

# MANTENIMIENTO PREDICTIVO

APLICACION DE MODELOS DE  
MACHINE LEARNING

Claudia Murillo

2023



# INDICE

Contexto

Preguntas de Interes

Estructura de Datos

Analisis Exploratorio de Datos

Entrenamiento & Evaluacion de Modelos de ML

Insights&Conclusiones

# CONTEXTO

Las empresas buscan como aumentar el tiempo de servicio de las maquinas, reducir las paradas inesperadas, reducir costos de mantenimiento.

Todo esto lleva a la necesidad de encontrar formas más precisa de conocer o predecir el comportamiento futuro de la maquinaria basándose desde la adquisición y análisis de datos e información que se puede obtener a través del monitoreo real y continuo de la condición de trabajo de los equipos.

# PREGUNTAS DE INTERES

1. Existe alguna relación entre los parámetros monitoreados de las maquinas que permita modelar el efecto que generan en la maquina?
2. Cuales son los componentes de las maquinas que mas fallas registra?
3. Es posible visualizar los comportamientos anormales de las maquinas que sean indicio de falla?

# ESTRUCTURA DE LOS DATOS

La composición del dataset es la siguiente:

```
telemetry_df.head(3)
```

	datetime	machineID	volt	rotate	pressure	vibration
0	2015-01-01 06:00:00	1	176.217853	418.504078	113.077935	45.087686
1	2015-01-01 07:00:00	1	162.879223	402.747490	95.460525	43.413973
2	2015-01-01 08:00:00	1	170.989902	527.349825	75.237905	34.178847

```
errors_df.head(3)
```

	datetime	machineID	errorID
0	2015-01-03 07:00:00	1	error1
1	2015-01-03 20:00:00	1	error3
2	2015-01-04 06:00:00	1	error5

```
failures_df.head(3)
```

	datetime	machineID	failure
0	2015-01-05 06:00:00	1	comp4
1	2015-03-06 06:00:00	1	comp1
2	2015-04-20 06:00:00	1	comp2

```
maint_df.head(3)
```

	datetime	machineID	comp
0	2014-06-01 06:00:00	1	comp2
1	2014-07-16 06:00:00	1	comp4
2	2014-07-31 06:00:00	1	comp3

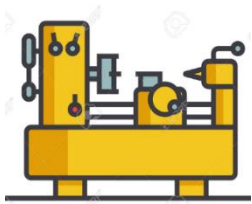
```
machines_df.head(3)
```

	machineID	model	age
0	1	model3	18
1	2	model4	7
2	3	model3	8

- **telemetry\_df** : Contiene una serie de tiempo con los registros de los parametros de funcionamiento de las maquinas. (876100 , 6)
- **errors\_df** : Es el registro de las alarmas generadas por los equipos durante el tiempo de supervision. Caba aclarar que estas alarmas no son consideradas como fallas al no implicar la perdida de funciones de los equipos. (3919 , 3)
- **Failures\_df** : Registro del componente de la maquina que present una falla. (761 , 3)
- **maint\_df** : Es el registro de los componentes de las maquinas a los cuales se les realizo algun tipo de mantenimiento, sin especificar si fue por rutina(preventivo) o debido a una falla (correctivo). (3286 , 3)
- **machines\_df** : Contiene el ID, el modelo y la edad de las maquinas. (100 , 3)

# ESTRUCTURA DE LOS DATOS

## Equipos registrados



**100**  
**Equipos**

## Parametros Evaluados



**Voltaje**  
**Rotación**  
**Presión**  
**Vibración**

Adicionalmente...

