



**TRABAJO DE GRADO
Opción Seminario-Diplomado.**

**ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE DECISIONES EN
EL MANTENIMIENTO PREDICTIVO DE MAQUINARIA INDUSTRIAL Y AGRICOLA,
UTILIZANDO ESTRATEGIAS DE MACHINE LEARNING**

CORPORACIÓN UNIVERSITARIA REMINGTON.

FACULTAD DE INGENIERÍA

INGENIERÍA INDUSTRIAL

ESTUDIANTES:

ANYI ZULEIMA LÓPEZ DOMÍNGUEZ

TUTOR: JUAN CARLOS BRIÑEZ DE LEÓN

OPCIÓN DE TRABAJO DE GRADO SEMINARIO-DIPLOMADO.

2024.

Dedicatoria

A mi querida familia, por su incondicional cariño y apoyo constante. Su paciencia, comprensión y sacrificio han sido mi mayor fuente de motivación y de gran fortaleza.

A mis abuelos, por creer en mí y por enseñarme el valor del esfuerzo y la perseverancia.

A mi esposo por ser mi compañero de vida y un aliento para seguir este camino académico.

A todos quienes que me han apoyado durante mi pregrado en Ingeniería Industrial, incluyendo amigos, profesores y compañeros. Su ayuda, consejos y palabras de aliento han sido fundamentales en mi formación académica y personal.

Este logro es tanto mío como de ustedes. ¡Gracias!

Agradecimientos

A mis profesores, por compartir su conocimiento, su tiempo y su pasión por la enseñanza. Gracias por guiarme y por inspirarme a ser la mejor versión de mí mismo.

A mis amigos y compañeros de clase, por su camaradería, su apoyo y por hacer de esta experiencia una etapa memorable y enriquecedora.

A todos aquellos que de alguna manera contribuyeron a mi formación durante mi investigación y desarrollo Académico", mis más sinceros agradecimientos. Su ayuda, consejos y palabras de aliento han sido fundamentales en mi camino.

Finalmente, agradezco a mi institución educativa por brindarme las herramientas y el entorno necesario para desarrollarme académica y profesionalmente.

Contenido

Dedicatoria	2
Agradecimientos.....	3
Tabla de Figuras.....	5
Resumen	6
Abstract.....	6
Palabras clave	7
Marco conceptual y contextual	8
Introducción.....	8
Mantenimiento Predictivo: Definición y Relevancia.....	8
Necesidad en la agricultura e industria	9
Importancia del Análisis Exploratorio de Datos (EDA).....	10
Preparación de los Datos.....	10
Selección y Evaluación de Modelos de Machine Learning.....	11
Aplicaciones del Mantenimiento Predictivo en Industria y agricultura.....	12
Desafíos en la Implementación del Mantenimiento Predictivo	13
Futuro del Mantenimiento Predictivo en la Agricultura	14
Pregunta problema:	15
Acercamiento a los datos.....	15
Aproximación a los datos con gráficos - analítica.....	20
Objetivos:.....	22
Objetivo general.	22
Objetivos específicos.....	22
Análisis de probabilidad de falla (Regresión logística)	23
Preparación de los datos (Regresión)	23
Definición de Variables y División del Conjunto de Datos.....	24
Normalización de los Datos.....	25
Modelo de toma de decisiones	25
Construcción del Modelo de Regresión.....	25
Validación del modelo	26
Uso del Modelo Acceso usuario con nuevos datos.....	27
Conclusiones.....	30
Análisis de probabilidad de falla (Redes Neuronales)	32
Preparación de los datos (Redes Neuronales)	33
Carga y Exploración del Conjunto de Datos.....	34
Exploración Inicial:	34
Transformación de Variables	34
División del Conjunto de Datos.....	35
Modelo de Redes Neuronales.....	36
Definición y Configuración del Modelo	36
Entrenamiento del Modelo	37
Validación del Modelo	37
Evaluación del Modelo en los Datos de Prueba	37

	5
Conclusiones Redes Neuronales	42
Conclusión.....	43
Referencias	45
Anexos	45
Documento del EDA	45
Script Modelo de regresión logística.....	45
Script modelo de redes neuronales	46

Tabla de Figuras

0-I PREDICTIVE-MAINTENANCE-CLASSIFICATION.....	16
0-II DESCRIPTION DE VARIABLES	18
0-III DISTRIBUTION DE VARIABLES TEMPETATURA.....	20
0-IV TORQUE VS RPM.....	20
0-V VARIABLE TIPOS DE FALLA.....	21
0-I TABLA DE PREPARACION DE DATOS	24
0-II DEFINICIÓN Y SUBCONJUNTOS.....	24
0-III NORMALIZACIÓN DE DATOS.....	25
0-IV REGRESIÓN LINEAL.....	26
0-V VALIDACIÓN DEL MODELO	27
0-VI IMPORTACIÓN DEL MODELO	28
0-VII IMPORTACIÓN DEL MODELO	29
0-VIII VALIDACIÓN DE DATOS Y REPORTE	30
0-IX RESULTADOS CONCLUYENTES	30

Resumen

Este proyecto se centra en el desarrollo de un algoritmo computacional avanzado para mejorar el mantenimiento predictivo de maquinaria industrial y agrícola utilizando técnicas de machine learning. El objetivo principal es identificar condiciones de fallo potenciales en equipos antes de que ocurran, optimizando así los procesos de mantenimiento y reduciendo costos operativos.

El algoritmo propuesto integra estrategias de machine learning, incluyendo modelos de regresión logística y redes neuronales, para analizar datos operativos en tiempo real. Se han utilizado conjuntos de datos detallados que incluyen variables como temperatura, velocidad de rotación, torque y desgaste de herramientas, entre otros, para entrenar y validar los modelos.

Abstract

This project focuses on developing an advanced computational algorithm to enhance predictive maintenance in industrial and agricultural machinery using machine learning techniques. The main objective is to identify potential failure conditions in equipment before they occur, thereby optimizing maintenance processes and reducing operational costs.

The proposed algorithm integrates machine learning strategies, including logistic regression models and neural networks, to analyze real-time operational

data. Detailed datasets have been used, including variables such as temperature, rotational speed, torque, and tool wear, to train and validate the models.

Palabras clave

Machine Learning, Mantenimiento Predictivo, Regresión Logística, Redes Neuronales, Preprocesamiento de Datos, Clasificación de Fallas, Dataset, Feature Engineering, Entrenamiento del Modelo, Evaluación del Modelo, Validación Cruzada, Predicciones, Estrategias de Mantenimiento, Optimización de Recursos, Tiempo de Inactividad, Redes Neuronales, Algoritmo de Clasificación, Datos Sintéticos, Pipeline de Machine Learning, Normalización, Escalamiento, One-Hot Encoding, Análisis Exploratorio de Datos.

Marco conceptual y contextual

Introducción

La agricultura e industrial en general en la actualidad se enfrentan a varios desafíos críticos que incluyen la necesidad de maximizar la eficiencia operativa y productiva y minimizar los costos de operación, producción y mantenimiento. Un aspecto fundamental que influye significativamente en estos factores es el mantenimiento de la maquinaria agrícola e industrial. Las fallas imprevistas de maquinaria pueden resultar en interrupciones significativas de la producción y diferentes procesos, llevando a pérdidas financieras y afectación de KPIs considerables. Para abordar estos desafíos, el mantenimiento predictivo puede convertirse en una estrategia indispensable en las operaciones agrícolas e industriales.

Mantenimiento Predictivo: Definición y Relevancia

El mantenimiento predictivo es una estrategia moderna la cual utiliza el análisis de datos y aprendizaje de maquina llamado Machine Learning para predecir en qué momento se requiere mantenimiento en un equipo. Este enfoque nos ayuda a realizar el mantenimiento justo a tiempo para prevenir fallas inesperadas y prolongar la vida útil de los equipos. A diferencia de un mantenimiento reactivo, que se lleva a cabo después ocurrir una falla, y del mantenimiento preventivo, que se realiza regularmente sin considerar el estado real del equipo, el mantenimiento

predictivo optimiza los tiempos de mantenimiento basándose en el estado actual, datos obtenidos y el rendimiento del equipo.

Beneficios del análisis predictivo

- Disponibilidad: Al predecir cuándo es probable que ocurra una falla, se puede programar el mantenimiento de manera proactiva, minimizando las interrupciones en la operación.
- Durabilidad: El mantenimiento oportuno puede prevenir daños mayores y extender la vida útil de la maquinaria.
- Costo: Al realizar mantenimiento solo cuando es necesario, se reducen los costos asociados con el mantenimiento preventivo innecesario y las reparaciones de emergencia.
- Seguridad: La identificación temprana de problemas potenciales puede prevenir accidentes y mejorar la seguridad operativa.

Necesidad en la agricultura e industria

En el ámbito de industria y agrícola, el uso de maquinaria avanzada y tecnología de sensores ha abierto nuevas posibilidades para el mantenimiento con análisis predictivo. La maquinaria agrícola e industrial en general moderna está equipada con sensores que monitorean continuamente diversas variables críticas, como la temperatura, la vibración y las horas de operación, así como la implementación de procesos BPM que contribuyen a una adquisición completa de

datos. Estos datos proporcionan una visión detallada del estado y rendimiento del equipo, lo que es fundamental para desarrollar modelos predictivos precisos.

