



Trends in
**Applied Sciences
Research**

ISSN 1819-3579



Academic
Journals Inc.

www.academicjournals.com

Using Process Mining for Customer Retention

¹N. Abedzadeh, ²N. Nematbakhsh and ²M.A. Nematbakhsh

¹Department of Computer Engineering, International Esfahan University, Iran

²Faculty of Computer Engineering, International University of Esfahan, Iran

Abstract: Recognizing churn customers and provoke them to stay in our store, is a way for customer churn management. This study does so not only through predicting churn probability but also by proposing a method to keep them. The historical information of customers and CLV, are used to form a churn model. We use the attributes identified in the churn model to divide all churners into distinct groups. Then, a churn predictor uses the churn model to predict the churn probability of a given customer. When the churn model finds that the customer has a churn probability, process mining is used to suggest specific retention policies to keep them. The Refah departments store dataset is used to analyze this method. This dataset consists of customer's data and one year sells of Refah department stores. This study's experiments show that our method has high evaluation accuracy as compared with other methods.

Key words: Customer churn, data mining, customer retention, process mining

INTRODUCTION

Customer churn is the loss of existing customers to a competitor. One way for customer churn management, is recognizing churn customers and provoke them to stay in our store. Today's several items influence the behavior of department store's customer's behavior and churning them, such as corporation's competition and changes in the needs. It is more profitable to keep and satisfy existing customers than to constantly attract new customers who are characterized by a high attrition rate (Reinartz and Kumar, 2003). So companies start to focus on customers, instead of corporations selling and services. In fact, stores increase their profits by decreasing the churn customers.

We are following to find better methods of ensuring that customers remain loyal. Predicting customers behavior is an important way to do this. Building lasting relationships with customers could be the best way for minimizing churning probability.

In order to effectively manage customer churn within a company, it is crucial to build an effective and accurate customer-churn model. To accomplish this, there are numerous predictive modeling techniques available. Some articles have predicted the probability of churn using traditional statistical models (Gerpott *et al.*, 2001). Several researchers have examined customer churn and the factors that influence churn (Aur lie and Croux, 2006). Moreover, there are some data mining techniques that can effectively assist with the selection of customers most prone to churn (Hung *et al.*, 2006). Decision Tree could be the best way that modeling churn (Petersen *et al.*, 2009; Bong-Horng *et al.*, 2007). Some other articles showed that CLV is highly correlated with firm value using a longitudinal analysis

Corresponding Author: N. Abedzadeh, Department of Computer Engineering,
International Esfahan University, Iran

of a firm's data. Process mining is the first time that is uses for customer retention. This article uses process mining to discover a concrete workflow schema modeling all possible execution patterns registered in a given log, which can be exploited subsequently to support further-coming enactments (Greco *et al.*, 2008).

This study works on two problems. The first problem uses data mining to predict the churning probability. It calculates CLV to constructs a churn model. Then it uses this model to find the best policy for retaining them. The second problem tries to predict the churn probability of every new customer. Then it uses process mining to keep them.

MODELING TECHNIQUES FOR CUSTOMER RETENTION

System Architecture

This system works on two modes: the training and testing. The training mode uses data mining to predict the probability of churning. Historical information of customers is using to calculate CLV. The algorithm that we will show in following section is used to divide customers into distinct groups. Then we use these information to constructs a churn model. This model is used to find the best policy for retaining them. The testing mode tries to predict the churn probability of every new customer. Then it uses process mining to keep them.

The following sections explain each component in these problems and shows how it works with a real-world sample data set as an example.

Training Mode

The churn model is constructed in this mode. Some information is used to calculate CLV for one year. The historical information moreover the CLV are used to construct the churn model. Then this churn model is used for process mining to discover the best policy for retaining customers.