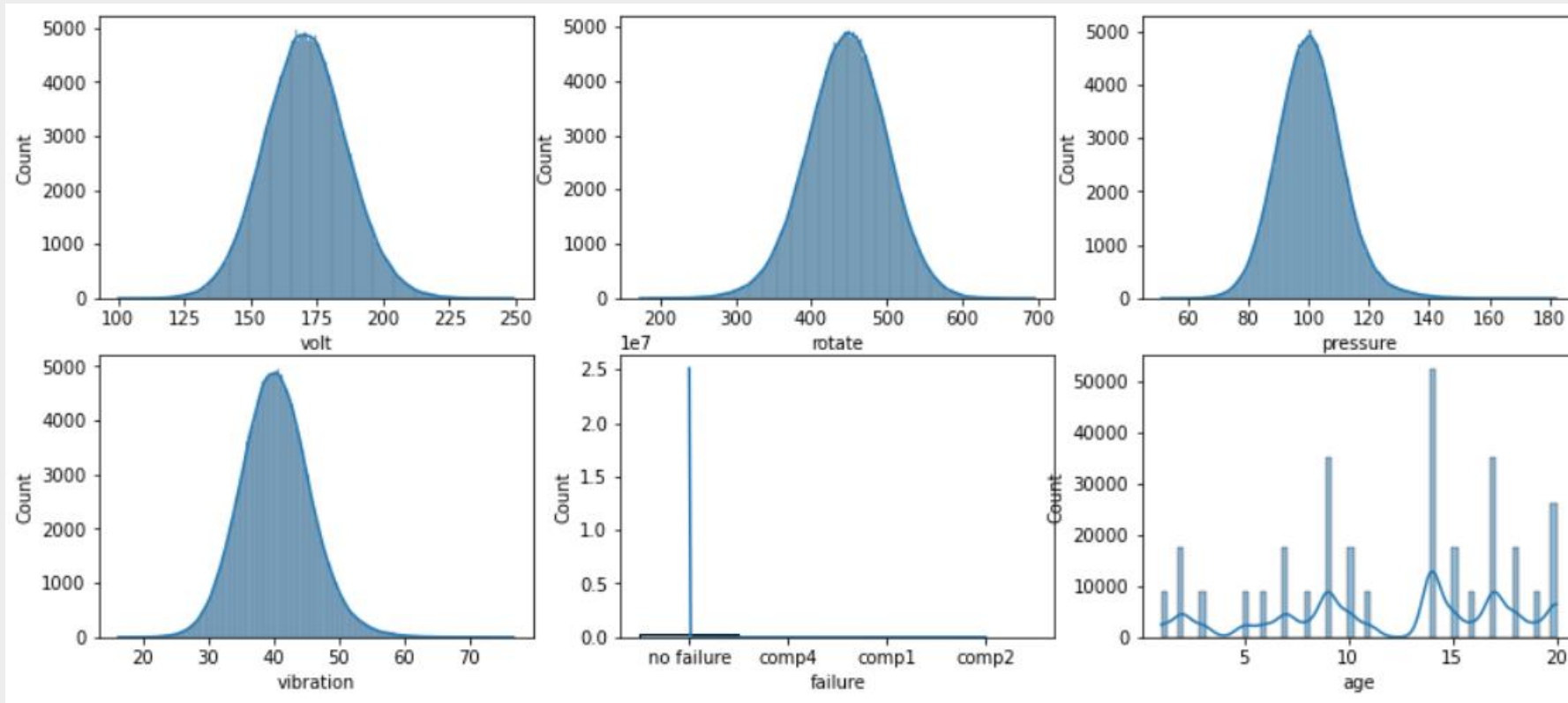
- Este dataset fue **creado artificialmente** para recrear el comportamiento de distintos tipos de maquinas dentro de un procesos productivo habitual.
- Se registran las **Fallas** de 4 componentes de las maquinas.
- El período de registro de parametros es entre **2015-2016**



