**Applying Machine Learning to predict Mobile Cost**

**Aim**

In this Project, on the basis of the mobile Specification like Battery power, 3G enabled, Wi-Fi, Bluetooth, Ram etc., we are predicting the mobile cost

The objective of this project is to learn and apply Machine learning model on the dataset and analyse the behaviour of it.

This kind of prediction will help companies estimate price of mobiles to give tough competition to other mobile manufacturer

Also it will be useful for consumers to verify that they are paying best price for a mobile.

**Problem Statement**

A mobile phone, cell phone, cell phone, or hand phone, sometimes shortened to simply mobile, cell or just phone, is a portable telephone that can make and receive calls over a radio frequency link while the user is moving within a telephone service area.

The first handheld mobile phone was demonstrated by John F. Mitchell and Martin Cooper of Motorola in 1973, using a handset weighing c. 2 kilograms (4.4 lbs).

In 1979, Nippon Telegraph and Telephone (NTT) launched the world's first cellular network in Japan. In 1983, the DynaTAC 8000x was the first commercially available handheld mobile phone. From 1983 to 2014, worldwide mobile phone subscriptions grew to over seven billion—enough to provide one for every person on Earth.

In first quarter of 2016, the top smartphone developers worldwide were Samsung, Apple, and Huawei, and smartphone sales represented 78 percent of total mobile phone sales. For feature phones (or "dumbphones") as of 2016, the largest were Samsung, Nokia, and Alcatel.

Source: <https://en.wikipedia.org/wiki/Mobile_phone>

We all may have faced the problem of identifying the important features from a set of given data and removing the irrelevant or less important features which do not contribute much to our decision making in order to achieve better accuracy for our model.

In machine learning and statistics, feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of reducing the number of input variables when developing a predictive model. Feature selection techniques are used for several reasons:

* It reduces model complexity by dropping some irrelevant features.
* Helps ML algorithm to train a model faster.
* Reduction of dimensionality helps in avoid overfitting.

**Learning Outcome**

* Extracting the data
* Data Manipulation and Cleaning
* Creating charts and bars
* Perform wrangling operations to draw more insights
* Implementing ML model that give you a good results based on problem.
* Comparing models to find the best one
* Performing hyper tuning technique to optimise the model

**Data Attributes**

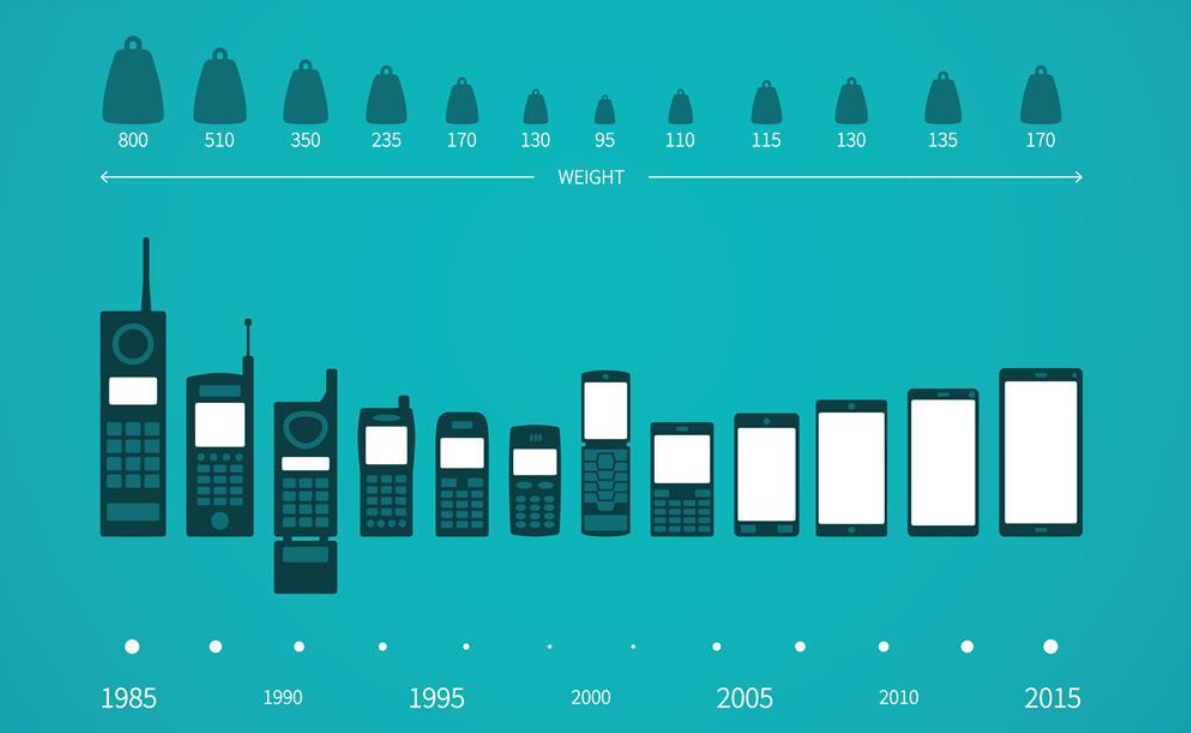
|  |  |
| --- | --- |
| **Column Name** | **Description** |
| battery\_power | Total energy a battery can store in one time measured in mAh |
| blue | Has bluetooth or not |
| clock\_speed | speed at which microprocessor executes instructions |
| dual\_sim | Has dual sim support or not |
| fc | Front Camera mega pixels |
| four\_g | Has 4G or not |
| int\_memory | Internal Memory 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 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 | price range of the mobile |

