

# Mobile Price Range Prediction

Aniket Raj, Kushal Kishor, A Balaji,  
Akhilesh Bhardwaj, Aswathaman,  
Data science trainees,  
AlmaBetter, Bangalore

## Overview:

In the competitive mobile phone market companies want to understand sales data of mobile phones and factors which drive the prices.

## Business Problem:

The objective is to find out some relation between features of a mobile phone (e.g.: - RAM, Internal Memory, etc.) and its selling price. In this problem, we do not have to predict the actual price but a price range indicating how high the price is? The aim is to develop a model to predict the price of a mobile when the specifications of a mobile are given and to find the ML algorithm that predicts the price most accurately. The usage of archival data to accurately forecast forthcoming instances is the essence of Predictive Analytics. One of the ways Predictive Analytics can be performed is by using Machine Learning. Predictive Machine Learning works by taking in data as input to develop and train a prediction model and the trained model is used

to predict the outcome of future data instances. Supervised Machine Learning algorithms make use of data that contains a pre-defined class label, which is the attribute that needs to be predicted. The class label is the price of a mobile in our case. The Mobile Price Class dataset sourced that categorizes mobiles into price ranges was used to train the prediction model. Python is used due to its readily accessible ML libraries. Various classification algorithms were used to train the model to try and find the algorithm that is able to predict the mobile price class most accurately. Metrics like accuracy score, confusion matrix, etc. are used to evaluate the trained model to determine the algorithm most suitable among the used ones.

## Data Description:

This dataset has around 2,000 observations in it with 21 columns and all are in numeric values.

## Variables description:

- **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).

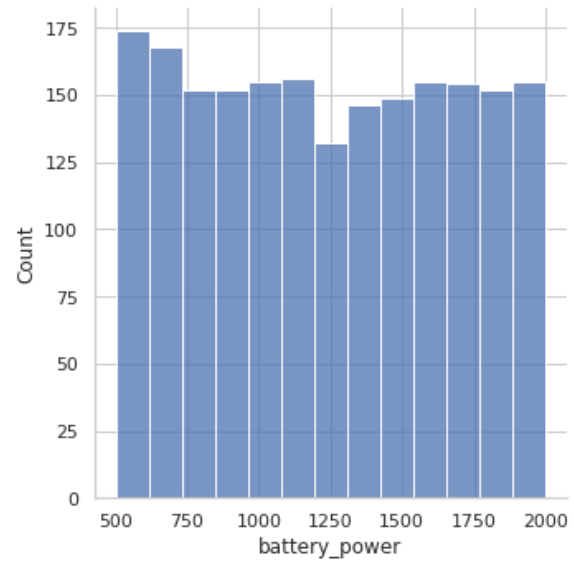
## Approach:

Our business wants to predict a price range indicating how high the price of a mobile when the specifications of a mobile are given. Problem is a classification one and we can follow the steps as given below to develop a model —

- **Perform exploratory data analysis** — Observe various features which are impacting a price range of a mobile.
- **Prepare Data** — Clean data — missing values, unknown values, Outlier detection and treatment to ensure that data is ready for algorithm to consume.
- **Split data** — Split our data into training and test data. I went for 80–20 split.
- **Choose an algorithm** — Identifying a right algorithm for the problem is a major task and mostly it just doesn't happen in one go. I went for

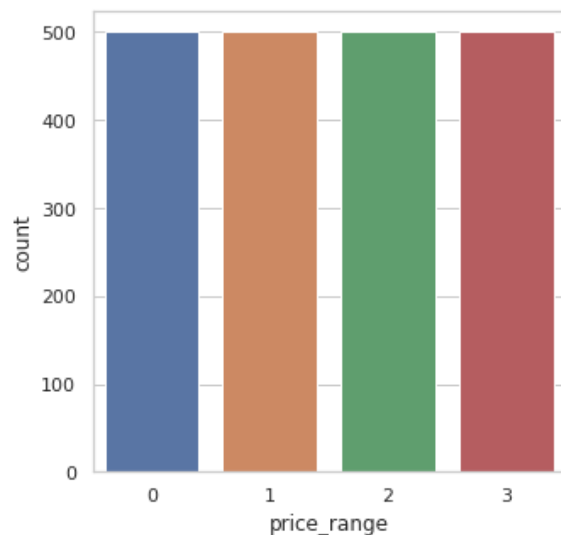
Classification algorithm as all features.

