

**COMSATS UNIVERSITY ISLAMABAD,**

**WAH CAMPUS**

|  |  |
| --- | --- |
| Subject:- | Machine Learning |
| Topic:- | Mobile Price Classification |
| Class/Section:- | BSE/2A |
| Submitted By:- | Muhammad Uzair(042) |
| Submitted to:- | Dr.Khalil Afzal |

Mobile Price Classification

# Introduction

## Introduction & Background Information.

In this era, mobile phone is necessary thing and mobile companies burn mid night oil to reach at peek mobile phone development. Due to this we have analyze and classify the existing mobile data for launching new cell device. And using this analyses and prediction mobile phone companies build his device accurately according to market level and compete his product as higher as possible.

* I share the link of my GitHub repository of this project on Reference section (please check my GitHub repository to visualize the train and test dataset and screenshots).

## About Dataset.

I am Android and IOS application developer. I very need to know which mobile phone is up on market that’s why I google something about mobile phone device and After scrolling; I found the data set of “Mobile Price Classification Data”. That’s upload by Abishek Sherma in 28 January 2018. It contains 21 columns of test and 21 columns of train dataset.

Data set contain all physical features of mobile devices including height, width, memory-size, ram etc. and mobile prices also.

* I share the link of dataset repository on Reference section.

# Statement of the Problem.

Bob has started his own mobile company. He wants to give tough fight to big companies like Apple, Samsung etc.

He does not know how to estimate 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.

Bob 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.

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

## Significance of the project.

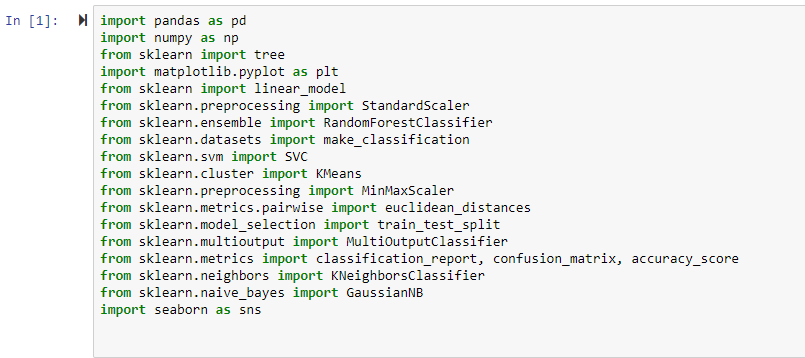
The Significance of this project is to show the current data of mobile devices in visual form for launching new device and make a future prediction for new devices and compare the features with prices.

Methodologies and algorithms.

* Linear regression
* Standard scaler
* Decision Tree Classifier
* Random Forest Classifier
* Support vector machine
* KMeans Cluster
* GaussianNB
* KNeighborsClassifier

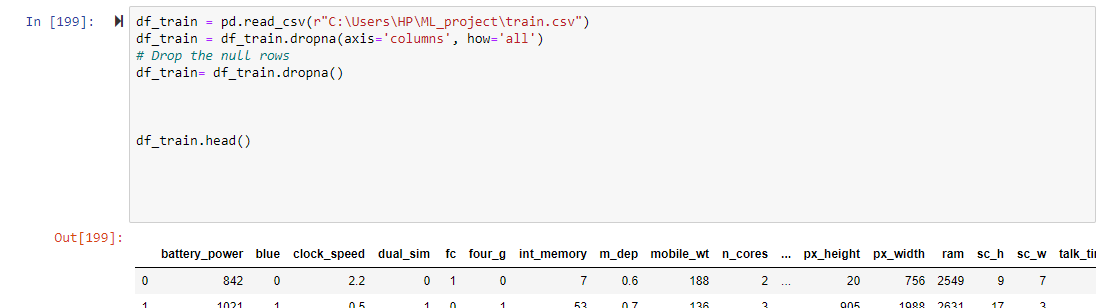
I Explain one by one using screenshots and description.