Calculate CLV

The CLV is generally defined as the present value of all future profits obtained from a customer over his or her life of relationship with a firm (Reinartz and Kumar, 2003). The CLV is similar to the discounted cash flow approach used in finance. However, there are two key differences (Gupta *et al.*, 2006). First, CLV is typically defined and estimated at an individual customer or segment level. This allows us to differentiate between customers who are more profitable than others rather than simply examining average profitability. Second, unlike finance, CLV explicitly incorporates the possibility that a customer may defect to competitors in the future. CLV for a customer (omitting customer subscript) is (Sunil *et al.*, 2004; Reinartz and Kumar, 2003).

$$CLV = \sum_{t=0}^T \frac{(p_t - c_t)r_t}{(1+i)^t} - AC$$

Where:

p_t : Price paid by a consumer at time t

c_t : Direct cost of servicing the customer at time t

I : Discount rate or cost of capital for the firm

r_t : Probability of customer repeat buying or being Aalive@ at time t

AC : Acquisition cost and

T : Time horizon for estimating CLV

The dataset that is used for this study is Refah departments store data. We filter data and delete the customers that their information isn't complete. Then the historical information of customers is used to calculate CLV for one year. The calculated CLV is the formula that is used for calculation CLV is pointed above. The parameters of pt , ct , i , rt , AC and T are calculated for CLV. The algorithm that is used for distinction of customers to 4 groups is pointed below:

The algorithm for finding the probability of churning

```
If((CLV1 * 1.1) <= CLV2) then churn=1.  
If((CLV1 * 0.9) <= CLV2 < (CLV1 * 1.1)) then churn=2.  
If((CLV1 * 0.6) < CLV2 < (CLV1 * 0.9)) then churn=3.  
If (CLV2 <= (CLV1 * 0.6)) then churn=4.
```

There are four groups of churning customers that are described here:

- Loyalty customers
- Almost loyalty customers
- Almost churn customers
- Churn customers

Customer loyalty describes the tendency of a customer to choose one business or product over another for a particular need.

The following section will describe the method that we have used to discover the common attributes of customer churning. We have used the historical information, the CLV of customers for one year and the groups that each customer is belonged to it.

Constructing the Churn Model

The historical information of customers, CLV and churn number are used to construct the churn model by decision tree. In the following sections we will describe historical information and decision tree to construct the churn model.

Historical Information

The historical information of customers, CLV and churn number are used to construct the churn model by decision tree. It has been shown over and over that past consumer behavior is the best predictor of future behavior. Past behavior is a much better predictor of future behavior than demographics ever will be. A visitor or buyer who repeats their behavior is more likely to continue repeating it, meaning their future value to the business is high. So when you look at a particular segment of customers, if repeating visitors or buyers are rising, then your future business with this segment of customer will be stronger than it is today. If repeaters are falling, business from this customer segment will be weaker in the future.

When you look at the repeating behavior of different groups of customers, you can make judgments about which customer groups will be most valuable in the future. You want to do everything you can to attract customers with high repeat behavior and reduce or eliminate any spending or other efforts on attracting customers with low repeat behavior.

There are about ten attributes in our dataset. We filtered these data and use just eight attributes for our dataset. Our historical information is pointed here:

- ID : Between 1-7000 customers
- Sex : 1 for men, 2 for women

- Zone : Value 12 for Esfahan city, value 13 for other cities
- Age : More than 17 years old
- CLV1 : More than 24
- CLV2 : More than 20
- ProdID : Value 1 for grain, value 2 for clothing, value 3 for household services, value 4 for kitchen services, value 5 for bread, value 6 for meat types, value 7 for pastry types, value 8 for meal types
- Churn : Value 1 for loyalty customers, value 2 for almost loyalty customers, value 3 for almost churn customers, value 4 for churn customers

Decision Tree

An important technique in machine learning, decision trees are used extensively in data mining. They are able to produce human-readable descriptions of trends in the underlying relationships of a dataset and can be used for classification and prediction tasks. The technique has been used successfully in many different areas, such as medical diagnosis, plant classification and customer marketing strategies.

Churn Model

The churn model learned by C5.0 is represented by a decision tree as illustrated in Fig. 1. Each leaf is associated with a parameter of (1/2/3/4), which means that every customer with these attributes is belonged to one group. Note that each path in the decision tree forms a rule. For example, the following path says that for customers whose $49 <= CLV2 < 66$, $69 < CLV1 <= 74$, Zone = 12, ProdID $<= 7$ and age $<= 74$, there is a likelihood that this customer is belonged to group 3 and would be almost churn.

