



Capstone project 3

MOBILE PRICE RANGE PREDICTION

Sangamesh chandankera

CONTENT

- Introduction
- Problem statement
- Data summary
- Exploratory Data Analysis (EDA)
- Data Visualization
- Feature Engineering & Selection
- Building and Evaluating Model
- Conclusion

INTRODUCTION

- To predict “If the mobile with given features will be Economical or Expensive” . Different classifiers are used to achieve as higher accuracy as possible.
- Results are compared in terms of highest accuracy achieved and minimum features selected. Conclusion is made on the base of best feature selection algorithm and best classifier for the given dataset.
- This work can be used in any type of marketing and business to find optimal product (with minimum cost and maximum features). We do not have to predict the actual price but a price range indicating how high the price is.

PROBLEM STATEMENT

Your task is to build a model that predicts price range indicating how high the price is. In the competitive mobile phone market companies want to understand sales data of mobile phones and factors which drive the prices. The objective is to find out some relation between features of a mobile phone (eg:- 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.

DATA SUMMARY

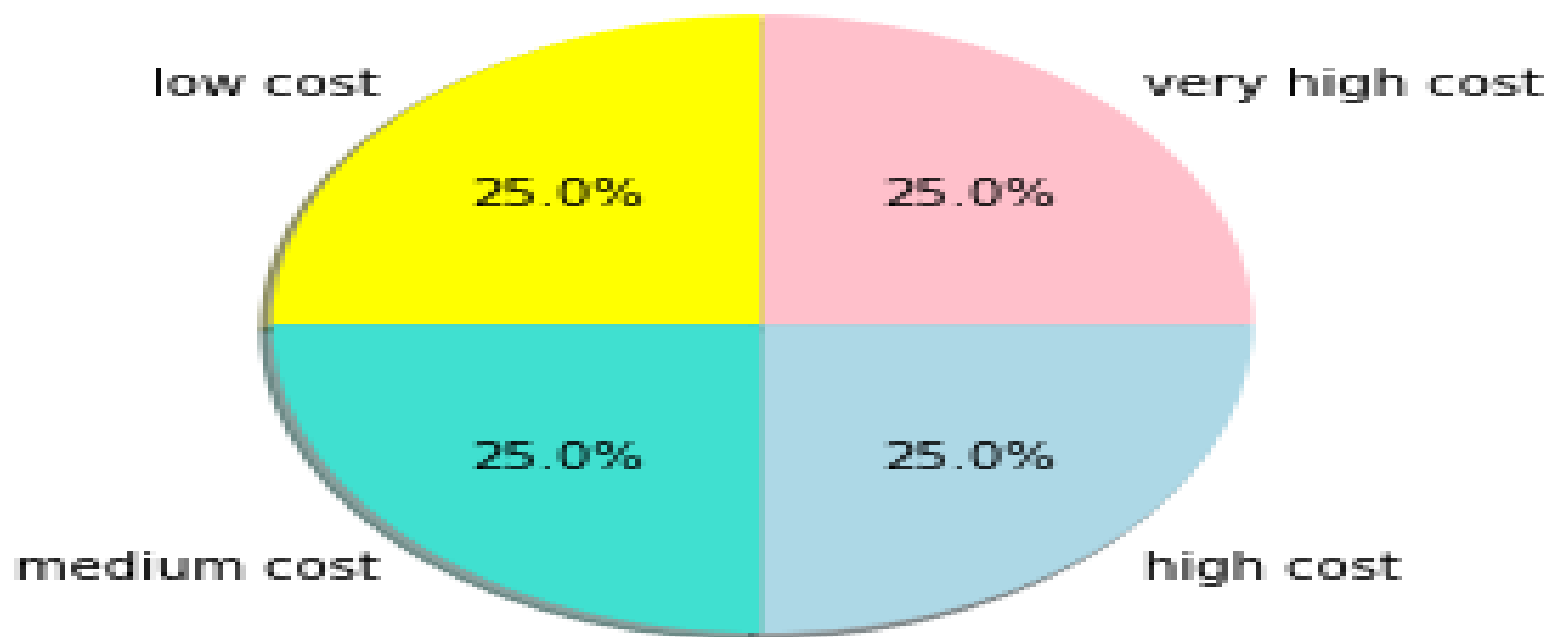
- Battery_power - Total energy a battery can store in one time measured in mAh
- Blue - Has bluetooth or not
- Clock_speed - speed at which microprocessor executes instructions
- Dual_sim - Has dual sim support or not
- Fc - Front Camera mega pixels
- Four_g - Has 4G or not
- Int_memory - Internal Memory 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).