Here below are my all the libraries that I used for performing my mobile classification project.

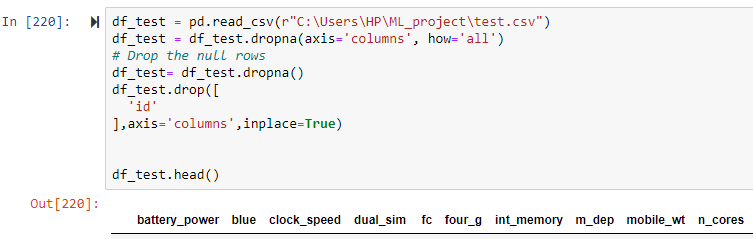


Firstly I import my training and testing dataset and using dropana() function to drop all the null rows and clean the dataset from Null rows. And use “r” in read\_CSV function to read the CSV file properly

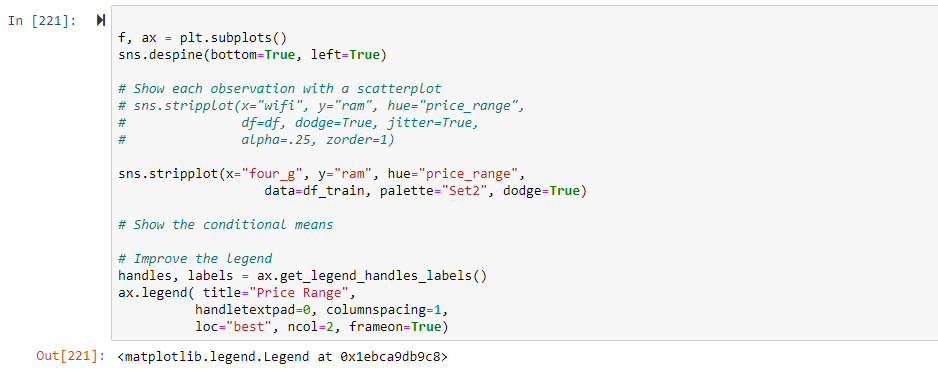
Here is my importation of Training Data-set.



Here is my importation of Testing Data-set.



At first, I evaluate the mobile phone price rang according to four\_g cell phones. That’s why I create a strip plot method of plotting with a different colors.





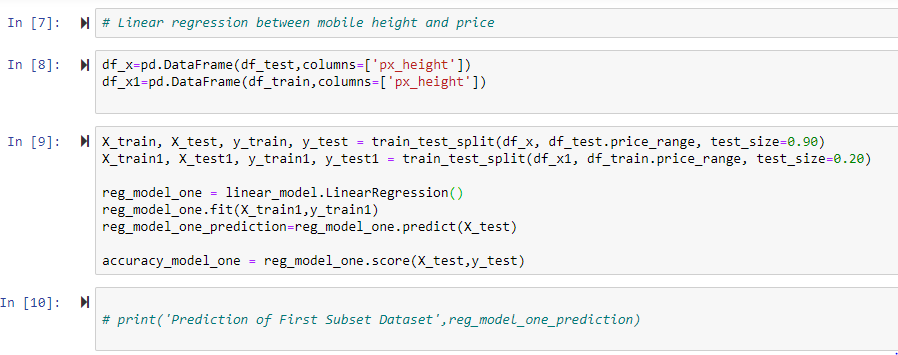
## Linear Regression.

I implement the linear regression algorithm for prediction, accuracy and also finding the regression line between accuracy and prediction of model of price range of mobile phone and height of cell phone.