$49 <= CLV2 < 66$ and $69 < CLV1 <= 74$ and Zone = 12 and ProdID $<= 7$ and age $<= 52$:- (3)

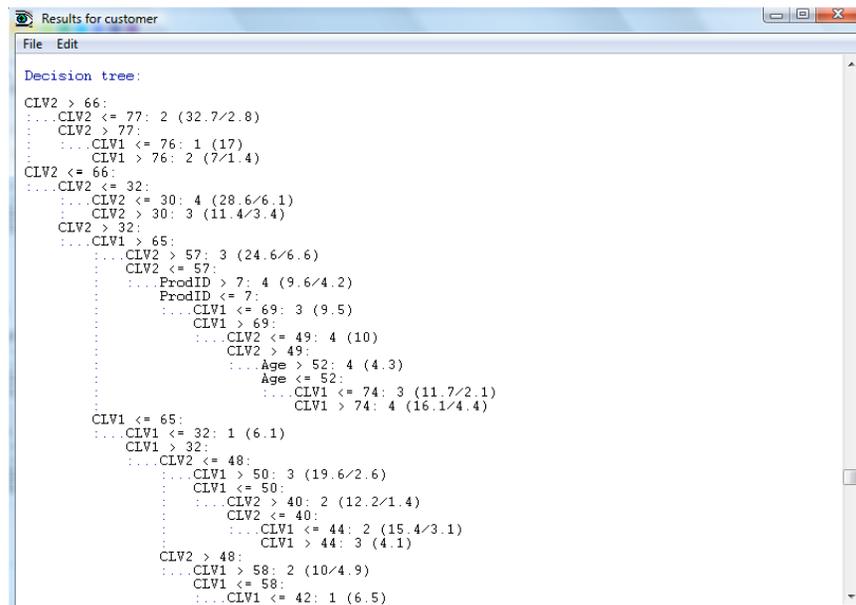


Fig. 1: A snapshot of part of churn model

Policy Model Constructor

This module constructs retention strategies by looking for proper retention incentives for potential churners. This involves two tasks. First, any implicit relationships among churners must be identified in order to create possible clusters. Second, the attributes of those clusters must be analyzed to interpret their significance. Based on the characteristics of clusters, specific policies like implement targeted telephone or mail campaigns (or policies) are proposed, to retain customers.

To accomplish the first task, the learned churn model is consulted to single out those attributes which are recognized by the model as having strong relationships with churning. Those attributes are used to segment all churners into different clusters and labels the most significant attributes in each cluster. Second, the attributes in each cluster, which conceptually represent a churner group, can be analyzed to interpret their significance. Based on the interpretation customers are divided to 4 groups and the set of policies are proposed in order to retain the customers of each group.

Using Process Mining for Customer Retention

Process mining targets the automatic discovery of information from an event log. This discovered information can be used to deploy new systems that support the execution of business processes or as a feedback tool that helps in auditing, analyzing and improving already enacted business processes. The main benefit of process mining techniques is that information is objectively compiled. In other words, process mining techniques are helpful because they gather information about what is actually happening according to an event log of an organization and not what people think that is happening in this organization.

The process mining is used to retain customers. Polices that are used to retain customers are a lot, but we can't discover the policy that is more efficient. Some polices have a different influences on different people (Chip, 1996; Nigel *et al.*, 2000). Some of the result of polices on customers can be considered for efficiency not all of them. Process mining is a method that is used to consider the efficiency of different polices.

For using process mining, two states should be considered. First, the different polices would be considered on customers in groups of 3 and 4. Second, the CLV of customers would be calculated again. The customers that their CLV is increased more than 20% would be considered for process mining. These customers would be selected to mine their process of retaining polices. The α algorithm is used to discover this process as shown in Fig. 2.

A small section of customer retention process is shown in Fig. 3. This process is done by the operator to retain customers and is proved to be the best policy for increasing CLV of group of 3. Now this process could be compare with other processes to check their conformance.

