Predicting User Churn

Calvin Goh

Jun 2014

Problem and Approach

Can we predict which customer will switch mobile operator?

- Some of the typical Information available about subscribers are demographics (age etc.), data and call usage, revenue information.

- At the time of renewing contracts, some subscribers do and some do not churn. It would be extremely useful to know in advance which customers have a higher propensity of churning, and to prevent it – especially in for high revenue subscribers.

- This is a prediction problem where a small sample data set is used to see who has churned and who has not in the past, and to predict which customer will churn and which will not.

Data: Network & Contract

Input Variables:

- Account details (Length, Subscribed Plans, Phone number)
- Call duration and charges during the Day, Evening, Night & International calls
- Call duration during the Day, Evening, Night & International calls
- Location (Area code, State)
- Current Status (Churned/Not Churned)

Source: Dataset from Ian Pardue at IanPardue.com

Analysis – Data Cleaning/Pre-Processing

Account Length	Integer	0	Min 1	Max 243	Average 101.065	Deviation 39.822
VMail Message	Integer	0	Min 0	Max 51	Average 8.099	Deviation 13.688
Day Mins	Numeric	0	Min 0	Max 350.800	Average 179.775	Deviation 54.467
Eve Mins	Numeric	0	Min 0	Max 363.700	Average 200.980	Deviation 50.714
Night Mins	Numeric	0	Min 23.200	Max 395	Average 200.872	Deviation 50.574
Intl Mins	Numeric	0	Min 0	Max 20	Average 10.237	Deviation 2.792
CustServ Calls	Integer	0	Min 0	Max 9	Average 1.563	Deviation 1.315
Churn	Integer	0	Min 0	Max 1	Average 0.145	Deviation 0.352
Int'l Plan	Integer	0	Min 0	Max 1	Average 0.097	Deviation 0.296
VMail Plan	Integer	0	Min 0	Max 1	Average 0.277	Deviation 0.447
Day Calls	Integer	0	Min 0	Max 165	Average 100.436	Deviation 20.069
Day Charge	Real	0	Min 0	Max 59.640	Average 30.562	Deviation 9.259
Eve Calls	Integer	0	Min 0	Max 170	Average 100.114	Deviation 19.923
Eve Charge	Numeric	0	Min 0	Max 30.910	Average 17.084	Deviation 4.311
Night Calls	Integer	0	Min 33	Max 175	Average 100.108	Deviation 19.569
Night Charge	Real	0	Min 1.040	Max 17.770	Average 9.039	Deviation 2.276
Intl Calls	Integer	0	Min 0	Max 20	Average 4.479	Deviation 2.461
Intl Charge	Numeric	0	Min 0	Max 5.400	Average 2.765	Deviation 0.754
State	Polynominal	0	Least CA (34)	Most WV (106)	Values WV (106), MN (84), ...[49 more]	
Area Code	Integer	0	Min 408	Max 510	Average 437.182	Deviation 42.371
Phone	Polynominal	0	Least 422-9964 (1)	Most 327-1058 (1)	Values 327-1058 (1), 327-1319 (1), ...[3331 more]	

- Examples in the dataset were relatively clean, ie. no missing data or abnormal data in all the attributes
- Except for State & Phone attribute, all others were either Integer or Numeric data types

- Churn attribute type changed from numerical to polynominal
- Churn examples replaced from (1/0) to (Yes/No)
- Churn attribute role set to label for prediction

