Telco Customer Churn prediction

Decision Tree with SPSS

Purpose

Maintaining customer loyalty is one of the key aspirations of any telecommunication company. Acquiring a new customer is more costly than retaining an old one. Customer churn is one of the issues that telecom company ('Telco') is facing today.

Therefore, predicting customer churn is more necessary for Telco to be able to effectively retaining customer.

This presentation aims at predicting customer churn using different decision tree algorithms in SPSS with two objectives:

1. Screening the existing customers which are likely to as churn cases.
2. Creating the model (extracting rule) to follow for future customer data.

Data Description

We will predict the criteria of Customer going to churn by examining relationships between the independent variables and the churn status variable.

The sample Telco data (in SPSS sample data) for this study has totaling of 1000 records. The data set consists of the following information:

- Customer demographic including age, gender, marital status, address…
- Service and equipment rental information: voice, pager, internet, callID, callwait, forwarding, conference, equipment, call card
- Servicing log from last month: log-call, log-equipment, long-distance…
- Customer who left within last month – Customer Churn.

Decision Tree

Growing methods

There are three growing method in the decision tree, namely:

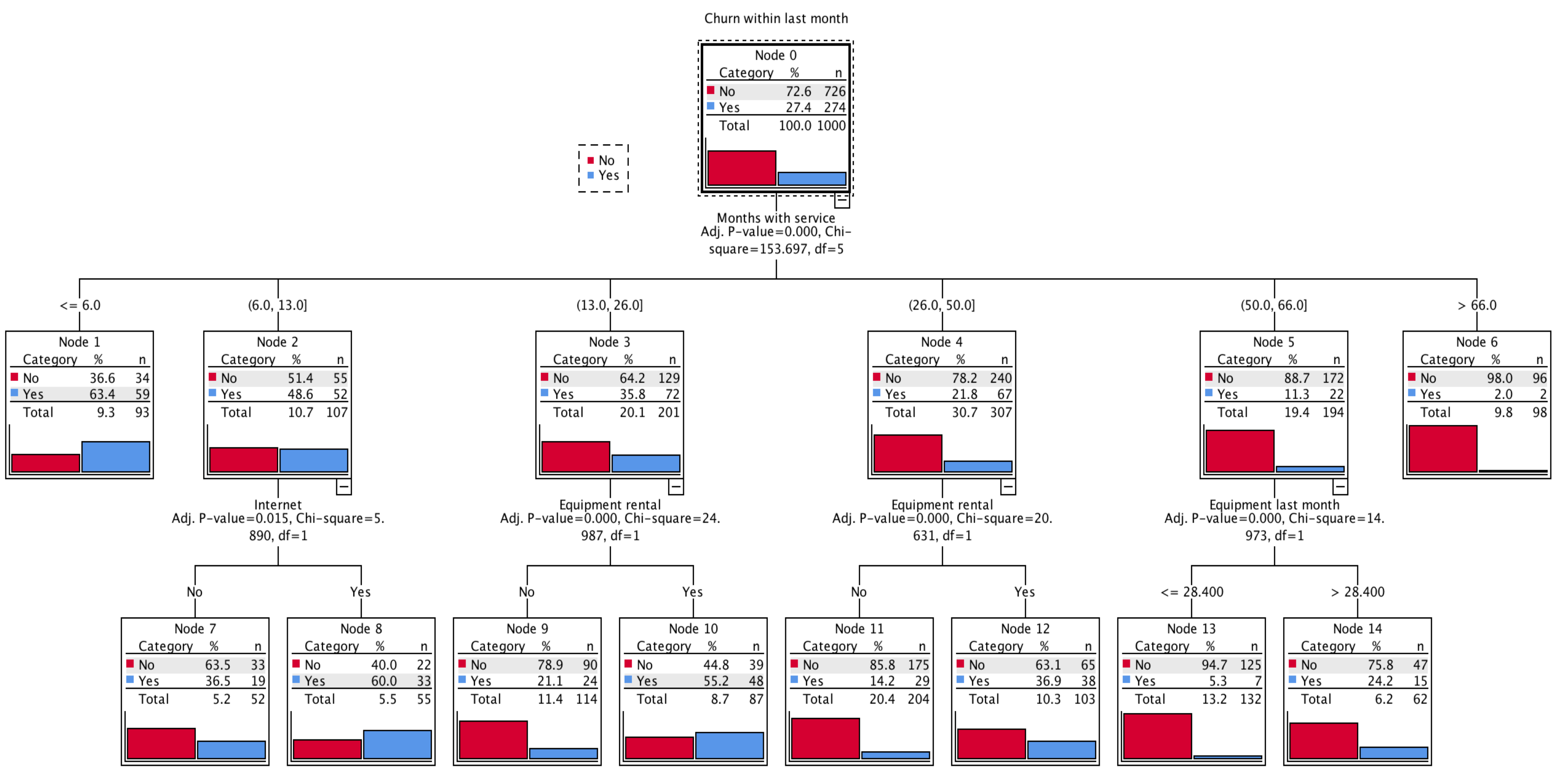
- **CHAID** (**CH**i-squared **A**utomatic **I**nteraction **D**etection)
- **CART** (**C**lassification **a**nd **R**egression **T**rees) – or ***CRT***
  - CART algorithm use Gini's impurity index to build the Decision Tree.
  - ***ID3* : (Iterative Dichotomiser):** Uses Information Gain as attribute selection measure.
  - ***C4.5, C.5* (Successor of ID3):** Uses Gain Ratio as attribute selection measure.
- **QUEST** (**Q**uick, **U**nbiased, **E**fficient **S**tatistical **T**ree)