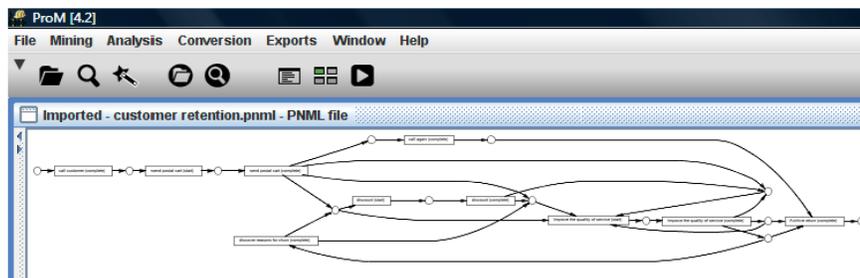


Fig. 2: A mined process for customer retention

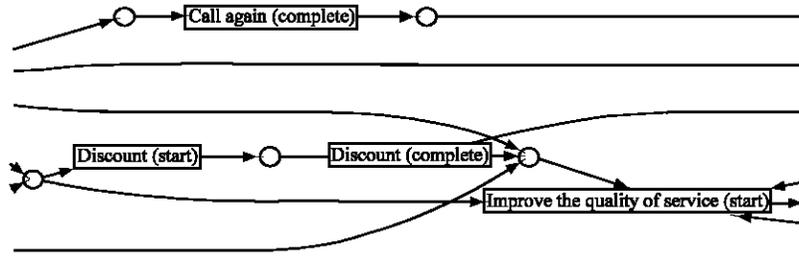


Fig. 3: A small section of customer retention process

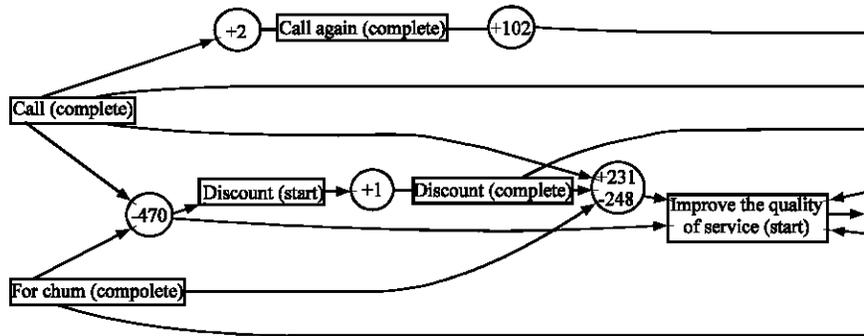


Fig. 4: A small section of conformance checking of processes

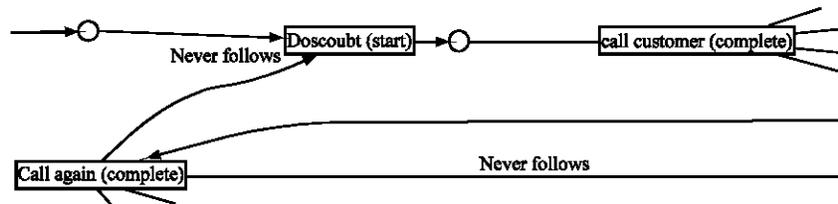


Fig. 5: A small section of conformance checking precision

The result of conformance checking is shown in Fig. 4 and 5. You can compare the best customer policy with the policies that applied for other customers. The compared model is shown that why this policy didn't have an efficient influence. As shown in Fig. 3-5, some sections of these policies never follow other sections and this is the reason that this policy didn't have an efficient influence. Now the best policy would be selected to retain customers.

Testing Mode

In the testing mode, a churner predictor prompts dialogue to solicit customer data from the user. The attributes of new customer would be entered to know that this customer is belonged to which groups as shown in Fig. 6. The result would be says that how much is the probability of churn of this customer. Figure 7 shows that this customer has 0.93 of probability of almost churn and 0.07 of probability of almost loyal.

