Notas explicativas:

1.Obtenemos desde el Scada un año de datos de procesos del funcionamiento de una turbina de gas. “TBM\_data.csv”. Este fichero consta de 169 variables y más de 10.000 registros.

2. En el Notebook Obtencion\_Ficheros.ipynb, y en el script get\_files.py, tenemos el código para obtener:

i. Un fichero de entrenamiento para el modelo con los seis primeros meses del año (u otro intervalo de tiempo deseado)

ii. Un obtención masiva de ficheros más pequeños para simular el control diario, que nos servirán para el análisis de patologías de la máquina, sensibilidad a las condiciones del sistema, etc.

3. Ejecutar la primera celda del Notebook Obtencion\_Ficheros.ipynb nos debe dar un fichero “data.train.csv” que lo usaremos en el siguiente para el proceso de aprendizaje automático. Ejecutar la segunda celda nos devolverá 211 ficheros “data\_train\_(i).csv” para la simulación de control diario.

4. En el Notebook MLconSeleccion\_r2.ipynb, obtenemos el modelo entrenado y con las variables seleccionadas, atendiendo al criterio de selección de características Forward de mejor r2, Igualmente en el Notebook MLconSeleccion\_K.ipynb, obtenemos el modelo entrenado con el criterio de selección de características SElectKBest. Se utilizan dos scripts ModReg\_TG.py y f\_Selection.py, que poseen diferentes funciones para el desarrollo del código y obtención de estadísticos.

5. Conclusiones sobre, qué modelo de selección es más eficiente.

6. En Notebook OYMconDL\_TBM.ipynb, entrenamos los modelos de control diario. En este caso sólo se utilizan las variables seleccionadas en el paso anterior. Utilizamos el criterio de selección de variables Forward de mejor r2, para obtener la sensibilidad del sistema a cada variable.

7. Por último, y mediante el mismo Notebook del paso 6., obtenemos un dataframe con la estructura mostrada en la siguiente tabla

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Fecha Análisis | Nombre Variable | R2 | Valor Medio | % desv Std | Coef Variable (W) |
| dd/mm/aa | Bias | - | - | Std Dev | - |
| dd/mm/aa | Variable 1 |  |  |  |  |
| dd/mm/aa | … | … | … | … | … |
| dd/mm/aa | Variable n |  |  |  |  |
| dd/mm/aa | Desv Tecnico % | - | % | - | - |
| dd/mm/aa | R2\_Train | - | 0.xx | - | - |
| dd/mm/aa | R2 Test | - | 0.xx | - | - |

Estos mismos datos son pasados a formato json, enviados a MongoDB y mediante una conexión entre Mongo y Tableau, pueden ser visualizados. (Los pasos de configuración serán explicados en al memoria del TFM).

El código se recoge en el Script ModDL.py, que contiene las funciones.