Final Capstone

Bank Marketing Campaign- Predicting Term Deposit Subscriptions

Goal

The goal of this project is to analyze the bank's most recent marketing campaign and identify patterns that will help the financial institution improve for future campaigns. This analysis will enable the bank to have greater effectiveness in future marketing campaigns.

Problem Identification

How can the financial institution have a greater effectiveness for future marketing campaigns?

Approach

I will follow the steps as in the first guided capstone as the following:

1. Problem Identification: this proposal
2. Data Wrangling: data collection, data organization, data definition, data cleaning
3. Exploratory data analysis: explore the data relationships of all the features and understand how the features compare to the response variable
4. Pre-processing: create a cleaned development dataset to use to complete the modeling step
5. Modeling: build two to three different models and identify the best one, and review model outcomes
6. Documentation: finalize code, review the results, finalize documentation
7. Presentation - a slide deck with 10-20 slides

Data sources

The data is available from [kaggle].

Marketing bank dataset uploaded originally in the UCI Machine Learning Repository. The dataset gives information about a marketing campaign of a financial institution in which we will have to analyze in order to find ways to look for future strategies in order to improve future marketing campaigns for the bank.

Attribute Description

Input Variables:

1. **bank client data:**

   1 - **age:** (numeric)
   2 - **job:** type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')
   3 - **marital:** marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)
   4 - **education:** (categorical: primary, secondary, tertiary and unknown)
   5 - **default:** has credit in default? (categorical: 'no','yes','unknown')
   6 - **housing:** has housing loan? (categorical: 'no','yes','unknown')
   7 - **loan:** has personal loan? (categorical: 'no','yes','unknown')
   8 - **balance:** Balance of the individual.

2. **Related with the last contact of the current campaign:**

   9 - **contact:** contact communication type (categorical: 'cellular','telephone')
   10 - **month:** last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
   11 - **day:** last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
   12 - **duration:** last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

3. **Other attributes:**

   13 - **campaign:** number of contacts performed during this campaign and for this client (numeric, includes last contact)
   14 - **pdays:** number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously

contacted)
15 - **previous:** number of contacts performed before this campaign and for this client (numeric)
16 - **poutcome:** outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

Output variable (desired target):
17 - **y** - has the client subscribed a term deposit? (binary: 'yes','no')

Deliverable

I will create a github repo containing the work I complete for each step of the project including Jupyter notebooks, a final report, and a slide deck.