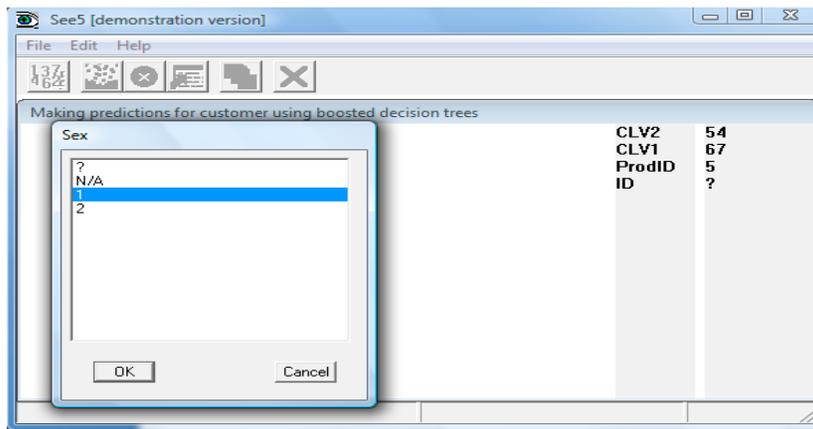


Fig. 6: Enter customer data to the churner predictor

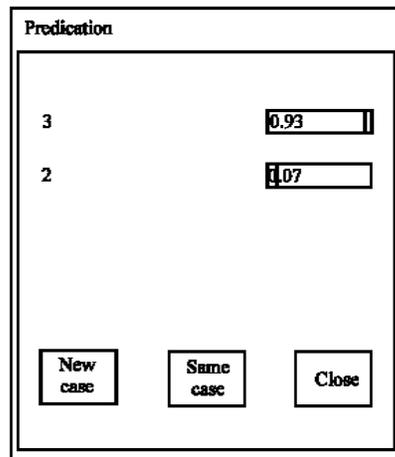


Fig. 7: Churn probability as predicted by the churner predictor

Since, the churn probability is above 90%, this customer is considered almost to churn. Such a result means that the policy maker is needed to retain it. The policy that considered as a best policy for each group is used for this customers to retain them.

Classification Technique

C5.0 is a well-known decision tree algorithm developed by Quinlan (1993). It introduces a number of extensions to the classical ID3 algorithm (Quinlan, 1986). With respect to a set of training cases T , the C5.0 algorithm can be written recursively as follows:

- If T is a leaf, the algorithm is finished
- Alternatively find a test X on some attribute such that the resulting partition T_1, T_2, \dots, T_m has the highest gain
- Apply C5.0 to T_1, T_2, \dots, T_m

This study uses C5.0 as a classification technique

Gain Ratio Criterion Vs. Absolute Gain

The classical ID3 algorithm uses a criterion called information gain (Kamber and Han, 2000). Suppose a test X is conducted on some attribute with n outcomes, which partitions the training set T into subsets of T1, T2, . . . and Tn. Let S be any set of cases and let freq(Cj, S) stand for the number of cases in S which belong to class Cj, j = 1, 2, . . . , k. Let |S| denote the number of cases in set S. To find the expected entropy of S, the surprise of each class is summed up and weighted by its occurrence frequency in S, giving Eq. 1.

$$\text{Entropy}(S) = - \sum_{j=1}^k \frac{\text{Freq}(C_j, S)}{|S|} \times \log_2 \left(\frac{\text{Freq}(C_j, S)}{|S|} \right) \tag{1}$$

After T is partitioned to n outcomes according to the test X, the expected entropy can be calculated as the weighted sum over the subsets, as defined by Eq. 2.

$$\text{Entropy}(T) = \sum_{i=1}^n \frac{|T_i|}{|T|} \times \text{Entropy}(T_i) \tag{2}$$

Now define the information gain for text X, denoted as gain(X), by Eq. 3 as a metric of gain by partitioning T according to test X. Finally select a test with the maximum information gain to be the root.

$$\text{Gain}(X) = \text{Entropy}(T) - \text{Entropy}_x(T) \tag{3}$$

This selection of a test is based on absolute information gain. Although it produces good results, the criterion has a serious deficiency: it is biased towards tests with more outcomes. In C5.0 this bias is rectified by a normalization factor defined by Eq. 4. In the equation, a split_entropy is calculated, which becomes larger if the outcome of a partition contains more subsets.

