

## **Management Problem**

The firm, a Portuguese bank, conducted a marketing campaign where they were able to document specific characteristics about their clients and if they subscribed to a term deposit or not. The firm hopes to find useful information in the data collected in order to create a new direct marketing campaign with high success rates than in the past.

## **Approach**

The first step in the process of identifying recommendations for what clients to target in the upcoming direct marketing campaign was to find a suitable predictive model. The model selection process approach included looking at four different models:

1. Logistic Regression
2. Gaussian Naive Bayes
3. Decision Tree
4. Linear Support Vector Machines (SVM)

Leveraging the full data set provided by the firm, we were able to split the dataset into a training set (70%) and test set (30%). We then ran this data through each model in order to calculate the Accuracy, Cross-Fold Validation Results, and the Area Under Curve (AUC) results. We leveraged these results in order to select the best model. After running the tests, we have determined a Logistic Regression to be the best model for this project as it yielded the highest AUC result for True Positive Rate vs False Positive Rate.

The final step in the analysis was to look at the most important attributes of the clients which resulted in a client purchasing a term deposit from the firm. In order to identify the most important attributes, we looked at the coefficients of the Logistic Regression Model. The next

step was to calculate the response rates (clients who purchased a term deposit / total clients in that category) for all categories within each important attribute.

### **Recommendation**

Based on the results of the analysis and through exploratory data analysis (EDA) and the Logistic Regression model, we have found the following recommendations for the firm to target in their direct marketing campaign which are predicted to yield the highest success rates:

- Less than the age of 20 and over the age 65 (students and retirees)
- Contact clients around the first and tenth day during the months of March, September, October and December
- Clients with at least a college level education
- Clients without a current loan (housing or other type of loan)
- Clients who do not have any debt in default

### **Attachments**

- ROC Chart for all Models used for model selection (roc\_chart/roc\_curve.png)
- Coefficient Charts used to determine important attributes (coef\_charts/\*.png)
- Response Rates for each attribute used during EDA (response\_charts/\*.png)
- Console Output (console\_output.rtf)