Shiraz Journal of System Management
Vol. 2, No. 1, Ser. 5, (2014), 39-49

Predicting Customer Churn Using CLV in Insurance Industry

Vahid Dust Mohammadi*
Department of Industrial Engineering,
Tarbit Modares University,
Tehran, Iran
Vahid.dostmohamadi@modares.ac.ir

Amir Albadvi
Department of Industrial Engineering,
Tarbit Modares University,
Tehran, Iran
albadvi@modares.ac.ir

Babak Teymorpur
Department of Industrial Engineering,
Tarbit Modares University,
Tehran, Iran
b.teimourpour@modares.ac.ir

Abstract. Today, increased level of customer awareness caused them to access to the other suppliers easily and they can get their services from the competitors with similar or even better quality and same price. Therefore, focusing on customers and preventing them to leave, has been the most important strategy for any company. Researches have shown that retaining former customers is cheaper than attracting new ones. In the proposed model in this article we first identified important factors causing customers in insurance industry, to have a specific behavior by using a k-means clustering algorithm, and then we tried to predict the future behavior of them by a logistic regression. Our model accuracy is 98%.

Keywords: Customer churn, customer lifetime value, k-means clustering, logistic regression, insurance industry.

Received: January (2014); Final Revision: May (2014)
*Corresponding author

1. Introduction

Changing economic and social characteristics caused organizations to do things in a different way.[1] Unlike the past, today customers are determining success factors of organizations. On the other hand, using new technologies like internet made making new customers and finding new suppliers easier and cheaper. In this situation new entrants can enter into the competing market with very low costs and lots of substitutes can be found just by visiting some sites in the internet. All of these features of using the internet have made the customers stronger than before, customers who can choose between different suppliers and different substitutes. On the other hand survining in the competing market is more complicated for organizations.[2] Organizations have known that their most important asset is their customers, so all of their effort is to reach to a larger market share and creating value for more valuable customers. One of the issues that organizations encounter is the way they should communicate with their customers, in the other words; they are supposed to maintain their old customers in addition to making new ones. In this situation the more successful organization is the one who can maintain its valuable customers and by making appropriate policies preventing them to leave ,because maintaining the old customers is more cheaper than making new ones.[3] So, different researches have been done recently in the field of detecting organization's valuable customers are very important and helpful to find good ways of preventing customers to churn.

In the next section the importance of churn analysis from the viewpoint of different researches is going to be studied, then we explain the characteristics of data we have used in this article and in the last section we propose our predicting model for analyzing customer behavior.

2. Customer Relationship Management (CRM)

Customer relationship management includes processes and systems which support a long and valuable relationship with special customers. [4] Customer behavior records and information technology is the base of any

CRM strategy. The increasing use of the internet and its related technologies has made marketing more easily and the way customers communicating with organizations are being changed due to these services. Although CRM is one of the important business strategies, but there is not already an international specified definition for it. [4, 5]
In different researches different definitions of CRM can be found:

• An organizational trend for analyzing and affecting customer behavior along a valuable period of communication for better acquisition and maintaining of loyal customers. [5]

• Strategic using of information, processes, technologies and people for managing the relationship with customers with organizations (including marketing, sales, service and support) during the customer lifecyle. [6]

• A general strategy and process of acquisition, maintaining and communicating with customers for creating value for organization and customer. CRM includes marketing, sales, services and organization's supply chain for better performance and efficiency in creating value for customers. [7]

CRM includes customer identification, customer attraction, customer retention and customer development [8]. Analyzing customer churn is one of the issues of CRM which is being classified in customer retention field.

Today in industrial countries the number of service firms is increasing and Iran is not an exception. The growth rate of the number of insurance companies is impressive in Iran. There are 20 insurance companies (one of them acts as a public company and the rest belongs to the private sector), 15,200 insurance agents and 270 brokerages act in Iran insurance industry [9]. The penetration rate of insurance in Iran is 1.5% of GDP and if social insurance premiums, pension funds and supportive insurance are counted in this ratio it is 4.6% of GDP [9]. So the importance of insurance industry cannot be denied in the country. The existence of different insurance companies and new entrants to the insurance industry in one hand, and using new technologies like the internet, which made more informed customers and competitors, in the other hand, made competition for customer attraction more intense in the first step and preventing customers to churn-which in important-in the second. Chu in his article in 2007 declares that cost of new customer attraction is 5-10 times more than the cost of current customer retention. Researches show that just 5% increase in customer retention rate can cause 25%-95% increase in net present value in the insurance industry. [10]

So today organizations focus on customer retention more than customer attraction, in fact, they found that the best strategy for them is to maintain their current customers and preventing them to churn. [11]

Modeling customer churn makes churn management easier and is essential for success, profitability and surviving in the competing market. If the insight of the customer behavior analysis is used, the customer retention rate can be increased and as a result by choosing appropriate policies more profit can be gained. In fact, predicting customer churn is not the last thing that churn management have to do, and it cannot decrease the churn rate by its own, so one would need more analysis to do that. Customer value has been studied as customer lifetime value, customer equity and customer profitability. Past researches define customer lifetime value in almost the same way. Different definitions of CLV are shown in table 1. [12]

