

A DUAL-STEP MULTI-ALGORITHM APPROACH FOR CHURN PREDICTION IN PRÉ-PAID TELECOMMUNICATIONS SERVICE PROVIDERS

Ali Tamaddoni Jahromi, Mehrad Moeini, Issar Akbari, Aram Akbarzadeh

Department of Industrial Engineering, Tarblat Modares University, Tehran, Iran

Abstract: Nowadays customer churn has become the main concern of companies which are active in different industries. Among all industries which suffer from this issue, telecommunications industry can be considered in the top of the list with approximate annual churn rate of 30%. Dealing with this problem, there exist different approaches via developing predictive models for customer churn but due to the nature of pre-paid mobile telephony market which is not contract-based, customer churn is not easily traceable and definable, thus constructing a predictive model would be of high complexity. Handling this issue, in this study, we developed a dual-step model building approach, which consists of clustering phase and classification phase. With this regard firstly, the customer base was divided into four clusters, based on their RFM related features, with the aim of extracting a logical definition of churn, and secondly, based on the churn definitions that were extracted in the first step, different algorithms were utilized with the intention of constructing predictive models for churn in our developed clusters. Evaluating and comparing the performance of the employed algorithms based on “gain measure”, we concluded that employing a multi-algorithm approach in the model constructing step, instead of single-algorithm one, can bring the maximum gain among the tested algorithms.

Keywords: Customer Churn, Data mining, Telecommunications Industry

Introduction

Nowadays customer churn has become the main concern of firms in all industries (Neslin, Gupta, Kamakura, Lu, & Mason, 2006), and companies, regardless of the industry that they are active in, are dealing with this issue. Customer churn can blemish a company by decreasing profit level, losing a great deal of price premium, and losing referrals from continuing service customers (Reichheld & Sasser, 1990).

Considering the churn rate of different industries, one can find that the telecommunications industry is one of the main targets of this hazard such that the churn rate in this industry ranges from 20 to 40 annually (Berson, Smith, & Therling, 1999; Madden, Savage, & Coble-Neal, 1999). Customer churn in mobile telecommunications refers to “the movement of subscribers from one provider to another” (Wei & Chiu, 2002).

Adopting a targeted and proactive approach toward managing customer churn, companies try to identify the close future churners and then target these customers with special programs or incentives to prevent them from churning. Targeted proactive programs have the potential advantages of having lower incentive costs, because the incentive may not have to be as high as when the customer has to be “bribed” not to leave at the last minute.

However, this system would be wasteful if churn prediction is inaccurate, because then companies are wasting incentive money on customers who would have stayed anyway, this threat elucidates the need for an accurate model for churn prediction (Coussement & Van den Poel, 2008b; Neslin, Gupta, Kamakura, Lu, & Mason, 2006). This model has to be able to recognize the customers which tend to churn in close future. With this regard the first step that has to be taken is to define the “Churn” and the “Churner”, but due to the nature of pre-paid mobile telephony market which is not contract-based, customer churn is not easily traceable and also definable, thus building a predictive model would be of high complexity.

The study at your disposal aims at developing a predictive model for customer churn in pre-paid mobile telephony companies. With this regard the first step is to give a sensible definition for churn in such companies and afterwards construct the predictive model based on the extracted definition.

1 - Literature Review

Model building for churn prediction is strongly dependent on data mining techniques due to the better performance of machine learning techniques than the statistical techniques for non-parametric dataset (Baesens, Viaene, Van den Poel, Vanthienen, & Dedene, 2002; Bhattacharyya & Pendharkar, 1998).

Data mining is “the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules” (Berry & Linoff, 2004) and it involves selecting, exploring and modeling large amounts of data to uncover previously unknown patterns, and finally comprehensible information, from large databases (Shaw, Subramaniam, Tan, & Welge, 2001).

What data mining tools do is to take data and construct a model as a representation of reality. The resulted model describes patterns and relationships, present in the data (Rygielski, Wang, & Yen, 2002).

The application of data mining tools in CRM is an emerging trend in global economy. Since most companies try to analyze and understand their customers' behavior and characteristic, for developing a competitive CRM strategy, data mining tools have become of high popularity (Ngai, Xiu, & Chau, 2009).

