



Bank Marketing Campaign

Predicting Term Deposit Subscriptions

Apr 2023

Alice Wang

Introduction

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.

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:

Problem Statement

How can the financial institution have a greater effectiveness for future marketing campaigns?

Goal

The goal of this project are:

1. Prediction of the results of the marketing campaign for each customer and clarification of factors which affect the campaign results. This helps to find out the ways how to make marketing campaigns more efficient.
2. Finding out customer segments, using data for customers, who subscribed to term deposit. This helps to identify the profile of a customer, who is more likely to acquire the product and develop more targeted marketing campaigns.

Data

Marketing bank dataset uploaded originally in the UCI Machine Learning Repository. The dataset gives information about a marketing campaign of a financial institution in which we will have to analyze in order to find ways to look for future strategies in order to improve future marketing campaigns for the bank.

Attribute Description:

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)
4. education: (categorical: primary, secondary, tertiary and 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')
8. balance: Balance of the individual.

Related with the last contact of the current campaign:

9. contact: contact communication type (categorical: 'cellular', 'telephone')
10. month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
11. day: last contact day of the month (numeric: 1-31)
12. 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:

13. campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
14. 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)
15. previous: number of contacts performed before this campaign and for this client (numeric)
16. poutcome: outcome of the previous marketing campaign (categorical: 'failure', 'nonexistent', 'success')

Output variable (desired target):

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

Data Wrangling

[Data Wrangling Notebook](#)

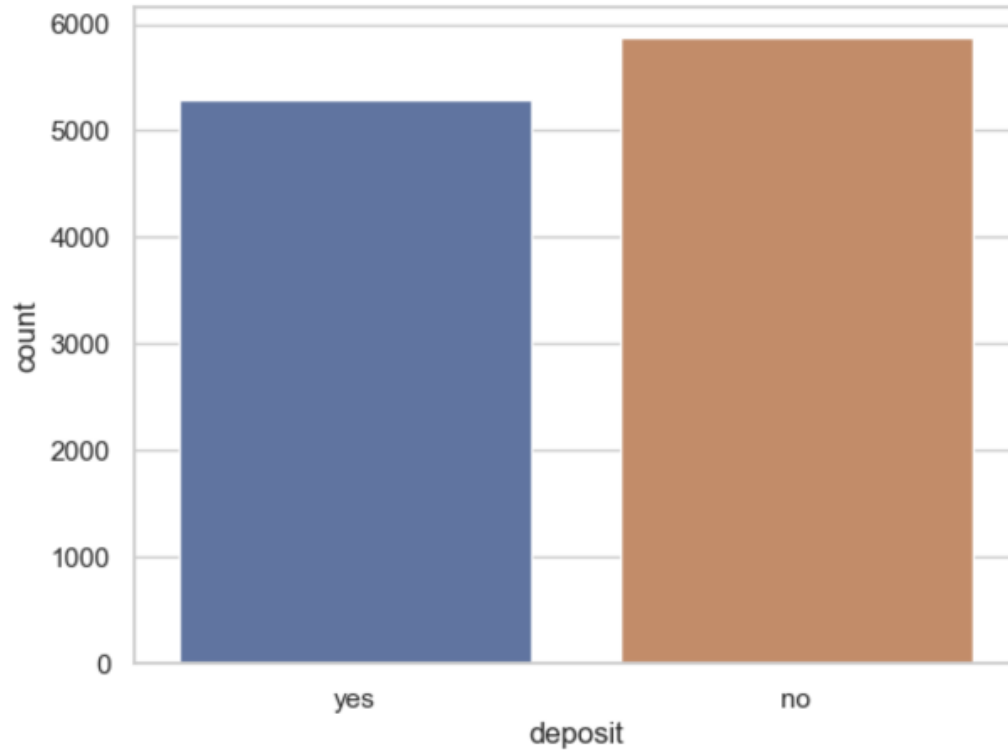
Major steps:

1. The dataset has a shape of (11162, 17)
2. Check for duplicates: There are 0 duplicates for this dataset
3. Check for missing values: There are 0 missing values
4. Will perform EDA and see if we need to drop any fields for pre processing

