se permite identificar por completo el comportamiento de la retroexcavadora. Para solucionar el problema de los datos, hay que enfocarse en la sensorización del motor, pues deben ser medidos una mayor cantidad de parámetros, ya sea de fuerza o torque, por ejemplo, para que se tenga un mayor entendimiento del comportamiento de la máquina, siendo analizar si las grandes variaciones en las vibraciones son debido al desgaste de los componentes o de alguna situación en particular que afecta la operación de la retroexcavadora.

Referencias

- [1] D. Kwon, M. R. Hodkiewicz, J. Fan, T. Shibusani, and M. G. Pecht, "IoT-Based Prognostics and Systems HealthManagement for Industrial Applications," *IEEE ACCESS*, 2016, doi: 10.1109/ACCESS.2016.2587754.
- [2] A. R. Sayadi, A. Lashgari, M. M. Fouladgar, and M. J. Skibniewski, "Estimating capital and operational costs of backhoe shovels," *Journal of Civil Engineering and Management*, vol. 18, no. 3, pp. 378–385, Jun. 2012, doi: 10.3846/13923730.2012.692705.
- [3] Svetozar Madzhov, "Research on the Reliability Level of Backhoe Loaders' Diesel Engines," *Journal of Environmental Science and Engineering A*, vol. 8, no. 4, Apr. 2019, doi: 10.17265/2162-5298/2019.04.004.
- [4] Q. Fan and H. Fan, "Reliability Analysis and Failure Prediction of Construction Equipment with Time Series Models," *Journal of Advanced Management Science*, pp. 203–210, 2015, doi: 10.12720/joams.3.3.203-210.
- [5] B. Ellis, P. Stephan Heyns, and S. Schmidt, "Diagnosis and Prognosis of Mechanical Components Using Hybrid Methods," in *Applied Condition Monitoring*, vol. 20, Springer Science and Business Media Deutschland GmbH, 2022, pp. 106–115. doi: 10.1007/978-3-030-85584-0_11.
- [6] P. Aqueveque, L. Radrigan, F. Pastene, A. S. Morales, and E. Guerra, "Data-Driven Condition Monitoring of Mining Mobile Machinery in Non-Stationary Operations Using Wireless Accelerometer Sensor Modules," *IEEE Access*, vol. 9, pp. 17365–17381, 2021, doi: 10.1109/ACCESS.2021.3051583.
- [7] C. R. Gagg and P. R. Lewis, "Wear as a product failure mechanism - Overview and case studies," *Eng Fail Anal*, vol. 14, no. 8 SPEC. ISS., pp. 1618–1640, 2007, doi: 10.1016/j.engfailanal.2006.11.064.
- [8] P. Poor, J. Basl, and D. Zenisek, "Predictive Maintenance 4.0 as next evolution step in industrial maintenance development; Predictive Maintenance 4.0 as next evolution step in industrial maintenance development," 2019.
- [9] M. Jasiulewicz-Kaczmarek, S. Legutko, and P. Kluk, "Maintenance 4.0 technologies - new opportunities for sustainability driven maintenance," *Management and Production Engineering Review*, vol. 11, no. 2, pp. 74–87, Jun. 2020, doi: 10.24425/mper.2020.133730.
- [10] Z. M. Çinar, A. A. Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, and B. Safaei, "Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0," *Sustainability (Switzerland)*, vol. 12, no. 19, Oct. 2020, doi: 10.3390/su12198211.
- [11] "CONCEPTOS DE VIBRACIONES MECÁNICAS."
- [12] Gonzalo Daza Hernández, "VIBRACIONES MECÁNICAS."