El análisis de datos es una herramienta esencial en el mantenimiento predictivo. Los datos recopilados por los sensores y seguimiento propios de la operación junto con bitácoras de uso pueden ser utilizados para identificar patrones y relaciones que no son evidentes a simple vista. Este proceso comienza con el Análisis Exploratorio de Datos (EDA), que ayuda a entender la estructura y características de los datos e información contenida e identificar tendencias y detectar anomalías.

Importancia del Análisis Exploratorio de Datos (EDA)

El EDA es el primer paso en cualquier proyecto de análisis de datos. Permite obtener una comprensión inicial de los datos, identificar relaciones entre variables y formular hipótesis iniciales. En el contexto del mantenimiento predictivo, el EDA puede revelar relaciones entre variables operativas (como temperatura, vibración, RPM, Desgaste) y la probabilidad de que se necesite mantenimiento. Este conocimiento inicial es crucial para la posterior construcción de modelos predictivos.

Preparación de los Datos

El preprocesamiento de datos es la parte más importante para asegurar que los datos sean de alta calidad y adecuados para el análisis. Este proceso incluye

la limpieza de datos, manejo de valores nulos, normalización y escalado de variables, y la transformación de variables según sea necesario. La calidad de los datos tiene un impacto alto en el desempeño de los modelos de Machine Learning.

Selección y Evaluación de Modelos de Machine Learning

Para el desarrollo de modelos predictivos, se pueden emplear diversos algoritmos de Machine Learning. Cada algoritmo tiene sus propias fortalezas y es adecuado para diferentes tipos de problemas y conjuntos de datos. Algunos de los algoritmos comúnmente utilizados en el mantenimiento predictivo incluyen:

- **Regresión:** Adecuada para problemas de clasificación binaria, como predecir la necesidad de mantenimiento.
- **Árboles de decisión:** Algoritmos basados en árboles que pueden manejar relaciones no lineales entre variables y proporcionar interpretabilidad.
- **Support Vector Machines (SVM):** Efectivas para problemas de clasificación con datos complejos y estructuras no lineales.
- **Redes neuronales:** Capaces de capturar patrones complejos en grandes volúmenes de datos.

La evaluación de los modelos es una etapa crucial para determinar cuál algoritmo ofrece el mejor desempeño. Las métricas y/o KPIs comunes para evaluar modelos predictivos incluyen precisión, recall, F1-score y ROC-AUC. Estas métricas permiten evaluar la capacidad del modelo para predecir

correctamente la necesidad de mantenimiento y balancear entre falsos positivos y falsos negativos.

Aplicaciones del Mantenimiento Predictivo en Industria y agricultura

La implementación de un sistema de mantenimiento predictivo en la agricultura puede transformar la manera en que se gestiona la maquinaria. Los pasos principales para la implementación incluyen:

- **Instalación de Sensores:** Equipar la maquinaria con sensores para monitorear variables operativas críticas.
- **Recolección y Almacenamiento de Datos:** Establecer sistemas para la recolección continua y almacenamiento seguro de los datos esto puede ser complemento de la implementación de modelos de gestión de proyectos que implique recolección de métricas para el análisis de calidad.
- **Modelo Predictivo:** Utilizar técnicas de Machine Learning para analizar los datos históricos y desarrollar un modelo predictivo con todos los parámetros claros y apoyo de áreas técnicas que contribuyan a una construcción completa de las necesidades.
- **Integración del Sistema:** Integrar el modelo predictivo con el sistema de gestión de mantenimiento para alertar automáticamente cuando se necesita mantenimiento.

- **Monitoreo y Mejora Continua:** Evaluar continuamente el desempeño del sistema y realizar mejoras basadas en nuevos datos y evaluar según la retroalimentación operacional.

Desafíos en la Implementación del Mantenimiento Predictivo

A pesar de los numerosos beneficios, la implementación del mantenimiento predictivo en la industria y agricultura presenta varios desafíos:

- **Calidad y Complejidad de los Datos:** Los datos de sensores pueden ser ruidosos o incompletos, lo que afecta la precisión de los modelos predictivos. Además, manejar grandes volúmenes de datos requiere infraestructura adecuada y técnicas avanzadas de análisis y recolección, evitar información irrelevante y evaluar constantemente estas necesidades.
- **Costo de Implementación:** La instalación de sensores y la infraestructura necesaria para el análisis de datos pueden ser costosas, lo que puede ser un obstáculo para operaciones industriales y agrícolas más pequeñas se requiere que sea escalable. (engeman.com, 2023)
- **Adaptabilidad y Actualización del Modelo:** Los modelos predictivos deben ser adaptables y capaces de actualizarse con nuevos datos y condiciones operativas que fluctúen, así como adaptarse a diferentes clases de industria y generalizar según el insumo de datos.
- **Interpretabilidad y Confianza en el Modelo:** Es crucial que los modelos sean interpretables para que los ingenieros y operadores puedan entender y

confiar en las predicciones, facilitando la toma de decisiones informadas por lo que el apoyo del área operativa es indispensable.

Futuro del Mantenimiento Predictivo en la Agricultura

El mantenimiento predictivo en la industria está evolucionando rápidamente con los avances tecnológicos. Algunas tendencias emergentes incluyen:

- Internet de las Cosas: La conectividad de dispositivos IoT permite una recolección de datos en tiempo real y una integración más completa de sistemas y procesos.
- Inteligencia Artificial y Aprendizaje Automático: El uso de técnicas avanzadas de inteligencia artificial para mejorar la precisión y eficiencia de los modelos predictivos.
- Big Data y Análisis Predictivo: El análisis de grandes volúmenes de datos permite identificar patrones y tendencias que no serían visibles en conjuntos de datos más pequeños.
- Automatización: La integración de sistemas predictivos con robótica y automatización para realizar mantenimiento sin o mínima intervención humana, aumentando la eficiencia y reduciendo errores humanos y mejorando procesos.

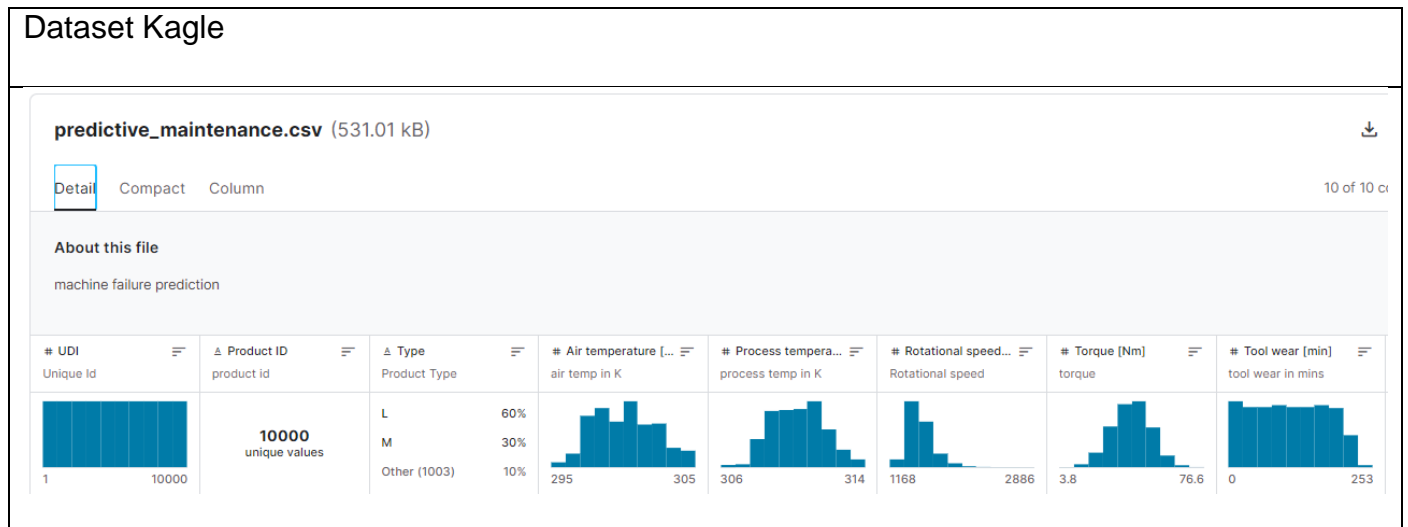
Pregunta problema:

¿Cómo pueden las técnicas de Machine Learning aplicadas al análisis de datos operativos de maquinaria industrial y agrícola mejorar la precisión y efectividad del mantenimiento predictivo, reduciendo los tiempos de inactividad y optimizando los costos operativos?

Acercamiento a los datos

Los datos utilizados en este proyecto provienen de un conjunto sintético diseñado para simular situaciones reales de mantenimiento predictivo en la industria y agricultura. Fueron recopilados y publicados en Kaggle por el usuario "shivamb". Este conjunto de datos específico, titulado "Machine Predictive Maintenance Classification Dataset", ofrece 10,000 puntos de datos con 14 características cada uno, que incluyen variables como temperatura del aire y del proceso, velocidad de rotación, torque, desgaste de herramientas, y etiquetas que indican la ocurrencia de fallas y su tipo. Este enfoque sintético permite explorar cómo las técnicas de Machine Learning pueden mejorar la precisión del

mantenimiento predictivo y reducir los costos operativos asociados mediante el análisis avanzado de datos operativos de maquinaria. (BANSAL, 2021)



0-i predictive-maintenance-classification

Descripción de variables.

