Week 8: Deliverables

Group Name: AI Boys

Date: 11/25/2022

Group Member ID	Name	Email	Country	College	Specialization
1	Islom Pulatov	islompulatov115@gmail.com	Poland	Epicode Global	Data Science
2	Ammar Sidhu	ammarsidhu@outlook.com	Canada	University of Toronto	Data Science

Project: Bank Marketing (Campaign)

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).

Business Understanding:

- 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.

  The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

Data Understanding

- 45,211 observations (rows)
- 17 features (16 predictors + 1 target)
- CSV data format (single file)

Types of Data for Analysis

- 9 Categorical Features

(1) job : type of job (categorical: "admin.","unknown","unemployed","management","housemaid","entrepreneur","student", "blue-collar", "self-employed", "retired", "technician", "services")

(2) marital : marital status (categorical: "married", "divorced", "single"; note: "divorced" means divorced or widowed)
(3) education (categorical: "unknown", "secondary", "primary", "tertiary")
(4) default: has credit in default? (Binary: "yes", "no")
(5) housing: has housing loan? (Binary: "yes", "no")
(6) loan: has personal loan? (Binary: "yes", "no")
(7) contact: contact communication type (categorical: "unknown", "telephone", "cellular")
(8) month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")
(9) poutcome: outcome of the previous marketing campaign (categorical "unknown", "other", "failure", "success")

- 7 Numerical Features
  (1) age (numeric)
  (2) balance: average yearly balance, in euros (numeric)
  (3) day: last contact day of the month (numeric)
  (4) duration: last contact duration, in seconds (numeric)
  (5) campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
  (6) pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)
  (7) previous: number of contacts performed before this campaign and for this client (numeric)
- 1 Target Feature (Categorical)
  (1) y - has the client subscribed to a term deposit? (Binary: "yes", "no")

Problems in the Data

- Number of NA Values: 0
- Outliers: N/A – Target Variable is Categorical and Binary for Classification
- Skewed: N/A – Target Variable is Categorical and Binary for Classification
- Potential Outliers and Skew in Numerical Features
- Classes (Predictor): 2
- Imbalanced Classes (over 80% 'No' and less than 20% 'Yes')

Approaches to Overcome Problems in Data

- Implement 2 Different Imbalanced Classification Techniques and Compare Predictions:
  (1) Cost-Sensitive Learning (Islom)
  (2) SMOTE (Ammar)
- Outliers in Numerical Features:
  (1) Train and Test Models with Outliers Present (Islom)
  (2) Train and Test Models without Outliers – Drop Outliers (Ammar)

Github: AI Boys