**Mobile Price Range Prediction**

Submitted by:- Subhasis Chattopadhyay

E-mail id: subhasischattopadhyaydbb@gmail.com

### **Problem Statement**

In the competitive mobile phone market companies want to understand sales of mobile phones and factors which drives the prices. The objective is to find out some relation between the 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 Description:**

* Battery\_power: Battery capacity in mAh
* Blue: Has a bluetooth or not
* Clock\_speed: Speed at which microprocessor executes instructions
* Dual\_sim: Has dual sim support or not
* Fc: Front camera megapixals
* Four\_g: Has 4G or not
* Int\_memory: Internal memory capacity
* M\_dep: Mobile depth in cm
* Mobile\_wt: Weight of Mobile phone
* N\_cores: No.of cores in the processor
* Pc: Primary camera megapixals
* Px\_height: Pixal resolution height
* Px\_width: Pixal resolution width
* RAM: Random Access Memory
* Sc\_h: Screen height
* Sc\_w: Screen width
* Talk\_time: Longest that a single battery can last over a call
* Three\_g: Has 3G or not
* Wifi: Has wifi or not
* Price\_range: This represents the value of mobile phone(0: low cost, 1: medium cost, 2: High cost, 3: Very high cost)

### **Data Pre-processing :**

### 1. Getting the dataset

### 2. Importing libraries

### 3. Importing datasets

### 4. Finding Missing Data

### 5. Encoding Categorical Data

## 6. Data Cleaning and Feature Engineering

7. Data Scaling

**Exploratory data analysis :**

1. Firstly Checked correlation between Each Features in dataset.
2. checked if bluetooth available or not???
3. Checked can Battery Power be affected in mobile Price?? and how the battery mAh is spread
4. Checked can Ram, Screen Height ,Screen width,pixel Height ,pixel width be affected in mobile Price??
5. Checked can 4G and 3G be affected in mobile Price??
6. Checked can FC (front camera megapixels) ,PC (Primary camera Megapixels) and Mobile Weight affected in mobile Price??

**Supervised Machine learning algorithms and implementation :**

1. Logistic regression

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

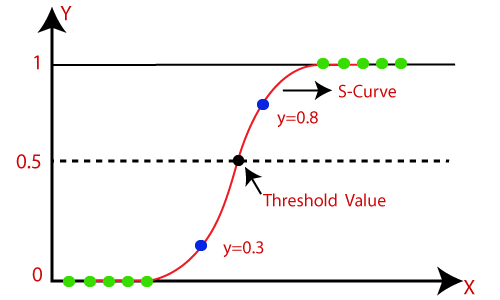
Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, True or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

Logistic Regression is much similar to Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.

In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic

Logistic Regression is a significant machine learning algorithm because it has the

Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification. The below image is showing the logistic function:



1. Decision tree

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

The decisions or the test are performed on the basis of features of the given dataset.

*It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions* .It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

A decision tree simply asks a question, and based on the answer (Yes/No), it further splits the tree into subtrees.

Below diagram explains the general structure of a decision tree:



