Predicting Customer Churn in Retail Business

Nafisseh Heiat, College of Business, Montana State University-Billings, 1500 University Drive, Billings, MT 59101. 406-657-2224, nheiat@msubillings.edu

ABSTRACT

Customer turnover or churn is a very important concern in many industries. In this study, Artificial Neural Network (ANN), Decision Tree (CHAID) and Classification & Regression Trees (C&RT) algorithms are applied to SAS dataset, and results are analyzed to determine the most efficient model for predicting customer churn in telecommunications.

INTRODUCTION

In many industries, customer turnover or churn is an important concern. Customer churn occurs when customers or subscribers stop doing business with a company or service, also known as customer attrition. It is also referred as loss of clients or customers. One industry in which churn rates are particularly useful is the telecommunications industry, because most customers have multiple options from which to choose within a geographic location.
Churn refers to the tendency of a subscriber to switch providers; it is one of the most common problems faced globally in the telecommunication industry. Reasons why a customer might churn include a competitive stimulation, unhappiness with service after the sale, dissatisfaction with quality of services, a move to another location, or disconnection by the provider due to account delinquency.

Customers switch from one provider to other results in loss of considerable profit. According to Lu Junxiang telecommunications industry experiences an average of 30-35 percent annual churn rate and it costs 5-10 times more to recruit a new customer than to retain an existing one.

In order to identify high risk customers that may switch to another provider and manage churning rate, we need to develop reliable models. Once a reliable and accurate churn model is developed and tested, companies may use the model for churn management. In this study, I apply Artificial Neural Network (ANN), Decision Tree (CHAID) and classification (C&RT) algorithms to a telecommunication dataset and compare the results obtained from ANN with results obtained from CHAID and C&RT to determine the most efficient model in terms of classification accuracy.

DATA PROCESSING AND METHODOLOGY

The dataset used in this study has 4708 entries and 15 attributes, including ID column and target churn. (1=churner, 0=non-churner). The attributes of the dataset are shown in table 1.

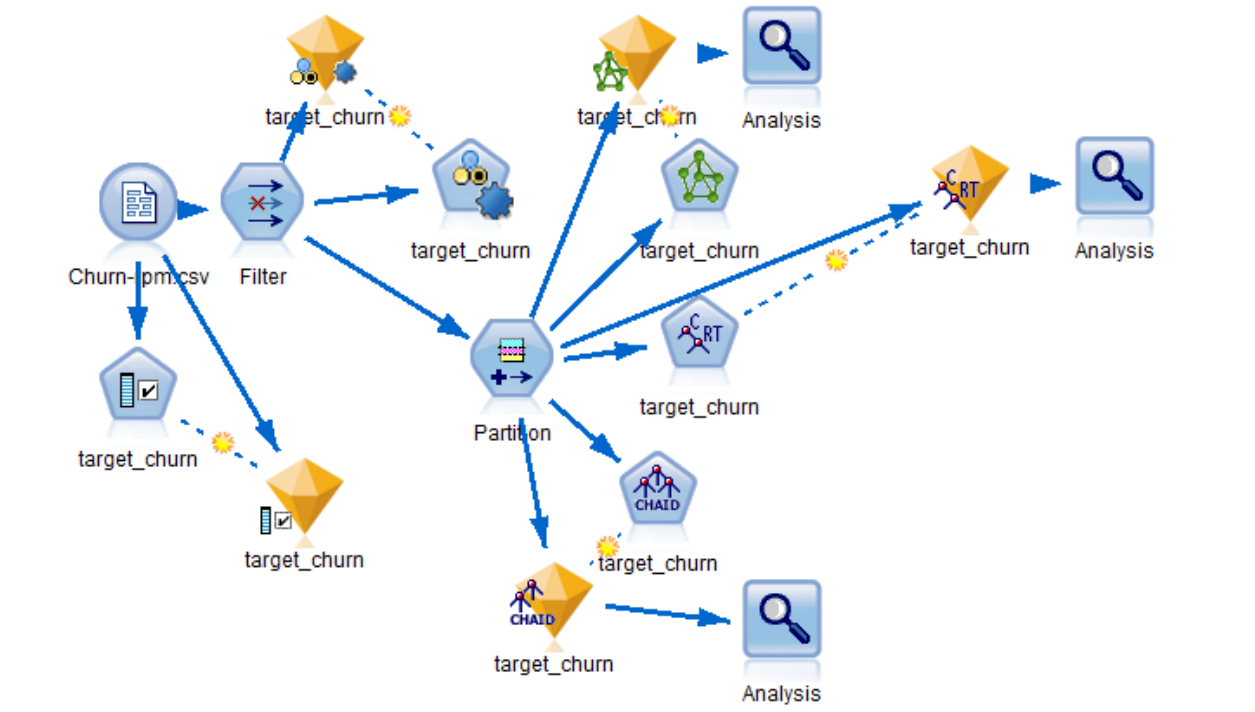
Table 1. Attributes in telecommunication dataset