Las variables en el conjunto de datos de mantenimiento predictivo de maquinaria industrial y agrícola incluyen diversas características operativas y de rendimiento de las máquinas. A continuación, una descripción de las principales variables presentes en los datos:

1. UID (Unique ID):
 - Identificador único para cada punto de datos en el conjunto.
2. Product ID:
 - Identificador del producto, indicando la calidad del producto (L, M, H) y un número de serie específico.

3. Type:

- Tipo de producto, categorizado como L (baja calidad), M (media calidad) o H (alta calidad).

4. Air temperature [K]:

- Temperatura del aire en Kelvin, generada mediante un proceso de caminata aleatoria y normalizada a una desviación estándar de 2 K alrededor de 300 K.

5. Process temperature [K]:

- Temperatura del proceso en Kelvin, generada mediante un proceso de caminata aleatoria normalizada a una desviación estándar de 1 K, añadida a la temperatura del aire más 10 K.

6. Rotational speed [rpm]:

- Velocidad de rotación en revoluciones por minuto (rpm), calculada a partir de la potencia de 2860 W y superpuesta con un ruido distribuido normalmente.

7. Torque [Nm]:

- Valor del par en Newton-metros (Nm), distribuido normalmente alrededor de 40 Nm con una desviación estándar de 10 Nm y sin valores negativos.

8. Tool wear [min]:

- Desgaste de la herramienta en minutos, afectado por la calidad del producto (H/M/L), donde los productos de alta calidad agregan 5

minutos, los de media calidad 3 minutos, y los de baja calidad 2 minutos.

9. Target:

- Variable objetivo que indica si ocurrió una falla en la máquina en ese punto de datos (0 = No Falla, 1 = Falla).

10. Failure Type:

- Tipo de falla si la máquina falló (por ejemplo, Falla en la Disipación de Calor u otro tipo de falla).

INFORMACIÓN DATASET

```

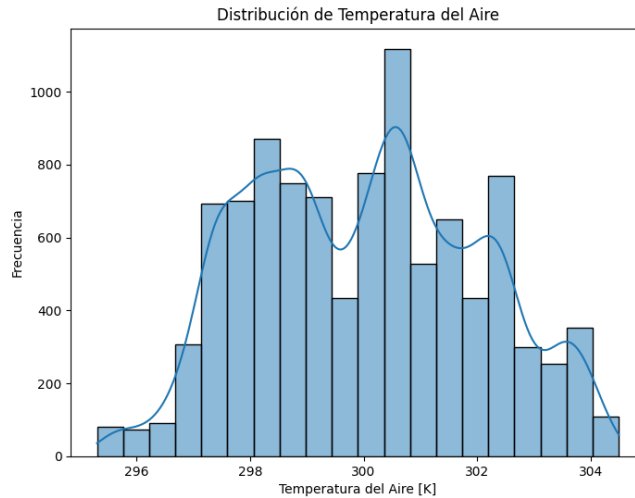
UDI Product ID Type Air temperature [K] Process temperature [K] \
0 1 M14860 M 298.1 308.6
1 2 L47181 L 298.2 308.7
2 3 L47182 L 298.1 308.5
3 4 L47183 L 298.2 308.6
4 5 L47184 L 298.2 308.7

Rotational speed [rpm] Torque [Nm] Tool wear [min] Target Failure Type
0 1551 42.8 0 0 No Failure
1 1408 46.3 3 0 No Failure
2 1498 49.4 5 0 No Failure
3 1433 39.5 7 0 No Failure
4 1408 40.0 9 0 No Failure
Index(['UDI', 'Product ID', 'Type', 'Air temperature [K]',
      'Process temperature [K]', 'Rotational speed [rpm]', 'Torque [Nm]',
      'Tool wear [min]', 'Target', 'Failure Type'],
      dtype='object')
UDI int64
Product ID object
Type object
Air temperature [K] float64
Process temperature [K] float64
Rotational speed [rpm] int64
Torque [Nm] float64
Tool wear [min] int64
Target int64
Failure Type object
dtype: object

```

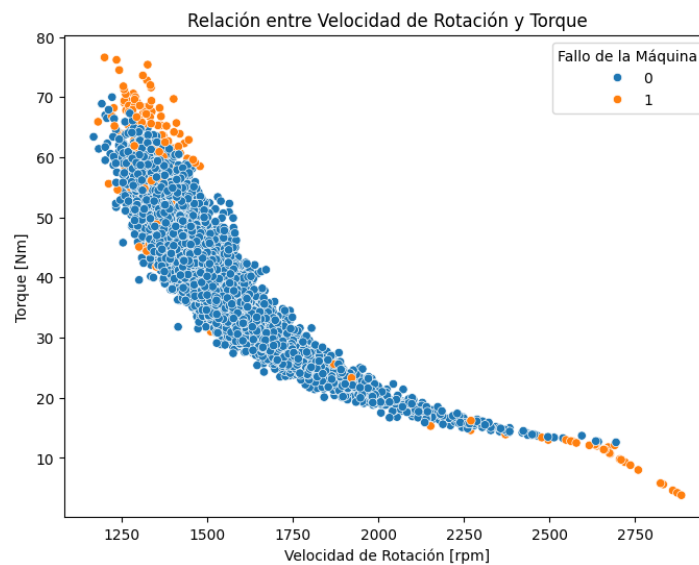
Variable	Descripción	Tipo
UID	Identificador único para cada punto de datos	Entero
Product ID	Identificador del producto, incluye una letra indicando la calidad (L, M, H) y un número de serie	Categórico
Type	Tipo de producto, categorizado como L (baja calidad), M (media calidad) o H (alta calidad)	Categórico
Air temperature [K]	Temperatura del aire en Kelvin	Numérico
Process temperature [K]	Temperatura del proceso en Kelvin	Numérico
Rotational speed [rpm]	Velocidad de rotación en revoluciones por minuto	Numérico
Torque [Nm]	Valor del par en Newton-metros	Numérico
Tool wear [min]	Desgaste de la herramienta en minutos	Numérico
Target	Indica si hubo o no una falla (0 = No Falla, 1 = Falla)	Binario
Failure Type	Tipo de falla si la máquina falló	Categórico

Aproximación a los datos con gráficos - analítica.



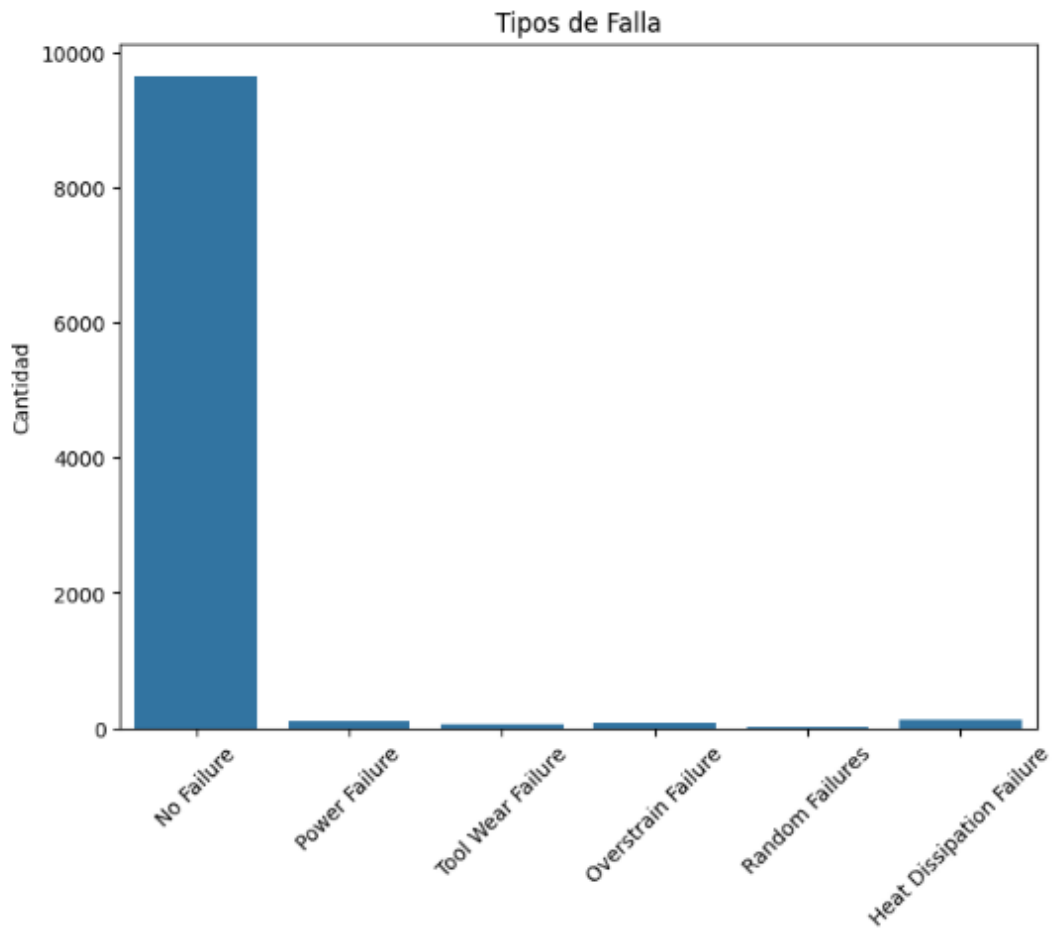
0-iii Distribution de variables tempetatura

Esta gráfica muestra la distribución de la temperatura del aire en el conjunto de datos. Al representar la frecuencia de los valores de temperatura, permite identificar la variabilidad y la tendencia central de esta variable.