$$\text{Split - Entropy}(X) = - \sum_{i=1}^n \frac{|T_i|}{|T|} \times \log_2 \left(\frac{|T_i|}{|T|} \right) \tag{4}$$

Now a gain_ratio can be defined as a new metric measuring the information gain with respect to the complexity of splits by Eq. 5.

$$\text{Gain - Ratio}(X) = \frac{\text{Gain}(X)}{\text{Split - Entropy}(X)} \tag{5}$$

Generally speaking, the use of this gain ratio criterion to select attributes for each test is robust and can consistently give a better choice of tests than the absolute gain criterion. Note that this gain ratio was originally proposed by C4.5/ C5.0. For detailed discussions please refer to reference (Kamber and Han, 2000).

Continuous Vs Categorical Attributes

The ID3 algorithm can only deal with attributes which contain categorical values. In C5.0, continuous attributes are properly dealt with. Specifically, if all the training cases in T are sorted according to the values of a continuous attribute, represented here as A, then there

will only be a finite number of values for A. These values are denoted in order as $\{v_1, v_2, \dots, v_m\}$. Now Eq. 6 can be used to calculate $m-1$ threshold values. For each threshold partition, the whole value range can be divided into two splits and evaluated on its appropriateness.

$$\frac{v_i + v_{i+1}}{2} \quad (6)$$

Process Mining Techniques

Process mining techniques allow for extracting information from event logs. For example, the audit trails of a workflow management system or the transaction logs of an enterprise resource planning system can be used to discover models describing processes, organizations and products. The α -algorithm is used to mine processes from log files. In the following section we will explain this algorithm.

α -Algorithm Definition

The α -algorithm is used to mine processes from log files (Jiafei *et al.*, 2007). Now we can give the formal definition of the algorithm followed by a more intuitive explanation.

Definition 2.16. (Mining algorithm α) Let W be a workflow log over T .

The $\alpha(W)$ is defined as follows:

$$T_w = \{t \in T \mid \exists_{\sigma} w^t \in \sigma\}$$

$$T_1 = \{t \in T \mid \exists_{\sigma} w^t = \text{first}(\sigma)\}$$

$$T_0 = \{t \in T \mid \exists_{\sigma} w^t = \text{last}(\sigma)\}$$

$$X_w = \{(A, B) \mid A \subseteq T_w, B \subseteq T_w \wedge \forall_{a \in A} \forall_{b \in B^* \rightarrow w^b} \wedge \forall_{a_1, a_2 \in A^*} \#_{w^{a_2}} \wedge \forall_{b_1, b_2 \in B^*} \#_{w^{b_2}}\},$$

$$Y_w = \{(A, B) \in X_w \mid \forall_{(A', B') \in X_w} A \subseteq A' \wedge B \subseteq B' \Rightarrow (A, B) = (A', B')\}$$

$$P_w = \{(p_{(A, B)}) \mid (A, B) \in Y_w\} \cup \{i_w, o_w\}$$

$$F_w = \{(\alpha, p_{(A, B)}) \mid (A, B) \in Y_w \wedge \alpha \in A\} \cup \{(p_{(A, B)}, b) \mid (A, B) \in Y_w \wedge b \in B\} \\ \cup \{(i_w, t) \mid t \in T_1\} \cup \{(t, o_w) \mid t \in T_0\}$$

$$\alpha(W) = (P_w, T_w, F_w)$$

The α -algorithm works as follows. First, it examines the log traces and (Step 1) creates the set of transitions (T_w) in the workflow, (Step 2) the set of output transitions (T_1) of the source place and (Step 3) the set of the input transitions (T_0) of the sink place. In steps 4 and 5, the α -algorithm creates sets (X_w and Y_w , respectively) used to define the places of the discovered workflow net. In Step 4, the α -algorithm discovers which transitions are causally related. Thus, for each tuple $(A;B)$ in X_w , each transition in set A causally relates to all transitions in set B and no transitions within A (or B) follow each other in some firing sequence. These constraints to the elements in sets A and B allow the correct mining of AND-split/join and OR-split/join constructs. Note that the OR-split/join requires the fusion of places. In Step 5, the α algorithm refines set X_w by taking only the largest elements with

respect to set inclusion. In fact, Step 5 establishes the exact amount of places the discovered net has (excluding the source place i_w and the sink place o_w). The places are created in Step 6 and connected to their respective input/output transitions in Step 7. The discovered workflow net is returned in Step 8.