Among all types of models that can be built via data mining, Classification is the most frequent one, especially in CRM, and it is capable of predicting the effectiveness or profitability of a CRM strategy through prediction of the customers' behavior (Ahmad, 2004; Carrier & Povel, 2003; Ngai, Xiu, & Chau, 2009), and can be defined as the process of finding a model that describes and distinguishes data classes, for the purpose of being able to use the model to predict the class of objects whose class label is unknown (Han & Kamber, 2006).

Considering all existing classification techniques Neural Network and Decision Tree are of high frequency of use, respectively (Ngai, Xiu, & Chau, 2009; Wei & Chiu, 2002).

Neural networks are a class of powerful, general-purpose tools readily applied to prediction, classification, and clustering. Neural networks have the ability to learn by example in much the same way that human experts gain from experience (Berry and Linoff, 2004).

On the other hand Decision Tree is a tree-shaped structure with different algorithms such as CART (Breiman, Friedman, Olshen, & Stone, 1984), CHAID (Kass, 1980), and C5.0 (Rulequest Research, 2008), that represents sets of decisions and is able to generate rules for the classification of a data set (Lee & Siau, 2001).

There exist different studies in various industries regarding developing computer assisted models for churn prediction (Coussement & Van den Poel, 2008a; Wei & Chiu, 2002; Yan, Fassino, & Baldasare, 2005; Coussement & Van den Poel, 2008b; Hung, Yen, & Wang, 2006; Zhao, Li, Li, Liu, & Ren, 2005).

In (2002) Wei & Chiu developed a model for subscriber churn prediction in telecommunications industry. They noted that subscriber churn is not an instantaneous occurrence that leaves no trace and they believed that before an existing subscriber churn, his/her call pattern might change.

Based on this assumption, they randomly considered a set of Observation, Retention, and Prediction time periods and by dividing the observation period into "S" sub-periods, they defined the following features:

1. Minutes of Use of a subscriber in the first sub-period (**MOU_{initial}**)
2. Frequency of Use of a subscriber in the first sub-period (**FOU_{initial}**)
3. Sphere of Influence of a subscriber in the first sub-period (**SOI_{initial}**)

4. **ΔMOU_s** : The change in MOU of a subscriber between the sub-period s-1 and s (for $s=2,3,\dots,n$) and is measured by $\Delta\text{MOU}_s = (\text{MOU}_s - \text{MOU}_{s-1} + \delta)/(\text{MOU}_{s-1} + \delta)$ where $\text{MOU}_1 = \text{MOU}_{\text{initial}}$ and $\delta = 0.01$.
5. **ΔFOU_s** : The change in FOU of a subscriber between the sub-period s-1 and s (for $s=2,\dots,n$) and is calculated as $\Delta\text{FOU}_s = (\text{FOU}_s - \text{FOU}_{s-1} + \delta)/(\text{FOU}_{s-1} + \delta)$
6. **ΔSOI_s** : The change in SOI of a subscriber between the sub-period s-1 and s (for $s=2,\dots,n$) and calculated as $\Delta\text{SOI}_s = (\text{SOI}_s - \text{SOI}_{s-1} + \delta)/(\text{SOI}_{s-1} + \delta)$.

In their model building phase they tested the role of different variables such as desired class ratio, number of sub-periods in the observation period, and also length of retention period, on model's accuracy. Their analysis revealed that the desired hit ratio equal to 1:2 (churner : non-churner), the retention length of 7 days, and the number of sub-periods of 2 can leverage the model accuracy to its optimum level.

2 - Methodology and Result

The raw data that were utilized in order to construct the churn predictive model, was the call record data (i.e. Date of call, Time of call, Duration of call, Type of Call, and Cost) of 34504 users of a mobile telecommunications service provider company in Iran in a time period from 1 November 2007 to 30 April 2008.

The first hurdle that we faced with in the initial steps of model building phase was the "Churn Definition" problem. In almost all previous studies, the customers of the service provider were its subscribers who had a contract with the company. Consequently, "Churn" in such conditions could be defined as the terminating the contract from customer's side or not renewing it after its expiry date. But circumstances would be different about pre-paid telecommunications service providers. In such companies there is no contract between the company and the clients to be expired or terminated. In other words churn in such cases happens with no tracking point such as terminating the contract or not renewing it and its recognition becomes complicated.