- **Predict and evaluate model** by using Metrics like accuracy score, confusion matrix, etc. used for Classification.

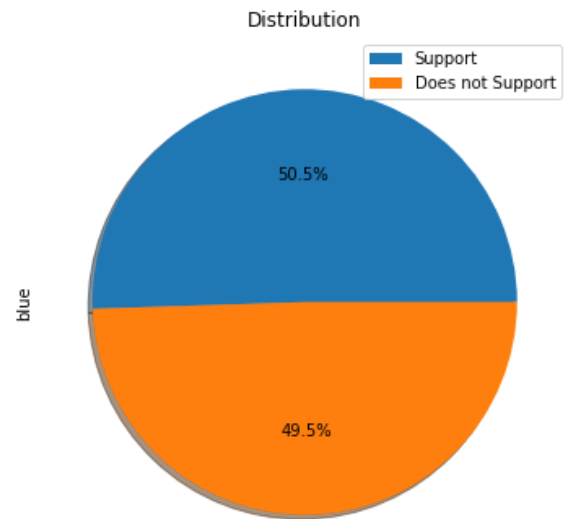


2) This plot shows how the battery mAh is spread.

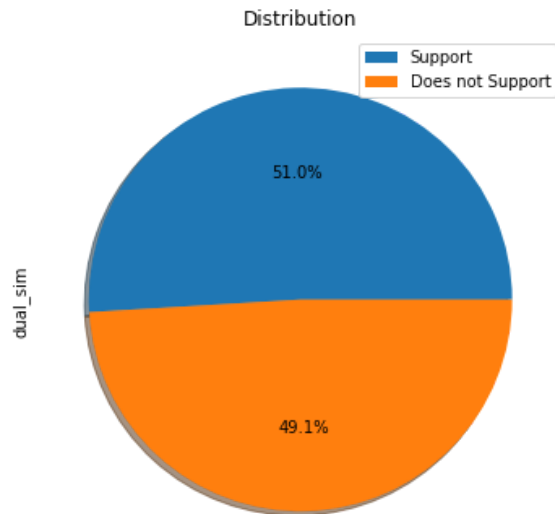
## Results of Exploratory Data Analysis (EDA):



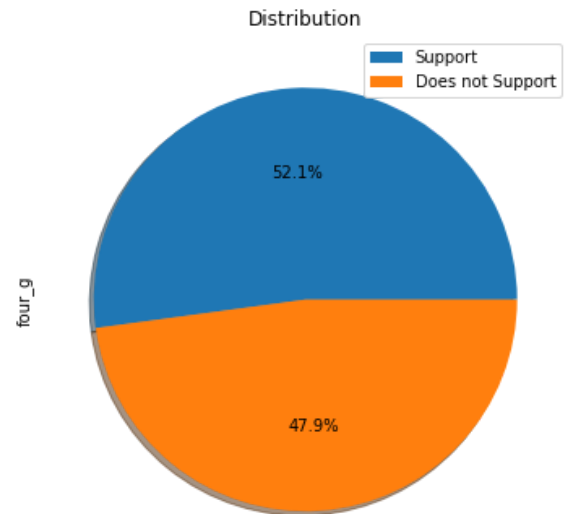
1) There are mobile phones in 4 price ranges. The number of elements is almost similar.