S No	Attribute Name

1	Account_ID
2	Current_Days_OpenWorkOrders
3	Method_of_Payment
4	Current_TechSupComplaints
5	Avg_Hours_WorkOrderOpenned
6	Avg_Days_Delinquent
7	Equipment_Age
8	current_billamount
9	Avg_Calls_Weekdays
10	Avg_Calls
11	Condition_of_Current_Handset
12	Account_Age
13	Sec_Out_PercentIncrease_MOM
14	acct_plan_subtype
15	target_churn

Figure 1 shows the model developed for this study.

Figure 1. Model developed for this study



I used the SAS churn dataset (available at SAS community website) in this study. The original data set had a number of categorical variables, some of which have been transformed into a series of binary variables so that they could be appropriately handled by the data mining software.

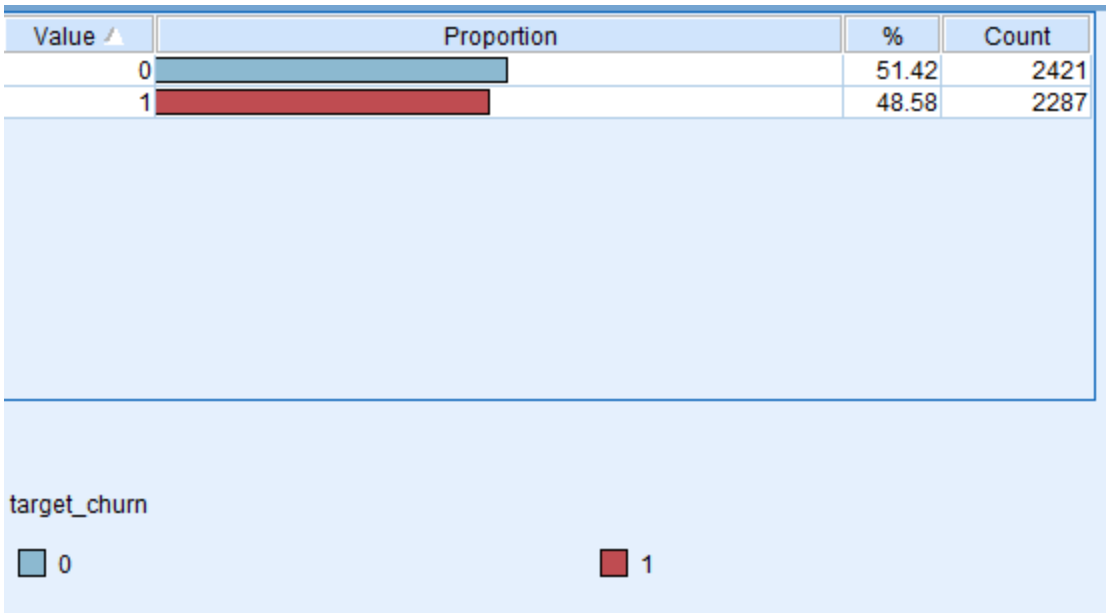
In Figure 1, my model starts with selecting the data set for the analysis. It follows with a feature selection node that identifies important variables for classifying the customer churn. Filter node includes the important variables and assigns the appropriate data type to the target/dependent and input/independent variables. Next, the dataset is partitioned to training and testing sets (70%, 30%). The original dataset has an even number of non-churned customers (51.42%) and churned customers (48.58%). Thus, addition of a balance node was not needed in here to make the distribution of churned and non-churned customers equal.

