# MOBILE PRICE CLASS CLASSIFICATION AND PREDICTION.

*A project report submitted to ICT Academy of Kerala*

*in partial fulfillment of the requirements*

*for the certification of*

**CERTIFIED PROFESSIONAL**

**IN**

**DATA SCIENCE & ANALYTICS**

submitted by

**Zareena Basheer Palakkal**

A close up of a sign

Description automatically generated

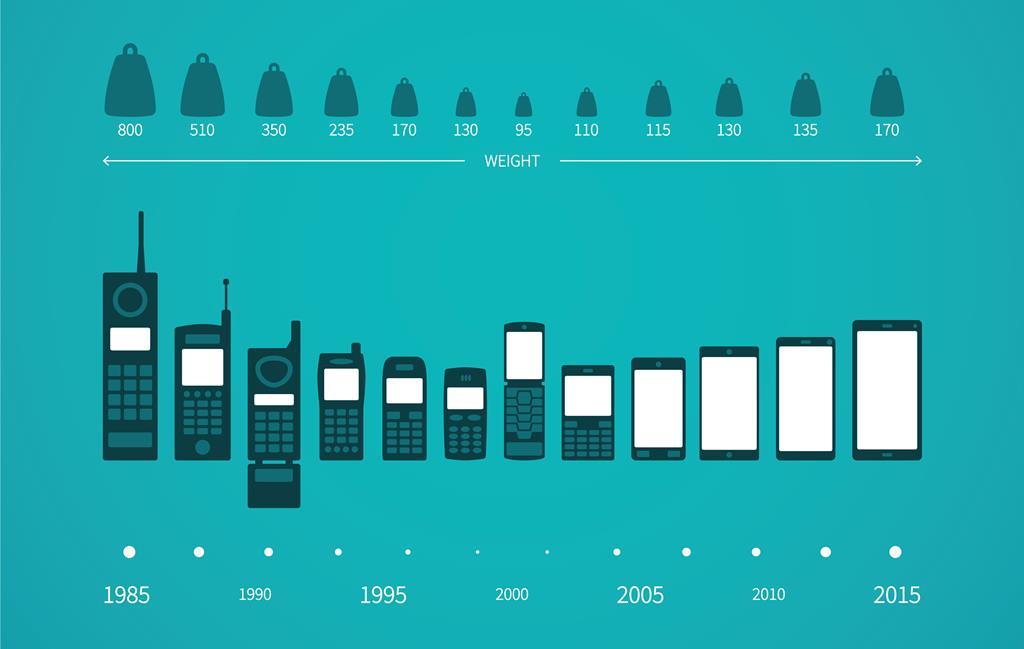
**ICT ACADEMY OF KERALA**

**THIRUVANANTHAPURAM, KERALA, INDIA**

**DECEMBER 2019**

**ABSTRACT**

One of the most selling and purchasing device is Mobile Phones. With each newer versions and features added to mobile phones, their prices cannot be simply assumed. The selling price of mobile phones plays a very important role in their effective marketing and business. So the relation between different features of a mobile phone (example – Random Access Memory, Battery Power, Screen Size etc) and its selling price is found out. In this paper, a classification algorithm classifies the selling price of mobile phones into various ranges, indicating how high the price is and do not predict their actual price. To achieve this, the price ranges or the sales details are collected from various mobile phone companies. The conclusion is that the selected classification algorithm can predict or find out the selling price range of an entirely new mobile phone, whose required features are given. This is a powerful tool which can be extended to other type of marketing and business.



**TABLE OF CONTENTS.**

**Introduction.**

**1. About the problem**.

**1.1.** What is the problem?

**1.2.** Why is the problem?

**1.3.** How can solve the problem?

**2. Data Analysis.**

**2.1.** Business Understanding.

**2.2.** Data Mining.

***2.2.1.*** *Importing Python Libraries.*

***2.2.2.*** *Reading the data.*

***2.2.3.*** *Data Preprocessing.*

**2.3.** Data Cleaning.

**2.4.** Data Exploration.

**2.5.** Feature Engineering.

**2.6.** Predictive Modeling.

***2.6.1.*** *Splitting the data into training set and testing set and Standardizing it.*

***2.6.2.*** *Fitting the models.*

***a)*** *Logistic Regression.*

***b)*** *Random Forest Classifier.*

***c)*** *Decision Tree Classifier.*

***d)*** *SVM.*

***e)*** *Gradient Boosting Classifier.*

***f)*** *KNN. ((a) Minkowski, (b) Euclidean, (c) Manhattan.)*

**3. Predicting the price range of Mobile phones using the best model.**

**4. Outcomes of the work.**

**5. Conclusions.**

**6. References.**

**1.INTRODUCTION**

Price is the most effective attribute of marketing and business. The very first question of costumer is about the price of items. All the costumers are first worried and thinks “If he would be able to purchase something with given specifications or not”. Artificial Intelligence - which makes machine capable to answer the questions intelligently - now a days is very vast engineering field. Machine learning provides us best techniques for artificial intelligence like classification, regression, supervised learning and unsupervised learning and many more. Different tools are available for machine learning tasks like MATLAB, Python, Cygwin, WEKA etc. We can use any of classifiers like, Linear Regression, Logistic Regression, Random Forest, Decision Tree, KNN and many more. Different types of feature selection algorithms are available to select only best features and minimize dataset. This will reduce computational complexity of the problem. As this is optimization problem so many optimization techniques are also used to reduce dimensionality of the dataset.

Mobile now a day is one of the most selling and purchasing device. Every day new mobiles with new version and more features are launched.

Hundreds and thousands of mobile are sold and purchased on daily basis. So here the mobile price class prediction is a case study for the given type of problem i.e. finding optimal product. The same work can be done to estimate real price of all products like cars, bikes, generators, motors, food items, medicine etc.

Many features are very important to be considered to estimate price of mobile. For example, RAM of the mobile. Battery timing is also very important in today’s busy schedule of human being. Size and thickness of the mobile are also important decision factors. Internal memory, Camera pixels, and video quality must be under consideration. Internet browsing is also one of the most important constraints in this technological era of 21st century. And so is the list of many features based upon those, mobile price is decided. So we will use many of above mentioned features to classify whether the mobile would be very economical, economical, and expensive or very expensive.

**PROBLEM STATEMENT**

To make a suitable machine learning algorithm to predict the price of a mobile phone (classify as price range 0, 1, 2, 3) using the given attributes or features of mobile phone.

**Data set Introduction:** Dataset has 22 features and 3000 entries. The meanings of the features:

**Id**: ID

**battery\_power:**Total energy a battery can store in one time measured in mAh

**blue:** Has bluetooth or not

**clock\_speed:** speed at which microprocessor executes instructions

**dual\_sim:** Has dual sim support or not

**fc:** Front Camera mega pixels

**four\_g:** Has 4G or not

**int\_memory:** Internal Memory in Gigabytes

**m\_dep:** Mobile Depth in cm

**mobile\_wt:** Weight of mobile phone

**n\_cores:** Number of cores of processor

**pc:** Primary Camera mega pixels

**px\_height:** Pixel Resolution Height

**px\_width:** Pixel Resolution Width

**ram:** Random Access Memory in Mega Bytes

**sc\_h:** Screen Height of mobile in cm

**sc\_w:** Screen Width of mobile in cm

**talk\_time:** longest time that a single battery charge will last when you are

**three\_g:** Has 3G or not

**touch\_screen:** Has touch screen or not

**wifi:** Has wifi or not

**price\_range: This is the target variable with value of 0(low cost), 1(medium cost), 2(high cost) and 3(very high cost)**

**1. ABOUT THE PROBLEM.**

1.1. What is the problem?

This dataset is all about the price range of different kind of mobile phones. As we know mobile phones are part of human lives now, and development in this field is done day by day, the importance of this kind of dataset is very huge.

