Customer Churn

Chris Berardi

October-November 2018

Predicting Customer Churn

The ability to accurately predict customer churn, that is, to predict which customers will cease to use a service, or a product, is of extreme importance to any business. Predicting churn allows for attempts to prevent churn. However, churn is generally difficult to predict for a variety of reasons, the most important being that churn is generally less common than non-churn. For this reason most model fitting will greatly underestimate churn, a non-optimum solution since churn is of such great importance. Thankfully there exist methods to improve the fit of minority binary classifiers. However the improved ability to predict real churn comes at the cost of greatly increased false churn prediction–incorrectly classifying non-churn events as churn events. But this is the price that must be paid to model churn.

The Data

The data used for this report contains information for 7043 customers each with 20 attributes and one target. The

The Target

Churn
No: 5174
Yes: 1869

The unbalanced nature of the binary target necessitates rare event modeling.

One attribute, CustomerID was not used to model the data as it was a unique ID for each customer and thus contains no information. The table below describe each attribute by data type.

Binary Attributes

Gender	Senior Citizen	Partner	Dependent	Phone Service	Paperless Billing
Male	0	Yes	Yes	Yes	Yes
Female	1	No	No	No	No

The binary attributes used in this report have been inconsistently coded, Senior Citizen uses an indicator variable in contrast to the Yes/No distinction for the other binary attributes. The binary attributes collected tells us various demographic and service related information about each customer.

Nominal Attributes

Multiple Lines	Internet Service	Online Security	Online Backup	Device Protection
Yes	Fiber Optics	Yes	Yes	Yes
No	DSL	No	No	No
No Phone Service	No	No Internet Service	No Internet Service	No Internet Service

Tech Support	Streaming TV	Streaming Movies	Contract	Payment Method
Yes	Yes	Yes	Month-to-Month	Bank Transfer (automatic)
No	No	No	One year	Credit Card (automatic)
No Internet Service	No Internet Service	No Internet Service	Two year	Electronic Check
				Mailed Check

All but one nominal attribute has three levels, Payment Method having four. These nominal attributes deal with various services offered by this telecom company.

Interval Attributes

Tenure	Monthly Charges	Total Charges
Time as Customer Months	Amount Owed per Month Dollars	Total Amount Paid as Customer Dollars

The data contains very few interval variables. However the interval variables that are included are of great importance since they are critical for defining the amount of money lost to churn, and the amount of money saved by the model.

The Model

As the ration of churn to non-churn events was 1:3, none rare event modeling was attempted to predict churn. This approach failed to adequately model churn, predicting far less churn than existed, and therefore failed in its primary purpose. For this reason rare event modeling with random undersampling was then carried out.

Using standard goodness of fit measure for binary classification, sensitivity, specificity and the like, would be not appropriate in this case. For that reason a savings function was created to assess the goodness of fit for models. The savings function was defined as follows: Total Savings = Amount of Churn Predicted - Cost to Prevent Churn - Amount of Churn not Predicted. The variables are defined as follows:

Amount of Churn Predicted: the sum of monthly charges for all churn events predicted.

Cost to Prevent Churn: a theoretical assumption that it will cost 25% of the monthly cost to prevent a churn event. Cost to prevent churn is then the sum of this amount for all churn predictions.

Amount of Churn not Predicted: the sum of monthly charges or all churn events incorrectly classified as non-churn events.

The best model will then predict the most churn correctly while not falsely predicting churn. The savings function will prioritize large amounts of churn over predicting more churn.

Given the nature of the data, a decision tree model was selected to model churn. The first step in training the model was determining the optimum RUS ratio for this data. The following ratios were used:

Ratio for RUS (Majority:Minority)
30:70
40:60
50:50
60:40
70:30

The optimum RUS ratio was then used to hyperparameterize a decision tree model. Each depth was run 1000 times to generate average and standard deviations for the savings at each depth. The following depths were used:

Decisions Tree Depths
5
6
7
8
10
12

Results

The optimum ratio for RUS was found to be 30:70. This model had the highest missclassification rate, but it also had the highest rate of correct churn prediction. The properties of the model are as follows:

30:70 RUS	
Missclassification Rate	.3108
Churn Classified as Non-Churn Cost	\$9001
Non-Churn Classified as Churn Cost	\$28687
Cost to Prevent Churn	\$26026
Savings from Preventing Churn	\$130129
Total Cost	\$63715 ± \$728.84
Net Savings	\$66415 ± \$1150.64

Using this 30:70 RUS ratio, hyperparameter optimization was conducted. A depth of 7 was found to be the optimum depth for this data. Following that, ensemble modeling was done to create a final model for this data. 1000 trees were created and averaged to produce the following results:

Ensemble Model Results	
Missclassification Rate	.3135
Churn Classified as Non-Churn Cost	\$906
Non-Churn Classified as Churn Cost	\$31210
Cost to Prevent Churn	\$27645
Savings from Preventing Churn	\$128225
Total Cost	\$59761
Net Savings	\$78464

Extracting feature importance from the ensemble models, the following predictors were found to be disproportionately important in predicting customer churn:

Important Features in the Ensemble Model
Contract
TotalCharges
MonthlyCharges

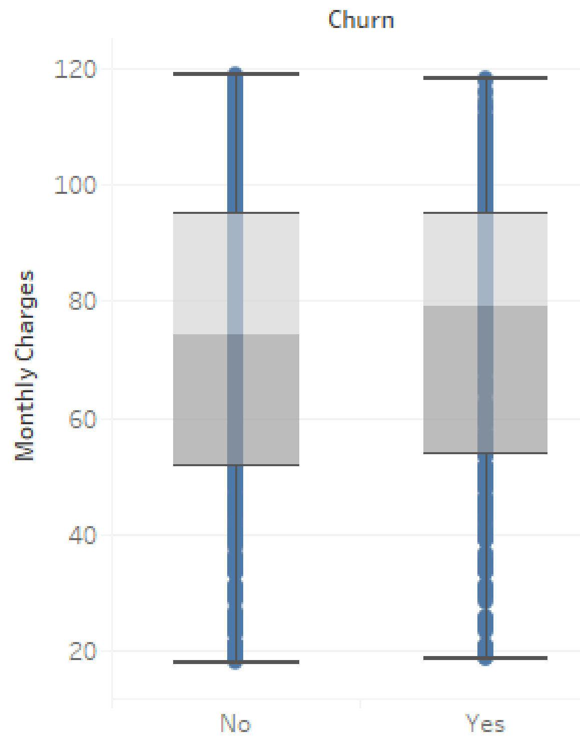
Conclusions

Various conclusions can drawn from the model. We can first see from the missclassification error rate that the model, essentially, predicts twice as many churn events as actually exist. This atrocious error rate is necessary to predict as many of the actual churn events as possible. However, this level of missclassification allowed for all but $906 out of $129131, .7%, of churn to be accurately classified. However classifying churn as well as that brings in a significant cost: $31210, 9.85% of total non-churn payments, spent to prevent churn for customers that would not churn. This cost allows for far more money, $78464, to be saved once churn is prevented.

Of more important is the identification of features that give rise to churn, specifically: Contract, TotalCharges, MonthlyCharges. The following charts show the relationship between the important features and churn.

Contract Type by Churn

Monthly Charge by Churn



Total Charges by Churn

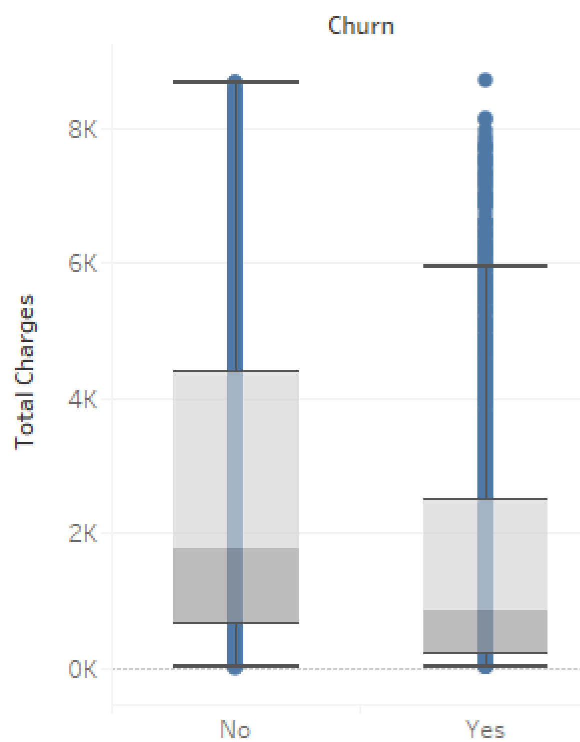


Figure 1 on the previous page contains plots that shows the relationship between the most important features and the target for this analysis. Conclusions, as well as possible methods for mitigating churn, will be draw from each plot.

Starting with the most important predictor, Contract, we can see a clearly, in the pie charts, distinctions between those with contracts, and those going month-to-month, specifically, those going month-to-month are far more likely to churn. This implies that one method of decreasing churn will be to increase contract use among customers. If possible, this would greatly decrease churn since month-to-month customers make up a majority of customers.

Continuing on to TotalCharges, we can see from the box-plot that the individuals that churn are, generally, those who have spent less with this company over time. The reasons for this counterintuitive result is consistent with the previous findings that month-to-month customers are more likely to churn. The conclusions that can be drawn from this plot on how to prevent are thus the same: convince customers to sign contracts to stay with the company longer, thus making them spend more with the company over time.

The final important feature, MonthlyCharges, is the last box-plot on the preceding page. Customers that churn spend, on average, more than those who don't. While we expect customers that pay the most to want to find different service, the reason why they pay more must be found. The interpretation of this plot is made clear when the contract plot is considered.As has been already established, customers that churn generally do not sign contracts, increasing the average amount they pay. This accounts for most of the increased charges.

Taken together, the model, as well as the feature importance, show a path that can be taken to minimize future churn. Work must go into convincing customers to switch from month-to-month to one-year or two-year contracts. This will decrease monthly charges as well as increase total charges i.e. make them customers for a longer period of time. Furthermore, more work needs to be done to determine the monthly charge under which customer cease churning at high rates. This amount can inform possible remediation to prevent churn in month-to-month customers.