MOBILE PRICE RANGE PREDICTION

**AIM:** classify the range of mobile prices based on some features.

**INTRODUCTION**: it is a very popular machine learning project. In this project estimate the price of mobiles his company creates. In this competitive mobile phone market, you cannot simply assume things. To solve this problem we are collects sales data of mobile phones of various companies. I have implemented a Mobile Price Prediction using different Machine Learning Algorithms. This project will classify the price range of the mobile price. The price ranges from 0-3. We’ll discuss the price range in the dataset. It's a classification problem. Now I have trained a mobile price classification using 3 ML algorithms. This model classifies the range of the mobile based on the different parameters like from camera, touch screen, cores, battery, clock speed, internal memory, battery capacity, etc. After training the model using 3 algorithms, I compared all the models using the graph.

**In this problem you do not have to predict the actual price but a price range indicating how high the price is**.

**FEATURES:**

1. battery\_power : Total energy a battery can store in one time measured in mAh
2. blue : Has bluetooth or not
3. clock\_speed : speed at which microprocessor executes instructions
4. dual \_sim : Has dual sim support or not
5. fc : Front Camera megapixels
6. four\_g : Has 4G or not
7. int\_memory : Internal Memory in Gigabytes
8. m\_dep : Mobile Depth in cm
9. mobile\_wt : Weight of mobile phone
10. n\_cores : Number of cores of processor
11. pc : Primary Camera megapixels
12. px\_height : Pixel Resolution Height
13. px\_width : Pixel Resolution Width
14. ram : Random Access Memory in Megabytes
15. sc\_h : Screen Height of mobile in cm
16. sc\_w : Screen Width of mobile in cm
17. talk\_time : longest time that a single battery charge will last when you are
18. three\_g : Has 3G or not
19. touch\_screen : Has touch screen or not
20. wifi : Has wifi or not
21. price\_range : This is the target variable with values of 0(low cost), 1(medium cost), 2(high cost) and 3(very high cost).

**STEPS TO PREDICT THE MOBILE PRICE RANGE.**

1.LOAD THE DATA

2.ANALYIZE AND VISUALIZE THE DATA SET

3.MODEL TRANING

4.MODEL EVALUATION

5.TESTING THE MODEL

**EXPLANATION:** will go through step by step.

STEP1: Loading the Data:

To load the data we have to import necessary packages like numpy, pandas, seaborn, %inline matplotlib .

* Numpy will be used for any computational operations.
* we’ll use matplotlib and seaborn for data visualization.
* we use pandas for loading data from various sources like local storages , databases , excel , csv files ……etc
* Next we load the data using pd.read\_csv() . and set the column name as for the iris data information
* pd.read\_cvs reads CSVfiles.csv transfer comma separated value
* df.head() only shows the first five rows from the dataset table -All the numeric values are in cm.

**STEP2: ANALYSIZE AND VISUALIZE DATASET:**

* Let’s see information.
* Before that we have done some basic operations like describe (). , info () , head functions and tail functions droping function , indexing and slicing ..etc
* Later we prepare the data in a perfect manner without any null values that means we cleared the data
* we have built some model for the dataset based on that data we do some basic operations and analyzed the data
* After that based on analization we visualized the data with seaborn pairplot,matplotlib….etc

**ABOUT DATASET**: **someone has started his own mobile company. He wants to give tough fights to big companies like Apple, Samsung etc.**

**He does not know how to estimate the price of mobiles his company creates. In this competitive mobile phone market you cannot simply assume things. To solve this problem he collects sales data of mobile phones of various companies.**

**he wants to find out some relation between features of a mobile phone(eg:- RAM,Internal Memory etc) and its selling price. But he is not so good at Machine Learning. So he needs your help to solve this problem.**

**STEP3: ANALYZING THE DATASET**

* Firstly to visualize the whole dataset we used seaborn pairplot method it plot the whole dataset information
* Now we ploted the avg of each features of each class
* Later we separated features from target values and calculated the average of each features of all classes, we use pairplot, jointplot ,histogram ,point plot ,boxplot we used to visuvalize the data.