- ***Exhaustive CHAID*** *is another variant of CHAID.*

Descriptive Analysis with CHAID

Tree Depth	5
Parent/Child	100/50
α	0.05
Bonferroni adj	YES

CHAID algorithm uses Chi-square statistic to find a significant different between child and parent nodes. A chi-square value closer to 0 indicates that there is a significant difference between the two classes which are being compared. Similarly, a value closer to 1 indicates that there is not any significant difference between the 2 classes.



The remaining rate increase over month of services:

- Node 1: Within *six month of service*, customer are more likely to switch than higher.
- Node 6: Customer with more than *66 months service* (~5 years) are more likely to remain.
- Node 2: with 6-13 months of service, switching rate seem to equal but if splitting with additional internet feature (Node 7 & 8) : customer more likely to change if they *use internet service*.
- Node 3: with 13-26 months, remaining rate is higher, however, it changes if customer *have equipment rental*.

CHAID

Tree Depth	5
Parent/Child	100/50
α	0.05
Bonferroni adj	YES

Gains for Nodes

Node	Node N	Node Percent	Gain N	Gain Percent	Response	Index
1	93	9.3%	59	21.5%	63.4%	231.5%
8	55	5.5%	33	12.0%	60.0%	219.0%
10	87	8.7%	48	17.5%	55.2%	201.4%
12	103	10.3%	38	13.9%	36.9%	134.6%
7	52	5.2%	19	6.9%	36.5%	133.4%
14	62	6.2%	15	5.5%	24.2%	88.3%
9	114	11.4%	24	8.8%	21.1%	76.8%
11	204	20.4%	29	10.6%	14.2%	51.9%
13	132	13.2%	7	2.6%	5.3%	19.4%
6	98	9.8%	2	0.7%	2.0%	7.4%

Growing Method: CHAID
Dependent Variable: Churn within last month

The top five nodes account for more or less 40% of data but they allow us to identify more or less 71% of those who are likely to switch providers

Classification

Observed	Predicted No	Predicted Yes	Percent Correct
No	631	95	86.9%
Yes	134	140	51.1%
Overall Percentage	76.5%	23.5%	77.1%

Growing Method: CHAID
Dependent Variable: Churn within last month

The model is 77.1% accuracy at making predictions whether customer switching.

Predictive Analysis

Tree Depth	5
Parent/Child	50/30

- Data set is splitted 80/20 into Training and Testing, and fed to all growing data mining technique CHAID, CART, QUEST and Exhaustive CHAID to find and validate Prediction models.
- The results were obtained and analysed.
- For each technique, the output include a Classification table – measuring the tree's predictive accuracy, an assiociated Risk table, and Decision Tree with Gain for nodes table.

