

Churn Prevention in Telecom Services Industry- A systematic approach to prevent B2B churn using SAS

Krutharth Peravalli, Dr. Dmitriy Khots

West Corporation

ABSTRACT

“It takes months to find a customer and only seconds to lose one” - Unknown. Though Business-to-Business (B2B) churn problem might not be as common as Business-to-Consumer (B2C) churn, it has become crucial for companies to address this effectively as well. Using statistical methods to predict churn is the first step in the process of retaining customers which also includes model evaluation, prescriptive analytics including outreach optimization and performance reporting. Providing visibility into model and treatment performance allows the data and Ops teams to tune models and adjust treatment strategy.

WEST Corporation's Center for Data Science (CDS) has partnered with one of the lines of business in order to measure and prevent B2B customer churn. CDS has coupled firmographic and demographic data with internal CRM and past outreach data to build a propensity to churn model using SAS. CDS has provided the churn model output to an internal Client Success Team (CST), who focuses on high risk / high value customers in order to understand and provide resolution to any potential concerns that may be expressed by such customers. Furthermore, CDS automated weekly performance reporting using SAS that not only focuses on model statistics, but also on CST actions and impact. This paper focuses on all of the steps involved in churn prevention process including building and reviewing the model, treatment design and implementation, as well as performance reporting.

INTRODUCTION

WEST Corporation Outbound Notification Platform

West Corporation's outbound notification platform is used by Healthcare companies to notify their patients about upcoming doctor appointments. Additionally, patients are provided with the flexibility to confirm, cancel or reschedule their appointments using the same notification platform. West uses its outbound notification platform to send Voice, SMS and Email notifications to its customers. Customers such as healthcare companies can configure the number of notifications, types of notifications, etc., within the platform. Notification process is initiated with loading of files consisting of customer's phone number and email address into outbound platform using SFTP. Once files are loaded, all records are scrubbed against “do not call” lists and other filtering criteria and notification date/time/type are configured within the notification platform.

DATA MINING PROCESS

Business Understanding

Over the years, many healthcare companies have subscribed to use West's powerful outbound notification platform to notify their patients. West offers various types of notifications (products) depending on customers' needs and one of the most popular type of notification is Housecalls. Housecalls are outbound messages to engage patients, including appointment confirmations, recalls and follow-ups, no-show rescheduling, health outreach and prevention, etc. Alerts are delivered via voice phone calls, email messages and two-way SmartText mobile text messages. Each customer can subscribe to multiple products and hence separate contracts are created for each product. Every year, West's account base has been increasing through the successful actions of the field sales teams; however, this growth has been staved off due to some churn that may not have been as evident given overall growth. CDS was involved in performing descriptive analysis of the Housecalls churn rate and as a result suggested to build a propensity to churn predictive model to identify the high probability to churn accounts.

Data Understanding

Propensity to churn predictive model has been built using account data from January 2013 to June 2015. Variables used for building the model are collected from different data sources and are classified into one of the below categories:

Customer Specific Variables: Account related variables provided by client such as customer's Account Class and US Census Region derived from customer's state of residence.

Account-Product Variables: Account-Product related variables maintained by West such as Product Number, Product Type, Product Status, Contract Dates, Contract Pricing, etc.

Notifications Related Variables: Historical notification related variables such as Total Notification Volume, Successful Notification Volume, and Notification Type for each of the Account-Products.

Customer Billing Variables: Billing related variables such as Minimum Notification Units and Minimum Amount.

Revenue Variables: Monthly revenue numbers for each account along with number of providers.

Incident Specific Variables: Information regarding support tickets opened by customers is captured in variables such as Incident Type, Number of Days Open, Status Code, Priority Code, Incident Origin, etc.

Demographic Variables: Demographic variables like Average Household Size, Average Household Income, Presence of children in the household, etc. are input to the model. This data is available to West from a third party vendor.

Dynamic Variables: In addition to the above variable categories, dynamic variables are created for each account that show the account snapshot at the end of each month in the analysis time frame.

Data Preparation and Target Variable Identification

Step 1: Pull the Account-Product data from the database for all Housecalls products.

Step 2: Join Account-Product data from Step 1 with Account data to pull all the account specific variables.

Step 3: Using the dataset in Step 2, create Dynamic Variables for each account that show the number of active products, inactive products, average days active, active products in last N months, , etc. for each month in the analysis time frame.

Step 4: Pull Notifications Related Variables and create dynamic variables for each account such as total volume in last N months, successful volume in last M months, etc. for each month in the analysis time frame.