To shed light on the issue, imagine a database of customers, consisting a number of customers with different calling behavior, some of them use their cell phones every day, but the others use it every 2, 3, ..., or 20 days. Now if we define a churner as "a person who has not used his/her cell phone for 7 days" a considerable part of our customers who use their cell phone occasionally (i.e. every 8, 9, ..., 20 days) would be considered as a "churner", mistakenly. On the other hand if we take a longer time span and define a churner as "a person who hasn't used his/her cell phone for 25 days" our model may suffer from inability in recognizing the real churners.

The above discussed wrong signals would increase the number of False Negatives (FN) and False Positives (FP) in our predictive model and consequently lower the level of model's accuracy.

Tackling this problem our model building phase was broken into two parts: (1) clustering phase for defining the churn (2) Classification phase for predicting the churn

With the aim of extracting a logical definition for churn, the raw data was employed for constructing the relevant features in accordance with the prior studies in this realm (Ansari, Kohavi, Mason, & Zheng, 2000; Hung, Yen, & Wang, 2006), and 5 individual interviews with telecom experts.

The outcome is the following features that were constructed to be used in clustering phase:

1. **Call Ratio:** proportion of calls which has been made by each customer with more than one day time distance to his/her total number of calls.
2. **Average Call Distance:** the average time distance between one's calls
3. **Max Date:** the last date in our observed time period in which a call has been made by a specific customer
4. **Min Date:** the first date in our observed time period in which a call has been made by a specific customer
5. **Life:** the period of time in our observed time span in which each customer has been active
6. **Max-Distance:** the maximum time distance between two calls of an specific person in our observed period
7. **No-of-days:** number of days in which a specific customer has made or received a call
8. **Total-no-in:** the total number of incoming calls for each client in our observed period
9. **Total-no-out:** the total number of outgoing calls for each client in our observed period
10. **Total-cost:** the total money that each customer has been charged for using the services in the specific time period under study
11. **Total-duration-in:** the total duration of incoming calls for a specific customer in our observed time span
12. **Total-duration-out:** the total duration of outgoing calls for a specific customer in our observed time span

UTILIZING THE ABOVEMENTIONED SET OF RFM RELATED 12 FEATURES, THE CUSTOMER BASE WAS DIVIDED INTO 4 INDIVIDUAL CLUSTERS BY THE USE OF TwOStEP (SPSS INC, 2007) CLUSTER TECHNIQUE, WITH DIFFERENT "MAX-DISTANCE"'S, AND BY CONSIDERING THE "PREDICTION PERIOD" LENGTH AS TWICE THE "MAX-DISTANCE" AND ALSO BASED ON WEI & CHIU (2002) FINDINGS, WE OBTAINED FOUR SETS OF OBSERVATION, RETENTION, AND PREDICTION PERIODS FOR OUR FOUR EXTRACTED CLUSTERS (SEE TABLE 1).

Cluster No.	Observation Length	Retention Length	Prediction Length	Cluster's Max-Distance
1	18 Mar to 16 Apr (30days)	17 Apr to 23 Apr (7 days)	24 to 30 Apr (7 days)	3.88 days
2	14 Mar to 12 Apr (30 days)	13 to 19 Apr (7 days)	20 to 30 Apr (11 days)	5.66 days

3	21 Feb to 22 Mar (30 days)	23 to 29 Mar (7 days)	30 Mar to 30 Apr (32 days)	16.33 days
4	4 Feb to 5 Mar (30 days)	6 to 12 Mar (7 days)	13 Mar to 30 Apr (49 days)	24.88 days

Table 3: Prediction, Retention, and Observation Period for Each Extracted Cluster of Customers

AFTERWARDS BY CONSIDERING TWO SUB-PERIODS OF 15 DAYS IN THE OBSERVATION PERIOD OF EACH CLUSTER, , WE CONSTRUCTED THE FOLLOWING FEATURES FOR EVERY SINGLE CLUSTER BASED ON WEI & CHIU (2002)'S STUDY AND BY THE USE OF THE CALL RECORDS DATA OF THE OBSERVATION PERIOD:

1. ***IMOU_{initial}***: Incoming MOU of a subscriber in the first sub-period
2. ***IFOU_{initial}***: Incoming FOU of a subscriber in the first sub-period
3. ***OMOU_{initial}***: Outgoing MOU of a subscriber in the first sub-period
4. ***OFOU_{initial}***: Outgoing FOU of a subscriber in the first sub-period
5. **Δ *IMOU₂***: The change in IMOU of a subscriber between the sub-period 1 and 2.
6. **Δ *IFOU₂***: The change in IFOU of a subscriber between the sub-period 1 and 2
7. **Δ *OMOU₂***: The change in OMOU of a subscriber between the sub-period 1 and 2.
8. **Δ *OFOU₂***: The change in OFOU of a subscriber between the sub-period 1 and 2.
9. **Churn**: binary churn labels for each client according to their churn status in prediction period. Regarding this a churner is a person who doesn't have a call record in the prediction period, otherwise he/she would be considered as a non-churner

Utilizing the abovementioned features and considering the hit ratio of 1:2 (churner : non-churner), we constructed different predictive models for each cluster, by employing the Neural Networks technique and also different algorithms of the Decision Tree technique. With this regard CART, C5.0, and CHAID algorithms among Decisions Tree algorithms were applied and their performance was compared with the Neural Networks based constructed model which was a three layer Neural Network including 8 neurons in input layer, 3 neurons in the hidden layer, and one neuron in its output layer. As the validation technique the single split model validation was adopted which has been proven to be an accurate validation method (Burez & Van den Poel, 2008b; Montgomery, Li, Srinivasan, & Liechty, 2004; Swait & Andrews, 2003).

Tables 2 to 5 represent the performance of our constructed models with different algorithms on our clusters, based on gain measure for top %10 and %20 clients of each cluster.

Cluster No	%Gain for Percentile = 10	%Gain for Percentile = 20
1	77.9	80.4
2	10	29.6
3	15	30.6

4	22.5	63.6
---	------	------

Table 2: Performance of Developed Decision Tree (C5.0) Predictive Models Based on Gain Measure

Cluster No	%Gain for Percentile = 10	%Gain for Percentile = 20
1	59	79.4
2	36	67.4
3	15	30.6
4	40	80

Table 3: Performance of Developed Decision Tree (CHAID) Predictive Models Based on Gain Measure

Cluster No	%Gain for Percentile = 10	%Gain for Percentile = 20
1	46.3	77.8
2	54	66.7
3	15	30.6
4	17	45.5

Table 0: Performance of Developed Decision Tree (CART) Predictive Models Based on Gain Measure

Cluster No	%Gain for Percentile = 10	%Gain for Percentile = 20
1	72.2	77.8
2	33.3	33.3
3	15	30.6
4	25	60

Table 5: Performance of Neural Networks Predictive Models Based on Gain Measure

Comparing the performance of our developed models based on gain measure, one can find that Decision Tree algorithms outperform the Neural Networks algorithm. Furthermore, examining the gain which has been brought by each of the Decision Tree based constructed models, we can conclude that maximum performance will be achieved by utilizing different algorithms of Decision Tree technique in different clusters. Table 6 depicts the most appropriate algorithm among the tested algorithms, for model building in each cluster.

Cluster No.	Appropriate Algorithm
Cluster 1	Decision Tree C5.0 Algorithm
Cluster2	Decision Tree CART Algorithm
Cluster3	All Tested Algorithms
Cluster 4	Decision Tree CHAID Algorithm

Table 6: The Appropriate Algorithm for Model Building in Each Cluster

Consequently, by applying this multi algorithm approach the gain factor for each cluster would be in accordance with table 7.

Cluster No	%Gain for Percentile = 10	%Gain for Percentile = 20
1	77.9	80.4
2	54	66.7
3	15	30.6
4	40	80

Table 7: Performance of the Multi-algorithm Model Building Approach on Our Developed Clusters Based on Gain Measure

While the gain factor of random sampling is %20 for the top %20 of the customer base in all clusters, table 7 depicts that the developed model is able to bring the gain factor of %80.4, %66.7, %30.6, and %80 for the top %20 of the customer base of our four developed clusters, respectively. This implies that by applying the developed multi algorithm predictive model, choosing a sample size of only %20 of each cluster’s customer base is enough for identifying %80.4, %66.7, %30.6, and %80 of the total number of churners in each of our four clusters, respectively.

