**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| * **Mounika Dontula (**[**mounikadontula2795@gmail.com**](mailto:mounikadontula2795@gmail.com)**)**   **• Data Cleaning**   * + - * Nearest Neighbors Imputation(KNNImputation)       * Handling null and duplicate values       * Dropping some columns   **• EDA-Exploratory Data Analysis**   * + - * Having a look at the target variable.       * Correlation of dependent variable with other independent variable.       * Univariate analysis of categorical columns.       * Relation between RAM,battery\_power,px\_height and px\_width.       * Checking which numerical feature is driving the price range the most.       * Mobile with both 3g and 4g.       * Mobile with no 3g and 4g.       * Checking the distribution of numerical columns and outliers.       * Outlier Treatment.       * Feature Selection.   **• One Hot Encoding**   * + - * The main aim of One Hot encoding is to produce binary integers of 0 and 1 to encode our categorical features.   **• Predictive Models**   * + - * A Classification model reads some input and generates an output that classifies the input into some category.   **• Machine Learning Algorithms(Classification)**   * + - * Decision Tree Classifier.       * Random Forest Classifier.       * Gradient Boosting Classifier.       * K-Nearest Neighbour Classifier.       * XGBoost ClassifierSupport Vector Machine. |
|  |
| **Please paste the GitHub Repo link.** |
| Github Link:-  https://github.com/MOUNIKADONTULA/Mobile-Price-Range-Prediction-Classification.git |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| First I decided to take up this project solely due to mutual interest inMobile Price Range Prediction. Then I decided to gain insights doing some EDA etc. I started to form the questions. After doing some random EDAs. I gained some confidence.  **Short summary:**   * In this project, We worked on a mobile price range prediction problem wherein we had to classify the price range with different classification algorithms to check the performance of the data. * The dataset contained about 2000 records, and 21 attributes. * This data set contains no null values, duplicate values and missing values and does exploratory data analysis(EDA). * price\_range is our target/dependent variable and there is no imbalance in the target variable. * 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 the Low cost category. * In Decision Tree Classifier the Train accuracy has been reduced to 97% from 100% and Test accuracy is increased by 2%. Thus, we somewhat reduced the overfitting by reducing the training accuracy. However this will not be a good model for use. * In Random Forest and Gradient Boosting Classifier the Train accuracy is same but the Test accuracy is decreased slightly. * In xGBoost classifier theTrain accuracy is decreased from 100 to 98 but the Test accuracy remains same,hence this model is slightly overfitted. * In Support Vector Machine the Train accuracy is same but the Test accuracy is increased from 90 to 97 percent.SVM performed very well as compared to other algorithms. * From the above algorithms we can conclude that Support Vector Machine with using hyperparameters we got the best results. |