

Evaluations of Data Mining Methods in Order to Provide the Optimum Method for Customer Churn Prediction: Case Study Insurance Industry

Reza Allahyari Soeini^{1, a} and Keyvan Vahidy Rodpys^{2, b}

¹Industrial Development &Renovation organization of Iran, Director Development &Renovation Investment, Tehran, Iran

²Department of e-commerce, Nooretouba University, Master's degree student, Rasht, Iran

^aAllahyari@idro.org , ^bkeyvanvahidy@yahoo.com

Abstract. Competitive advantage for survival and maintenance of the old companies to new companies need to identify accurately understand behavior customers. So many different ways for organizations to predict the company's customers churn. The most common methods of predicting customer churn, data mining methods. Data mining methods to determine the optimal method of prediction is of special importance. So in this article using Clementine software and the database contains 300 records of customers Iran Insurance Company in the city of Anzali, Iran will be collected using a questionnaire. First, determine the optimal number of clusters in K-means clustering and clustering customers based on demographic variables. And then the second step is to evaluate binary classification methods (decision tree QUEST, decision tree C5.0, decision tree CHAID, decision trees CART, Bayesian networks, Neural networks) to predict customers churn.

Keywords: Data mining, Customer churn prediction, K-means clustering, classification binary, insurance

1. Introduction

In the present competitive world market is the main factor for the same customer is expanding day by day. Nowadays with the expansion of market share due to the consumption market demands and desires of customers makes it possible for companies to be able to devote more market share. It is essential to the concept of customer relationship management (CRM). Effective CRM to increase customer value reflects the company's revenues are contributed to customer throughout the process is the customer [11]. CRM, the exchange value between customers and companies to create value in this connection Weber insists, therefore, corporate effort to develop long term relationships with clients based on value creation for both sides of the main goals CRM. One of the main examples of customers in every industry, especially in competitive markets is very important that customer churn. Customer churn very important issue, because the lack of customers, new customers must be guaranteed through the problems because it costs too much trouble attracting a new customer, the high cost process leading to the termination of customer service and reduce the negative impact on revenue and customer loss is not companies [14]

Behavioral patterns of customers turned away from the existing data is something that is long lasting in some industries such as telecommunications, banking, newspaper, film industry, retail industry has been made [16] Thus provide guidelines for predicting churn customer can help companies in any industry to identify any aspects of customer churn, with appropriate strategies to deal with this phenomenon to take steps.

One of the valuable tools for exploring data mining tool for extraction of knowledge from data is the data mining process that uses smart techniques to extract knowledge from data collection. Knowledge extracted in

the form of models, patterns or rules will provide the templates, models and extracted regulations to provide various forms of knowledge.

This knowledge can be a criterion for future decision making, the next function or system changes are required in today's most important commercial data mining companies because they know they have found. Through which we can identify the characteristics and behavior of their customers and the companies with the need to regulate them. In this paper we determine the optimal number of clusters in K-Means clustering and clustering customers based on demographic variables. Methods to evaluate the optimal method for identifying customer churn, we provide results for the company to adopt their strategies and make the right decisions in dealing with them.

2. Literature Review To Predict Customer Churn with Data Mining

Customer relationship management, and customer churn prediction in particular, have received a growing attention during the last decade[16]. Tab.1 provides an overview of the literature on the use of data mining techniques for customer churn prediction modeling. The table summarizes the applied modeling techniques,

Inherent tendency for a withdrawal of continued customer churn commercial relations with customers in a period when a company. Simply churn indicates that customers of a company, go to another company. According to this definition, the customer is turned away anyone who will stop all its activities with the company [15] Tab.1 Summary of literature review of the applications of data mining is expected to provide customer churn.

As seen in Tab.1, modeling techniques to predict the frequency churn customers using data mining These include methods such as clustering, decision trees, logistic regression, neural networks, Bayesian networks, random forests, association rule, support vector machines Modeling capability to provide customer churn with data analysis is created. Research literature in order to understand the reasons for the low hand churn customer [12] [15] the necessity of achieving customer churn approval to use the optimal method of data analysis to understand why customers may churn [19].Evaluation methods to compare results from the over classification is to predict customer churn [16] [20] shows the necessity of research in this area.

