DATA MINING FINAL PROJECT

Prediction of the Success of Bank Telemarketing

Business Question:

Marketing team in a bank is one of the core teams for the upliftment of the organization. They will create a lot of impact on the development of the bank and the first step of the growth in any bank will be started by the marketing team.

Marketing is the process of presenting a product or service to the market, promoting it, and encouraging consumers to make purchases. The key to a successful business nowadays is effective marketing, and marketing profoundly affects our day-to-day lives.

The value of marketing is acknowledged to the point that it is claimed that marketing proficiency is a prerequisite for financial success. If there is insufficient demand for the company's goods and services, finance, operations, accounting, and other business functions will be irrelevant.

Banks should also constantly innovate and offer newer, more compatible services and products to their serving marketing team.

There are a lot of issues that were faced by the marketing team in a bank. Some of them are Time, Competition and speed.

Time: Time management is without a doubt one of the most urgent issues. On most days, it could seem as though there simply aren't enough hours in the day to finish the never-ending lists of things to accomplish. Bank marketing managers are currently responsible for more tasks than ever before, therefore setting priorities is essential to moving forward.

Competition: In this brave new world, banks not only compete fiercely with one another but also with nascent fintech companies. In order to remain competitive and relevant, it is crucial to grasp what their advantage is. A strategy is required by the banks to generate a greater revenue. The model we will build will help the banks to target the necessary customers and move ahead in the competition.

Speed: The speed at which a bank marketer must make decisions is fundamentally different than it was a few years ago. They need to change their decisions according to the results they get in a financial year and those decisions need to be relatively fast and optimal. The outcomes majorly depend on the decisions that were made. So, this model helps to build strategies in a fast-paced way and provide the quintessential results.

The data refers to direct marketing initiatives carried out by a bank in Portugal, the marketing initiatives were centered on phone calls, it was occasionally essential to have many conversations with the same client to ascertain if the product bank term deposit would be subscribed (yes) or not (no).

Financial products and services like debit cards, checking and savings accounts, etc. are provided to customers by the bank. To increase overall revenue, the bank employs a variety of marketing strategies for its financial products, such as credit cards, term deposits, loans, etc. Current customers of the bank are the intended market for these initiatives. The marketing efforts must be economical if the bank is to increase total sales and profits.

The bank conducted a telemarketing campaign for one of its products, "Term Deposits," to develop long-term ties with its current clients. The dataset contains information on all customers who were contacted to open term deposit accounts within the specified time period.

Our goal is to put more effort into raising the rate of favorable responses. The results we will extract from our model will help to choose and target the customers accordingly, which will help them to overcome the issues.
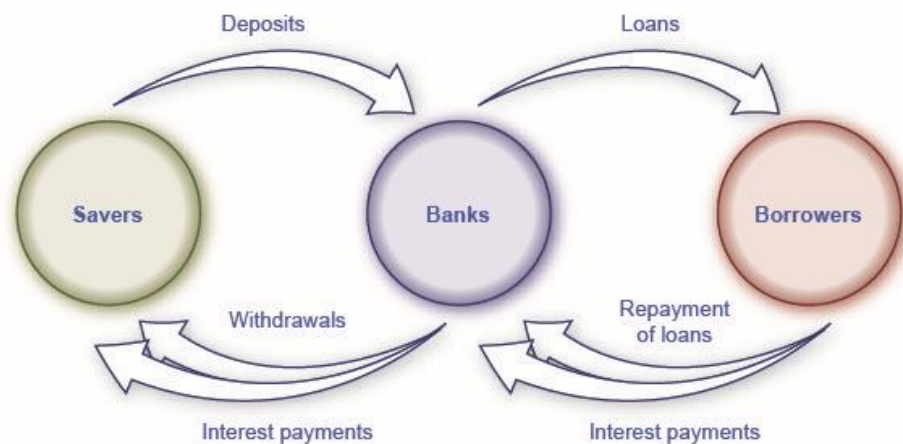
Motivation:

Term deposit funds are held by the bank for a set period and are used by banks to generate income by investing in marketing campaigns to expand their customer base for credit cards, term deposits, loans, and other financial products. Banks need to attract customers for term deposit plans. If the client base declines, the bank's reserves will ultimately run dry, the bank won't be able to make money, it will go bankrupt, the balance sheet will collapse, and there won't be any cash flow. The banking business relies heavily on cash flow. Hence, it is a consideration for the problem to be solved.
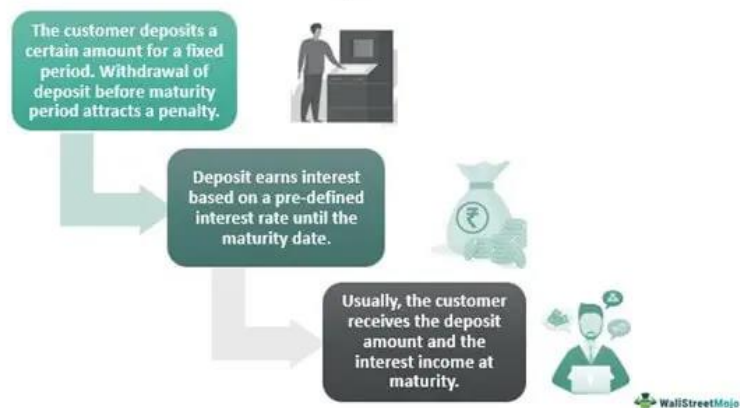
Here, we're looking at a prediction that would be advantageous to several parties, including the bank, the marketing team, and the clients. Knowing who to target in this case will reduce marketing expenses for the company, its employees, and the marketing firm. Additionally, customers who don't require that product won't see any advertisements relating to it.

