**PRCP-1000-ProtugeseBank**

**Problem Statement**

**Task 1:** Prepare a complete data analysis report on the given data.

**Task 2:** Create a predictive model which will help the bank marketing team to know which customer will buy the product.

**Task3:** Suggestions to the Bank market team to make customers buy the product.

**Dataset Link:**

This dataset is about the direct phone call marketing campaigns, which aim to promote term deposits among existing customers, by a Portuguese banking institution from May 2008 to November 2010.

**Link:** <https://d3ilbtxij3aepc.cloudfront.net/projects/CDS-Capstone-Projects/PRCP-1000-ProtugeseBank.zip>

**Note: Kindly use Data>bank-additional>bank-additional-full.csv for this use case.**

**Attribute Information:**

**Input variables:**

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 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 – **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')

**Model Comparison Report**

Create a report stating the performance of multiple models on this data and suggest the best model for production.

**Report on Challenges faced**

Create a report which should include challenges you faced on data and what technique used with proper reason.

Note: All above tasks have to be done on a single Jupyter notebook and share the same for the final submission.