Mobile Price Classification Business Case

The most crucial aspect of promoting a product is its pricing. A smartphone is one of those goods where price is important since it has so many functions that a corporation must consider how to price it in such a way that the features are justified while still covering the marketing and manufacturing expenditures. People constantly updating their cell phones anytime they find new features in a new gadget, making mobile phones the best-selling electronic equipment. Thousands of mobile phones are sold every day, making determining the price of a mobile phone a challenging process for someone who wants to start their own mobile phone business.

In this project, I'm attempting to help a start-up mobile company owner named Bob. He aspires to take on major corporations such as Apple, Samsung, and others. He has no idea how to estimate the cost of the mobile phones his company makes. We can't just assume anything in today's competitive mobile phone market. Bob has collected sales data from multiple brands' mobile phones to solve this problem. Bob is now looking to find correlation between a mobile phone's features (e.g., RAM, Internal Memory, etc.) and its selling price. He is, however, not very good at Machine Learning. As a result, he requires assistance in resolving this issue.

In order to help Bob, I have implemented a Mobile Price Prediction using different Machine Learning Algorithms like Decision tree, K- nearest neighbours and Logistic regression. This project will classify the price range of the mobile price based on the different parameters like camera, touch screen, cores, battery, clock speed, internal memory, battery capacity, etc. The price ranges from 0 (low cost), 1(medium cost), 2 (high

cost),3 (very high cost). After training the model using 3 algorithms, I will be comparing all the models using the graph.

Description of dataset

train.csv

- This data contains 21 column and 2000 rows with no null and categorical values which includes 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 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 Wi-Fi or not).

test.csv

- This data contains 21 column and 1000 rows with no null and categorical values which includes 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 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 Wi-Fi or not).

ML algorithm application

I have mainly used 3 ML algorithm in this project, decision tree classifier, K-nearest neighbours, and Logistics regression. For the decision tree classifier, I found that accuracy is 84 %. However, I also noted that, there is a 10 % difference between actual and predicted data from the confusion matrix (for example: price range 2 predictions).

When it comes to the second algorithm, which is K nearest neighbours. The accuracy that I obtained is 88%. And the actual versus predicted difference is around 4%. For example, for the price range 4, the actual data 40 and the predicted data is 42.

But when I tried with the third algorithm, the logistic regression. It showed an accuracy of 95.5 % and the actual versus predicted difference is 0 to 2%. In the case of price range 0, the actual data is 50 and the predicted data is also 50.

So, after comparing all the three algorithms, I found out that, Logistic regression is the best method to follow for this case and the predicted the price range using the same.

Risks and benefits

The main risk associated with this application is that if the businessperson doesn't have any knowledge about phone and its price. There is a chance that, he might end with giving high price to phone that should fall into cheap ones and vice versa.

When it comes to those entrepreneurs who have a basic knowledge about the phone, this algorithm helps to set price for their phones.

Other Risks

In the current scenario, data used contains no null values, outliers and is a balanced data. If our training data contains too much info, that will lead to noise and inaccurate data entries. Then the model does not categorize the data correctly. So, if that the case with the data, then it will result in overfitting. So, in such cases always follow linear algorithm.