0-iv Torque vs RPM

Un gráfico de dispersión que visualiza la relación entre la velocidad de rotación (rpm) y el torque (Nm), diferenciando entre fallos y no fallos de la máquina. Se identifica patrones y posibles correlaciones entre la velocidad de rotación y el torque, y cómo estas variables pueden influir en la probabilidad de fallos.



0-v Variable Tipos de falla

Un gráfico de barras que muestra la cantidad de ocurrencias para cada tipo de falla en el conjunto de datos. Visualiza la frecuencia de los diferentes tipos de

fallas para entender cuáles son más comunes y planificar estrategias de mantenimiento adecuadas.

Objetivos:

Objetivo general.

Desarrollar un algoritmo computacional que utilice técnicas de Machine Learning para analizar y tomar decisiones en el mantenimiento predictivo de maquinaria industrial y agrícola, con el fin de mejorar la precisión y efectividad, reducir los tiempos de inactividad y optimizar los costos operativos.

Objetivos específicos.

- Integrar y preprocesar datos operativos de maquinaria, asegurando la calidad y consistencia de los datos.
- Realizar un análisis exploratorio de los datos para identificar patrones y relaciones significativas entre las variables.
- Implementar y evaluar modelos de clasificación para predecir la probabilidad de fallas en la maquinaria.
- Desarrollar modelos de regresión para predecir valores continuos, como el desgaste de las herramientas y el tiempo hasta la próxima falla.
- Validar los modelos desarrollados utilizando técnicas como la validación cruzada y ajustar los hiperparámetros para mejorar su rendimiento.
- Integrar los modelos predictivos en un sistema de toma de decisiones que ayude a planificar el mantenimiento preventivo de manera más eficiente.

Análisis de probabilidad de falla (Regresión logística)

Se analiza mediante el primer algoritmo cuantos minutos de uso soporta la maquinaria antes de fallar mediante la técnica de regresión este modelo podrá predecir el desgaste y optimizar cuando implementar el reemplazo o evaluar su presupuesto con antelación.

Preparación de los datos (Regresión)

Inicialmente debemos cargar el conjunto de datos, verificar sus columnas, tipos de datos y convertir las variables categóricas en variables dummy para su uso en el modelo. Esta preparación permite identificar los datos relevantes para el modelo y limpiar para su uso optimizado. (Buitrago, 2020)

Importación y validación	Resultados
<pre> # Importar librerías necesarias import pandas as pd from sklearn.model_selection import train_test_split from sklearn.preprocessing import StandardScaler, OneHotEncoder from sklearn.compose import ColumnTransformer from sklearn.pipeline import Pipeline from sklearn.linear_model import LogisticRegression from sklearn.metrics import classification_report # Paso 1: Cargar el conjunto de datos file_path = 'synthetic_data.csv' data = pd.read_csv(file_path) </pre>	<pre> <bound method DataFrame.info of 0 1 P1 L 297.3 308.1 1 2 P2 M 298.3 309.2 2 3 P3 M 298.3 309.1 3 4 P4 M 298.8 309.6 4 5 P5 L 297.6 308.1 ... 14995 14996 P14996 M 298.3 309.6 14996 14997 P14997 L 297.3 308.7 14997 14998 P14998 M 298.9 309.5 14998 14999 P14999 M 298.2 309.2 14999 15000 P15000 L 298.5 308.6 Rotational speed [rpm] Torque [Nm] Tool wear [min] Target \ 0 1428 37.1 31 0 1 1052 34.0 4 0 2 1436 23.4 12 0 3 1968 26.0 39 1 4 1798 25.5 6 0 ... 14995 1343 45.3 45 0 14996 1066 22.3 24 0 14997 1972 34.7 5 1 14998 1055 27.1 21 0 14999 1031 36.0 30 1 </pre>
<p>Carga de dataset Validación de datos Verificación de tipos</p>	<p>El conjunto de datos incluye identificadores únicos (UDI), el ID del producto, y el tipo de producto (L, M, H).</p>

Validación de nulos Conversión de variables categóricas a dummy	Además, se registran mediciones como la temperatura del aire y del proceso (en Kelvin), la velocidad de rotación (en rpm), el torque (en Nm), y el desgaste de la herramienta (en minutos)
--	--

0-i Tabla de Preparacion de datos

El siguiente paso implica utilizar estos datos preparados para entrenar y validar un modelo de regresión, con el fin de predecir el desgaste de la herramienta y optimizar el mantenimiento predictivo.

Definición de Variables y División del Conjunto de Datos

Definimos la variable objetivo (en este caso, podemos usar Tool wear [min] para la regresión) y las variables independientes. Luego, dividimos el conjunto de datos en conjuntos de entrenamiento y prueba. (datos.gob.es, 2023)

Definición y División

```
# Paso 2: Preprocesamiento de datos
# - Dividir en características y variable objetivo
X = data.drop(['UDI', 'Target', 'Failure Type'], axis=1)
y = data['Target']

# - Definir características numéricas y categóricas
numeric_features = ['Air temperature [K]', 'Process temperature [K]', 'Rotational speed [rpm]', 'Torque [Nm]']
categorical_features = ['Product ID', 'Type']

# - Preprocesamiento con ColumnTransformer y OneHotEncoder
preprocessor = ColumnTransformer(
    transformers=[
        ('num', StandardScaler(), numeric_features),
        ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_features)])
✓ 0.0s
```

Definiendo la variable objetivo y dividimos los datos en subconjunto para las respectivas validaciones dentro del mismo dataset

0-ii Definición y subconjuntos

Normalización de los Datos

Normalizamos los datos para mejorar el rendimiento del modelo de regresión para eso usaremos sklearn su función Pipeline

El escalado de datos es una técnica de preprocesamiento que ajusta los valores de las características (features) para que tengan una media (mean) de 0 y una desviación estándar (standard deviation) de 1. (4geeks.com, 2022)

Con esto se busca también que todas las características contribuyan de manera equitativa en el modelo y con esto terminaríamos la preparación de los datos para el caso de validación de falla probables en la maquinaria industrial y agrícola.

```
Uso de StandardScaler

# Paso 3: Construir el pipeline con el modelo de Regresión Logística
pipeline_lr = Pipeline(steps=[('preprocessor', preprocessor),
                              ('classifier', LogisticRegression(random_state=42))])
✓ 0.0s
```

0-iii Normalización de datos

Modelo de toma de decisiones

Construcción del Modelo de Regresión

Se utilizará concretamente un modelo de regresión lineal de “sklearn” librería que nos permitirá validar la variable con los datos previamente tratados y divididos en subconjuntos para su posterior validación.

Regresión logística

```
# Paso 4: División en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

✓ 0.0s

```
# Paso 5: Entrenamiento del modelo
pipeline_lr.fit(X_train, y_train)
```

✓ 0.1s

0-iv Regresión Lineal

Mediante la regresión logística analizaremos el desgaste de la herramienta o maquinaria y se clasificara según su probabilidad de falla, identificando cuales de los registros product ID requieren de mantenimiento.

Validación del modelo

A continuación, apoyados en la librería continuamos con la validación usando los datos de prueba y entrenamiento separados para la validación y entrenamiento, adicional a esto realizamos pruebas con nuevos datos y vemos los valores para entender que tan bien está funcionando el modelo.

Validación del modelo con “mean_squared_error” y “r2_score”

```

y_pred = pipeline_lr.predict(X_test)
print(classification_report(y_test, y_pred))

```

✓ 0.0s

	precision	recall	f1-score	support
0	0.87	0.98	0.92	2208
1	0.91	0.61	0.73	792
accuracy			0.88	3000
macro avg	0.89	0.79	0.83	3000
weighted avg	0.89	0.88	0.87	3000

```

from sklearn.model_selection import cross_val_score
# Validación cruzada para evaluar el rendimiento del modelo
cv_scores = cross_val_score(pipeline_lr, X, y, cv=5)
print("Mean Cross-Validation Accuracy:", cv_scores.mean())

```

✓ 0.7s

Mean Cross-Validation Accuracy: 0.8806666666666667

0-v Validación del modelo

Uso del Modelo Acceso usuario con nuevos datos

Con la ayuda de joblib exportamos el modelo para su uso con nuevos datos e identificación de los mantenimientos requeridos a continuación un paso a paso del uso del modelo para identificar que equipos requieren mantenimiento y validación.

Importación del modelo

```
import joblib

# Guardar el modelo entrenado como archivo pickle
joblib.dump(pipeline_lr, 'modelo_regresion_logistica.pkl')
print("Modelo guardado como 'modelo_regresion_logistica.pkl'.")
```

✓ 0.0s

Modelo guardado como 'modelo_regresion_logistica.pkl'.

