Mobile Price Range Prediction

Sandhy kumari sah
Data science trainees,
AlmaBetter, Bangalore

Abstract:

People fail to make correct decisions, due to the non-availability of necessary resources to cross-validate the price. During the purchase of mobile phones, various features like memory, display, battery, camera, etc., are considered. 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, and GridsearchCV 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 purposes like connecting with our family, playing games, and taking photos to keep our memory. So specifications such as RAM, internal memory, Wi-Fi, 3G/4G connectivity, etc. play an important role to buy a mobile. To the 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

features so the maximum amount of sales will be possible.

We have been provided with the dataset containing the 2000 rows and 21 features. From the given dataset we have to predict the mobile price range for the mobile company to make more sells. We consider the important feature which is 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, and GridsearchCV to find the best modules which give us the best price range prediction.

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 - the 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).

3. Factors Affecting:

Following are the factors affecting the 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 also gradually increases.

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

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

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

6. **Wifi**: mobiles that have wifi facilities tend to move in the high price range mobile group.

4. Steps involved:

The following steps are involved in the project

1. Exploratory Data Analysis:

After loading and reading the dataset in a notebook, we performed EDA. Comparing the target variable which is price range with other independent variables.

This process helped us figure out various aspects and relationships between the target and the independent variables and also we observed the distribution of variables.

It gave us a better idea that how the feature behaves with the target variable.

2. Preprocessing data:

The dataset contains 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 features which are affecting the most.

4. Correlation Analysis:

We plot the heatmap to find the correlation between both the dependent variable and independent variables.

5. Train test Split:

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

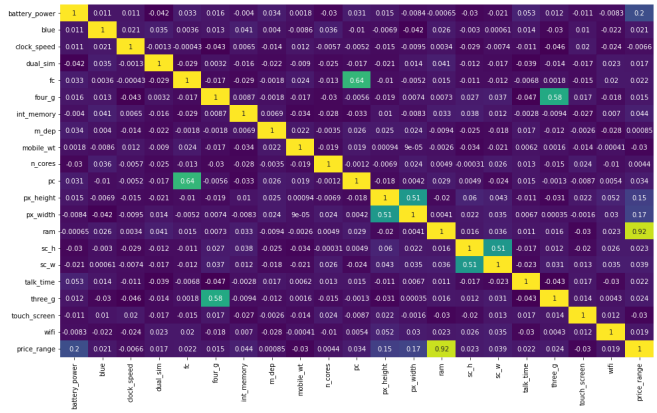
6. Models:

We use 7 modeling to train the data and for predicting the accuracy, RMS and R2. 1. KNN Classifier

2. Random Forest
3. Gradient Boosting Classifier
4. XGBoost Classifier
5. Logistic Regression
6. Decision Tree Classifier
7. SVM

5. 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 increases the Price Range also increases.
2. Also as the Battery Power size increase the cost of the mobile price range increase. 3. The dual sim and Four_G facility have a positive correlation with the Price range.
4. There is a negative correlation between the mobile weight with respect to the Price Range of the mobile as the weight increases the price decreases.
5. Primary camera and the Front camera has a correlation with a price range.

6. Modeling Results:

1. KNN Classifier:

We train the model by KNN Classifier and we get results as follows:

- Train score - 0.94
- Test Score - 0.94

1. Random Forest:

By performing Random Forest we get the results as follows:

- Train score - 0.89
- Test Score - 0.83

2. Gradient Boosting Classifier:

By performing Gradient Boosting Classifier we get the results are as follows:

- Train score – 0.90

• Test Score - 0.90

3. XGBoost Classifier:

By performing XGBoost Classifier we get the results as follows:

• Train score – 1.0

• Test Score - 0.91

4. Logistic Regression:

By performing Logistic Regression
The classifier we get the results are as follows:

• Train score – 0.78

• Test Score – 0.76

5. Decision Tree Classifier:

By performing the Decision Tree Classifier we get the results as follows :

• Train score – 0.91

• Test Score – 0.83

6. Support Vector Machine(SVM):

By performing SVM we get the results as follows:

• Train score - 0.95

• Test Score – 0.95

7. Conclusions:

1. We build a predictive model, which could help companies to estimate the price of mobiles in a much more effective way.
2. To predict the cost of various different types of products, the same procedure can be performed.
3. RAM in mobile phones is a 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 an important role to attract the customer .
5. Customer prefers the longer Battery backup for long-lasting.
6. Kneighbors and Xgboost are given the best accuracy score 94%, 91% test respectively, roc_auc score for kneighbors is 99%.