Group Name: Pattern Pros

Github Repository Link:
[DataGlacier/Group_Project at main · danielkingswood/DataGlacier (github.com)]

Team Member Details:

Name	Email-ID	Country	University	Specialisation
Jay Panara	jay.panara@gmail.com	Canada	University of Waterloo	Data Science
Shreya Dwivedi	shreyad@usc.edu	USA	University of Southern California	Data Science
Sarah Sindeband	ssindeband2018@fau.edu	USA	Florida Atlantic University	Data Science
Daniel Kingswood	ddk727@gmail.com	UK	University of Bristol	Data Science

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 helps 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).

Business Understanding:

The bank wants to use the ML model to shortlist customers whose chance of buying the product is more so that their marketing channel (telemarketing, SMS/email marketing, etc.) can focus only on those customers who have a greater chance of buying the product.
This will save resources and their time (which is directly involved in the cost (resource billing)).
We need to develop a model with duration and without duration features and report the performance of the model.
The data is related to 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 assess 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) to a term deposit (variable y).

Attribute Information:

Input variables:
bank client data:
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:
'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course', 'university.degree','unknown')
5 - default: has credit in default? (categorical: 'no','yes','unknown')
6 - housing: has a housing loan? (categorical: 'no','yes','unknown')
7 - loan: has a personal loan? (categorical: 'no','yes','unknown')

related with the last contact of the current campaign:
8 - contact: contact communication type (categorical: 'cellular','telephone')
9 - month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
10 - day_of_week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
11 - duration: last contact duration, in seconds (numeric).
Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

other attributes:
12 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
13 - 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)
14 - previous: number of contacts performed before this campaign and for this client (numeric)
15 - poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')
social and economic context attributes
16 - emp.var.rate: employment variation rate - quarterly indicator (numeric)
17 - cons.price.idx: consumer price index - monthly indicator (numeric)
18 - cons.conf.idx: consumer confidence index - monthly indicator (numeric)
19 - euribor3m: euribor 3 month rate - daily indicator (numeric)
20 - nr.employed: number of employees - quarterly indicator (numeric)

Output variable (desired target):
21 - y - has the client subscribed a term deposit? (binary: 'yes','no')

Project Lifecycle:

Week	Task	Deadline
Week 7	Business Understanding	19th October, 2022
Week 8	Data Understanding	26th October, 2022
Week 9	Data Cleansing and Transformation	2nd November, 2022
Week 10	EDA with Recommendation	9th November, 2022
Week 11	Presentation of EDA for business and recommended models for technical user	16th November, 2022
Week 12	Model selection and building with dashboard	23rd November, 2022
Week 13	Final Project submission and report with powerpoint	30th November, 2022

Data Intake Report

Name: Bank Marketing Campaign
Report date: 10/17/2022
Internship Batch: LISUM13: 30
Version:1.0
Data intake by: Sarah Sindeband
Data intake reviewer:<Jay Panara>
Data storage location:
DataGlacier/Group_Project at main · danielkingswood/DataGlacier (github.com)

Tabular data details: bank.csv

Total number of observations	45211
Total number of files	1
Total number of features	17
Base format of the file	.csv
Size of the data	450 KB

Tabular data details: bank-full.csv

Total number of observations	4521
Total number of files	1
Total number of features	17
Base format of the file	.csv
Size of the data	439 MB

Tabular data details: bank-additional.csv

Total number of observations	4119
Total number of files	1
Total number of features	21
Base format of the file	.csv
Size of the data	570 KB

Tabular data details: 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	5.56 MB