```
import pandas as pd
import joblib

# Cargar el modelo guardado
model_path = 'modelo_regresion_logistica.pkl'
pipeline_lr = joblib.load(model_path)
```

✓ 0.0s

0-vi Importación del modelo

Importación del modelo

```
import pandas as pd

# Cargar los datos nuevos desde el archivo CSV
csv_file = 'new_data_to_predict.csv'
df_new = pd.read_csv(csv_file)

print("Datos cargados desde el archivo CSV:")
print(df_new)
```

✓ 0.0s

Datos cargados desde el archivo CSV:

	Product ID	Type	Air temperature [K]	Process temperature [K]	\
0	P1	M	298.0	309.6	
1	P2	M	298.9	309.3	
2	P3	H	298.7	309.0	
3	P4	L	299.0	309.3	
4	P5	M	298.2	308.8	
..	
95	P96	L	298.5	309.4	
96	P97	L	298.5	308.5	
97	P98	L	297.6	309.7	
98	P99	L	297.8	309.0	
99	P100	L	298.6	309.6	

	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]
0	1489	36.4	14
1	1351	40.1	4
2	1089	32.3	50
3	1359	24.5	2
4	1241	35.2	8
..
95	1424	40.7	19
96	1670	44.3	42
97	1177	40.7	50

0-vii Importación del modelo

Luego de la importación del modelo y datos nuevos realizamos el análisis y obtenemos el reporte con las conclusiones y necesidades de mantenimiento.

Validación de datos y reporte

```
import pandas as pd
from sklearn.metrics import classification_report
import datetime

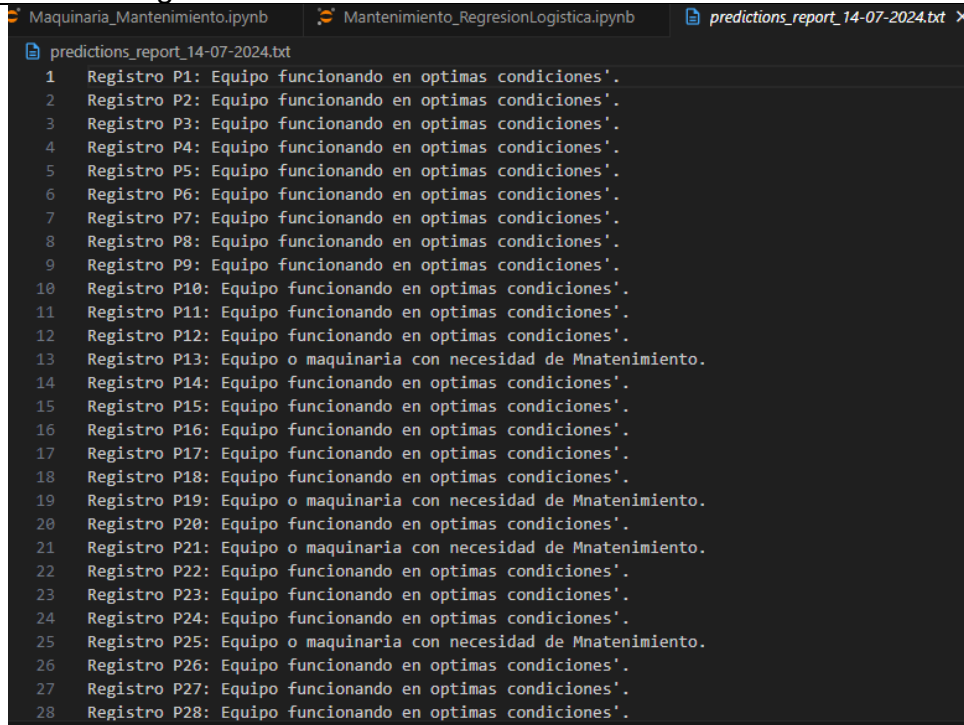
current_date = datetime.datetime.now().strftime('%d-%m-%Y')
# Generar predicciones para los datos nuevos
predictions = pipeline_lr.predict(df_new)

# Mostrar las predicciones y su interpretación
predictions_result = []
for i, (product_id, pred) in enumerate(zip(df_new['Product ID'], predictions)):
    if pred == 0:
        result = f"Registro {product_id}: Equipo funcionando en optimas condiciones'."
    else:
        result = f"Registro {product_id}: Equipo o maquinaria con necesidad de Mntenimiento."
    predictions_result.append(result)

# Guardar las predicciones en un archivo TXT
txt_file = f'predictions_report_{current_date}.txt'
with open(txt_file, 'w') as f:
    for line in predictions_result:
        f.write(line + '\n')
print(f"Reporte guardado exitosamente como '{txt_file}'")
```

0-viii Validación de datos y reporte

Resultados concluyentes claro de toma de decisiones basada en el análisis machine learning



```

Maquinaria_Mantenimiento.ipynb  Mantenimiento_RegresionLogistica.ipynb  predictions_report_14-07-2024.txt X
predictions_report_14-07-2024.txt
1  Registro P1: Equipo funcionando en optimas condiciones'.
2  Registro P2: Equipo funcionando en optimas condiciones'.
3  Registro P3: Equipo funcionando en optimas condiciones'.
4  Registro P4: Equipo funcionando en optimas condiciones'.
5  Registro P5: Equipo funcionando en optimas condiciones'.
6  Registro P6: Equipo funcionando en optimas condiciones'.
7  Registro P7: Equipo funcionando en optimas condiciones'.
8  Registro P8: Equipo funcionando en optimas condiciones'.
9  Registro P9: Equipo funcionando en optimas condiciones'.
10 Registro P10: Equipo funcionando en optimas condiciones'.
11 Registro P11: Equipo funcionando en optimas condiciones'.
12 Registro P12: Equipo funcionando en optimas condiciones'.
13 Registro P13: Equipo o maquinaria con necesidad de Mntenimiento.
14 Registro P14: Equipo funcionando en optimas condiciones'.
15 Registro P15: Equipo funcionando en optimas condiciones'.
16 Registro P16: Equipo funcionando en optimas condiciones'.
17 Registro P17: Equipo funcionando en optimas condiciones'.
18 Registro P18: Equipo funcionando en optimas condiciones'.
19 Registro P19: Equipo o maquinaria con necesidad de Mntenimiento.
20 Registro P20: Equipo funcionando en optimas condiciones'.
21 Registro P21: Equipo o maquinaria con necesidad de Mntenimiento.
22 Registro P22: Equipo funcionando en optimas condiciones'.
23 Registro P23: Equipo funcionando en optimas condiciones'.
24 Registro P24: Equipo funcionando en optimas condiciones'.
25 Registro P25: Equipo o maquinaria con necesidad de Mntenimiento.
26 Registro P26: Equipo funcionando en optimas condiciones'.
27 Registro P27: Equipo funcionando en optimas condiciones'.
28 Registro P28: Equipo funcionando en optimas condiciones'.
  
```

0-ix Resultados concluyentes

Conclusiones

El modelo de regresión logística entrenado mostró un desempeño sólido en la clasificación de equipos según su necesidad de mantenimiento, con una precisión general del 88%. A continuación, se detallan las conclusiones y hallazgos principales:

Se evaluaron un total de 100 registros de equipos, de estos registros, el modelo identificó que:

- El 87% de los equipos fueron clasificados como funcionando en óptimas condiciones (No Failure).

- El 13% de los equipos fueron identificados como necesitando mantenimiento (Failure).

Los registros clasificados como No Failure se consideran funcionando en condiciones normales y no requieren intervención inmediata de mantenimiento, los registros clasificados como Failure indican equipos que presentan condiciones que sugieren la necesidad de mantenimiento preventivo o correctivo.

- Precisión: El modelo logró una precisión del 87% en la clasificación de equipos funcionando correctamente y del 91% en la identificación de equipos que necesitan mantenimiento.
- Recall: La capacidad del modelo para identificar correctamente los equipos que necesitan mantenimiento fue del 61%, lo que sugiere una mejora potencial en la capacidad de detección de estas condiciones.
- F1-Score: El puntaje F1 promedio del modelo fue del 73%, indicando un buen equilibrio entre precisión y recall.

Basado en los resultados obtenidos, se recomienda lo siguiente:

- Priorizar el Mantenimiento Preventivo: Dado que el modelo puede identificar correctamente la mayoría de los casos de Failure, se sugiere implementar un programa de mantenimiento preventivo para los equipos clasificados como Failure para evitar fallos catastróficos.

- Monitoreo Continuo: Es fundamental monitorear continuamente las métricas de desempeño del modelo y ajustar los parámetros según sea necesario para mejorar su precisión y recall.

Análisis de probabilidad de falla (Redes Neuronales)

Exploramos en profundidad el uso de redes neuronales como una herramienta avanzada para predecir la probabilidad de falla en maquinaria industrial y agrícola, este enfoque representa un avance significativo respecto a métodos tradicionales, permitiendo una clasificación más precisa y detallada de los equipos según su estado de operación y necesidades de mantenimiento.

La maquinaria industrial y agrícola opera en entornos dinámicos y exigentes, donde la gestión eficaz del mantenimiento puede significar la diferencia entre la productividad y el tiempo de inactividad costoso. Las técnicas de mantenimiento predictivo han evolucionado considerablemente, y las redes neuronales ofrecen la capacidad de analizar múltiples variables complejas de manera simultánea, permitiendo una evaluación más holística del estado de los equipos.

Utilizaremos un modelo de redes neuronales (Escobar, 2010) diseñado específicamente para la clasificación binaria de equipos en función de su proximidad a la falla. Este modelo se basa en datos históricos detallados, incluyendo mediciones como la temperatura del aire y del proceso, velocidad de rotación, par, desgaste de herramienta, entre otros. Estos datos son cruciales para entrenar el modelo y garantizar una precisión óptima en la predicción de fallas.

El uso de redes neuronales en este contexto no solo mejora la precisión de las predicciones, sino que también permite una optimización proactiva de las estrategias de mantenimiento. Al identificar con mayor exactitud los equipos propensos a fallas, las empresas pueden priorizar y programar intervenciones de mantenimiento preventivo de manera más eficiente, reduciendo costos operativos y aumentando la disponibilidad de la maquinaria.

Preparación de los datos (Redes Neuronales)

En esta sección, detallamos los pasos necesarios para preparar los datos antes de entrenar el modelo de redes neuronales. La calidad y la adecuación de los datos son fundamentales para garantizar la precisión y la efectividad del modelo en la predicción de la probabilidad de falla en maquinaria industrial y agrícola.

Carga y Exploración del Conjunto de Datos

Carga del Conjunto de Datos: inicialmente, cargamos los datos históricos que incluyen variables como la temperatura del aire y del proceso, velocidad de rotación, par, desgaste de herramienta, entre otros.

Cargue el dataset