# ANALISIS EXPLORATORIO DE DATOS

EDA

# Distribución de los Parámetros de Operación

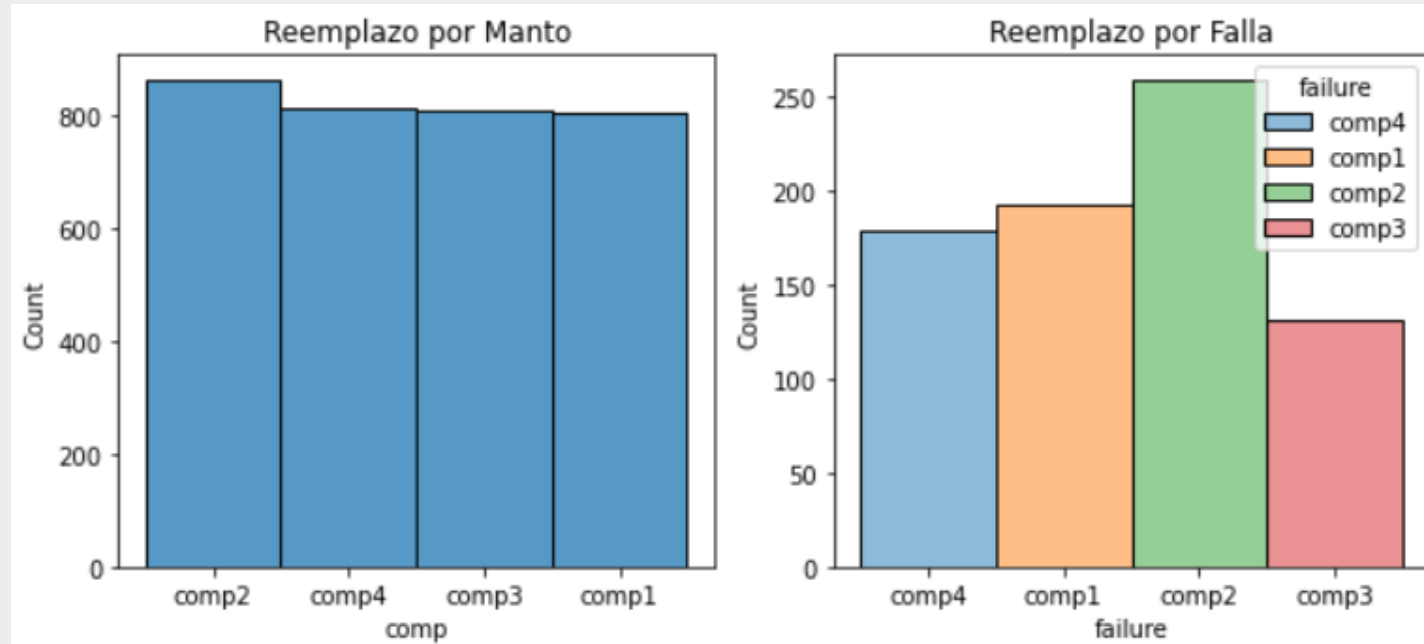


Gráfica 1

Se puede observar una distribución casi normal en todos los parámetros evaluados



# Distribución de los Parámetros de Operación

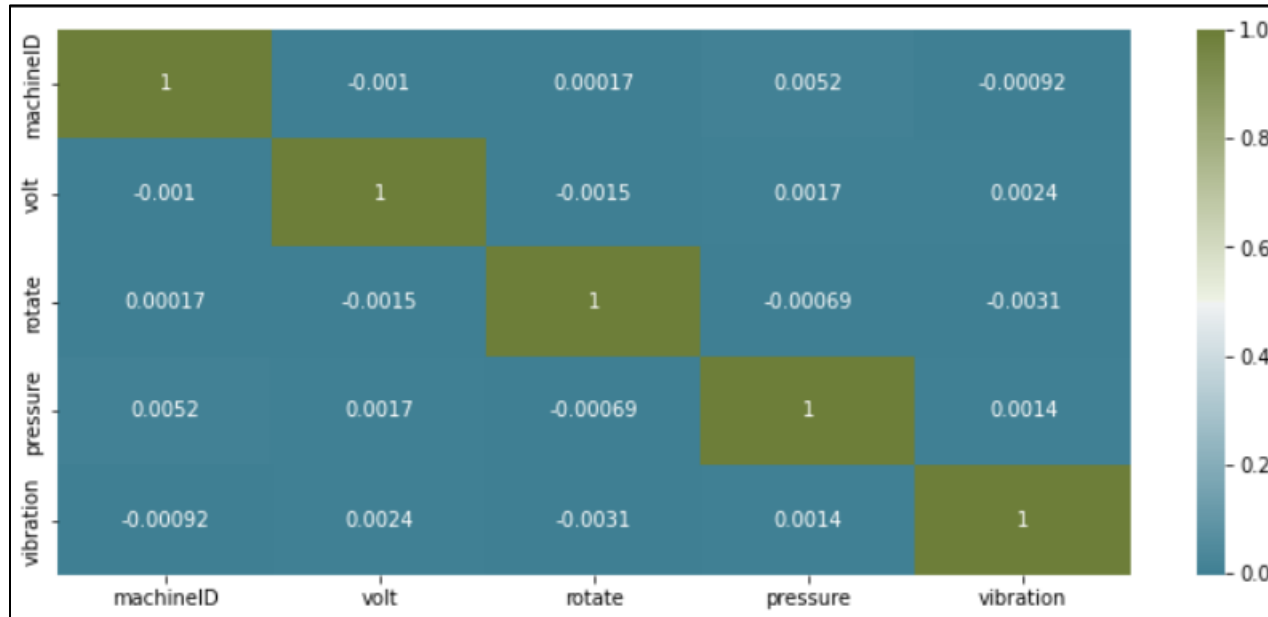


Gráfica 2

Gráfica 3

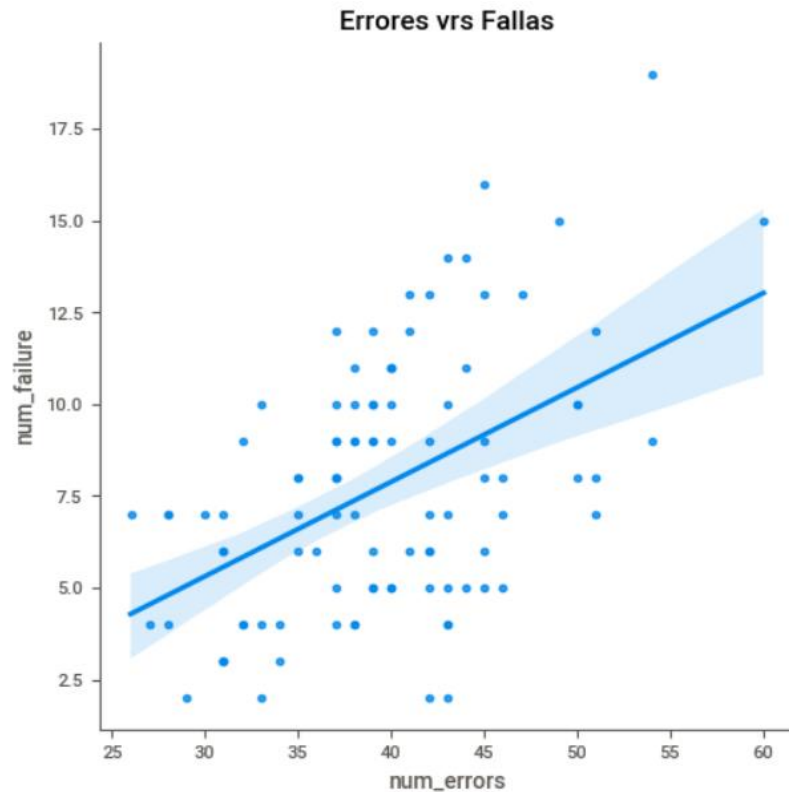