Tab.1 Application of data mining in customer churns the research literature

Action	Data mining techniques	Business	References
C4.5 decision tree to model customer churn	Decision tree C4.5	Telecommunication	[17]Wei et al ,2002
Toward the use of data mining techniques to predict customer churn	Decision tree C4.5, Neural network and evaluate the results	Telecommunication	[1]Au et al ,2003
Analysis of customers churn	Association Rule Apriori	Banking	[4]chiang et al,2003
Identify and predict customer churn	Clustering RFM, Decision tree, SVM	Insurance	[25]Morik and Kopck,2004
Models to predict customer churn As part of the customer lifetime value	Logistic regression, Decision trees, Neural network	Telecommunication	[24] Hwang et al.,2004
Comparison of techniques for prediction and focus on profitable customers in a non-contractual	Logistic regression, Neural networks, Random forests	Retail	[20]Buckinx et al,2005
Analysis model with the focus on predicting customer churn	Logistic regression, Decision trees, Neural networks, Bayesian	Telecommunication	[13] Neslin et al,2006
Assessment of classification methods for predicting customer churn	Decision trees (C5.0 CART, Tree Net,)	Telecommunication	[21]Chandar et al,2006
Applying data mining in managing customer churn	Neural networks, decision tree	Telecommunication	[9]Hung et al,2006
Provide a model compound for retaining current customers	Clustering, Decision tree C5.0	Pay TV company	[30] Chu et al 2007
Evaluation of data mining methods to maintain current customers	Markov chain, Random forest, Logistic regression	Pay TV company	[3] Burez et al,2007
Assessment of classification methods for predicting customer churn	SVM, logistic regression, Random forests	Newspaper	[6]Coussement et al,2008

Comparison of neural network techniques to predict customer churn	Neural network, SOM clustering	telecommunication	[28] Tsai et al,2009
identify customer churn	Fuzzy C-means clustering	Banking	[26]Popović et al,2009
Improve the marketing structure prediction customers churn	neural networks, Decision trees, Logistic regression	Newspaper	[22]Coussement et al,2010
identify and predict customer churn	Neural networks, Decision trees, Association Rule Apriori	Telecommunication	[29] Tsai et al,2010

Tab.2 strengths weaknesses of each customer churn modeling techniques are shown. What was said to be acknowledged. Evaluation of data mining methods in order to provide the optimum method for predicting churn something that improves Marketing, CRM companies are current customers toward transition from the company is expected to rival companies

Tab.2 Strengths and weaknesses of various data mining techniques in modeling customer churn

Data mining techniques	Strengths	Data mining techniques
<ul style="list-style-type: none"> the difficulty of extracting classification rules the stability of their steady the optimal solution [8] 	<ul style="list-style-type: none"> very simple technique provide reliable results provide concrete rules [27] 	Decision tree
<ul style="list-style-type: none"> The difficulty of performing construction lack of transparency interpret the output results[1] 	<ul style="list-style-type: none"> The ability to predict precisely [1] 	Neural Networks
<ul style="list-style-type: none"> Inability to express behavior patterns hidden in data the inability of the behavioral patterns of behavioral phenomena [1] 	<ul style="list-style-type: none"> Ease of application performing model very rich literature on the use of model[24] 	Regression
<ul style="list-style-type: none"> Total amount of items that do not frequent [4] 	<ul style="list-style-type: none"> Ability to discover hidden relationships among data behavioral the ability to sequence the events, phenomena customer behavior [4] 	Association Rules
<ul style="list-style-type: none"> Does the Time Being [7] 	<ul style="list-style-type: none"> accuracy of data with much better results on The error rate can be controlled [7] 	Support vector machine
<ul style="list-style-type: none"> method's performance alone is not sufficient to predict customer behavior [3] 	<ul style="list-style-type: none"> The most widely used method Initial assessment of customer data [3] 	Clustering
<ul style="list-style-type: none"> The difficulty of performing construction [23] 	<ul style="list-style-type: none"> stable and steady The data subject has a good performance [23] 	Random Forest
<ul style="list-style-type: none"> New Bayes method for the case of binary characters fare much less accuracy [13] 	<ul style="list-style-type: none"> The number of nominal variables than is the case New Bayes for better performance[13] 	New Bayes

3. Research Methodology

In Fig.1 the conceptual structure that we have to provide the optimum method to predict the show will churn customers. First, determine the optimal number of clusters in K-means clustering and clustering customers based on demographic variables. And then the second step is to evaluate binary classification methods (decision tree Quest, decision tree C5.0, CHAID decision tree, decision trees, CART, Bayesian networks, neural networks) to predict customers churn.

