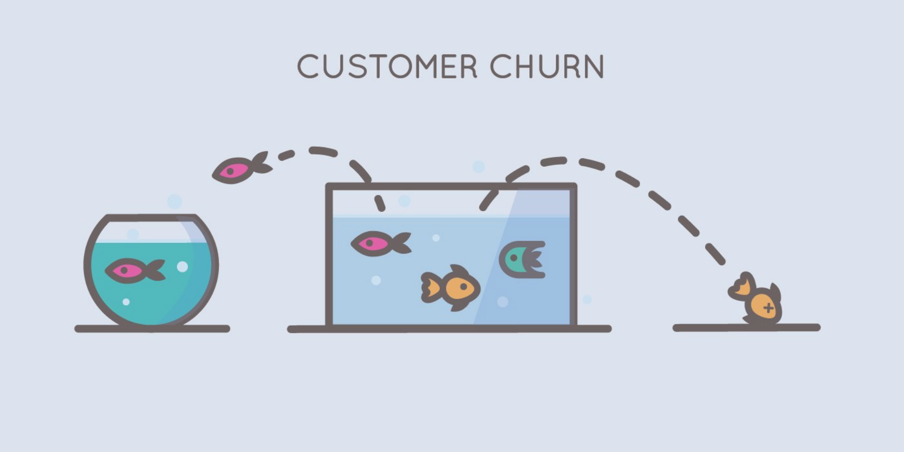
Predicting and Preventing Churn of High Value Customers



Analysis and Presentation by Jeremy Silva

The Problem

Teleworld, a mobile phone and telecommunication company, has been struggling with a high Churn rate for the past couple of years.

They are looking for a way to reduce their Churn rate and retain more customers.

The Proposed Business Solution

In order to solve Teleworld's problem we need to accomplish two major things:

1) Develop a way for Telecom to identify customers who are at high risk of Churning so Teleworld can intervene with retention incentives.

2) Identify and rank which factors are contributing most heavily to Churn.

The Proposed Data Mining Solution

We will seek to accomplish the two previously proposed business solutions with the follow two data mining tasks:

1) Develop a classification model which can predict whether or not a customer will Churn in the coming three months.

2) Use coefficient analysis or some other means to Identify and rank the factors which contribute to Churn.

The Data

Telecom has provided us a dataset with information on 3100 randomly selected customers.

After cleaning the dataset looks like so…

	Call Failure	Complains	Subscription Length	Charge Amount	Seconds of Use	Frequency of use	Frequency of SMS	Distinct Called Numbers	Age Group	Tariff Plan	Status	Customer Value	Churn
0	8	0	38	0	4370	71	5	17	3	1	1	197.640	0
1	0	0	39	0	318	5	7	4	2	1	2	46.035	0
2	10	0	37	0	2453	60	359	24	3	1	1	1536.520	0
3	10	0	38	0	4198	66	1	35	1	1	1	240.020	0
4	3	0	38	0	2393	58	2	33	1	1	1	145.805	0

Final Modeling Results

After many modeling iteration I settled on a Random Forest Classifier using Synthetic Minority Oversampling to balance the classes of the minority class (Churn Cases) during training. Classification Report Below:

Classification Report for FINAL MODEL: Random Forest with 0.5 SMOTE Ratio Model

	precision	recall	f1-score	support
0	0.98	0.97	0.98	797
1	0.84	0.91	0.87	148
accuracy			0.96	945
macro avg	0.91	0.94	0.92	945
weighted avg	0.96	0.96	0.96	945

Notes: Results are from unseen TESTING DATA

Modeling Iterations

I tried a number of different models and a number of different class balancing strategies

Models Tried

Linear: Logistic Regression

Ensemble Tree: Random Forest, Gradient Boosting, Ada Boost

Class Balancing Techniques

Oversampling Minority Class: SMOTE, ADASYN (both with various class ratios)

Modeling Iterations Cont...

I tried each model with each oversampling techniques and added the results to a dataframe. Pictured here:

