Machine Learning - Mini Project

(Task 2)

Group Members:

Umang Kirit Lodaya - 60009200032

Jay Bhanushali - 60009200047

Data Description

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 with the same client was required, to assess if the product (Bank Term Deposit) would be subscribed (1) or not subscribed (0).

Train Set

The train set contains the data to be used for model building. It has the true labels for whether the customer subscribed for term deposit (1) or not (0)

Test Set

Set of calls for which the prediction needs to be done regarding the subscription status of the customer for term deposit post-campaign.

Attribute Information:

Input variables (Feature):

Numeric:

1. Id

2. Age

3. Balance:

   - The bank balance of the customer.

4. Day_of_Month

   - The day of the month on which communication took place

5. Duration:

   - States the duration of the call (in sec) regarding the campaign. This is an important feature as this attribute highly affects the target label (e.g., if duration=0 then y='no'). Thus, this input will be included for benchmark purposes and will be discarded if the intention is to have a realistic predictive model.

6. Num_contacts_in_campaign:

   - Number of contacts performed during the campaign with the client for subscription

7. Days_since_prev_campaign_contact:

   - Gap (in number of days) between two contacts

8. Num_contacts_prev_campaign:

   - Number of contacts performed for promoting the campaign beforehand

Categorical:

1. Job_type:

   {'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown'}

2. Marital: States the marital status of customer

   {'divorced', 'married', 'single', 'unknown'}

3. Education :

   {'high_school', 'illiterate', 'professional_course', 'unknown', 'university_degree'}

4. Default: Is the customer a defaulter or not?

   {'yes', 'no', 'unknown'}

5. Housing_loan: Has a housing loan or not?

   {'yes', 'no', 'unknown'}

6. Personal_loan: Has a personal loan or not?

   {'yes', 'no', 'unknown'}

7. Communication_type: Mode of communication during the campaign

   {'cellular', 'telephone', 'unknown'}

8. Month: Month in which contact with the customer took place for the campaign

   {'jan', 'feb', … , 'dec'}

9. Prev_campaign_outcome:

   The outcome of promotional contact with the client beforehand for attending/interest in the campaign {'failure', 'unknown', 'success', 'other'}

Output variable (Label):

Term_deposit_subscribed:

Categorical target label to be predicted {'0, '1'}