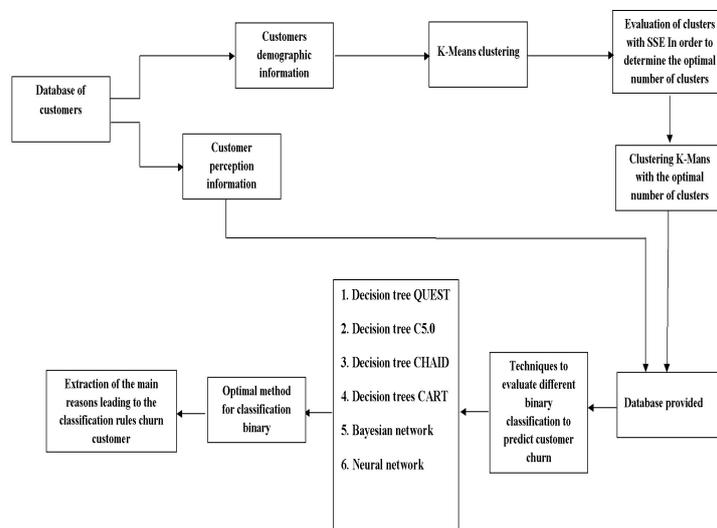


Fig.1 Research methodology

3.1 Case study

During the research database that includes customers in time interval of 23JUL-23SEP 2011, Car insurance customers by questionnaire from Iran Insurance Company in the city of Anzali, Iran has been collected and included demographic variables and customer perception, used data mining.

3.2 K-Means Clustering

The most famous and applicable method of clustering is K-Means which introduced by Mc. Queen in 1967. K-Means Clustering steps are as follows:

First, it randomly selects K of the objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is the most similar, based on the distance between the object and the cluster mean. It then computes the new mean for each cluster. This process iterates until the criterion function converges. Typically, the square-error criterion is used for cluster evaluation [18].

3.3 Decision tree CART

CART decision tree in 1984 by a group of statistical classification and regression was developed. The above algorithm for a comprehensive study on decision trees, providing technical innovations, debate on a complex data structure tree and a strong management on large sample theory for trees is important. CART decision tree and a recursive binary segmentation procedure is capable of processing particular attributes with continuous and discrete values. The data are managed in a row and need not be binary operations. Trees without any law to the greatest extent possible, grown and then by the algorithm, the cost-complexity to root pruning are two-dimensional case, pruning is placed divide that the overall performance of the tree on the data to test the least role play. More than one division at a time may be removed by this operation. The overall goal of this algorithm produces a series of nested trees, and pruning trees, each of which has been

Optimized and are candidates. Tree of appropriate size by calculating the predicted performance of each tree in the pruning process is determined up. Tree performance on independent test data are evaluated and selected based tree after the evaluation is not performed. The cross validation test continues. If there are no test data are necessary to determine the best tree in one step, this algorithm will not be able to. [5][2]

3.4 Decision tree QUEST

QUEST stands for Quick, Unbiased, and Efficient Statistical Tree. It is a relatively new binary tree-growing algorithm. It deals with split field selection and split-point selection separately. The univariate split in QUEST performs approximately unbiased field selection. That is, if all prediction fields are equally informative with respect to the target field, QUEST selects any of the predictor fields with equal probability. QUEST affords many of the advantages of CART. You can apply automatic cost-complexity pruning to QUEST tree to cut down its size. QUEST uses surrogate splitting to handle missing values [31]

3.5 Decision Tree Chaid

CHAID stands for Chi-squared Automatic Interaction Detector. It is a highly efficient statistical technique for segmentation, or tree growing, developed by (Kass, 1980). Using the significance of a statistical test as a criterion, CHAID evaluates all of the values of a potential predictor field. It

Merges values that are judged to be statistically similar with respect to the target variable and maintains all other values that are dissimilar [31].