Step 5: Pull Customer Billing Variables from the database.

Step 6: Pull Monthly Revenue for each account from the database.

Step 7: Pull Support Ticket Specific Variables and create historical variables for each account such as total tickets in last N months, active tickets in last N months, resolved tickets in last N months, cancelled tickets in last N months, etc. for each month in the analysis time frame.

Step 8: Join the datasets from Steps 3-7 and create one final dataset that has the variables for each account for each month in the analysis time frame

Step 9: Join data from Step 8 with demographic information available at West using Zip9 or Zip5 codes of the accounts.

Step 10: Target Variable Creation - If an account has all inactive products, then target=1 else target=0 (Binary Target).

Predictive Modeling in SAS Enterprise Miner

Sampling: Prepared dataset is split into Model (70%) and Holdout (30%) sample using PROC SURVEYSELECT procedure with option of Simple Random Sampling technique available in SAS.

```
/*Create Holdout Data*/
```

```
Proc surveyselect data=FinalData  
method=srs n=XXXX  
seed=YYYY out=FinalData_Holdout;  
run;
```

```
/*Create Model Data*/
```

```
Proc SQL;  
Create table FinalData_Model as  
Select * from FinalData where account_id not in  
(Select distinct account_id from FinalData_Holdout);  
Quit;
```

Imputations & Transformations: Model Data is then loaded into SAS Enterprise Miner. Data Partition node is used to split the data into Train (70%) and Test (30%) sample. Less than 1% of data has missing values and they are imputed using impute node. Interval Variables are imputed by replacing missing information with mean value and Class variables are imputed by replacing missing information with mode value of the respective variable. Best Transformation technique has been used on Interval variables and dummy indicators are created for class variables using Transform Node available in SAS Enterprise Miner.

Model Building and Evaluation: Decision Tree, Logistic Regression and Neural Network techniques were tried on the transformed dataset available from the previous step. Variables with insignificant chi-square values have been removed allowing significant variables to be selected by the model. The output from the three model nodes is then fed into the Model Comparison node in SAS EM. Logistic Regression technique with Stepwise Selection turned out to be the best modeling technique based on KS Statistic. Final Model has a KS of 46. Stability of the model is tested by running the final model on Test Data (KS=45). Figure 1 shows the SAS Enterprise Miner screenshot of the entire data and model flow.

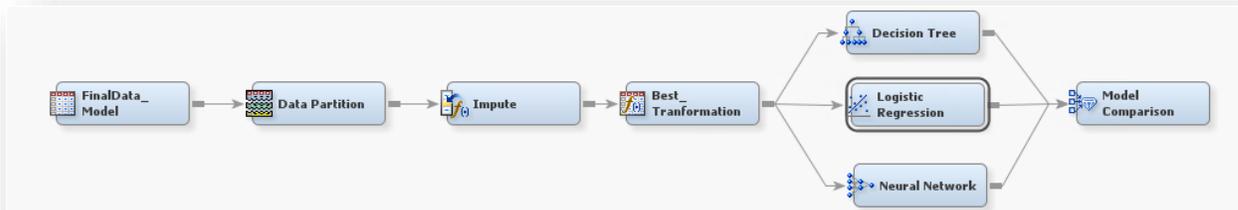


Figure 1. Model Diagram

Logistic Regression Results

The logistic regression module used the Chi Square test to evaluate variable importance and selected N% of available predictors into the final model. Variables that are most important in predicting churn were the dynamic variables, especially those involved in historical issue resolution, while pricing variables turned out to be least important. Figure 2 lists a redacted variable output and chi-square tests for each variable.

Variable Name	DF	Number In	Chi-Square	Pr > ChiSq
Variable1	2	1	700.7762	<.0001
Variable2	2	2	532.2973	<.0001
Variable3	2	3	460.4003	<.0001
Variable4	2	4	185.9131	<.0001
Variable5	1	5	106.6098	<.0001
Variable6	1	6	104.3332	<.0001
Variable7	1	7	30.6424	<.0001
Variable8	1	8	15.0838	0.0001
Variable9	1	9	10.1462	0.0014
Variable10	1	10	11.1623	0.0008
Variable11	1	11	10.4678	0.0012
Variable12	2	12	11.2743	0.0085
Variable13	1	13	10.2543	0.041
Variable14	1	14	9.9542	0.019

Figure 2. Variable List

Cumulative Lift Chart - Using Proc Rank procedure in SAS, each account is assigned a decile based on their probability to churn score. Cumulative Lift charts are used to demonstrate the model performance on the train and test dataset. From the cumulative lift charts of Train (Figure 3) and Test (Figure 4) datasets it is evident that the top 4 deciles contribute to 85% of the churns.

