**Capstone Project**

**Mobile Price Range Prediction**

**Rohit Madame**

**Abstract:**

Mobile phone market is very competitive market .Mobile phones comes in all sort of prices, features, specifications etc. It is important for the customer to have a mobile price prediction as per their price range and best features.

**Problem Statement:**

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.

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)

**Steps Involved:**

* Data Preparation
* Data Cleaning and Feature Engineering
* EDA
* Model selection and Evaluation
* Comparing Different Models
* Conclusion

**Data Preparation:**

Importing all necessary libraries and mobile price range data. This dataset contains 2000 rows and 21 columns.

**Data Cleaning and Feature Engineering:**

This dataset doesn’t contain null values or duplicate values. But we found that pixel height and screen width has minimum value 0. That is practically not possible. So in screen width we have replaced it with mean value and pixel height has 0 value in only two rows so we dropped it.

**EDA:**

It is a good practice to understand the data first and try to gather as many insights from it. EDA is all about making sense of data in hand, before getting them dirty with it. First we have plotted Heatmap which shows the correlation of dependant variable to independent variable. Ram, Battery\_power, pixel height and pixel width are correlated with the price range. We saw that as ram increase the price range of mobile also increases. Battery power between 1200 and 1300 mAh falls under price range of 1 and 2.For battery more than 1300 mAh has in to the very high cost price group. As pixel height higher than 700 falls into the very high cost price group and for less than 700 it falls under medium cost. For pixel width more than 1300 it is in the very high cost price group and for more than 1200 and less than 1300 it is in the 1 and 2 group. Then we have plotted pie chart for Bluetooth, dual sim, four\_g, three\_g, touch\_screen and wifi. Then we prepared pie chart for our dependant variable and found that our target variable is uniformly distributed. Finally we plotted box plot for numerical value and obtained outliers in the pixel height column and front camera column. We removed outliers by IQR method.

Upper\_limit=Q3+1.5\*IQR.

IQR=Q3-Q1

**Model selection and Evaluation:**

1. **Gradient Boosting Classifier:**

We trained model and got the accuracy as follows:

* Train score: 1.0
* Test score: 0.8943

1. **Gradient Boosting Classifier(Hyperparameter Tunning):**

* Train score: 1.0
* Test score: 0.8909

1. **Random Forest Classifier:**

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

* Train score – 1.0
* Test Score - 0.8705

1. **Random Forest Classifier(Hyperparameter Tunning):**

* Train score: 1.0
* Test score: 0.8807

1. **KNearestNeighbors:**

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

* Train score: 0.9509
* Test score: 0.9182

1. **KNearestNeighbors(Hyperparameter Tunning):**

* Train score: 0.9491
* Test score: 0.9097

**Conclusions:**

* As Ram increase the price range of mobile also increases.
* As the primary camera increases the megapixels, the price range also increases.
* KNearestNeighbors got the best accuracy whereas Random Forest Classifier got least accuracy.
* We have checked for the feature importance's of each model. RAM, Battery Power, pixel height and pixel width contributed the most while predicting the price range.

**References:**

* GeeksforGeek
* Almabetter
* stackoverflow