

Capstone Project – 3

Supervised Machine Learning(Classification)

MOBILE PRICE RANGE PREDICTION

BY: Nisarga C

CONTENTS

1. Problem statement
2. Data Description
3. Introduction
4. EDA and Feature Engineering
5. Feature Selection
6. Preparing Dataset for Modelling
7. Applying Model
8. Model Validation and Selection
9. Conclusion



PROBLEM STATEMENT

- A man began his own versatile organization. He needs to give intense battle to large organizations like Apple, Samsung and so forth He doesn't have a clue how to appraise cost of mobiles his organization makes.
- In this serious cell phone market you can't just expect things. To take care of this difficult he gathers deals information of cell phones of different organizations.
- He needs to discover some connection between highlights of a cell phone (eg: RAM , Internal Memory)and its selling cost. However, he leaves something to be desired at Machine Learning. So he needs your assistance to tackle this issue.
- In this difficult you don't need to anticipate real cost however a value range showing how high the cost is

DATA 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
- **Touch_screen** - Has touch screen or not
- **Wifi** - Has wifi or not
- **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
- **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).

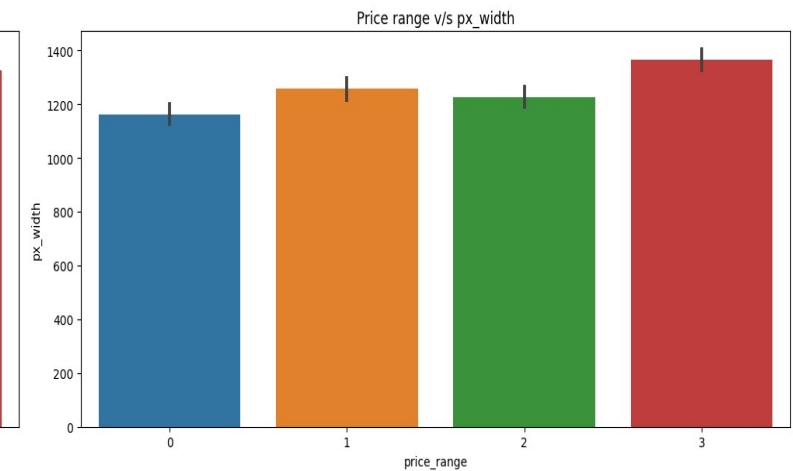
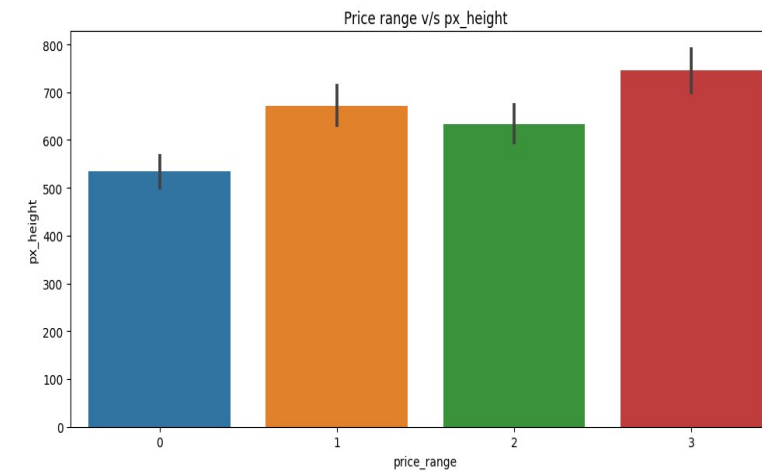
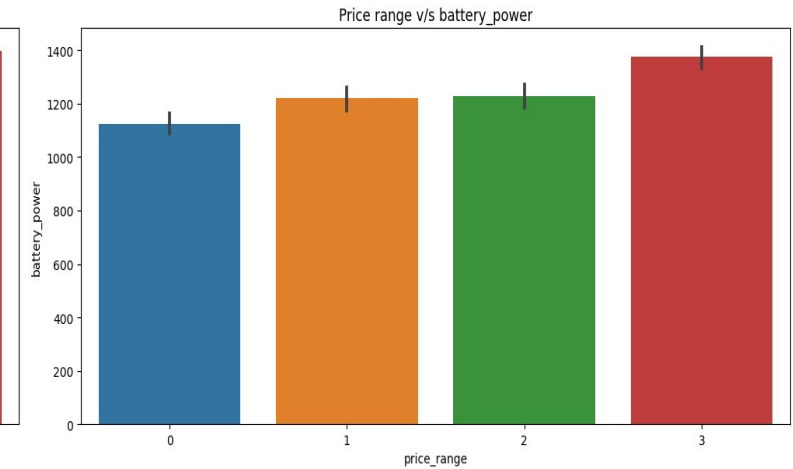
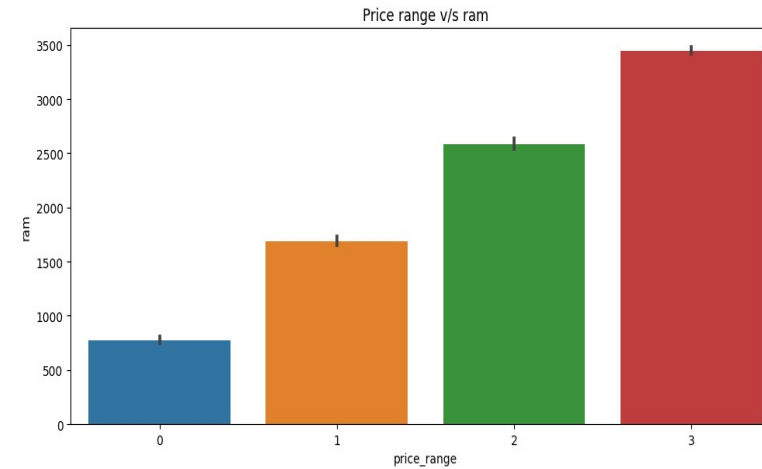
INTRODUCTION