```
from sklearn.metrics import confusion_matrix, classification_report

# Cargar el dataset
data = pd.read_csv('data/predictive_maintenance.csv')

# Mostrar los primeros registros del dataset cargado
print("Primeros registros del dataset:")
print(data.head())
```

Exploración Inicial:

Realizamos una exploración inicial del conjunto de datos para verificar la estructura, dimensiones y tipos de variables presentes. Esto nos permite comprender la naturaleza de los datos y planificar las transformaciones necesarias.

Transformación de Variables

Conversión de Variables Categóricas: Convertimos las variables categóricas, como el tipo de producto y el identificador único, en variables dummy. Esta transformación asegura que todas las características sean numéricas y adecuadas para el modelado con redes neuronales.

Convertir variables categóricas a variables dummy

```
# Convertir variables categóricas a variables dummy
data = pd.get_dummies(data, columns=['ProductID', 'Type', 'FailureType'], drop_first=True)
```

0-i Convertir variables categóricas a variables dummy

Normalización de Datos: Utilizamos técnicas de normalización, como StandardScaler de sklearn, para escalar todas las características a un rango uniforme. Esto es crucial para mejorar la convergencia y el rendimiento del modelo durante el entrenamiento.

División del Conjunto de Datos

- **Definición de Variables:** Identificamos la variable objetivo (por ejemplo, falla/no falla) y las variables independientes que se utilizarán para entrenar el modelo.
- **División en Conjuntos de Entrenamiento y Prueba:** Dividimos el conjunto de datos en conjuntos de entrenamiento y prueba, utilizando una proporción adecuada (por ejemplo, 80% entrenamiento y 20% prueba). Esta división nos permite evaluar la capacidad del modelo para generalizar a datos no vistos.

Normalización de Datos y División en conjuntos

```
# Convertir variables categóricas a variables dummy
data = pd.get_dummies(data, columns=['ProductID', 'Type', 'FailureType'], drop_first=True)

# Definir la variable objetivo
y = data['Target']

# Definir las variables independientes
X = data.drop('Target', axis=1)

# División en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Inicializar el escalador
scaler = StandardScaler()

# Ajustar y transformar los datos de entrenamiento y prueba
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

0-ii Normalización de Datos y División en conjuntos

Modelo de Redes Neuronales

Definición y Configuración del Modelo

Implementamos un modelo de redes neuronales con capas densas, optimizado para la clasificación binaria basada en datos escalados.



```
Add to chat (Ctrl+I) | Edit highlighted code (Ctrl+I).

# Crear el modelo
model = Sequential()
model.add(Dense(64, input_dim=X_train_scaled.shape[1], activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(1, activation='sigmoid'))

# Compilar el modelo
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])

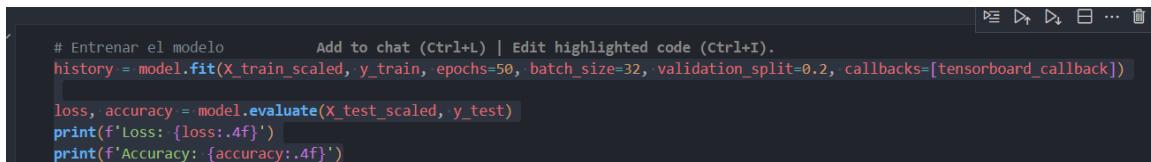
# Configurar TensorBoard
log_dir = "logs/fit/" + datetime.datetime.now().strftime("%Y%m%d-%H%M%S")
tensorboard_callback = TensorBoard(log_dir=log_dir, histogram_freq=1)

# Resumen del modelo
model.summary()
```

0-i Crear Modelo

Entrenamiento del Modelo

Entrenamos el modelo utilizando datos preparados y validamos su rendimiento utilizando técnicas como TensorBoard para monitorizar y mejorar su precisión.



```
# Entrenar el modelo
history = model.fit(X_train_scaled, y_train, epochs=50, batch_size=32, validation_split=0.2, callbacks=[tensorboard_callback])

loss, accuracy = model.evaluate(X_test_scaled, y_test)
print(f'Loss: {loss:.4f}')
print(f'Accuracy: {accuracy:.4f}')
```

0-ii Entrenar el modelo

Validación del Modelo

Para evaluar la efectividad del modelo entrenado, utilizamos métricas estándar como precisión, recall y F1-score. Estas métricas nos proporcionan una comprensión detallada de cómo el modelo está realizando las predicciones sobre la probabilidad de falla en la maquinaria industrial y agrícola.

Evaluación del Modelo en los Datos de Prueba

El modelo entrenado mostró los siguientes resultados en los datos de prueba:

- **Loss:** 0.2178
- **Accuracy:** 0.9770

El valor de **Loss** (0.2178) indica la magnitud del error promedio del modelo en los datos de prueba. Un valor bajo es deseable, lo que sugiere que el modelo tiene un buen desempeño en la tarea de clasificación.

El valor de **Accuracy** (0.9770) representa la precisión general del modelo en la clasificación de las muestras de prueba. Con un 97.70% de precisión, el modelo clasifica correctamente la mayoría de los casos.

```

32/32 — 0s 3ms/step - accuracy: 0.9736 - loss: 0.2544
Loss: 0.2178
Accuracy: 0.9770
32/32 — 0s 3ms/step
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is consid
Confusion Matrix:
[[961  0]
 [ 23 16]]
Classification Report:
              precision    recall  f1-score   support

     0       0.98         1.00         0.99         961
     1       1.00         0.41         0.58          39

 accuracy          0.98         1000
 macro avg         0.99         0.71         0.78         1000
 weighted avg         0.98         0.98         0.97         1000

```

0-iii Evaluación de Modelo

Para analizar la pérdida y la precisión del modelo durante el entrenamiento y la validación utilizamos la siguiente graficas