Here, we are given by 21 features of different mobiles which may or may not affect the price range of a mobile. Our aim is to classify the price range by using different machine learning algorithms and hence predicting the price range of another mobile dataset with same features by the light of the best algorithm we have found out.

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

1.2. Why is the problem?

If someone has started his/ her own mobile company and he/ she wants to give tough fight to big companies like Apple, Samsung etc., but, 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/ she needs to collect sales data of mobile phones of various companies and do some explorations in it.

So, exactly they want to find out some relation between features of a mobile phone (e.g.:- RAM, Internal Memory etc.) and its selling price.

It is not only for mobile companies but also each individual have the benefits of this kind of classification. Because, by observing this they can decide whether they are giving away their money on their desired mobile phone which has all their expectations.

1.3. How can solve this problem?

In order to solve this kind of problem, we will use ‘Classification’ mechanism. **Classification** is a central topic in [machine learning](https://brilliant.org/wiki/machine-learning/) that has to do with teaching machines how to group together data by particular criteria. Classification is the process where computers group data together based on predetermined characteristics.

A common example of classification comes with detecting spam emails. To write a program to filter out spam emails, a computer programmer can train a machine learning algorithm with a set of spam-like emails labeled as spam and regular emails labeled as not-spam. The idea is to make an algorithm that can learn characteristics of spam emails from this training set so that it can filter out spam emails when it encounters new emails.

Classification is an important tool in today’s world, where data is used to make all kinds of decisions in government, economics, medicine, and more. Researchers have access to huge amounts of data, and classification is one tool that helps them to make sense of the data and find patterns.

While classification in machine learning requires the use of (sometimes) complex [algorithms](https://brilliant.org/wiki/algorithm/), classification is something that humans do naturally every day. Classification is simply grouping things together according to similar features and attributes. When you go to a grocery store, you can fairly accurately group the foods by food group (grains, fruit, vegetables, meat, etc.) In machine learning, classification is all about teaching computers to do the same.

So, here we will classify the price range of mobile phones according to the features given(like ram, battery power etc.). The different classifications are given by 0,1,2,3.

0 - Very low price range.

1 - Low price range.

2 - High price range.

3 - Very high price range.

First we will split our data into train set (to train the machine), and test set (in which the price range is not given, and we need to predict its price range.) and then use different machine learning algorithms to achieve the goal.

**2. DATA ANALYSIS.**

We are going to analyze and interpret the data according to the cycle of data analysis.

**2.1. Business Understanding**

Defining the objectives for the problem that needs to be tackled.

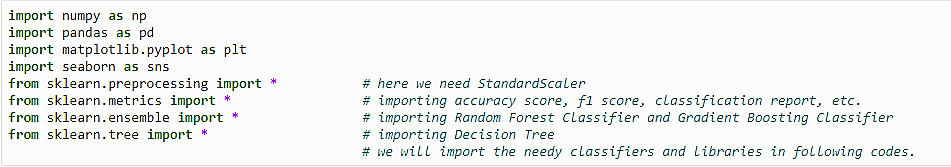
In this project, we are attempt to study following things,

* The features which affects the mobile price in a very high manner.
* How those features varying according to the price range.
* The way this kind of classification be useful for mobile companies as well as common individuals for achieving their goals.

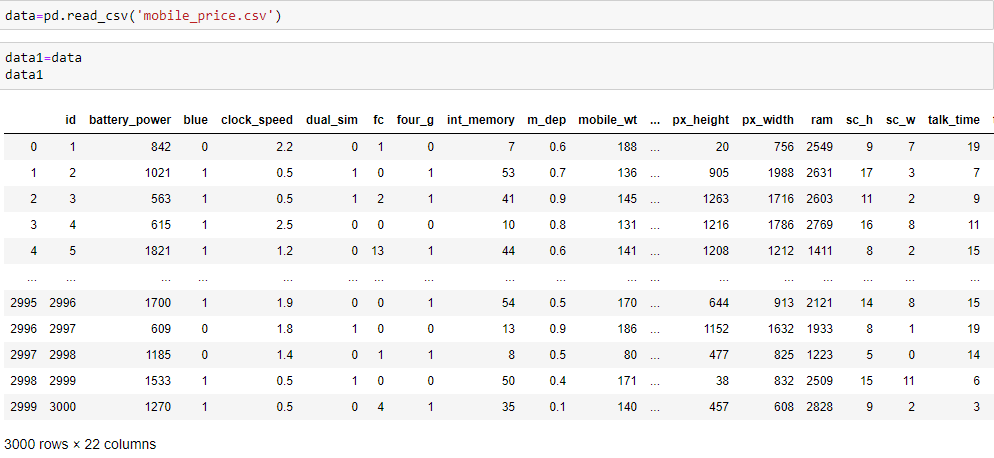
**2.2. Data Mining.**

Gather and scrape the data necessary to the project.

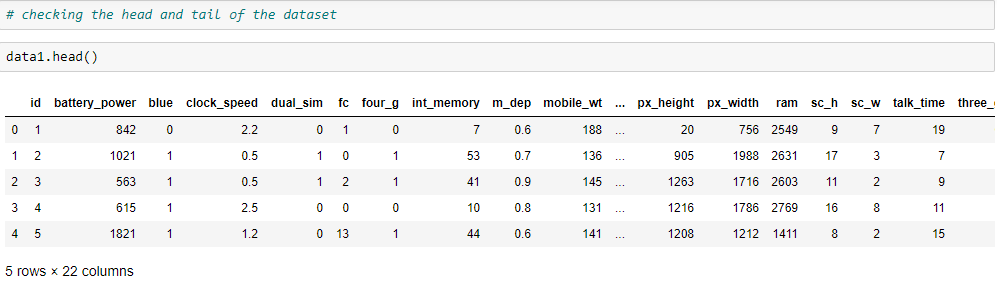
2.2.1. Importing Python Libraries.

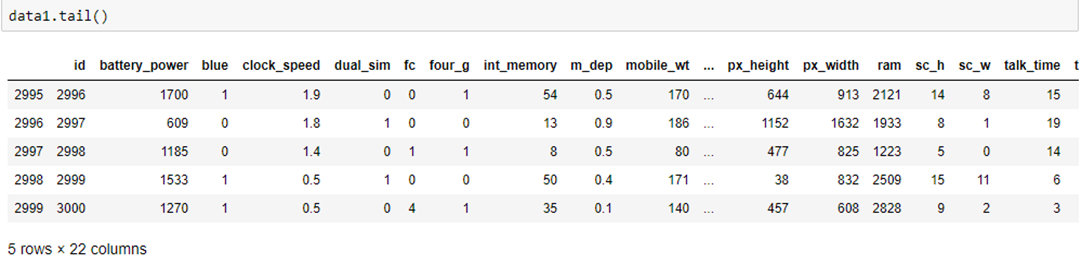


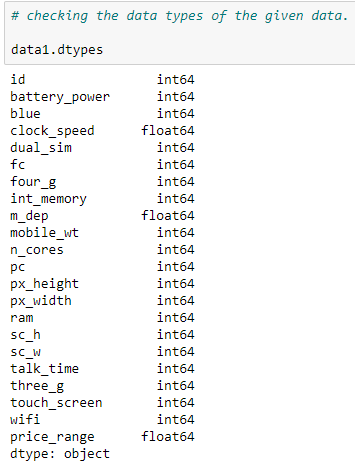
2.2.2. Reading the data.