Note that no mining algorithm is able to find names of places. Therefore, we ignore place names, i.e., α is able to rediscover N if and only if $\alpha(W) = N$ modulo renaming of places.

RESULTS AND DISCUSSION

To evaluate the performance of the proposed method, we apply it to a real-world database. The Refah departments store provided the database for this study. The data set, as extracted from the store's data warehouse, included records of more than 15,000 customers described by 8 variables. The customers whose their attributes weren't complete had been removed.

This dataset is divided in two sections: training and testing. The training dataset was first fed to C5.0 to construct the churn model. The testing dataset was then used to assess the accuracy of the built model. In the training dataset, CLV is calculated for one year. The CLV of first six months is calculated for discovering that this customer is belonged to which one of the groups. Then C5.0 is used to generate rules to know the common attributes as shown in Fig. 8. The tertiary CLV is calculated for process mining to discovering the best method for retaining customers.

Process mining is used to mine the best process that is used to retain customers. The policies that are considered for each group are presented in Fig. 9.

This study has described an architecture to deal with on modeling churn and retaining churning customers. Two modes had been worked on this architecture, namely, the training and testing. The training mode had been used data mining to predict the probability of churning. Historical information of customers used to calculate CLV and divide customers into distinct groups to construct churn model. Then this model used to find the best policy for retaining them. The testing mode tried to predict the churn probability of every new customer and keep them.

```

Results for customer
File Edit
Rule 11: (22/1, lift 3.2)
  CLV1 > 49
  CLV1 <= 65
  CLV2 > 32
  CLV2 <= 46
  -> class 3 [0.917]

Rule 12: (16/1, lift 3.1)
  CLV1 > 42
  CLV1 <= 65
  CLV2 > 32
  CLV2 <= 40
  -> class 3 [0.889]

Rule 13: (13/1, lift 3.0)
  Age <= 52
  CLV1 > 65
  CLV1 <= 74
  CLV2 > 49
  CLV2 <= 57
  ProdID <= 7
  -> class 3 [0.867]

Rule 14: (5, lift 3.0)
  Age > 56
  CLV1 <= 51
  CLV2 <= 32
  -> class 3 [0.857]

Rule 15: (12/1, lift 3.0)
  CLV1 > 65
  CLV1 <= 69
  CLV2 <= 57
  -> class 3 [0.857]

Rule 16: (17/2, lift 2.9)
  CLV1 <= 48
  CLV2 <= 32
  -> class 3 [0.842]
    
```

