

Machine learning Engineer Nanodegree

Capstone Proposal

Madhu Babu

(madhukukkuluri@gmail.com)

March 1st,2019

Proposal

Domain Background: Out of all the machine learning models(Supervised ,Unsupervised ,reinforced learning),Supervised learning is one of the most promising field where the maximum development has taken place and is currently using in the real world application.

Our dataset is mobile dataset.This dataset was derived to predict the prices range for the mobiles.This dataset is taken from kaggle.Here we are supposed to train a model that predicts the price ranges by using the features of the mobiles.

Problem Statement:This project is about the Mobile Database. A man started his own mobile company.He doesnot know how to estimate the mobile prices.So he collected sales data of various mobile companies .He wants to find some relation between features of mobile and the selling price of the mobile.So he choose you as his employee.Our work is to predict the rance of mobile price but not the exact price of the mobile.

Hence this project comes under classification.So we should apply a classification model to the dataset so that we can get the best model to predict the range of mobile price . Furthur more the model should be very best useful for the real world applications.

Datasets and Inputs: The datasets or inputs used in the project are thoroughly described.The dataset is named as train.csv contains the following attributes.

- id:ID
- battery_power:Total energy a battery can store in one time measured in mAh
- blue:Has bluetooth or not
- clock_speed:speed at which microprocessor executes instructions
- dual_sim:Has dual sim support or not
- fc:Front Camera mega pixels
- four_g:Has 4G or not
- int_memory:Internal Memory in Gigabytes
- m_dep:Mobile Depth in cm
- mobile_wt:Weight of mobile phone

- n_cores: Number of cores of processor
- pc: Primary Camera mega pixels
- px_height: Pixel Resolution Height
- px_width: Pixel Resolution Width
- ram: Random Access Memory in Megabytes
- sc_h: Screen Height of mobile in cm
- sc_w: Screen Width of mobile in cm
- talk_time: longest time that a single battery charge will last when you are
- three_g: Has 3G or not
- touch_screen: Has touch screen or not
- wifi: Has wifi or not
- price_range: mobile price range(0,1,2,3)

Here the target variable is price_range which we have to predict.

Solution Statement: To solve this problem, we will use multi class classification algorithms to predict the range of price classes i.e 0,1,2,3. It needs a very big dataset in which this approach will try to learn to predict the mobile prices range on its own. We will train the dataset to different models and the model with the highest score can predict the target variable i.e price_range. This type of obtaining a solution is a promising process of training in which the best model is found out and can take the best prediction. Hence the problem can be solved.

Benchmark Model:

The benchmark model for this mobile datasets model can be SVM classifier which can be derived from sklearn.svm. The solution acquired by this will be related to results of the learned algorithm. This is the most benchmark because our learned evaluation will be far better than the benchmark values produced by the benchmark model. Also this benchmark model can help us in many ways because by this benchmark model, we can evaluate our model, if we got the result which has value lower than the value of benchmark model, then we can perfectly say that our model is incorrect. The difference between the learned evaluation and the benchmark model score should be very huge so that our learned model can do prediction well on its own.

Evaluation Metrics:

The evaluation metric used in our learning algorithm and the benchmark model is accuracy score. The accuracy score decides which model is best required for the prediction of the mobile price ranges. The fbeta score can also be used but we are not taking fbeta score because there are four different classes for the target variable for our project and the number

of each classes in the whole dataset are equal i.e, the classes 0,1,2 and 3 having same number of instances in the dataset.The accuracy_score can be imported from sklearn.metrics.

Project Design:

Importing .csv file:

I do my project in the jupyter notebook of python3.Firstly,I will import numpy,pandas for the evaluation of accuracy,for representation purposes,for mathematical calculations etc..And I have put the csv file downloaded in the corresponding project files and now,I will import my csv file by uploading it to the jupyter notebook.In the jupyter notebook,by read_csv function,firstly I will import the .csv file in which mobile dataset is present.

Data Preprocessing:

In our Mobile Dataset,all the attributes are numerical values.In the step of preprocessing, we don't need to do any scaling method because from the pairplot which drawn on this project we have seen that there is no outliers for our data.And also we don't have to change any values because we have all attributes are numerical values.

Train_Test_Split:

Now the preprocessing ends here,and lets jump into the training and testing part.now the dataset should be divided into training and testing part in the ratio of 80% and 20%.

For splitting the data into training and testing parts,we want to import train_test_split from sklearn.cross_validation.

Now,I want to train the training data,so that the trained model can predict price range of the given mobile.

Applying Different Types Of Models:

I choosed some classification algorithms in which I think these models can best predict the model.But,at First,the benchmark model is taken first i.e.,SVM so as to predict the values of the best learned models in which the values are compared with the values of the values of the benchmark model.

The classifiers are choosen for our dataset are DecisionTreeClassifier ,RandomForestClassifier,KNeighbors classifierLogistic Regression, Adaboost classifier etc.. are used to train the model to get the best results.

For the model,which has higher accuracy score,then for that hyperparameter combination is applied to get more accurate results.Hence after applying hyperparameter combination,the model can predict mobile price range on its own i.e 0,1,2,3.