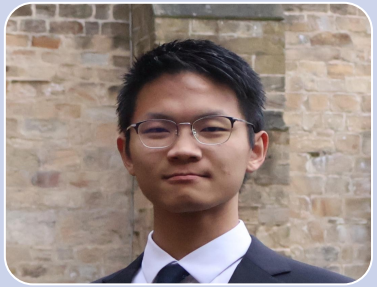
BANKBYTES

Mark Kimutai Kitur
kiturmark@gmail.com
Kenya
Dedan Kimathi university
Of technology
Data Science



Yue hu
hy1550278246@gmail.com
The United Kingdom
Durham University
Data Science

Rakshith Nagaraju
rakshith.nagaraj6@gmail.com
The United Kingdom
University of Liverpool
Data Science


Data Understanding

Pillars of the data

- The Bank dataset columns were divided into four main pillars .
- These pillars made it easier to understand factors that might greatly influence the outcome of the marketing campaign.
- The pillars are:

- Personal Information
- Contact Information
- Campaign Information
- Economic Indicators

1. Personal Information

- These attributes provide demographic and financial details about the clients which are essential for understanding their profiles and predicting their profiles.

a. Age: Different age groups show varying levels of interest in the product.

b. Job: The type of job indicates client's financial stability thus influencing their decision.

c. Marital status: This affects financial responsibilities and priorities.

d. Education: Level of education often correlates with income and financial literacy.

e. Loan: Personal loans reflect financial behavior and risk.

f. Default: This attribute shows if a client has credit. This illustrates creditworthiness and financial reliability.

g. Housing: This attribute shows if a client has housing loan which indicates financial commitments and ability to invest.

2. Contact Information

- This characteristics provide details about the method and timing of the last contact with the client, which can influence the client's response.

a) Contact: This describes means of communication, which their effectiveness greatly vary.

b) Month: The time of the year can affect client responses due to seasonal factors and financial cycles.

c) Day of the week: client availability and mood might differ on different days of the week.

d) Duration: This is duration of call in seconds. Longer call might indicate more interest from the client, but this is known only after the call.

3. Campaign Information

- These attributes track the number of contacts made and the outcomes of previous interactions, providing insights into the client's engagements history.

a) Campaign: Number of contacts during this campaign. Frequent contacts might indicate persistence but could also lead to fatigue.

b) Pdays: Days since the client was last contacted (999 means not contacted before).

c) Previous: This column shows number of contacts before this campaign. Past engagement level could affect the current campaign's success.

d) Poutcome: Outcome of the previous campaign. Previous campaign success can be a strong predictor of the current campaign success.

4. Economic Indicators

- These attributes provide contextual economic data which can affect the overall success of the marketing campaign.

a) Emp.var.rate: This column shows employment variation rate (quarterly indicator). This reflects economic stability and job market conditions.

b) Cons.price.idx: Consumer price index (monthly indicator). This illustrates inflation and cost of living, affecting disposable income.

c) Cons.conf.idx: Consumer confidence index is really important as it reflects consumer optimism or pessimism about the economy, influencing spending behavior.

d) Euribor3m: This column provides information on Euribor 3-month rate (daily indicator). This affects loan interest rates, influencing financial decisions.

e) Nr.employed: Number of employees (quarterly indicator). This reflects labour market size and the economic health.

Problem Description

Objective

The main objective of this project is to develop a predictive model for the bank to identify potential customers who are likely to buy their term deposit product. The model aims to save resources and time by focusing marketing efforts on customers with a higher likelihood of purchasing the product, thus increasing the efficiency and success rate of telemarketing campaigns.

Specific Problems

(1) Predictive Factors Identification: Identify the key factors that influence a customer's likelihood of purchasing a term deposit. These factors could include demographic information, financial status, previous interactions with the bank, and past purchase behavior.

(2) Model Development: Develop and validate a machine learning model that can accurately predict the probability of a customer buying a term deposit. The model should be able to handle the complexity and variety of data available, ensuring high predictive performance.

(3) Resource Optimization: Utilize the predictive model to prioritize and focus marketing efforts on high-potential customers. This will help the bank save time and resources while improving the overall success rate of their telemarketing campaigns.

Data Resources

To develop a predictive model for identifying potential customers likely to buy the bank's term deposit product, a variety of data types have been utilized. The data is collected by the bank.

The data is expected to be in structured formats, which is in CSV files.

Type of Data

There are two type of data:

a. Categorical data:

   job, marital, education, default, housing, loan, contact, month, day_of_week, poutcome

b. Numeric data:

   Age, duration, campaign, pdays, previous, emp_var_rate, cons_price_idx, cons_conf_idx, euribor3m, nr_employed

Problems in the data & Proposed Solutions.

Null (NA) values.

Null value is when there is no record of data in a particular instance in the database.

It represents the absence of a value.

The key to good data analysis is having complete data.