IST 707 – DATA ANALYTICS (M003)
Prediction for client term deposit subscription

Team 11: Adheesh Phadnis, Lakshya Gupta, Vidushi Mishra

Final Project Proposal

Context

The core business of a financial institution can be broadly classified as lending and borrowing. Lending generates revenue to the bank in the form of interest from customers with some level of default risk involved. A low risk strategy for the banks is to attract public savings into the bank. To achieve this goal, one of the popular ways for the banks is telemarketing. It is an interactive technique of direct marketing via the phone which is widely used by banks to sell long-term deposits. The bank telemarketing data used here is related with direct marketing campaigns of a Portuguese bank institution.

Problem Statement

The aim of this study is to predict whether a client is going to subscribe a long-term deposit through telemarketing strategy. Further, this study will analyze the characteristics of the clients who are predicted to invest in the long-term deposits. The bank can then utilize this information to allocate resources to focused customers and thus increase their revenue.

Dataset Description:

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 (bank term deposit) would be ('yes') or not ('no') subscribed.
There are two datasets:
1) bank-additional-full.csv- It has 41188 instances and 20 features.
2) bank-additional.csv – It contains 10% of the examples.

The above datasets contain the following features:

- Bank client data such as age, type of job, marital status, education, has credit in default, balance, housing, loan.
- Attributes related with the last contact of the current campaign such as contact communication type, contact day of the months, last contact month of year, last contact duration.
- Social and Economic context attributes such as employment variation rate, consumer price index, no. of employees.

- Other attributes such as number of contacts performed during this campaign and for this client, number of days that passed by after the client was last contacted from a previous campaign, number of contacts performed before this campaign and for this client.

Link

https://archive.ics.uci.edu/ml/datasets/bank+marketing#

Approach

We have planned to leverage CRISP-DM methodology (Cross-industry standard process for data mining). It is the most widely used analytics model. CRISP-DM breaks the process of data mining into major phases:

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling (classification techniques)
5. Evaluation
6. Deployment

Data Preprocessing

- Data standardization and normalization
- Handling missing values and outliers
- Feature engineering
- Using techniques like SMOTE oversampling to deal with imbalanced data set

Data mining algorithms

- Logistic regression
- K-Nearest Neighbors(KNN)
- Support Vector Machine (SVM)
- Ensemble learning methods
    - Random forest classifier
    - Light Gradient Boosting Machine (LGBM)
    - XGBoost

Evaluation Metrics

We will be using evaluation metrics such as Area Under Curve, Precision, Recall, Balanced Accuracy and f1-score to evaluate the model.