BANK MARKETING CAMPAIGN

NAME: CATHERINE SANDA
EMAIL: sandacate@gmail.com
COUNTRY: KENYA
SPECIALIZATION:DATA SCIENCE
GITHUB REPO LINK: https://github.com/cate6495/bank-week7

PROBLEM DESCRIPTION

We are given data related to direct marketing campaigns i.e. phone calls of a bank in Portugal.

The classification goal is to predict whether a client will subscribe or not(yes/no) to a term deposit (variable y).

BUSINESS UNDERSTANDING

The aim is that the bank gains a clearer understanding of its customer base and predict customers response to its telemarketing campaigns. By analyzing customer features the bank will be able to predict customer saving behaviors and identify the customers who are more likely to subscribe to term deposits. The bank can then focus its marketing campaigns on those customers. This will not only allow the bank to secure term deposits more effectively but also reduce the cost incurred in the marketing campaigns and increase customer satisfaction by reducing undesirable advertisements to certain customers.

PROJECT LIFECYCLE (3/09/2022-4/20/2022)

Below is a summary of the steps:

a) Business Problem

The bank aims to find out the characteristics that help to determine how customers successfully subscribe for deposits, which helps in increasing campaign efficiently and selecting high value customers. As such the goal is to build Machine Learning model that learns the unknown patterns, maps and several input features classifying whether client will subscribe for longer deposits or not.

b) Data Collection

We will pick our dataset from the UCI Machine Learning repository.

c)Data preparation

In this step we will:

- Select the relevant data.
- Handle the missing values by either removing the rows containing them or by imputing them with relevant data.

- Remove erroneous data if any.
- Check for outliers and handle them.
- Remove unwanted columns and features.
- Derive new features from existing ones by using feature engineering.
- Perform EDA. The insights from EDA could help in finding the right set of features and algorithm to be used in model building.

d) Model Building and Selection

Use the prepared data as input to get our desired output i.e. the target variable. We then define a set of metrics to compare the models and choose the top performing one. The data will determine the appropriate machine learning algorithm that is best suited for the model.

Then we evaluate the model by testing the accuracy (how well does our model perform? Does it describe the data accurately?) and the relevance (Does our model answer the original question)

DATA INTAKE REPORT

Bank-additional-full.csv

Total number of observations	41188
Total number of files	1
Total number of features	21
Base format of the file	csv
Size of the data	4.70MB