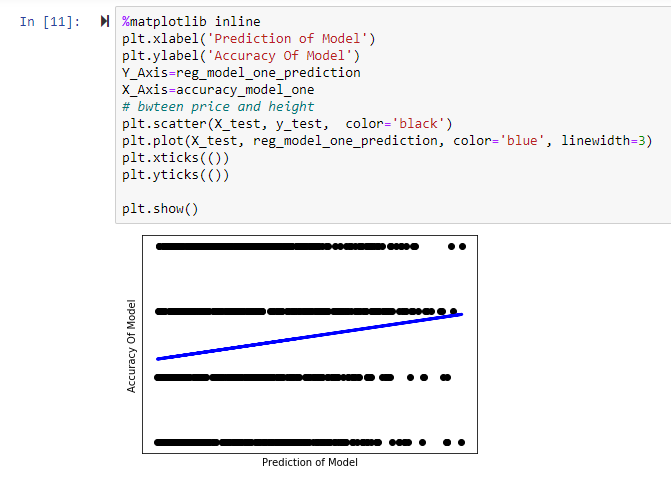
I use train test split method for training data and set testing size is 20 percent then I also use this same method for my testing data and set the testing size is 90 percent.

X train and Y train pick from training dataset.

X test and y test from testing dataset.



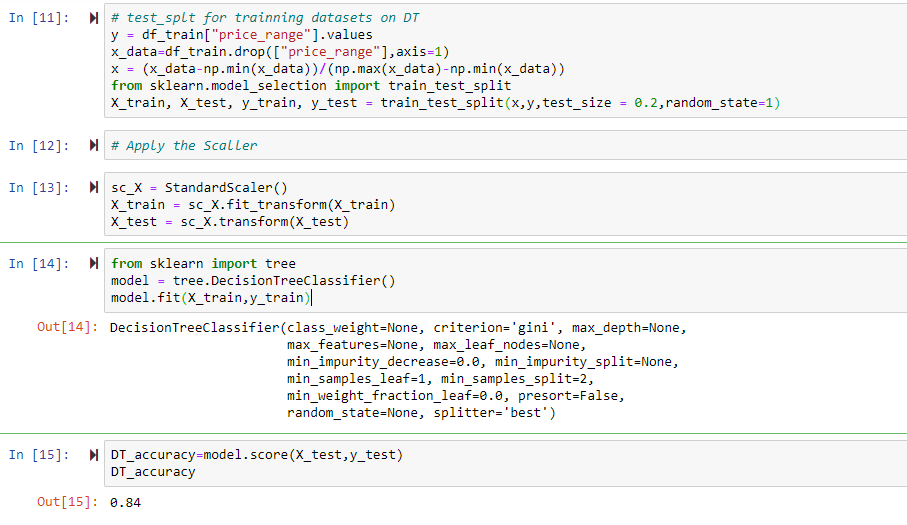
After finding the accuracy and prediction. I plot the regression line between prediction and accuracy of height and price.



## Decision tree Algorithm.

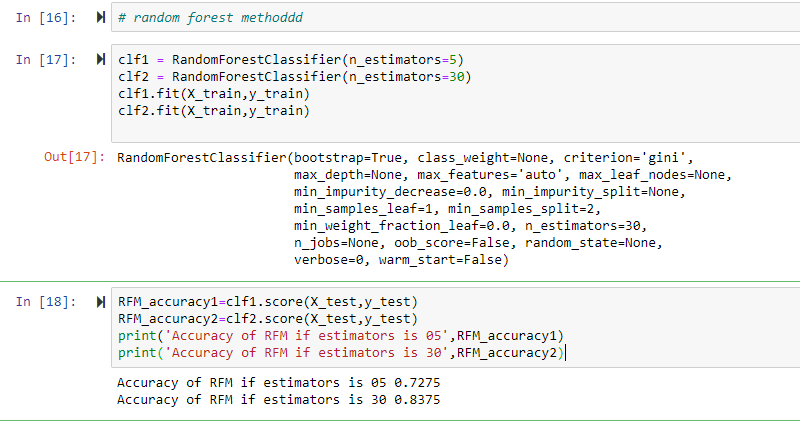
I split the data using test train split method of price range because I implement the decision tree on price range and set the price range as a target or y on test and split method.