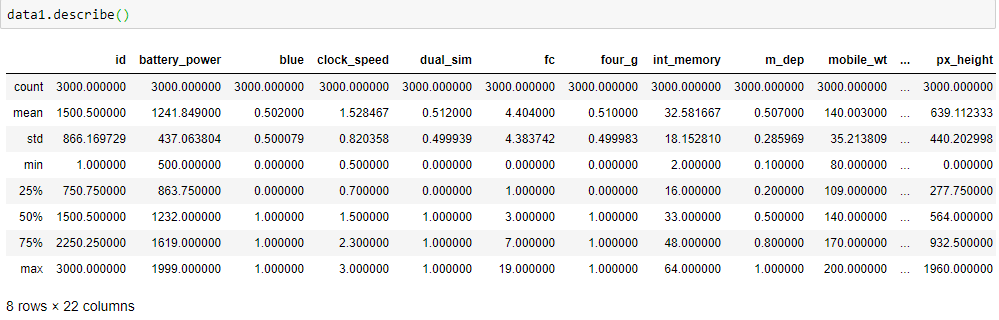
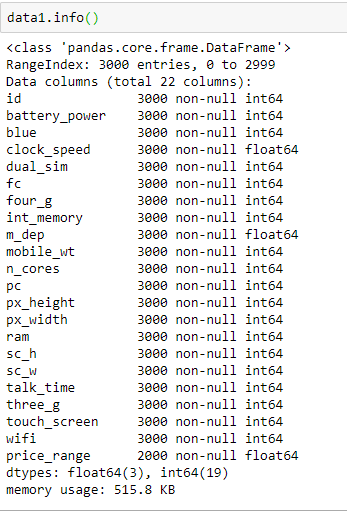


2.2.3. Data Preprocessing.



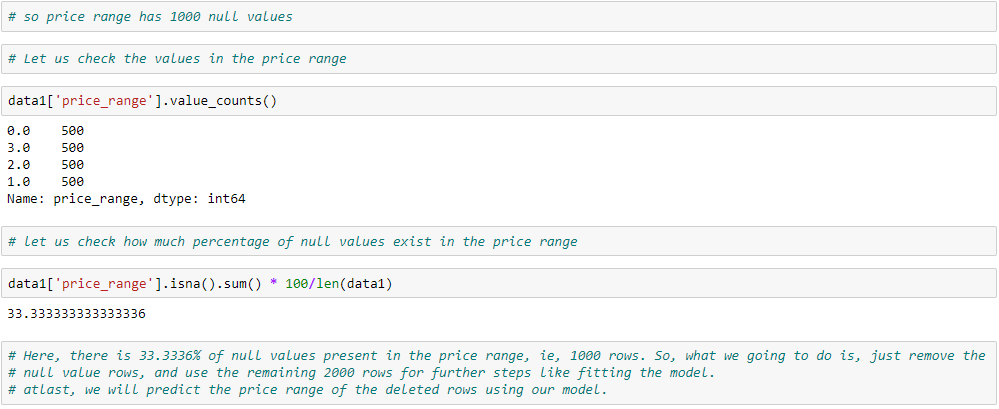
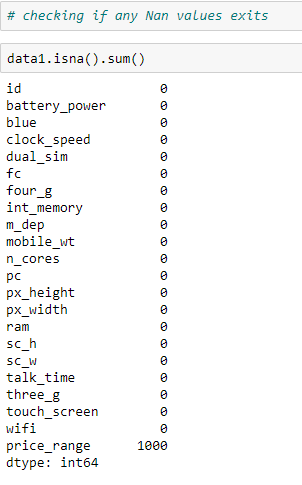


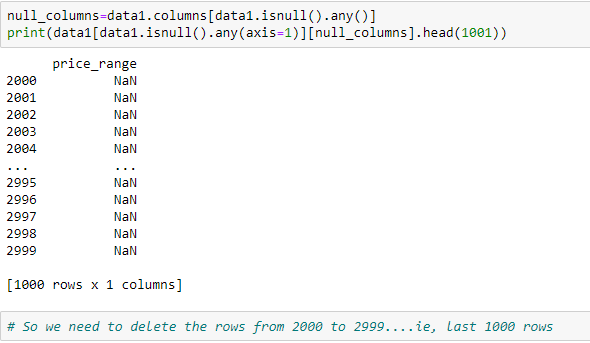


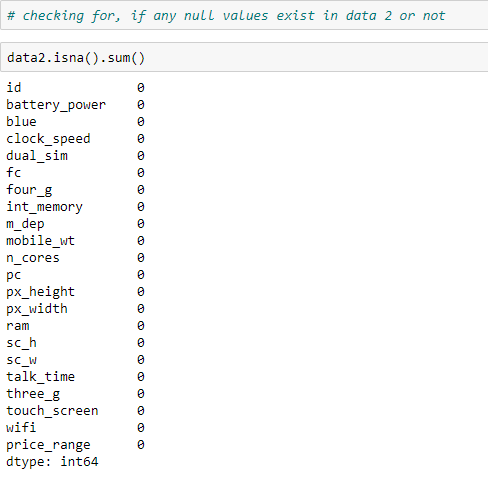
**2.3. Data Cleaning.**

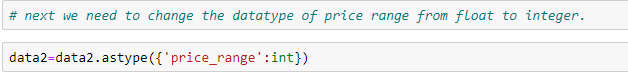
Fix the inconsistencies within the data and handle the missing values.







