

¿Maquinas Eternas?

Maximiliano Belotti
CODER HOUSE Informe

Base de datos

- Los datos pertenecen a un DATABASE publico, <https://www.kaggle.com/shivamb/machine-predictive-maintenance-classification> .
- Dentro del directorio https://github.com/maxi879/proyecto_final_coder/blob/c6ad029288470c03c1b1730be33d6b686970f367/DATOS se encuentra los datos en formato .xlsx y un .txt con los detalles tecnicos necesarios para realizar un analisis.

Bibliotecas externas

- Las bibliotecas utilizadas son :
 - Pandas ==> Utilizado el objeto DataFrame para la manipulacion de datos.
 - Seaborn ==> Utilizado para la creacion y visualizacion de datos.
 - Matplotlib ==> Llamando especificamente a pyplot para graficar.
 - Warnings ==> Para filtrar errores relativamente ignorables.
 - Numpy ==> Para manipulacion de los datos.
 - Sklearn ==> Machine Learning y metricas.
 - Imblearn ==> Para solucionar problemas de balanceo de datos.
 - pandas_profiling ==> Profile

Objetivo

Implementar un modelo de MACHINE LEARNING que puede predecir la rotura o no, de una herramienta de produccion masiva. En caso de rotura anticipar la razon de la rotura.

Con estas inteligencias de negocio, se podran realizar mantenimientos preventivos, optimizando el uso de materiales y repuestos. Esto tendra un impacto notorio en la linea de produccion, logrando estabilidad, minimizando fallas y tiempos de reparacion.

Aspectos importante de los datos

El data set contiene 10 variables, dos son usadas como index (UDI,Product_ID). El data set contiene dos variables, Target y Failure_Type, ambas son las variables a predecir. Target es una variable binaria, mientras que Failure_Type es es multiclass.

Dentro de las restantes 6 se encuentran Type, es una variable categorica, mientras que las otras 5 son variables numericas, que representan datos continuos proveniente de sensores instalados en las maquinas.

Usamos definiciones matematicas para explicar relaciones mas complejas entre las features:

- $Potencia[W] = T[Nm] * Vang[rad/s]$
- $Sobre\ esfuerzo = Uso[min] * T[Nm]$
- $Despeje\ de\ Temp = |Temp_ambiente - Temp_Trabajo| * rpm$

Insight: Estos graficos describen el comportamiento de las variables durante los momentos de falla:



- La primera columna de graficos muestra los puntos de Power Failure y como se comportan las diferentes variables, es importante notar como las variables Rotational_speed y Torque varian en comparación de los graficos de los momentos sin fallas, estos se distribuyen lejos de la media correspondiente (Torque = 40Nm, Rotational_speed = 1500 rpm), lo cual es esperable teniendo problemas de potencia. Es importante notar que esta falla se presenta una mayor cantidad de veces en los primeros minutos de uso, siendo esta una razon para creer que puede ser un problema de calibración de la herramienta.
- La segunda columna representa la fallas por desgaste, por uso de la misma, esta falla es particular ya que solo se puede concluir que esta falla se presenta en momentos donde los valores de Tool_wear(min) son altos, es por eso que esta variable debe ser analizada pensando en cuales son los valores de Tool_wear maximos que se deben llegar antes de realizar el remplazo de piezas para evitar fallas.
- La columna del medio refiera a los problemas por sobreesfuerzo, esta falla presenta tres características marcadas en los datos que explican la falla, lo primero que se observa es bajos valores de rpm que se explica por valores altos de Torque, a esto se le suma la tercer característica, determinante para esta tipo de rotura, altos valores de Tool_wear, la combinacion entre alto Torque y gran cantidad de desgaste resulta en una falla por Overstrain.
- Las datos de Random failure no muestran ningun comportamiento marcados.
- Heat Dissipation failure, esta falla presenta bajas rpm, altos niveles torque y altos niveles de temperatura. Esta falla a diferencia de las anteriores, es independiente de la variable Tool_wear, ya que se produce en cualquier momento según los graficos.