- Price always has an important impact factor in the product buying aspect and also in the mindset of the buyer who would consider “what is the worth and is it good to buy within this range”. In Mobile there are many specification and features like camera, video, quality of processor, quality of the material.
- There are many constraints in consideration of the price, as the product should be economical and reachable with overall consideration. Mobile Prices and Specification is mainly considered for selection and comparison.
- In this Prediction, There are many multiple variables to be considered to get the précised results of the price and other features. Of the mobile dataset this will help the buyer and also the marketer and the developer to get precise information from historical data of mobile phones and help them to decide are fine and satisfactory.
- In this project KNN Model algorithm, it is used to find the distances between K models and model for testing the dataset. The accuracy is also calculated using K model and the training model is used. The KNN model is used to predict the model for price.

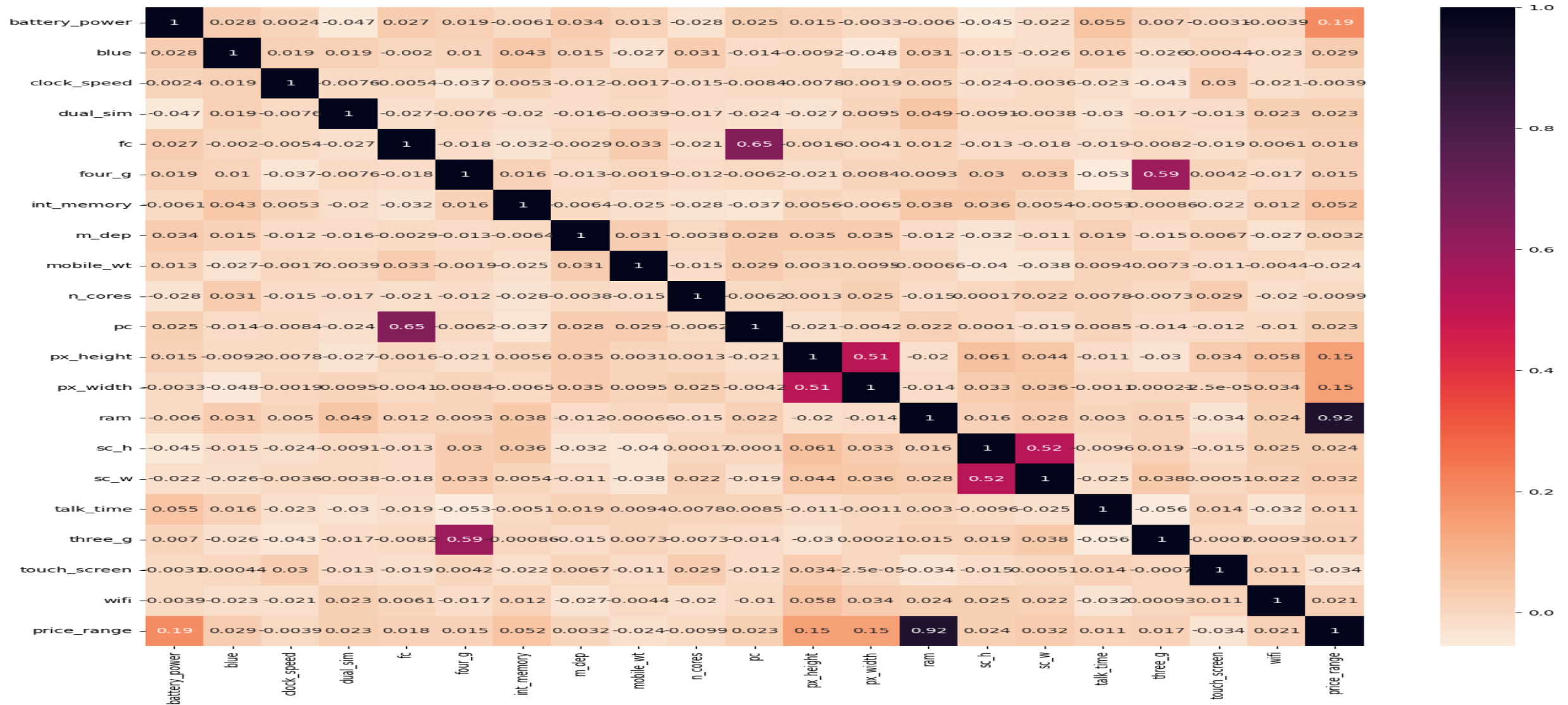
EDA And Feature Engineering

Relation Between Price,Range,Ram,Px_Height,Px_width

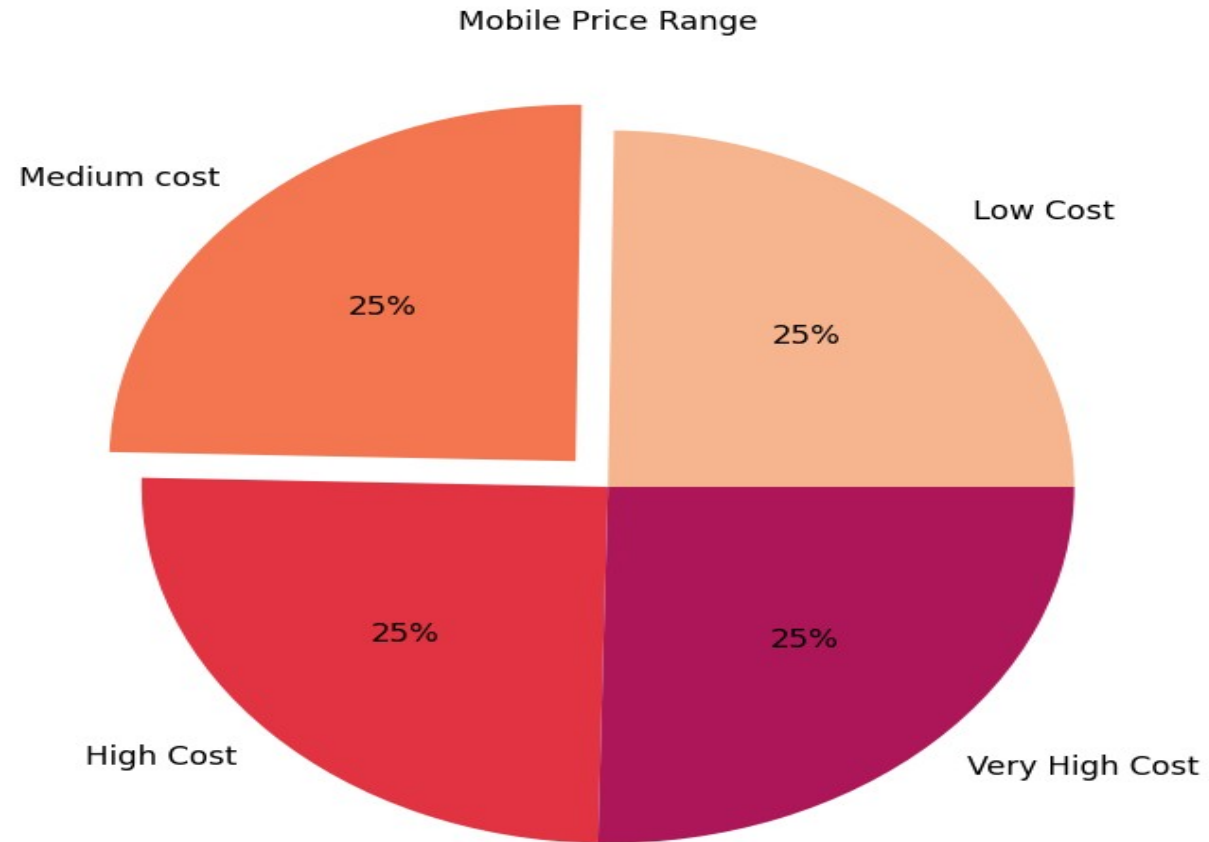
- Mobiles having RAM more than 3000 MB falls under Very high cost category. As RAM increases price range also increases.
- Mobiles having RAM less than 1000 MB falls under low cost category.
- Mobiles with battery power more than 1300 mAh has very high cost. And Mobiles with battery power between 1200 and 1300 mAh falls under medium and high cost category.
- Mobiles with more than 700 pixel height and width more than 1300 has very high cost.