Table 1. Difinitions of CLV

Year	Researcher	Definition
2003	Gupta and Lehmann	Net present value of all future profits that can be gained from the customer
1998	Berger and Nasr	Net profit or loss that an organization gains from the customer along contract period.
1996	Blattberg and Deighton	The expected gain of a customer without counting the customer management costs.
1996	Bitran and Mondschein	Gaining net profits that a customer makes along his lifetime.
1996	Pearson	NPV can be gained along customer interactions
1994	Jackson	The NPV of profits, can be expected of customer future interactions
1989	Roberts and Berger	NPV of all customer interactions

According to the above definitions, CLV can be defined as the collection of revenues from customers of the organization along their interaction period, which attraction, sale and service costs are subtracted from the, and is declared in terms of time value of money. What is obvious in the above definitions is the historical attitude to the customer purchases and they don't define any potential for the customer, this defect has been corrected in the new models. [13] CLV is calculated as:

$$CLV = \sum_{i=1}^{n} \frac{(R_i - C_i)}{(1+d)^{i-0.5}} \tag{1}$$

R_i= the amount of customer revenue that the organization gains;

C_i= the cost of services that the organization provides; D= the amount of discount rate;

I= the number of periods that customer have transaction with the organization.

Logistic regression is one of the applicable techniques for analyzing classified data. For example, if the result of an experiment is defined as loose or win, then the dependent variable is not continues and will be a categorical variable. One kind of logistic regressions is binary logistic regression which it has two classes of dependent variables. [14]

This model is defined as:

$$logit(p) = ln(\tfrac{p}{1-p}) = \beta_0 + \beta_1 x_{1,i} + ... + \beta_k x_{k,i} \tag{2}$$

$$p = pr(y_i = 1|X) = \tfrac{e^{\beta_0 + \beta_2 x_{2,i} + ... + \beta_k x_{k,i}}}{1+e^{\beta_0 + \beta_2 x_{2,i} + ... + \beta_k x_{k,i}}} \tag{3}$$

β_0= the constant of the equation
β= the coefficient of the predictor variables

3. Research Data Features

In this article the data of four years of third party insurance of one of the insurance companies of Iran used for generating a predictive model. 2,365,567 customer records which their base contract year was 1387 were chosen to analyze, in the other words customers who their contract date was after 1388 were not considered in the analysis (By considering this limitation the number of records decreased to 1,447,064). 75% of the records were used to generate the predictive model and the rest was used to test the accuracy of the generated model. In the next step 15 attributes which were used in the database of the company were used to cluster the customers into two clusters using k-means method (churn customer's cluster and non-churn customer's cluster) then the accuracy of the clustering method was analyzed by comparing the clustering result with the expectations of the experts of the company (the customer who has not used any services of the company is a churn one) and in 89% cases the results were matched. After insuring the accuracy of the model the important variables in the clustering were extracted. Figure 1 shows the clustering output of the Rapidminer:

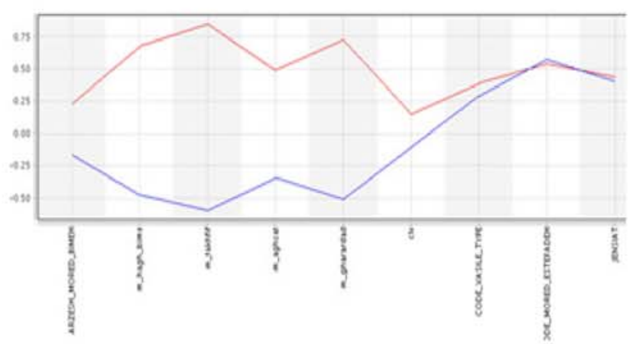

Figure 1. The most important attributes in determining customer clusters

As it is shown in figure 1 four attributes have the greatest impact in the clustering which is:

- The number of contracts of the customer
- The number of installments
- The insurance premium

- Customer lifetime value
- The value of the insured car

To be insured about the accuracy of choosing the right attributes, clustering the customers was done again using the above attributes and in 94% cases customers were in the correct clusters. As a result, these attributes figure an overall image of the customer.

4. Logistic Regression Model for Predicting Customer Churn

After choosing the appropriate attributes, they were used to generate the model. The cluster, which customers were in, was chosen as the dependent variable and the six chosen attributes were used as predictive attributes. In the other words, our target in the modeling was to predict the dependent variable knowing the six attributes. We used IBM SPSS to generate the model and the coefficients in table 2 were the result of the software.

Table 2. Coefficients for logistic regression

Coefficient	Abbreviation	Factor
-3.206	X_1	n_installment
-7.947	X_2	n_contract
-47.937	X_3	Payment
-7.925	X_4	Discount
0.357	X_5	Value
41.617	X_6	CLV
3.442	β_0	constant

According to model output if p-value is more than 0.5 then the customer is more probable to churn and if not the probability of leaving the company is less. This model was tested with the rest 25% of test records and in 95% of them it predicted the right behavior.

