Capstone Project

Supervised machine learning – Classification

Mobile Price Prediction

Contributor -Neha Pasi

Abstract

Mobile phones come in all sorts of prices, features, specifications and all. Price estimation and prediction is an important part of consumer strategy. Deciding on the correct price of a product is very important for the market success of a product. A new product that has to be launched, must have the correct price so that consumers find it appropriate to buy the product

Problem Statement

The data contains information regarding mobile phone features, specifications etc. and their price range. The various features and information can be used to predict the price range of a mobile phone.

Defaulters

1. Defining Problem Statement
2. EDA
3. Feature Selection
4. Preparing dataset for modeling
5. Applying model
6. Model selection

Data Summary

The dataset is of shape(1820, 21)

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

Data Summary

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(Dependent Variable) - This is the target variable with value of 0(lowcost), 1(medium cost),2(high cost) and 3(very high cost).

Data Pipeline

Data Processing - We have checked for null values in any columns and found that we have preety much clean data and we don't have any null values in our dataset

after checking statistical summary we found that the column of Sc_w(screen width) has min value zero which is unjustified. So, we have dropped those rows.

EDA – In this part we have done Exploratory data analysis using some charts on the features of dataset.

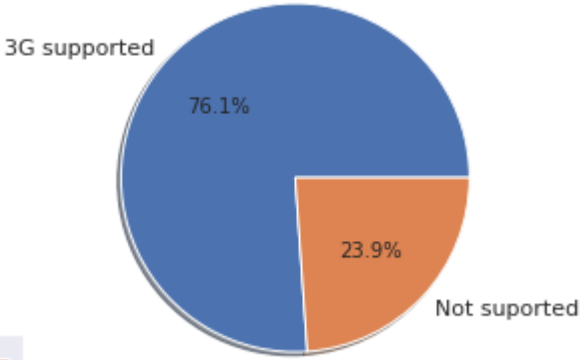
Fitting different Models – for modelling we tried various classification algorithms

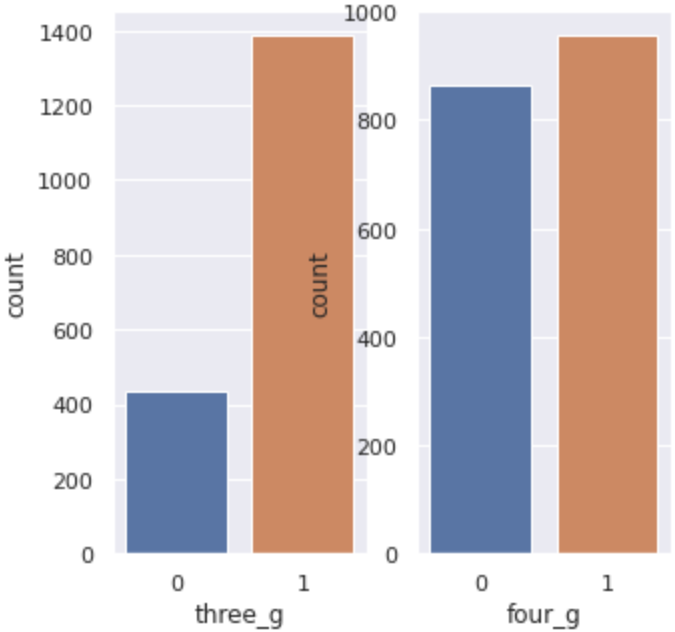
- Random Forest Algorithm
- Naive Bayes
- KNN Classifier
- SVM(support vector machine) Algorithm

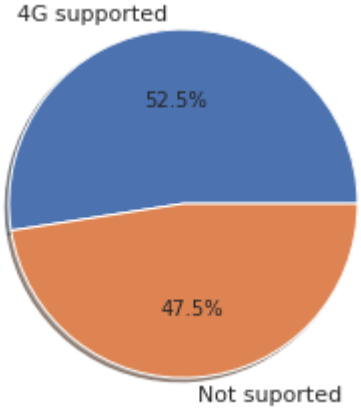
Hyperparameter tuning(SVM Classifier)