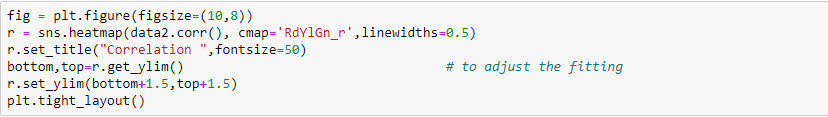


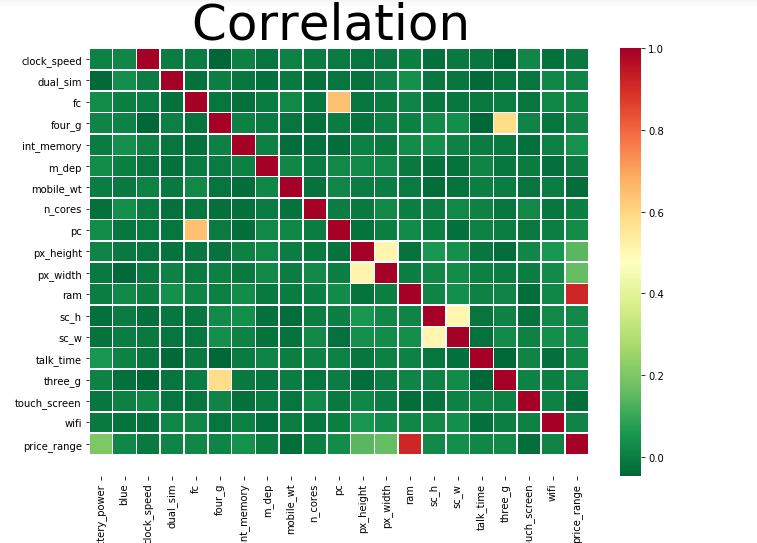


**2.4. Data Exploration.**

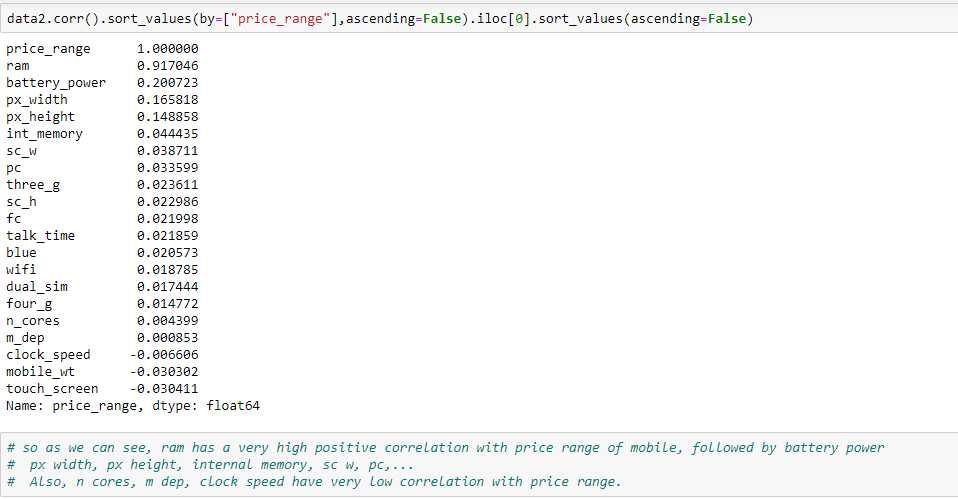
**Visualizing our Data, finding correlation among features and target label (Price Range – 0, 1, 2, 3)**







Correlation values in descending order.



We observe ram has the highest correlation value i.e., as ram value increases price also increases and negative correlation between weight of phone and price i.e., as weight increases price decreases etc.

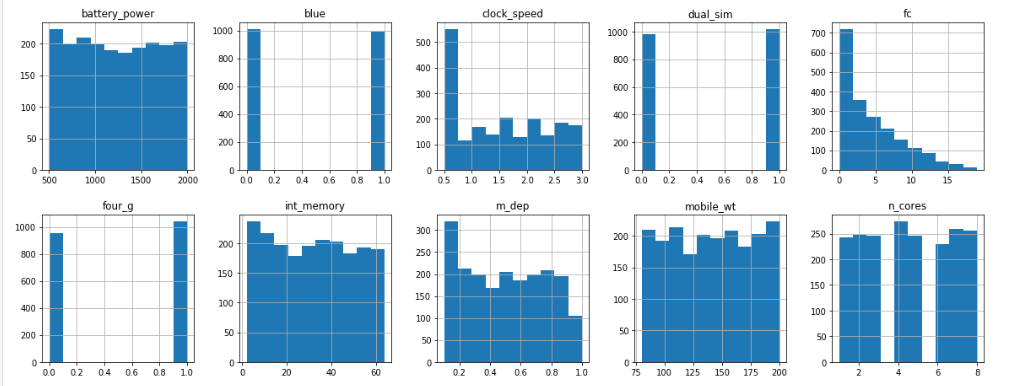
**Also we find the price of mobile phone to be highly depended of the following features in their decreasing order:**

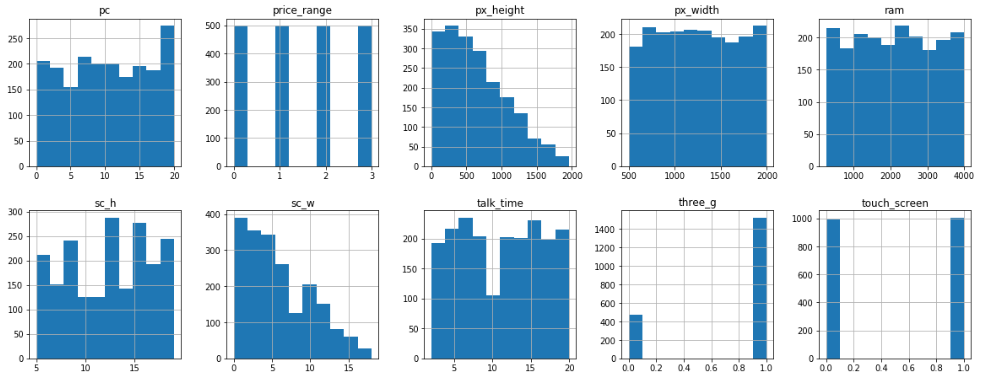
1. **RAM (Random Access Memory) Size of Mobile Phone.**
2. **Battery Power of Mobile Phone.**
3. **Screen Size of Mobile Phone.**

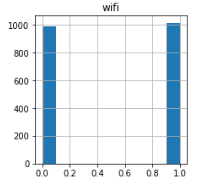
Plotting the distribution of values/ data for all the given features of the mobile phone.

This is done with the help of a histogram. A histogram is an accurate representation of the distribution of numerical data. It is a diagram consisting of rectangles whose area is proportional to the frequency of a variable and whose width is equal to the class interval.









Next let us plot different graphs to depict the relationship between our (target label) price range and other features which are correlated to it.

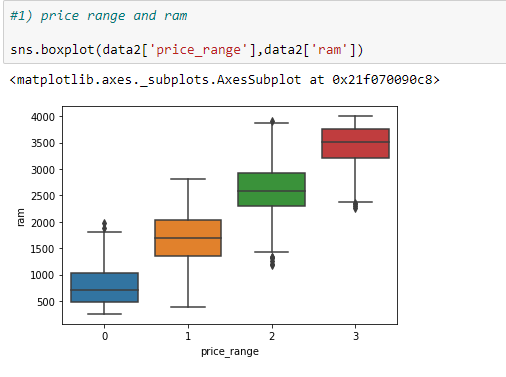
**Box Plot between Price Range And RAM of Mobile Phones**

A box plot—displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum.In a box plot, we draw a box from the first quartile to the third quartile. A vertical line goes through the box at the median. The whiskers go from each quartile to the minimum or maximum.

RAM is becoming a more reliable indicator of a phone’s performance than CPU specs.

RAM stands for Random Access Memory. This tells you any part of the data it stores can be accessed directly. The phone doesn’t have to scan through sequentially-stored data as you might do with a CD, an old tape cassette or, most importantly, a hard drive. It’s effectively instant-access.

RAM is also what lets you flick between apps without having to start from scratch each time. Even if an app isn’t actively doing anything in the background, RAM is used to create a saved state that the phone can zap back into as if you’d never left.