BASIC EXPLORATION

- The dataset contains 2000 rows and 21 columns.
- The dataset has variables only numerical features in it.
- No null values present.
- No duplicate values present.
- It is a multi-class classification problem.

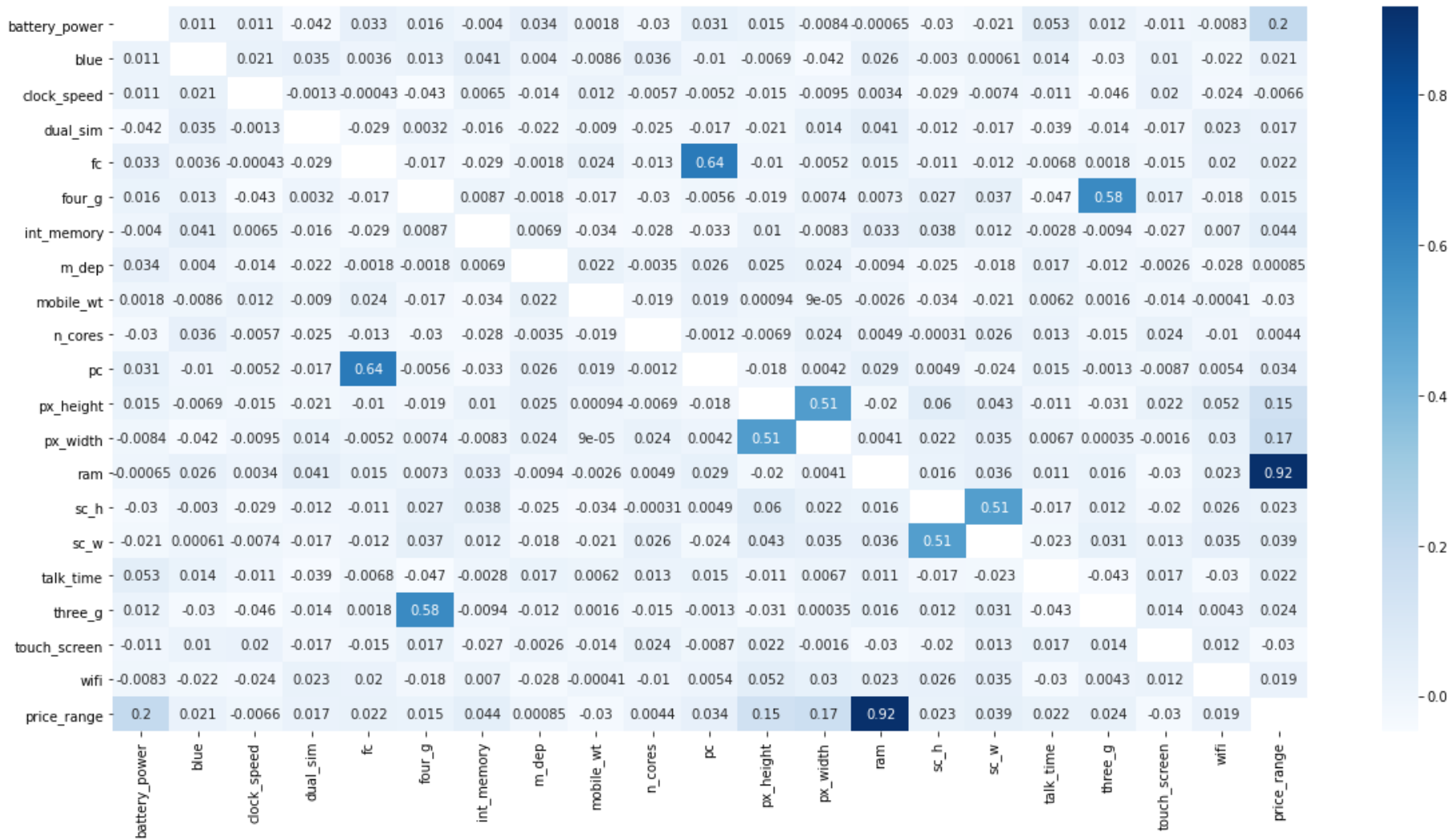
Dataset is balanced?

