Capstone Three - Project Proposal

What is the problem you want to solve?

Bob has started his own mobile company. He wants to give tough fights to big companies like Apple,Samsung etc. He does not know how to estimate the price of mobiles his company creates. In this competitive mobile phone market you cannot simply assume things. To solve this problem he collects sales data of mobile phones of various companies. Suppose that as a data scientist working for Bob, I need to find a certain relationship between the functions of mobile phones and their selling prices for Bob to predict the price range of mobile phones produced by the company.

Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis?

Obviously Bob, the boss of the company, is my client. Because as a data scientist I have to work for him and help him deal with real business problems. As the owner of the company, Bob needs to consider the issue of pricing and selling mobile phones after production, because in today's society there are various types of mobile phones. If our company wants to succeed in the market, it needs to compare its functions with other mainstream mobile phone brands to launch the most popular mobile phone brand. A price range that is conducive to our company's profitability. According to my final analysis, Bob will get a price range and the impact of the use of different types of mobile phone functions on customers through machine learning and modeling.

What data are you using? How will you acquire the data?

I'm going to use the data set called"Mobile Price | Multiclass classification" from Kaggle.com, and this dataset contains all the functions of current mainstream mobile phones and the frequency of customer use. I'm going to use the pandas library, and functions like read_csv() to collect the data from the database.

Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

First, I will perform data collection and extract the csv file of the database into my notebook. Next I'll create a folder to store any data visualizations I'll get to later. Then I will go to understand the data characteristics. Use the formulas .columns, .dtypes, .info() to grasp the main characteristics of this data, build different classification models to predict the mobile phone price range. Also Imputing missing values , converting cat to num and then do the data visualization to make more clearly version.

After that choosing models and test them, and also training the model and evaluate it. After these all make predictions and final conclusions.

What are your deliverables?

A code, and there are also data visualization charts in this notebook file.