

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Neha Pasi
nehapasi05@gmail.com

Please paste the GitHub Repo link.

Github Link:- <https://github.com/haynapasi050505/supervised-machine-learning-classification-capstone-project>

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

Mobile phones come in all sorts of prices, features, specifications and all. Price estimation and prediction is an important part of consumer strategy. Deciding on the correct price of a product is very important for the market success of a product. A new product that has to be launched, must have the correct price so that consumers find it appropriate to buy the product.

For this project I have the Mobile Price dataset with approximately 2000 rows and 11 columns.

First I checked null values and found that there are no numerical column with missing data.

We have found that we have pretty much clean data.

Then I checked the statistical summary of numerical columns and screen width column has minimum value zero which is unjustified. So, I removed them.

After that I have done EDA (Exploratory Data Analysis) on the features of the data to find out some relation between features of a mobile phone (e.g.:- RAM, Internal Memory, etc.) and to find the effect of these features on target variable price of mobile phone.

There is not much difference in count of (has or not) in all these two categorical features except three_g.

76.1% mobile phones supported 3g and 23.9% doesn't.

There is similar distribution of mobile weight in all price range of mobile phones.

The value of ram increased with the increase of price ranges of mobile phones.

Internal memory is more in 1 (medium cost) and 4 (very high cost) of mobile phones.

Battery power increases with increase in price ranges.

After EDA I split the dataset into the Training set and Test set.

Then I defined a function for creating a confusion matrix to know the accuracy score.

After splitting I implemented the four classification algorithms.

1. Gaussian NB classifier.
2. KNN Classifier
3. Random Forest Classifier
4. SVM classifier

Then I done hyperparameter tuning on SVM classifier and found not no improvement in accuracy of model.

I looked at Classification,

When i compared the accuracy score of all the models, the SVM

(support vector machine) model has high accuracy of 96%. So, finally this model is best for predicting the Mobile price Classification.