MINOR PROJECT

Project Name:

Machine Learning December Minor Project

Project Description:

Problem statement: Create a classification model to predict whether price range of mobile based on certain specifications

Context: An entrepreneur has started his own mobile company. He wants to give tough fight to big companies like Apple, Samsung etc.

He does not know how to estimate price of mobiles his company creates. In this competitive mobile phone market, one cannot simply assume things. To solve this problem, he collects sales data of mobile phones of various companies.

He wants to find out some relation between features of a mobile phone (e.g., RAM, Internal Memory etc) and its selling price. But he is not so good at Machine Learning. So, he needs your help to solve this problem.

In this problem you do not have to predict actual price but a price range indicating how high the price is

Dataset:

https://drive.google.com/file/d/1z76hFc-oEeA33BWP6YniV9HlaWsQnHK-/view?usp=share_link

Details of features:

The columns are described as follows:

Dataset as 21 features and 2000 entries. The meanings of the features are given below.

1. battery_power: Total energy a battery can store in one time measured in mAh
2. blue: Has bluetooth or not
3. clock_speed: speed at which microprocessor executes instructions
4. dual_sim: Has dual sim support or not
5. fc: Front Camera mega pixels
6. four_g: Has 4G or not
7. int_memory: Internal Memory in Gigabytes
8. m_dep: Mobile Depth in cm
9. mobile_wt: Weight of mobile phone
10. 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 Mega Bytes
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: This is the target variable with value of 0(low cost), 1(medium cost), 2(high cost) and 3(very high cost).

Steps to consider:

1. Remove handle null values (if any).
2. Split data into training and test data.
3. Apply the following models on the training dataset and generate the predicted value for the test dataset
   a) Logistic Regression
   b) KNN Classification
   c) SVM Classifier with linear and rbf kernel
   d) Decision Tree Classifier
   e) Random Forest Classifier
4. Predict the price range for test data
5. Compute Confusion matrix and classification report for each of these models.
6. Report the model with the best accuracy.