



Data Science Internship at Data Glacier

Week 8

Project: Bank Marketing (Campaign)

Team Name: The Data-driven Dreamers

Name: Rayhanul Islam Rumel
Email: rayhanul.islam.rumel@gmail.com
Country: Germany
University: Saarland University
Specialization: Data Science

Name: Uday Vikram Singh
Email: udaykymset@gmail.com
Country: Scotland
University: University of Edinburgh
Specialization: Data Science

Name: Amogh Vig
Email: vigamogh@gmail.com
Country: United States
University: Vanderbilt University
Specialization: Data Science

Batch Code: LISUM17

Date: 26/02/2023

Submitted to: Data Glacier

Table of Content

| | |
|---|----------|
| Problem Description | 3 |
| Data Understanding | 3 |
| The type of data we have got for analysis | 3 |
| The problems in the dataset | 4 |
| The approaches we are trying to apply to our data set to overcome problems | 4 |

Problem Description

ABC Bank is planning to introduce a new term deposit product for its customers, and they need to develop a model that can predict whether a customer will buy their product or not. To achieve this goal, they want to analyze the past interaction of customers with the bank or other financial institutions. The bank wants to use this model to understand the customers' behavior and preferences regarding term deposits, and to determine which customers are most likely to buy the product. This will help the bank to focus its marketing efforts on those customers who are more likely to buy, resulting in a more efficient and effective sales process. Ultimately, the bank hopes to increase its sales and revenue through the successful launch of its new term deposit product.

Data Understanding

About Dataset: This dataset pertains to the direct marketing efforts of a Portuguese bank. The campaigns relied on phone calls, and it was often necessary to make multiple attempts to contact the same customer in order to determine whether or not they would subscribe to the bank's term deposit product. Responses were either 'yes' or 'no'.

The main dataset titled *bank-full.csv* contains **45211 observations** and **17 features**. We used methods for example *head()*, *info()*, and *describe()* to understand the dataset. We observed that the dataset is well-processed and clean. Among 17 features 7 features are *int64*, and 10 features are *objects*.

All of the *object* type values are already converted to lowercase letters. All the numeric values are the type of *int64*. So, no conversion/ transformation is required.

The type of data we have got for analysis

The dataset contains two types of data: *int64*, and *object*

The problems in the dataset

Luckily, the dataset is already processed and clean. Therefore, the dataset doesn't contain any Missing values, Outliers, or Duplicate values.

The approaches we are trying to apply to our data set to overcome problems

Although minimal data processing is required for the dataset we are handling, we have taken into account the following methods to mitigate potential issues if necessary:

- Identifying, imputing, and/ or removing missing values
- Identifying, and removing duplicate values
- Identifying, and removing outliers from the dataset
- Converting columns to appropriate data types
- Rename column names if necessary

Github Repo link

Please visit the following link to see our project:

https://github.com/rirumel/data_science_bank_marketing_Campaign/tree/main