

Predicción del valor de un coche usado

Para este problema utilizaremos el conjunto de datos de dominio público 'vw.csv', obtenido de kaggle.com.

Disponemos de información de vehículos usados (marca Volkswagen), cada uno de ellos con su precio de valor de mercado.

Los datos que contiene el dataset son los siguientes:

- model: Variable categórica con el modelo de coche.
- year: Año de fabricación del modelo.
- price: Precio de valor de mercado.
- transmission: Tipo de caja de cambio.
- mileage: Numero de kilómetros.
- fuelType: Tipo de combustible usado
- tax : Impuesto de circulación.
- mpg: Consumo del vehículo, en galones por milla.
- engineSize: Tamaño del motor, en litros

Hay que realizar las siguientes tareas:

1. eliminar los registros (filas) que contengan algún valor nulo,
2. Codificar las variables categóricas (transmission, fuelType y model) con el método 'get_dummies' de pandas.
3. La variable objetivo es 'price', obtener variables predictoras y objetivo para entreno y testeo (20%), con el 'random_state' que cada persona escoja.
4. Entrenar los siguientes modelos, obtener los 'scores' y comparar los resultados:
 - Árbol de decisión
 - K-vecinos (n_neighbors = 20)
 - Support Vector Machines (SVR)
 - Random Forest (n_estimators = 100)
 - Bagging (n_estimators = 150)
 - AdaBoost
 - Stacking usando como modelos base los tres primeros modelos de esta lista, y como meta-modelo la regresión lineal.

Para ello, crear una lista con los modelos y hacer un bucle que vaya recorriendo la lista. Cada vez hay que imprimir en pantalla un mensaje de que ha comenzado el entrenamiento de cada modelo. Ir metiendo en una lista los acrónimos de los nombres de los algoritmos que se van entrenando, y en otra lista los scores de los modelos resultantes de cada algoritmo.

Al final hay que hacer un bar-plot con los acrónimos de los algoritmos en el eje x y los valores de los 'score' en el eje y.

Indicar el algoritmo que consigue el mejor 'score'.