3.6 Decision tree C5.0

Algorithm decision tree C5.0 because this algorithm features a new non-categorized variable cost offers. There are errors in the algorithm is the same treatment to all. With this feature, this algorithm attempts to reduce the error rate is predicted in some recent applications of data mining, data volume is greatly increased. In some cases, hundreds or even thousands of properties is observed. C5.0 Prior to class, capable of screening the attribute is automatically and The practice of removing attributes that are relevant and less

dependent than other attributes in applications with high data volume, resulting in screening of smaller classification and prediction accuracy higher [2]

3.7 Neural Network

For the best rated models, are models of neural network a simplified model of the field of neural networks and brain nerve cells? It is designed for computers. The main objective of this study is to find a suitable set of weights for different categories of participants. The neural network learning the way that the records are tested. And when an incorrect estimate of the weight adjustment is performed. This process continues to improve so estimates are conditional upon. Neural networks are powerful estimators as well as other methods of estimating. And sometimes they do best. The main disadvantage of this method is much time spent on various parameters is selected [2]

3.8 Decision Bayesian Network

The Bayesian networks through mathematical rules based on new information combined with knowledge exists. The Bayesian network is based Bayes theory, uncertainty is a powerful tool for determining circumstances a very simple form of a New Bayesian classification [18]

3.9 Evaluation Methods

One of the important subjects in K-Means clustering is determining number of optimum clusters. Measuring Euclid distance is one of the benchmarks for determining. The algorithm continues until the other cluster centers do not change or in other words, the elements in each cluster does not move in the other iterations. And if the convergence criteria to a predetermined threshold is

Reached, the algorithm ends. One of these criteria, the Sum squared errors or SSE. The SSE represents the best clustering (optimal number of clusters) [5]

Measures of performance that we can verify it with the CART decision tree classification method to assess the evaluation criteria that is mentioned in the following: Overall accuracy: Percentage of records that have been correctly predicted records. Profit: function of a set of coefficients revenue costs associated with weight coefficients is made. A good model this function must be started from zero to a maximum point and come back and modeling in a weak form of the line of ascending, descending or direct to be seen. ROC curve: ROC curve is an indicator for measuring the performance of a model of the area under the curve is more accurate indication. Index Lift: Sample rate depending on the sort of confidence, predict the parameters of the basic units of society as a whole shows Lift [10]

4. Implementation Model

Place In order to implement the model, First step to determine the optimal number of clusters in K-means clustering and clustering customers based on demographic variables. And then the second step is to evaluate binary classification methods (decision tree QUEST, decision tree C5.0, decision tree CHAID, decision trees CART, Bayesian networks, Neural networks) to predict customers churn.

4.1 K-Means Clustering Implementation

Clustering to our data, the demographic variables that are more descriptive aspects of our data, it is also the K-Means clustering due to the ease of application we used. The clustering of the K-Means, the number of clusters is of great importance and will affect the results of our optimal. Therefore, the SSE criterion for evaluating the quality of clustering is to evaluate the number of clusters given the low volume of data to compare the number of clusters to 2 clusters, we start. Results for clusters 2, 3, 4, 5 and 6 represent respectively SSE index rate 2.198561, 2.156033, 1.859767, 1.908669, 1.933044. Clustering with 4 clusters of less than SSE Materials cluster with cluster 2,3,5 and 6, are, in fact, will show better performance. After the K-means clustering with four clusters, we used.

Reconstructing A characteristic of each cluster is as follows:

Cluster 1: a cluster of customers or employees, mostly engineers. And the level of monthly income between 5 to 10 million Rial, with an average age between 30 to 40 years (more than 97 %), Cluster 2:

Customer educated (82 % or higher degree) And high income between 10 to 20 million Rial (95 %) and employee jobs, medical markets, Cluster 3: consumers, workers or farmers (with more than 78 %) and having income between 2.50 to million Rial (about 95%) are mostly high school graduates and school education (90 %), Cluster 4: The customer has the job market (95 %) with age between 30 to 40 years), about 72% and education diploma (91%) and income between 5 to 1 million Rial (over 81%)

4.2 Evaluation Of Classification Methods For Predicting Customer Churn

The clustering of these variables, variables perception customer of the target variable, Churn Index. Binary classification evaluation methods to provide the optimum method for predicting which customers will churn. In this regard, Mel Shaw binary decision tree classification techniques to assess the Quest, the decision tree QUEST, decision tree C5.0, decision tree CHAID, decision trees CART, Bayesian networks, neural networks, which we. As we see in Tab.3. Decision tree CART technique has better performance than other techniques, perhaps unusual that algorithm shows a better performance but due to the fact that the data collection results are not far-fetched.