Next CHAID (Chi-squared Automatic Interaction Detector), ANN (Artificial Neural Network) and C&RT (Classification & Regression Trees) algorithms are applied to the dataset and analysis and evaluation node are added to analyze the results.

ANALYSIS

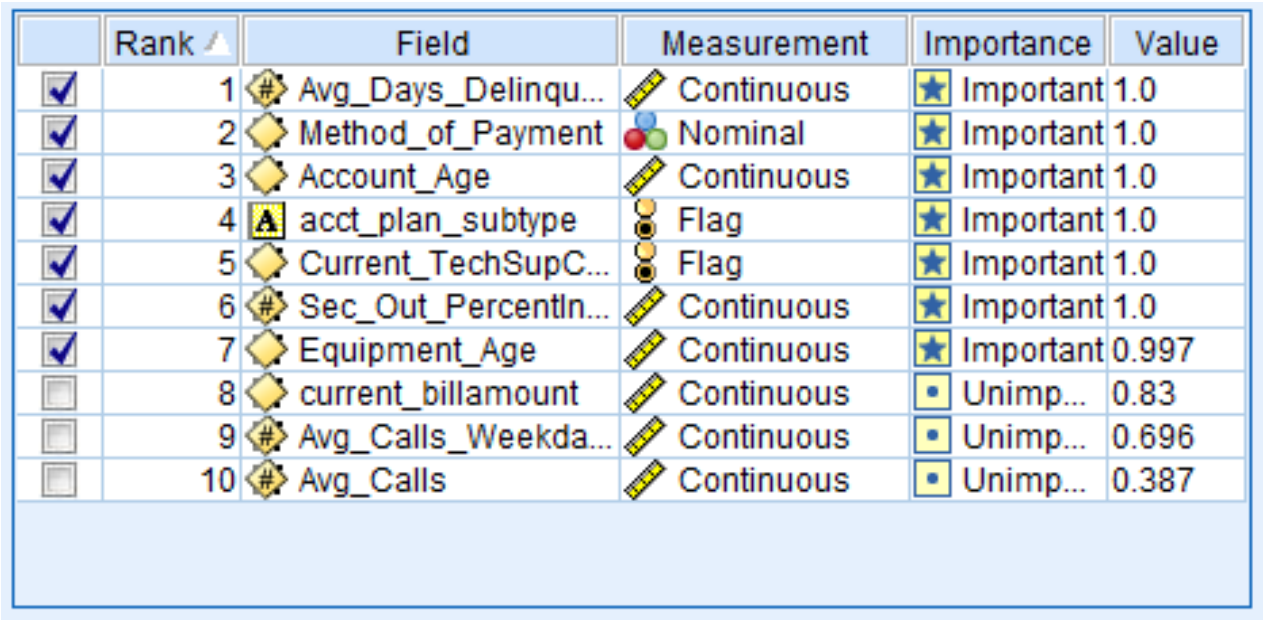
Figure 2 illustrates the distribution of target-churn, which is almost equal for non-churners (0) and churners (1).

Figure 2. Churn Distribution



Next, Figure 3 shows Feature Selection, which determines important variables in this study.

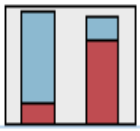
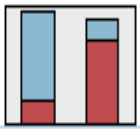
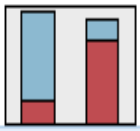
Figure 3. Feature Selection



	Rank	Field	Measurement	Importance	Value
☑	1	Avg_Days_Delinqu...	Continuous	Important	1.0
☑	2	Method_of_Payment	Nominal	Important	1.0
☑	3	Account_Age	Continuous	Important	1.0
☑	4	acct_plan_subtype	Flag	Important	1.0
☑	5	Current_TechSupC...	Flag	Important	1.0
☑	6	Sec_Out_PercentIn...	Continuous	Important	1.0
☑	7	Equipment_Age	Continuous	Important	0.997
☐	8	current_billamount	Continuous	Unimp...	0.83
☐	9	Avg_Calls_Weekda...	Continuous	Unimp...	0.696
☐	10	Avg_Calls	Continuous	Unimp...	0.387

Figure 4 shows most accurate algorithms that are determined by Auto classifier.

