Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Ranjith K (ranjith26.ganesh@gmail.com): • **Exploratory data Analysis** • **Outlier detection** • **Outlier treatment** • **Handling null values** • **Label encoding** • **Correlation Heatmap** • **Train test split on dataset** • **Machine learning model implementation** • **Hyper parameter tuning** • **Regression evaluation** • **Model Explainability** • **Summary, technical documentation**
Please paste the GitHub Repo link.
GitHub Link: - https://github.com/RanjithK2608/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)

During the purchase of mobile phones, various features like memory, display, battery, camera, etc., are considered. People fail to make correct decisions, due to the non-availability of necessary resources to cross-validate the price. To address this issue, a machine learning model is developed using the data related to the key features of the mobile phone. The developed model is then used to predict the price range of the new mobile phone.

Three machine learning algorithms namely Support Vector Machine (SVM), Random Forest Classifier (RFC), Logistic Regression are used to train the model and predict the output

We imported the dataset and read the csv file, null values were taken care of at first, Distribution was created using histogram for numerical features, later outliers were found out with the help of box plot and treated by log transformation.

Exploratory data analysis was done where few insights were obtained from the data using visualizations. Correlation heat map was generated to understand the correlation among the variables and removed the features which has high correlation.

The dataset was then split into train and test data and various regression machine learning algorithms like Logistic Regression, K Nearest Neighbors Classifier and XGBoost Classifier were applied on the data followed by hyper parameter tuning using Randomized Search CV on XGB Classifier model to obtain the best possible results.

In K nearest Neighbors classification model, we have got the knn sc ore as 56.25%, accuracy score of 67% for training set and 65% for test set.
During 'elbow method' we have got the insight that the optimum value of k is 22 with least error rate.

In Logistic Regression Model, we have got the log score as 91%. accuracy score of 98% for training set and 91% for test set.

In XGBoost model the score was 89% before hyper parameter tuning.

RandomizedSearchCV is used for hyperparameter tuning in XGBoost classifier and the accuracy obtained after hyper parameter tuning was 86% for training set and 80% for test set. Finally, in the model explaininabilty we have used shap and we got the insight that 'ram', 'battery power', and phone dimensions are the features which is deciding as key factor for the price range pre diction.