The output of the SPSS is used to test the goodness of fit and as shown in table 3 almost 98% cases the predicted behavior was matched to the reality.

Table3. Output of SPSS for goodness of fit Model Summary

step	-2 log likelihood	Cox& snell R square	Negelkerke R Square
1	30260.758	0.720	0.970

a. Estimation terminated at iteration number 11 because parameter estimates changed by less than .001

Table 4. Output of SPSS for classification table

observed			predicted		
			cluster		Percentage correct
			a	b	
Step 1	cluster	a	146962	2514	98.3
		b	2663	209627	98.7
Overall percentage					98.6

Table 5. Output of SPSS for performance of the model

		B	S.E	Wald	df	Sig.
Step 1	CLV	41.613	0.484	7402.6	1	.000
	Nu-in	-3.206	0.332	10354.9	1	.000
	Nu-co	-7.947	0.083	9196.9	1	.000
	Premium	-47.939	0.527	8279.4	1	.000
	Discount	-7.935	0.068	13641.6	1	.000
	value	0.357	0.029	149.2	1	.000
	constant	3.442	0.033	10740.4	1	.000

As shown in the table 3 attributes were not significantly correlated and no multicollinearity existed between them (standard error for all attributes was less than 2). Also, according to p-value column null hypothesis for logistic regression model was rejected for all attributes ($p < 0.05$). These attributes all together had a meaningful share in predicting the probability of churn.

The coefficients show that which attributes increase or decrease the probability of leaving the company. Attributes with negative coefficient has negative impact of not being churned and attributes with positive coefficients has positive impact in being a churn customer. In the other words number of installments, payment, discount and the number of contracts has negative impact of being a churn customer and the rest attributes has positive impact in being a churn one.

For analyzing the goodness of fit we used Cox & Snell and Nagelkerke which they were 0.72 and 0.97 in the SPSS output. As shown, 72% to 97% of variability of the dependent variable can be predicted by the independent variables.

5. Conclusion

As shown in the previous paragraph, 5 attributes (payment, installment, value, CLV, number of contracts and discount) are the most important attributes in the churn management modeling in Iran insurance industry. In the other words, attributes like demographic, application of the insured car doesn't have a meaningful impact on customer churn. Among the extracted attributes number of installment and payments and the amount of discount have an inverse impact in customer churn and the rest attributes have a direct impact. Among these attributes having good policies for discount rate, installment and making customers to have more contracts can be very helpful for success in churn management.

Acknowledgments:
The research reported herein was supported by the insurance research center affiliated to the central insurance of the Islamic Republic of Iran. The data used for this research was provided by the Iran insurance

company. We gratefully acknowledge all our sponsors. We also thank Mr. Hazraty for his helpful suggestions and feedback.

References

[1] Berry, M. J. and Linoff, G. S. (2004), Data mining techniques: for marketing, sales, and customer relationship management: John Wiley & Sons.

[2] Turban, E., Sharda, R., Delen, D., and Efraim, T. (2007), Decision support and business intelligence systems: Pearson Education India.

[3] Tsai, C. F. and Lu, Y. H. (2009), "Customer churn prediction by hybrid neural networks," Expert Systems with Applications, vol. 36, pp. 12547-12553.

[4] Ling, R. and Yen, D. C. (2001), "Customer relationship management: An analysis framework and implementation strategies," Journal of Computer Information Systems, vol. 41, pp. 82-97.

[5] KhakAbi, S., Gholamian, M. R., and Namvar, M. (2010), "Data mining applications in customer churn management," in Intelligent Systems, Modelling and Simulation (ISMS), International Conference on, pp. 220-225.

[6] Zhang, Y., Liang, R., Li, Y., Zheng, Y., and Berry, M. (2011), "Behavior-Based Telecommunication Churn Prediction with Neural Network Approach," in Computer Science and Society (ISCCS), International Symposium on, pp. 307-310.

[7] Benoit, D. F. and Van den Poel, D. (2012), "Improving customer retention in financial services using kinship network information," Expert Systems with Applications, vol. 39, pp. 11435-11442.

[8] Kracklauer, A., Mills, H. D. Q., and Seifert, D. (2004), "Customer management as the origin of collaborative customer relationship management," in Collaborative Customer Relationship Management, Ed: Springer, pp. 3-6.

[9] Kajvary, (2013), "important factors in brand equity in insurance industry in customer perspective," insurance journal.

[10] Chu, B. H., Tsai, M. S., and Ho, C. S. (2007), "Toward a hybrid data mining model for customer retention," Knowledge-Based Systems, vol. 20, pp. 703-718.

[11] Kim, H. S. and Yoon, C. H. (2004), "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market," Telecommunications Policy, vol. 28, pp. 751-765.

[12] Bahman, A. (2014), "modeling CLV in insurane industry," engineering, tarbiat modares, Iran.

[13] Malthouse, E. C. and Blattberg, R. C. (2005), "Can we predict customer lifetime value?" Journal of Interactive Marketing, vol. 19, pp. 2-16.

[14] Amin, R. (2011), "using predicted improvement factors for communication creteria in binary logistic," operation management.