Title: Bank Marketing Based Deposit Prediction

First name	Last Name	IIT Email
Aarish Salam	Memon	amemon2@hawk.iit.edu

Category 1 project: data mining/machine learning task

Category 2 project: data mining/machine learning application

Use underline above to indicate your project category

Important Notes:

- Each group must submit ONLY one copy by a single team member!

1. Introduction

I've picked a dataset from Kaggle Titled "Bank marketing campaigns dataset | Opening Deposit" which has 21 columns having different data types. We would be performing supervised learning task to get the prediction of the label of unseen data. The dataset has 41188 records, also has a very valuable columns which almost all of them would be used for the prediction models.

2. Data Sets (or Applications if it is category 2)

As already mentioned, the dataset is picked from Kaggle Titled "Bank marketing campaigns dataset | Opening Deposit".

It is a dataset that describing Portugal bank marketing campaigns results. Conducted campaigns were based mostly on direct phone calls, offering bank client to place a term deposit.
If after all marking affords client had agreed to place deposit - target variable marked 'yes', otherwise 'no'

These are the features/columns in the dataset

1. Age (Numeric) – Age of the customer
2. Job - Type of job (Categorical)
3. Marital - Marital status (Categorical)
4. Education (Categorical) – States the education level
5. Default: has credit in default? (Categorical)
6. Housing: has housing loan? (Categorical)
7. Loan: has personal loan? (Categorical)

8. Contact: contact communication type (Categorical)
9. Month: Last contact month of year (Categorical)
10. Dayofweek: Last contact day of the week (Categorical)
11. Duration: Last contact duration, in seconds (numeric).
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)
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")

3. Research Problems

There has been a revenue decline for the Portuguese bank and they would like to know what actions to take. After investigation, we found out that the root cause is that their clients are not depositing as frequently as before. Knowing that term deposits allow banks to hold onto a deposit for a specific amount of time, so banks can invest in higher gain financial products to make a profit. In addition, banks also hold better chance to persuade term deposit clients into buying other products such as funds or insurance to further increase their revenues. As a result, the Portuguese bank would like to identify existing clients that have higher chance to subscribe for a term deposit and focus marketing effort on such clients

4. Evaluations

We will be using multiple evaluation methods, like creating a confusion matrix consisting of accuracy, precision, recall and f1-score and ROC Curve.

5. Expected Outcomes

Our goal is to create a most accurate prediction model to get identify if the marketing campaign worked in the favor or company or not, also once the prediction model is created with higher accuracy same model is used for different campaigns and different companies using the same technique. Though there are many columns we'll work on feature selection and reduction so we can check if there are any features to be removed.