Then After splitting I apply the Standard Scaler method to transform or scale my training and testing data in case of mean and multivariate data. After fit\_transformation on training and simple transformation on testing data set; I find the accuracy of my model.

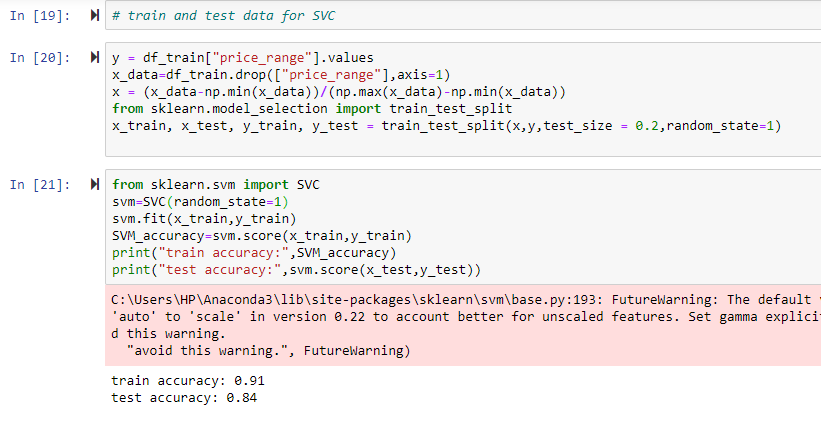


Then in tree library I used random forest method to check the accuracy difference to decision tree and changing n\_estimators for accuracy difference.

Firstly I set the n\_estimator to 5 and then change into 30. The accuracy difference shown below.