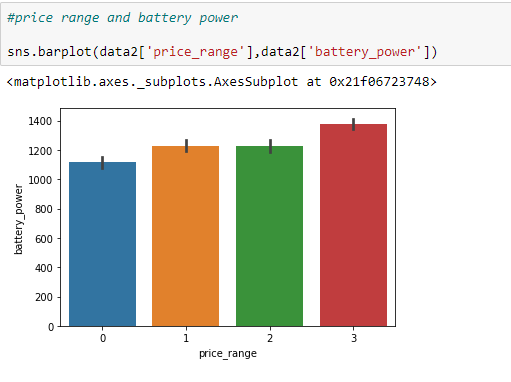


According to price range (0,1,2 and 3), we have :

* In 0 class(low cost) RAM values are changing between 0- 2000 megabytes
* In 1 class(medium cost) RAM values are changing between 0-3000 megabytes
* In 2 class(high cost) RAM values are changing between 1000-4000 mb
* In 3 class(very high cost) RAM values are changing between 2000 and 4000 mb( mostly 3500-4000 mb)

The capacity of the RAM increases as the price\_range also increase. That is the mobile phones falling under the category of highest price\_range ‘3’ has the maximum RAM. And vice-versa, lower RAM is found in mobile phones of price\_range ‘0’.

**Plot between Price Range and Battery Power**

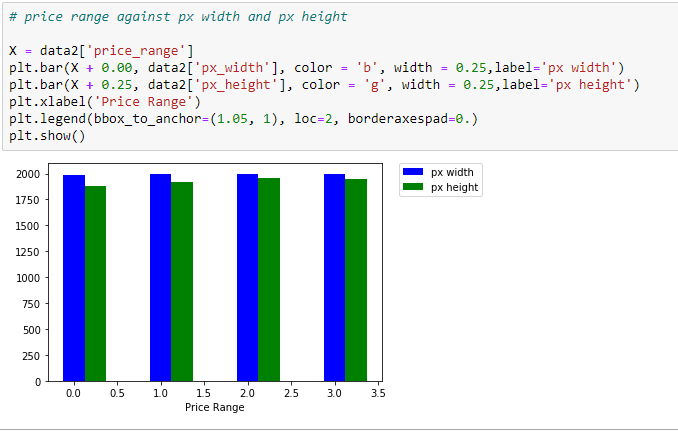


The battery\_power means theTotal energy a battery can store in one time is measured in mAh.

According to price range (0,1,2 and 3), we have :

* In 0 class(low cost) battery power is 1100 mAH. The battery power for 0 class is a really good and supportive feature to purchase that class, when compared with price range 1 and 2 which are higher in prices but offers only a slight increase in battery power.
* In 1 class(medium cost) and in 2 class(high cost) battery power is 1200 mAH. For price range 2 which is higher than price range 1, the battery power is the same. For a buyer considering only higher battery power, can get a better buy/ purchase in the mobile phones of price range 1.
* In 3 class(very high cost) battery power is the maximum .

**Bar Plot between Price Range and Pixel width and Pixel height**

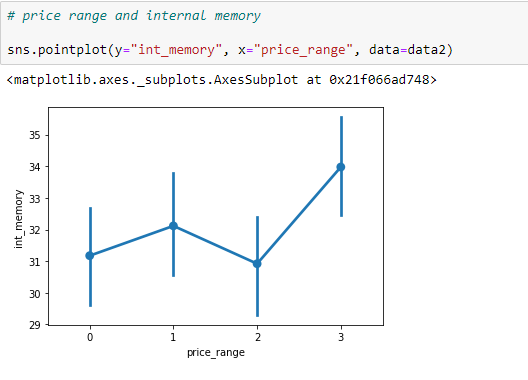


All the mobile phones under the various price \_ranges have the same pixel width.

But pixel height has slight variations, as follows:

* In 0 class(low cost) pixel height is less or lowest of other price ranges.
* In 1 class(medium cost) pixel height is higher or more than 0 class.
* In 2 class(high cost)pixel height is the maximum. Even slightly higher than higher price range 3 class.
* In 3 class(very high cost) pixel height is slightly lesser than in price range 2 class.

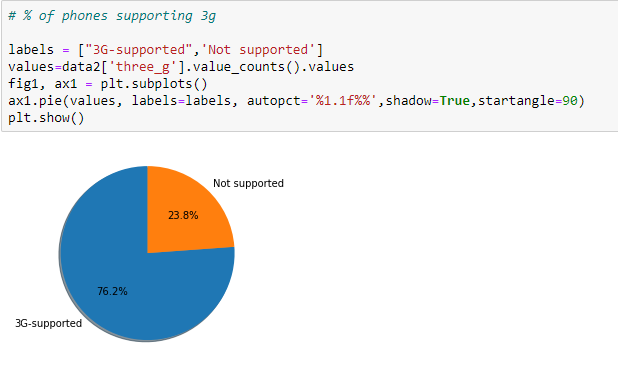
**Plot between Price Range and Inernal memory**



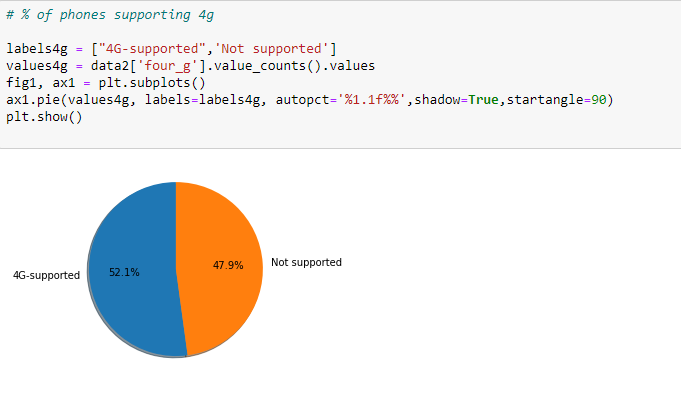
Internal memory of mobile phones in Giga Bytes is as follows:

* In 0 class(low cost) , the total internal memory is of the same size as that of much higher price range class 2. Thus price range 0 class mobile phones offers a good/ substantial amount of internal memory when in comparison with mobile phones of much higher cost, price range class 2.
* In 1 class(medium cost) , the total internal memory is more/ higher than price range class 2. Thus for lower price, more internal memory is obtained in the mobile phones in price range class 1.
* In 2 class(high cost) offers same internal memory as that of price range class 0.
* In 3 class(very high cost) offers maximum internal memory. Thus highest price range mobile phones offer the maximum or largest internal memory of 34 Gigabytes.

**Pie chart of all the mobile phones, for the 3G and 4G facilities.**

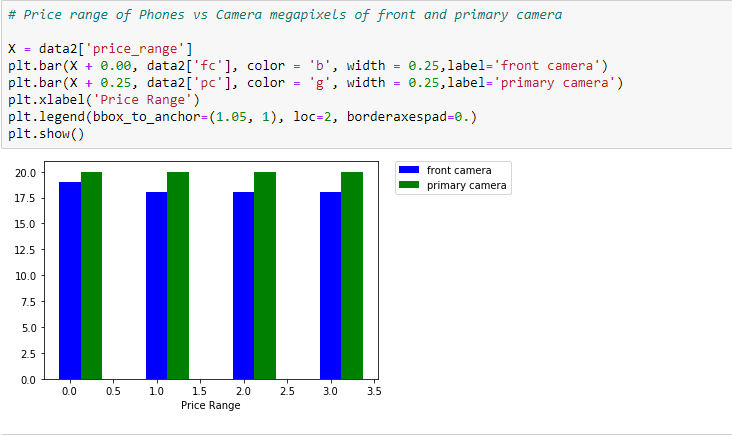


From the above graph it is visible that a huge category (76.2%) of mobile phones, support 3G and only 23.8% of mobile phones, do not support the facility of 3G.



