# Capstone Project Submission

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member's Name, Email and Contribution:** |
|---|
| Rahul Shah: shahrahool010@gmail.com<br>1. Exploratory data analysis – univariate and multivariate analysis.<br>2. Data Wrangling – checking missing values, outliers, features modification.<br>3. Fitting Models – splitting the data, applying algorithms, hyperparameter tuning , model evaluating, model explanation.<br>4. Presentation, Technical documentation. |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/rahool010/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)** |
| Nowadays, Mobile phones have become a part and parcel of everyone's life. People want more features and best specifications in a phone and that too at affordable prices. In the competitive mobile phone market companies want to understand sales data of mobile phones and factors which drive the prices. The data provided had features such as battery power, ram, internal memory, mobile weight, camera resolutions, pixel dimensions, screen dimensions, 3g, 4g, touchscreen, dual sim, talktime, wifi, bluetooth and target variable price range.<br><br>The problem statement was to build a machine learning model that could predict the price range values, given other variables.<br><br>The first step in this exercise involved exploratory data analysis, where we tried to identify patterns, trends, correlation, relationships of each variable with our dependent variable. We tried to figure out impact of each variable in determining the price range values.<br><br>The second step involved data wrangling in which we tried to check data integrity, missing values, outliers and performed feature modifications.<br><br>Next step was to implement machine learning algorithms on our splitted and standardized data and evaluate the performance using several evaluation metrics. Three algorithms were used namely; Random Forest classifier, XGBoost classifier and Support Vector Classifier. We also did hyperparameter tuning to improve model performance and evaluated the performance of each model using several evaluation metrics. The best performance was given by the Support Vector Classifier.<br><br>Next, we implemented shap techniques to understand the working of the model. We saw ram, battery power and pixels were the top contributors in determining price ranges. Higher the values of these, led to higher price range.<br><br>The best model was SVM. The accuracy of our best model was 0.95 and 0.9275 for training and test set respectively. |