

*Summary Report – Bank Marketing Analysis*

Course #: ITC 6040

Course Name: Informatics Capstone

Submitted By:

Harshil Donga

Ronit Patel

* **Analysis**

The given dataset was about marketing campaign data ran by the the Portuguese bank institution. To analyse the data, we have been given two different CVS files named “Bank-Full” and “Bank”. We have imported this both files as a train and test dataset from UCI machine learning repository. Then we have looked into the structure of the data and count all the observations according to the positive and negative responses for each variable. Moreover, we also checked if there is any missing value in the given dataset but the dataset was very structured. Hence, we couldn’t find any missing elements.

Additionally, we also count all the observations according to the Month, job status, marital status, loan, contact, poutcome and many other variables.

This project can help banks or any other customer-based institution to target their audience who would potentially subscribe for their services and run their marketing campaign around this audience. It can save their money and time as they could detect potential customer and run a campaign around them instead of the whole population.

This target customer marketing analysis and prediction have been carried out for a long time. Not only in the banking sector, but also in many other businesses like retail or any other customer-oriented industry target customer marketing has become essential. For instance, the grocery giant Target created Machine Learning Algorithm to predict the pregnancy of women based on their shopping. The reason behind this was to increase the sales of baby products. This marketing strategy worked for Target, it was successful in predicting the pregnancy and they increased the revenue from sales of baby products. In the finance sector, banks or stock brokers use this marketing strategy to increase their customer base and save their money and time by focusing on a small number of the population who are more likely to subscribe for their services

Our project revolves around three major stakeholders that are customers, client and marketing campaigner. The ideal client for our would be bank and marketing campaign company because Bank and marketing campaign company would be able to identify potential customers to run the marketing campaign. Eventually, that will help to save money and time for them.

We are using Python programming language to conduct analysis and predictions and Jupyter notebook to run python code. After importing dataset to the jupyter notebook, we conducted explanatory data analysis using graphs and charts to explore some relationship within the dataset. But, we couldn’t find any convincing results and so, after conducting feature engineering on the dataset we were able to find the attributes that were affecting output attribute the most and attributes which weren’t affecting the output variable much. Although, explanatory data analysis using graphs and charts helped us somehow in determining some of the relationships within the dataset.

After conducting explanatory data analysis using graphs and charts, we did data preprocessing to make data compatible with Machine Learning Algorithm. For data preprocessing, we encoded multiple categorical attributes. Then, we did feature engineering on data to find out the least significant attributes that don’t affect the outcome attribute. We found out that age and pdays were the two attributes that least affects the outcome variable and so, we dropped that attributes.

After doing data preprocessing and feature engineering, we began with running Machine Learning classification algorithms. We used four Machine Learning Classification methods Decision Support tree, Random Forest, Logistic Regression and K-Nearest neighbor. We trained training data with these models and predicted the outcome of the test dataset which had 4521 records. Then, we used a classification report and confusion matrix to evaluate the outcome of these models based on their prediction accuracy. For decision support tree and K-Nearest neighbor we got 100% accuracy, that is, these models precisely predicted whether the customer whose record is in test dataset will make a one-month deposit with a bank or not. The confusion matrix and classification report for both are shown below. While random forest algorithm predicted the outcome with 99% accuracy and logistic regression method with 89% accuracy.

(Decision Support tree and K-Nearest neighbor)

[[4000 0]

[ 0 521]]

precision recall f1-score support

no 1.00 1.00 1.00 4000

yes 1.00 1.00 1.00 521

avg / total 1.00 1.00 1.00 4521

(Random Forest tree)

[[3999 1]

[35 486]]

precision recall f1-score support

no 0.99 1.00 1.00 4000

yes 1.00 0.93 0.96 521

avg / total 0.99 0.99 0.99 4521

(Logistic Regression)

[[3904 96]

[351 170]]

precision recall f1-score support

no 0.92 0.98 0.95 4000

yes 0.64 0.33 0.43 521

avg / total 0.89 0.90 0.89 4521

* **Business Recommendation**

This project is about helping any industry like bank or commerce to attract their targeted customers via advertising campaign in a most effective way that can save their money and time in the campaign. For these project, I would suggest the Portuguese bank that they should focus ojn customers that are doing technical, admin or management related jobs during their next telemarketing campaign as they are most likely to make a one-month deposit with the bank. Also, the duration of the call is the most significant factor pertaining to someone will make a one-month deposit with a bank or not. So, they should try to convince customer with a long conversation on phone and explain them about the benefits as we discovered that the more is the duration of call the more are chances of the customer to make a one-month deposit with the bank. Moreover, the bank can use our decision tree or K-nearest neighbor algorithm during their next campaign, input the details about their customer one of the models, determine the customers that are most likely to make a one-month deposit and only contact those customers which will save their time and money during a future marketing campaign.