The results obtained from the above rules, we can conclude the following:

First cluster (customers or employees, mostly engineers): Churn largest in this group of customers that although more activity is now in Iran Insurance. But degree of dissatisfaction with the answer to problems is more important and will encourage them to transition from. Second cluster (customers educated): For this group of customers is more important than the amount the insurer. Third cluster (our farmers and workers): For this group of little interests to customers who have insurance, currently, the company's obligation to pay compensation and duration of insurance other than the insurer is very important. Four clusters (consumer market): For this group of customers, the major. Current level of mutual trust and reputation and is recognized This group of customers and generally do not tend to change in future influencing the choice of insurer If it be the satisfaction of other insurance companies What insurance company and willingness to work as a factor in this group of customers churn.

Tab .3 Evaluate the predictive binary techniques

Prediction method	Overall accuracy	Profit	ROC curve	Index Lift
decision trees CART	99.66	215	1	3.33
Bayesian networks	99.33	215	1	3.33
neural networks	99.66	210	1	3.33
decision trees Quest	99.33	205	1	3.33
decision trees C5.0	98.66	200	0.99	3.33
decision trees CHAID	98.66	200	0.99	3.33

5. Summaries

From what was said to be acknowledged that the use of data mining methods to predict churn customers is something common reason is the diversity, strength and flexibility of these techniques. Given the optimal method to achieve correct results in predicting customer churn it is undeniable. Therefore, we determine the optimal number of clusters in K-Mans clustering and clustering customers based on demographic variables and then the second step of the binary classification methods (decision tree QUEST, decision tree C5.0, decision tree, align, decision trees, CART, Bayesian networks, neural networks) to predict churn paid customers. The results show better performance than other techniques CART decision tree technique ,perhaps that algorithm shows a better performance but due to the fact that the data collection results are not far-fetched. Patterns were extracted by decision tree and show that most churn customers are in officers or engineers. Most of their activity is in Iran Insurance but are not satisfy with services provided and eager to churn. In educated group, the main reason for churning was the behavior of insurer and for those who work in market, reputation of insurer and well-knowing was very important. In worker group with lower tendency toward insurance, time intervals of policy and insurer commitments are important factors. Results of data mining methods provide an opportunity for managers and marketing professionals to make decision and choose suitable strategies to prevent churn of customers and let them go to other companies