Con estas graficas lo que se puede observar es que una cantidad considerable de reemplazo de los componentes se realiza de forma prematura.

# EVALUACION DE CORRELACION

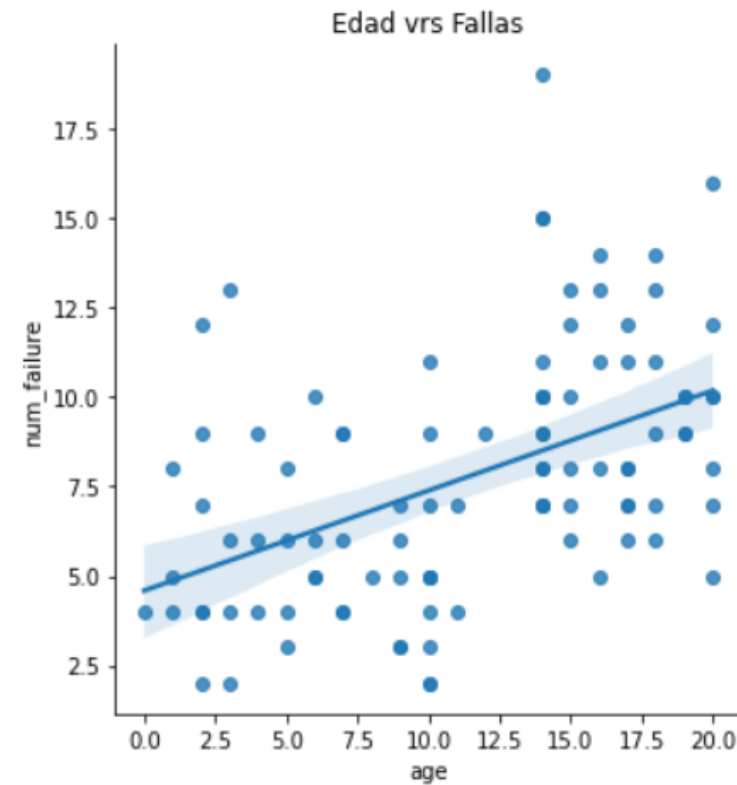


Los valores de correlación entre las variables de funcionamiento de las maquinas se acercan a cero, lo que indica la **no dependencia** entre estas variables.