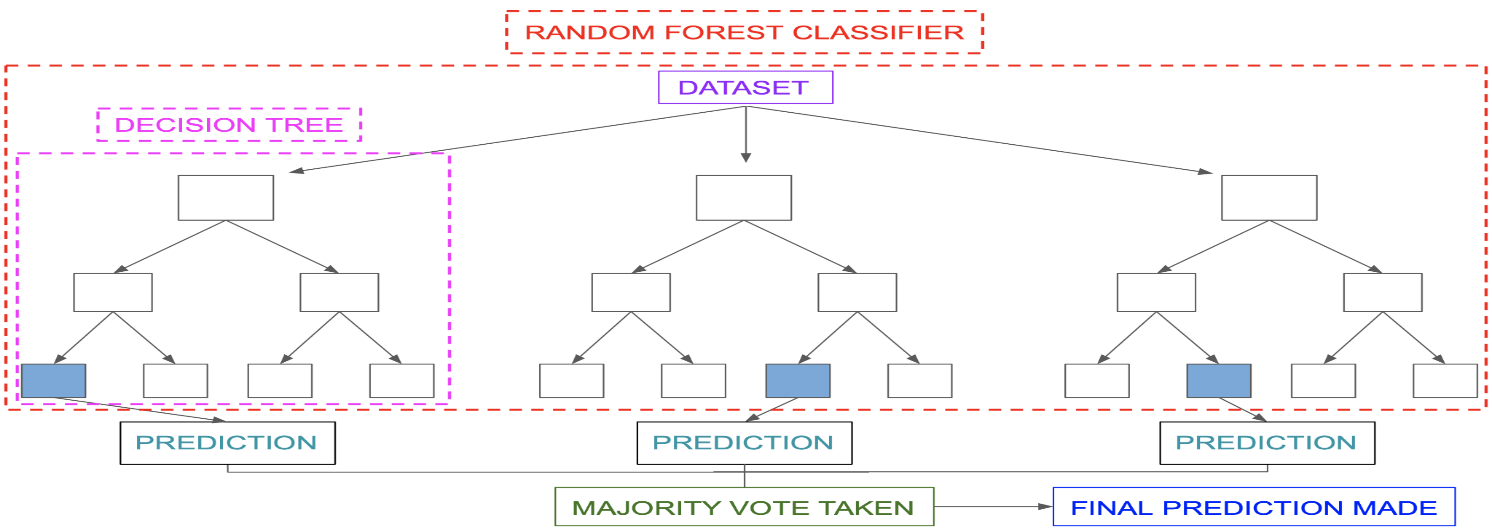
1. Random forest classifier

A random forest is a supervised machine learning method built from decision tree techniques. This algorithm is used to anticipate behavior and results in a variety of sectors, including banking and e-commerce.

A random forest is a machine learning approach for solving regression and classification issues. It makes use of ensemble learning, which is a technique that combines multiple classifiers to solve complicated problems.

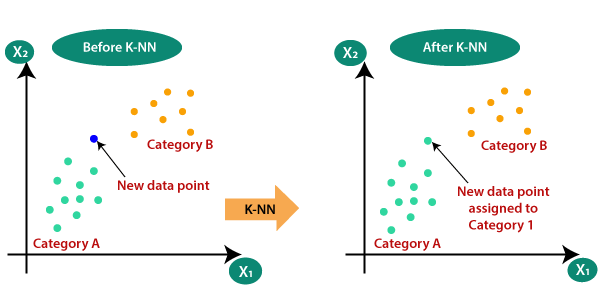
A random forest method is made up of a large number of decision trees. The random forest algorithm’s ‘forest’ is trained via bagging or bootstrap aggregation. Bagging is a meta-algorithm ensemble that increases the accuracy of machine learning algorithms.

The outcome is determined by the (random forest) algorithm based on the predictions of the decision trees. It forecasts by averaging or averaging the output of several trees. The precision of the outcome improves as the number of trees grows.



1. K-Nearest Neighbour

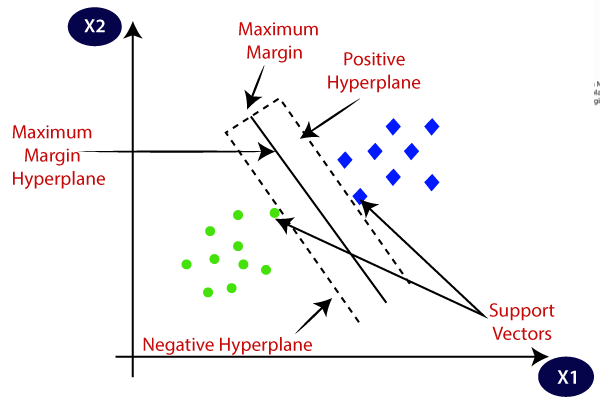
K-nearest neighbors (KNN) is a type of supervised learning algorithm used for both regression and classification. KNN tries to predict the correct class for the test data by calculating the distance between the test data and all the training points. Then select the K number of points which is closet to the test data. The KNN algorithm calculates the probability of the test data belonging to the classes of ‘K’ training data and class holds the highest probability will be selected. In the case of regression, the value is the mean of the ‘K’ selected training points.



1. SVM

Support Vector Machine, or SVM, is a prominent Supervised Learning technique that is used for both classification and regression issues. However, it is mostly utilized in Machine Learning for Classification purposes.

The SVM algorithm’s purpose is to find the optimum line or decision boundary for categorizing n-dimensional space so that we may simply place fresh data points in the proper category in the future. A hyperplane is the optimal choice boundary.



## **Conclusion:**

1. From EDA we can see that there are mobile phones in 4 price ranges. The number of elements is almost similar.

2. half the devices have Bluetooth, and half don’t.

3. 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.

4. costly phones tends to be lighter.

5. RAM, battery power, pixels played more significant role in deciding the

price range of mobile phones.

1. form all the above experiments we can conclude that logistic regression gives the best results

**References**

1.Stack\_overflow

2.GeekforGeeks

3.Scikit-learn

**GitHub Link:**

[**https://github.com/SubhasisChattopadhyay/Almabetter-Capstone-Projects/blob/main/Mobile\_Price\_Range\_Prediction\_Capstone\_Project.ipynb**](https://github.com/SubhasisChattopadhyay/Almabetter-Capstone-Projects/blob/main/Mobile_Price_Range_Prediction_Capstone_Project.ipynb)