Bank Marketing (Campaign)

Assignment

Week 8 Assignment

Prepared by:

Aly Medhat Moslhi

Submitted to:

Data Glacier Internship Program

Patch No.:

LISUM11

Aug. 25th, 2022

Table of Content

Team Members' Details

Track: Data Science
Batch No.: *LISUM11*

Name: Aly Medhat Moslhi
Country: Egypt
Email: alymedhat10@yahoo.com
Company: AAST – Teaching assistant

Problem Description

ABC Bank wants to sell it's 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).
The bank wants to use ML model to shortlist customer whose chance of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc) can focus only to those customers whose chance of buying the product is more.

i. Problem statement

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, 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)

Data Understanding

i. Dataset Columns & Data Types

The Main dataset (Bank-additional-full data) is formed of 41188 rows and 21 columns, the columns are formed of 10 numeric columns and 11 categorical columns defined as:

A. Customer Information:

1 - Age (numeric)

2 - Job (categorical): "type of job"

- admin
- blue-collar
- technician
- services
- management
- retired
- entrepreneur
- self-employed
- housemaid
- unemployed
- student
- unknown

3 - Marital Status (categorical): "marital status"

- married
- single
- divorced
- unknown

4 - Education (categorical):

- university Degree
- high school
- basic.9y
- professional Course
- basic.4y
- basic.6y
- unknown
- illiterate

5 - default (categorical): “Does the customer has credit in default?”

- yes
- no
- unknown

6 - housing (categorical): “Does the customer has a housing loan?”

- yes
- no
- unknown

7 - loan (categorical): “Does the customer have a personal loan?”

- yes
- no
- unknown

B. Communication Information:

8 - contact (categorical): communication type

- cellular
- telephone

9 - month (categorical): “last contact month of the year”

10 – day of week (categorical): “last contact day of the week (in working days)”

11 - duration (numeric):” last contact duration, in seconds”

C. Campaign Information:

12 - campaign (numeric): number of contacts performed during this campaign and for this client

13 - pdays (numeric): number of days that passed by after the client was last contacted from a previous campaign

14 - previous (numeric): number of contacts performed before this campaign and for this client

15 - poutcome (categorical): outcome of the last campaign marketing

- nonexistent

- failure
- success

D. social and economic context attributes

16 - emp.var.rate (numeric): employment variation rate - quarterly indicator

17 - cons.price.idx (numeric): consumer price index - monthly indicator

18 - cons.conf.idx (numeric): consumer confidence index - monthly indicator

19 - euribor3m (numeric): Euro Interbank Offered 3-month rate - daily indicator

20 - nr.employed (numeric): number of employees - quarterly indicator

E. Target:

21 - y - has the client subscribed to a term deposit? (binary: 'yes','no')

ii. Problems within the data

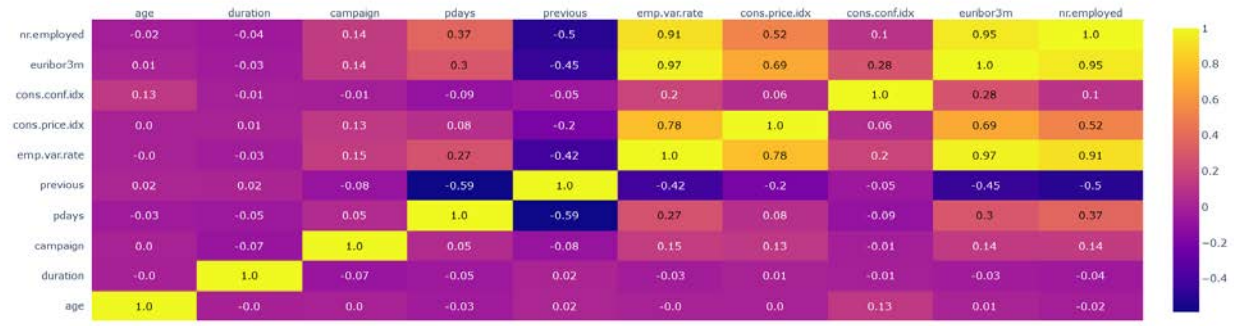
After analyzing the data, the following comments were noted:

1. There is no NaN values within the datasets
2. There are some duplicate rows

Analysis Approach

i. Multivariate analysis of the data

The heatmap showed in figure 1 shows the correlation between the numeric variables. It shows that there are verry high correlation between employment variation rate, number of employees, Euro Interbank Offered 3-month rate and the consumer confidence index.



ii. Categorical vs Categorical

It is difficult to analyze most of the features manually since most of the features are categorical. Therefore it is better to compare them using feature engineering after data cleaning.

iii. Numeric vs Categorical analysis:

There are 9 numeric columns and 11 categorical columns. Showing the correlation between them gives an insight of the data. For instance comparing the job and the age of a client and the how it can affect the decision.

Data Intake Report

Name: Bank Marketing (Campaign)
Report date:
Internship Batch: LISUM11: 30 June - 30 Sept 2022
Version: 1.0
Data intake by: Aly Medhat Moslhi
Data intake reviewer:
Data storage location:

Bank-additional-full data

Total number of observations	41188
Total number of files	1
Total number of features	20
Base format of the file	.csv
Size of the data	5.8 MB

Bank-additional data

Total number of observations	4119
Total number of files	1
Total number of features	20
Base format of the file	.csv
Size of the data	586 KB

Bank-additional-name data

Total number of files	1
Base format of the file	.txt
Size of the data	8 KB

GitHub Repository link

https://github.com/alymedhat10/Bank-Marketing-Campaign-.git