

Applied Machine Learning

SECOND SEMESTER 2021-22



Assignment 2

SUBMITTED BY

NAME: Saksham Jain **ID:** 2020hs70033

Observation Section

This data is used 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 competitive mobile phone market, you cannot simply assume things. To solve this problem, we have collected sales data of mobile phones of various companies.

This kind of prediction will help companies estimate price of mobiles to give tough competition to other mobile manufacturer

Also, it will be useful for Consumers to verify that they are paying best price for a mobile.

About The Data

In This Data:

- 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 Wi-Fi or not

About The Algorithms

Chosen Algorithm

- Linear Regression
- Random Forest

Linear Regression

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.

Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables getting used.

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.

Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

Conclusion

Random Forest Algorithm has the best accuracy than Linear Regression Algorithms.

Regression Algorithm	Accuracy (%)
Random Forest	93%
Linear Regression	91%

So, from the above techniques used we can conclude that Random Forest is better than Linear Regression as it shows the accuracy of 93.00% on contrast of Linear Regression that shows 91.00%

Why Random Forest is Better

Random Forests are another way to extract information from a set of data. The appeals of this type of model are: It emphasizes feature selection — weighs certain features as more important than others. It does not assume that the model has a linear relationship — like regression models do.

Random forest is an ensemble of decision trees. This is to say that many trees, constructed in a certain “random” way form a Random Forest.

- Each tree is created from a different sample of rows and at each node, a different sample of features is selected for splitting.
- Each of the trees makes its own individual prediction.
- These predictions are then averaged to produce a single result.