

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 higher success rates than in the past. The goal of this assignment is to provide recommendations for what clients and the time of year the firm's next marketing campaign should be geared towards.

Approach

Similar to the previous assignment, the approach was to create a training and test data set (70/30) and run that data through multiple supervised learning algorithms in order to identify important features in the data. In the previous assignment, we failed to create dummy variables for the categorical attributes and thus, we have rectified this in assignment 4. Since the data used in the previous assignment did not include the dummy variables, we still decided to run the data through the same four supervised learning models as listed below:

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

Based on the results after training and testing the four models listed above, Logistic Regression still produced the greatest Area Under the ROC curve (AUC) of 0.91. Due to these attractive results, we chose to stick with the Logistic Regression model and look to the coefficients to identify important features for the successful outcomes of our target response variable. Secondly, we also decided to leverage the Decision Tree model due to its high

accuracy (0.898), AUC (0.88), and cross validation results (0.9). We then reviewed the important features derived from the results of the Decision Tree model.

Recommendations

Based on the results of the analysis of the Logistic Regression model's coefficients as well as the important features from the Decision Tree model, we have identified the following recommendations for the firm to leverage when designing their next direct marketing campaign:

- For a timing perspective, the firm should hold their next campaign in the months of March, October, and December
- Contacting potential clients either on their telephone or cellular
- Clients who purchased previous services from the firm from the prior campaign
- Clients with either white collar jobs, retired or are currently students
- Clients with at least a tertiary education

Attachments

- Python Script (BittingNate_Assignment_4.py)
- Logistic Regression Coefficients Chart (coef_chart/logreg_coef.png)
- Decision Tree Important Features Chart (coef_chart/tree_important_features.png)
- ROC Chart (roc_chart/roc_curve.png)
- Graphviz Decision Tree Chart (tree/dt.dot)