It can be said that, out of all the total, given mobiles phones, nearing half of them do not support 4G facility and the other half (52%) supports 4G facility.

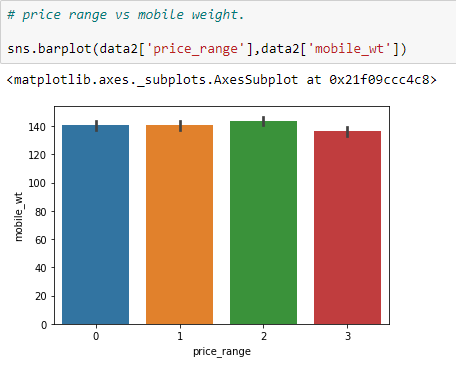
**Bar Graph for price range and available cameras in mobile phone.**



According to price range (0,1,2 and 3) we have :

* (In 0 class low cost) the number of mobile phones with front camera mega pixels is the maximum, than that of all other higher price ranges (1, 2 and 3)
* The front camera mega pixel rate (green bar in graph) is the same for all the price range classes.

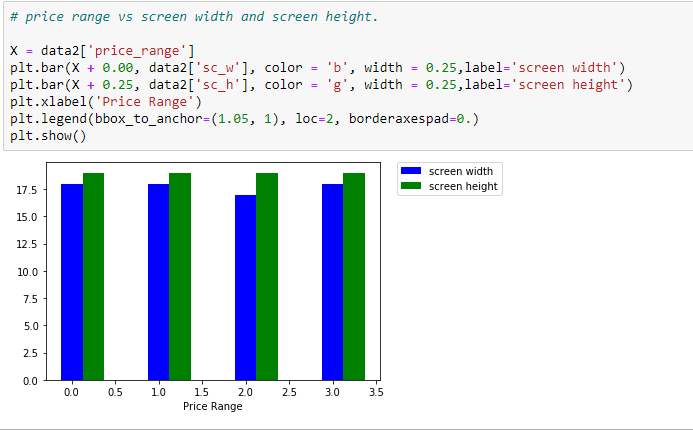
**Barplot showing the relationship between price range and mobile weight.**



According to price range (0,1,2 and 3) we have :

* From class 3 price\_range (highest cost) (red bar) the weight of mobile phone decreases a the price increases.
* Heavier weight mobile phone is for price\_range class 2 (Green bar)
* For price\_range 0 and 1, the mobile phone weights are the same.

**Bar graph of price range and screen dimensions.**

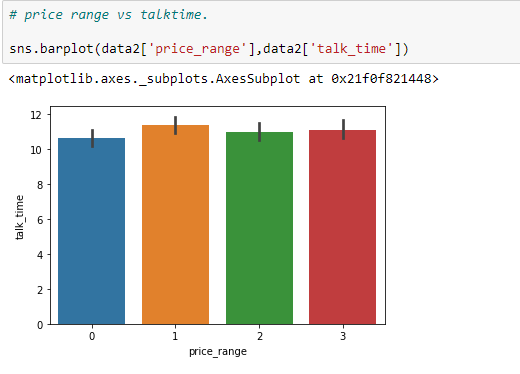


According to price range (0,1,2 and 3) we have :

* The scree width is lowest for class 2 price\_range (higher cost) mobile phones. The screen width is same for all other price\_ranges.
* The screen height is same for all price\_ranges (Green Bar)

**Bar graph of price range and mobile phone talktime.**

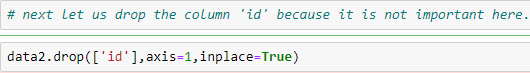
Talk time means the longest time a single battery charge will last when you are using it.



* Talktime is highest for the mobile phones of price\_range 1 (Orange bar in the graph).
* Talktime is lowest for the mobile phones of price\_range 0 (Blue bar in the graph).
* Talktime is slightly more for the mobile phones of price\_range 3 (Red bar in the graph) than the talktime for the mobile phones of price\_range 2 (Green bar in the graph).

**2.5. Feature Engineering.**

Selecting important features and constructing more meaningful conclusions.

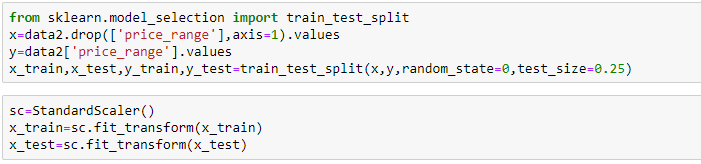


Now our data have 21 columns which include 20 features and the target variable – price\_range.

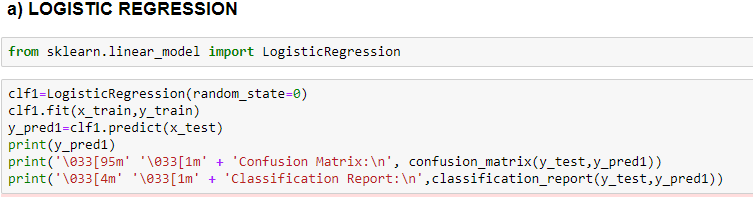
**2.6. Predictive Modeling.**

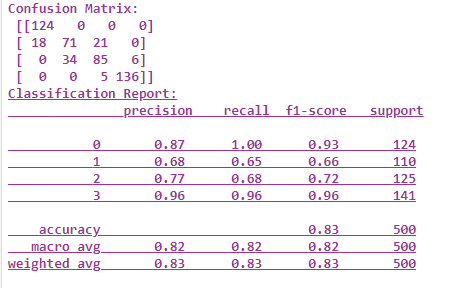
Here, we will train machine learning models, evaluate their performance and use them for predictions.

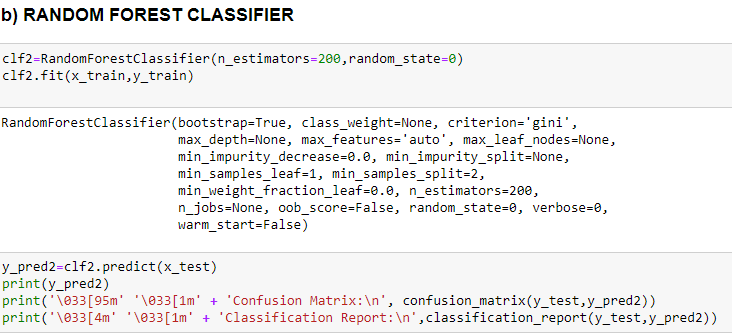
2.6.1. Splitting the data into training set and testing set and Standardize it.