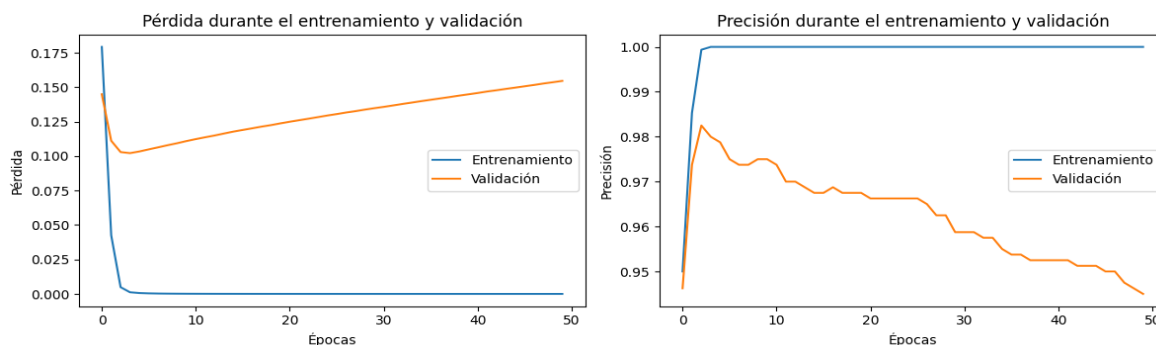
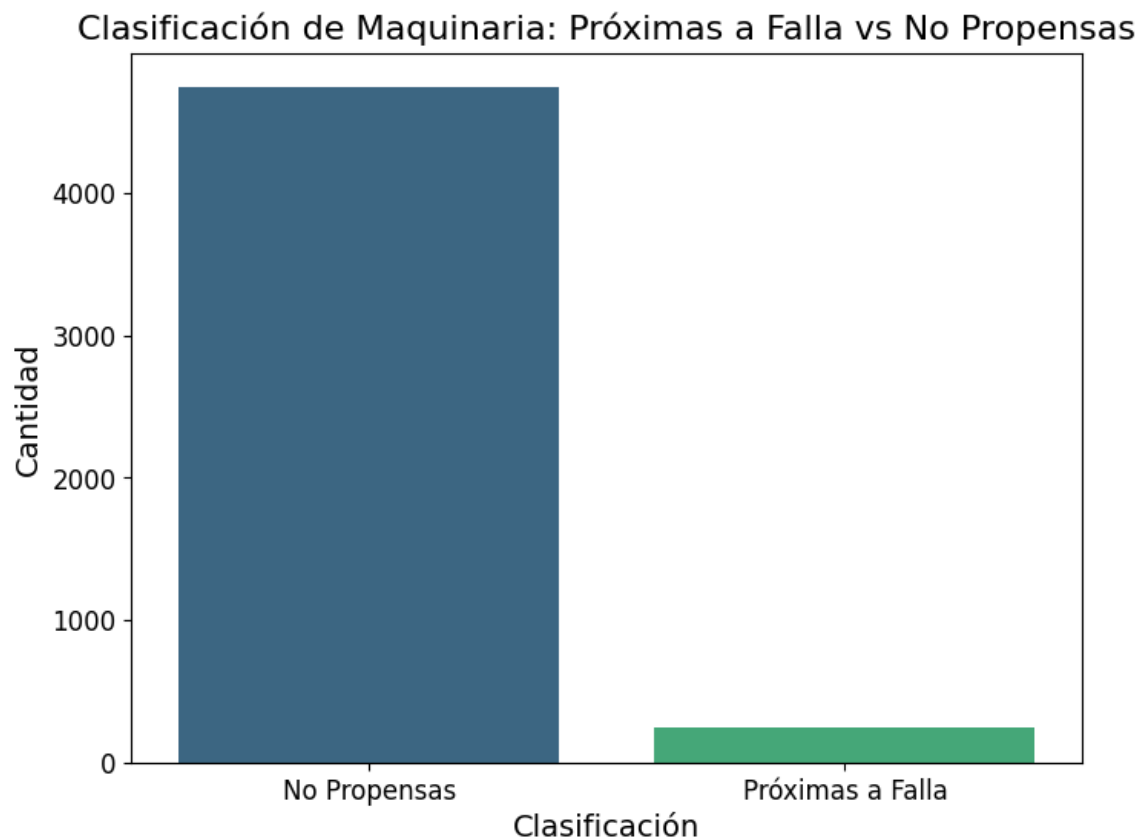


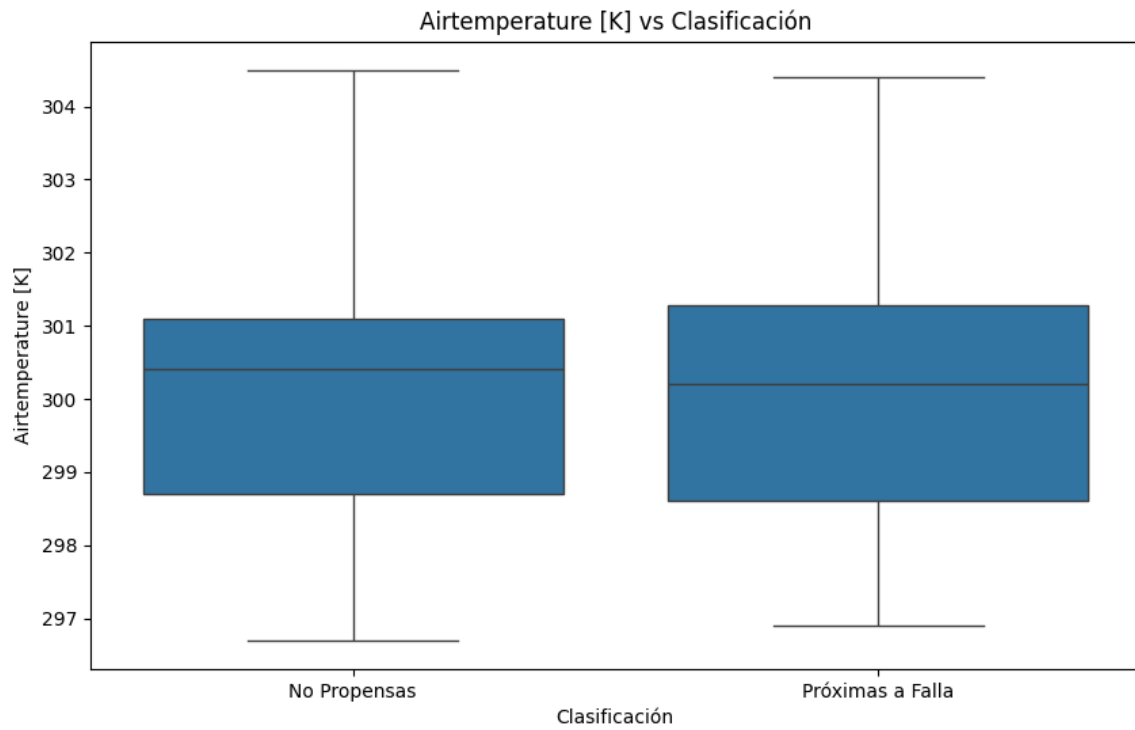
Ilustración 0-iv Perdida Precisión

Como podemos observar en la gráfica anterior en el periodo de entrenamiento y validación en el entrenamiento la pérdida es 0 en la validación aumenta hasta 0.150 de pérdida en las 50 épocas del entrenamiento y validación y por otro lado

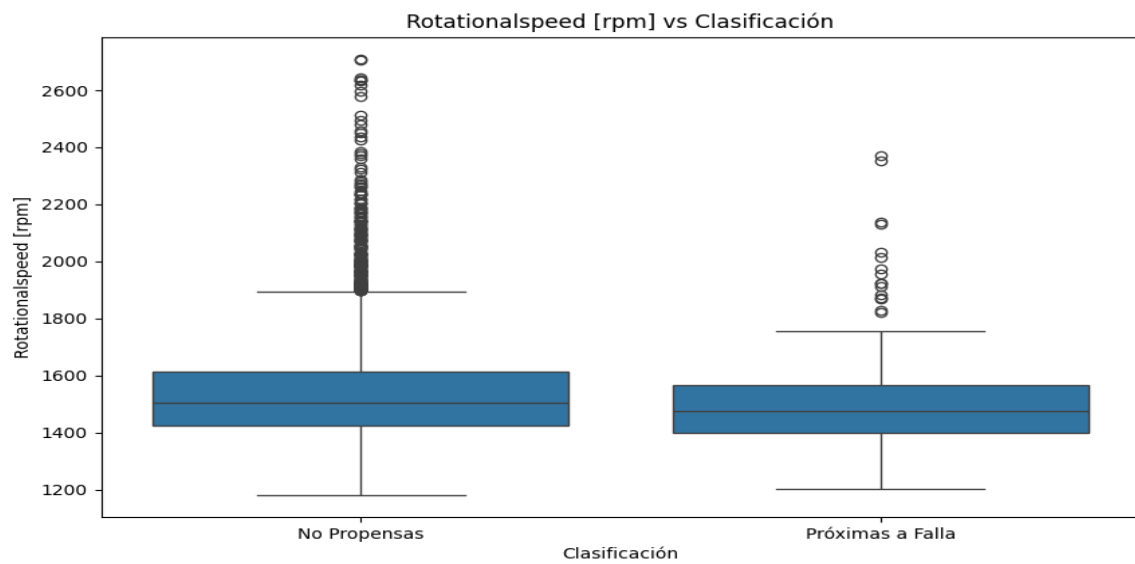
la precisión en el entrenamiento es 1 en los datos de validación se redujo un 0.05 el cual el modelo tiene una buena tasa de predicción y garantiza que el modelo es viable para predecir si la maquinaria va a tener fallas o no prontamente.