- [13] N. Sakib and T. Wuest, “Challenges and Opportunities of Condition-based Predictive Maintenance: A Review,” 2018.
- [14] Kwon D, Hodkiewicz M, Fan J, Shibutani T, and Pecht M, “IoT-Based Prognostics and Systems Health Management for Industrial Applications,” *IEEE Access*, 2016.
- [15] Q. Butler, Y. Ziada, D. Stephenson, and S. Andrew Gadsden, “Condition Monitoring of Machine Tool Feed Drives: A Review,” *J Manuf Sci Eng*, vol. 144, no. 10, Oct. 2022, doi: 10.1115/1.4054516.
- [16] B. Ellis, P. Stephan Heyns, and S. Schmidt, “Diagnosis and Prognosis of Mechanical Components Using Hybrid Methods,” in *Applied Condition Monitoring*, vol. 20, Springer Science and Business Media Deutschland GmbH, 2022, pp. 106–115. doi: 10.1007/978-3-030-85584-0_11.
- [17] A. Bronshtein, “Train/Test Split and Cross Validation in Python,” *Towards Data Science*, May 17, 2017. <https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6>
- [18] “Qué es overfitting y underfitting y cómo solucionarlo,” *Aprende Machine Learning*, Dec. 12, 2017. <https://www.aprendemachinelearning.com/que-es-overfitting-y-underfitting-y-como-solucionarlo/>
- [19] T. Boyle, “Dealing with Imbalanced Data,” *Towards Data Science*, Feb. 03, 2019. <https://towardsdatascience.com/methods-for-dealing-with-imbalanced-data-5b761be45a18>
- [20] “Glossary of Common Terms and API Elements,” *scikit learn*. <https://scikit-learn.org/stable/glossary.html#term-parameter>
- [21] “Pronóstico de series de tiempo mediante deep learning,” *MathWorks*, 2022. <https://la.mathworks.com/help/deeplearning/ug/time-series-forecasting-using-deep-learning.html>
- [22] M. S. Kan, A. C. C. Tan, and J. Mathew, “A review on prognostic techniques for non-stationary and non-linear rotating systems,” *Mechanical Systems and Signal Processing*, vol. 62, Academic Press, pp. 1–20, Oct. 01, 2015. doi: 10.1016/j.ymsp.2015.02.016.
- [23] Y. Peng, M. Dong, and M. J. Zuo, “Current status of machine prognostics in condition-based maintenance: A review,” *International Journal of Advanced Manufacturing Technology*, vol. 50, no. 1–4, pp. 297–313, Sep. 2010, doi: 10.1007/s00170-009-2482-0.
- [24] J. Z. Sikorska, M. Hodkiewicz, and L. Ma, “Prognostic modelling options for remaining useful life estimation by industry,” *Mech Syst Signal Process*, vol. 25, no. 5, pp. 1803–1836, Jul. 2011, doi: 10.1016/j.ymsp.2010.11.018.

