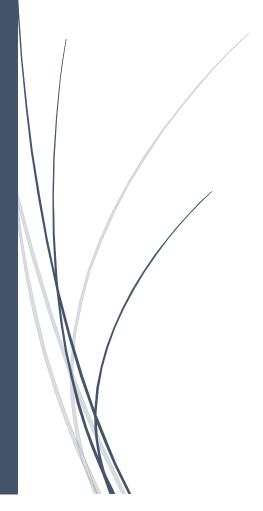# Bank Marketing Data

Methodology

Prepared By
KEERTHI

OBJECTIVE:

Term deposit is a fixed-term investment that includes the deposit of money into an account at a financial institution. Term deposit investments usually carry short-term maturities ranging from one month to a few years and will have varying levels of required minimum deposits.

The data is related with direct marketing campaigns of a Portuguese 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.

The objective is to train a ML model that returns the probability of a customer to accept the offered product. Here, the classification goal is to predict if the client will subscribe a term deposit (variable y) i.e. 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.

Overall, using this case study will try to study for determining the success of Bank Telemarketing.

Basic Information:

UCI Dataset URL: https://archive.ics.uci.edu/ml/datasets/Bank+Marketing

Tool: Python SAS and Jupyter notebook (Data Cleansing)

Domain: Banking and Marketing

Area: Data Science and Predictive Modelling

APPROACH

This research aims at improving the performance of predicting the willingness of bank clients to apply for a term deposit in highly imbalanced datasets

1. *Define Problem Statement and Import Data*
   a. *Import Dataset:*
2. *Data Cleaning & Exploratory Data Analysis*
   a. *2.1.1 Data Description*
   b. *2.1.2 Checking Values for Categorical attributes & Numerical Values*
      i. *2.1.2.1 Exploring Categorical Values*
      ii. *2.1.2.2 Exploring Numerical Values*
      iii. *2.1.2.3 Correlation Matrix*
   c. *2.1.3 Data Cleansing Activity*
      i. *2.1.3.1 Checking Missing values*
      ii. *2.1.3.2 Checking duplicate values*
      iii. *2.1.3.3 Encoding Categorical values to numerical codes.*
      iv. *2.1.3.4 Dropping less meaningful columns*
      v. *2.1.3.5 Removing Outliers*
   d. *2.1.4 Exploratory Data Analysis*
   e. *2.1.5 Creating Dummy Variables and splitting data in Training and Testing set*
3. *Data Modelling & Evaluation*
   a. *3.1.1 Define Model data including dummy variable, X, Y, Training, Test Dataset*
   b. *3.2 Function to build and asses the model which can be used generally for all the Classification model*
   c. *3.3 Building Model*
      i. *3.3.1 Logistic Regression*
      ii. *3.3.4 Decision trees*
      iii. *3.3.5 Random Forest*
   d. *3.4 Model Evaluation*
      i. *ROC Curve for all models*
      ii. *Confusion matrix*
4. *Exporting Final Cleaned data, Training data, testing data and Prediction generated by the model on Testing Data*
5. *CONCLUSION (INSIGHTS):*
6. *Appendix/Reference*