**STEP4: MODEL EVAULATION:**

We trained the data by using the featuers of mobile. By using plots we analyze the data based on comparisions and we implement knn algorithem for knowing k value and nearest values. Based on k value we find the mobile price ranges.

**TESTING THE VALUES:**

After training the data we test the data by using logistic regression and decision making and decision trees. Then, We testing the data. In the way of testing we predict the values. After prediction we get the accurate value**.**

**TO CALCULATE AVERAGE OF EACH FEATURES**

I worked on a dataset , we get accurate values means by prediction of featuers we get mobile

Cost range. From the output we get accurate value, like based on ram, weight, like other featuers we pay some extra money.

Here we learned the machine learned concepts like linear regression making, knn algorithm, decision tree, logistic regression in machine learning.

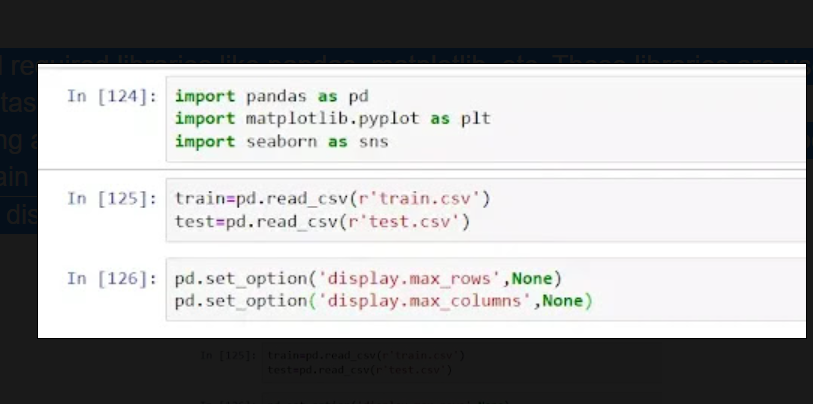
**Based on our data set when we are visualizing the data with the help of sea born we predict the value based on featuers. So we get the accurate data.**

**Steps to predict the mobile prizes:**

**Step:1** First of all, import all required libraries like pandas, matplotlib, etc. These libraries are used to load, preprocess and visualize the dataset.

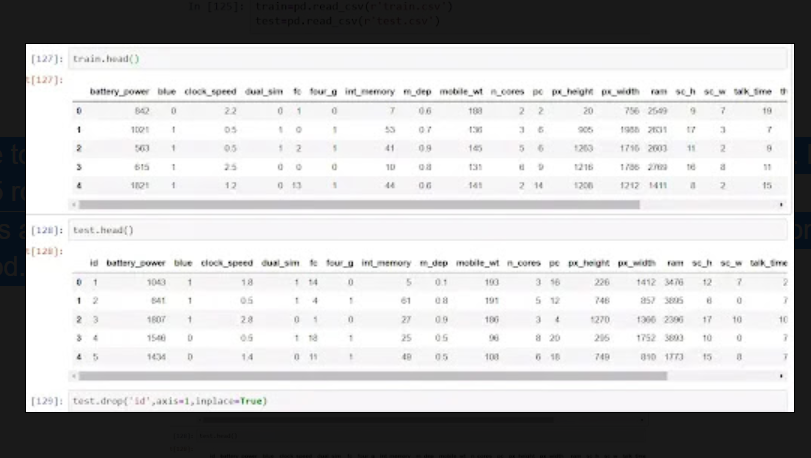
Then load the training and testing dataset using the read\_csv function of the pandas module and store into the separate variable train and test.

Now write a code to display the all number of rows and columns.



