# CAPSTONE PROJECT SUBMISSION

## INSTRUCTIONS

I. Please fill in all the required information.

II. Avoid grammatical errors.

## TEAM MEMBER'S NAME, EMAIL AND CONTRIBUTION:

## TEAM - DATA DEFENDERS

S No Member Name Email Contribution 1. Lokesh Tokas lokesh.you@gmail.com Colab Notebook

2. Saraswat Mukherjee mae21saraswat@gmail.com Technical documentation 3. Charan C S ccharancs543@gmail.com Summary 4. Shubham Sartape shubhamns19.pumba@gmail.com Presentation

## GITHUB REPO LINK.

GitHub link: https://github.com/Donein/ML-Classification-mobile-price-prediction-

# PROJECT SUMMARY

In this project we have implemented Mobile Price Range predictions on the basis of different mobile specifications like RAM, WIFI, Bluetooth, Battery, etc, by using different Machine Learning models. So we use many of the features to classify whether the mobile is very economical, economical, expensive or very expensive.

Different classification models are used to obtain high accuracy and conclusion is made on the basis of best Machine Learning algorithm and classifier.

The given data set has 2000 entries and 21 columns which contains one target variable 'price-range' which is classified into 4 types from least expensive to most expensive.
Data pre-processing and Data cleaning is done to known the feature importance and finding out missing values.

In EDA, analysis is done on variable columns if outliers are present by visualizing with boxplot and there are no extreme outliers present. Then feature engineering is done on different variable column and price-range column to known the frequency distribution by using bar plots. Checking multicollinearity by plotting heat map to know the strength of correlation and observing the feature importance of each variable columns. It is known RAM and battery power has the highest impact on price range.
Size and thickness of the mobile are also important factors to consider.

We compared 4 different classification models like Linear Regression, Random Forest Classifier, KNN, SVM (Support Vector Machine) that performs well and accurate predictions on new data is made. Hence splitting the data into training set and testing test to avoid overfitting and to estimate the performance of the ML model on new data.

For each classification model, accuracy of both train and test data is obtained then confusion matrix and classification report is evaluated. The AUC-ROC score gives the best predictions, then optimizing the model that has been trained by applying cross validation and hyperparameter tuning for the best performance.

Both Logistic Regression and SVM (Support Vector Machine) algorithm almost gave best accuracy after hyper-parameter tuning with 91.6% train accuracy and 89.2% test accuracy. Random Forest was overfitting. Hence to conclude the best ML model for any marketing and businessrequirement, optimal product is obtained by using Logistic

Regression Classification model.