Figure 4. Best models determined by Auto classifier

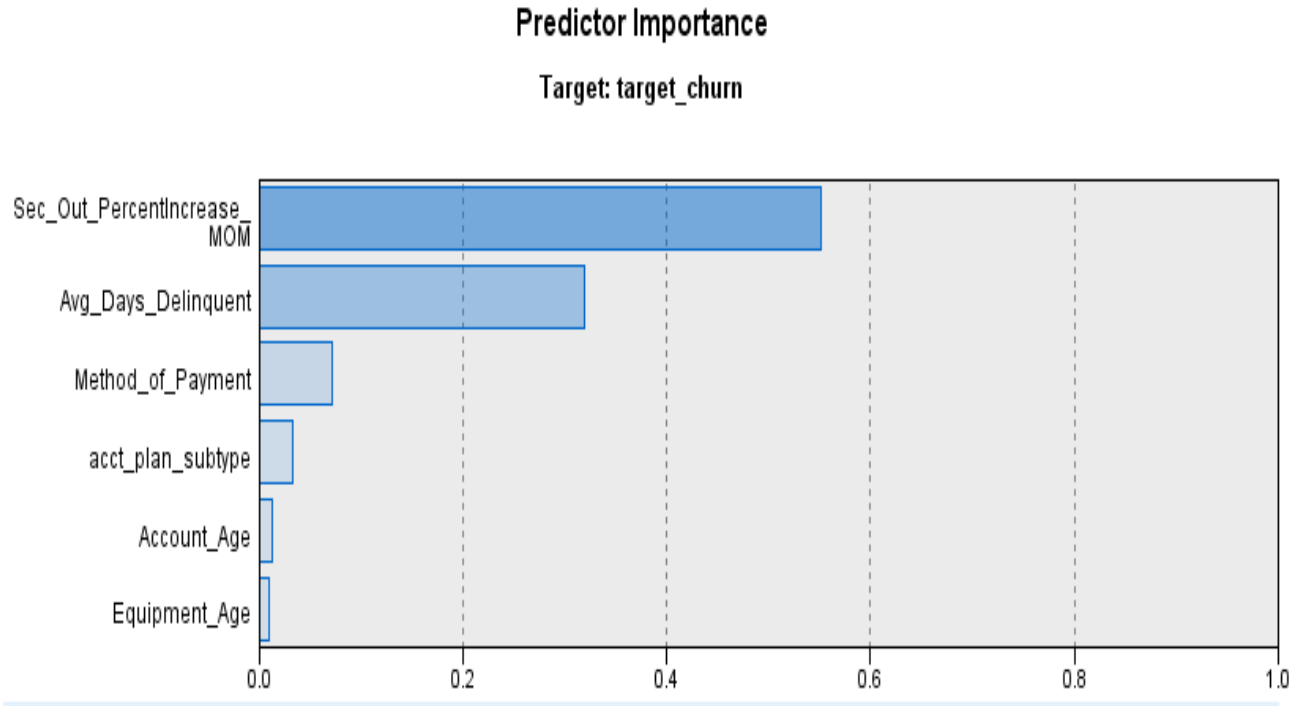
Use?	Graph	Model	Build Time (mins)	Max Profit	Max Profit Occurs in	Lift(Top 3...	Overall Accuracy	No. Fields Used	Area Under
☑		Neural Net 1	< 1	6,945.0	47	1.833	80.799	7	0.882
☑		CHAID 1	< 1	6,565.37	50	1.813	79.333	5	0.871
☑		C&R Tree 1	< 1	6,564.268	50	1.777	79.312	6	0.845

As shown above, overall accuracy of ANN is higher than CHAID and C&RT. The last two models have almost equal accuracies.

Decision Tree (CHAID) Analysis Results

Figure 5 shows the important variables in classifying customer churn according to CHAID algorithm.

Figure 5. Important Variables Determined By CHAID



In Figure 6, confusion matrix of CHAID indicates the accuracy of the classification of the algorithm.