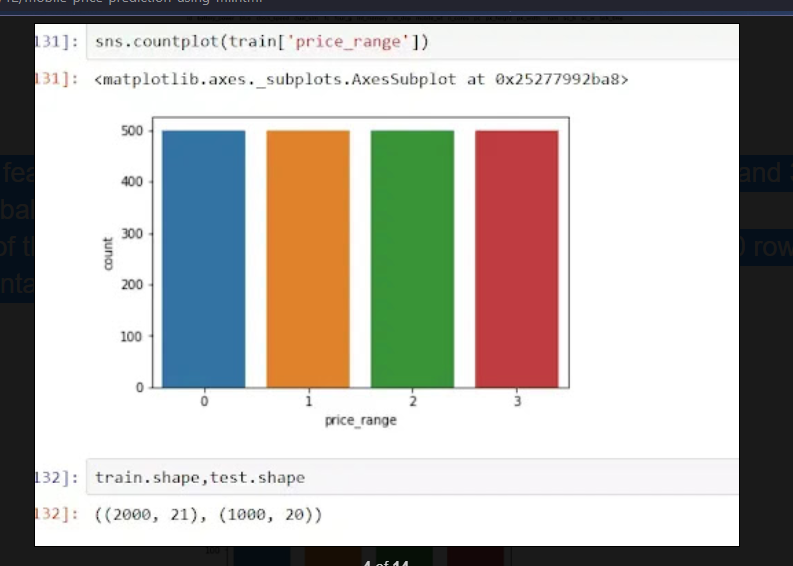
**Step2:** Now this displays the top 5 rows of training and testing data set using the head function. Head function by default displays the 5 rows. We can increase it by passing an integer value.

Test data set contains an extra feature “id” which is of no use, so drop the “id” feature from the testing data set using the drop method.

****

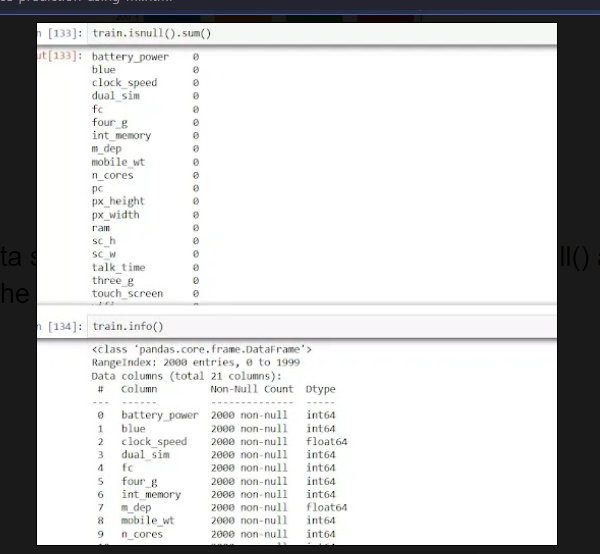
**Step3:** Now display our target feature “price range” and as you can see in the graph, 0, 1, 2 and 3 all have 500 rows. It means there is no imbalanced dataset.

Now check the shape of the data set. As you can see, the train dataset contains 2000 rows and 21 columns and the test dataset contains 1000 rows and 20 columns.

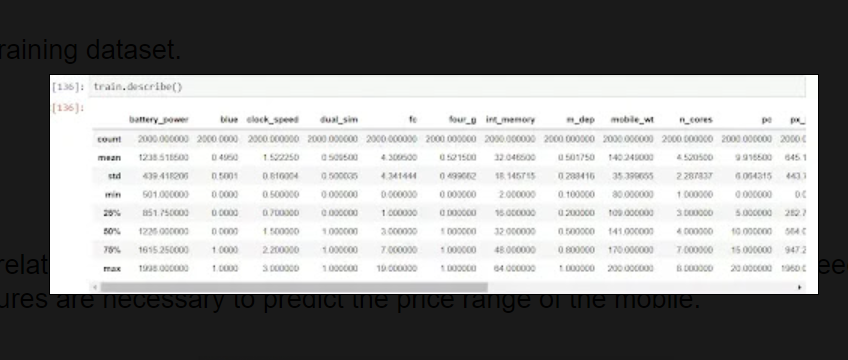
****

**Step4:** Now check that the training data set contains null values or not using the isnull() and sum() method.

Now check the information of the data set using the Info() method.