# OTRAS RELACIONES

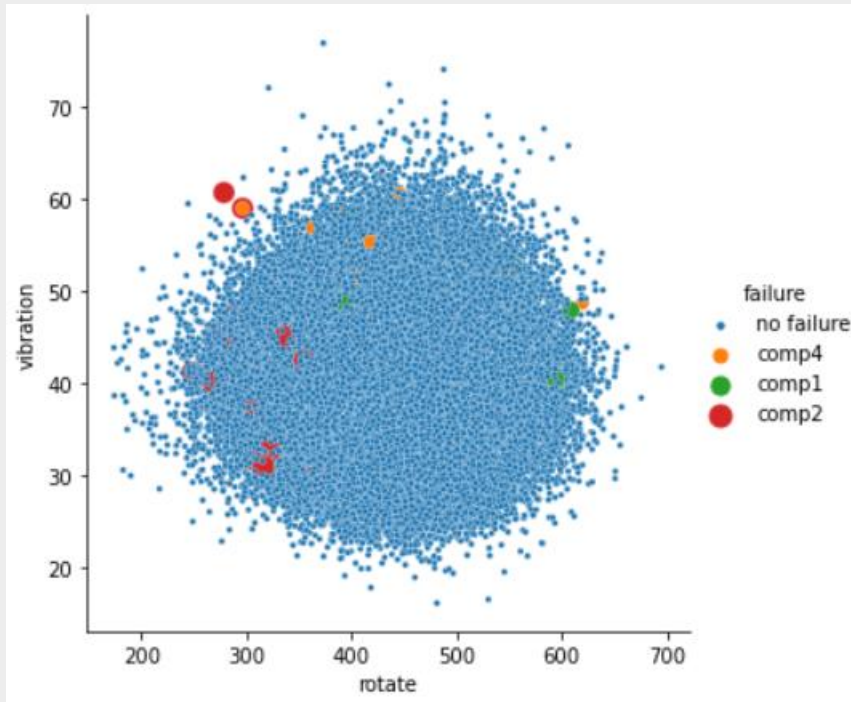


Aca si se observa una correlacion positiva entre los eventos de errores y las fallas.

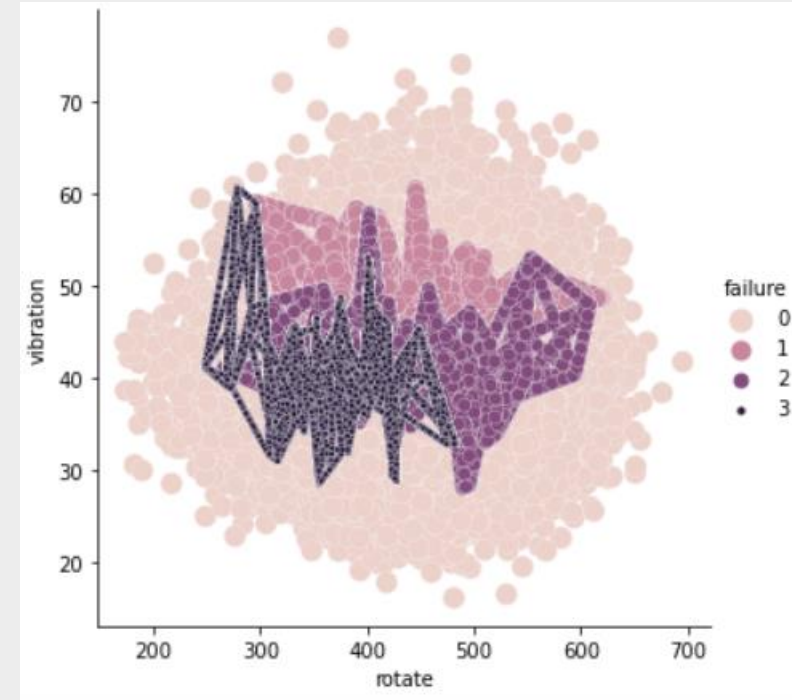


Como es de esperar, las maquinarias con mayor edad y por lo tanto con un esperado mayor desgaste, tenderan a presentar mas fallas.