Figure 6. CHAID Confusion Matrix

Comparing $R-target_churn with target_churn

'Partition'	1_Training		2_Testing	
Correct	2,627	80.21%	1,123	78.37%
Wrong	648	19.79%	310	21.63%
Total	3,275		1,433	

Coincidence Matrix for $R-target_churn (rows show actuals)

'Partition' = 1_Training	0	1
0	1,361	316
1	332	1,266

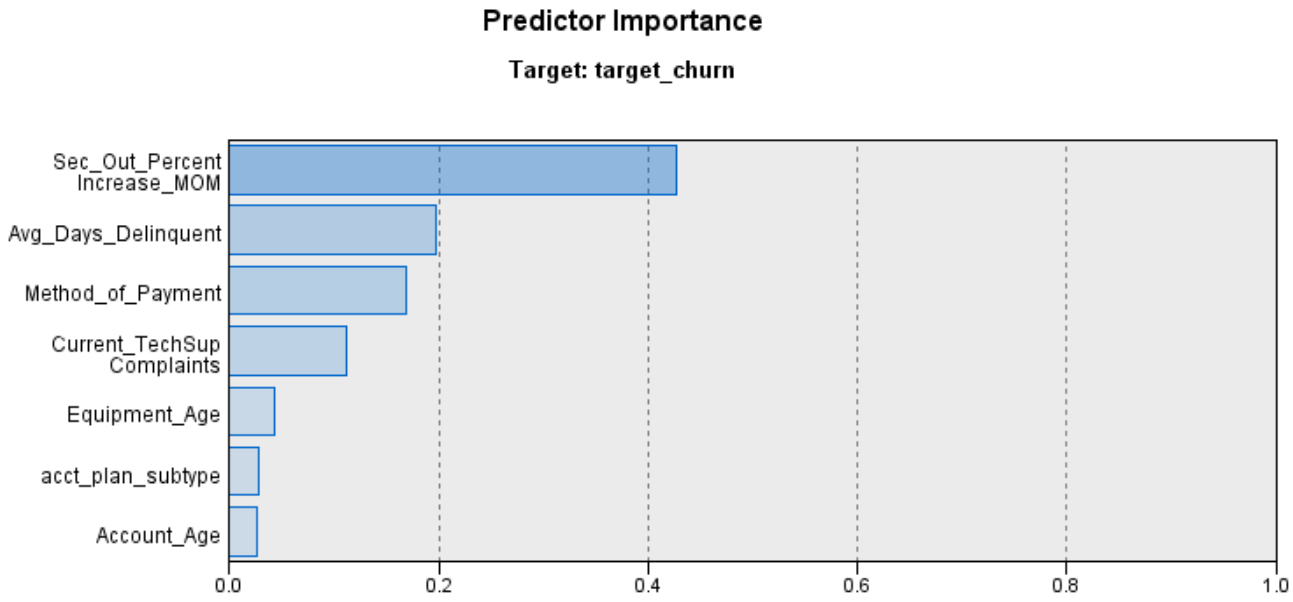
'Partition' = 2_Testing	0	1
0	599	145
1	165	524

Classification accuracy based on validation (testing) dataset for the non-churned customers is 80.5% and for the churned customers is 76%.

Artificial Neural Network (ANN) Results

Figure 7 displays the important variables in classifying customer churn according to ANN algorithm.

Figure 7. ANN Important Variables



In Figure 8, confusion matrix of ANN indicates the accuracy of the classification of the algorithm.

Figure 8. ANN Confusion Matrix

Comparing $N-target_churn with target_churn

'Partition'	1_Training		2_Testing	
Correct	2,521	76.98%	1,122	78.3%
Wrong	754	23.02%	311	21.7%
Total	3,275		1,433	

Coincidence Matrix for $N-target_churn (rows show actuals)