balanced or imbalanced?

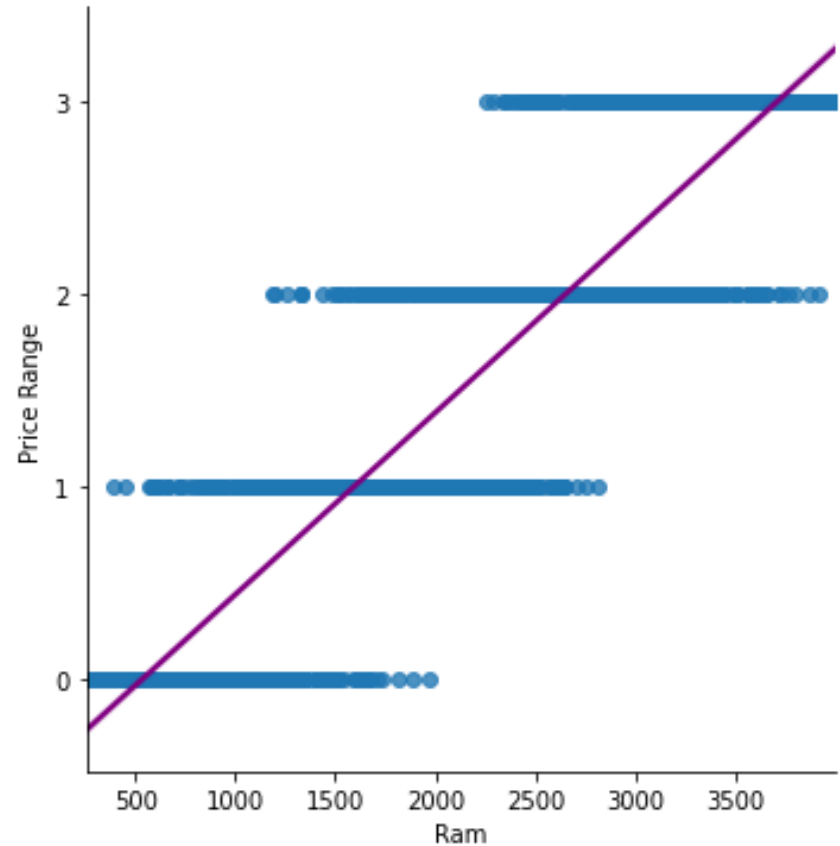
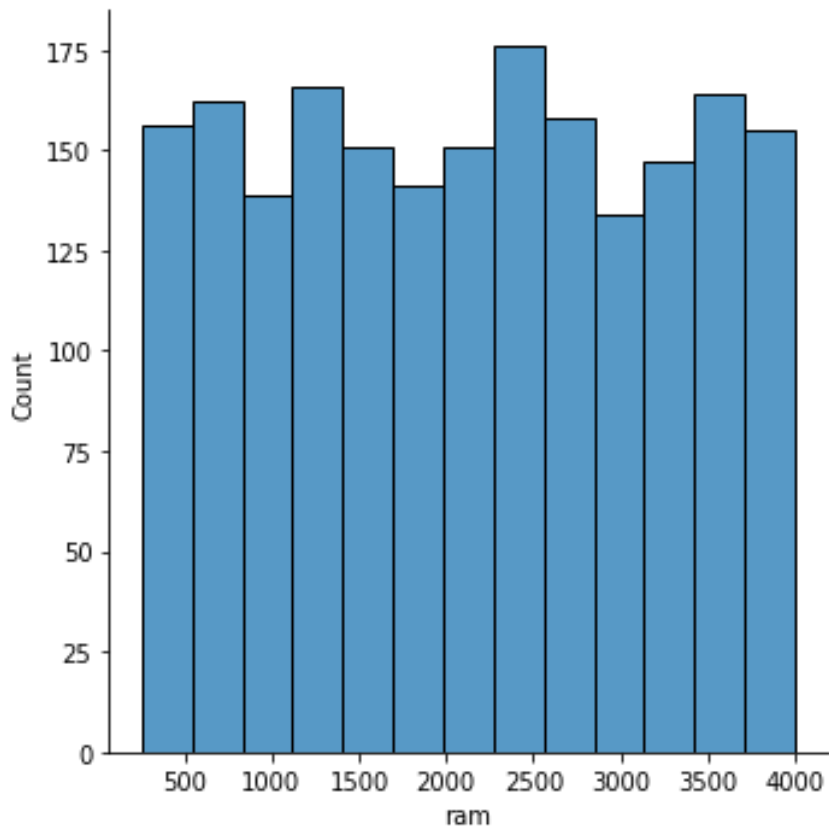


