**CLASSIFYING THE PRICE OF MOBILE PHONES BY APPLYING ML MODELS WITH A MULTINOMIAL APPROACH USING PYTHON LANGUAGE**

**Updated SOW (V2)**

**Data Acquisition & Understanding**

**NAME: HEET SONI**

**STUDENT NUMBER: 100768241**

**PROFESSOR: MARCOS BITTENCOURT**

**EXECUTIVE SUMMARY**

Mr.Bush started his new mobile company. For better profit, he wanted to check which mobile feature is affecting the price of the mobile phone. Mobile features like battery, Wi-Fi, 3g/4g support, sim support, camera, clock speed, etc are changing the price of the phone. With the help of various machine learning models, we will identify which mobile feature is more important. Moreover, in the end, we will classify the mobile price range.

**BUSINESS PROBLEM STATEMENT**

Classify the mobile pricing in comparison with mobile set features. Additionally, identify which mobile features affecting the price most for a better mobile selling point.

**DATA SOURCE**

Mobile price classification:

<https://www.kaggle.com/iabhishekofficial/mobile-price-classification>

**DATA DESCRIPTION**

The dataset contains various types of independent variables and one dependent variable. Dataset has 2001 rows and 21 columns. All the variables are in numerical formats.

Independent variables:

Battery power, Blue, Clock speed, Dual sim, Fc, Four g, Int memory, M deep, Mobile w, N cores, Pc, Px height, Px width, Ram, Sc h, Sc w, Talk time, Three g, Touch screen, wi-fi.

Dependent variables:

Price range

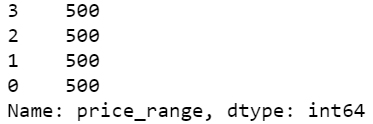
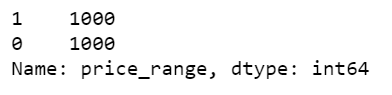
**ABOUT THE MOBILE DATA**

* Here the data have all the values in numerical formate.
* There are no null values available.
* There are no outliers presented in the data.
* In the dataset, the dependent value "price\_range" is divided into four categories.
* All the categories have the same value counts.

**EDA**

* Most of the independent variables are either right-skewed or left-skewed.
* There are no normally distributed variables available.

We converted four dependent variables into two dependent variables.



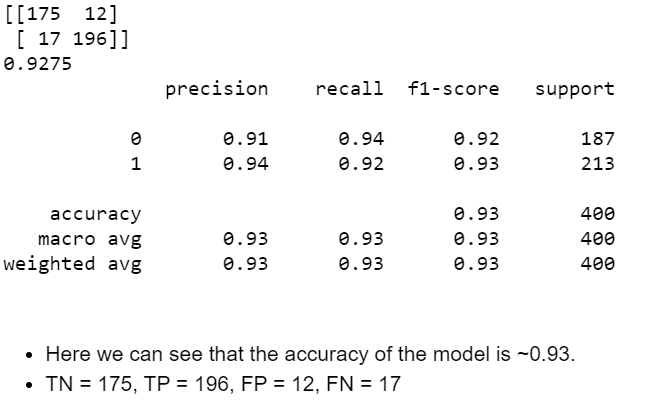
Here some independent variables have positive and some variables have negative effect on the target variable.

Positive effect: dual sim, ram, Px height, Px width, battery power, Pc, Fc, talk time

Negative effect: mobile width

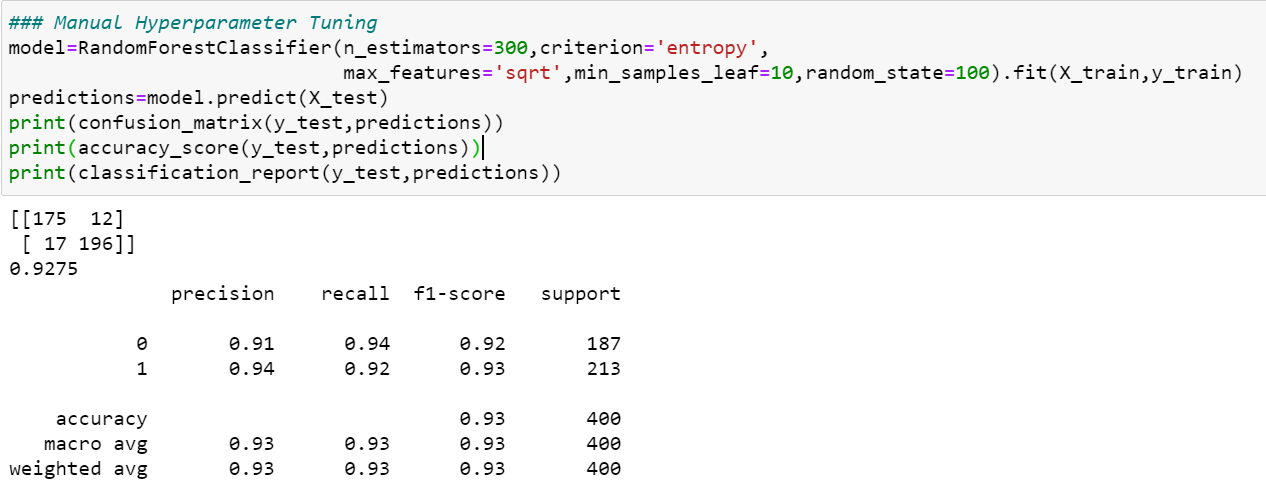
No effect: Bluetooth, wifi, 3G, 4G, internal memory, sc\_h, sc\_w

**BASIC MODEL: RANDOM FOREST**



* + Here we can see that the accuracy of the model is ~0.93.
  + TN = 175, TP = 196, FP = 12, FN = 17

Manual Hyperparameter Tuning:



- Here we Manually added some Hyperparameters for better tuning.

- But there is not much effect on the matrix.

**FUTURE WORK**

* + For future work we will apply Random search CV and Grid search cv for better performance.
  + will try to see what happen to the matrix if we apply PCA method.