'Partition' = 1_Training	0	1	$null$
0	1,198	479	0
1	273	1,323	2

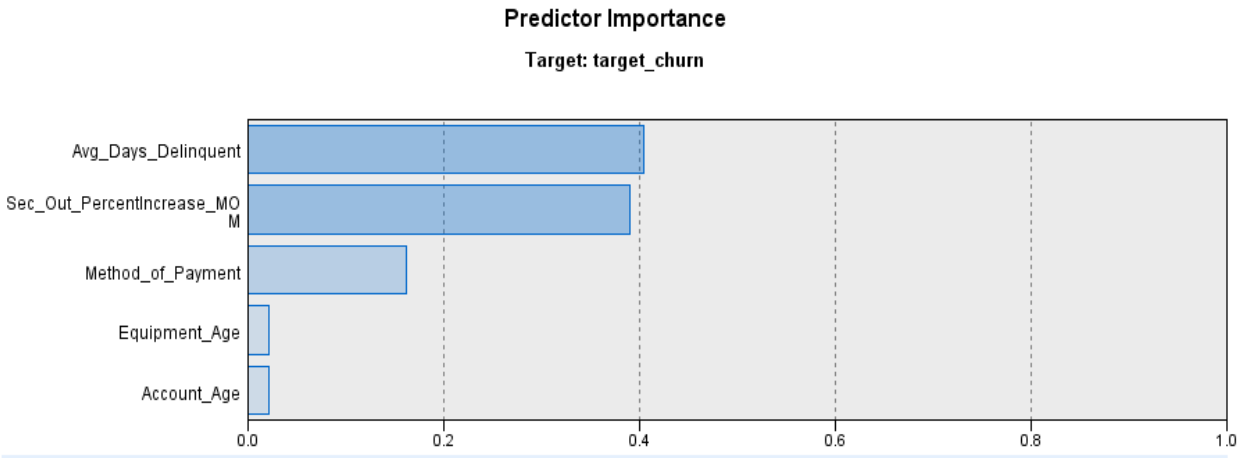
'Partition' = 2_Testing	0	1	$null$
0	554	188	2
1	121	568	0

Classification accuracy based on validation (testing) dataset for the non-churned customers is 74.7% and for the churned customers is 82.4%.

Classification (C&RT) Analysis Results

Figure 9 shows the important variables in classifying customer churn according to C&RT algorithm.

Figure 9 Important Variables Determined By C&RT



In Figure 10, confusion matrix of C&RT indicates the accuracy of the classification of the algorithm.

Figure 10. C&RT Confusion Matrix

Comparing $R-target_churn with target_churn

'Partition'	1_Training		2_Testing	
Correct	2,613	79.79%	1,121	78.23%
Wrong	662	20.21%	312	21.77%
Total	3,275		1,433	

Coincidence Matrix for $R-target_churn (rows show actuals)

'Partition' = 1_Training	0	1
0	1,211	466
1	196	1,402

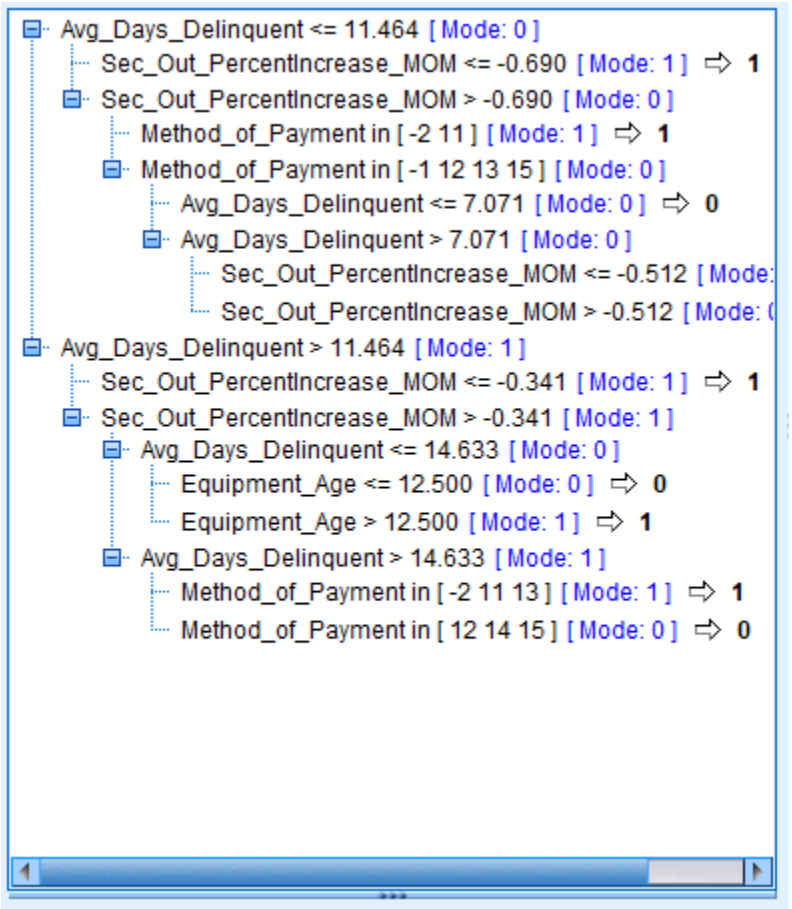
'Partition' = 2_Testing	0	1
0	536	208
1	104	585

Classification accuracy based on validation (testing) dataset for the non-churned customers is 72% and for the churned customers is 84.9%.

