

Capstone Project-3

MOBILE PRICE RANGE PREDICTION

Supervised Machine Learning (Classification)

BY: NISARGA C

Table of Content:-

1. Abstract
2. Introduction
3. Problem Statement
4. Data Description
5. Exploratory Data Analysis
6. Conclusion

ABSTRACT:

Mobile phone has become a common commodity and usually the most common purchased item. Thousands of types of mobiles are released every year with new features and new specification and new designs. So the real question is prediction is that what is the real price of the mobile and to estimate the price of the mobile within the market for optimal marketing and successful launch of the product. Price has become a major factor for development of any product and its sustainability in the market. Mobile prices also impact the marketing of the mobile and also its popularity with other competitors. With the available specifications and desired designs, money is also an important factor to survive within the market. Customer usually sees that they are able to buy with the specification with the given estimated price or not. So to estimating the price is an important factor before releasing the mobile and also to know about the market and competitors. In this Prediction, Dataset is collected from the existing market and different algorithms are applied to reduce the complexity and also identify the major selection features and get the best comparison within the data .This Tool is used to find the best price with maximum specifications.

Keywords: Machine Learning, Data Collection, Forward Selection, Backward Selection.



INTRODUCTION

Price always has an important impact factor in the product buying aspect and also in the mindset of the buyer who would consider “what is the worth and is it good to buy within this range”. During any product launch into the market, there is a lot of variables and factors are considered and especially in mobiles many features and specification like memory is considered and also the impacting of the cost also may have impact with the competition in the market place. In Mobile there are many specification and features like camera, video, quality of processor, quality of the material. There are many constraints in consideration of the price, as the product should be economical and reachable with overall consideration. Mobile Prices and Specification is mainly considered for selection and comparison. Different tools and Classifiers are used select best features and select the dataset for comparison. Since thousands of mobiles are released each Manuscript received on October 20, 2021. Year so dataset is complex to collect. So with selective feature, it is used to reduce the complexity of the dataset and get the estimate price to get an idea to release the product in the market. In this Prediction, There are many multiple variables to be considered to get the précised results of the price and other features. Of the mobile dataset this will help the buyer and also the marketer and the developer to get precise information from historical data of mobile phones and help them to decide are fine and satisfactory. In this project KNN Model algorithm, it is used to find the distances between K models and model for testing the dataset. The accuracy is also calculated using K model and the training

model is used. The KNN model is used to predict the model for price.

PROBLEM STATEMENT

A man began his own versatile organization. He needs to give intense battle to large organizations like Apple, Samsung and so forth He doesn't have a clue how to appraise cost of mobiles his organization makes.

In this serious cell phone market you can't just expect things. To take care of this difficult he gathers deals information of cell phones of different organizations.

He needs to discover some connection between highlights of a cell phone (egg: RAM, Internal Memory) and its selling cost. However, he leaves something to be desired at Machine Learning. So he needs your assistance to tackle this issue.

Data Description

Battery power - Total energy a battery can store in one time measured in mash

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

Touch_screen - Has touch screen or not

Wifi - Has wifi or not

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

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)

This project consists of five main steps that are mentioned below

- Data Collection.
- Analysis of Data.
- Visualizing the Data.
- Classification of Data.
- Testing.

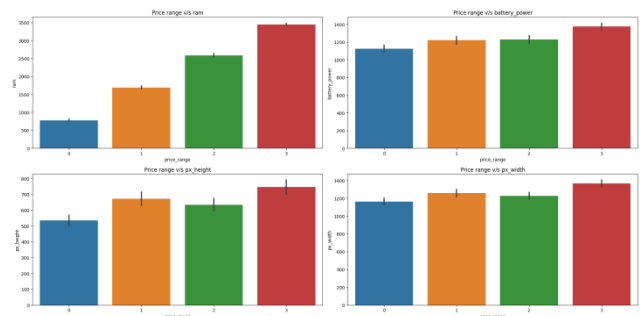
A. Data Collection: Different features and data about mobile are collected. Type of mobile used-Android or Any other Type, Size of the Screen, Memory of RAM, Camera Pixels, thickness and length of the mobile and also the durability of the battery is collected.

B. Analysis of Data: The Data collected is analyzed and it takes the range of minimum and maximum into consideration and takes the count and derives its mean value. The price value is divided into different categories as shown below and also it's divided into training data set and instances data set for more accuracy.

