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

Multiple Lines	Internet Service	Online Security	Online Backup	Device Protection
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

Conclusions