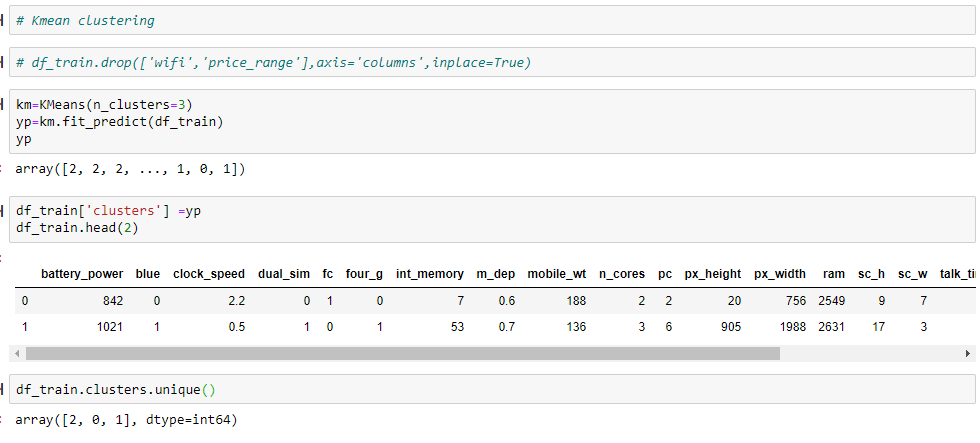


## Support Vector Machine.

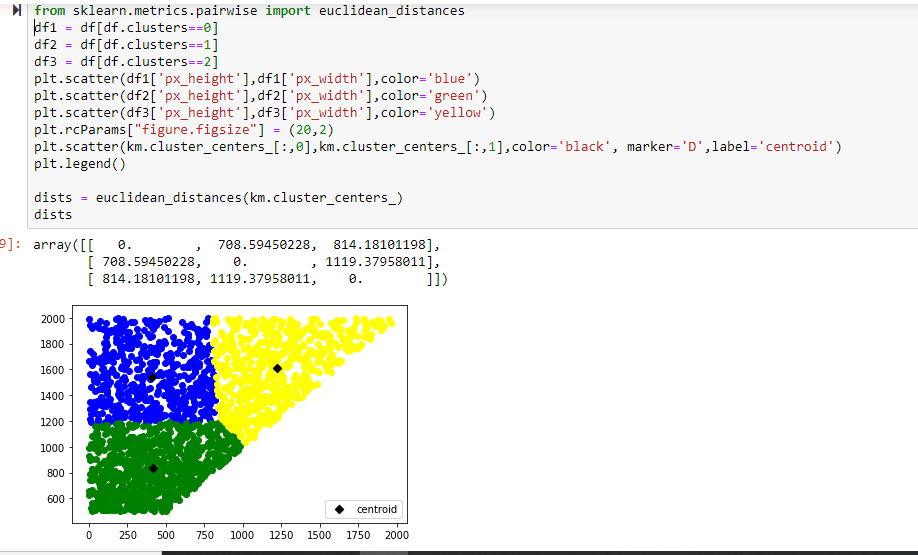


## KMEAN Clustering Algorithm.

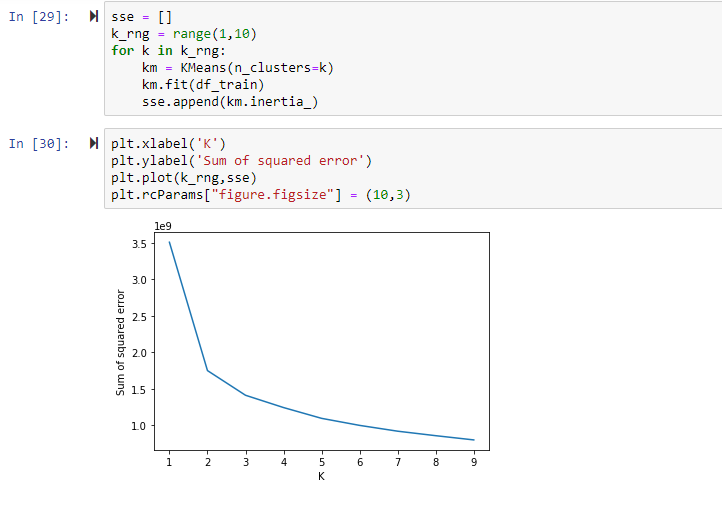
* After performing support vector machine, I also apply the KMean clustering algorithm to find the clustering and its centroid points that show in below figure.
* Firstly I create the Three clusters from my training datasets and then add this clustering column to my original training data sets.
* The I use a unique () function use to know about how many cluster I made. I show the unique values of cluster column.



* After making the cluster I create a data frames of these clusters and performing the scatter plot between mobile height and mobile device width using different colors.
* Then I Cluster center function that input the X and Y value of cluster center and label as centroid on graph.
* And I also find the Euclidean distances between cluster adding cluster centers.

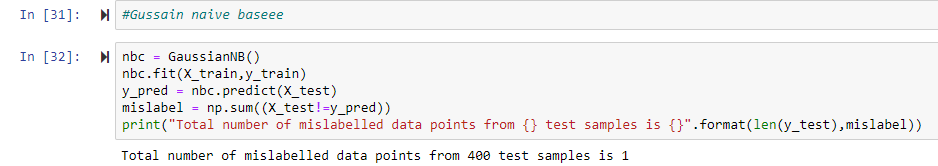


After successful application of KMean algorithm and clustering. I made a decision to find the sum of square error of my training dataset. And I set the K range in a graph 1 to 10.



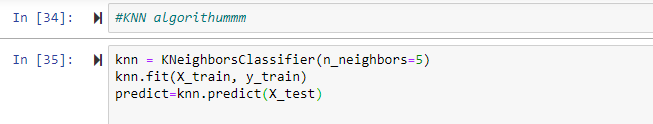
## Gaussian Naive Bayes.