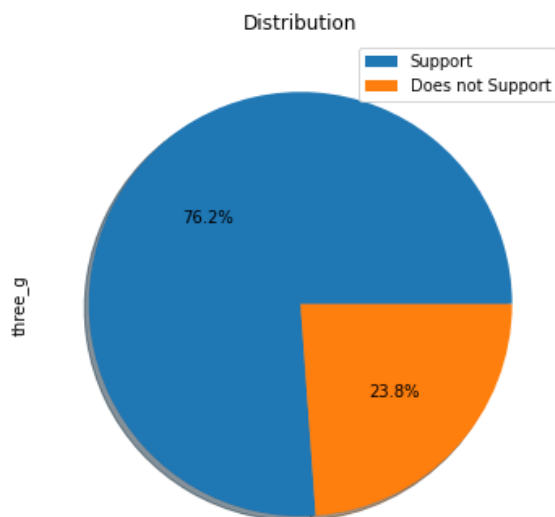
3) The percentage count of how many devices support Bluetooth and how many does not support.



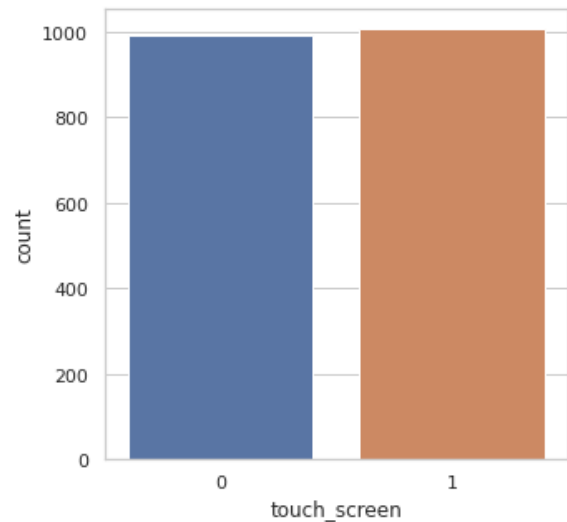
4) The percentage count of how many devices support Dual sim and how many does not support.



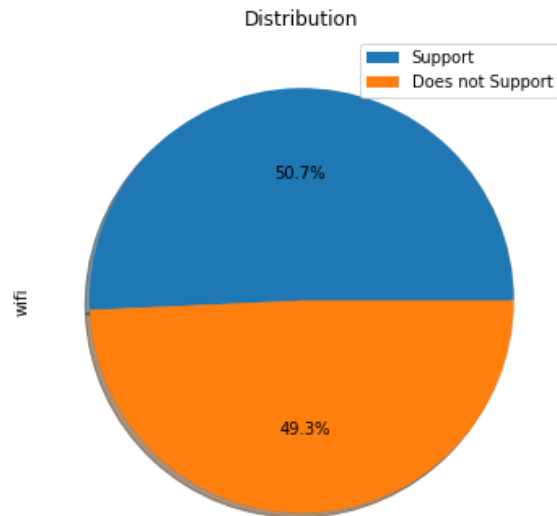
6) The percentage count of how many devices support 4G and how many does not support.



5) The percentage count of how many devices support 3G and how many does not support.



7) The count of how many devices have Touch screen and how many does not.



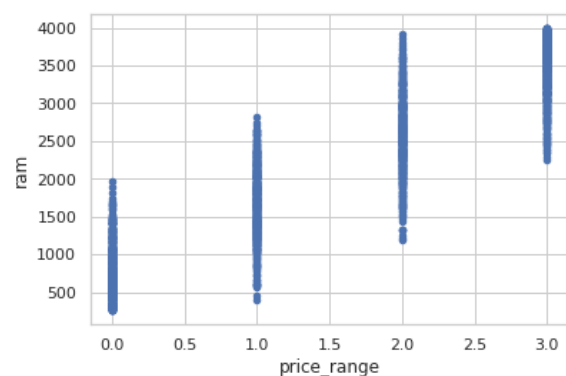
8) The percentage count of how many devices support WIFI and how many does not support.

### Target Variable:

Price range is the target variable with value of 0(low cost), 1(medium cost), 2(high cost) and 3(very high cost).

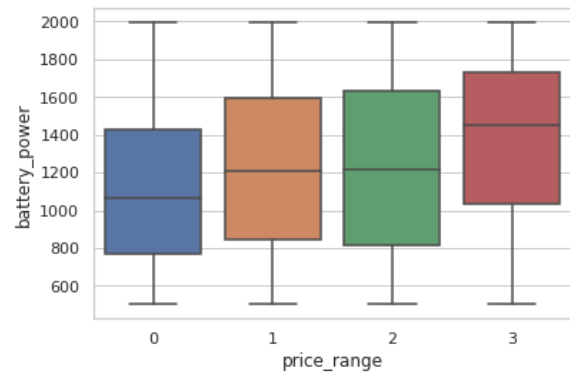
### Preparing Data for training:

1.

