

PEC2

Introducción

Dados los siguientes datasets:

1. Wine.csv → definir la calidad (clasificación)
2. Realstate.csv → definir el coste de una casa (regresión)
3. Titanic.csv → definir quién sobrevive (clasificación)
4. Insurance.csv → definir el valor del seguro (regresión)
5. Cualquier dataset que quieran

Elegir una de las 5 opciones para realizar el ejercicio. Vale la pena mencionar que si se elige la opción 5 se restan 10 pts del máximo total.

Primera Parte (20 pts)

Crear el pipeline de ML (con transformers y estimators) para dar un resultado (los labels están definidos arriba)

Segunda Parte (20 pts)

Empaquetar el algoritmo, dos formas posibles a elección:

- Con un archivo de tipo pkl
- Con una librería disponible en PiPy

Tercera Parte (20 pts)

Crear el contenedor y subirlo a Docker Hub (hacerlo en un repositorio público)

Cuarta Parte (20 pts)

Disponibilizar el algoritmo como API utilizando Flask dentro de un Contenedor y subirlo a Docker Hub como un versión 2, es decir, con un tag diferente

Extras (20 pts)

- Crear un entorno tox de pruebas dentro del repositorio que se empaquetó (se debe subir a Git)
- Definir un Pod con un NodePod para poder ejecutarlo en Kubernetes.

Entregables:

- **Obligatorios:**
 - El comando docker para hacer pull del repositorio con una presentación explicando como funciona el algoritmo
 - URL de Git con la información del DockerFile (y lo que se utilizó para crear la imagen)
- Para los extras url al repositorio Git con el código del paquete python y los yamls del Pod y NodePort