

Collin Gros
Zac Holt
Nha Quynh Nguyen
Stephanie Salayandia
Phillip Sauers

STAGE 2: PROJECT PROPOSAL

1. Problems and Motivations

This problem seeks to predict whether a customer will subscribe to a long-term deposit for a bank after the bank has used telemarketing advertisement methods. Term deposit is defined as a type of investment that allows clients to deposit their money into a financial institution account with the aim of earning higher interest. Generally, during this fixed period, money cannot be withdrawn from the account. The term of long-term deposit ranges from one year to five years ^[3]. The dataset itself was compiled in the years following the financial crisis of 2008 ^[1] and will provide a look into the effectiveness of reaching financial customers through automated and distance-based methods.

We believe the results could be particularly useful given the current climate of COVID restrictions. Not only would it model how effective telemarketing methods are in general, but also to identify which customers are most likely to subscribe to the long-term deposit.

Based on the results, we may be able to build a potential customer profile using the features that are most strongly correlated with a subscription. Marketing strategy could also be modified and improved in order to enhance the effectiveness and attract more clients to the program (Term Deposit).

2. Dataset

The dataset is made up of data collected from a Portuguese bank in the 5 years following the financial crisis of 2008. The original team compiling the dataset used feature engineering to reduce the total amount of features from 150 to the 22 most relevant.^[1]

The dataset contains the following features:

<i>age</i>	- Customer's age, in years.
<i>job</i>	- Customer's job, subcategorized.
<i>marital</i>	- Customer's marital status (married/divorced/single).
<i>education</i>	- Customer's level of education (secondary/tertiary/primary/unknown).
<i>default</i>	- Whether a customer has a default in their financial history (true/false).
<i>balance</i>	- The customer's current bank balance.
<i>housing</i>	- Whether the customer has a housing loan (true/false).
<i>loan</i>	- Whether the customer has a personal loan (true/false).
<i>contact</i>	- What method was used to contact the customer (cellular/unknown/telephone).

<i>day</i>	- The date of contact (1-31).
<i>month</i>	- The month of contact (jan, feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec).
<i>duration</i>	- The duration of the call-in seconds.
<i>campaign</i>	- The number of contacts performed during this campaign and for this client.
<i>pdays</i>	- The number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted).
<i>previous</i>	- The number of contacts performed before this campaign and for this client.
<i>poutcome</i>	- The outcome of the previous marketing campaign (categorical: ['unknown' 'other' 'failure' 'success'])

These features are used to predict the output variable:

<i>deposit</i>	- Whether the customer subscribed to the long-term deposit. (true/false)
----------------	--

There are a total of more than 11,100 instances in this dataset, with each entry containing 17 attributes. We believe that this dataset will be large enough for training a learning model to predict whether a customer is subscribed to the long-term deposit or not.

References

- [1] [Moro et al., 2014] S. Moro, P. Cortez and P. Rita.
A Data-Driven Approach to Predict the Success of Bank Telemarketing.
Decision Support Systems, Elsevier, 62:22-31, June 2014
- [2] Bank Marketing Dataset. Retrieved from
<https://www.kaggle.com/janiobachmann/bank-marketing-dataset>
- [3] Chen, J. (March 2020). Term Deposit Definition. Retrieved from
<https://www.investopedia.com/terms/t/termdeposit.asp>