Problem 4: Understanding the Task

After examining the description and data, your goal now is to clearly state the Business Objective of the task. State the objective below.

4.1.0 Business Objectives and Background

The client is the sales and marketing leadership of a Portugese bank.

This bank's marketing department has conducted multiple phone call based marketing campaigns contacting prospects one or more times in order to determine if the prospect would make a term deposit.

Based on the data collected, the client would like to understand what type of customer is most likely to subscribe to a term deposits, and, possibly, segment these customers into different populations (target markets) based on their characteristics (demographics).

The bank's marketing department can then use this information to determine how to best increase sales of their term deposit product

4.1.1 Business success criteria

The banks success is determined by the profit is is able to generate from its current and future client base by selling to the client its 'products'. In this case, the product is a term deposit, however, there are other products that could also be sold to the same prospect. In the case of this specific product, term deposits, the success criteria is being able to identify which prospects, whether currently customers of the bank or not, are most likely to purchase the product. Once this is known, the

bank can then focus its limited resources on selling to these likely product customers, thus reducing sales expenses spent on customer who are not likely to buy the product AND prioritize sales activities on customer who are likely to by the product, thus maximizing revenue and profit for the product.

4.2.0 Inventory of resources

Data Resources: The data resource available are two data sets, one one with all observations and on, a random subset of 10% of these total observations located at https://archive.ics.uci.edu/dataset/222/bank+marketing.

Technical Resources: Anaconda, Jupyter notebooks, Google Colab, Python libraries.

Manpower Resources: One individual

4.2.1 Requirements, assumptions, constraints

Requirements:

- Identify possible characteristics / demographics of prospects who will or who will not purchase the product (primary)
- Further segment possible purchases into target markets (secondary)

Assumptions

- Determinations and recommendations will be made on complete data
- The scope of the analysis is restricted to the data provided
- Both current bank customers and non-customers are included in both data sets

Constraints

- As an exercise, it is not possible to verify business

objectives or requirements
- The business objectives, requirements, assumptions and constraints can only be assumed
- While there are four data sets total available online, only two will be used for this analysis, and one of these sets is a randomized subset of the other
- Computation Restraints... the number of rows of data and features make it to computationally intensive to use the full data set
-

4.3 Determine the data mining goals

4.3.0 Data Mining Goals

Primary Goal - Segment the prospects into two groups, group A - respondents who have indicated they will subscribe to (purchase) the product and group B - those who will not.
Secondary Goal - Characterize these two groups' features
Tertiary Goal - Segment group A into target markets

4.3.1 Data Mining Success Criteria

The respondents can be segmented into Group A and Group B

4.4 Project Plan

4.4.0

The data set will be accessed, understood, prepared, modeled and evaluated in order to meet the data mining success criteria, and, thusly, the business objectives

4.4.1 Project plan steps

1   Import data

Recommendation

While the Best Model is a Random Forest, and the SHAP values show Duration of Contact and Days from Last Contact are the most relevant features for making predictions, the actual performance of the model is not significantly different from that of the baseline.

The Best Model's accuracy was 0.897167 and a precision of 0.52427, compared to the baseline score for the full data set of 0.87335. and baseline score for the small data set is: 0.88667.

First, the model should be run and fitted against the entire dataset. Once this is done, the accuracy can be computed again.

The full set baseline Random Forest model, without performing hyper tuning (due to the compute resources being insufficient to run GridSearch) an Accuracy of 0.8974. The difference from best performing model could be compared to the 0.87335 baseline, and this could be used to determine the new improvement. However, given the low Precision scores, the ability to use the model to correctly identify whom to target is not possible

Rather, since the most highly correlated and best indicator of whether a prospect will purchase is the duration of the contact (where a contact of 0 means there is no purchase, which in itself needs to be explored), the bank could calculate the cost per minute of the campaign per contact and then compare this cost against the additional % increase in product sales. However, we will not that this possible ~1% increase in sales, is only the assertion of a sale by the respondent, not the actual sale itself, which is probably lower than 1%.

If the dollar value of increased sales is greater than the cost of the campaigns AND the opportunity cost of the campaign is less than other alternative opportunities to increase sales and profitability, then continuing the marketing campaigns may make sense. However, an initial look would recommend NOT continuing the marketing campaigns.