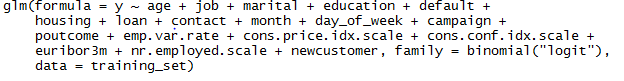
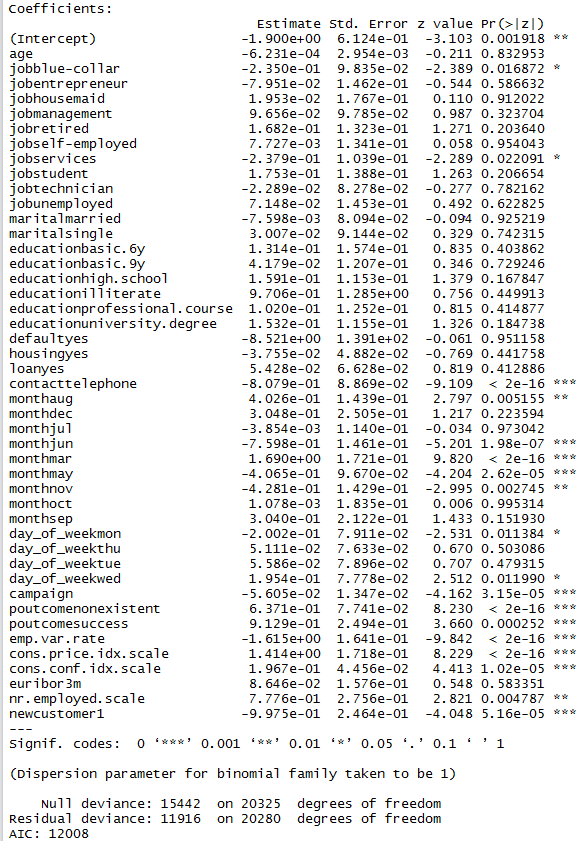
Analysis Results and Interpretation:

Both logistic regression and decision tree methods were utilized for model creation. The first model (Model 1) was created by using glm function with all independence variables from the training dataset which was created in the data pre-processing step.



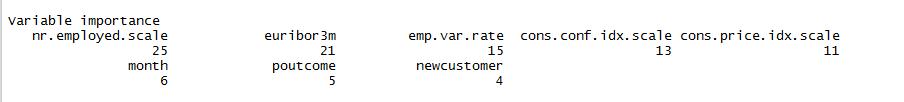
The summary of the Model 1 is listed below:

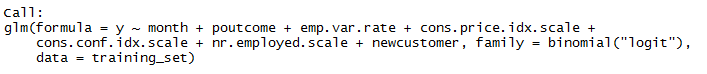


According to the coefficients matrix, job type, contact type, month, day, number of contacts during this campaign, outcome of previous campaign, employment variation rate, consumer price index, consumer confidence index, number of employees, and customer type were variables with high significance. The AIC of the model was 12008.

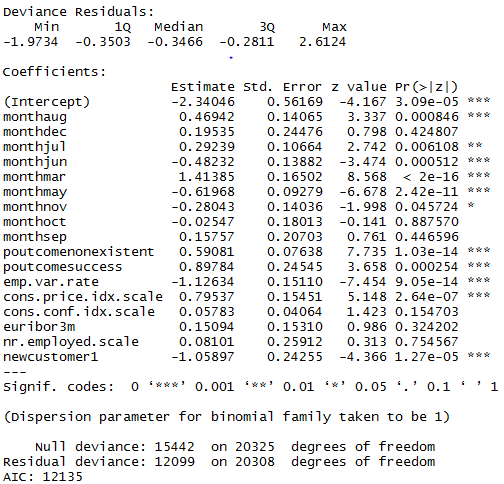
The next logistic regression model was based on the result from Decision Tree model.

The Decision Tree model was created by importing all independent variables from the training data set into rpart function from rpart package. The team wanted to use the Variable Importance matrix from the Decision Tree model to select variables for the second logistic regression model. The team wanted to see if a dimension-reduced logistic regression model will have a better accuracy ratio.





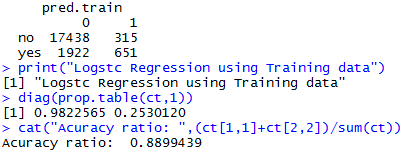
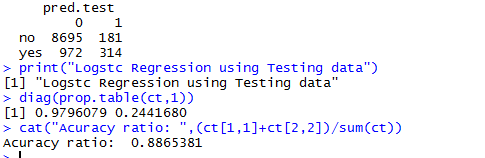
The second model (Model 2) was based on only independence variables which were suggested by the Decision Tree model.



The result has a higher AIC value then Model 1, and the significant variables are moth, previous campaign output, employee variance rate, customer price index, number of employees, and customer type.

Since a lower AIC sore means a better model, the team chose Model 1 as the final model.

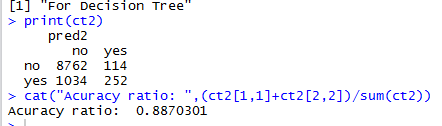
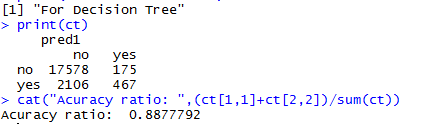
Blow are the predict results from the training set and testing set using the Model1



The accuracy ratio of the Model 1 was 0.8865 for the testing set, and 0.8899 for the training set, which were extremely close, and the hit rate from both results were close to 98% and 25%.

To find out if the regression model has the best prediction result, the team compared the result from Model 1 with the decision tree model using both training set and test set, and found out that Model 1 with all independent variables is better when predict using using test set

Training Set and Test Set



Findings

The overall accuracy of the regression model is 89%, and it is almost the same compare to the decision model. Both Decision Tree and Logistic Regression models suggests that the prediction of weather a customer will open a new account in the Bank is really depends on the economic condition of the time, such as customer price index, employment variation, and number of employees. Also, a new customer or a customer who did not open an account with the bank is more likely to open a new account this time. From the perspective of time, customers are more likely to open a new account from March through August. These findings suggest the managers of the bank should start campaign at the beginning of the first half of the year and should focus on customers who did not have existing account with the bank.