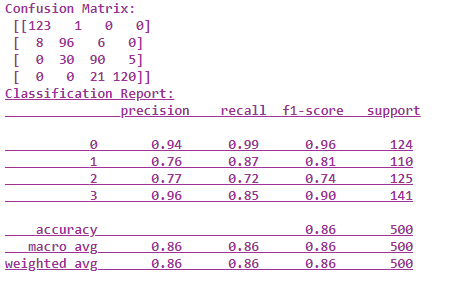


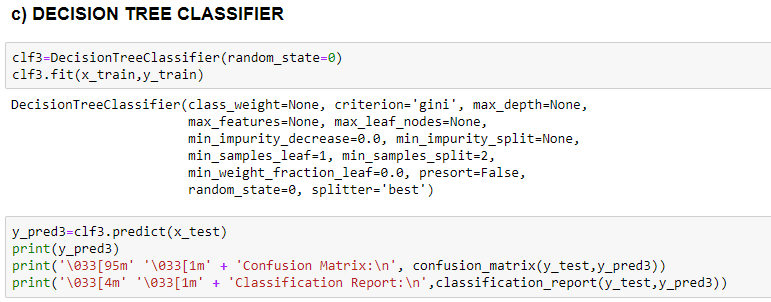
2.6.2. Fitting the models.

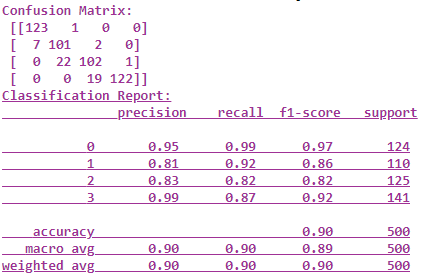
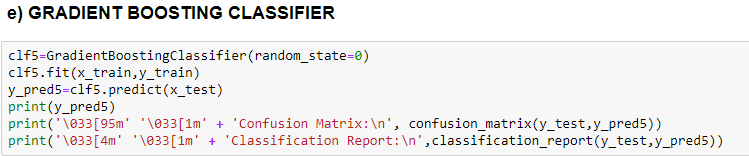
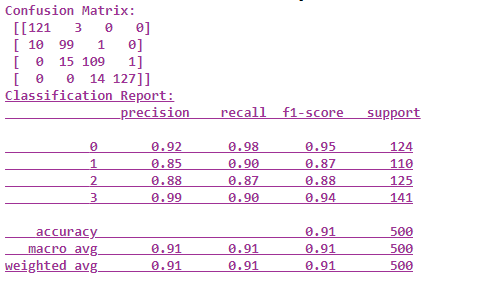
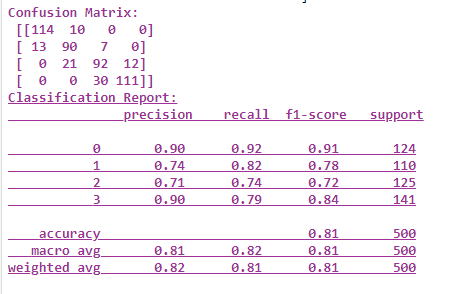


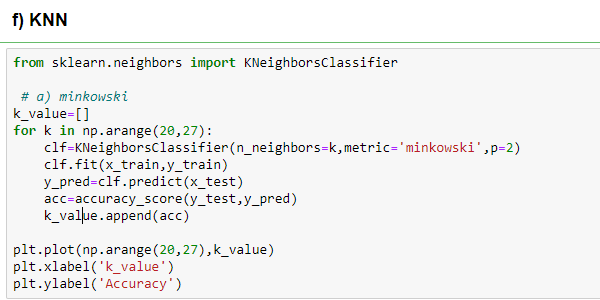


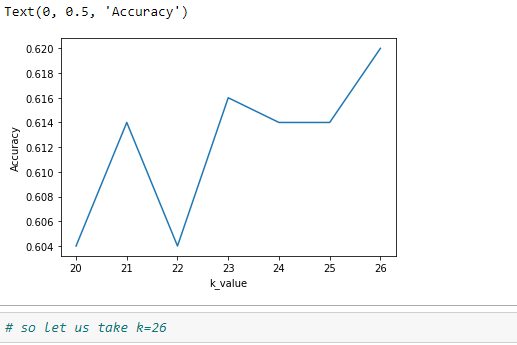


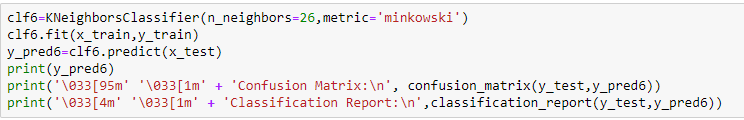


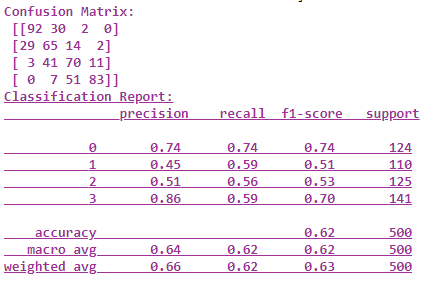






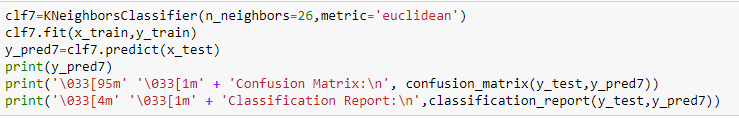


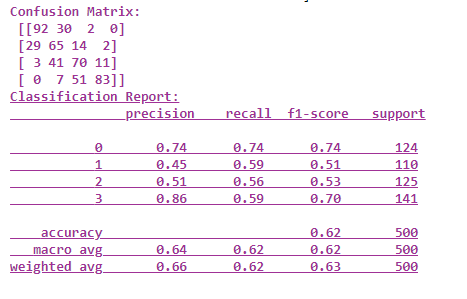


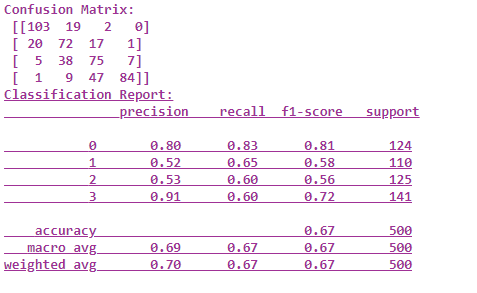
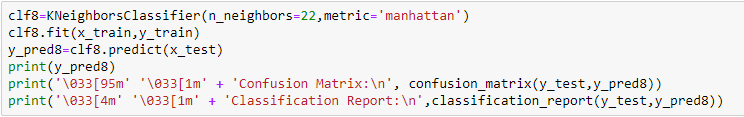
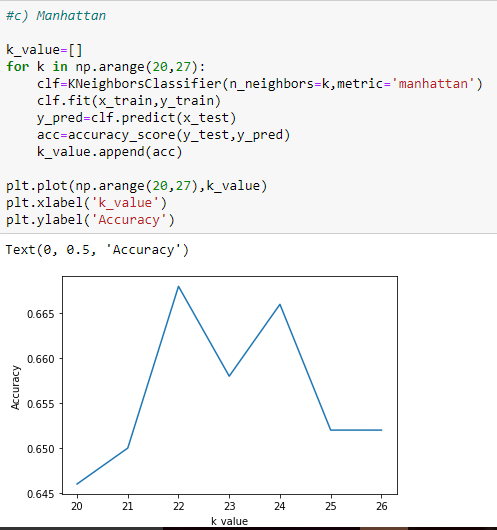