Best Model Selection

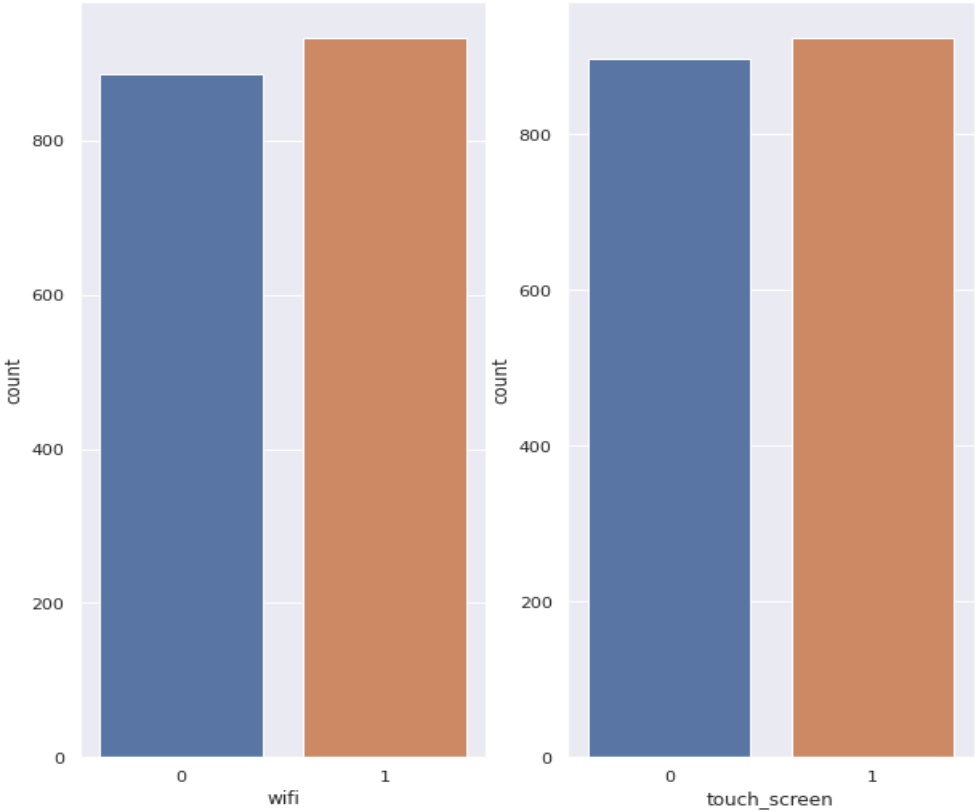
EDA

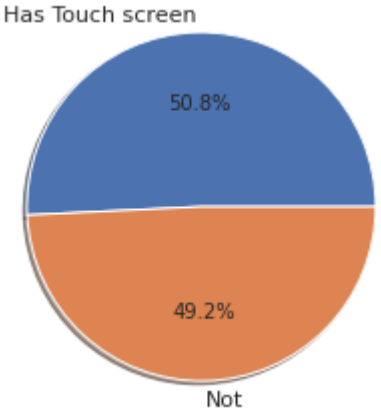

3G supported
76.1%
23.9%
Not suported

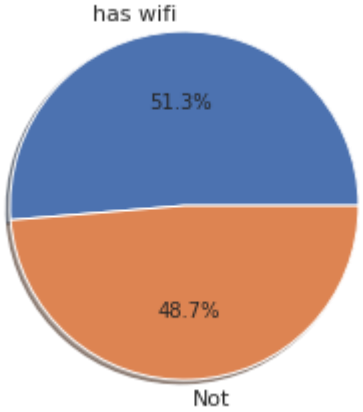

count
1400
1200
1000
800
600
400
200
0
0
1
three_g
count
1000
800
600
400
200
0
0
1
four_g