DATA VISUALIZATION

AI

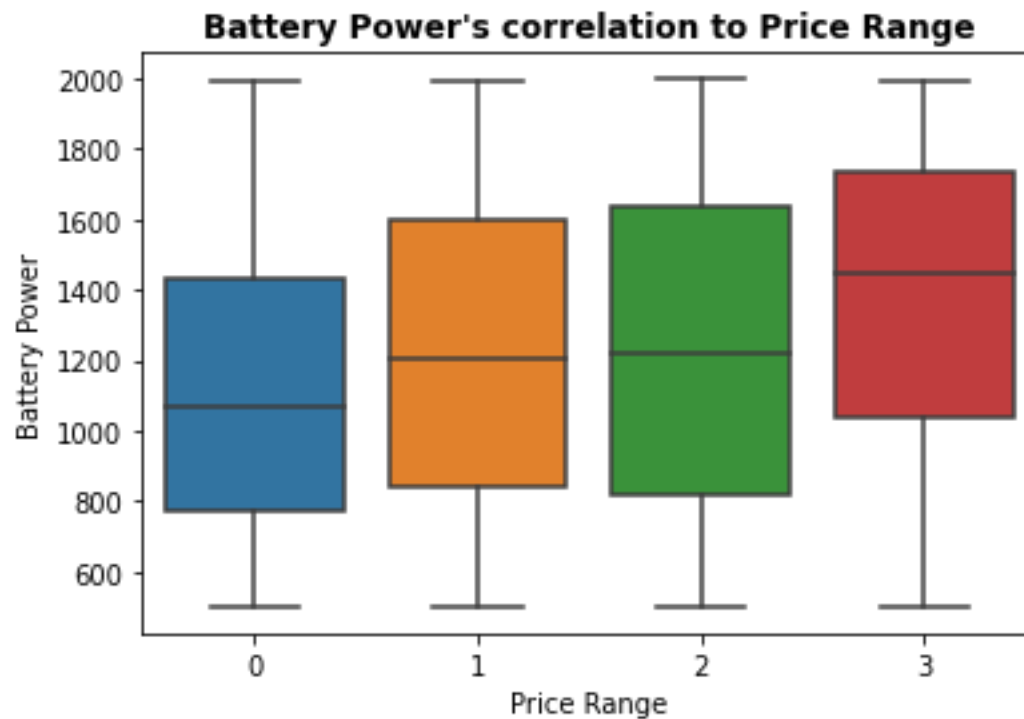


- The most influential variable is ram.
- Most of the variables have very little correlation to price range.
- Primary camera mega pixels and front Camera mega pixels have correlation but they do not effect prige range.
- 3G and 4G is somewhat correlated.