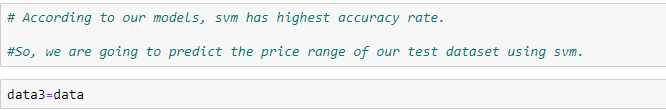
#So take k = 26

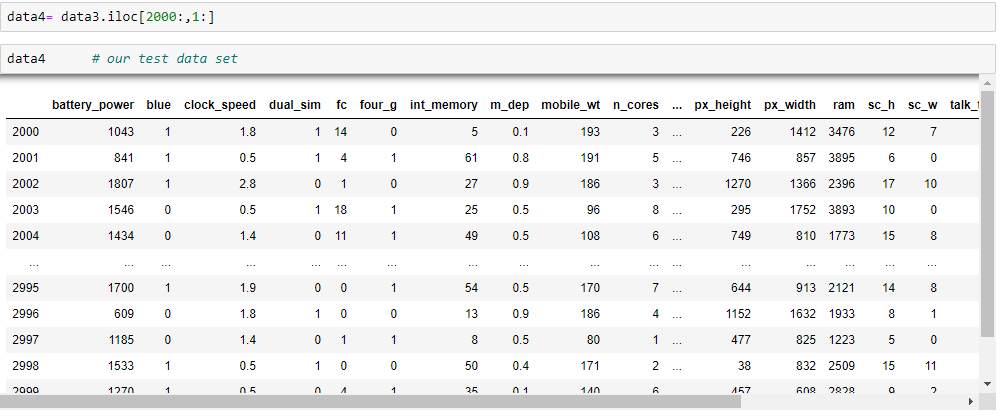


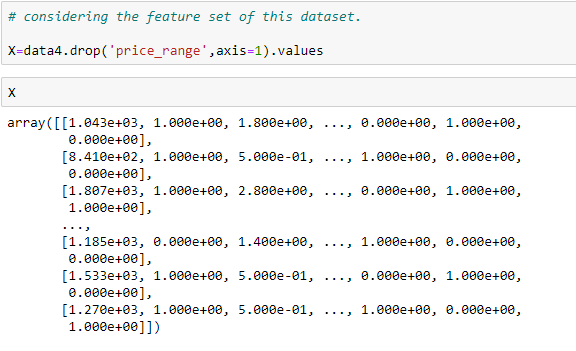


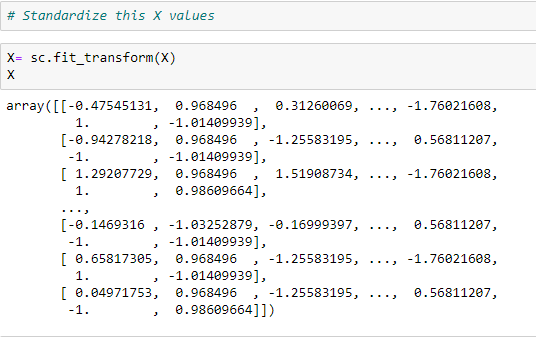


**3. Predicting the price range of Mobile phones by using the best model.**

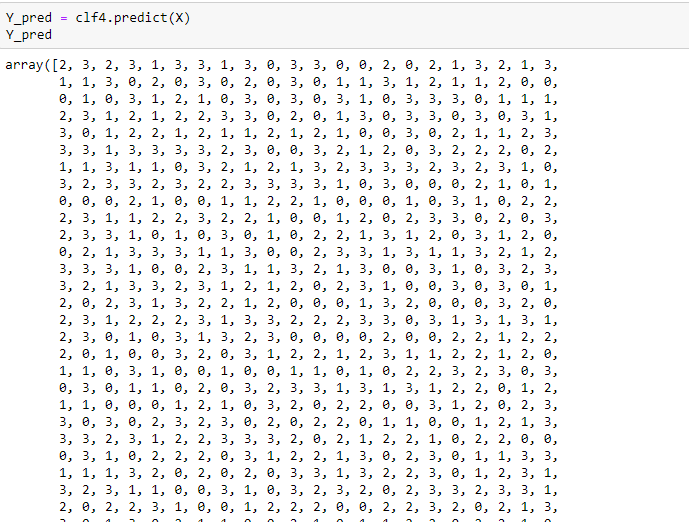


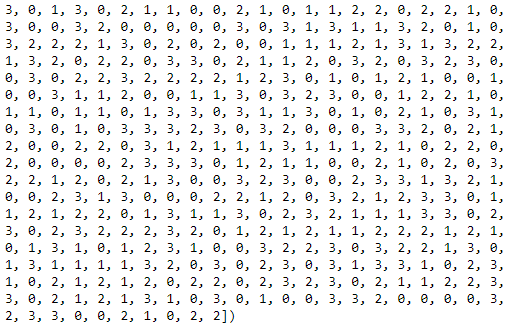






Now, let us predict the price ranges of the 1000 test set of mobile phones by using our best classification model – SVM.





**4.OUTCOMES OF THE WORK.**

Cost prediction is the very important factor of marketing and business. To predict the cost or price range, same procedure can be performed for all types of products for example Cars, Foods, Medicine, and Laptops etc.

Best marketing strategy is to find optimal product (with minimum cost and maximum specifications). So products can be compared in terms of their specifications, cost, manufacturing company etc. So, this kind of prediction will help the companies estimate price of mobiles to give tough competitions to other mobile manufacturers.

By specifying economic range a good product can be suggested to a costumer. I.e. it is useful for the costumers to verify that they are paying best price for the mobile.

**5. CONCLUSIONS.**

The various price\_ranges of the mobiles exist due to differences like: The size of their Random Access Memory. It can be said that, costlier mobile phones have more RAM, which is a core technical feature.

Higher battery power increases the mobile price.

Mobile phones in the highest price\_range class 3, has larger RAM ,more battery power,Larger Internal Memory. They also have Wifi and Dual Sim.. Also the weight of the mobile is less for the highest price\_range mobile phones.

In 0 class(low cost) battery power is 1100 mAH. The battery power for 0 class is a really good and supportive feature to purchase that class, when compared with price range 1 and 2 which are higher in prices but offers only a slight increase in battery power.

In 0 class(low cost) , the total internal memory is of the same size as that of much higher price range class 2. Thus price range 0 class mobile phones offers a good/ substantial amount of internal memory when in comparison with mobile phones of much higher cost, price range class 2.

In 1 class(medium cost) and in 2 class(high cost) battery power is 1200 mAH. For price range 2 which is higher than price range 1, the battery power is the same. For a buyer considering only higher battery power, can get a better buy/ purchase in the mobile phones of price range 1.

Highest accuracy was obtained from SVC Support Vector Classifier model in the Machine Language.

**6. REFERENCES.**

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

2) <https://github.com/abhinav23dixit/Mobile-Price-Classification>.

3) <https://medium.com/@sai.teja667edu/mobile-price-prediction-using-machine-learning-classification-techniques-1893262704e6>.

*4) Mobile Price prediction using Machine Learning Techniques* by B.Balakumar , P.Raviraj , V.Gowsalya.

*5) International Journal of Computer Applications (0975 –8887),Volume 179 –No.29, March 2018,*

Mobile Price Class prediction using Machine Learning Techniques by Muhammad Asim, Zafar Khan.