ENABLING CHURN REDUCTION USING ANALYTICS

A FINAL REPORT WAS SUBMITTED

TO

KENT STATE UNIVERSITY

MASTER OF SCIENCE

IN

BUSINESS ANALYTICS

SUBMITTED BY

MEGHANA UDIGA

SNIGDHA HARI

KARTHIKEYAN RAMESH

INSTRUCTOR

DAMBAR UPRETY; PH.D.

PROFESSOR

Group members and contributions:

Meghana Udiga	Prepared the code for the project
Snigdha Hari	Prepared the code for the project
Karthikeyan Ramesh	Assisted in making Presentation

PROJECT GOAL

The Goal of this Project is to collaborate with ABC Wireless Inc. to assist them to overcome their customer churn problem. We're working as part of a team to build a model that can predict/identify users who are likely to churn utilizing previous data from ABC Wireless Inc. We use the modelling technique/approach we learned in class. For this model, we've added a variety of variables from the dataset. The goal of this initiative is to use analytics to reduce turnover. Churn (customer loss due to competition) is a concern for telecom businesses since acquiring a new client is more expensive than keeping a current one from leaving.

The majority of telecom companies undergo voluntary turnover. The churn rate has a substantial impact on the customer's lifetime value since it determines the length of service and future income of the firm. Telecom companies spend hundreds of dollars to acquire a new customer, and when that customer leaves, the company not only loses that customer's prospective revenue, but also the resources spent to acquire that customer. Churn cuts away profitability. To prevent churn, telecom companies have used two strategies: (a) an untargeted strategy and (b) a targeted strategy. To develop brand loyalty and thus retain clients, the untargeted strategy relies on great products and extensive advertising. The targeted strategy focuses on identifying consumers at risk of leaving and responding accordingly to maintain them. ABC Wireless Inc uses a focused strategy and a predictive analytic technique to identify customers who are likely to churn ahead of time. We then provide customized programs or incentives to certain clients. If attrition forecasts are off, this strategy can cost a company a lot of money since the company wastes incentive money on customers who would have stayed anyhow. R is being used to provide data mining, visualization, and analytics for churn prediction as well as define a modelling approach for a given data collection and assess the strategy's performance.

DATA EXPLORATION AND ANALYSIS

To obtain a sense of our data set, we'll start with a summary of the data frame. When you initially upload the data and check the summary of the data, we can see there are 3333 observations of 20 distinct variables. With the exception of the columns (variables)'state, 'area code, 'international plan, 'voice mail plan, 'total night calls,' and 'churn,' the summary reveals that many columns contain 'NA' values. We can determine a bell curve distribution of data for the greatest quantity of data or variables from the above result. We can also see that "Total day minutes" and "Total evening minutes" have a small proportion or many outliers, respectively. The abnormal skewness of "Customer Service calls" is also obvious. # We'll figure out how many clients there are based on the churn statistics.

We have used, Barplot(Churn Count,xlab="Churn",ylab="Count",col = "darkblue",main = "Number of Customers based on churn data") We can see from the given figures that 2850 consumers did not switch, while 483 users went to another supplier. We can deduct from the above that 483 clients only migrated to other providers for various reasons. We will calculate the number of subscribers depending on the State because 2850 clients have stayed with the present provider.

Details of your modelling strategy (i.e. what technique and why):

The ggplot shows a significant inverse relationship between the total day fee and the number of customer support calls, total foreign costs, and total evening charges for churned customers. th the current provider. We use the ggplot to depict the correlation between the variables where churn is equivalent to Yes. ggcorrplot (Corr churn cust, method = "circle", type = "lower", ggtheme = theme linedraw) r ggcorrplot (Corr churn cust, method = "circle

According to the statistics, churn is equal to yes when calls are made to customer support because the charges are higher.

Model Selection: The influence of various factors and their relevance in predicting the result of the dependent variable is illustrated using a predictive model based on regression and Decision Tree Models.