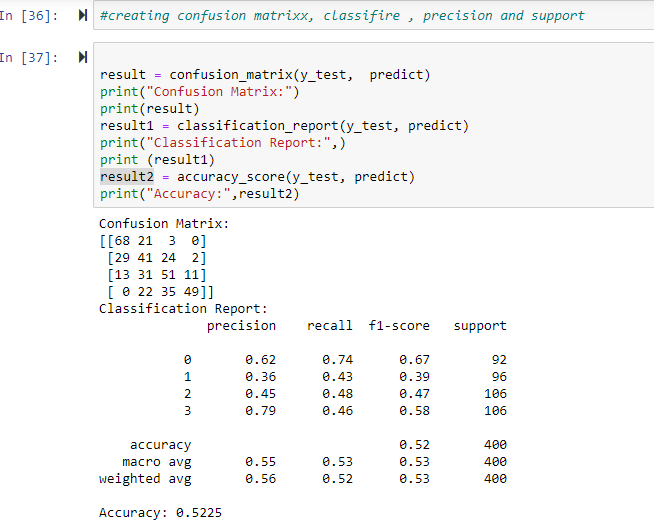
I used the Gaussian naive bayes for making the prediction and then using this prediction value I calculate the total mis-labelled data points from my test dataset.



## KNN algorithm.

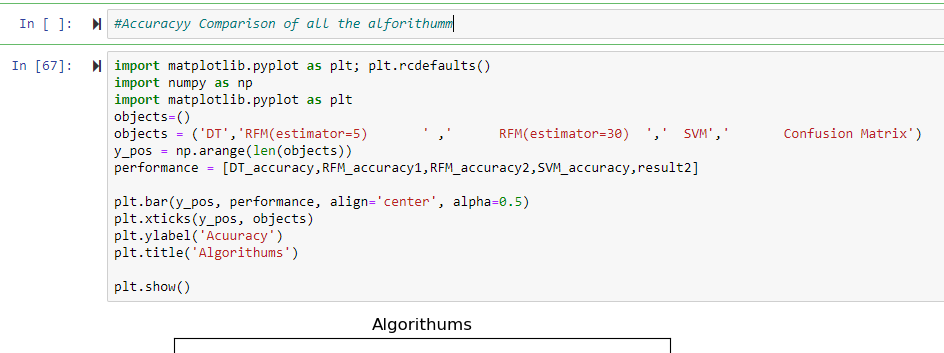
I apply the KNN algorithm and find the prediction, accuracy and also find the confusion matrix and classification report that contain the precision, recall and support and then at the end tell us about the total accuracy that obtained from the confusion matrix.

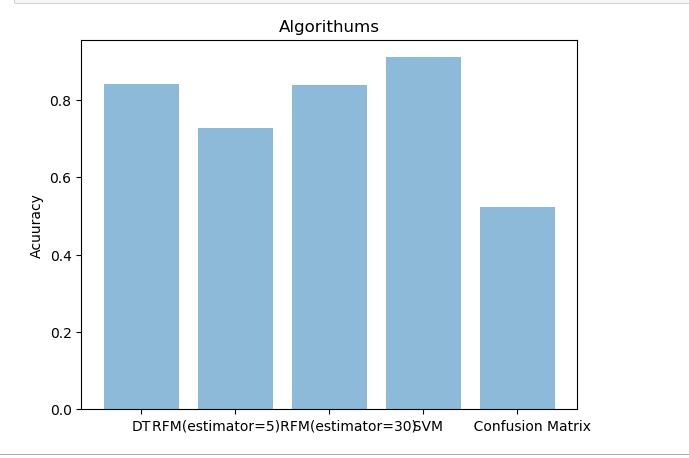




## Conclusion of All Algorithms.

At the end of my project I want to show you my model accuracy of all the algorithms by using bar graph.





## References:

### My GitHub repository link of this project:

https://github.com/Muhammad-uzair029/ML\_project.git

### My dataset repository link of this project:

<https://www.kaggle.com/iabhishekofficial/mobile-price-classification>

### Helping material collected by:

<https://codebasicshub.com/tutorial/machine-learning/support-vector-machine-svm>

<https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>

### For Error solving:

<https://stackoverflow.com/questions/tagged/machine-learning>

<https://github.com/>

# The End