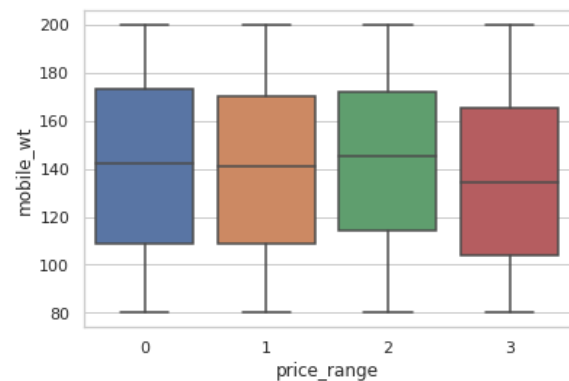


1) Ram has continuous increase with price range while moving

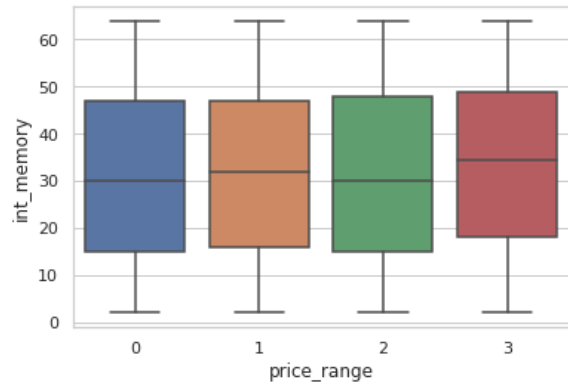
from Low cost to Very high cost.



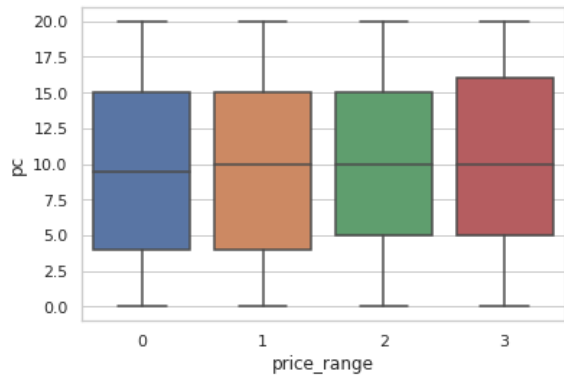
2) Battery power has continuous increase with price range while moving from Low cost to Very high cost.



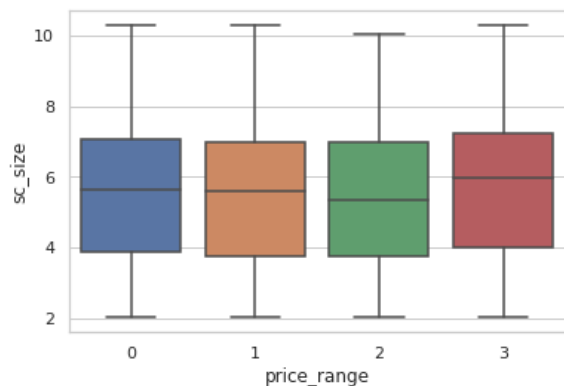
3) There is not a continuous increase in Mobile weight as we move from Low cost to Very high cost. Mobiles with very High cost has lower mobile weight.



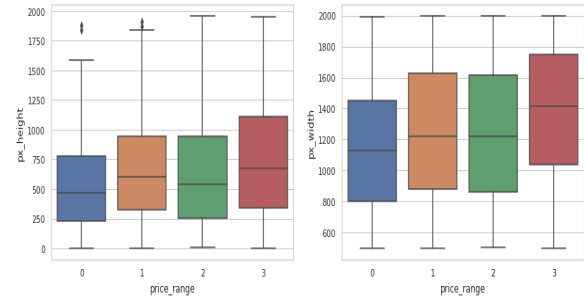
- 4) As we move from Low cost to Very high cost, Mobiles has almost equal Internal memory.