	Model Type	Avg Accuracy	Churn Class Precision	Churn Class Recall	Churn Class F1-Score
0	Logistic Regression	0.90	0.81	0.43	0.56
1	RandomForestClassifier	0.95	0.93	0.76	0.83
2	GradientBoostingClassifier	0.94	0.88	0.75	0.81
3	AdaBoostClassifier	0.93	0.81	0.71	0.76
4	RandomForestClassifier with SMOTE'> 0.25	0.96	0.90	0.80	0.85
5	RandomForestClassifier with SMOTE'> 0.5	0.96	0.86	0.89	0.87
6	RandomForestClassifier with SMOTE'> 0.75	0.96	0.81	0.93	0.87
7	RandomForestClassifier with SMOTE'> auto	0.95	0.79	0.92	0.85
8	RandomForestClassifier with ADASYN'> 0.25	0.95	0.91	0.78	0.84
9	RandomForestClassifier with ADASYN'> 0.5	0.95	0.81	0.92	0.86
10	RandomForestClassifier with ADASYN'> 0.75	0.95	0.79	0.93	0.86
11	RandomForestClassifier with ADASYN'> auto	0.95	0.77	0.95	0.85
12	GradientBoostingClassifier with SMOTE'> 0.25	0.95	0.88	0.77	0.82
13	GradientBoostingClassifier with SMOTE'> 0.5	0.93	0.74	0.89	0.81
14	GradientBoostingClassifier with SMOTE'> 0.75	0.93	0.70	0.93	0.80
15	GradientBoostingClassifier with SMOTE'> auto	0.91	0.65	0.95	0.77
16	GradientBoostingClassifier with ADASYN'> 0.25	0.94	0.88	0.71	0.79
17	GradientBoostingClassifier with ADASYN'> 0.5	0.93	0.74	0.89	0.80
18	GradientBoostingClassifier with ADASYN'> 0.75	0.93	0.71	0.92	0.80
19	GradientBoostingClassifier with ADASYN'> auto	0.92	0.66	0.96	0.78
20	AdaBoostClassifier with SMOTE'> 0.25	0.93	0.80	0.76	0.78
21	AdaBoostClassifier with SMOTE'> 0.5	0.90	0.65	0.83	0.73
22	AdaBoostClassifier with SMOTE'> 0.75	0.89	0.61	0.91	0.73
23	AdaBoostClassifier with SMOTE'> auto	0.88	0.58	0.92	0.71
24	AdaBoostClassifier with ADASYN'> 0.25	0.93	0.79	0.76	0.78
25	AdaBoostClassifier with ADASYN'> 0.5	0.92	0.69	0.87	0.77
26	AdaBoostClassifier with ADASYN'> 0.75	0.89	0.61	0.88	0.72
27	AdaBoostClassifier with ADASYN'> auto	0.86	0.54	0.89	0.67

Applying the Model

By implementing the model, Teleworld can get a report on a daily or weekly basis that identifies customers who are of high value to the company and at high risk of Churning in the coming months.

These customers can then be given incentives to stay on either an automated or personalized basis.

We can define high value and high risk anyway we want. But let’s say we want to see customers who are of higher than average value and at above a 65% chance of churning in the coming months. The resulting data frame is picture on the next slide.

