Capstone Project Submission

Team Member's Name, Email and Contribution:

1) Kajal Dhun

Email ID: kajaldhun@gmail.com

Contribution:

a) Data Wrangling - handling mismatched columns.

b) EDA = Distribution of n_cores(no of core processor) and pc(primary camera megapixel).

c) EDA = Percentage distribution of mobile with respect to 'three_g','four_g', 'blue', 'touch_screen', 'dual_sim'.

d) EDA = Distribution of mobile features(ram, battery power, px_height, px_width) wrt price_range.

e) EDA = Count of mobiles contain 3G and 4G features (Having 3G and 4G data).

h) Data Preprocessing and Training model with 5 Classification algorithms.

f) Technical Documentation.

g) PPT making.

h) Technical Documentation's conclusion.

2) Navinkumar Sambari

Email ID: nnnghvs143@gmail.com

Contribution:

a) Data Wrangling - Outlier treatment

b) EDA = Price_Range VS ram, battery power, px_height, px_width.

c) EDA = Count of mobiles in each price range with supporting or not supporting specifications.

d) Correlation Matrix analysis

e) Data Preprocessing and Training model with 5 Classification algorithms

f) PPT making.

g) Coding Conclusion and Technical Documentation's conclusion.

4) Tanu Rajput

Email ID: tanuurajput689@gmail.com

Contribution:

a) Data Wrangling - handling mismatched columns

b) EDA = Count of mobiles for each n_cores and m_dep wrt price _range.

c) EDA = Count of mobiles does not contain 3G and 4G features (Not having 3G and 4G data).

d) EDA = How price effects based on number of cores.

e) EDA = Line plot which shows the trend of the major features which are considered wrt to Price range.

f) Data Transformation and Training model with 5 Classification algorithms.

f) PPT making.

h) Technical Documentation's Conclusion

❖ **We set meetings when we started working on implementing Classification algorithms and contributed equally and also equally contributed for Project Summary.**

Please paste the GitHub Repo link.
GitHub Link: - https://github.com/TanuRajput110/Mobile-price-range-prediction

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

This Mobile price range prediction dataset contains 21 features and 2001 observations including column headers. This project is of Classification analysis because our target feature is the price_range which has categorical labels 0-3. This Classification analysis helps to achieve our goal of predicting mobile price range by considering the other features of the dataset. The crucial part is the prediction of mobile price range for the market success of a product.

After loading the dataset, firstly, we performed data preprocessing, we did some data exploration by checking types, missing values, duplicate values and data description. In this dataset there are neither null values nor duplicate values. But there are some columns which have mismatched values which were founded by statistical information and also there are some outliers in some features. So, we did some data wrangling and handled them. Outliers were handled by using the Quantile method.

After that, we performed Exploratory Data Analysis to obtain the insights of our dependent variable "Price_range". Various graphs were constructed comparing the "Price_range" column with other columns. List of insights were obtained. We observed that the price range is mostly affected by RAM, 4G features, battery_power, primary camera and front camera megapixels, pixel resolution height and pixel resolution width. The higher the RAM, higher the PRICE.

Next step was feature engineering in which we used Standard Scalar for transforming data and also used category_encoders for encoding the labels. Then we splitted our data into train data and test data.

Now the modeling part begins, here we used 5 Classification algorithms, viz., Logistic Regression, K-Nearest Neighbor, Decision Tree Classifier, Random Forest Classifier, SuperVectorMachine. So, after fitting the models and evaluating metrics (such as precision score, recall score, and F1 score). We also did hyperparameter tuning and we came to the result and conclusion. We got the Test Recall Score among all the models, SupeVectorMachine and Logistic Regression gave the best scores . But among them we decided to choose SuperVectorMachine as our predictive model having recall score = 0.957. Therefore, this model is the best for predicting the mobile price range.