

Banking Dataset

Term deposits serve as a significant revenue stream for banks, representing cash investments held within financial institutions. These investments involve committing funds for a predetermined period, during which they accrue interest at an agreed-upon rate. To promote term deposits, banks employ various outreach strategies including email marketing, advertisements, telephonic marketing, and digital marketing.

Despite the advent of digital channels, telephonic marketing campaigns persist as one of the most effective means of engaging customers. However, they necessitate substantial investment due to the requirement of large call centers to execute these campaigns. Therefore, it becomes essential to pre-identify potential customers likely to convert, enabling targeted outreach efforts via phone calls.

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

Content

The data is related to the 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 by the customer or not. The data folder contains two datasets:-

Banking_data.csv: 45,211 rows and 18 columns ordered by date (from May 2008 to November 2010)

Detailed Column Descriptions:

age: This column represents the age of the bank client. It's a numeric variable indicating the age in years.

job: This column indicates the type of job the client has. It's a categorical variable with options such as "admin.", "unknown", "unemployed", "management", etc.

marital: This column represents the marital status of the client. It's a categorical variable with options such as "married", "divorced", or "single".

education: This column indicates the level of education of the client. It's a categorical variable with options such as "unknown", "secondary", "primary", or "tertiary".

default: This column indicates whether the client has credit in default. It's a binary variable with options "yes" or "no".

balance: This column represents the average yearly balance in euros for the client. It's a numeric variable.

housing: This column indicates whether the client has a housing loan. It's a binary variable with options "yes" or "no".

loan: This column indicates whether the client has a personal loan. It's a binary variable with options "yes" or "no".

contact: This column represents the type of communication used to contact the client. It's a categorical variable with options such as "unknown", "telephone", or "cellular".

day: This column represents the last contact day of the month. It's a numeric variable.

month: This column represents the last contact month of the year. It's a categorical variable with options such as "jan", "feb", "mar", etc.

duration: This column represents the duration of the last contact in seconds. It's a numeric variable.

campaign: This column represents the number of contacts performed during this campaign and for this client. It's a numeric variable.

pdays: This column represents the number of days that passed by after the client was last contacted from a previous campaign. It's a numeric variable where -1 means the client was not previously contacted.

previous: This column represents the number of contacts performed before this campaign and for this client. It's a numeric variable.

poutcome: This column represents the outcome of the previous marketing campaign. It's a categorical variable with options such as "unknown", "other", "failure", or "success".

y: This column is the target variable and indicates whether the client has subscribed to a term deposit. It's a binary variable with options "yes" or "no".

Questions:

- What is the distribution of age among the clients?
- How does the job type vary among the clients?
- What is the marital status distribution of the clients?
- What is the level of education among the clients?
- What proportion of clients have credit in default?
- What is the distribution of average yearly balance among the clients?
- How many clients have housing loans?
- How many clients have personal loans?
- What are the communication types used for contacting clients during the campaign?
- What is the distribution of the last contact day of the month?
- How does the last contact month vary among the clients?
- What is the distribution of the duration of the last contact?
- How many contacts were performed during the campaign for each client?
- What is the distribution of the number of days passed since the client was last contacted from a previous campaign?
- How many contacts were performed before the current campaign for each client?
- What were the outcomes of the previous marketing campaigns?
- What is the distribution of clients who subscribed to a term deposit vs. those who did not?
- Are there any correlations between different attributes and the likelihood of subscribing to a term deposit?