CHAID algorithm created 13 rules set for classifying non-churned customers and 21 rules set for classifying churned customers. While C&RT algorithm created 4 rules set for classifying non-churned customers and 6 rules set for classifying churned customers. In case of developing a decision support system or an expert system, some of the rules that are redundant or are not used by decision makers should be removed or consolidated.

Figure 11 displays example of C&RT rules for churn.

Figure 11. Example of Rules for Churn



CONCLUSIONS

The performance of Artificial Neural Network (ANN) in classifying non-churned customers and churned customers (74.7%, 82.4%) is better than performance of Decision Tree (CHAID) in classifying non- churned customers and churned customers (80.5%, 76%). ANN performance is a little better than C&RT in classifying non- churned customers and churned customers (72%, 84.9%) too. In addition, CHAID and Classification & Regression Trees (C&RT) algorithms create rule sets that may be used to develop a decision support or an expert system.

Comparing the results of this study with two previous studies by this author indicates the following: Applying Decision Tree (C5) and Artificial Neural Network (ANN) to an IBM dataset in 2018 showed that performance of C5 in classifying non-churned customers and churned customers (96%, 98%) is better than performance of ANN in classifying non- churned customers and churned customers (89%, 84%).
Applying a number of classification algorithms to the same IBM dataset in 2017 showed that overall accuracy of C5 in classifying non-churned customers and churned customers (89.2%) is better than overall accuracies of CHAID (81.4%) and ANN (81.25%) in classifying non- churned customers and churned customers.

In future studies, Revenue Churn may be added to this study. Revenue Churn is different from Customer Churn. Customer Churn refers to the number of customers that have discontinued their subscription on a

given period. Revenue Churn (usually referred as MRR Churn), is how much those lost customers represents in revenue.

REFERENCES

Lu, Junxiang, "Predicting Customer Churn in the Telecommunications Industry — An Application of Survival Analysis Modeling Using SA", SUGI 27 Conference of SAS, Orlando, FL, 2002.

Lu, Junxiang," Detecting Churn Triggers for Early Life Customers in the Telecommunications Industry – An Applications of Interactive Tree Training," Proceedings of the 2nd Data Mining Conference of DiaMondSUG 2001, Chicago, IL, 2001.

Wayne Thompson and David Duling, "Rapid Predictive Modeling for Customer Intelligence", SAS Global Forum 2010, Cary, NC, SAS Institute Inc.

Sascha Schubert, Susan Haller and Taiyeong Lee, "It's About Time: Discrete Time Survival Analysis Using SAS® Enterprise Miner™", SAS Global Forum 2012, Cary, NC, SAS Institute Inc..

Wielenga, Doug. 2007. "Identifying and Overcoming Common Data Mining Mistakes." Proceedings of the SAS Global Forum 2007 Conference. Cary, NC: SAS Institute Inc.

Potts, W. 2005. "Predicting Customer Value." Proceedings of the SAS Global Forum 2005 Conference. Cary, NC: SAS Institute Inc. Available at: http://www2.sas.com/proceedings/sugi30/073-30.pdf

Carlos Andre Reis Pinheiro, Markus Helfert, "Creating a customer influence factor to decrease the impact of churn and to enhance the bundle diffusion in telecommunications based on social network analysis", Proceedings of the SAS Global Forum 2010, Cary, NC, SAS Institute Inc.

Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Navati, A.,"Social ties and their relevance to churn in mobile telecom networks." In: 11th International Conference on Extending Database Technology: Advances in Database Technology. No. 261, pp. 668-677 (2008).

Nkululeko Ngcongo, "Finding The Best Statistical Model To Predict Customer Defection In Telecommunication Retail Setting" A Research Report submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science in Mathematical Statistics, February 11, 2014.