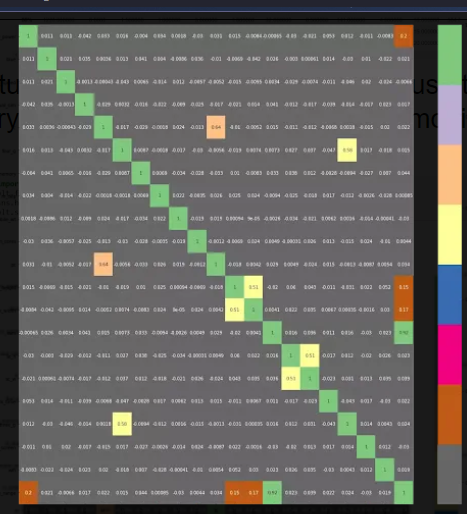


**Step5:** Now describe the training dataset.

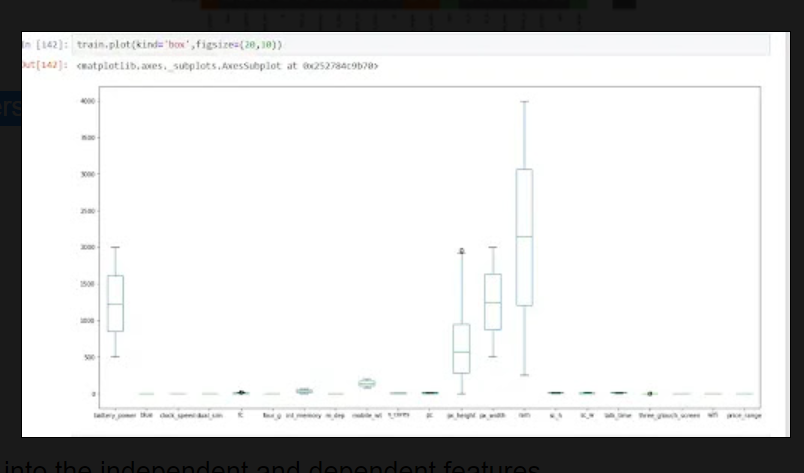
****

**Step6:** Now check the correlation of the features for the knowledge only. Because there is no need to check correlation. All features are necessary to predict the price range of the mobile.





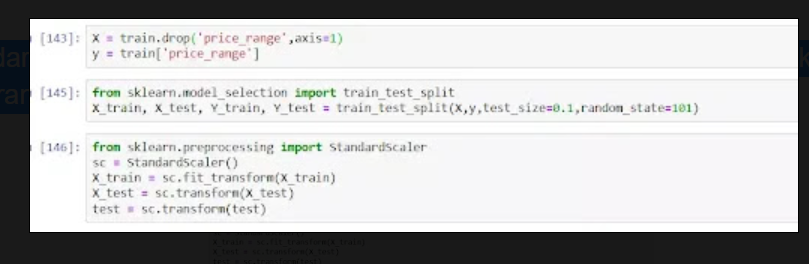
**Step7:** Then check the outliers in the dataset. But no outliers are present.

****

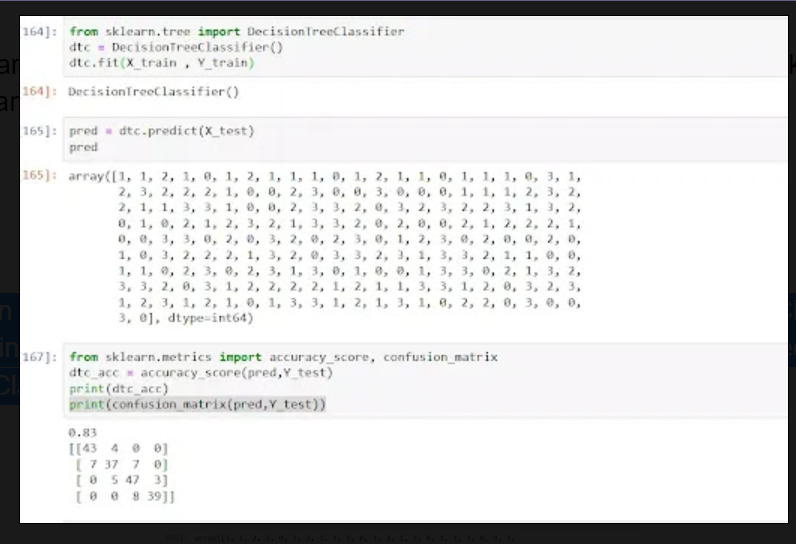
**Step8:** Now split the dataset into the independent and dependent features.