Analysis – Correlation Tests

Attributes	Account Length	VMail Message	Day Mins	Eve Mins	Night Mins	Intl Mins	CustServ Calls	Churn	Int'l Plan	VMail Plan	Day Calls	Day Charge	Eve Calls	Eve Charge	Night Calls	Night Charge	Intl Calls	Intl Charge	State	Area Code	Phone
Account Length	1	-0.005	0.006	-0.007	-0.009	0.010	-0.004	0.017	0.025	0.003	0.038	0.006	0.019	-0.007	-0.013	-0.009	0.021	0.010	0.001	-0.012	0.037
VMail Message	-0.005	1	0.001	0.018	0.008	0.003	-0.013	-0.090	0.009	0.957	-0.010	0.001	-0.006	0.018	0.007	0.008	0.014	0.003	-0.003	-0.002	-0.018
Day Mins	0.006	0.001	1	0.007	0.004	-0.010	-0.013	0.205	0.049	-0.002	0.007	1.000	0.016	0.007	0.023	0.004	0.008	-0.010	-0.010	-0.008	-0.021
Eve Mins	-0.007	0.018	0.007	1	-0.013	-0.011	-0.013	0.093	0.019	0.022	-0.021	0.007	-0.011	1.000	0.008	-0.013	0.003	-0.011	0.010	0.004	0.014
Night Mins	-0.009	0.008	0.004	-0.013	1	-0.015	-0.009	0.035	-0.029	0.006	0.023	0.004	-0.002	-0.013	0.011	1.000	-0.012	-0.015	-0.002	-0.006	0.011
Intl Mins	0.010	0.003	-0.010	-0.011	-0.015	1	-0.010	0.068	0.046	-0.001	0.022	-0.010	0.009	-0.011	-0.014	-0.015	0.032	1.000	-0.006	-0.018	0.006
CustServ Calls	-0.004	-0.013	-0.013	-0.013	-0.009	-0.010	1	0.209	-0.025	-0.018	-0.019	-0.013	0.002	-0.013	-0.013	-0.009	-0.018	-0.010	-0.004	0.028	0.010
Churn	0.017	-0.090	0.205	0.093	0.035	0.068	0.209	1	0.260	-0.102	0.018	0.205	0.009	0.093	0.006	0.035	-0.053	0.068	0.022	0.006	0.040
Int'l Plan	0.025	0.009	0.049	0.019	-0.029	0.046	-0.025	0.260	1	0.006	0.004	0.049	0.006	0.019	0.012	-0.029	0.017	0.046	0.033	0.049	-0.008
VMail Plan	0.003	0.957	-0.002	0.022	0.006	-0.001	-0.018	-0.102	0.006	1	-0.011	-0.002	-0.006	0.022	0.016	0.006	0.008	-0.001	-0.005	-0.001	-0.014
Day Calls	0.038	-0.010	0.007	-0.021	0.023	0.022	-0.019	0.018	0.004	-0.011	1	0.007	0.006	-0.021	-0.020	0.023	0.005	0.022	-0.026	-0.010	0.000
Day Charge	0.006	0.001	1.000	0.007	0.004	-0.010	-0.013	0.205	0.049	-0.002	0.007	1	0.016	0.007	0.023	0.004	0.008	-0.010	-0.010	-0.008	-0.021
Eve Calls	0.019	-0.006	0.016	-0.011	-0.002	0.009	0.002	0.009	0.006	-0.006	0.006	0.016	1	-0.011	0.008	-0.002	0.017	0.009	0.016	-0.012	0.009
Eve Charge	-0.007	0.018	0.007	1.000	-0.013	-0.011	-0.013	0.093	0.019	0.022	-0.021	0.007	-0.011	1	0.008	-0.013	0.003	-0.011	0.010	0.004	0.014
Night Calls	-0.013	0.007	0.023	0.008	0.011	-0.014	-0.013	0.006	0.012	0.016	-0.020	0.023	0.008	0.008	1	0.011	0.000	-0.014	0.002	0.017	0.001
Night Charge	-0.009	0.008	0.004	-0.013	1.000	-0.015	-0.009	0.035	-0.029	0.006	0.023	0.004	-0.002	-0.013	0.011	1	-0.012	-0.015	-0.002	-0.006	0.011
Intl Calls	0.021	0.014	0.008	0.003	-0.012	0.032	-0.018	-0.053	0.017	0.008	0.005	0.008	0.017	0.003	0.000	-0.012	1	0.032	-0.030	-0.024	-0.011
Intl Charge	0.010	0.003	-0.010	-0.011	-0.015	1.000	-0.010	0.068	0.046	-0.001	0.022	-0.010	0.009	-0.011	-0.014	-0.015	0.032	1	-0.006	-0.018	0.006
State	0.001	-0.003	-0.010	0.010	-0.002	-0.006	-0.004	0.022	0.033	-0.005	-0.026	-0.010	0.016	0.010	0.002	-0.002	-0.030	-0.006	1	0.005	0.033
Area Code	-0.012	-0.002	-0.008	0.004	-0.006	-0.018	0.028	0.006	0.049	-0.001	-0.010	-0.008	-0.012	0.004	0.017	-0.006	-0.024	-0.018	0.005	1	-0.018
Phone	0.037	-0.018	-0.021	0.014	0.011	0.006	0.010	0.040	-0.008	-0.014	0.000	-0.021	0.009	0.014	0.001	0.011	-0.011	0.006	0.033	-0.018	1