4G supported
52.5%
47.5%
Not suported

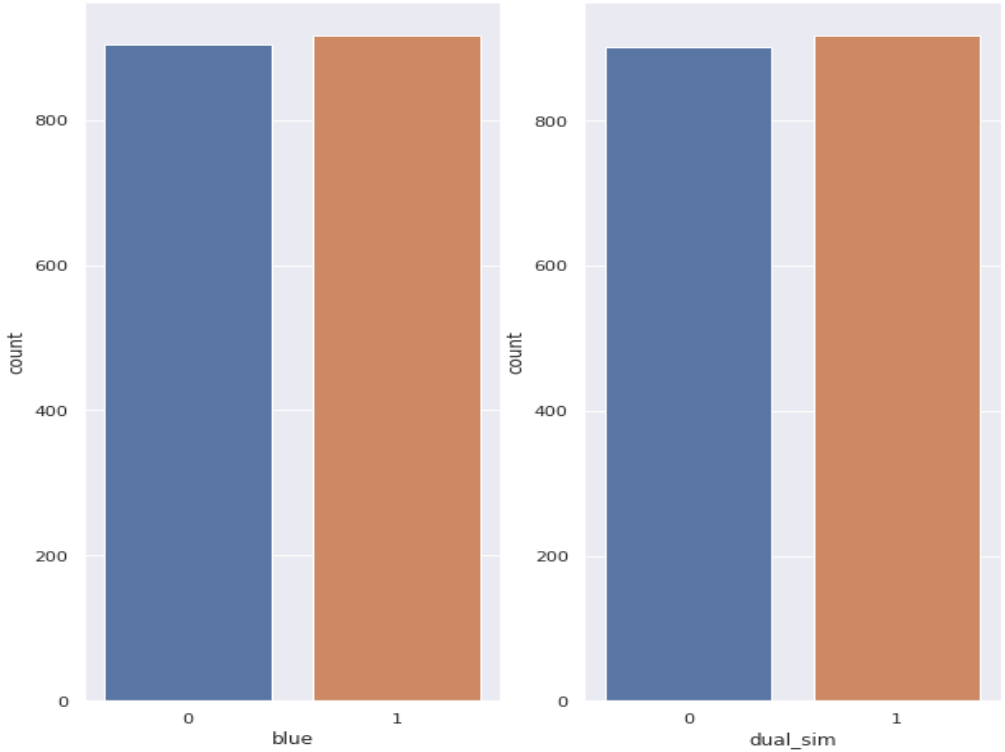
EDA

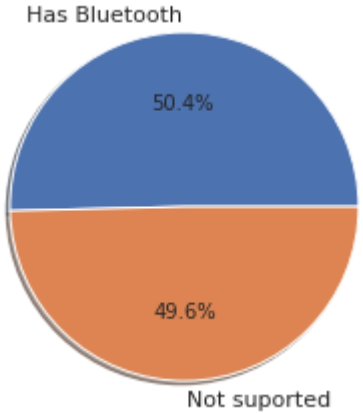

800
600
400
200
0
count
0
1
wifi
800
600
400
200
0
count
0
1
touch_screen

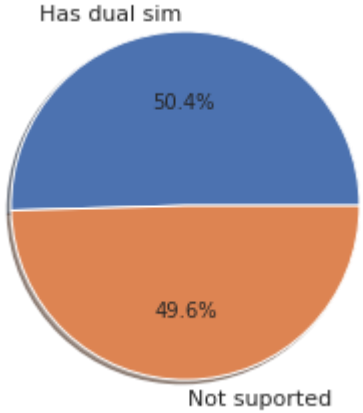

Has Touch screen
50.8%
49.2%
Not


has wifi
51.3%
48.7%
Not

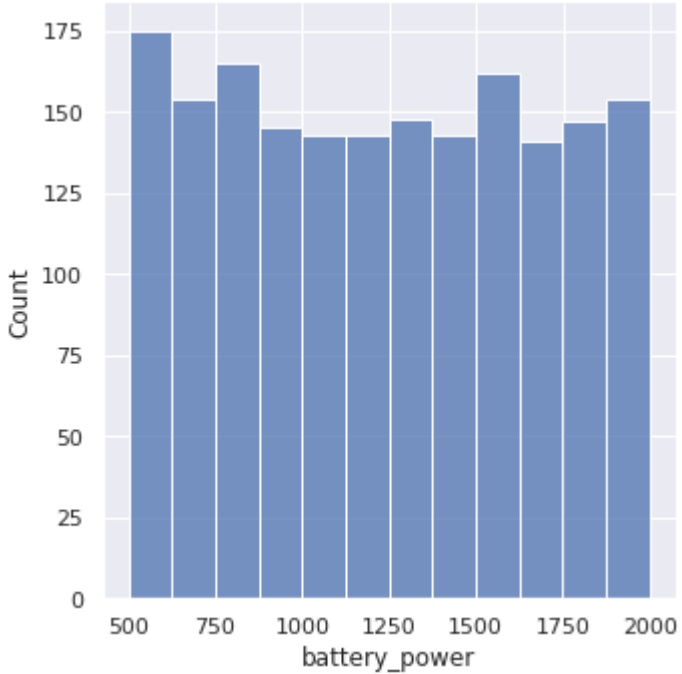
EDA

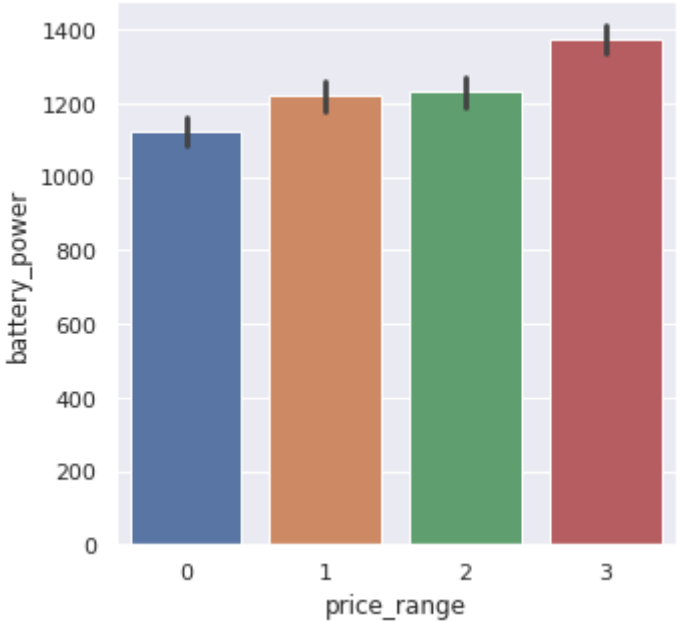

800
600
400
200
0
count
0
1
blue
800
600
400
