

Analytics Startup Plan

Synopsis: *This document provides a high-level walkthrough of the activities required to guide completion of the analysis.*

Project	The marketing campaign
Requestor	Portuguese banking institution
Date of Request	04/07/2022
Target Quarter for Delivery	28/08/2022
Epic Link(s)	N/A
Business Impact	Predicting client's subscription to a term deposit

1.0 Business Opportunity Brief

The specific ask:

To properly exploit the business opportunity mentioned above, the bank needs to determine the characteristics of the customers who are most likely to take up the product. The task is therefore to provide a set of customer characteristics that best predict their likelihood to take up the product based on historical data and defining traits.

1.1 Supporting Insights

There are two methods through which bank marketing campaigns are carried out. The first method involves mass marketing campaigns targeted at the general population; the second method is a targeted marketing campaign that is aimed toward a select group of individuals. The mass marketing campaign results in a low positive response rate of product purchases or service subscriptions. In contrast, targeted marketing has a comparatively higher positive response. Even though commodity is sold to the wider population, the amount of losses as a result of implementing this approach is unsustainable for the bank. The primary motivation for carrying out the campaign is to sell the product and this is effectively achieved using the direct marketing approach. The advent of data-driven management has seen more marketing managers apply statistical strategies in the identification of potential service subscribers and product consumers. The statistical strategies are key in determining the customers who are most likely to invest in the bank and thus avoid potential impasses.

The tough international economic outlook of the current decade coupled with the impact of the recent COVID-19 pandemic has meant that the need for banks to sell more long-term deposits is more pressing as they look to improve their reserves. This goal by the banks means marketing managers have to increase their campaigns to sell more to the general public. The managers have therefore to be strategic in their approach; utilise the limited resources available to them to attract as many clients as possible. With data from their customer base and previous campaigns, a multivariate data classification method can be applied to identify potential customers.

1.2 Project Gains

Accurate classification of the data from the bank means that the institution can reduce losses that occur from targeting customers who are less likely to purchase the product or subscribe to a service. The marketing manager can therefore reduce the efforts placed in targeting the whole sample of bank customers in their database and only do so for a portion of the sample. This approach increases the rate of positive response and the bank spends less on the whole exercise.

Customers are more likely to maintain their accounts with the bank because of their perceived positive targeting by the bank as opposed to random product advertisements that may not be of interest to them.

2.0 Analytics Objective

- Which variables are essential in the determination of a customer's likelihood to subscribe?
- Does a customer's financial health determine their potential to subscribe to the long-term deposit product?
- Does a customer's demographic characteristics determine their potential to subscribe to the long-term deposit product?

2.1 Other related questions and Assumptions:

The data set is imbalanced and therefore the chance of experiencing an accuracy paradox is high.

2.2 Success measures/metrics

- Increase in the number of subscribes for the term deposit product.
- Decrease in costs for marketing campaigns

2.3 Methodology and Approach

Type of Analysis: PCA Factor analysis, logistic regression, Correlation analysis

The PCA and factor analysis are necessary for the selection of a subset of features that best produce the most optimally efficient model.

The target variable is categorical measuring whether or not the customer has subscribed to a term deposit. This being a binary variable, the logistic regression method is most preferred and this is used to determine the predictive capability of deciding whether a customer would sign-up for a long-term deposit product offering. The accuracy of the regression model will be compared with other machine learning models to determine if it is the best to apply to the case analysis.

The correlation analysis is used to identify the most accurate rules to be applied by the marketing manager to identify the customers who are most likely to subscribe to the long-term deposit. The correlation reveals the subgroup within the different classes that are more likely to subscribe to the service.

Methodology:

The first step would be to assign the designation for customers willing to subscribe to the long-term deposit product (1) or not (0). The whole data set was pre-processed with some continuous variables turned to categorical for ease of use with the logistic regression. Variables with a large number of categories were merged into fewer attributes. The variable age had a wide range that resulted in many classes therefore it was binned into fewer

classes based on the minimum and maximum age of the customers. The output class was ignored in the PCA analysis since it is only applicable to unsupervised machine learning approaches. In addition, character values were omitted when running the PCA analysis as it does not recognize such values. The factor analysis step reduced the highest common variance of all variables into a single rating. It was therefore critical when looking for essential variables from the entire dataset. The machine learning model (logistic regression) was tested using various evaluation metrics. The confusion matrix was used to test the accuracy, sensitivity and specificity of the model.

Output: The output will consist of a number of classes of customers that are likely to subscribe to the service. In addition, the additional analysis would reveal the characteristics of customers in subgroups of the classes that are more likely to subscribe to the service. This provides a more targeted sample of the bank's customer base.

3.0 Population, Variable Selection, considerations

Audience/population selection: banking customers (Portugal)

Observation window:

Inclusions:

Exclusions:

Data Sources: <https://www.kaggle.com/datasets/prakharrathi25/banking-dataset-marketing-targets>

Audience Level: Marketing Team

Variable Selection:

Derived Variables:

Assumptions and data limitations:

Assumptions:

- The data sampling sample is representative of the actual population of term depositors.
- Unbiased data are employed in the analytical modeling process.
- The term deposit analytical project is connected to the data that was gathered.

4.0 Dependencies and Risks

Risk	Likelihood (based on historical data)	Delay (based on historical data)	Impact
False-negative results. The model predicts that the customer is not likely to buy and so the customer is not a target	Low		The opportunity is lost as the customer potential is not realized.

of the campaign while in reality, the customer was likely to purchase			
False positive results. The model predicts the customer is likely to purchase and the bank targets the customer who ends up not purchasing the product.	High		Monetary losses and an annoyed customer for marketing a product that they would otherwise not purchase

5.0 Deliverable Timelines

Item	Major Events / Milestones	Description	Scope	Days	Date
1.	Kick-off / Formal Request	Formally present the case analysis and its need as a contributing factor to the banks' success.		4	08/07/2022
2.	Assessment / Triage	Assessment of the proposal and the credibility of the objectives formed.		2	17/07/2022
3.	Prioritization	Assign resources to data collection and compilation		14	17/07/2022
4.	Data Exploration & Analysis <ul style="list-style-type: none"> Issues with duplicates Issues with Spend data 	<p>Explore the data and pre-process it for analysis.</p> <p>Remove any duplicate data and deal with spent data.</p> <p>Bin continuous variables, grouping base don't the minimum and maximum values.</p> <p>Partition of the data eliminating character values for some of the analysis steps</p>		7	24/07/2022
5.	Story Board 1	Present the analysis plan to the panel for critique		1	25/07/2022

6.	QA Output	Implement critiques from storyboard 1 to better present findings and capture the KPIs of interest to the bank management		4	27/07/2022
7.	Internal team Presentation	Present the findings to the analysis team for further critique		1	02/08/2022
8.	Go/No Go	Decide on elements to include or exclude from the final presentation. Or redo the entire analysis, and present it to the internal team before presenting the final findings to the bank management		7	09/08/2022
9.	Story Board 2	Present the findings to the bank management		1	19/08/2022
10.	Pilot	Carry out a pilot study using the model developed. Collect data and analyze it to improve the model.		7	26/08/2022
11.	Delivery & sign-off	Deliver the results and sign off on the final model adopted.		1	31/08/2022