- [25] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, “Remaining useful life estimation based on nonlinear feature reduction and support vector regression,” *Eng Appl Artif Intell*, vol. 26, no. 7, pp. 1751–1760, Aug. 2013, doi: 10.1016/j.engappai.2013.02.006.
- [26] M. Fonte, P. Duarte, L. Reis, M. Freitas, and V. Infante, “Failure mode analysis of two crankshafts of a single cylinder diesel engine,” *Eng Fail Anal*, vol. 56, pp. 185–193, Oct. 2015, doi: 10.1016/j.engfailanal.2015.02.014.
- [27] C. Wang and Q. Zhang, “Reliability Analysis of Heavy-Duty Truck Diesel Engine Based on After-Sales Maintenance Data,” *Journal of Failure Analysis and Prevention*, vol. 21, no. 3, pp. 993–1001, Jun. 2021, doi: 10.1007/s11668-021-01145-3.
- [28] B. Kareem, “Mechanical failure analysis of automobile crankshafts under service reconditioned modelling approach,” *Eng Fail Anal*, vol. 80, pp. 87–101, Oct. 2017, doi: 10.1016/j.engfailanal.2017.05.036.
- [29] S. K. Nithin, K. Hemanth, and V. Shamanth, “A review on combustion and vibration condition monitoring of IC engine,” in *Materials Today: Proceedings*, 2020, vol. 45, pp. 65–70. doi: 10.1016/j.matpr.2020.10.093.
- [30] A. Moosavian, G. Najafi, B. Ghobadian, M. Mirsalim, S. M. Jafari, and P. Sharghi, “Piston scuffing fault and its identification in an IC engine by vibration analysis,” *Applied Acoustics*, vol. 102, pp. 40–48, Jan. 2016, doi: 10.1016/j.apacoust.2015.09.002.
- [31] U. DE Concepcion and P. Nelson Saavedra, “LABORATORIO DE VIBRACIONES MECANICAS DEPARTAMENTO DE INGENIERIA MECANICA FACULTAD DE INGENIERIA TUTORIAL SEVERIDAD VIBRATORIA. PARTE I.” [Online]. Available: www.dim.udec.cl/lvmUniversidaddeConcepción.CHILEhttp://www.dim.udec.cl/lvm
- [32] A. Saini, “An Introduction to Random Forest Algorithm for beginners,” *Analytics Vidhya*, Oct. 19, 2021. <https://www.analyticsvidhya.com/blog/2021/10/an-introduction-to-random-forest-algorithm-for-beginners/>
- [33] Guruchan M, “Machine Learning Basics: Support Vector Machine (SVM) Classification,” *Towards Data Science*, Aug. 03, 2020. <https://towardsdatascience.com/machine-learning-basics-support-vector-machine-svm-classification-205ecd28a09d>
- [34] “Naive Bayes,” *scikit learn*. https://scikit-learn.org/stable/modules/naive_bayes.html#gaussian-naive-bayes
- [35] Saini A, “Decision Tree Algorithm – A Complete Guide,” *Analytics Vidhya*, Aug. 29, 2021. <https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/>
- [36] Aliyev V, “Gradient Boosting Classification explained through Python,” *Towards Data Science*, Sep. 05, 2020. <https://towardsdatascience.com/gradient-boosting-classification-explained-through-python-60cc980eeb3d>

- [37] I. Chelliah, “Bagging Decision Trees — Clearly Explained,” *Towards Data Science*, May 07, 2021. <https://towardsdatascience.com/bagging-decision-trees-clearly-explained-57d4d19ed2d3>
- [38] Sarang Narkhede, “Understanding Confusion Matrix,” May 09, 2018. <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>
- [39] “Descripción de relaciones matemáticas y realización de predicciones a partir de datos experimentales,” *MathWorks*, 2022. <https://la.mathworks.com/help/deeplearning/ug/time-series-forecasting-using-deep-learning.html>
- [40] S. Bhattacharyya, “Ridge and Lasso Regression: L1 and L2 Regularization,” *Towards Data Science*, Sep. 26, 2018. <https://towardsdatascience.com/ridge-and-lasso-regression-a-complete-guide-with-python-scikit-learn-e20e34bcbf0b>
- [41] T. Sharp, “An Introduction to Support Vector Regression (SVR),” *Towards Data Science*, Mar. 03, 2020. <https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2>
- [42] T. Masui, “All You Need to Know about Gradient Boosting Algorithm – Part 1. Regression,” *Towards Data Science*, Jan. 10, 2022. <https://towardsdatascience.com/all-you-need-to-know-about-gradient-boosting-algorithm-part-1-regression-2520a34a502>
- [43] A. Prasad, “Regression Trees | Decision Tree for Regression | Machine Learning,” *Medium*, Aug. 08, 2021. <https://medium.com/analytics-vidhya/regression-trees-decision-tree-for-regression-machine-learning-e4d7525d8047>
- [44] N. Khandelwal, “A Brief Introduction to XGBoost,” *Medium*, Jul. 2020. <https://towardsdatascience.com/a-brief-introduction-to-xgboost-3eae2e3e5d6#:~:text=XGBoost%20vs%20Gradient%20Boosting,can%20be%20parallelized%20across%20clusters.>