**Capstone Project- Submission**

**Mobile Price Range Prediction** BY

#### MD MOINUDDIN

GitHub Link: -

https://github.com/Moinuddin-developer/Netflix-Movies-and-TV-Shows-Clusterin

###### Abstract:

The key purpose of this research work is to determine "If the mobile with given features would be under a certain price range." Specific feature selection algorithms are used to recognize and delete features that are less necessary and redundant and have minimal complexity in computation. Different classifiers are used to achieve the best possible accuracy. Results are measured in terms of achieving the maximum accuracy and choosing the minimum features. Statement is made based on the algorithm for best selection of features and best classifier for the given dataset. This work can be used to find the optimal product (with minimum cost and maximum features) in any form of marketing and industry. It is suggested that future work will extend this research and find a more sophisticated solution to the given problem and a more accurate tool for estimating prices.

###### Introduction:

Price is the most important component in the marketing of any product and is often the definitive factor in its sale to a consumer. In a constantly evolving and volatile market, the price is often the factor that makes or breaks a product. Setting an optimal price before the release of a product is imperative for any company. A tool that gives the estimated price of a product after weighing in the features it provides can come in handy and can help the company in making an informed decision while setting the market price for a product. Such a tool can also be used by a consumer to get an estimated price based on the features they are looking for in the product. Machine learning algorithms can perform various tasks which are to be chosen with respect to the data to be worked with and the motive of the task. Various tools and languages like Python and its libraries are available to perform machine learning tasks. Examples of some frequently used algorithms include Logistic Regression, Random Forest, Decision Tree, xgBoost etc. Feature selection algorithms can be used to select and extract only the best parameters to train a model to optimize the accuracy and lessen the computational time of the model. Any of these methods can be used to perform the task of predicting the price of a product depending on the type of data available to train the model. Nowadays, a cellphone is an essential accessory for a person. It is the fastest evolving and moving product in the technology market space. New mobiles with updated versions and new features are introduced into the market at a rapid pace. Thousands of mobiles are sold each day. In such a fast-paced and volatile market, a mobile company needs to set optimal prices to compete with its rivals. The first step in fixing a price is to estimate the price based on the features. The objective of this research is to develop an ML model capable of estimating the price of a mobile phone based on its features. A potential buyer can also make use of the model to estimate the price of a mobile by inputting just the features they require into the tool. The same approach to create a prediction model can be used to develop a price estimation model for most products that have similar independent variable parameters. The price of a mobile is dependent on many features, for example, the processor, battery capacity, camera quality, display size and thickness, etc. These features can be used to classify phones into various categories like entry-level, mid-range, flagship, premium, etc. Supervised ML algorithms are used in this paper as the dataset used has a definitive class label for price range.

###### Data 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 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).

###### Problem Statement:

As the demand for smartphones increases, consumers face the challenge of selecting the best smartphone that meets their requirements and budget. With the constantly changing market and new technologies being introduced, it becomes difficult for consumers to keep up with the price ranges of smartphones. This is where Mobile Price Range Prediction comes into play. The problem is to predict the price range of a smartphone based on its features, specifications, and brand. With accurate predictions, consumers can make informed decisions and purchase the best smartphone within their budget.

###### Objective

The key purpose of this research work is to determine "If the mobile with given features would be under a certain price range." Specific feature selection algorithms are used to recognize and delete features that are less necessary and redundant, and have minimal complexity in computation. Different classifiers are used to achieve the best possible accuracy. Results are measured in terms of achieving the maximum accuracy and choosing the minimum features. Statement is made based on the algorithm for best selection of features and best classifier for the given dataset. This work can be used to find the optimal product (with minimum cost and maximum features) in any form of marketing and industry. It is suggested that future work will extend this research and find a more sophisticated solution to the given problem and a more accurate tool for estimating prices.

