Project Proposal

Case Study:

Predicting Customer Churn for a Telecommunication Company

Using R

Group #

Names :

- A
- B
- C
- D

Motivation for Choosing the Case

Compared to other business case, the customer churn case is relatively the most important issue to deal with. Every company relies heavily on its ability to retain customers. Losing customers would mean the reduction in overall customer lifetime value, and it directly affect the probability of the company. Moreover, the loss would be translated into higher customer acquisition cost since the firm should get more new customers to balance off the number of customers leaving the company's service.

While customer churn issue is relevant in every industry, it is particularly a vital issue in a telecommunication industry. Companies in telecommunication industry can have an average annual churn rates to be as high as 67 percent, and 75% new customers are churners of other companies (Hughes, 2010). Preventing churn can save millions of dollars, and hence predicting churn before it happens is crucial. Developing a solid model to predict churn will provide the company with insight on customer profile that has higher possibility to churn and then create a quick prevention measures.

The target of this project is to create a logistic regression model and the decision tree model to predict the customer churn. We can split the data to make a training set, which we will make the model from, and test set, which we will test the model's accuracy. We can also compare the two models and determine which model gave the more accurate result.

Initial data overview & exploratory analysis

The data for the case consists of 2,114 observations with 14 variables, ranging from customers' profile, such as ID and gender, to their interaction with the company's services, such as tenures, payment method, and monthly charges, including the Churn variable which will be the target variable. Using built in function in R, we can see that there are two incomplete or missing value in TotalCharges. These two observations can be considered insignificant compared to total 2,114 observations, so we can eliminate the two and get the total of 2,112 observations. We can also eliminate unnecessary variables, such as the customerID, which has nothing to do in determining whether a customer will churn or not. Hence, we will analyze 2,112 observations, 12 predictor variables, and the target variable Churn.

For initial data overview, we can look at its data structure. Using *str* function in R, we can see that most of the variables have character data type, with the exception of SeniorCitizen and tenure being integer, and MonthlyCharges and TotalCharges being numerical. Most of these variables should be converted to factors so that R can understand the data for further analysis. For instance, the SeniorCitizen is comprised of 1 if the customer is a senior citizen, and 0 if the customer is not. We need to tell R that it's

a factor using *as.factor* function. The target variable Churn should also be made as a factor and numerical 0 and 1 for the predictive model.

The next step is to look at descriptive statistics of each variables and get a quick summary. We may also want to see correlation between the numeric variables, MonthlyCharges and TotalCharges, to see if we can eliminate one of the two. We can also create grouping for the tenure variable and create another variable that based on the group. However, for the primary analysis, we will let the variables be and provide no further changes.

For better understanding, we can create bar graph with overlay for each predictor variable and the target variable. The overlay bar graph will depict how each variable in the data corresponds with the Churn variable. For instance, we can see that month-to-month contract has about 50% Churn, which is much higher than customer with one or two year contract. It can be the first sign to see that this variable will be an important variable to predict the Churn variable.

References

Hughes, A. (2010, March). *Churn reduction in the telecom industry | Database Marketing Institute*. Database Marketing Institute. http://www.dbmarketing.com/2010/03/churn-reduction-in-the-telecom-industry/