PROPOSAL

Bank Term Deposit Subscription

Here comes your name
Here comes you Email

BANK TERM DEPOSIT ANALYSIS AND PREDICTION

Motivation:

The aim of my project is to extract useful insights from the dataset like how age, bank balance, job title and other factors contribute to subscription of bank term deposit. In that way we can setup bank term deposit marketing campaigns very effectively. I will also build machine learning model to predict whether a particular customer is likely to subscribe to term deposit or not. By analyzing the dataset we can setup effective campaigns to target specific customer and we can also increase revenue and reduce the cost as well.

IMPORTANT ANALYSIS

Questions we will ask from Data:

- Effect of Marital status on term deposit
- Effect of job title on term deposit
- How education effects term deposit
- In which age bracket most of customer lies
- Which variables have strong correlation
- Is there any effect of age and balance on term deposit subscription

DATA INFO

There are 9500 records in the dataset, with 17 columns/features. Some of the dataset's columns are categorical, while others are numerical. In comparison to those who subscribed, there were more people who did not sign up for a term loan.

Source of dataset:

https://www.kaggle.com/edith2021/bank-marketing-campaign

DATA COLUMNS

Bank client data:

1 - age (numeric)
2 - job : type of job (categorical:
"admin.","unknown","unemployed","management","housemaid","entrepreneur","s
tudent",
"blue-collar","self-employed","retired","technician","services")
3 - marital : marital status (categorical: "married","divorced","single"; note:
"divorced" means divorced or widowed)
4 - education (categorical: "unknown","secondary","primary","tertiary")
5 - default: has credit in default? (binary: "yes","no")
6 - balance: average yearly balance, in euros (numeric)
7 - housing: has housing loan? (binary: "yes","no")
8 - loan: has personal loan? (binary: "yes","no")

Related with the last contact of the current campaign:

1 - contact: contact communication type (categorical: "unknown","telephone","cellular")
2 - day: last contact day of the month (numeric)
3 - month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")
4 - duration: last contact duration, in seconds (numeric)

Other attributes:

1 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
2 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)
3 - previous: number of contacts performed before this campaign and for this client (numeric)
4 - poutcome: outcome of the previous marketing campaign (categorical: "unknown","other","failure","success")
Output variable (desired target):

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

MACHINE LEARNING MODELS AND TOOLS

To solve the classification problem I will use following machine Learning models

1.k-nearest neighbors (KNN)
2. Decision Tree
3. Logistic Regression
4. Support Vector Machines
5.Naïve Bayes Classifier

Tools:

Pandas: Pandas to load and basic exploration of dataset.
Seaborn: Seaborn to visualize and find insights from dataset.
Imblearn: for balancing unbalanced data
Sklearn: sklearn for model building and validation

MVP GOAL:

Basic Overview of Dataset
Exploratory Data Analysis
Data Preparation for model building
Balancing unbalanced dataset
Feature selection
Model validation and selection