Multivariate Analysis

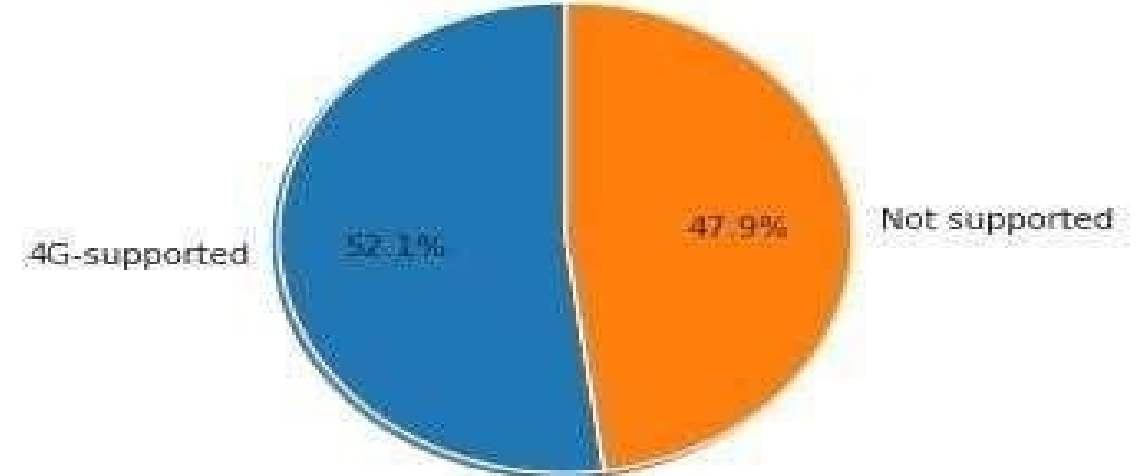
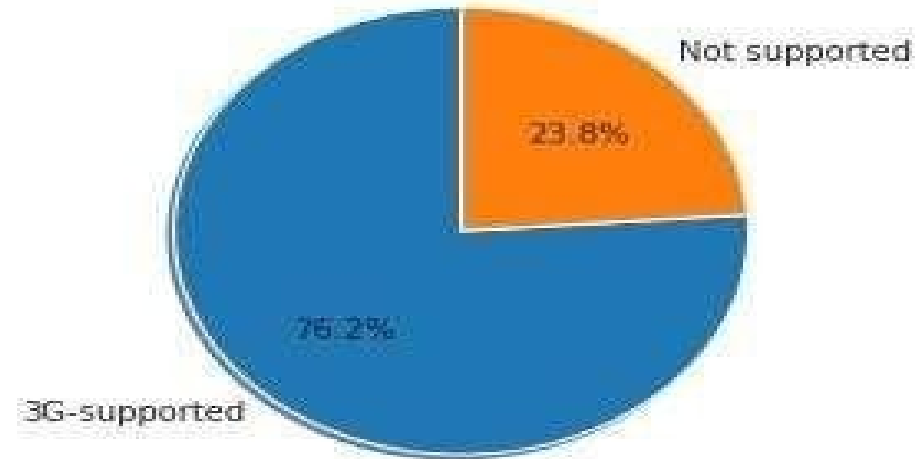


Checking Price



- There are mobile phones in 4 price ranges. so according this pie chart percentages of all Mobile price range almost similar.
- It will not help much. Because costs are almost similar.