Puntos de interes

- Este analisis parte de la necesidad de brindar la capacidad de prevenir y evitar roturas de maquinas de produccion masiva.
- Buscaremos identificar cuales son las variables mas significativas al momento de rotura segun su tipo. Esto permitira a la empresa disminuir los momentos de falla, logrando asi una mayor estabilidad de la linea de produccion y evitar costos mayores por fallas serias causadas por problemas constantes sin resolver. Se fijaran parametros de funcionamiento normales.

Valores medianos según rotura y valores normales.

		Air_tempera	Process_tem	Rotational_s	Torque_[Nm]	Tool_wear_[min]	Power_[W]	Overstrain_[minNm]	Heat dissipation_[rpmK]
Target	Failure_Type								
0	No Failure	300.00	310.0	1507.0	39.80	107.0	6.243.035.753	3951.20	15196.80
	Random Failu	300.75	311.1	1490.0	44.60	142.0	6.915.074.782	5645.40	14837.85
	Heat Dissipati	302.45	310.7	1346.0	52.35	106.0	7.298.925.044	5549.65	11050.85
	No Failure	300.50	309.9	1438.0	45.20	119.0	6.811.266.088	4449.90	15465.90
	Overstrain Fa	299.45	310.1	1362.5	56.75	207.0	8.088.454.402	11545.40	13862.65
1	Power Failure	300.40	310.2	1386.0	63.60	100.0	9.100.377.103	3608.80	14188.20
	Tool Wear Fa	300.40	310.3	1521.0	37.70	215.0	5.986.933.120	7843.20	14706.0

Estos datos pueden ser observados en los cuadros de medianas. Dentro del directorio \\EDA.

- Falla por desgaste de la herramienta (TWF): la herramienta se reemplazará por falla en un tiempo de desgaste de la herramienta seleccionado aleatoriamente entre 200 y 240 minutos (120 veces en nuestro conjunto de datos). En este momento, la herramienta se reemplaza 69 veces y falla 51 veces (asignadas aleatoriamente).

- Falla por disipación de calor (HDF): la disipación de calor provoca una falla en el proceso, si la diferencia entre la temperatura del aire y la del proceso es inferior a 8,6 K y la velocidad de rotación de la herramienta es inferior a 1380 rpm. Este es el caso de 115 puntos de datos.

- Falla de energía (PWF): el producto del par y la velocidad de rotación (en rad/s) es igual a la potencia requerida para el proceso. Si esta potencia está por debajo de 3500 W o por encima de 9000 W, el proceso falla, que es el caso 95 veces en nuestro conjunto de datos.

- Fallo por sobreesfuerzo (OSF): si el producto del desgaste de la herramienta y el par supera los 11 000 minNm para la variante de producto L (12 000 M, 13 000 H), el proceso falla debido al sobreesfuerzo. Esto es cierto para 98 puntos de datos.

Machine Learning

El modelo optimo para la prediccion de ambas variables Target, es RandomForestClassifier. Se comprobaron varios algoritmos de clasificacion para la variable Target Binaria. En cuanto la variable Target Multiclas se realizo un hipertuning, logrando un aumento de F1_Macro_avr del 0.67 al 0.81 .

Antes del hipertuneo.

	precision	recall	f1-score	support
Heat Dissipation Failure	0.49	1.00	0.66	112
No Failure	1.00	0.93	0.97	9652
Overstrain Failure	0.62	1.00	0.76	78
Power Failure	0.79	1.00	0.88	95
Random Failures	0.39	1.00	0.56	18
Tool Wear Failure	0.10	1.00	0.18	45
accuracy			0.94	10000
macro avg	0.57	0.99	0.67	10000
weighted avg	0.98	0.94	0.96	10000

Post hipertuneo.

	precision	recall	f1-score	support
0	0.58	1.00	0.74	112
1	1.00	0.97	0.98	9652
2	0.92	1.00	0.96	78
3	1.00	0.99	0.99	95
4	1.00	0.50	0.67	18
5	0.33	1.00	0.49	45
micro avg	0.98	0.97	0.98	10000
macro avg	0.80	0.91	0.81	10000
weighted avg	0.99	0.97	0.98	10000
samples avg	0.97	0.97	0.97	10000