200
0
count
0
1
dual_sim


Has Bluetooth
50.4%
49.6%
Not suported

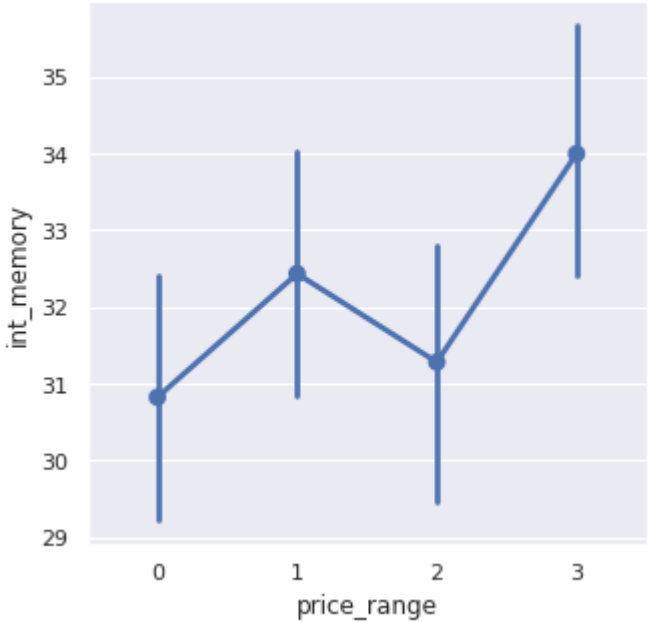

Has dual sim
50.4%
49.6%
Not suported

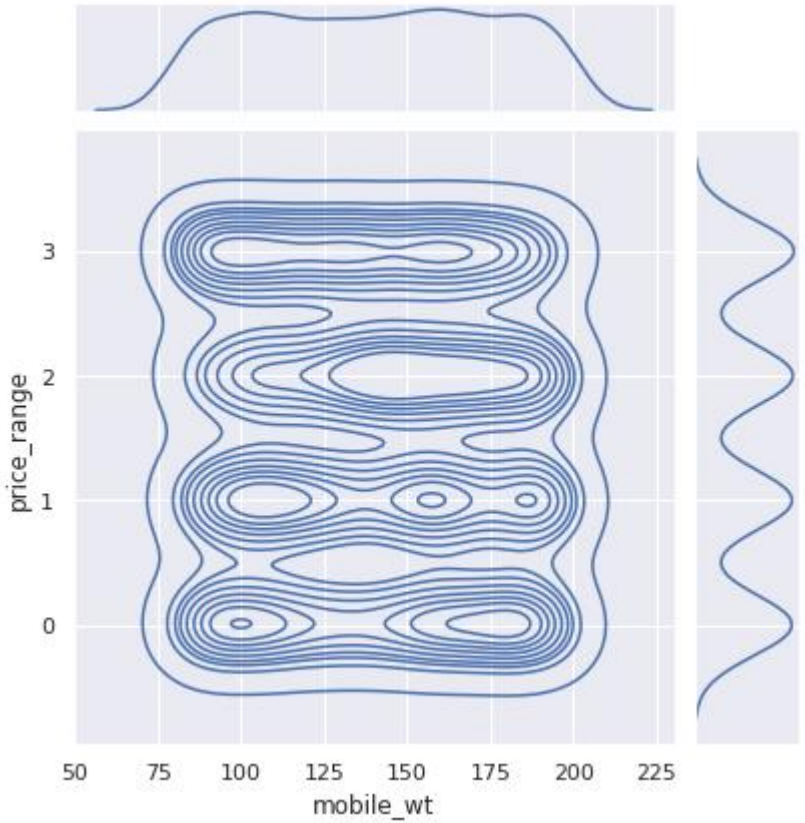
EDA


175
150
125
100
75
50
25
0
Count
500
750
1000
1250
1500
1750
2000
battery_power

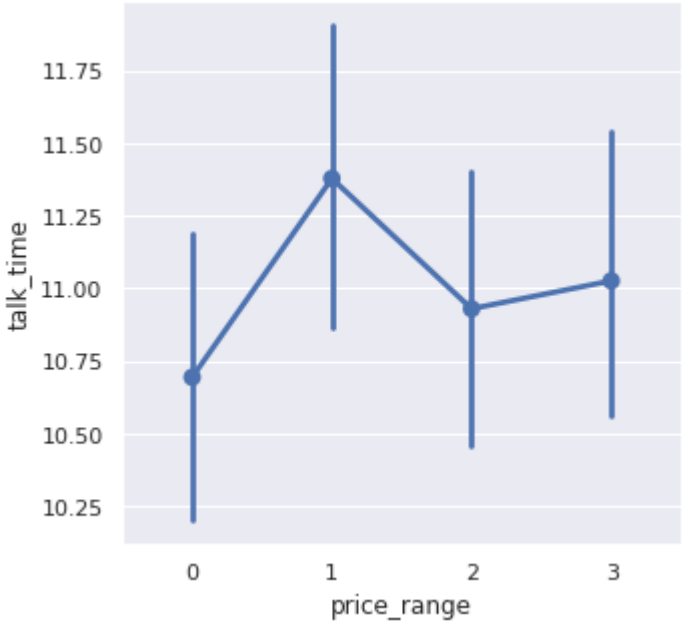

1400
1200
1000
800
600
400
200
0
battery_power
0
1
2
3
price_range

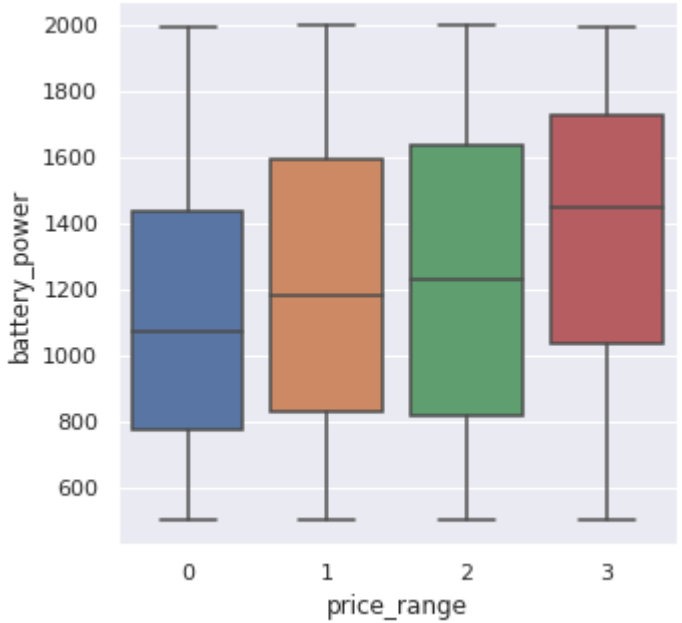
EDA


