

Week 7

Data Science Intern at Data Glacier
Project: Bank Marketing (Campaign)

Name: Adrian Baysa
Email: adrianbaysa2@gmail.com
Country: Philippines
Batch Code: LISUM16

1. Problem Description

ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

2. Business Understanding

Bank wants to use ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc) can focus only to those customers whose chances of buying the product is more.

This will save resource and their time (which is directly involved in the cost (resource billing)).

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

3. Project Lifecycle along with Deadlines

- a. Week 7 (Jan 19, 2023):
 - i. Problem Description
 - ii. Business Understanding
- b. Week 8 (Jan 26, 2023):
 - i. Data Understanding and Data Type Analysis
 - 1. The data is related with direct marketing campaigns of a Portuguese banking institution.
 - 2. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.
 - 3. The bank-full.csv dataset has all examples and 17 inputs, ordered by date (older version of this dataset with less inputs).
 - 4. Attribute Information:
 - a. 1 - age (numeric)
 - b. 2 - job : type of job (categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
 - c. 3 - marital : marital status (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)
 - d. 4 - education (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')
 - e. 5 - default: has credit in default? (categorical: 'no', 'yes', 'unknown')
 - f. 6 - housing: has housing loan? (categorical: 'no', 'yes', 'unknown')
 - g. 7 - loan: has personal loan? (categorical: 'no', 'yes', 'unknown')
related with the last contact of the current campaign:
 - h. 8 - contact: contact communication type (categorical: 'cellular', 'telephone')
 - i. 9 - month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
 - j. 10 - day_of_week: last contact day of the week (categorical: 'mon', 'tue', 'wed', 'thu', 'fri')
 - k. 11 - 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.
other attributes:
 - l. 12 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

- m. 13 - 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)
 - n. 14 - previous: number of contacts performed before this campaign and for this client (numeric)
 - o. 15 - poutcome: outcome of the previous marketing campaign (categorical: 'failure', 'nonexistent', 'success')
 - # social and economic context attributes
 - p. 16 - emp.var.rate: employment variation rate - quarterly indicator (numeric)
 - q. 17 - cons.price.idx: consumer price index - monthly indicator (numeric)
 - r. 18 - cons.conf.idx: consumer confidence index - monthly indicator (numeric)
 - s. 19 - euribor3m: euribor 3 month rate - daily indicator (numeric)
 - t. 20 - nr.employed: number of employees - quarterly indicator (numeric)
 - u. Output variable (desired target): 21 - y - has the client subscribed a term deposit? (binary: 'yes', 'no')
- c. Week 9 (Feb 2, 2023):
 - i. Data Cleansing/Features Engineering
- d. Week 10 (Feb 9, 2023):
 - i. Exploratory Data Analysis
- e. Week 11 (Feb 16, 2023):
 - i. Exploratory Data Analysis for Business Users
 - ii. Exploratory Data Analysis for Recommended Models
- f. Week 12 (Feb 23, 2023):
 - i. Handling Data Imbalance Constraints
 - ii. Comparing Models
 - iii. Final Evaluation
 - iv. Converting ML Metrics into Business metrics and Explanation of Results
- g. Week 13 (Feb 28, 2023):
 - i. Final Solution/Recommendation (Slide Presentation)

4. Data Intake Report

- a. Name: Bank Marketing Data Set (bank-full.csv)
- b. Report Date: 1/19/2023

- c. Data storage location:
 - i. <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
 - ii. https://raw.githubusercontent.com/agbaysa/dataglacier_week7/main/bank-additional-full.csv
- d. Tabular data details: bank-additional-full.csv with all examples (41188) and 20 inputs, ordered by date (from May 2008 to November 2010)
- e. Total Number of observations: 41,188
- f. Total Number of Files: 1
- g. Total Number of features: 21 columns, 1 target variable
- h. Base format of the file: csv
- i. Size of data: 6.6MB

5. Github Repolink

- a. https://github.com/agbaysa/dataglacier_week7/