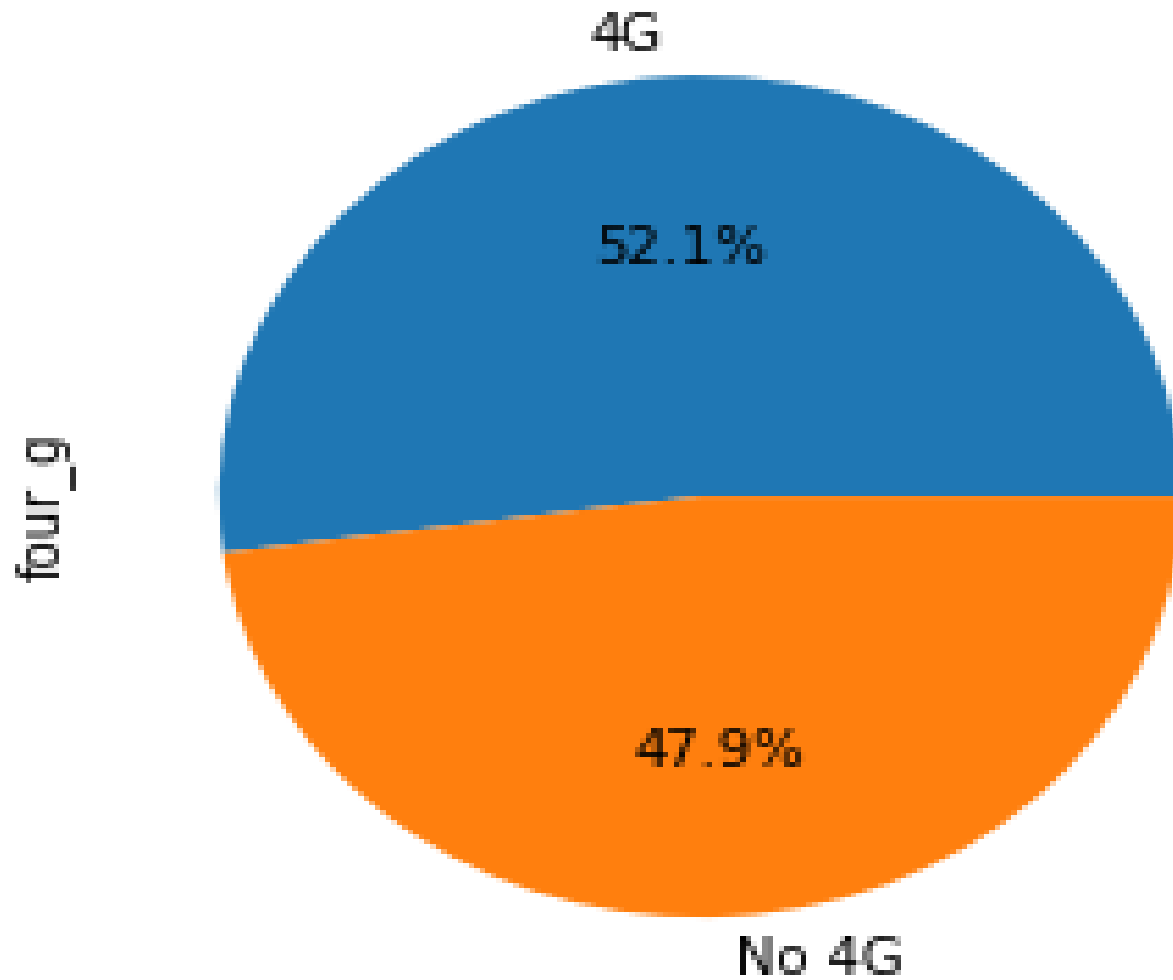


- Most of the mobile phones used are around 2500mb RAM.
- It shows high correlation between ram and price range.
- As Ram increases, mobile's price also increases.