- 5) Primary camera megapixels are showing a little variation along the target categories.



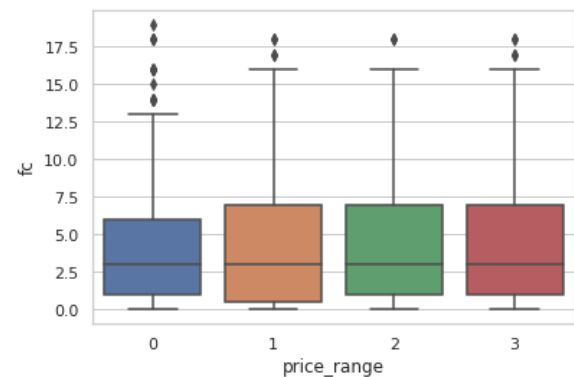
- 6) Screen Size shows little variation along the target variables.



- 7) Pixel height is almost similar as we move from Low cost to Very high cost, little variation in pixel height.

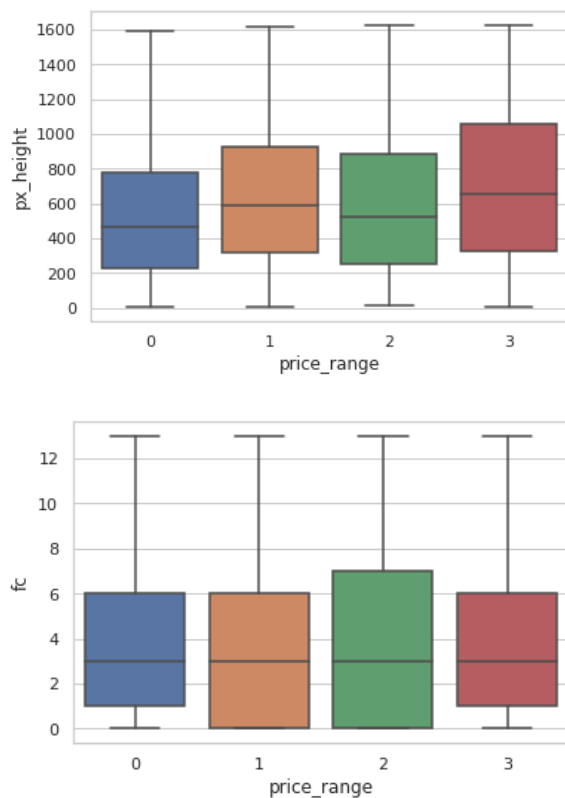
There is not a continuous increase in pixel width as we move from Low cost to Very high cost. Mobiles with 'Medium cost' and 'High cost' has almost equal pixel width.

so, we can say that it would be a driving factor in deciding price range.



- 8) we can see front camera shows little variation along the target variables.

As we can see from above graph there are Outliers! Because outliers can markedly affect our models that's why we will do treatment of outliers.



The shape of the final data set is (1865,19)

- 2.Perform train, test split
- 3.Create Model object and train model
- 4.Perform Predictions

Metrics used to evaluate the algorithms in this paper are confusion matrix, classification report and accuracy score.

A confusion matrix has the total count of the accurately grouped occurrences along its cross and the count of the incorrectly classified instances in the rest of the matrix. We have used 4 class values; so, the matrix generated is a 4\*4 matrix.

A classification report gives the full report of the classification with parameters like recall, precision, f1-score, etc.

Accuracy score gives the accuracy of the trained model after evaluating it using test data, for which we have sampled 20% of the dataset.