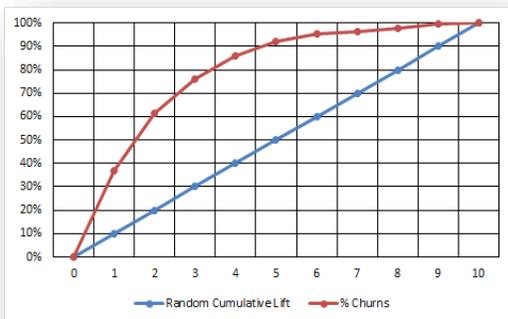


Figure 3. Train Lift Chart

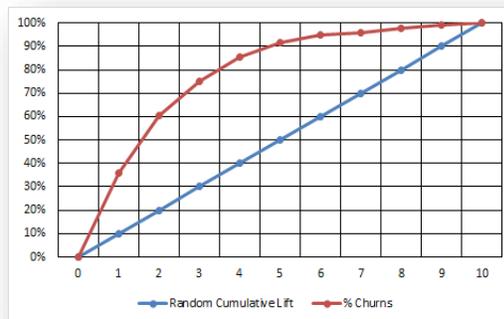


Figure 4. Test Lift Chart

Scoring/Deployment process

All of the variables that are created in the data prep part of the model building process are created for the new accounts before they are scored. After creating all variables for new accounts, score code is created using SAS Enterprise Miner Score node. As part of the process, each account is assigned a decile based on its propensity to churn. Accounts in the top 5 deciles that were not touched by West's Client Success Team (CST) previously are selected for outreach to save customers from churning. From the selected accounts, Test List (90%) and Control List (10%) are created by randomly selecting accounts using PROC SURVEYSELECT. Accounts in Control List are not touched by Client Success Team and are used to measure impact of CST activities.

Performance Reporting

It is imperative to measure performance once the model is built and deployed. Measuring model performance, utilization, and possible decay in a timely manner allows making necessary improvements to treatment design and strategy to improve churn mitigation. CDS provides monthly account scoring in order to reflect changes to dynamic attributes and distributes a performance evaluation report that focuses on model statistics and Client Success Team impact. Using various SAS macros and procedures, CDS has created and automated a weekly performance evaluation report that highlights various KPIs such as the CST Worked and Non-Worked (test and control) accounts churn rates, revenue churn rates, accounts saved, revenue saved, etc. In particular, the performance dashboard contains the following elements:

Churn Summary Report. This report shows the year-over-year churn comparison break down by CST worked and non-worked accounts. Additional saved accounts and revenue saved are calculated using the worked accounts churn and revenue churn rates.

Additional Saved Accounts= (# of Worked Accounts + # of Non-Worked Accounts) / (Non-Worked Accounts Churn Rate-Worked Accounts Churn Rate)

Additional Revenue saved= (Monthly \$ Amount Worked + Monthly \$ Amount Non-Worked) / (Non-Worked Monthly Revenue Churn Rate – Worked Monthly Revenue Churn Rate)

Figure 5 shows the churn rates comparison between 2015 and 2016 along with additional saved accounts and saved revenue

	2015	2016
CST Worked Accounts		
CST Not Worked Accounts		
CST Worked Accounts Churn Rate	10.53%	6.50%
CST Not Worked Accounts Churn Rate	19.58%	17.54%
Additional Saved Accounts		
Additional Saved Revenue		
Additional Annual Revenue Saved		

Figure 5. Churn Summary

List Evaluation Report. For each set of lists (Test and Control), the SAS evaluation system publishes a report showing churn rates, revenue churn rates, saved accounts and revenue saved. Additional saved accounts and revenue saved are calculated using the same logic as stated in Churn Summary Report section. This report provides a comparison between Test and Control lists. Figure 6 shows churn rates, revenue churn rates, saved accounts and saved revenue for both test and control lists.

