Group name: Laaroussi Saadeddine

Name: Laaroussi Saadeddine

Email: laar.saad.eddine@gmail.com

Country: Morocco

Specialization: Data science

Project: Bank Marketing (Campaign)

Batch code: LISUM09

Table of contents

Problem description

- ABC Bank wants to sell its term deposit product to customers.

- Before launching the product, they want to develop a model that will 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).

- ABC Bank wants to use ML (machine learning) model to shortlist customers whose chances of buying the product are higher.

- They want their marketing channel (tele marketing, SMS/email marketing etc) to focus only on those customers whose chances of buying the product are higher.

- 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).

Data understanding

- Features and values

Features	Types	Description	Values	Null ?	Outliers ?
Age	Int64	Age of the person	Between 19 and 95	No	No
Job	Object	Job of the person	['admin.' 'blue-collar' 'entrepreneur' 'housemaid' 'management' 'retired' 'self-employed' 'services' 'student' 'technician' 'unemployed' 'unknown']	No	No
Marital	Object	Marital situation	['divorced' 'married' 'single']	No	No
Education	Object	Education	['primary' 'secondary' 'tertiary' 'unknown']	No	No
Default	Object	Has a default credit	['no' 'yes']	No	No
Balance	Int64	Amout of balance	Between -8019 and 102127	No	Yes
Housing	Object	Has a house	['no' 'yes']	No	No

Loan	Object	Took a loan	['no' 'yes']	No	No
Contact	Object	Was contacted with	['cellular' 'telephone' 'unknown']	No	No
Day	Int64	Number of day in a month	From 1 to 31	No	No
Month	Object	Months of a year	['apr' 'aug' 'dec' 'feb' 'jan' 'jul' 'jun' 'mar' 'may' 'nov' 'oct' 'sep']	No	No
Duration	Int64	Last contact duration, in seconds	Between 0 and 4918	No	Yes
Compaign	Int64	Number of contacts performed for this campaign for this client	Between 1 and 63	No	Yes
Pdays	Int64	Number of days before last contact	Between -1 and 871	No	Yes
Previous	Int64	Number of contacts performed before this compaign for this client	Between 0 and 275	No	Yes
Poutcome	Object	Outcome of the previous marketing campaign	['failure' 'other' 'success' 'unknown']	No	No
Y	Object	Has subscribed or not	['no' 'yes']	No	No

First analysis (Outliers, Skewness, NA values)

Numeric features that might contain outliers are :

- Age
- Balance
- Day
- Duration
- Compaign
- Pdays
- Previous

All features aside from Age are skewed to the right.

There are no NA values in data, however from the data and the skewness, we can see that the null values in many features are either -1 or 0, or ‘unknown’ for the object values.

Outliers are present in data as well. For example for the feature previous max value 275 while the value before 275 is 58.

Removing outliers with IQR (interquartile range) will help fix the skewness and overall data.

Github repo link:

https://github.com/Saad-code13/Internship_Assignment/tree/main/week%208