

# TERM DEPOSIT LEAD PREDICTION

## EXLEQ 22

Team : Champions NSUT  
Lakshay Kapoor 4572  
Vikas 4555

# Marketing Strategy



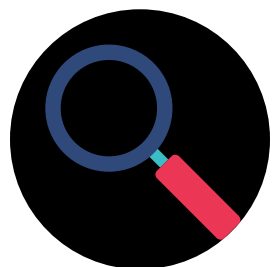
## Background

EXL's client ABC is a financial institution based out of India. They offer a range of banking services like Savings accounts, Current accounts, Investment products, Credit products etc.



## Problem Statement

In 2018, the firm launched a new investment product, "Term Deposit", which has observed a rapid growth since its inception. The increase in term deposit accounts has provided the firm with additional funds to create other offerings. Hence, based on this performance, it further wants to increase its term deposit accounts amongst existing customers. In order to do so, the firm plans to run a campaign and wants EXL to identify a list of customers who could be contacted via telecommunication channels to open a term deposit account with ABC. Since the campaign will be run for a short period of time, and owing to budget constraints, the firm wants a list of only 1000 customers from EXL for the target list.



## Important Factor

Using various statistical and analytical method such as correlation statistics and variable transformation following 8 factor were considered for our analysis:

- **customer\_age**
- **job\_type**
- **balance**
- **housing\_loan**
- **day\_of\_month**
- **month**
- **last\_contact\_duration**
- **prev\_campaign\_outcome**

# Process Overview

## Term Deposit Customer Analysis

### **Data profiling**

Exploring data with respect to the number of features, data points, null values, and their statistical description

### **Exploratory Data Analysis**

Graphical and statistical analysis was done on training data to find patterns and insights about data

### **Feature Engineering**

- Null value handling
- Categorical handling
- Imbalance handling
- Feature selection
- Feature Combining

### **Merge result of various models**

The outcome of various models were combined to provide list of top 1000 term depositor.

### **Evaluate different models**

Various models were evaluated based on Accuracy, Recall, Precision and F1 Score.

# Data Preprocessing

## 01 Data Understanding

We used various statistical and graphical techniques to understand our data in context with the provided data dictionary to provide unbiased analysis.

## 02 Data Preparation

The provided training data(Historical) had a large number of null values and outliers which were filled using suitable imputation methods. The Categorical features were mapped to numerical counterparts by ordinal number encoding.

## 03 Data Balancing

The Provided training data was highly skewed with respect to the decision variable. Hence we handled it using both under and over sampling.

# Model Selection

We evaluated different models and used the most accurate to extract results.

**Original data with selected features based on info gain**

Random Forest with hyper parameter tuning

**745**

**Original data with selected combined features based on info gain**

Union of AdaBoost, XGBoost, and Decision tree

**343**

**Over Sampled Data using Random Sampling**

Random Forest with hyper parameter tuning

**677**

**Under Sampled Data using Near Miss Sampling**

Adaboost with hyper parameter tuning and applying other model on the predicted result

**548**

# Final Outcome

## Selecting Top 1000 enties

- The original data set with top 8 feature based on information gain was used with **Random forest** to provide a list of **745** potential term depositor.
- Common feature such as housing loan and personal loan were combined to make a new feature "having loan". This updated dataset was used to derive result using various model to provide list of **343** potential term depositor.
- Data was oversampled using **random sampling method** and subjected to **Random forest** to yield a list of **677** potential term depositor.
- These three results were combined by taking a union to provide a exhaustive list of **947 customer**.
- The remaining 53 entries were aggregated by applying **Adaboost** on under sampled data using **Near-Miss** method to finalise the result.