Regression can be expressed in two ways: • Linear Regression • Logistic Regression Because the dependent variable (target variable) is categorical, a logistic regression model is more suited than other models in this circumstance. While a linear regression model is appealing, it is inadequate for forecasting the performance of a binomial feature. The optimal outcome for this model, as determined by logistic regression, is a probability or likelihood of possibilities that lie between 0 and 1. Furthermore, because classification is our major aim, we examined the dataset and selected Logistic Regression and Decision Models. We'll test the two models on our dataset and pick the best one to be the final model to predict the test dataset. Predictive Ability Determination Using Logistic Regression and Decision Tree Models:

The following methods were followed before selecting a model:

1. The dataset has been separated into training and validation sets to avoid overfitting the model.

2. Using a logistic regression model to predict the validation set outcomes.

3. Testing the model's performance with a confusion matrix.

4. Create a decision tree model and anticipate the validation set outcomes.

5. Verifying the model's accuracy Creating a Logistic Regression Model: Logistic regression is a statistical analytic technique for predicting a binary result, such as yes or no, based on past data set observations. A logistic regression model investigates the relationship between one or more independent factors in order to predict a dependent data variable.

A logistic regression, for example, might be used to predict whether a political candidate will win or lose an election, or if a high school student would be admitted to a specific institution. Simple comparisons between two options are available with these binary outcomes.

Estimation of the model's performance:

We're dealing with a collection of data that contains a list of clients whose future attrition we need to forecast. We forecasted 97 users switching from ABC Wireless to another network 1000. The accompanying box plot distribution shows that customers who contact customer service more than 2-4 times are considerably more likely to transfer suppliers. It may be estimated that around 64% of all customers who contacted customer service for the first time churned. The following are the outcomes of using the confusion matrix:

1. Accuracy: 86.19 per cent 2. Sensitivity 97.02 per cent 3. Specificity: 21.88 per cent

Creating a Decision Tree Model: Making a tree-shaped diagram to map out a plan of action or a statistical probability analysis is what decision tree analysis is all about. It's a strategy for breaking down big issues or splitting them into smaller portions. Each branch of the decision tree indicates a possible conclusion.

The Confusion Matrix revealed the following findings: Accuracy: 91.14 per cent Sensitivity: 97.02 per cent, Specificity: 56.25 per cent Choosing the best model: When comparing the two models, it is evident that the Decision Tree Model is the preferable alternative since it is more accurate than the logistical regression model. Although the sensitivity of both models is equivalent, the Decision Tree has greater specificity. As a consequence, the Decision Tree Model is the most appropriate and best model to use. Now, in the final model, we'll utilize the test data and the Decision Tree Algorithm to forecast Churn.

INSIGHTS AND CONCLUSIONS

We discovered that decision tree models have higher AUC and accuracy values than logistic regression models. As a result, the option We're working with a set of data that includes a list of clients for whom we need to predict future attrition. We predicted that 97 customers would switch from ABC Wireless to another network out of 1000.mal model for future predictions is the decision tree. To enhance their customer retention percentage, we recommend that ABC Wireless strengthens their customer care call center and maintains regular contact with their consumers. There is a dipolar link in the graph; those who did not churn had lower total evening minutes and total day calls, resulting in greater totals, evening minutes, and day calls.

Those who churned experienced the opposite effect. ABC Wireless should create more competitive packages that take advantage of these dipolar connections, according to us. Those that do not churn and do not have an international plan make the most overall foreign calls, with the highest total international costs at first (we suspect these customers to be new ones). There appears to be a steady intermediate range on the x-axis between 3 and 8, however those who churned and did not have an international plan (green curve) had higher overall foreign costs.

We recommend that ABC Wireless sell its foreign plans more actively in order to reduce churn. Due to time restrictions, we focused primarily on the numerical variables 'total international charge,' 'total international calls,' 'total eve minutes,' 'total eve charge,' and categorical variable 'international plan,' as well as the sign and magnitude of their coefficients and statistical significance.