EDA

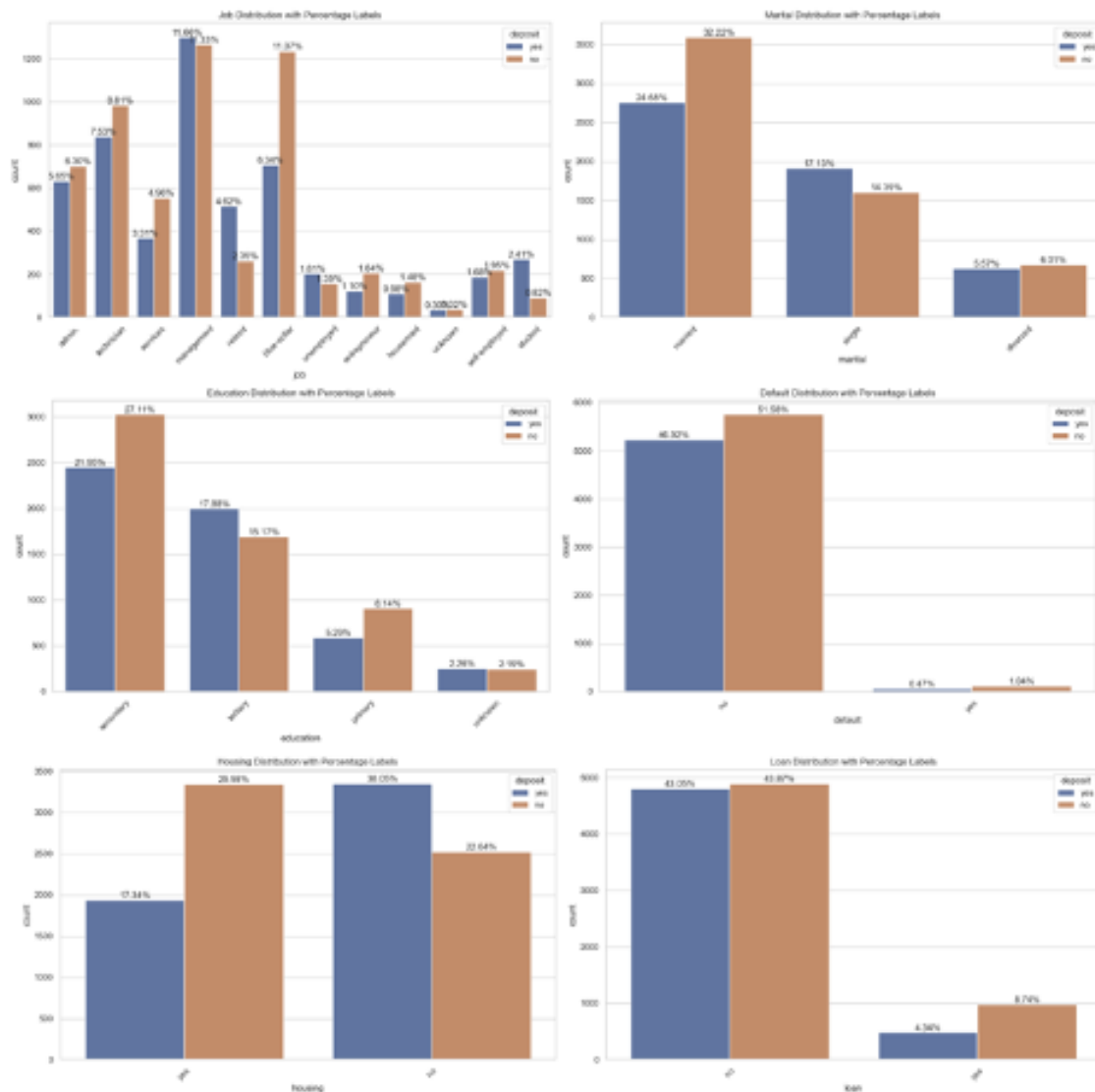
[EDA Notebook](#)

Term Deposit is what we are predicting, EDA is all related to deposit (our target feature). Summary:



- Target variable 'deposit' is balanced, we can use accuracy as a metric for a model, which predicts the campaign outcome.