Figures 1 to 4 illustrate the gain chart of the constructed model on each cluster.

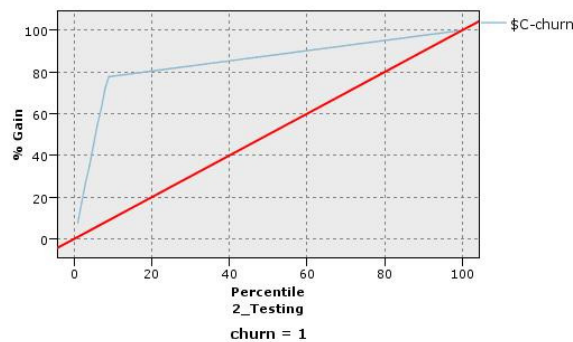


Figure 1: Gain Chart of Decision Tree C5.0 Algorithm for Cluster 1

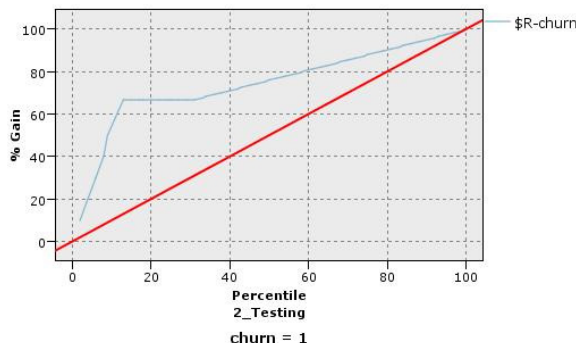


Figure 2: Gain Chart of Decision Tree CART Algorithm for Cluster 2

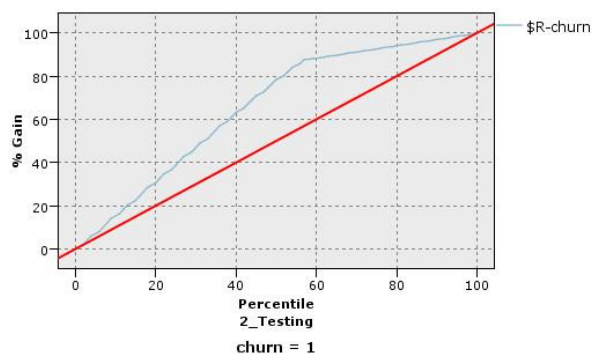


Figure 3: Gain Chart of Decision Tree CART Algorithm for Cluster 3

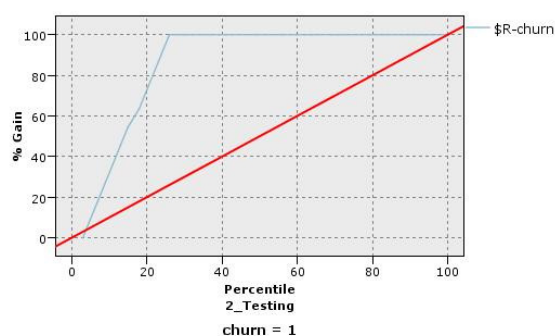


Figure 4: Gain Chart of Decision Tree CHAID Algorithm for Cluster 4

As presented in the above figures, the developed predictive model on each cluster has a considerable better performance than the random sampling (diagonal line).

Conclusion

In this study we aimed at constructing a predictive model for customer churn in pre-paid mobile telephony companies. With this regard firstly, the customer base was divided into four clusters based on their RFM related features, with the intention of extracting a logical definition of churn and secondly, based on the churn definitions that were extracted in the first step, we went through the model building phase and tested the performance of different model building algorithms including Neural Networks, Decision Tree (C5.0), Decision Tree (CART), and Decision Tree (CHAID).

Evaluating and comparing the performance of the employed algorithms based on “gain measure”, we concluded that Decision Tree algorithms in all clusters outperform Neural Networks, based on “Gain Measure” for top 10% and 20% of our customer base.