Predictive Analysis with CART

- After analyse the obtained results, it is observed that the CART has give the most accurate result as well as minimum risk error:

	Accuracy in Training	Accuracy in Testing
CART	79.1%	76.7%
CHAID	78.4%	76.7%
QUEST	78.6%	75.6%
Exhaustive CHAID	77.4%	76.2

Classification

Sample	Observed	Predicted No	Predicted Yes	Percent Correct
Training	No	519	61	89.5%
	Yes	108	119	52.4%
	Overall Percentage	77.7%	22.3%	79.1%
Test	No	134	12	91.8%
	Yes	33	14	29.8%
	Overall Percentage	86.5%	13.5%	76.7%

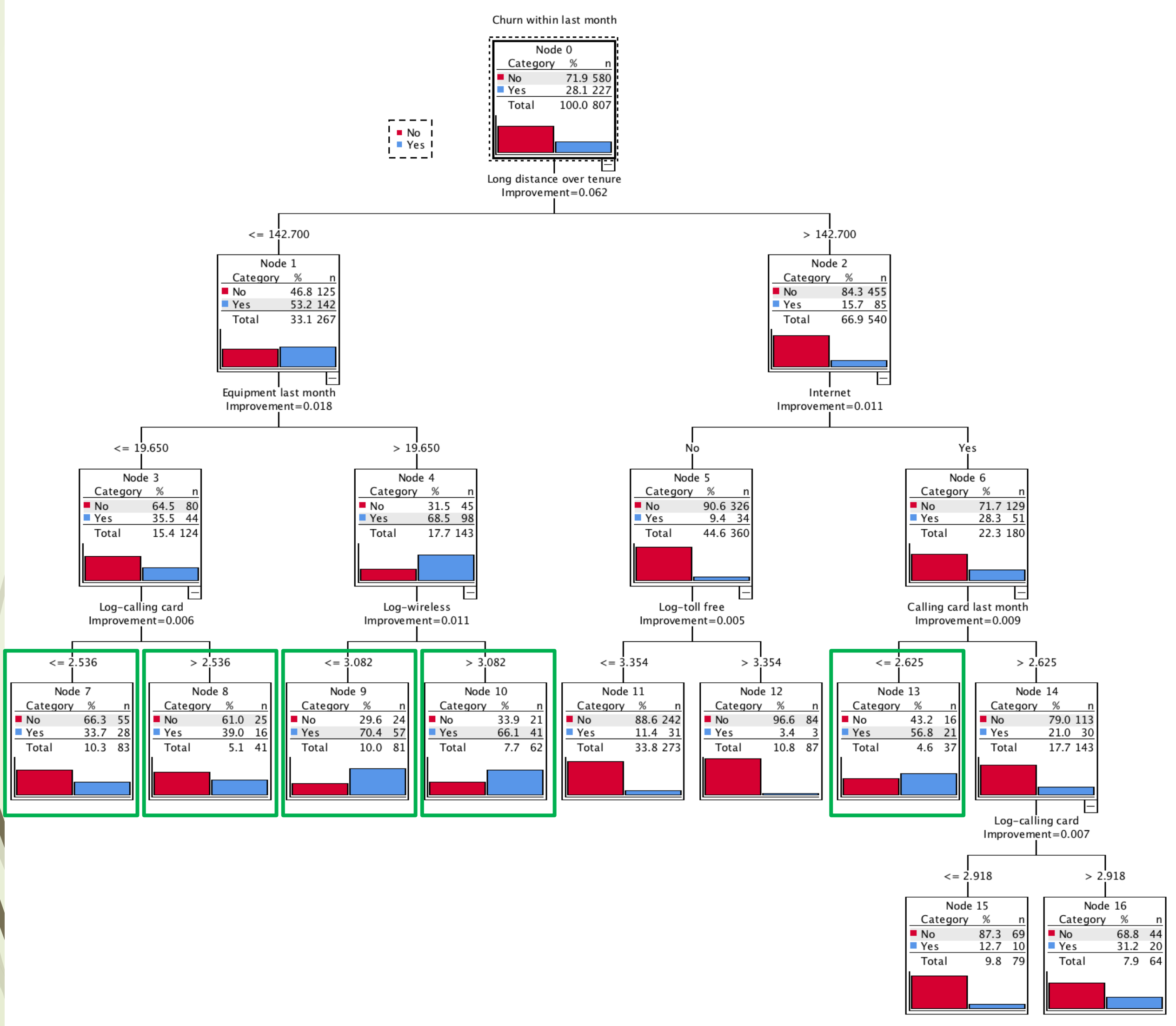
Growing Method: CRT
Dependent Variable: Churn within last month