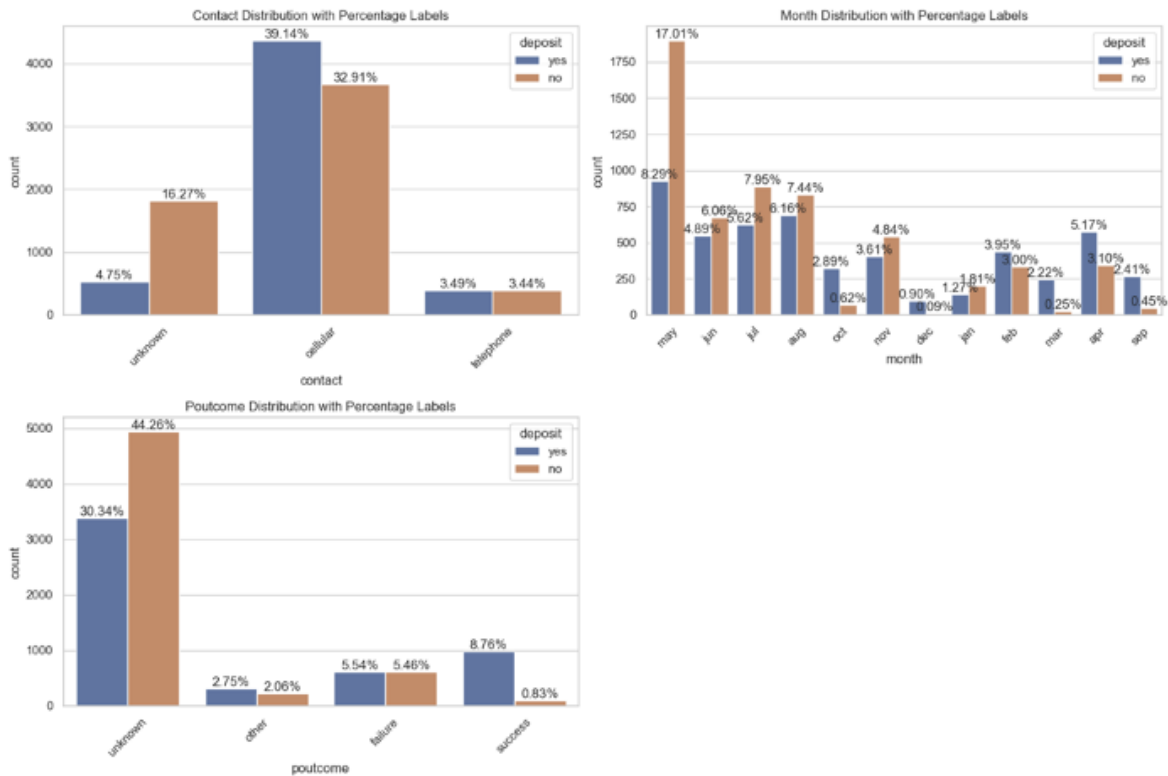
Client Demographic Analysis



- 36% of clients are in the age 30's-40's
- Top 3 job: 20% of clients work in management, followed by 16% in blue-collar, and 15% technician.
- More than half with 52% are married
- 45% have a secondary education

5. Mostly have no loan and no default credit card.
6. It's 50-50 in terms of having housing
7. 90% of clients have balance below \$20,000

Campaign Analysis



1. Some numerical features have outliers (especially 'pdays', 'campaign' and 'previous' columns), will impute the outliers by mean for pre-processing.
2. Contact by celler have 39% deposit rate
3. May & August have higher deposit rate
4. Unknown outcome of the previous marketing campaign has higher deposit rate

Feature Engineer & Modeling

[Feature engineering & ML Notebook](#)

Feature Engineering Major Steps:

