**BANK MARKETING EFFECTIVENESS PREDICTION**

**Jayanth V, Gowthaam Kumarasamy,**

**Data science trainees,**

**Almabetter, Bangalore.**

**Abstract:**

The Portuguese banking institution is effectively conducting marketing campaigns in order to make people subscribe to their term deposit through direct phone calls. We are provided with various independent features which is the reason for the results of the marketing campaigns in the dataset.

Our experiment can help understand what could be the reason for the successful and unsuccessful marketing campaigns influenced by the different features by data analysis, feature selection and engineering, and implementing machine learning models, by taking into account previous trends to find the correct prediction (classification).

**Keywords: Machine learning, classification, marketing, feature engineering, data cleaning, data exploration.**

**Problem Statement:**

### The data is related with direct marketing campaigns (phone calls) 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. The classification goal is to predict if the client will subscribe a term deposit (variable y).

You are provided with bank marketing data of 45000+ records. The task is to predict whether a client will subscribe the term deposit for the test set.

### **Bank Client data:**

### age (numeric)

### job: type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')

### marital: marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)

### education (categorical: 'basic.4y','basic.6y','basic.9y','high. school','illiterate','professional. Course','university. Degree','unknown')

### default: has credit in default? (Categorical: 'no','yes','unknown')

### housing: has housing loan? (Categorical: 'no','yes','unknown')

### loan: has personal loan? (Categorical: 'no','yes','unknown')

### **Related with the last contact of the current campaign:**

### contact: contact communication type (categorical: 'cellular','telephone')

### month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')

### day\_of\_week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')

### duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

### **Other attributes:**

### campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

### 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)

### previous: number of contacts performed before this campaign and for this client (numeric)

### poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success').

### **Output variable (desired target):**

### y - has the client subscribed a term deposit? (Binary: 'yes','no').

**2. Introduction:**

The banking institutions conduct marketing campaigns for their new policies and schemes regularly for their customers (people). Here a Portuguese banking institution conducting direct marketing campaigns through the phone calls to their clients. Often the bank approaches to more than one contact for a client. This is to ensure whether the client will subscribe to their term deposit or not.

Our goal here is to build a predictive model (classification) which could help the Portuguese bank to predict whether their customer will subscribe term deposit or not in the future proactively.

**3. Types of term deposits:**

* Fixed deposit
* Recurring deposit

A Fixed Deposit is kept for a longer period and hence it earns a higher rate of interest. A Recurring Deposit takes a defined sum and invests it every defined period. This means each instalment earns interest for a lesser period than the previous instalment. The interest on a Fixed Deposit for the same maturity is more than that on a Recurring Deposit.

However, a Recurring Deposit is a convenient way of investment for people who have a fixed investment amount per month. As such, the investment type depends on the goals and funds available.

**4. Major Factors influencing subscription:**

The reasons for the clients to subscribe for this bank’s term deposit are:

* Age
* Job
* Loans
* Contact
* Previous outcome

**5. How Term deposit works and factors influences:**

Financial institutions e.g., Banks generate their revenue through lending and borrowing. Lending generates profits in for of interest from customers but some level of risk is involved, that is why machine learning algorithms come in handy in predicting clients who are eligible for loans. Another form that generates revenue for financial institutions is borrowing or attracting public' savings into the bank which is a bit less risky than lending.

Borrowing works this: the bank invests the client’s long-term deposits into other sectors which brings better returns, where some is paid to the customers. However when a client does fixed-term deposit, the company gets good returns than savings account as the customer or the client is deprived off the rights to access the money prior to the maturity unless the client is ready to compensate the bank.

Due to this reason, there is a stiff competition between banks to convince clients to do term deposits in their banks, and due to this marketing campaigns, a huge amount of money is spent by the banks in reaching out to clients, prospective subscribers and non-prospective ones since the bank doesn’t know who is and who is not. With advancement in data science and machine learning and availability of data, the banks are adapting to data-driven decisions and this will help in reducing the cost of marketing thus increasing the revenue of the bank.

**6. Steps Involved:**

* **Data Overview and Null values Treatment:**

Our dataset contains more than 45000 records and it has almost 17 features in which there are 16 independent features and 1 dependent feature which is the client’s answer (yes/no). The dataset has zero null values and it also has zero duplicated values, thus there is no need for the null values treatment.

* **Exploratory Data Analysis:**

After importing the dataset, we performed univariate and multivariate analysis