C. Visualizing of Data: Data would be visualized with the price prediction classification model using Elbow Method. Complexity of the model is reduced using principal variables. If there are higher no of constraints and features and variables

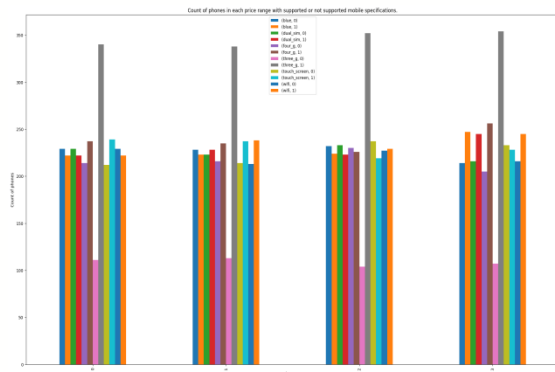
there would be more complexity and harder to visualize the data .so if the features are related we could use reduction algorithms to reduce the complexity. Features selection is used to select a particular dimension and eliminate other features and keep only features which give the most specific information. Extraction of dataset is used to find the specific dataset and features by forward and backward selection in which forward selection we start with no features then we add the important features which would give out the important information and In Backward Selection we eliminate the feature which does not give any information and keep only the main features.

Chart – 1-Relation between RAM, Battery power, x-height and px_width



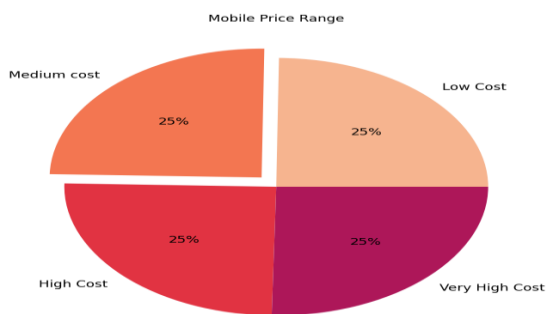
- Mobiles having RAM more than 3000 MB falls under Very high cost category. As RAM increases price range also increases.
- Mobiles having RAM less than 1000 MB falls under low cost category.
- Mobiles with battery power more than 1300 mash have very high cost. And Mobiles with battery power between 1200 and 1300 mash falls under medium and high cost category.
- Mobiles with more than 700 pixel height and width more than 1300 have very high cost.

Chart - 2 - Specification Check



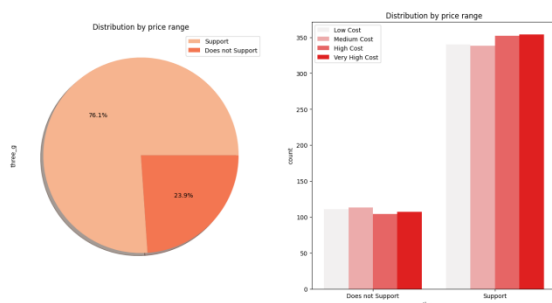
We can see that each price range category has equal number of mobiles phones having both supporting and non supporting specifications.

Chart - 3 - Checking the Price



There are mobile phones in 4 price ranges. So according to these pie chart percentages of all Mobile price range almost similar.

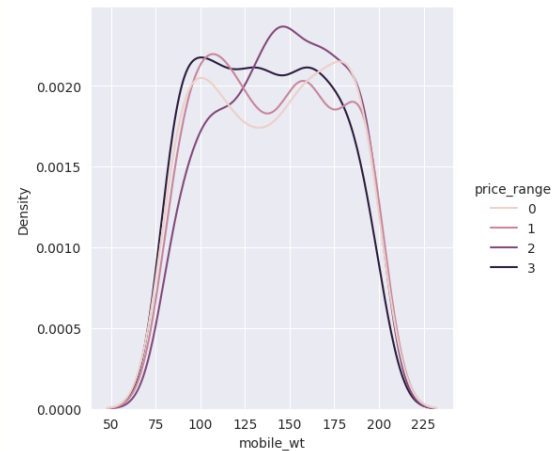
Chart - 4 - 3G & 4G Connectivity



Distribution of price range almost similar of supported and non supported feature in 4G. So that is not useful of prediction.

- Feature 'three' plays an important feature in Price prediction.

Chart - 5 - Mobile Weight



We can see that in this box plot, the costly phones are lighter in weight.