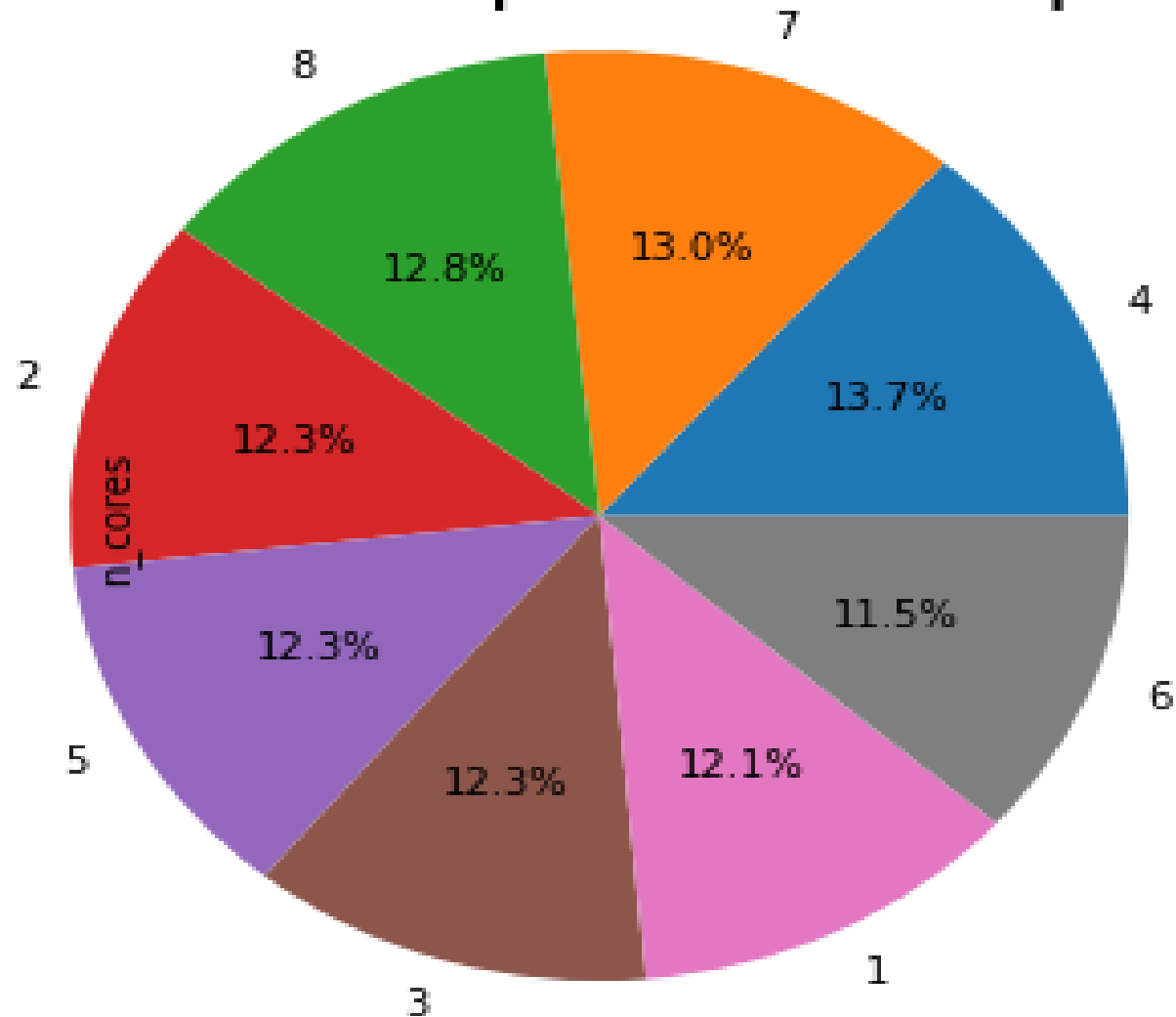
Analysis Details



Percentage of Mobiles with 4G

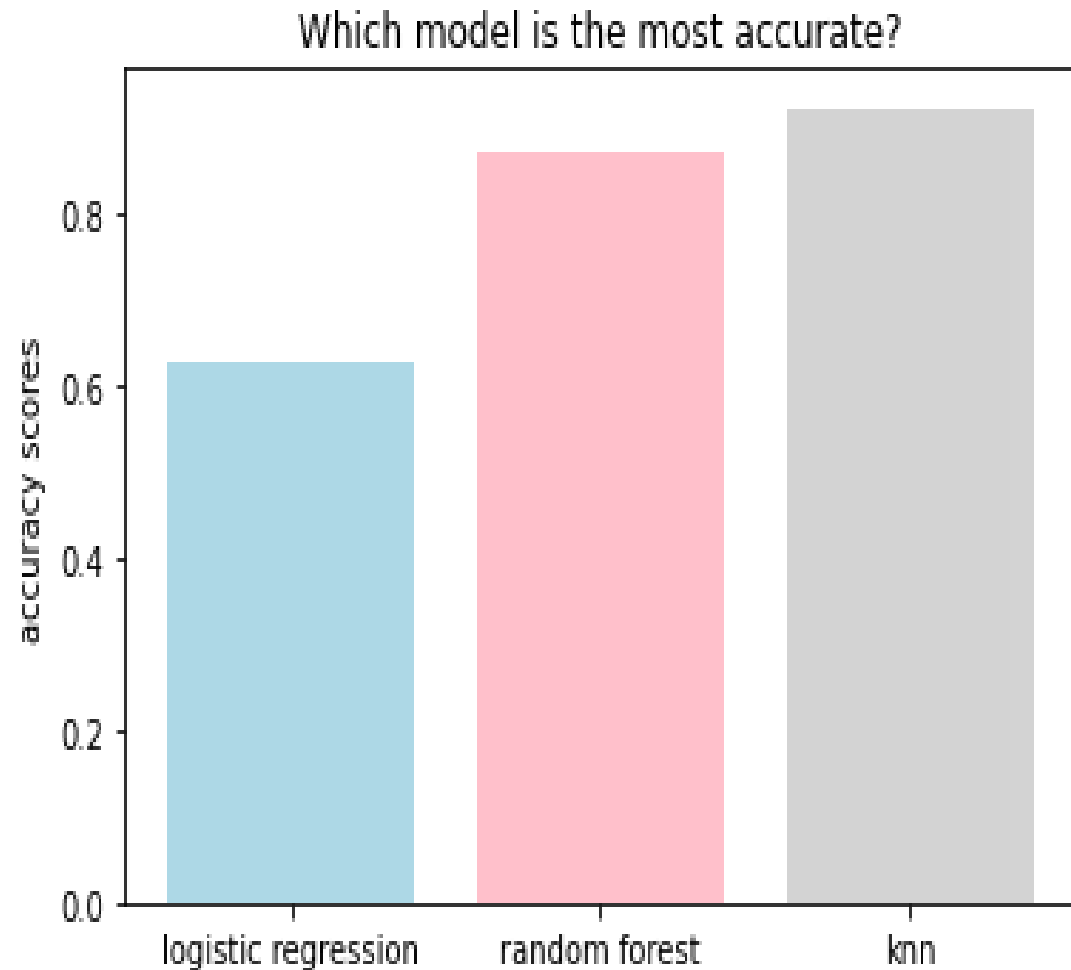


Number of cores of processor in mobile phones



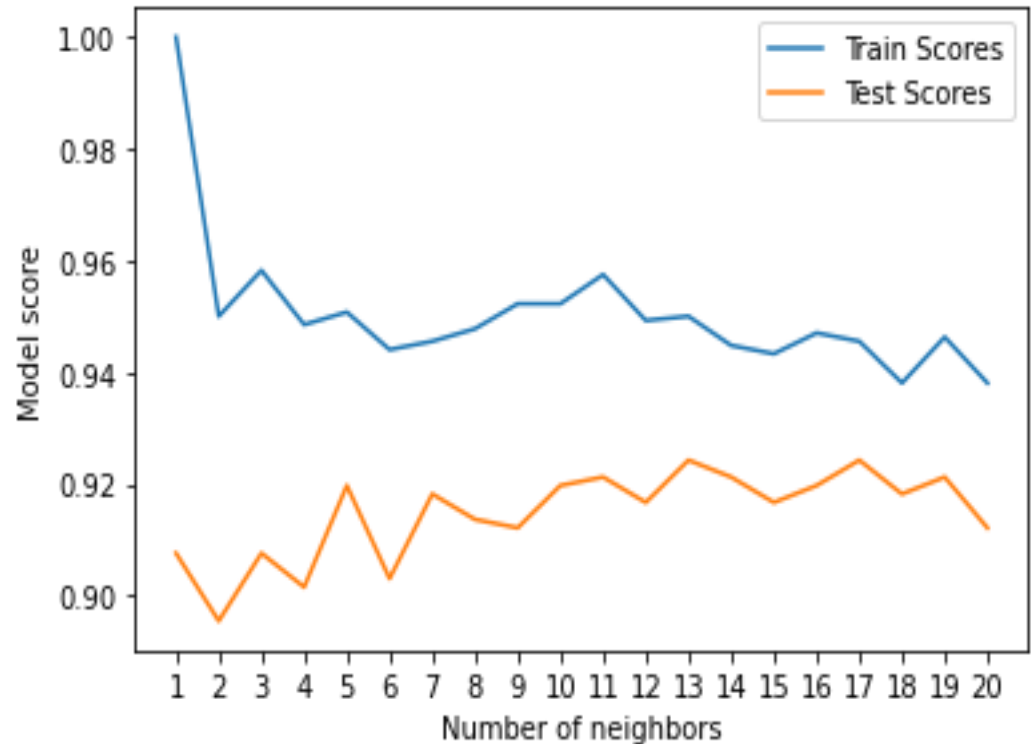
ML ALGORITHMS

- I have used 3 machine learning classification algorithms such as Logistic regression, Random Forest, KNN.
- The accuracy of each algorithm came out to be
 1. KNN : 0.919,
 2. Logistic Regression : 0.627
 3. Random Forest : 0.868



HYPERPARAMETER TUNING

- After applying hyperparameter tuning, the accuracies of algorithms are
KNN : 92.42%
- Random Forest : 87.72%



XGBoost

- XGBoost comes under boosting and is known as extra gradient boosting.
- GBM first calculates the model using X and Y then after the prediction is obtain.