The reason for building a model here is to predict how likely a consumer would be to respond to this new offer based on prior data on customers who have been targeted with comparable items. This will help to provide a clever technique for creating a solid plan that would accurately categorize the clients.

Because marketing strategies rely significantly on previous campaign results, integrating historical campaign data in modelling for future efforts is critical. Existing liabilities that can be used to evaluate performance include personal loans, mortgages, and loan defaults. A machine learning model is needed here to target a client who has a history of bad creditworthiness and investing the marketing efforts on this may not be a wise idea. In this way, a Machine learning model will be helpful for the reliable predictions.

DEPOSITORS OFTEN PREFER TERM DEPOSITS BECAUSE THEY PAY MORE INTEREST THAN TRADITIONAL SAVINGS ACCOUNTS, YET ARE JUST AS SAFE.

All bank marketing campaigns are dependent on customer's huge electronic data. The size of these data sources is impossible for a human analyst to come up with interesting information that will help in the decision-making process. The model we build will help in achieving this feat and help to target the potential customers for the bank. The purpose is increasing the campaign effectiveness by identifying the main characteristics that affect the success of the scheme. With the higher scoring of these metrics, we will be able to judge the success of these models in predicting the best campaign contact with the clients for subscribing deposit. The aim of the marketing campaign was to get customers to subscribe to a bank term deposit product. Whether they did this or not is variable 'y' in the data set. The bank in question is considering how to optimize this campaign in future.

A term deposit is typically a bank asset from which the bank generates money through various services. So, targeting all clients in the campaign will take a lot of time, and contacting a customer numerous times when the customer is not interested in investing in term deposits would harm the bank, and the consumer may switch banks. Furthermore, contacting the right consumer who is willing to invest in term deposits would contribute additional assets to the bank and improve bank services by bringing in more term deposits. As a result, this model will assist us in predicting the correct customer to target, as well as increasing the potential of the bank's campaign, in which the campaign will invest more time on the correct customers while reducing unnecessary advertisement on customers who are not willing to invest in term deposits. It will also allow truly interested targeting customers, which will have a great influence on bank assets and help create significant income.

About Dataset:

We will analyze the data collection with 45800 observations with 16 columns. The following features of the data allow us to predict whether a customer will make a term deposit or not:

Column Name	Data Type
Age	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')
Defaullt	has credit in default? (categorical: 'no','yes','unknown') After 30 days of late payments, this occurs. Your credit card is severely past due if six consecutive months go by during which you fail to make at least the minimum payment required.
Balance	Amount in his account
Housing	has housing loan? (categorical: 'no','yes','unknown')
Loan	has personal loan? (categorical: 'no','yes','unknown')
Contact	contact communication type (categorical: 'cellular','telephone')
day	last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
month	last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
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)
y	the client subscribed a term deposit? (binary: 'yes', 'no')

Past Analysis on the Dataset:

The author of this dataset has already taken this challenge and proposed a data mining (DM) approach to predict the success of telemarketing calls for selling bank long-term deposits.

- He has applied a semi-automatic feature selection scheme to reduce the number of features for the ease of computation.

- Later, he had performed model evaluation on the data by analysing it with neural network (NN) and support vector machine (SVM).

- He found out that the results to be 79% accurate by doing all the computations.

Our Analysis to Conduct on the Dataset:

	Predicted No	**Predicted yes**
Actual No	Customer will not subscribe to the product as predicted. **(True Negative)**	- Actual - customer will not subscribe the product. - Predicted- customer will subscribe the product. Here, prediction is wrong. **(False Positive)**
Actual Yes	- Actual – customer will subscribe the product. - Predicted- customer will not subscribe the product Here, prediction went wrong **(False Negative)**	Customer will subscribe to the product as predicted. **(True Positive)**

False Positive: The actual outcome was that the client did not subscribe to the product, but the model predicted that the customer would. The concern here is that the marketing team.

will target consumers that do not need to be targeted. The campaign team wastes time and effort, while the client is bombarded with advertisements that are unnecessary.

False Negative: The actual outcome was that the client subscribed to the product, but the model predicted that the customer would not subscribe to the product. As a result, the bank will suffer a significant loss because they are missing one of their investors for the product.

Here, FP is preferable over FN since we have less loss with false positive than with false negatives, thus we will concentrate on 'Recall'.

Based on the chosen dataset,

- We are dropping columns that are unwanted for the analysis.
- Perform data cleaning, converting column datatypes if necessary, encoding predictor and target variables if necessary and making the dataset balanced.
- Then, we are performing binary classifiers (classification models) like KNN model, Decision Tree (pruning), Adaboost, Gradient Boost, Random Forest, XGBoost, Logistic Regression and we are predicting which model best fits the dataset with a great recall score.