int_memory
35
34
33
32
31
30
29
0
1
2
3
price_range

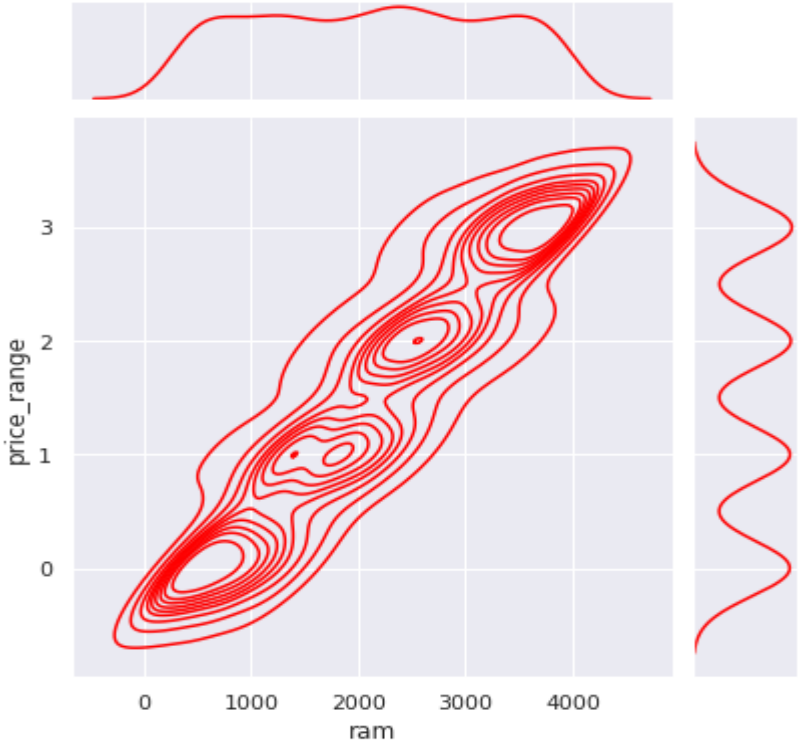

price_range
3
2
1
0
50
75
100
125
150
175
200
225
mobile_wt

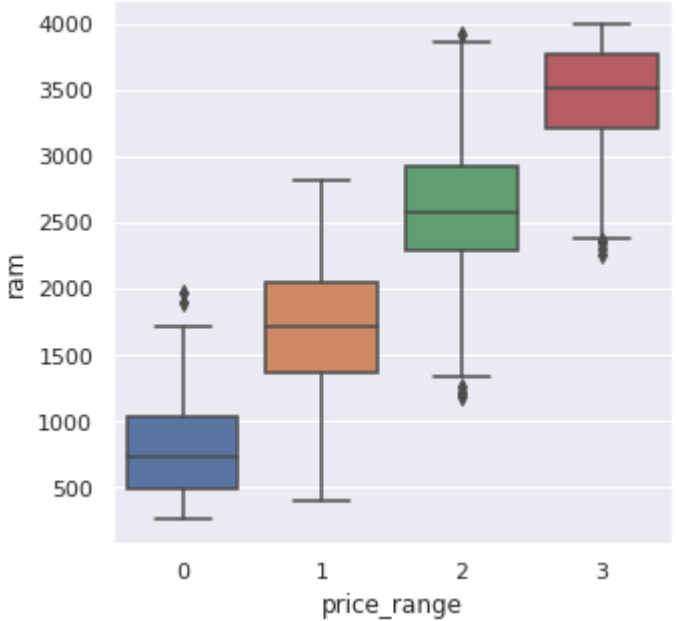
EDA


11.75
11.50
11.25
11.00
10.75
10.50
10.25
talk_time
0
1
2
3
price_range

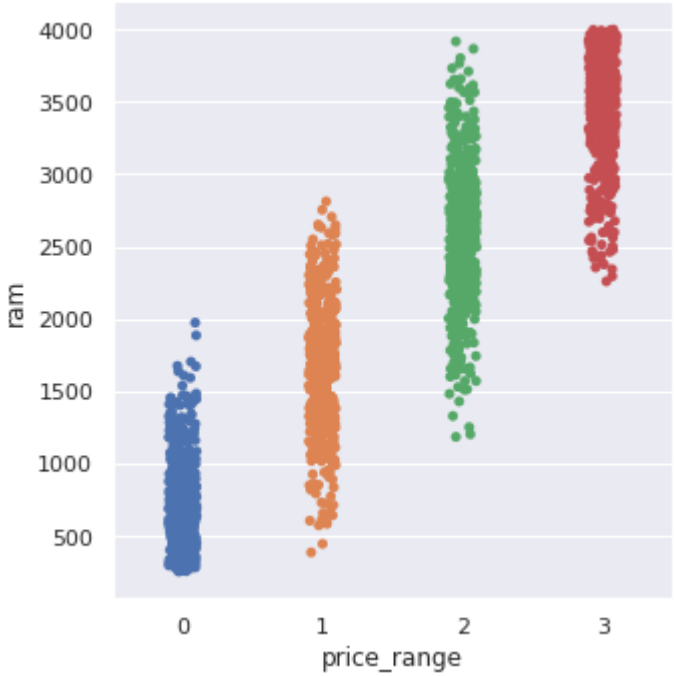

2000
1800
1600
1400
1200
1000
800
600
battery_power
0
1
2
3
price_range

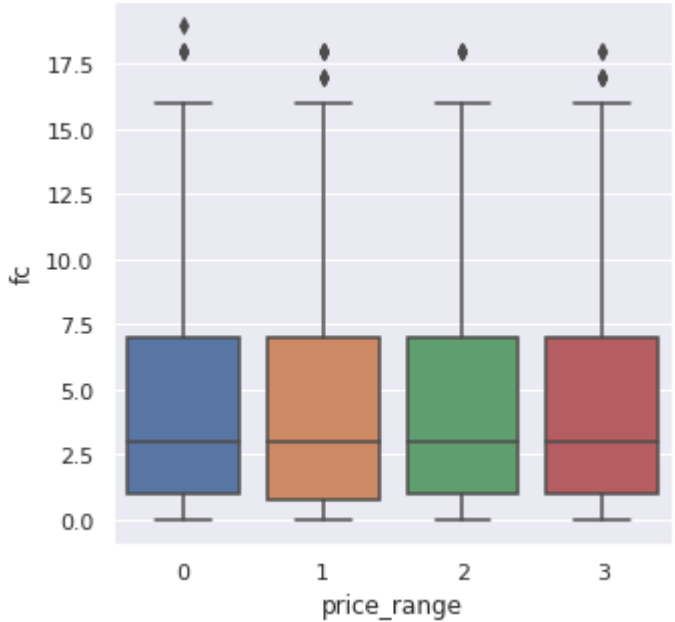
EDA