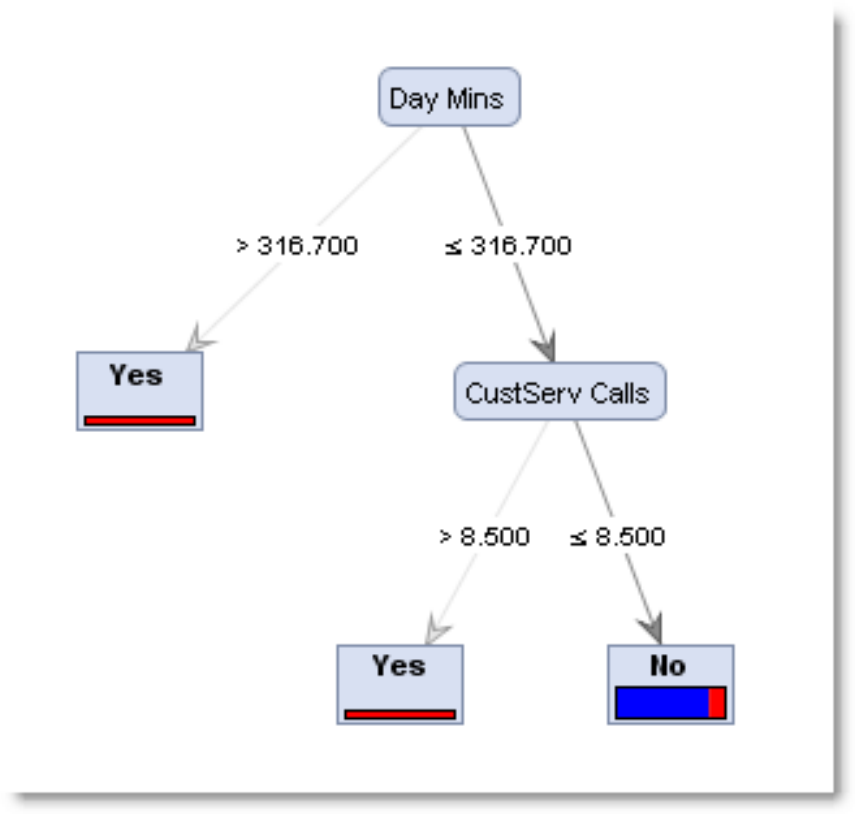
Based on RM's correlation matrix, there was a strong correlation for

- Call minutes
- Call charged
- Vmail Plan, Messages

While there was a fair churn correlation with

- Call minutes during the day
- Call charged during the day
- Number of customer service calls
- International plan

Analysis – Decision Trees



- Max depth = 20 (Def)
- Split ration = 0.8
- Split = relative
- Accuracy = 85.31%

Conclusions

- Subscribers who make calls more than 316 minutes during business hours tend to churn!
- Heavy users who make more than 8 customer service calls will switch to another mobile operator
- Not Important: Account length, International plan, charges

Challenges & Successes

Successes

- Visualization
- Correlation analysis
- Decision tree modeling

Challenges

- Explore more variables

Key Takeways & Next Steps

- Correlation does not mean causation
- Use graphics
- Open source materials available

Next Steps:

- Explore more variables
- Clustering – detect types of users