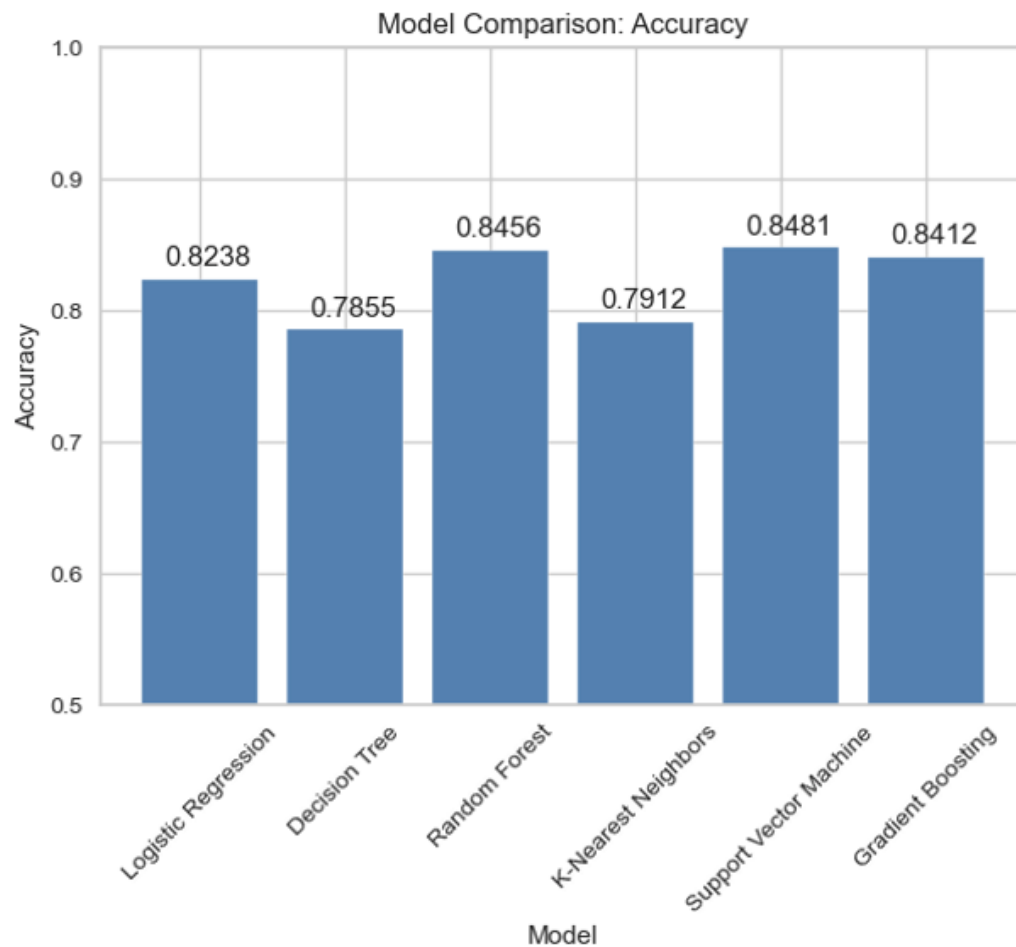
1. Encode some categorical features as ordered numbers when there is information in the order
2. Log transform of the skewed numerical features to lessen impact of outliers
3. Transformation of categorical features via one-hot encoding
4. Convert target variable to boolean
5. Split into testing and training datasets
6. Standardize the magnitude of numeric features using a scaler

Modeling Major Steps: Created below models:

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. K-Nearest Neighbors
5. Support Vector Machine
6. Gradient Boosting

Model Validation Metric:

1. Accuracy
2. Precision
3. Recall
4. F1-score
5. AUC-ROC



Note: Model Tuning on random forest using both GridSearchCV and RandomizedCV did not improve performance, so will use RandomForestClassifier without tuning.

Recommendations

What Actions should the Bank Consider? Solutions for the Next Marketing Campaign (Conclusion):

1) Age Category: The next marketing campaign of the bank should target potential clients from age categories below 30 years old and above 50 years old.

2) Occupation: Not surprisingly, potential clients that were students or retired were the most likely to subscribe to a term deposit. Retired individuals, tend to have more term deposits in order to gain some cash through interest payments. Remember, term deposits are short-term loans in which the individual (in this case the retired person) agrees not to withdraw the cash from the bank until a certain date agreed between the individual and the financial institution. After that time the individual gets its capital back and its interest made on the loan. Retired individuals tend to not spend bigly its cash so they are more likely to put their cash to work by lending it to the financial institution. Students were the other group that used to subscribe term deposits.

3) Balance: The customer's account balance has a huge influence on the campaign's outcome. People with account balance above 1490\$ are more likely to subscribe for term deposit, so future address those customers.

4) Seasonality: Potential clients opted to subscribe term deposits during the seasons of fall and winter. The next marketing campaign should focus its activity throughout these seasons.