price_range
ram
0
1
2
3
0
1000
2000
3000
4000


ram
price_range
4000
3500
3000
2500
2000
1500
1000
500
0
1
2
3

EDA


4000
3500
3000
2500
2000
1500
1000
500
ram
0
1
2
3
price_range


17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
fc
0
1
2
3
price_range

Model Fitting

Gaussian NB classifier

Gaussian NB Classifier Accuracy Score: 0.8461538461538461

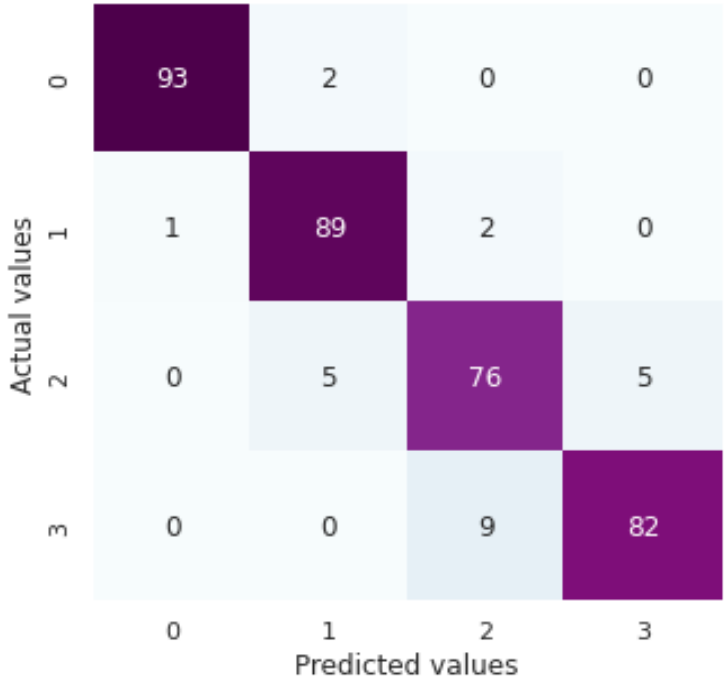
	precision	recall	f1-score	support
0	0.93	0.92	0.92	95
1	0.79	0.73	0.76	92
2	0.74	0.80	0.77	86
3	0.92	0.93	0.93	91
accuracy			0.85	364
macro avg	0.84	0.85	0.84	364
weighted avg	0.85	0.85	0.85	364

Model Fitting

KNNClassifier

kNN Classifier Accuracy Score: 0.9340659340659341

	precision	recall	f1-score	support
0	0.99	0.98	0.98	95
1	0.93	0.97	0.95	92
2	0.87	0.88	0.88	86
3	0.94	0.90	0.92	91
accuracy			0.93	364
macro avg	0.93	0.93	0.93	364
weighted avg	0.93	0.93	0.93	364