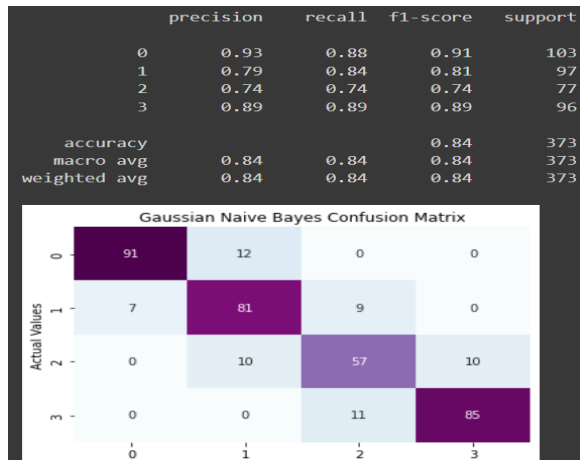
## • LOGISTIC REGRESSION

	precision	recall	f1-score	support
0	0.93	0.99	0.96	103
1	0.93	0.89	0.91	97
2	0.91	0.88	0.89	77
3	0.96	0.96	0.96	96
accuracy			0.93	373
macro avg	0.93	0.93	0.93	373
weighted avg	0.93	0.93	0.93	373

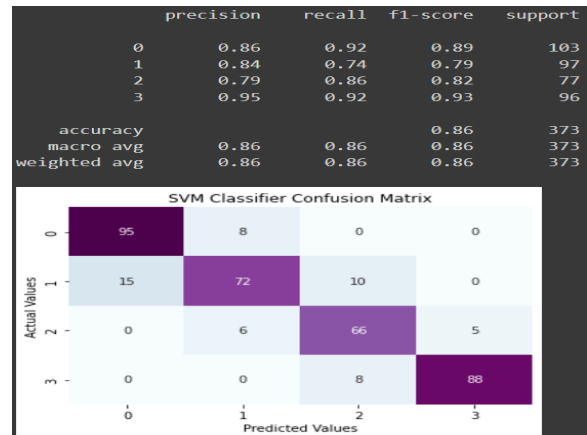
  

Logistic Regression Confusion Matrix				
Actual Values	0	1	2	3
0	102	1	0	0
1	8	86	3	0
2	0	5	68	4
3	0	0	4	92
	0	1	2	3
		Predicted Values		