3G and 4G Supported or Not

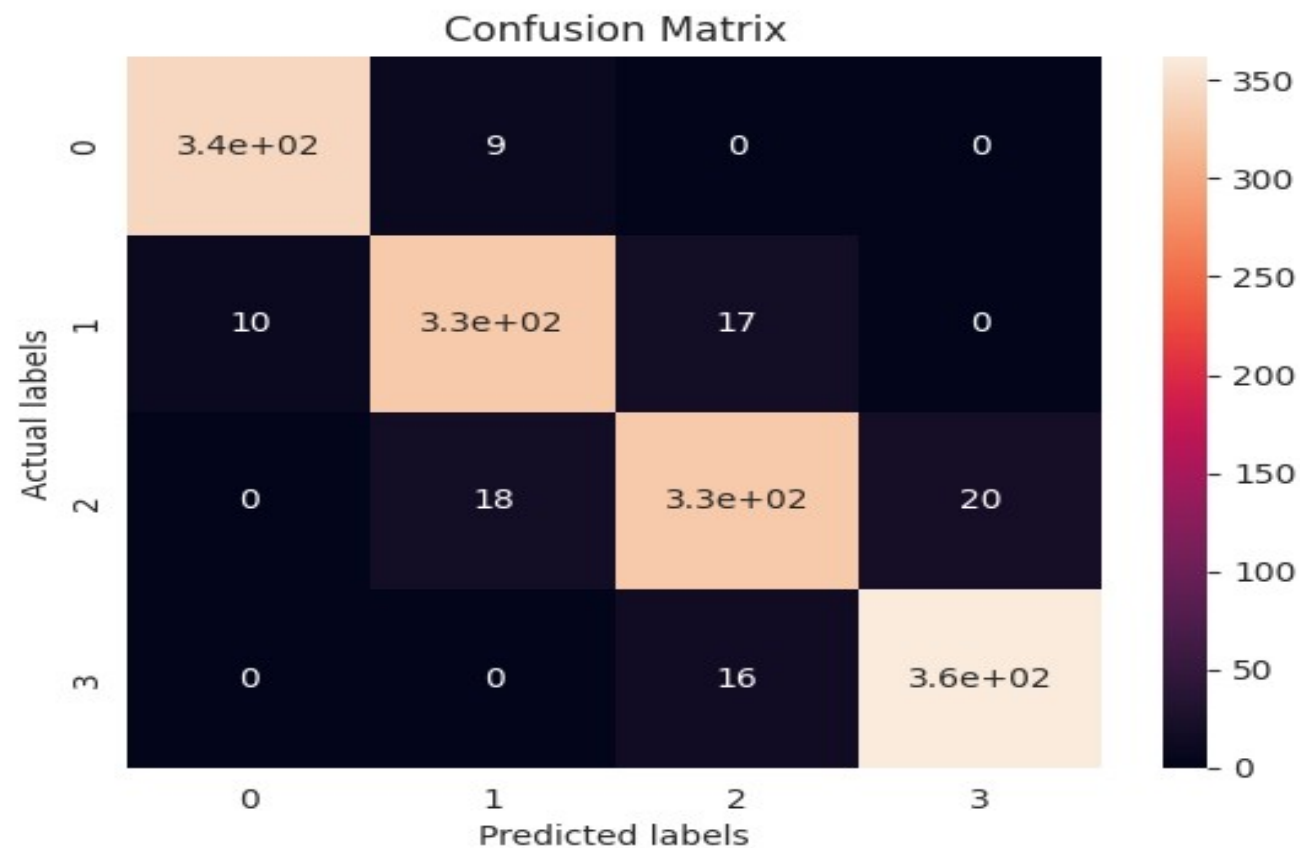


Model Selection And Evaluation

- I have implemented 3 classification models and achieved a fairly good result for all the algorithms.
- Logistic Regression
- Random Forest
- XG Boost
- Logistic regression shows approx accuracy of 90% and the other two shows approx 80% accuracy, which shows that the data were properly classified.
- Logistic Regression performed better than the other two model.

ML Model 1 – Building a Logistic Regression Model

- The precision on test dataset is 0.9419049940445509
- The recall on test dataset is 0.9423076923076923
- The roc auk score on test dataset is 0.9934859517458531



Implementing The Logistic Regression

Train Metrics

	precision	recall	f1-score	support
0	0.90	0.88	0.89	228
1	0.70	0.71	0.70	212
2	0.72	0.67	0.70	229
3	0.85	0.90	0.87	228
accuracy			0.79	897
macro avg	0.79	0.79	0.79	897
weighted avg	0.79	0.79	0.79	897

