Bank Marketing Dataset:-

Marketing Introduction:

The process by which companies create value for customers and build strong customer relationships to capture value from customers in return. - Kotler and Armstrong (2010).

Marketing campaigns are characterized by focusing on the customer needs and their overall satisfaction. Nevertheless, there are different variables that determine whether a marketing campaign will be successful or not. There are certain variables that we need to take into consideration when making a marketing campaign.

Approach: -

Import data from dataset and perform initial high-level analysis: look at the number of rows, look at the missing values, look at dataset columns and their values respective to the campaign outcome.

Clean the data: remove irrelevant columns, deal with missing and incorrect values, turn categorical columns into dummy variables.

Use machine learning techniques to predict the marketing campaign outcome and to find out factors, which affect the success of the campaign.

Categorical exploration:

Chart, bar chart

Description automatically generated

Flowchart: -

Diagram

Description automatically generated

The 4 Ps:

* Segment of the Population: To which segment of the population is the marketing campaign going to address and why? This aspect of the marketing campaign is extremely important since it will tell to which part of the population should most likely receive the message of the marketing campaign.
* Distribution channel to reach the customer's place: Implementing the most effective strategy in order to get the most out of this marketing campaign. What segment of the population should we address? Which instrument should we use to get our message out? (Ex: Telephones, Radio, TV, Social Media Etc.)
* Price: What is the best price to offer to potential clients? (In the case of the bank's marketing campaign this is not necessary since the main interest for the bank is for potential clients to open depost accounts in order to make the operative activities of the bank to keep on running.)
* 4) Promotional Strategy: This is the way the strategy is going to be implemented and how are potential clients going to be address. This should be the last part of the marketing campaign analysis since there has to be an in-depth analysis of previous campaigns (If possible) in order to learn from previous mistakes and to determine how to make the marketing campaign much more effective.

What is a Term Deposit?

A Term deposit is a deposit that a bank or a financial institution offers with a fixed rate (often better than just opening deposit account) in which your money will be returned back at a specific maturity time. For more information with regards to Term Deposits please click on this link from Investopedia:

Machine Learning Steps Involved:

1. XGBoost

Steps Involved:

* Import data from dataset and perform initial high-level analysis:
* Data Exploration (look at the number of rows, look at the missing values, look at dataset columns and their values respective to the campaign outcome).
* Data Cleaning (Cleaning the data: remove irrelevant columns)
* dealing with missing and incorrect values,
* turning categorical columns into dummy variables.
* Use machine learning techniques to predict the marketing campaign outcome and to find out factors, which affect the success of the campaign.

Dataset Information: -

The data is related with 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, 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 (yes/no) a term deposit (variable y).

Use Case: -

The objective is to train a ML model that returns the probability of a customer to accept the offered product. This is a binary classification task, therefore F1-score is a good metric to evaluate the performance of this dataset as itweights recall and precision equally, and a good retrieval algorithm will maximize both precision and recall simultaneously.

Attribute Information:-

Input variables: bank client data:

1-age (numeric)

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

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

4education(categorical:'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown')

5-default: has credit in default? (categorical: 'no','yes','unknown')

6-housing: has housing loan? (categorical: 'no','yes','unknown')

7-loan: has personal loan? (categorical: 'no','yes','unknown')

related with the last contact of the current campaign:

8-contact: contact communication type (categorical: 'cellular','telephone')

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

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

11-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')
* social and economic context attributes
* emp.var.rate: employment variation rate - quarterly indicator (numeric)
* cons.price.idx: consumer price index - monthly indicator (numeric)
* cons.conf.idx: consumer confidence index - monthly indicator (numeric)
* euribor3m: euribor 3 month rate - daily indicator (numeric)
* nr.employed: number of employees - quarterly indicator (numeric)

Output variable (desired target):

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

References:-

This dataset is publicly available in the UCI dataset repository as "Bank Marketing".