Capstone Project Submission

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Name: Jaya Vishwakarma **Email Id:** jayavishwa2@gmail.com **Contribution:** ➢ Data Wrangling ➢ Multicollinearity ➢ KNN ➢ XG Boost ➢ Random forest **Name: Kavya Sharma** **Email Id**: harrycr713@gmail.com **Contribution:** ➢ Data Wrangling ➢ Feature scaling ➢ Train and test data ➢ Decision tree ➢ Logistic regression **Name: Priyvrat Sharma** **Email Id**: priyvratsharma97@gmail.com **Contribution:** ➢ Data Wrangling ➢ Preprocessing of data ➢ Checking correlation ➢ Creating numerical and categorical columns ➢ Naïve Bayes **Name**: **Richa Pandya** **Email Id:** richapandya471@gmail.com **Contribution:** ➢ Data Wrangling ➢ Checking Outlier ➢ Univariate analysis ➢ Cross validation ➢ Append matrix
GitHub Repo link.

https://github.com/JayaVishwa/Mobile-Price-Range-Prediction

The most successful marketing and commercial attribute is price. It is the primary variable affecting how well that product sells. Mobile technology is a technology that follows its users wherever they go. One of the most major technical revolutions in human history had started by the mobile phone. In this project, we have tried to find out the best machine leaning model which can predict the price range of mobile phone with respect to the features, which are screen height and width, ram, internal memory, battery etc. given in our data set.

Data preparation & Visualization:

- We started with importing all the important libraries to help us explore the problem statement, perform EDA to draw conclusion on the basis of the data set, implemented classification algorithms and evaluation metrics.
- Our next step was to perform data wrangling over the raw data. We wrote some codes to find out the duplicate and null values. And we found that our data is clean. There are no null and duplicate values.
- After all the processing of the data, we get 2000 rows and 21 columns.
- Preprocessing data.
- Creating numerical and categorical columns.
- Checking Outlier and correlation.
- Univariate analysis and deal with Multicollinearity
- EDA

Model Building:

We have applied supervised machine learning algorithms like Decision Tree, Random Forest, Naïve Bayes, KNN, Logistic Regression and XG Boost. Accuracy score for these models have been mentioned below.

Model Performance (Accuracy Score):

Model	**GNB**	**KNN**	**DT**	**RM**	**LR**	**XG**
Training set1	0.82	0.74	0.92	0.92	0.98	1.00
Test set1	0.82	0.60	0.84	0.84	0.96	0.93

Most of the models are not able to get good accuracy for each class of target variable.
With hyper parameter tuning, even after assigning different parameters values XG boost performed not so good on test data but It works really well on training set. It was seen that ram has the highest impact on the price of the mobile.

Surprisingly, Logistic regression performed well in this classification problem. It was also found that there's some over fitting in case of XG Boost. Random Forest and Decision Tree models were tested which performed well as compared to the other 2 models like KNN and Naive Bayes.

Hence, it was concluded that the **Logistic model was the best among all**.