

Part 1- Supervised Learning

30 marks

Given is the 'Portugal Bank Marketing' dataset:

Bank client data:

- 1) **age** (numeric)
- 2) **job**: type of job
(categorical: "admin.", "bluecollar", "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**: education of individual (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 housing loan? (categorical: "no", "yes", "unknown")
- 7) **loan**: has 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) **dayofweek**: 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) **concave points_se:** standard error for number of concave portions of the contour

20) **euribor3m:** euribor 3 month rate - daily indicator (numeric)

21) **nr.employed:** number of employees - quarterly indicator (numeric)

Output variable (desired target):

22) **y:** has the client subscribed a term deposit? (binary: "yes","no")

Perform the following tasks:

Q1. Perform EDA on the given data. What does the primary analysis of several categorical features reveal? □5 marks

Q2. Perform the following pre-processing tasks: □10 marks

- a. Missing Value Analysis
- b. Label Encoding wherever required
- c. Selecting important features based on Decision Trees
- d. Handling unbalanced data using SMOTE
- e. Standardize the data using any one of the scalers provided by sklearn

Q3. Build the following Supervised Learning models: □5 marks

- a. Logistic Regression
- b. Decision Trees

Q4. Tabulate the performance metrics of all the above models, perform tuning of models and tell which model performs better in predicting if the client will subscribe to term deposit or not → 10 marks

Part 2 - ANN

20 marks

A famous 30-year-old pizza brand which has outlets in more than 90 countries started home delivery services a couple of years ago and the business has grown much faster than expected. However, outlet vendors are very much disappointed with few customers for their cheating activities. This is because vendors, shockingly, came to know that few customers after receiving the delivery are raising tickets for refund in the name of burnt pizzas. Even though customers received a good pizza but still few customers are trying to cheat vendors. To overcome this issue, Franchise has come up with an idea to integrate a pizza detection model in their application where customers can upload images for the burnt pizzas delivered. For example, if I have received a burnt pizza then I can upload a couple of images of the pizza to the application and it will classify the pizza as burnt or good in order to process my refund ticket.

Goal: You are hired as Deep Learning Engineer by a famous pizza franchise. You are asked to build a model where it accepts the images of pizza and detects as burnt pizza or good pizza.

Constraints: You should be using only ANN and shouldn't be using CNN or any other rule based model to generate results.

Data Description: Data is in the form of images collected from multiple sources of the internet.

Provided Files:

Train set: Train set is divided into burnt pizza and good pizza categories. While training the model you can label images of good pizza as 1 and burnt pizza as 0.

Test: Test set contains mixed images of both burnt pizzas and good pizzas.

Instructions:

Train set should be used to feed the model.

Test set should be used to predict labels for test data.

Evaluation Criteria: The evaluation metric for this problem statement is the Accuracy score where each image label is matched with the actual image label.