Test Metrics

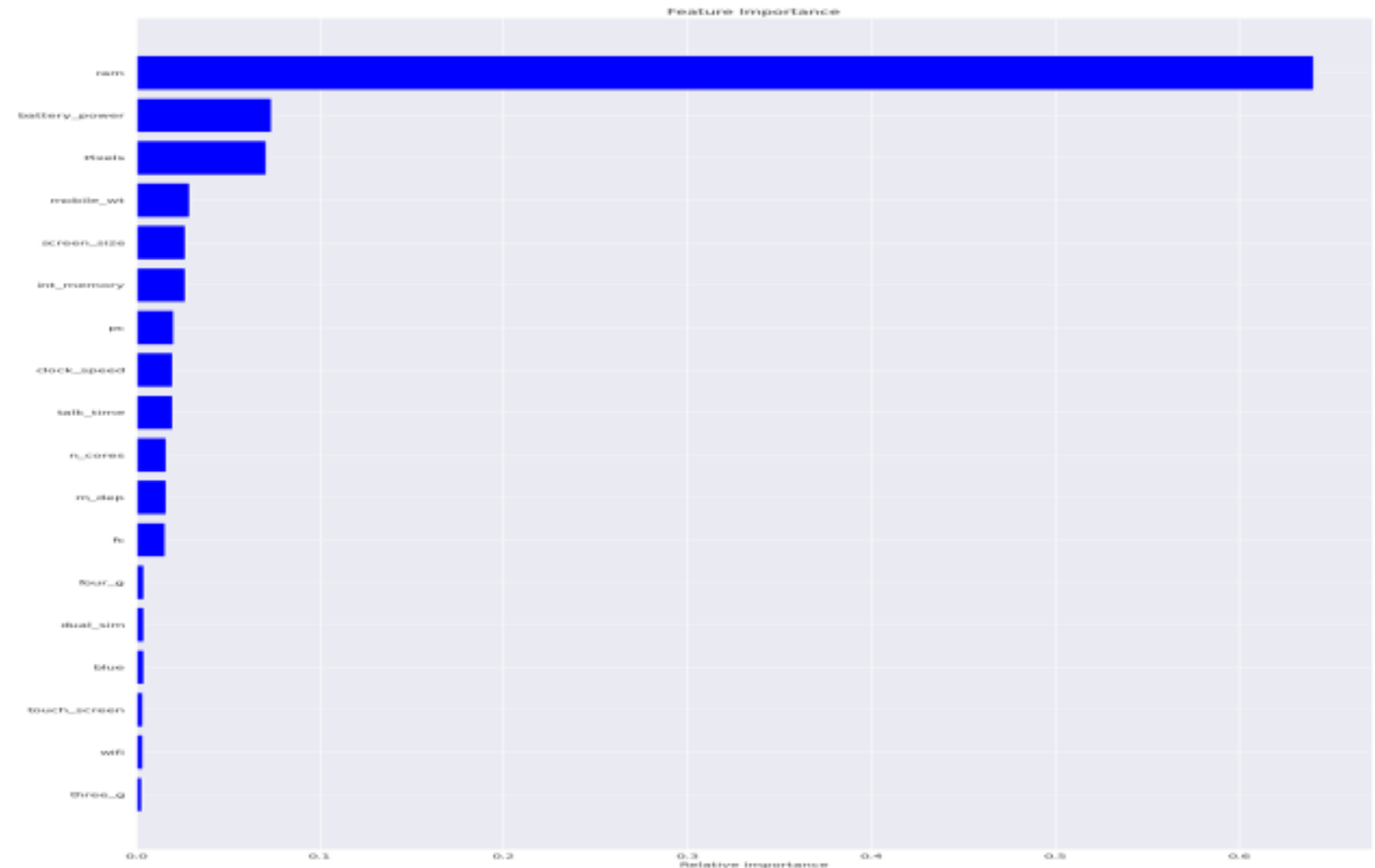
	precision	recall	f1-score	support
0	0.88	0.82	0.85	165
1	0.60	0.57	0.59	165
2	0.58	0.63	0.60	165
3	0.83	0.85	0.84	165
accuracy			0.72	660
macro avg	0.72	0.72	0.72	660
weighted avg	0.72	0.72	0.72	660



ML Model 2 – Building a Random forest classifier Model

By performing Random Forest we get the results are as follows:

- TrainScore-0.87
- TestScore-0.83



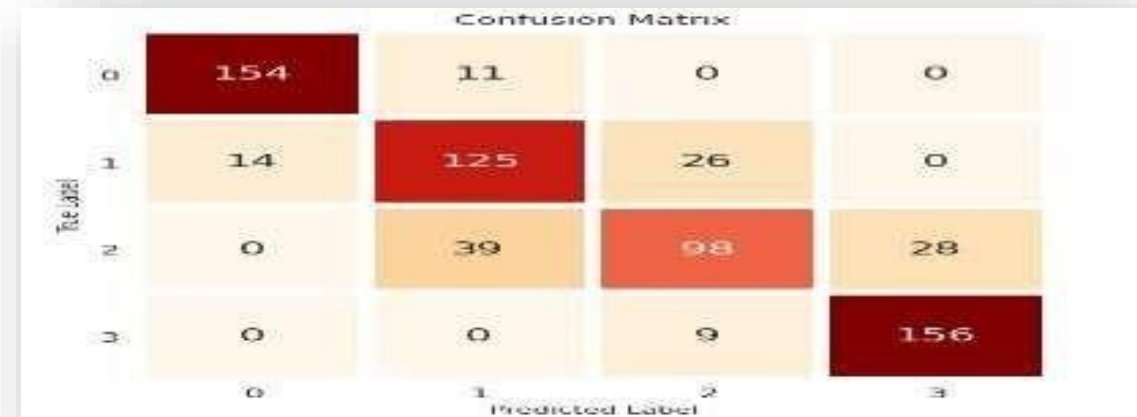
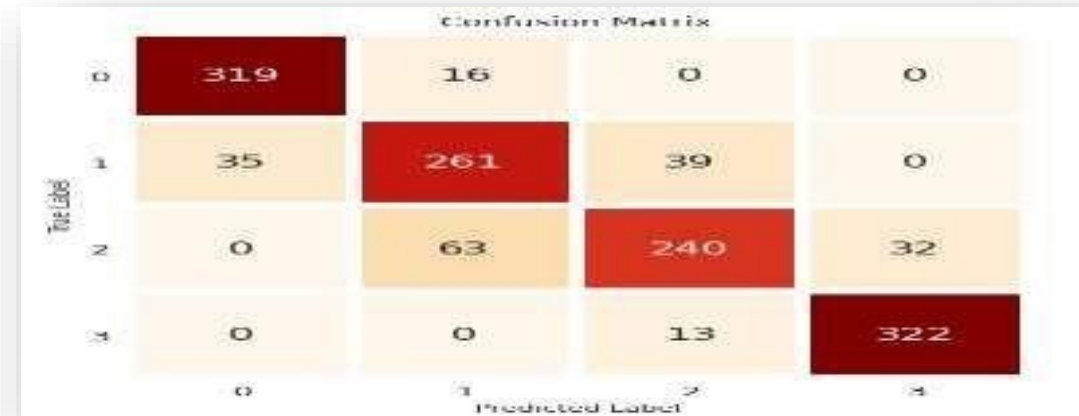
Implementing Random Forest Classifier