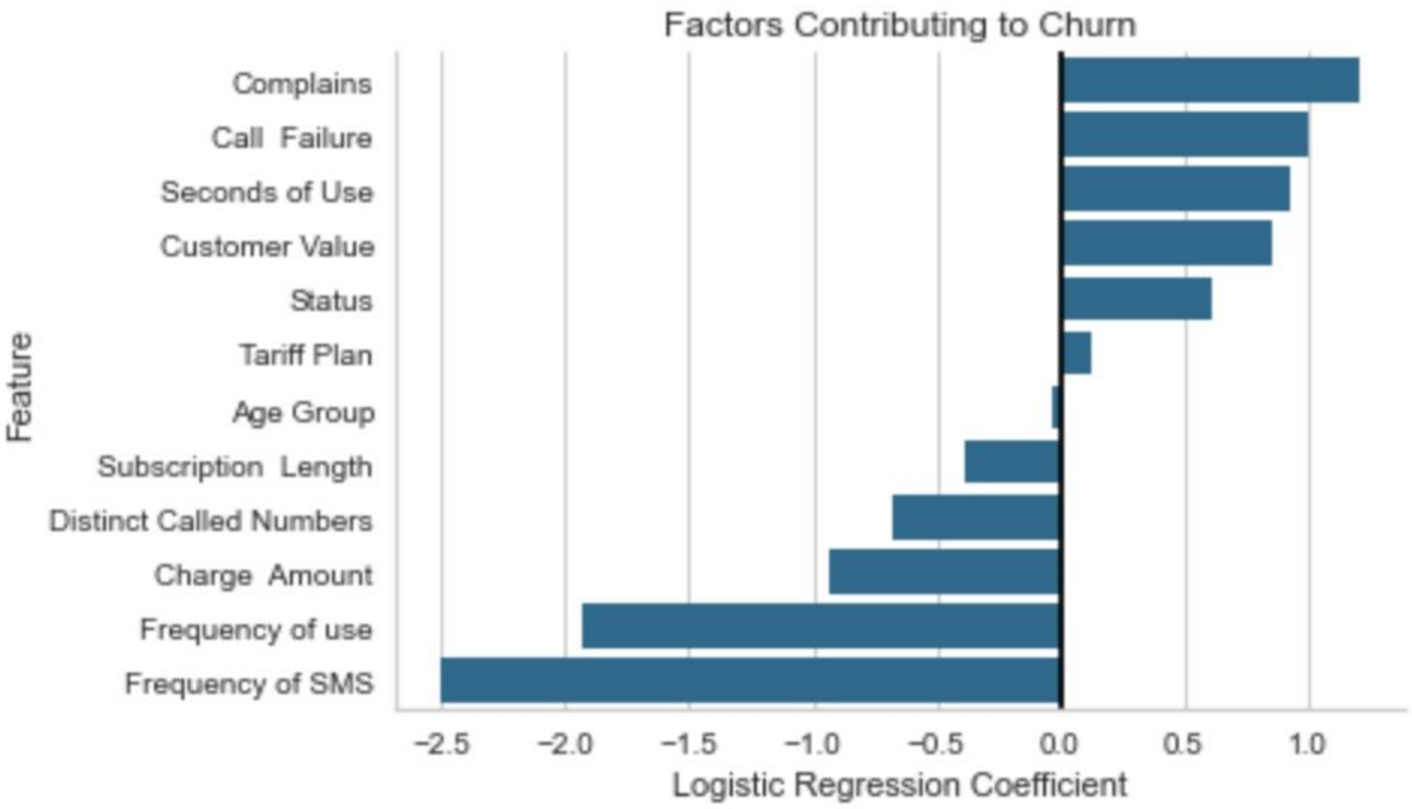
	Call Failure	Complains	Subscription Length	Charge Amount	Seconds of Use	Frequency of use	Frequency of SMS	Distinct Called Numbers	Age Group	Tariff Plan	Status	Customer Value	Churn Prediction	Churn Probability
1138	2	1	27	1	1070	33	191	11	2	2	1	909.135	1	0.95
1724	10	0	34	0	5703	93	28	25	2	1	1	386.820	1	0.91
24	13	1	36	1	5818	98	26	24	2	1	1	383.220	1	0.98
974	13	0	36	1	5818	98	26	24	2	1	1	383.220	1	0.91
1874	10	0	39	1	5513	98	27	23	2	1	1	373.995	1	0.79
2074	8	0	34	0	5513	92	19	22	2	1	1	337.725	1	0.99
477	13	1	36	0	2868	54	55	46	3	1	2	336.880	1	0.91
224	12	1	32	0	5738	96	16	20	2	1	1	334.530	1	1.00
1627	4	0	35	0	2078	44	59	33	3	1	2	320.880	1	0.75
576	9	1	36	0	3105	47	41	20	3	1	2	290.080	1	0.98
677	11	1	36	0	2458	45	43	34	3	1	2	272.120	1	0.99
2376	19	1	43	0	3555	57	31	24	3	1	2	268.480	1	0.75
476	13	1	35	0	3590	51	30	31	3	1	2	265.640	1	0.95
524	9	1	36	0	5218	88	5	9	2	1	1	261.270	1	0.81
1526	8	0	33	0	3285	45	23	29	2	1	2	253.350	1	0.81
2572	32	1	35	0	1860	63	44	34	3	1	2	252.920	1	0.89
2476	16	1	40	0	3475	55	27	23	3	1	2	249.200	1	0.94
2776	16	1	37	0	3485	48	25	21	3	1	2	241.320	1	0.99

Understanding how each Feature Contributes to Churn

I chose to look at the coefficients from the Logistic Regression model to rank the features.

While Logistic Regression wasn't the best classifier, using it for coefficient analysis provides a high interpetable ranking of the features and their relationship to the target variable (Churn)

Results on next slide!



The higher the coefficient, the more influence that feature has over the prediction of the target class.

Factors with positive coefficients make a customer more likely to Churn.

Factors with negative coefficients make a customer less likely to Churn.

Applying the Coefficient Analysis

Now that Teleworld can see how each factor contributes to Churn. They can prioritize and address those factors in a more systematic way.

Why use Logistic Regression Coefficient Analysis?

Sure, we could have visualized how each feature differs between the Churn and Non Churn group but by looking at the coefficients we can actually rank their overall influence.

Further information

This was a brief summary of the project and its findings.

For those interested in learning more, all code can be found in this Github Repository

https://github.com/jeremysilva1098/predicting_churn