Mobile Price Range Prediction

Girish R
Data science trainee, AlmaBetter,
Bengaluru

Abstract:

During the purchase of mobile phones, various features like memory, display, battery, camera, etc., are considered. People fail to make correct decisions, due to the non-availability of necessary resources to cross-validate the price. To address this issue, a machine learning model is developed using the data related to the key features of the mobile phone. The developed model is then used to predict the price range of the new mobile phone. Machine learning algorithms namely KNN Classifier, Random Forest Classifier (RFC), Gradient boosting Classifier, XGBoost classifier, Logistic Regression, Decision Tree Classifier, SVM are used to train the model and predict the output as low, medium, high or very high. In order to improve the classification accuracy feature selection method is used.

1. Problem statement:

There are many things we consider before buying a mobile as we used our mobile for various purpose like connecting with our family , playing games ,taking a photos to keep our memory. So this such specifications such as RAM, internal memory ,Wi-Fi , 3G/4G connectivity etc. plays important role to buy a mobile. To analysis of this important factor from time to time and come up with the best setoff specifications and price ranges so that people will buy the mobile. Hence through the various ML modules we will help the company to estimate the price range of mobiles according to feature so the maximum amount of Sell will pe possiblle.

2. Data set 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 inGigabytes
- ❖ 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 MegaBytes
- ❖ 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).

3. Introduction:

We have been provided with the dataset containing the 2000 rows and 21 feature. From the given dataset we have to predict the mobile price range for the mobile company to make the more sells. We consider the important feature which are actually affecting the price of the mobile and build the ML module. For that we first perform the EDA and also the correlation of the feature with the target feature. We build the following modules KNN, Random forest, Gradient Boosting Classifier, XGBoost Classifier, Logistic Regression, Decision Tree Classifier, SVM, to find the best modules which gives us the best price range prediction.

4. Factors Affecting:

Following are the factors affecting the for price range of mobile:

1. **RAM:** As the ram size increase the cost of the mobile price also increases.

2. **Battery Power** : As the Battery Power increases the price range is also gradually increases.

3. **Internal Memory** : We observed that there is high price range as the Internal memory size increases.

4. **Four_G** : 3G /4G mobile phone are categories in high range price.

5. **Clock speed:** Clock speed has negative correlation with the target variable.

6. **Wi-Fi:** mobiles which having wifi facility tend to move in high price range mobile group.

5. Steps involved:

The following steps are involved in the project.

1. **Exploratory Data Analysis:**
   After loading and reading the dataset in notebook, we performed EDA. Comparing target variable which is price range with other independent variables.
   This process helped us figuring out various aspects and relationships among the target and the independent variables and also, we observed the distribution of variables.
   It gave us a better idea that how feature behaves with the target variable.

2. **Preprocessing data:**
   Dataset contains a no null values also no duplicate values are found to disturb the accuracy.
   Dropping the unwanted columns from the dataset.

3. **Features selection:**
   With the help of exploratory data analysis, we analyzed the categorical as well as numerical features in the dataset. And selecting six main feature which are affecting the most.

4. **Correlation Analysis:**
   We plot the heatmap to find the correlation between both dependent

variable and independent variables.

5. Train test Split:

In train test split we take 'x' as dependent variables and 'y' take as independent variable then train the model.
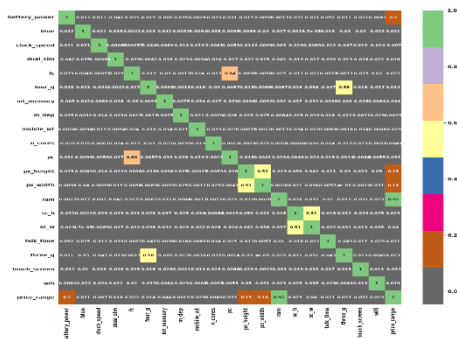
6. Models:

We use 7 modeling to train the data and for predicting the accuracy, RMS and R2.

a. KNN Classifier
b. Random Forest
c. Naïve Bayes Theorem
d. XGBoost Classifier
e. Logistic Regression
f. Decision Tree Classifier
g. SVM

6. Correlation Analysis:

We plot the heatmap to find the correlation between all the columns and observed that:

1. The highest correlation with respect to the price range is RAM which shows as the RAM increase the Price Range is also increases.
2. Also as the Battery Power size increase the cost of mobile price range increase.
3. The dual sim and Four_G facility

having positive correlation with the Price range.
4. There is negative correlation of the mobile weight with respect to the Price Range of mobile as the weight increases the price decreases.
5. Primary camera and the Front camera have a correlation with price range.

7. Modeling Results:

1. KNN Classifier:

We train model by KNN Classifier, and weget results as follows:

- Train score – 63%
- Test Score – 40%

2. Random Forest:

By performing Random Forest, we getthe results are as follows:

- Train score – 99%
- Test Score – 85%

3. XGBoost Classifier:

By performing XGBoost Classifier we get the results are as follows:

- Train score - 100
- Test Score - 88

4. Naïve Bayes Theorem:

By performing Naïve Bayes theorem, we get the results as follows:

- Train Score – 81%
- Test score - 78%

5. Logistic Regression:

By performing Logistic Regression

Classifier we get the results are as follows:

- Train score – 91%
- Test Score – 90%

6. Decision Tree Classifier:

By performing Decision Tree Classifier, we get the results are asfollows:

- Train score – 93%
- Test Score – 85%

7. Support Vector Machine (SVM):

By performing SVM we get the results are as follows:

- Train score – 95%
- Test Score – 81%

6. kNeighbors and Xgboost are given best accuracy score 95% test ,93% train and 91% train, 88% test respectively , roc_auc score for kneighbors is 99%.

Conclusions:

1. We build a predictive model, which could help companies to estimate price of mobiles in much effective way.
2. To predict the cost of various different types of products, same procedure can be performed.
3. RAM in the mobile phones is very important feature for the price range prediction of mobile as the ram and battery power increases the price range increases.
4. According to the user specifications the camera plays the important role to attract the customer .
5. Customer prefers the longer Battery backup for long lasting.