###### Approach:

Mobile price range prediction is a classic machine learning problem that involves predicting the price range of a mobile phone based on its features. The following is an approach which I used for mobile price range prediction:

1. Data Collection: The first step is to collect data on various mobile phones, their features, and their prices. The data can be collected from various sources such as online marketplaces, manufacturer websites, and reviews.
2. Data Cleaning: The collected data may contain missing values, outliers, and inconsistencies. It is essential to clean the data by removing the missing values, outliers, and inconsistencies.
3. Feature Engineering: Feature engineering is the process of selecting the most relevant features from the dataset. In the case of mobile price range prediction, the features that are usually considered include the mobile's brand, RAM, internal storage, screen size, camera quality, and battery capacity.
4. Data Preprocessing: The data needs to be preprocessed to normalize the feature values, scale them, and convert categorical variables into numerical variables.
5. Model Selection: Various machine learning algorithms can be used for mobile price range prediction. The commonly used algorithms include Decision Trees, Random Forest, Support Vector Machines (SVM), and Neural Networks.
6. Model Training: Once the model is selected, the next step is to train the model using the preprocessed data. The training process involves dividing the data into training and testing sets, fitting the model on the training set, and evaluating the model's performance on the testing set.
7. Model Evaluation: The performance of the model can be evaluated using metrics such as accuracy, precision, recall, and F1-score.
8. Model Deployment: Once the model is trained and evaluated, it can be deployed in the production environment. The model can be integrated with a mobile application or a web-based platform to predict the mobile's price range based on its features.

Overall, mobile price range prediction is a challenging problem that requires a combination of data science and machine learning techniques. With the right approach and techniques, it is possible to build an accurate and robust model that can predict the mobile's price range based on its features.

**Tools Used**

The whole project was done using python, in google Collaboratory. The following libraries were used for analyzing the data and visualizing it and to build the model .

* Pandas: Extensively used to load and wrangle with the dataset.
* Matplotlib: Used for visualization.
* Seaborn: Used for visualization.
* Datetime: Used for analyzing the date variable.
* Warnings: For filtering and ignoring the warnings.
* NumPy: For some math operations in predictions.
* Sklearn: For analysis and prediction.

Steps Involved

The following steps are involved in the project

1. **Data Collection:** The first step is to collect data related to mobile phone models and their prices. This data collected from Alma better.
2. **Data Cleaning:** Once the data is collected, the next step is to clean and preprocess it. This involves removing any irrelevant or duplicate data and formatting the data in a consistent manner.
3. **Feature Selection:** In this step, the most important features that influence the price of a mobile phone are identified. These features can include the brand, model, specifications, availability, and popularity.
4. **Data Splitting:** The collected and cleaned data is split into training and testing datasets. The training dataset is used to train the machine learning model, while the testing dataset is used to evaluate its performance.
5. **Model Selection:** A suitable machine learning algorithm is selected based on the problem statement and the available data. Regression algorithms such as linear regression, decision trees, and random forest can be used to predict the price range of mobile phones.
6. **Model Training:** The selected machine learning algorithm is trained using the training dataset. The algorithm learns the patterns and relationships between the features and the price of the mobile phone.
7. **Model Evaluation:** The performance of the trained machine learning model is evaluated using the testing dataset. The model’s accuracy, precision, recall, and other metrics are calculated to ensure that it is performing well.
8. **Model Deployment:** Once the machine learning model is trained and evaluated, it can be deployed to predict the price range of new mobile phone models. This can be done using a web application or a mobile app, which takes input from the user and returns the predicted price range of the mobile phone.

Conclusion

* From EDA we can see that here are mobile phones in four 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.
* form all the above experiments we can conclude that logistic regression and, XGboosting with using hyperparameters we got the best results.

**REFERENCES:**

<https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c#:~:text=Silhouette%20Coefficient%20or%20silhouette%20score%20is%20a%20metric%20used%20to,each%20other%20and%20clearly%20distinguished.&text=a%3D%20average%20intra%2Dcluster%20distance,each%20point%20within%20a%20cluster>.

<https://machinelearningmastery.com/clustering-algorithms-with-python>

<https://towardsdatascience.com/introduction-to-machine-learning-algorith>