Furthermore it was noticed that a Decision Tree algorithm does not possess the same performance in all developed clusters. Thus it was concluded that adopting a combination of Decision Tree algorithms for model building in our developed clusters can bring the maximum model performance. Comparing the performance of this multi-algorithm approach (see table 7), with the performance of other tested single algorithm approaches (see tables 2 to 5), one can find that adopting the multi-algorithm approach can bring a considerable superior performance for our predictive model.

Bibliography

- Ahmad, S. (2004). Applications of data mining in retail business. *Information Technology*, 2, 455-459.
- Ansari, S., Kohavi, R., Mason, L., & Zheng, Z. (2000). Integrating e-commerce and data mining: architecture and challenges. *WEBKDD'2000 Workshop on Web Mining for E-Commerce Challenges and Opportunities*.
- Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., & Dedene, G. (2002). Bayesian neural network learning for repeat purchase modeling in direct marketing. *European Journal of Operational Research*, 138 (1), 191-211.
- Berry, M., & Linoff, G. (2004). *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management* (2nd Edition ed.). Indianapolis: Wiley Publishing Inc.
- Berson, A., Smith, S., & Therling, K. (1999). *Building data mining applications for CRM*. New York: McGraw-Hill.
- Bhattacharyya, S., & Pendharkar, P. (1998). Inductive, evolutionary and neural techniques for discrimination: A comparative study. *Decision Sciences*, 29, 871-900.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. New York: Chapman & Hall / CRC.
- Carrier, C., & Povel, O. (2003). Characterising data mining software. *Intelligent Data Analysis*, 7, 181-192.
- Coussement, K., & Van den Poel, D. (2008a). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34, 313-327.
- Coussement, K., & Van den Poel, D. (2008b). Integrating the voice of customers through call center emails into a decision support system for churn prediction. *Information & Management*, 45, 164-174.
- Han, J., & Kamber, M. (2006). *Data Mining: concepts and techniques* (2nd Edition ed.). San Francisco: Elsevier.
- [12] Hung, S., Yen, D., & Wang, H. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, 31, 515-524.
- Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29 (2), 119-127.
- Lee, S., & Siau, K. (2001). A review of data mining techniques. *Industrial Management and Data Systems*, 101 (1), 41-46.
- Madden, G., Savage, S., & Coble-Neal, G. (1999). Subscriber churn in the Australian ISP market. *Information Economics and Policy*, 11, 195-207.
- Montgomery, A., Li, S., Srinivasan, K., & Liechty, J. (2004). Modeling online browsing and path analysis using click stream data. *Marketing Science*, 23 (4), 579-595.
- Neslin, S., Gupta, S., Kamakura, W., Lu, J., & Mason, C. (2006). Defection Detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, XLIII, 204-211.
- Ngai, E., Xiu, L., & Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36, 2592-2602.
- Reichheld, F., & Sasser, W. (1990). Zero defection: Quality comes to services. *Harvard Business Review*, 68 (5), 105-111.
- Rulequest Research. (2008). *Information on see 5/C5.0*. Retrieved from Rulequest.com: <http://www.rulequest.com/see5-info.html>

- Rygielski, C., Wang, J., & Yen, D. (2002). Data mining techniques for customer relationship management. *Technology in Society* , 24 (4), 483-502.
- Shaw, M., Subramaniam, C., Tan, G., & Welge, M. (2001). Knowledge management and datamining for marketing. *Decision Support Systems* , 31, 127-137.
- SPSS Inc. (2007). *Clementine 11.1 Algorithms Guide*. Integral Solutions Limited.
- [25] Swait, J., & Andrews, R. (2003). Enriching scanner panel models with choice experiments. *Marketing Science* , 22 (4), 442-460.
- Wei, C., & Chiu, I. (2002). Turning telecommunications call details to churn prediction: a data mining approach. *Expert Systems with Applications* , 23, 103-112.
- Yan, L., Fassino, M., & Baldasare, P. (2005). Predicting Customer Behavior via Calling Links. *Proceedings of International Joint Conference on Neural Networks*, (pp. 2555-2560). Montreal.
- Zhao, Y., Li, B., Li, X., Liu, W., & Ren, S. (2005). Customer churn prediction using improved one-class support vector machine. *Lecture Notes in Computer Science* , 3584, 300-306.