# Main Case Study Set-1 (Evaluation)

Total points 15/15

Email *

futureemperor2k12@gmail.com

0 of 0 points

Your Name *

Adithya D

Registered email-id with Intellipaat *

futureemperor2k12@gmail.com

Data Description 15 of 15 points

Bank Marketing Case-Study:

Relevant Information about the data:
The data is related with direct marketing campaigns of a 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 (or not) subscribed.

Data set link:
https://drive.google.com/drive/folders/1KnOqzvd2qGivJHZLNJWB9m8RFrih0J1T

Objective:
The objective is to predict if the client will subscribe (yes/no) to a term deposit, by building classification model using Machine Learning algorithms.

| Parameter | Description |
|---|---|
| age | Age of the clients (numeric) |
| job | The job type of the clients (categorical) |
| marital | Marital status of clients (categorical) |
| education | Education level of clients (categorical) |
| default | has credit in default? (categorical: "yes", "no") |
| balance | average yearly balance, in euros (numeric) |
| housing | has housing loan? (categorical: "yes", "no") |
| loan | has personal loan? (categorical: "yes", "no") |
| contact | contact communication type (categorical: "unknown", "telephone", "cellular") |
| day | last contact day of the month (numeric) |
| month | last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec") |
| duration | last contact duration, in seconds (numeric) |
| campaign | number of contacts performed during this campaign and for this client (numeric, includes last contact) |

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)
previous number of contacts performed before this campaign and for this client (numeric)
poutcome outcome of the previous marketing campaign (categorical: "unknown", "other", "failure",
"success")
deposit has the client subscribed a term deposit? (categorical: "yes", "no")

✓ 1.The total number of missing values in the dataframe are: * 1/1

- a.8
- b.10
- c.0 ✓
- d.50

✓ 2.The total number of duplicated values in the dataframe area. * 1/1

- a.0 ✓
- b.50
- c.15
- d.21

✓ 3.What is the shape of the data after dropping the feature “Unnamed: 0”, missing values and duplicated values? *1/1

- a.(5000,17)
- b.(5581,17) ✓
- c.(5580,18)
- d.(4581,18)

✓ 4.What is the average age of the clients those who have subscribed to deposit? *1/1

- a.39
- b.49
- c.32
- d.41 ✓

✓ 5.What is the maximum number of contacts performed during the campaign for the clients who have not subscribed to deposit? *1/1

- a.63 ✓
- b.32
- c.10
- d.5

✓ 6.What is the difference between the maximum balance (in euros) for the clients who have subscribed to deposit and for the clients who have not subscribed to the deposit? *1/1

- a.1747
- b.1514
- c.24373 ✓
- d.75054

✓ 7.What is the count of unique job levels in the data and find out how many clients are in the management level? *1/1

- a.10 & 1318
- b.12 & 3134
- c.13 & 2000
- d.12 & 1318 ✓

✓ 8.What is the percentage split of the categories in the column “deposit”? * 1/1

- (●) a.Yes- 47% & No- 53% ✓
- ( ) b.Yes - 40% & No- 60%
- ( ) c.Yes – 50% & No- 50%
- ( ) d.Yes – 30% & No- 70%

✓ 9.Generate a scatter plot of “age” vs “balance” and choose which of the following interpretation is correct? *1/1

- ( ) a.Across all ages, most of client’s average yearly balance is greater than 20000 euros
- ( ) b.As the age increases the bank balance of client increase
- (●) c.Across all ages, most of the client’s average yearly balance is less than 20000 euros ✓
- ( ) d.As the age decreases the bank balance of client decrease

✓ 10.How many unemployed clients have subscribed to deposit? * 1/1

- ( ) a.100
- ( ) b.85
- ( ) c.78
- (●) d.92 ✓

✓ 11.The command used to convert the categorical variables to indicator variables is: *1/1

- ( ) a.pandas.get_indicator(data)
- (●) b.pandas.get_dummies(data) ✓
- ( ) c.pandas.reshape(data)
- ( ) d.pandas.reshape.module(data)

✓ 12.The code below is used to get a list of unique column names excluding 'deposit' column. Fill in the blanks in the order of the blanks (1st blank, 2nd blank) with appropriate data types as given in the options *1/1

features= ____ (____(data.columns)-set(['deposit']))

- a.set and list
- b.set and set
- c.list and list
- d.list and set ✓

✓ 13.The command to predict the logistic regression model 'model' on test dataset (test) is: *1/1

- a.model.fit(test)
- b.model.prediction(test)
- c.model.predict(test) ✓
- d.model.LogisticRegression(test)

Please read below:

Starting afresh load the given dataset and prepare the dataset by following the steps given below:

- Ensure the datatypes of the columns are appropriate
- Drop the variable "Unnamed: 0", missing values, duplicated values
- Convert the categorical variables into dummy variables
- Split the data into train (70%), test (30%), and set the random state as 0
- Build a logistic regression model using the train dataset. Using the model that has been built, answer the following question.

✓ 14.What is the value of accuracy of the model on the test dataset? (Choose the appropriate range) *1/1

- a.20% to 60 %
- b.61% to 70%
- c.71% to 90% ✓
- d.91% to 100%

Please read below:

Starting afresh load the given dataset and prepare the dataset by following the steps given below:

- Ensure the datatypes of the columns are appropriate
- Drop the variable "Unnamed: 0", missing values, duplicated values
- Convert the categorical variables into dummy variables
- Split the data into the train (70%), test (30%), and set the random state as 0
- Build a KNN model using the train dataset and. Using the model that has been built, answer the following question

✓ 15.What is the value of accuracy of the model on the test dataset? (Choose the appropriate range) 1/1

- a.20% to 40%
- b.41% to 60%
- c.61% to 80% ✓
- d.81% to 100%

This form was created inside of Intellipaat.

Google Forms