**BASE CODE RESULT**

**Import Libraries:**  numpy, pandas, sklearn, matplotlib, seaborn,lightbgm etc.

# Dataset Name: Used Cars Price Prediction

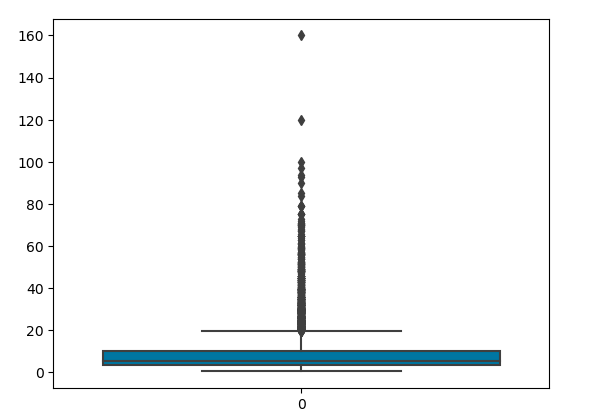
**Dataset Link:** <https://www.kaggle.com/datasets/avikasliwal/used-cars-price-prediction?select=train-data.csv>

**Data preprocessing:**

* We read our csv through “car\_price (1).csv”
* Then we have describe the data ,find the number of rows and columns then used to find the information of data
* Now, we have replace all the zero values of column ’seats’ with the median and drop the unnamed column.
* We use a lambda function taking x value applies to each value of the column mileage, Engine, Power and splits it using whitespace as delimiter
* After that we have replace the nan and null values of the column mileage, Engine, Power and convert it into float values, after that we have fill the null values with the median.
* We have draw the boxplots of the data to make it more clear
* At last, we have apply the Label Encoder to convert the categorical data into numerical data and provide the index to Owner\_Type column and use groupby for the location and price column

**Data Visualization:**

**Boxplot**

****

**Model: -**

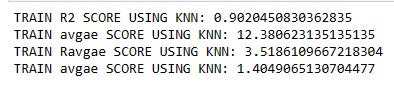
**KNN:** We have use the knn (KNearestNeighbors) model to predict the R2score,mean squared error, root mean squared error and mean absolute error ,first of all we have divide the data into train and test after that we have fit the train and test and after that we have predict the data,first of all we have divide the data into 75% training and 25% testing and calculate the r2score,mse,rmse,and mae of the knn,then we have divide the data into 80% training and 20% testing and calculate the same ,at last we have divide data into 85% training and 15% testing .

We have also apply the cross validation , fivefold and tenfold cross validation for the different k values 2,3,4,5.

**Training Results:**

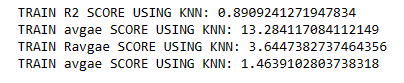
**For 25% Test Data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| KNN | R2score | MSE | RMSE | MAE |
|  | 0.90 | 12.38 | 3.51 | 1.40 |

****

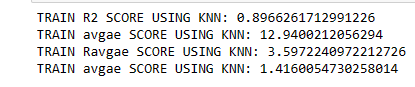
**For 20% Test Data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| KNN | R2score | MSE | RMSE | MAE |
|  | 0.89 | 13.28 | 3.64 | 1.46 |

****

**For 15% Test Data**

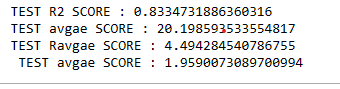
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| KNN | R2score | MSE | RMSE | MAE |
|  | 0.89 | 12.94 | 3.59 | 1.41 |

****

**Testing Results:**

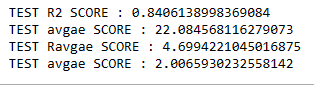
**For 25% Test Data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| KNN | R2score | MSE | RMSE | MAE |
|  | 0.83 | 20.19 | 4.49 | 1.95 |

****

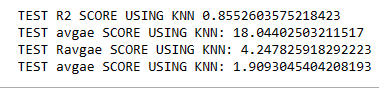
**For 20% Test Data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| KNN | R2score | MSE | RMSE | MAE |
|  | 0.84 | 22.08 | 4.69 | 2.00 |

****

**For 15% Test Data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| KNN | R2score | MSE | RMSE | MAE |
|  | 0.85 | 18.04 | 4.24 | 1.90 |

****

**For 15% Test Data**

|  |  |  |
| --- | --- | --- |
| Cross validation | K values | Average score |
| 10 | 3 | 0.8310 |
| 10 | 4 | 0.8317 |
| 10 | 5 | 0.8240 |
| 5 | 3 | 0.8226 |
| 5 | 4 | 0.8209 |
| 5 | 5 | 0.8141 |

# For Cross Validation=10 and k=3

# 

# For Cross Validation=10 and k=4

# 

# For Cross Validation=10 and k=5

# 

# For Cross Validation=5 and k=3

****

# For Cross Validation=5 and k=4

# 

# For Cross Validation=5 and k=5

# 