**Target**

Need to perform ML model to find the best cost range of the mobile.

**Skill Requirement**

* numpy
* pandas
* matplotlib
* seaborn
* datetime
* sklearn



# **Phase 1**

First and foremost task that every person shall do towards starting career in Data Science is to extract the relevant data and understand its characteristics.

This is what we expect you to perform by going through following tasks.

You can perform your own analysis also that can be noteworthy. Make a note of it somewhere and share it back with the team.

From this phase, you will learn to extract the datasets and perform necessary cleaning actions.

## Task 1

1. Import the relevant packages
2. Import the datasets
3. Walkthrough the characteristics of the dataset
4. Check out if there is any gap

## Task 2

1. Perform the data cleaning actions
2. Rename the attributes with below values for further operations;

|  |  |
| --- | --- |
| **Column Name** | **New Name** |
| battery\_power | Battery Power |
| blue | Bluetooth |
| clock\_speed | Clock Speed |
| dual\_sim | Dual Sim |
| fc | Front Camera MP |
| four\_g | 4G |
| int\_memory | Internal Memory(GB) |
| m\_dep | Mobile Depth |
| mobile\_wt | Weight |
| n\_cores | Number of cores |
| pc | Primary Camera MP |
| px\_height | Pixel R. Height |
| px\_width | Pixel R. Width |
| ram | RAM(MB) |
| sc\_h | Screen Height |
| sc\_w | Screen Width |
| talk\_time | Longest Battery Charge |
| three\_g | 3G |
| touch\_screen | Touch Screen |
| wifi | WIFI |
| price\_range | Price Range |

# **Phase 2**

You have successfully performed data cleaning operations and now we are ready to perform in detail analysis.

This is the most time consuming task that each data enthusiast shall perform and invest as much time as possible.

Perform all types of analysis and solve all the questions which when asked.

Expected Dataset Size: 3000 records, 21 attributes

Task

1. Univariate analysis of each variable
2. Bivariate Analysis of categorical vs numerical variables
3. Multivariate Analysis of categorical and numerical variables
4. Check distribution of variables
5. Create the charts

# **Phase 3**

As you have completed analysis of cost with respect to all the parameters, you are now ready to apply ML models.

Make a note of which attributes are important to identify the price range.

Task

1. Keeping the random state as 51, evaluate the models on the basis of Accuracy Score.
2. Import all the relevant models from the machine learning package.   
   (Compare Models: Logistic, KNN, SVC, Decision tree, Random Forest)
3. Create a chart to compare the accuracy score of all the models and share the best model.

# **Phase 4**

This is an incredible job done. You have identified the best model on the basis of accuracy.

Machine learning engineers also dive to get extra performance by changing the model parameters. It can enhance your accuracy and that can also be taken as more proof to support your hypothesis.

This enhanced performance not only supports your hypothesis under ideal conditions about the target problem but also support your decision about the parameters you have taken while making the predictions

Task

1. Apply Grid search CV method on the best model.
2. You have to cover the hyper tuning method on the model parameters