Train Metrics

	precision	recall	f1-score	support
0	0.90	0.95	0.93	335
1	0.77	0.78	0.77	335
2	0.82	0.72	0.77	335
3	0.91	0.96	0.93	335
accuracy			0.85	1340
macro avg	0.85	0.85	0.85	1340
weighted avg	0.85	0.85	0.85	1340

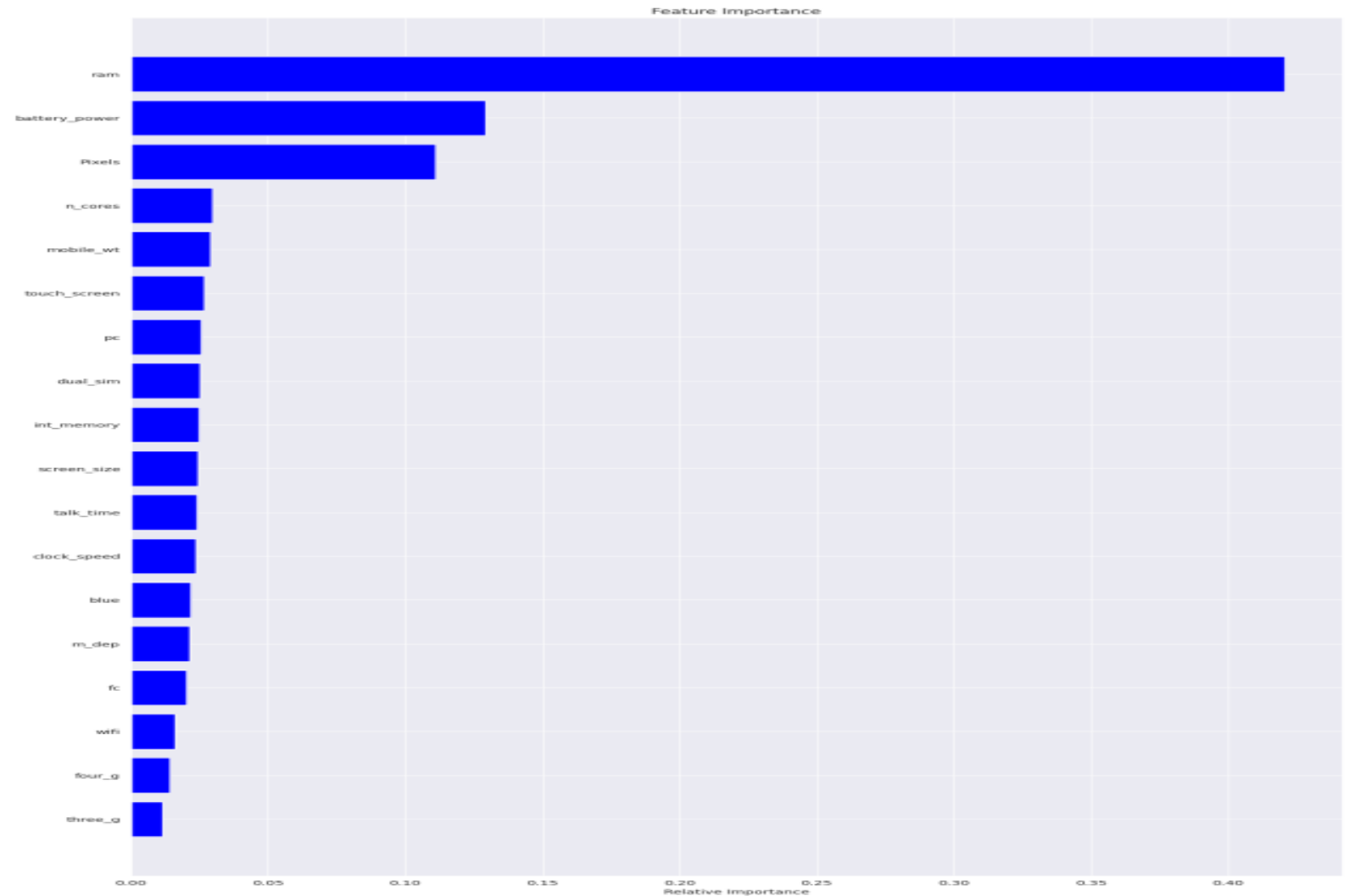
Test Metrics

	precision	recall	f1-score	support
0	0.92	0.93	0.92	165
1	0.71	0.76	0.74	165
2	0.74	0.59	0.66	165
3	0.85	0.95	0.89	165
accuracy			0.81	660
macro avg	0.80	0.81	0.80	660
weighted avg	0.80	0.81	0.80	660