Then split the dataset into the training and testing to evaluate the model using the train\_test\_split method.

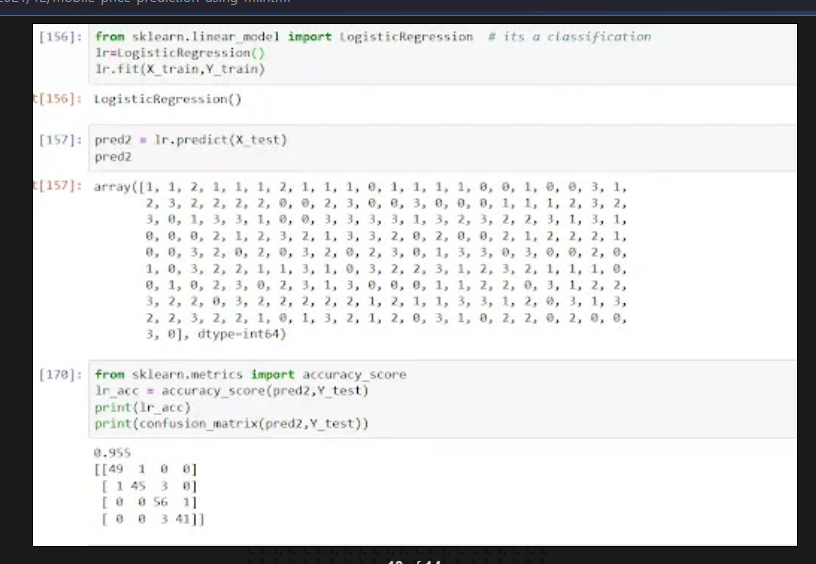
Then apply the standardization on the training and testing datasets. Standardization makes all the features' value in a particular range (0-1).



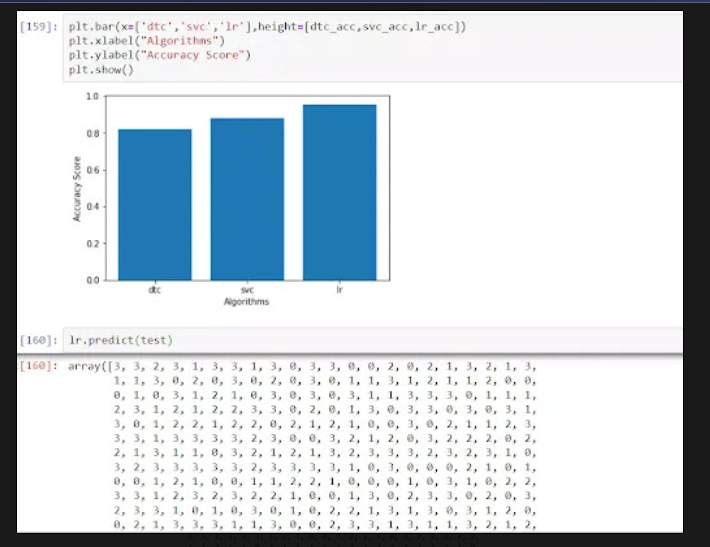
**Step9:** Now load the Decision Tree Classifier from sklearn library and define the Decision Tree Classifier and train with the X\_train and Y\_train dataset. Then test the model using the X\_test dataset. Then check the accuracy score of the Decision Tree Classifier. As you can see, the accuracy score is approx 83%.

****

**Step10:** Now load the Support Vector Classifier and define the SVC and train with the X\_train and Y\_train dataset. Then test the model using the X\_test dataset. Then check the accuracy score of the Support Vector Classifier. As you can see, the accuracy score is approx 88%.

****

**Step11:** Now visualize the accuracy score using the bar plot method of matplotlib. The final best performing model is Logistic Regression. Then test Logistic Regression with a separate test dataset.

****

**Conclusion:**

In this project I have predict the mobile price ranges . and using some ml algorithms. Testing the datasets and also finally predict the ranges.

That ranges is from 0-3 accuracy. And based on mobile features prices are changed.

**WORK DONE BY:**

**Rajeswari Davala**

**Satish Gudala**