**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Team Member’s Name: -** Mrunal Badgujar  **Email: - mrunalbadgujar18@gmail.com**  **Contribution: -**   1. Imported all the libraries for data exploration, Sorting, Cleaning and Visualization. 2. Imported and mounted data set required for analysis from google drive to google co lab. 3. Exploring data sets like number of columns and rows with heading and what is data type of each value using python libraries methods like Info (), shape, describe (), Head (), Tail (). 4. Checked Null Value and enriched/dropped/replaced with appropriate value using basic business knowledge and requirement with help of mathematics rules. 5. With all the group members brainstorming finalizes some observation and their visualization with interpretation. 6. With the group help prepare a presentation. 7. Gradient Boosting ML algorithm 8. Conclusion   **Team member name:** Chetan Patil  **Email:** chetanpatil4160@gmail.com  **Team contributions**   1. Find the question 2. Which feature impacts most on price? 3. Which type of phones have max price? 4. How many phones have Bluetooth? 5. Did some work on data visualization 6. Plot Regression charts 7. Did some work on data cleaning and changed the column names? 8. Find solutions for questions 9. Worked on conclusion 10. Logistic regression model 11. Decision Tree ML model 12. Plot Distribution of data   **Team member name:** Rajesh Patil  **Email:** rajesh.patil775607@gmail.com  **Team contribution:**   1. Question forming 2. Which phone has a maximum price in 3g or 4g? 3. Is ram effect on price? 4. Is data balanced? 5. data cleaning (outliers’ removal) 6. data visualization by using box plot 7. technical document preparation 8. Outlier detection and removal 9. Helped to make Presentation ppt 10. Decision Tree 11. Random Forest 12. conclusion   **Team member name:** Sachin Chaudhari  **Email: sachinchaudhari4141@gmail.com**  **Team contributions**   1. Correlation between data 2. With the group help prepare presentation 3. Worked on Presentation preparation and its interpretation. 4. Feature selection 5. K nearest neighbor model 6. SVM model |
| **Please paste the GitHub Repo link.** |
| GitHub Link: - https://github.com/chetanpatil4160/Machine-Learning-Mobile-Price-Range-Predicition |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| Mobile phones have 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 specifications and new designs. So, the real question is prediction is 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. Customers usually see that they are able to buy with the specification with the given estimated price or not. So, 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.  There is a need for data pre-processing because the data may be incomplete or inconsistent or noisy. There are many ways to deal with un-processed data. Firstly, we cleaned the data. By cleaning, we mean filling in the missing values in the data, identifying and removing outliers in the data, and smoothening the filling. After cleaning the data, we moved to normalization of the data. Furthermore, some features are reduced for simplicity's sake. After reduction, it's time to get insight from the dataset, so EDA is the best process for getting data insight. We came to know that our data is well balanced and RAM has the most important features among all. RAM has a positive correlation with mobile prices.  We applied some classic machine learning algorithms like Decision Tree, Random Forest classifier, Gradient Boosting Classifier-nearest Neighbor classifier Boost Classifier, Support Vector Machine (SVM), and Logistic regression. Logistic regression is basically a supervised classification algorithm. In a classification problem, the target variable (or output), y, can take only discrete values for a given set of features.  Contrary to popular belief, logistic regression IS a regression model. The model builds a regression model to predict the probability that a given data entry belongs to the category numbered as "1". Just like linear regression assumes that the data follows a linear function, logistic regression models the data using the sigmoid function. Logistic regression becomes a classification technique only when a decision threshold is brought into the picture. The threshold value is a critical aspect of logistic regression and is determined by the classification problem itself. Logic regression gave 95% accuracy. On test data, Random Forest achieved an accuracy of around 88%. The Gradient Booster had an accuracy of 91%. KNN gave the lowest accuracy of all, which is around 70%. SVM is the best fit algorithm because it has the highest accuracy, which is 97% accuracy.  The conclusion is that the SVM algorithm has the highest accuracy, which is 97%. SVM performed very well as compared to other algorithms. In terms of feature importance, RAM, battery power, px\_height, and px\_weight are the important features. The F1 scores for individual classes are also very good. The area under the curve for each class prediction is also almost 1. |