## Proporción de los registros de interes



Los registros de las fallas son muy inferior a los registros de no falla, lo cual nos crea un dataset desbalanceado.



Despues de la aplicación de la tecnica SMOTETomek para el balanceo del dataset, se crean registros sintéticos de los registros minoritarios.



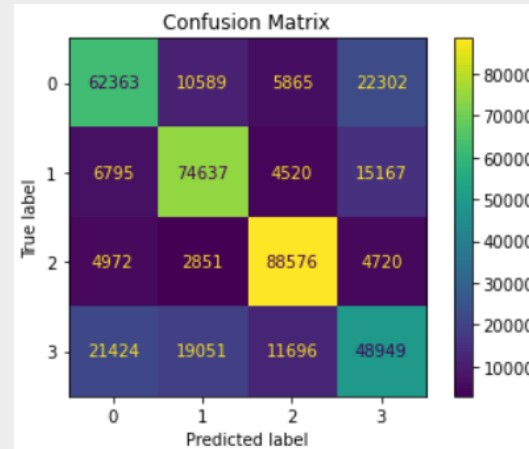
# ENTRENAMIENTO & EVALUACION DE LOS MODELOS DE ML

# 1. REGRESION LOGISTICA (Simple)

```
Training Accuracy : 67.92 %
Model Accuracy Score : 67.87 %
-----
Classification_Report:
      precision    recall  f1-score   support

0         0.65       0.62       0.63      101119
1         0.70       0.74       0.72      101119
2         0.80       0.88       0.84      101119
3         0.54       0.48       0.51      101120

 accuracy          0.68      404477
  macro avg         0.67       0.68      404477
 weighted avg         0.67       0.68      404477
-----
```



En una primera instancia, se entreno un modelo simple de regresión logistica el cual no entrego un buen desempeño.

## 1.1 REGRESION LOGISTICA + Ajuste de Hiperparametros

Luego del ajuste de los hiperparametros al modelo y el reentrenamiento obtuvimos solo una leve mejora en la precision de las predicciones del modelo.

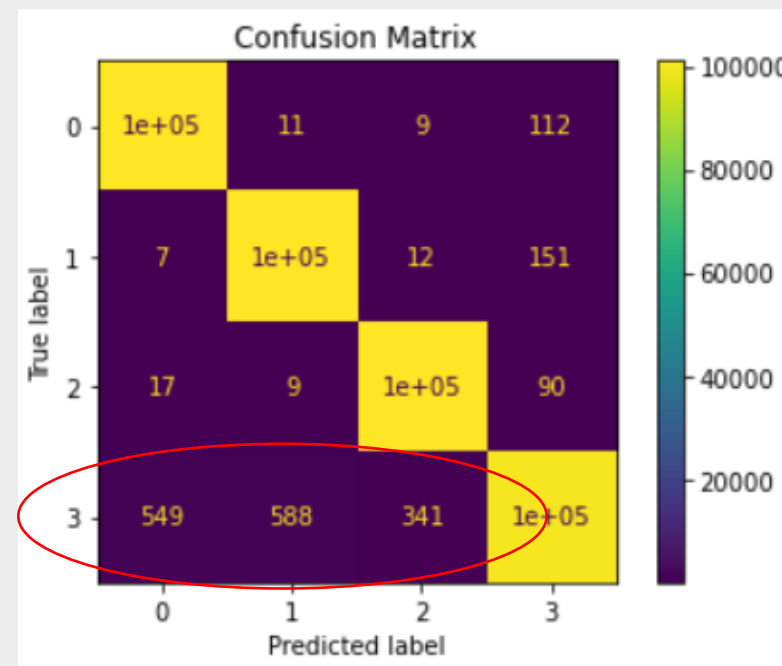
```
Tuned Hyperparameters : {'C': 0.001, 'penalty': 'l2', 'solver': 'liblinear'}
Accuracy : 0.6826751790701392
```