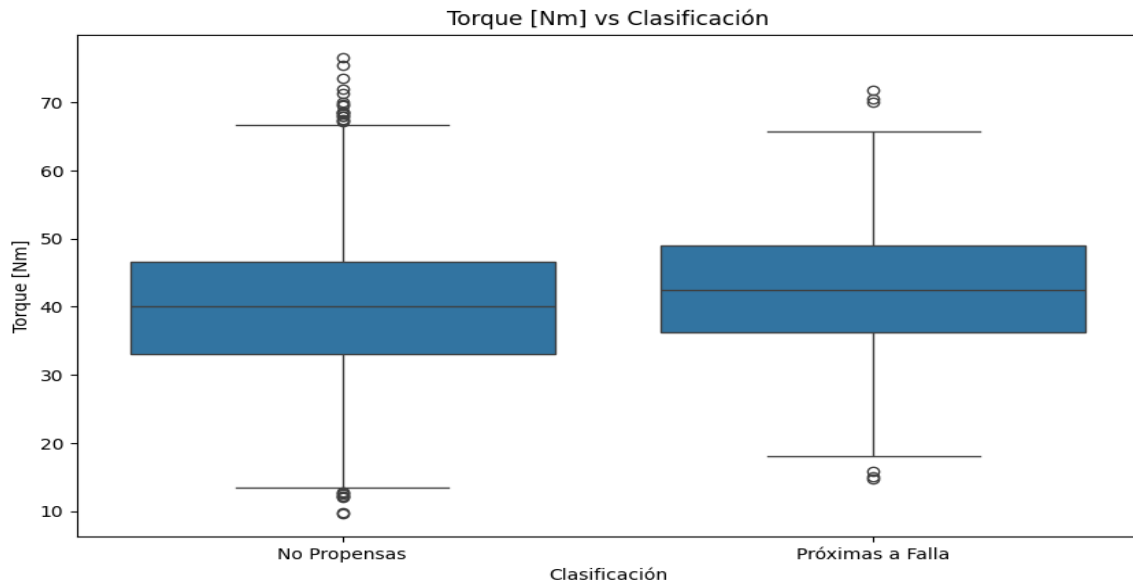
La grafica anterior nos demuestra que las maquinas que estas próximas a falla son muy pocas en comparación a las que nos propensas, es de resaltar que para que este modelo sea eficaz se debe estar ejecutando constantemente ya que los datos de las maquinas se va actualizando cada día y depende de cómo las operen. Como vamos a observar mas adelante



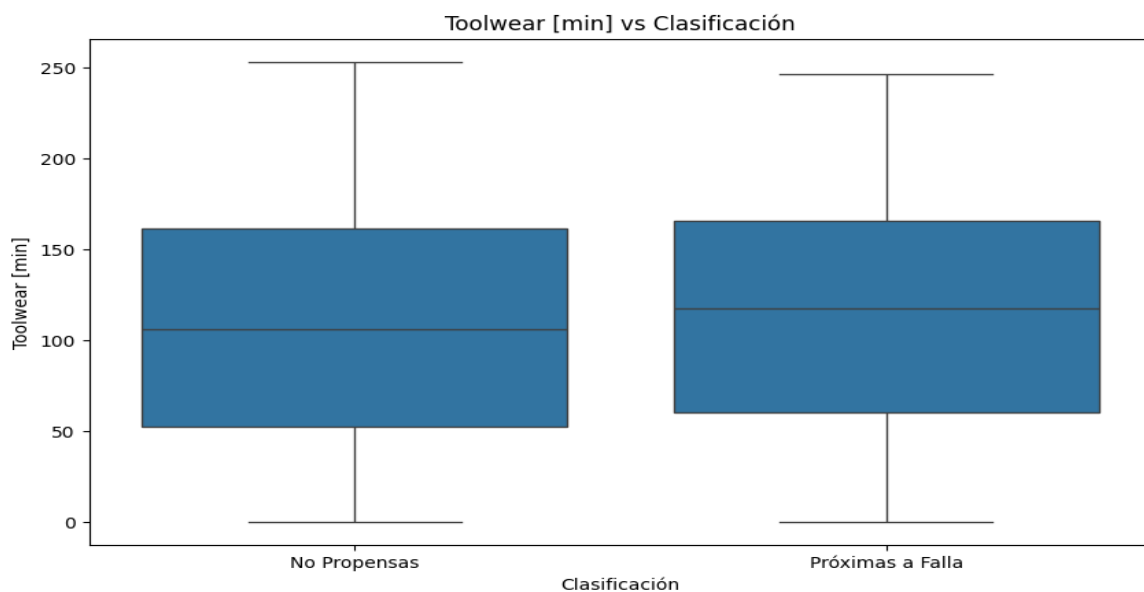
Esta grafica nos demuestra que cuando la temperatura del aire es ligeramente alta las maquinas son menos propensas a fallas.



Como podemos observar cuando las maquinas mantiene un margen de rotación en rpm por debajo del 1500 rpm son más propensas a fallas



Con esta grafica demostramos que cuando el torque este promedio a 40 nm las maquinas son menos propensas a fallas



Esta última grafica nos muestra que la maquina cuando las herramientas de trabajo estan desgastadas generan que las maquinas sea más promesas a fallar

Conclusiones Redes Neuronales

Las redes neuronales mejoran significativamente la precisión en la predicción de fallas en maquinaria industrial y agrícola en comparación con métodos tradicionales, permitiendo una clasificación detallada de los equipos según su estado operativo. Estas redes permiten una gestión proactiva del mantenimiento, reduciendo costos operativos y aumentando la disponibilidad de la maquinaria al identificar equipos propensos a fallas y priorizar intervenciones preventivas.

La preparación de datos es crucial para el éxito del modelo. La conversión de variables categóricas a dummies y la normalización de datos aseguran que todas las características sean adecuadas para el modelado. Una adecuada división en conjuntos de entrenamiento y prueba garantiza la capacidad del modelo para generalizar a datos no vistos. El modelo de redes neuronales implementado, optimizado para clasificación binaria, mostró un excelente rendimiento con una pérdida (loss) de 0.2178 y una precisión (accuracy) de 0.9770 en los datos de prueba, confirmando su viabilidad para predecir fallas.

Los análisis específicos revelaron que variables como la temperatura del aire, la velocidad de rotación, el par y el desgaste de herramientas influyen significativamente en la probabilidad de falla de la maquinaria. Máquinas son menos propensas a fallas con temperaturas ligeramente altas, mientras que las

que operan por debajo de 1500 rpm son más propensas a fallar. Un torque promedio de 40 nm reduce la probabilidad de falla, y herramientas desgastadas incrementan el riesgo de fallas.

Para mantener la eficacia del modelo, es crucial su ejecución continua con datos actualizados, ya que la operación diaria de las máquinas afecta los parámetros monitoreados. Las redes neuronales son, en resumen, herramientas poderosas que mejoran significativamente las estrategias de mantenimiento predictivo y optimizan la operación de equipos en entornos industriales y agrícolas.

Conclusión

El desarrollo de un algoritmo avanzado para el mantenimiento predictivo de maquinaria industrial y agrícola, basado en técnicas de machine learning, representa un avance significativo en la gestión eficiente de equipos. Este proyecto demuestra que la integración de modelos predictivos, como la regresión logística y las redes neuronales, puede mejorar la capacidad de anticipar fallos potenciales, permitiendo una intervención oportuna y optimizando los procesos de mantenimiento.

Los beneficios del mantenimiento predictivo son evidentes: mejora la disponibilidad de los equipos, prolonga su vida útil, reduce costos operativos y aumenta la seguridad. La implementación de este sistema requiere un enfoque integral que incluye la recopilación de datos en tiempo real mediante sensores, un

preprocesamiento riguroso para asegurar la calidad de los datos, y la selección cuidadosa de modelos de machine learning que se ajusten a las características específicas de los datos.

A lo largo de este trabajo, se han abordado los desafíos inherentes a la implementación de sistemas predictivos, como la gestión de datos complejos, los costos asociados y la necesidad de que los modelos sean interpretables y confiables. La evaluación continua y la adaptación del modelo son esenciales para mantener su precisión y relevancia en contextos operativos cambiantes.

En conclusión, la adopción de algoritmos de mantenimiento predictivo basados en machine learning no solo transforma la gestión de maquinaria, sino que también posiciona a las empresas agrícolas e industriales en la vanguardia de la innovación tecnológica. La capacidad de prever y prevenir fallos antes de que ocurran no solo mejora la eficiencia operativa, sino que también contribuye a la sostenibilidad y la competitividad a largo plazo. Este proyecto sienta las bases para futuras investigaciones y aplicaciones en el campo del mantenimiento predictivo, invitando a una colaboración continua entre ingenieros, científicos de datos y profesionales de la industria para seguir perfeccionando y ampliando estas tecnologías.

Referencias

- 4geeks.com. (2022). *4geeks.com*. Retrieved from <https://4geeks.com/es/lesson/escalado-de-caracteristicas>
- BANSAL, S. (2021). *Machine Predictive Maintenance Classification*. Retrieved from <https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification>
- Buitrago, B. (2020). *iWannaBeDataDriven*. Retrieved from <https://medium.com/iwannabedatadriven/machine-learning-modelos-de-regresi%C3%B3n-i-d293ae235e9a>
- datos.gob.es. (2023). *datos.gob.es*. Retrieved from <https://datos.gob.es/es/blog/como-preparar-un-conjunto-de-datos-para-machine-learning-y-analisis>
- engeman.com. (2023). *Desafíos en la implementación de la gestión del mantenimiento*. Retrieved from <https://blog.engeman.com/es/implementacion-de-la-gestion-del-mantenimiento/>
- Escobar, L. (2010). *redes Neuronales en prediccion de series de tiempo. santiago de chile*: <https://www.palermo.edu/ingenieria/Pdf2010/CyT9/02.pdf>.

Anexos

Documento del EDA



Analisis_datos_mod
elos.html

Script Modelo de regresión logística



Mantenimiento_Re
gresionLogistica.htr

Script modelo de redes neuronales



Maquinaria_Clasificacion.html