D. Classification of Data: After the data is visualized, the model is designed to find the accuracy where it is modeled with number of correct samples with total number. Classification is done to find the correct accuracy and correct samples instances. Classification is also done with predicated price and features and percentage of accuracy of positive samples with total samples.

E. Testing: With the preprocessed data, the prediction of test data is done with different features where preprocessed data is split for testing and training dataset. The data set is used to find the Accuracy and also the performance and also the predicated value of the product.

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. Pre processing 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. six main feature which area affecting the most.

4. Correlation Analysis:

We plot the heat map 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:

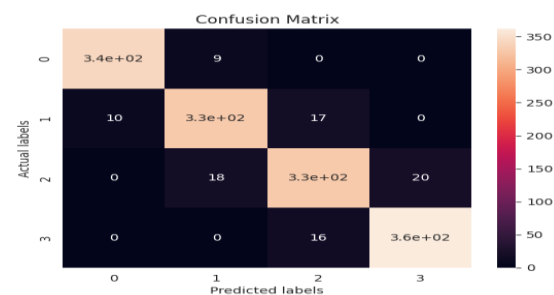
We uses 7 modeling to train the data and for predicting the accuracy, RMS and R2.

- Random Forest
- XG Boost Classifier
- Logistic Regression

ML Model Implementation

ML Model 1 – Building a Logistic Regression Model

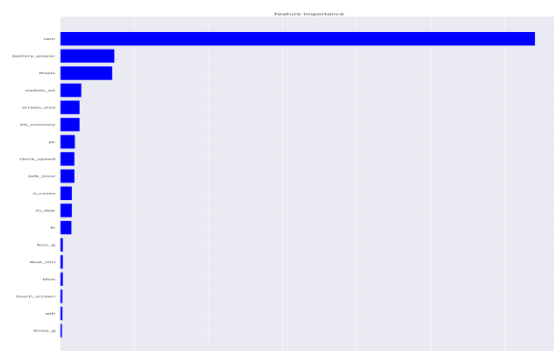
- The precision on test dataset is 0.9419049940445509
- The recall on test dataset is 0.9423076923076923
- The roc auk score on test dataset is 0.9934859517458531



ML Model 2 – Building a Random forest classifier Model

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

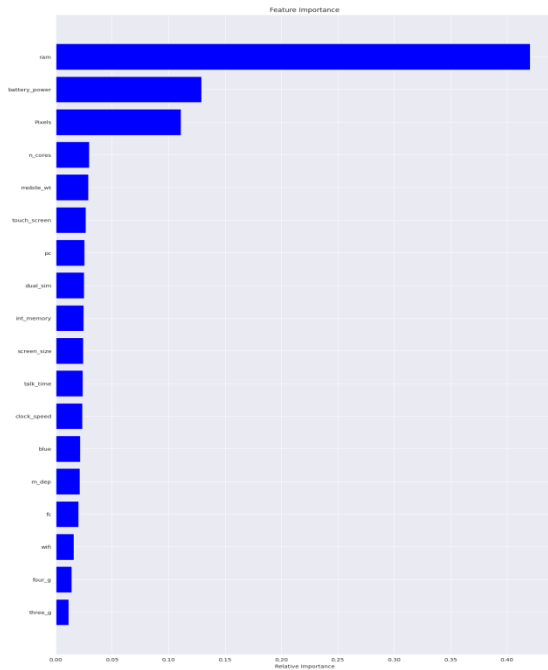
- Trainscore-0.87
- TestScore-0.83



ML Model 3 – Building a XG Boost Model

By performing XG Boost Classifier we get the results areas follows :

- Trainscore-0.91
- TestScore-0.88



- Random Forest
- XG Boost
- Logistic regression shows approx accuracy of 90% and the other two shows approx 80% accuracy, which shows that the data were properly classified.
- Logistic Regression performed better than the other two models.

Logistic Regression performs the best among the three classification methods used.

Random forest and XG Boost perform almost same.

CONCLUSION

- From EDA we can see that, there are mobile phones in 4 price ranges. The number of elements is almost similar.
- Half the devices have Bluetooth, and half don't.
- There is a gradual increase in battery as the price range increases.
- Ram has continuous increase with price range while moving from Low cost to Very high cost.
- Costly phones are lighter.
- RAM, battery power, pixels played more significant role in deciding the price range of mobile phone.

I have implemented 3 classification models and achieved a fairly good result for all the algorithms.

- Logistic Regression