Risk

Sample	Estimate	Std. Error
Training	.209	.014
Test	.233	.030

Growing Method: CRT
Dependent Variable: Churn within last month

Predictive Analysis with CART



- The Tree with CART algorithm is selected as predictive models, the top 5 nodes can be considered as the **Extracting rules** for selected model :

Node	Rule	Predicted Category	Probability
9	Long-distance over tenure <= 142.7 AND Equipment last month > 19.65 AND log-wideless <=3.082	YES	70.4%
10	Long-distance over tenure <= 142.7 AND Equipment last month > 19.65 AND log-wideless > 3.082	YES	66.1%
13	Long distance over tenure > 142.7 AND Internet=Yes AND Calling-card last month<=2.265	YES	56.8%
8	Long-distance over tenure <= 142.7 AND Equipment last month <= 19.65 AND log-calling card > 2.536	NO	60.9%
7	Long-distance over tenure <= 142.7 AND Equipment last month <= 19.65 AND log-calling card <= 2.536	NO	66.3%

Conclusion

The predictive model with CART algorithm reaches an accuracy 76.7%. It means that it predicts nearly 8 out of 10 times correctly whether customer is likely to churn or not.

The decision tree solves two goals as mentioning in the objectives:

1. Screening existing customers: We noted from Classification table, there are 61 False Positive ('FP') cases in Training and 12 FP in Testing.
   - *False Positive is a case still being customer but predicted as churn. We then need to pay attention to have appropriate actions to keep them satisfied with service !*
   - From False Positive cases, we can review the Predicted probability for churn (=1) and Variable importance *(or other variables according to business requirement)* to decide improvement of services.
2. Creating the model: **Extracting rules** from predictive model can also be used to predict potential churn cases for future data.

Independent Variable Importance

Independent Variable	Importance	Normalized Importance
Long distance over tenure	.066	100.0%
Calling card over tenure	.055	82.5%
Months with service	.055	82.4%
Calling card last month	.044	66.3%
Calling card service	.041	61.4%
Log-long distance	.039	58.8%
Long distance last month	.039	58.8%
Years with current employer	.031	46.0%

This is the feature/service we should pay attention for improvement.

False Positive

	Long distance over tenure	Calling card over tenure	Months with service	Calling card last month	Calling card service	Log-long distance	Long distance last month	Predicted Probability for churn=1
1	75.25	.00	13	.00	No	1.71	5.55	.70
2	96.30	.00	16	.00	No	1.89	6.60	.70
3	16.65	.00	8	.00	No	1.15	3.15	.70
4	41.85	.00	10	.00	No	1.58	4.85	.70
5	40.30	155.00	13	14.00	Yes	1.12	3.05	.70
6	9.25	40.00	4	13.75	Yes	.92	2.50	.70
7	16.35	.00	7	.00	No	.90	2.45	.70
8	125.05	.00	22	.00	No	1.96	7.10	.70
9	7.00	.00	3	.00	No	.99	2.70	.70
10	4.60	.00	2	.00	No	1.36	3.90	.70
11	55.05	75.00	13	6.75	Yes	1.35	3.85	.70
12	114.30	.00	17	.00	No	1.86	6.45	.70
13	52.85	95.00	12	8.50	Yes	1.34	3.80	.70
14	124.45	.00	20	.00	No	1.86	6.40	.70
15	14.95	.00	4	.00	No	1.28	3.60	.70
16	79.45	.00	19	.00	No	1.40	4.05	.70
17	11.20	.00	3	.00	No	1.54	4.65	.70
18	5.30	60.00	4	16.50	Yes	1.03	2.80	.70
19	53.50	.00	13	.00	No	1.64	5.15	.70
20	112.00	.00	29	.00	No	1.56	4.75	.70

Thank you