	# Accounts Worked by CST	# Accounts Non-Worked by CST	# Churned Accounts Worked by CST	# Churned Accounts Non-Worked by CST	\$ Monthly Revenue Worked by CST	\$ Monthly Revenue Non-Worked by CST	\$ Churned Monthly Revenue Worked by CST	\$ Churned Monthly Revenue Non-Worked by CST	Worked Accounts Churn Rate	Non-Worked Accounts Churn Rate	Worked Monthly Revenue Churn Rate	Non-Worked Monthly Revenue Churn Rate
Test									7.1%	32.0%	18.3%	26.9%
Control									0.0%	15.0%	0.0%	13.3%
Test												
Additional Saved Accounts												
Additional Saved Revenue												
Control												
Additional Saved Accounts												
Additional Saved Revenue												

Figure 6. List Evaluation Report

Churned Accounts List. This report contains the list of churned accounts along with the names of the inactive products, Client Success Team Agent Name, Cancellation Date, Average Revenue, etc. Accounts churned in the past 1 week are appended to the existing list. Figure 7 shows the list of churned accounts along with the products cancelled, CST Member Name, etc. This report is used by CST to evaluate missed opportunities.

Account Number	Account Name	Cancellation Date	CST Member	Average Revenue	InactiveProduct_1	InactiveProduct_2	InactiveProduct_3
		23-Sep-2015			HouseCalls ASP		
		29-Jul-2015			HouseCalls	HouseCalls ASP	HouseCalls ASP Unlimited
		31-Jan-2016			HouseCalls ASP	HouseCalls ASP Unlimited	

Figure 7. Churned Accounts List

CST Cases Overview. This report shows the number of cases each Client Success Team Member has opened for accounts and the status of the accounts thus providing an insight on the Members with high churn rates. Figure 8 shows the number of cases opened for each account and the status of the account.

Account Number	CST Member Name	# of Cases	Churned
1111	Agent1	1	Y
2222	Agent1	2	Y
3333	Agent2	1	N
4444	Agent3	2	Y
5555	Agent3	1	N

Figure 8. CST Cases Summary

CST Churn Rate Effectiveness Scorecard. This is a scorecard showing the churn rate effectiveness of all Housecalls Accounts, Model effectiveness chart, CST Contacted Accounts overview and Month-over-Month Housecalls Product Cancellations graph.

Figure 9 shows weekly Worked Accounts Churn Rate vs Non-Worked Accounts Churn Rate for all of the Housecalls Accounts

Figure 10 shows weekly Worked Accounts Churn Rate vs Non-Worked Accounts Churn Rate for all of the lists provided to Client Success Team. Weekly performance of CST on individual lists can be assessed using this graph.

Figure 11 shows # of overall CST Contacted Accounts and # of Accounts Churned after Contact.

Figure 12 shows a graph that highlights Month-over-Month Housecalls Product Cancellations.

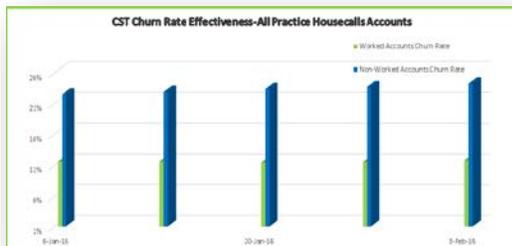


Figure 9. CST Churn Rate Effectiveness-All Housecalls Accounts



Figure 10. Model Churn Rate Effectiveness



Figure 11. CST Contacted Accounts Overview

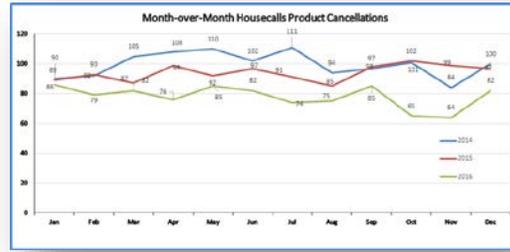


Figure 12. Month-over-Month Product Cancellations Graph

Figure 12 is especially important showing overall year-over-year impact of the joint efforts of the predictive modeling group and CST on churn mitigation.

CONCLUSION

Center for Data Science team at West not only uses SAS Enterprise Miner and Predictive Analytics for building robust predictive models but also uses the powerful SAS Macros effectively to generate and automate performance reports. Propensity to Churn model built by CDS has been performing consistently well for over one and half years. Using the weekly Performance reports, CDS is able to track the model performance as well as help CST track their agents' performance in a timely manner. Since model deployment, CDS has continuously worked with CST in designing a better treatment strategy based on the team's performance. Scorecard in the performance report especially helps executives in assessing the overall CST performance and provides a high level churn overview.

ACKNOWLEDGMENTS

The authors would like to thank Shruti Palasamudram for her guidance and suggestions throughout this study. The authors would also like to thank Dr. Goutam Chakraborty for inviting the team to present this topic at SAS Global Forum 2017.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please contact the authors at:

Krutharth Peravalli, West Corporation

E-mail: kperaval@west.com

Dr. Dmitriy Khots, West Corporation

E-mail: dkhots@west.com