- It will again calculates the model based on residual of previous model.
- loss function will give more weightage to error of previous model. and this process continuous until MSE gets minimizes.
- Accuracy before hyperparameter tuning on XGBoost : 89.69%.
- Accuracy After hyperparameter tuning on XGBoost : 90.90%.

Final Metrics Conclusion

	precision	recall	f1-score	support
0	0.96	0.98	0.97	178
1	0.91	0.90	0.90	163
2	0.87	0.88	0.88	161
3	0.95	0.93	0.94	158
accuracy			0.92	660
macro avg	0.92	0.92	0.92	660
weighted avg	0.92	0.92	0.92	660

Conclusion

- The most important features in predicting a mobile phone's price are ram, battery power, pixel height and pixel width! We figured this out by using a correlation matrix, specifically looking at the most highly correlated variables to price range.
- We used machine learning to predict mobile phones price's using all of the features in our dataset. Found that the best performing model was KNN outperforming Logistic Regression, Random Forest, and even XGBoost. Even able to improve our KNN model's score by tuning its hyperparameters (n_neighbors).
- KNN model using other metrics and saw that it performed very well by those metrics, as well. The fact that KNN was the best performing model to the idea that sometimes the more complicated models might not be the best model for a given dataset

Challenges

- Hyper parameter tuning part was bit complicated because even after applying it i didn't much bigger difference in the model performance.
- I was expecting the XGboost model would give more accuracy score but it turned out to be with KNN classifier as high accuracy.