ML Model 3 – Building a XG Boost Model

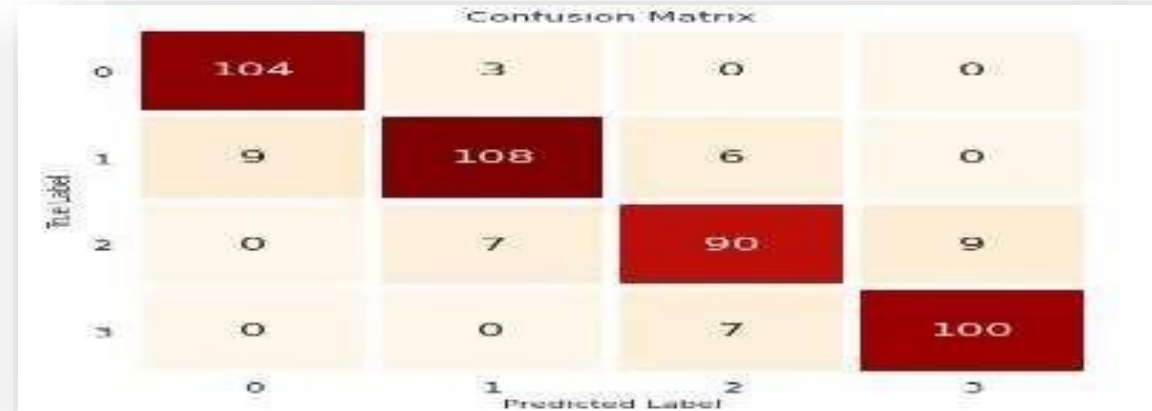
- By performing XG Boost Classifier we get the results areas follows :
- Trainscore-0.91
- TestScore-0.88



Implementing XG boost Classifier

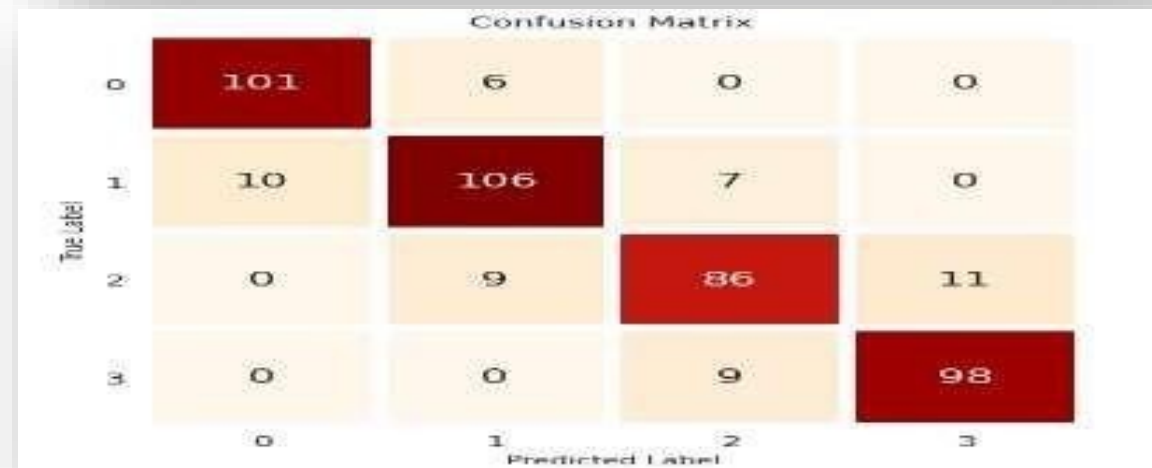
Train Metrics

Classification Report				
	precision	recall	f1-score	support
0	0.92	0.97	0.95	107
1	0.92	0.88	0.90	123
2	0.87	0.85	0.86	106
3	0.92	0.93	0.93	107
accuracy			0.91	443
macro avg	0.91	0.91	0.91	443
weighted avg	0.91	0.91	0.91	443



Test Metrics

Classification Report				
	precision	recall	f1-score	support
0	0.91	0.94	0.93	107
1	0.88	0.86	0.87	123
2	0.84	0.81	0.83	106
3	0.90	0.92	0.91	107
accuracy			0.88	443
macro avg	0.88	0.88	0.88	443
weighted avg	0.88	0.88	0.88	443



CONCLUSION

- From EDA we can see that, there are mobile phones in 4 price ranges. The number of elements is almost similar.
- Half the devices have Bluetooth, and half don't.
- There is a gradual increase in battery as the price range increases.
- Ram has continuous increase with price range while moving from Low cost to Very high cost.
- Costly phones are lighter.
- RAM, battery power, pixels played more significant role in deciding the price range of mobile phone.