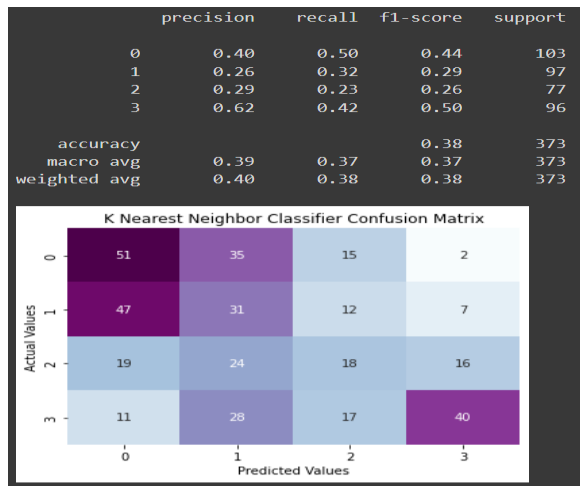
- GAUSSIAN NAIVE BAYES



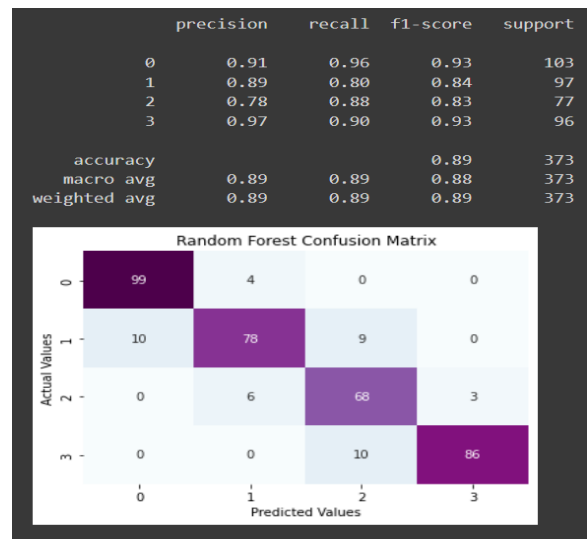
- SVM CLASSIFIER



- KNN CLASSIFIER

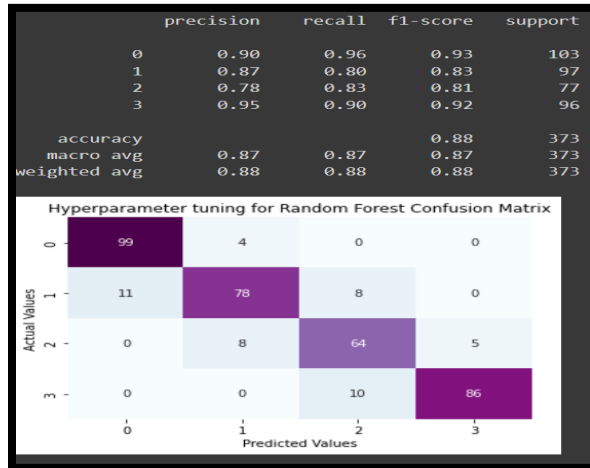


- RANDOM FOREST

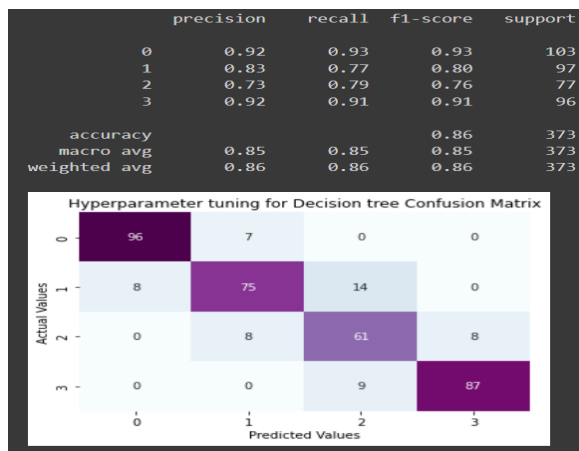




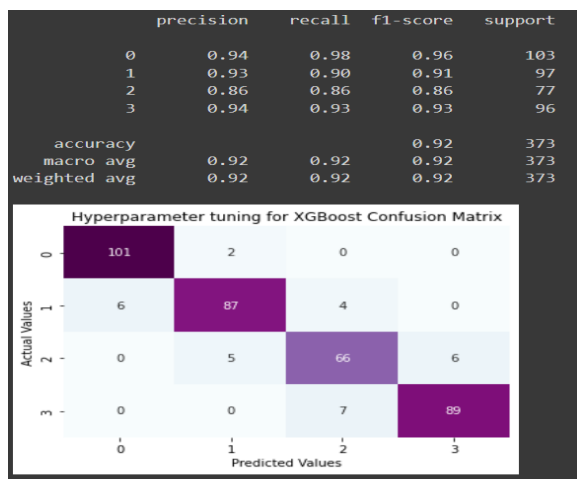
## Hyperparameter Tuning: 1. Grid Search CV on Random Forest:



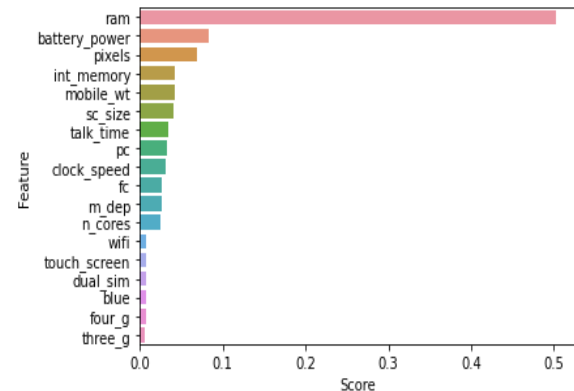
## 2. Grid search CV on Decision tree



## 3. Grid search CV on XGBoost



## Feature Importance:



The price of the mobile majorly depends on the ram, followed by battery power and camera pixel size.

## Conclusion:

From all the above experiments we can conclude that logistic regression and XGboosting with using hyperparameters we got the best results.

This would help organizations and consumers alike to make more educated decisions when it comes to price.