## 2. DECISION TREE

```
Training Accuracy   : 100.0 %
Model Accuracy Score : 99.53 %
-----
Classification_Report:
      precision    recall  f1-score   support

     0       0.99       1.00       1.00      101119
     1       0.99       1.00       1.00      101119
     2       1.00       1.00       1.00      101119
     3       1.00       0.99       0.99      101120

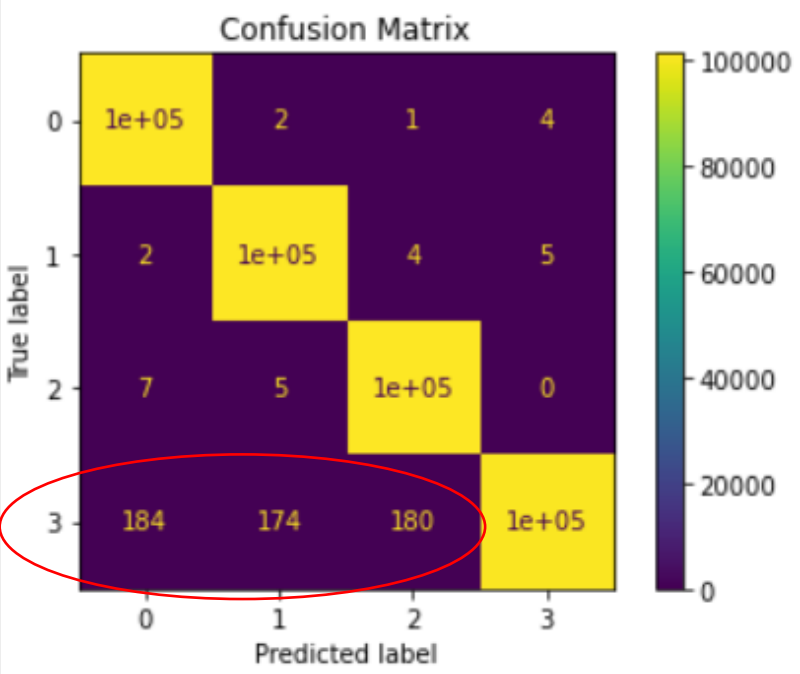
   accuracy                1.00      404477
  macro avg       1.00       1.00       1.00      404477
 weighted avg     1.00       1.00       1.00      404477
-----
```



Es evidente el salto en la accuracy que se logró con la aplicación de este modelo con respecto a la regresión logística. Este modelo presenta un alto valor de fallos en la predicción de la **etiqueta 3**

### 3. RANDOM FOREST

Training Accuracy : 100.0 %				
Model Accuracy Score : 99.86 %				
-----				
Classification_Report:				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	101119
1	1.00	1.00	1.00	101119
2	1.00	1.00	1.00	101119
3	1.00	0.99	1.00	101120
accuracy			1.00	404477
macro avg	1.00	1.00	1.00	404477
weighted avg	1.00	1.00	1.00	404477
-----				



Se alcanzó una mejoría en el accuracy del modelo que parece insignificante pero si observamos los valores predichos de la **etiqueta 3**, los errores en la predicción disminuyeron hasta un 66% en algunos casos,





# INSIGHTS & CONCLUSIONES

# INSIGHTS & CONCLUSIONES

1

Todas las variables de funcionamiento siguen una distribución normal pero no existe una correlación aparente entre ellas. Al ser registros temporales es posible que la relación sea mas con el registro anterior que con otras variables.

2

Se inicio el trabajo con la idea de realizar una clasificación binaria pero luego se decidió realizar una clasificación multivariada porque se considero de importancia conocer con precisión cual componente es el que fallara.

3

Basado en lo observado durante la exploración de los datos, se podrian estar reemplazando componentes que aun podrian entragar mas vida util. Aca se presenta una oportunidad para optimizar el programa de mantenimiento al crear un modelo predictive.

4

Se probaron 3 modelos de los cuales el que presento un mejor desempeño fue Random Forest seguido del Decision Tree. Aquel que peor se desempeño para esta tarea en particular de clasificación fue la regresión logística.

A series of white, thin, overlapping geometric lines on a black background, creating a complex, abstract pattern on the left side of the slide.

# GRACIAS!

Claudia Murillo