6. References

- [1] Wai-Ho Au; Keith C. C. Chan; Xin Yao. A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction. *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, Vol. 7, No. 6, Dec 2003, PP: 532-545
- [2] Micheal J. A. Berry ; Gordon S.linoff. *EBook Data Mining Technique for marketing Sales and CRM*: Wiley Publishing, Inc., Indianapolis. Indiana, 2004
- [3] Jonathan Burez; Dirk Van den Poel. CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services, *Expert Systems with Applications*, Vol.32, 2007, PP: 277–288
- [4] Ding-An Chiang; Yi-Fan Wang; Shao-Lun Lee; Cheng-Jung Lin. Goal-oriented sequential pattern for network banking churn analysis. *Expert Systems with Applications*, Vol. 25, 2003, PP: 293-302
- [5] Daniel Westreich ; Justin Lessler ; Michele Jonsson Funk . Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology*, Vol.63, 2010, PP: 826-833
- [6] Kristof Coussement; Dirk Van den Poel. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques, *Expert Systems with Applications*, Vol.34,2008, PP : 313-327
- [7] XIA Guo-en; JIN Wei-dong. Model of Customer Churn Prediction on Support Vector Machine. *Systems Engineering Theory & Practice*, Vol. 28, 2008, PP: 71-77
- [8] John Hadden ; Ashutosh Tiwari ;Rajkumar Roy; Dymitr Ruta. Computer assisted customer churn management: State-of-the-art and future trends. *Computers&OperationsResearch*, Vol. 34, PP:2902-2917
- [9] Shin-Yuan Hung; David C. Yen; Hsiu-Yu Wang. Applying data mining to telecom churn management. *Expert Systems with Applications*, Vol. 31, 2006, PP: 515–524
- [10] Chih-Ping Wei ; I-Tang Chiu. Turning telecommunications to Churn prediction:a data mining approach. *Expert Systems with Applications*, Vol. 23, 2002, PP: 103–112
- [11] liang ; Wei-wei ;Yuan-yuan. An Empirical Study of Customer Churn in E-Commerce Based on Data Mining. *Management and Service Science (MASS)*, 2010 International Conference on, 24-26 Aug2010, PP: 1 - 4
- [12] Lima; Mues ; Baesens. Domain knowledge integration in data mining using decision tables: case studies in churn prediction. *Data Mining and Operational Research*, Vol. 60 , 2009,PP:1096-1106
- [13]Neslin, Gupta, Kamakura, Mason, C. (2006). Defection Detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, Vol.43 (2), 2006,PP: 204-211
- [14] Saradh ; Palshikar. Employees churn prediction. *Expert Systems with Applications*, Vol. 38, PP: 1999–2006
- [15]Dirk Van den Poel; Bart Lariviere. Customer attrition analysis for financial services using proportional hazard models. (*European Journal of Operational Research*, Vol.157, 2004, PP: 196–217
- [16]Wouter Verbeke ; David Martens ; Christophe Mues ; Bart Baesens. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, Vol.38 ,2011,PP: 2354–2364
- [17] Yaya Xie ; Xiu Li; E.W.T. Ngai ; Weiyun Ying. Customers churn prediction using improved balanced random forests. *Expert Systems with Applications*, Vol.36, 2009, PP: 5445–5449
- [18]Han; Kamber. *Data Mining: Concepts and Techniques*, Second. Morgan Kaufman Publisher, 2006, PP: 383-407 .
- [19]Martens, D., De Backer, M., Haesen, R., Baesens, B., Mues, C., & Vanthienen, J. (2006). Ant-based approach to the knowledge fusion problem. In M. Dorigo, L. Gambardella, M. Birattari, A. Martinoli, R. Poli, & T. Stützle (Eds.), *Ant colony optimization and swarm intelligence, fifth international workshop. ANTS 2006* (Vol. 4150, pp. 84–95). Berlin, Germany: Springer-Verlag
- [20] Buckinx, Van den Poel. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *Europeaz Journal of Operational Research*, Vol.164,2005, pp.252-268

- [21] Chandar , Laha , Krishna. Modeling churn behavior of bank customers using predictive data mining techniques , 2006,Institute for Development and Research in Banking Technology(IDRBT) Castle Hills, Hyderabad-500057,2006
- [22] Coussement , Benoit , Van den Poel. Improved marketing decision making in a customer churn prediction context using generalized additive models , 2010(Expert Systems with Applications, Vol.37 ,2010,pp. 2132–2143
- [23]]K.W.De Bock and D.Van den Poel, ”Ensembles of Probability Estimation Trees for Customer Churn Prediction”, TRENDS IN APPLIED INTELLIGENT SYSTEMS, PT II, PROCEEDINGS Book Series: Lecture Notes in Artificial Intelligence , Vol.6097,2010,pp.57-66
- [24] .Hwang,T. Jung and E.Suh ,”An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry”,Expert Systems with Applications, Vol.26,2004, pp.181–188
- [25] K.Morik , H.opcke ,” Analyzing Customer Churn in Insurance Data”,Knowledge Discovery in Databases: PKDD 2004 Lecture Notes in Computer ScienceVol.3202,2004, pp.325-336, doi: 10.1007/978-3-540-30116-5_31
- [26] D.Popović and B.Dalbelo Bašić ,” Churn Prediction Model in Retail Banking Using Fuzzy C-Means Algorithm”,University of Zagreb, Faculty of Electrical Engineering and Computing, Vol.33, 2009,pp.243-247
- [27] K.M.Osei-Bryson, ” Evaluation of decision trees: a multi-criteria approach. Computers & Operations Research, Vol.31 , 2004,pp.1933-1945
- [28] C.F.Tsai and Y.H.Lu ,”Customer churn prediction by hybrid neural networks”, Expert Systems with Applications, Vol.36, 2009,pp.12547-12553
- [29] C.F.Tsai and M.Y. Chen ,” Variable selection by association rules for customer churn prediction of multimedia on demand”, Expert Systems with Applications, Vol.37 , 2010,pp.2006–2015
- [30] B.H.Chu, M.S.Tsai and C.S. Ho ,” Toward a hybrid data mining model for customer retention” , Knowledge-Based Systems ,Vol.20,2007, pp. 703-718
- [31] SPSS Inc, Clementine 12.0 Algorithms Guide, 2007