**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Name**: Kabeer Pande  **Email**: [kabeerpande7075@gmail.com](mailto:kabeerpande7075@gmail.com)  **Contribution:**  Colab notebook  Project summary  Technical documentation  Project presentation  Presentation video |
| **Please paste the GitHub Repo link.** |
| Github Link:- **https://github.com/kabeerrrh/Mobile-price-range-prediction.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)** |
| Businesses in the very competitive mobile phone market are interested in knowing what factors affect prices and how many phones are sold. Finding a correlation between a mobile phone's features and price is the objective of our project. The pricing range should be displayed to show how high the cost is.  We started by importing the libraries numpy and Pandas that will aid in our exploratory data analysis. There were 2000 rows and 21 columns in the dataset. As the dataset does not contain any duplicate rows or null values, we moved on to exploring the data with the help of the visualisation libraries seaborn and matplotlib. There, we discovered several fascinating facts about the data, such as the proportion of mobile phones with Bluetooth, dual SIM, and Wi-Fi, among many other things.  Then we used a heat map to build a correlation map on the dataset, and we discovered that there were no correlated data inputs in our dataset, indicating that there was no multicollinearity concern. Next, we executed a train-test split for developing machine learning models before moving on to data modeling.  We implemented Logistic Regression, Decision Tree, Random Forest, XGBoost  and K-Nearest Neighbor.  The accuracy score of :   Logistic Regression : 76%   Decision Tree : 82%   Random Forest : 88%   XGBoost : 89%   K-Nearest-Neighbor: 93%  So we conclude that K-Nearest-Neighbor is our best fit model for our Dataset. |