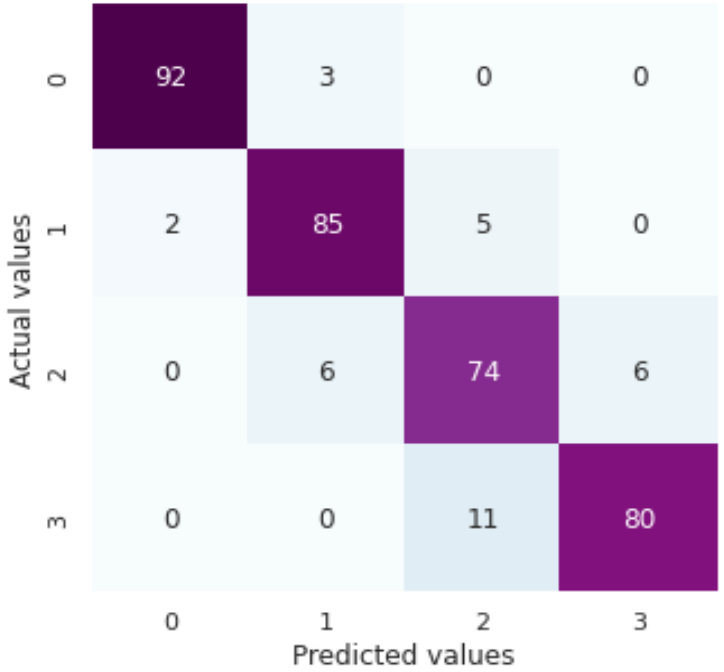


Model Fitting

Random Forest

Random Forest Classifier Accuracy Score: 0.9093406593406593

	precision	recall	f1-score	support
0	0.98	0.97	0.97	95
1	0.90	0.92	0.91	92
2	0.82	0.86	0.84	86
3	0.93	0.88	0.90	91
accuracy			0.91	364
macro avg	0.91	0.91	0.91	364
weighted avg	0.91	0.91	0.91	364

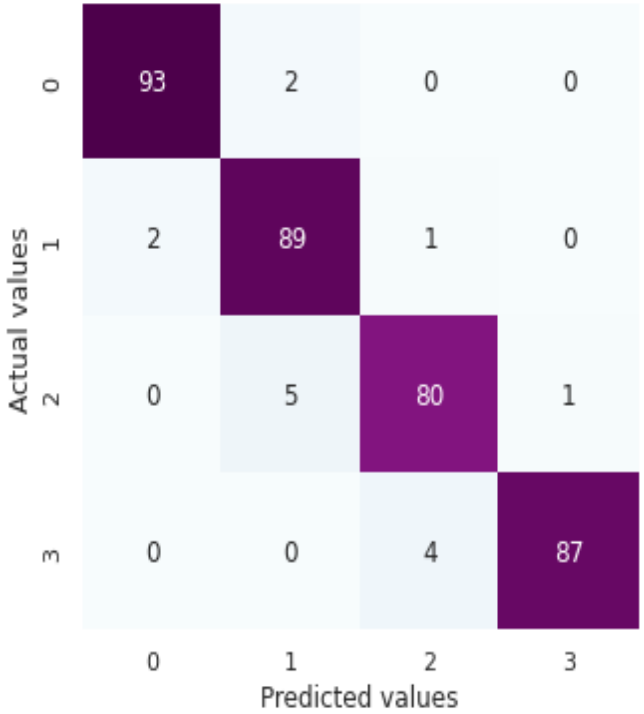


Model Fitting

SVM Classifier

SVM Classifier Accuracy Score: 0.9587912087912088

	precision	recall	f1-score	support
0	0.98	0.98	0.98	95
1	0.93	0.97	0.95	92
2	0.94	0.93	0.94	86
3	0.99	0.96	0.97	91
accuracy			0.96	364
macro avg	0.96	0.96	0.96	364
weighted avg	0.96	0.96	0.96	364



Hyperparameter tuning(SVM Classifier)

Gridsearch CV

grid prediction Accuracy Score: 0.9587912087912088

	precision	recall	f1-score	support
0	1.00	0.97	0.98	95
1	0.95	0.96	0.95	92
2	0.92	0.94	0.93	86
3	0.97	0.97	0.97	91
accuracy			0.96	364
macro avg	0.96	0.96	0.96	364
weighted avg	0.96	0.96	0.96	364

Model Selection

Observation1 - as seen in the table, Gaussian NB classifier is not giving great results and Random Forest are had better than Gaussian NB classifier but still have less accuracy and having lesser recall value.

Observation2 – KNN Classifier & SVM Classifier performed good in terms of test accuracy.

Observation3 - There is no difference here, in between SVM classifier modeling and after hyperparameter tuning in SVM classifier.

The SVM(support vector machine) classifier model has high accuracy of 96% . so, finally this model is best for predicting the Mobile price Classification.

Conclusion

79.1% of mobile phones support 3G and 29.9% of mobile phone doesn't.

Ram of phones has increased with the cost of mobile phones.

Battery power increased with the price ranges.

There is almost similar distribution in internal memory of all price ranges.

We looked at Classification, When we compare the accuracy score of all the models, the SVM(support vector machine) model has high accuracy of 96% . so, finally this model is best for predicting the Mobile price Classification.