Data Science Internship

Week 7: Data Science Project: Bank Marketing (Campaign)

Group Name: One
Name: Hira Fahim
Email: hirashahidd26@yahoo.com
Country: United Kingdom
Company: Unemployed
Specialization: Data Science
Batch Code: LISUM19
Submission Date: 5th April 2023
Submitted to: Data Glacier

Table of Contents

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

2.1. Objectives

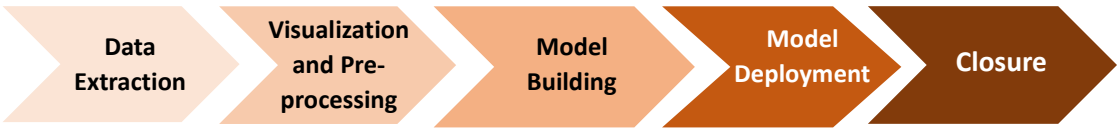
The goal is to build a binary classification model to predict whether the client will subscribe a term deposit or not.

2.2. Strategy

The analysis consists of four parts:

- Data Understanding.
- Perform exploratory analysis.
- Data Visualisation and Pre-processing.
- Model building.
- Model deployment.

3. Project lifecycle



4. Dataset Information

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.

5. Attribute Information

5.1. Input Variables

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', 'unknown')
5. default: has credit in default? (Categorical: 'no', 'yes')
6. balance: average yearly balance, in euros (numerical)

7. housing: has housing loan? (Categorical: 'no', 'yes')
8. loan: has personal loan? (Categorical: 'no', 'yes')
   # related with the last contact of the current campaign:
9. contact: contact communication type (categorical: 'cellular', 'telephone', Unknown)
10. day: last contact day of the month (numeric)
11. month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
12. duration: last contact duration, in seconds (numerical)
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', 'unknown', 'success', 'other')

5.2. Output variable (desired target)

17. y - has the client subscribed a term deposit? (Binary: 'yes', 'no')

6. Import Data Set

```
In [1]: import pandas as pd
        import numpy as np

In [2]: d= pd.read_csv("bank-full.csv")

In [3]: d.head()
```

Out[3]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
0	58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
1	44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
2	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
3	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
4	33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no

7. Dataset Details

- Shape of the dataset (Number of rows and columns)

```
In [18]: # no of rows and columns
         d.shape

Out[18]: (45211, 17)
```

Number of rows = 45211
Number of columns = 17

- Datatype of Columns and Non-null values

```
In [19]: # Datatypes of columns and non-null values
         d.info()

         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 45211 entries, 0 to 45210
         Data columns (total 17 columns):
          #   Column     Non-Null Count  Dtype
         ---  ------     --------------  -----
          0   age        45211 non-null  int64
          1   job        45211 non-null  object
          2   marital    45211 non-null  object
          3   education  45211 non-null  object
          4   default    45211 non-null  object
          5   balance    45211 non-null  int64
          6   housing    45211 non-null  object
          7   loan       45211 non-null  object
          8   contact    45211 non-null  object
          9   day        45211 non-null  int64
          10  month      45211 non-null  object
          11  duration   45211 non-null  int64
          12  campaign   45211 non-null  int64
          13  pdays      45211 non-null  int64
          14  previous   45211 non-null  int64
          15  poutcome   45211 non-null  object
          16  y          45211 non-null  object
         dtypes: int64(7), object(10)
         memory usage: 5.9+ MB
```

- Numerical and categorical Features

```
Numeric Features:
Index(['age', 'balance', 'day', 'duration', 'campaign', 'pdays', 'previous'], dtype='object')
================================================================================
Categorical Features:
Index(['job', 'marital', 'education', 'default', 'housing', 'loan', 'contact',
       'month', 'poutcome', 'y'],
      dtype='object')
```

- Null values

```
# total null values in the dataset
d.isnull().sum()

age          0
job          0
marital      0
education    0
default      0
balance      0
housing      0
loan         0
contact      0
day          0
month        0
duration     0
campaign     0
pdays        0
previous     0
poutcome     0
y            0
dtype: int64
```

There are no null values in the dataset.

8. Reference

This dataset is publicly available for research. The details are described in [Moro et al., 2011].
[Moro et al., 2011] S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimaraes, Portugal, October 2011. EUROSIS.
Available at:
[pdf] http://hdl.handle.net/1822/14838

[bib] http://www3.dsi.uminho.pt/pcortez/bib/2011-esm-1.txt

Source:
Created by: Paulo Cortez (Univ. Minho) and Sergio Moro (ISCTE-IUL) @ 2012