Fig. 8: A snapshot for generating rules

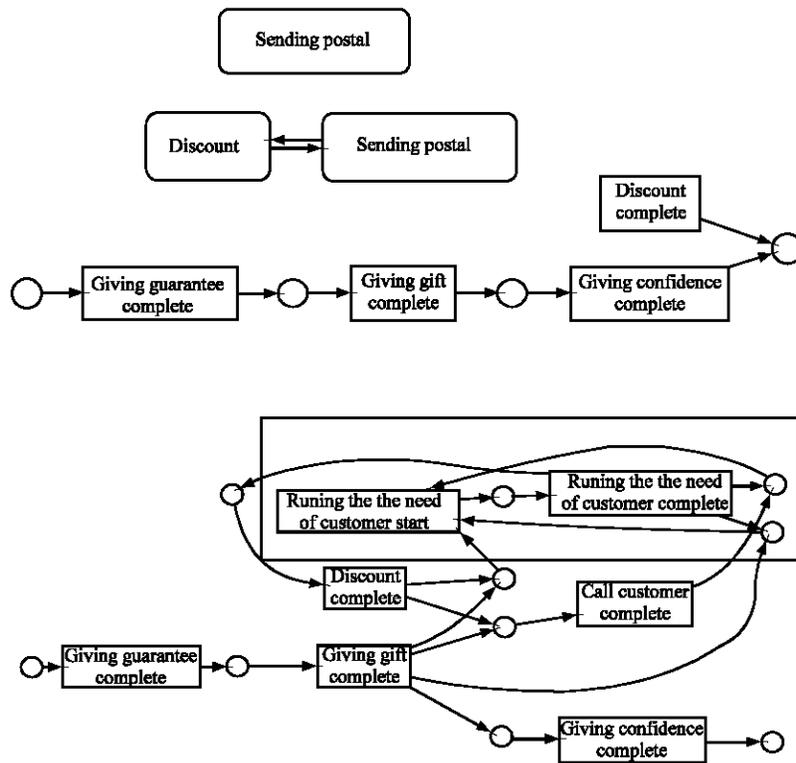


Fig. 9: Policies for each group

As a result of our article, four groups are known. The C5.0 is used to know that every new customer belonged to which group. Now we use the method that had shown in results section to retain this customer. As shown just sending postal cart is enough for the first group. They would be stay in our store and would be loyal forever. The second group is almost loyal and they may leave our stores. The discount after sending postal cart is necessary for them.

The process for the group of three is shown that sending postal cart or discount wouldn't be efficient for these customers. The best way for retaining this group is that first giving them the guarantee that the products of the company are the best and if one of the products would be damaged, we give back their money to them. We can give a note to them with the mean that we at least replace our products with a bad quality. Then we can give a gift to satisfy them. By giving confidence and discount to them, they would be come back to our store. This is a kind of persuasion that we do for retaining this group.

The group of four is really important. The considered method wouldn't be efficient for this group. After that our customer leave our company, this is the time that we must discover the reason of their churning and run their needs by calling them. Giving the guarantee and gift are the first methods of retaining them. But this process must be continuing by calling them and discovering their needs. This is the time for running their needs as possible.

As shown in our article, process mining is used for retaining customers for the first time. Until now we couldn't find the main process that is used in our stores. We had done a lot of works to retaining customers but we couldn't mine the best work. We didn't know which of the works is better than others or which of them would be more efficient for customers.

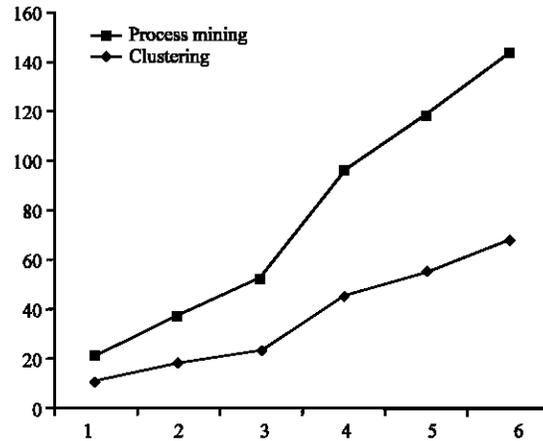


Fig. 10: The comparison of process mining and clustering

As shown in the Fig. 10, present results has a bout % 76 improvements in retaining churn customers of the Refah store while the result of clustering (Bong-Horng *et al.*, 2007) has about 68% improvements on it. In this Fig. 1 means 100 customers, 2 means 200 customers and so on. The compared model shows that by increasing customers, the ability of process mining in retaining customers is increased more than clustering.

CONCLUSIONS

This study has described a hybridized architecture to deal with to customer retention. Two modes had been worked on this architecture, namely, the training and testing. This is accomplished not only through predicting churn probability, but also by proposing retention policies. The training mode had been used data mining to predict the probability of churning. Historical information of customers used to calculate CLV and divide customers into distinct groups to construct churn model. Then this model used to find the best policy for retaining them. The testing mode tried to predict the churn probability of every new customer. Then it used process mining to keep them.

Directions for future research are given by the fact that nowadays there are a lot of attributes that are important on churning. Considering more attributes like behavior and taught of the customers would be efficient on retaining customers. Moreover, dividing customers into more distinct groups could be more efficient.

ACKNOWLEDGMENTS

N. Abedzadeh, M.A. Nematbakhsh and N. Nematbakhsh worked on customer retention. N. Abedzadeh used data mining for customer churn and process mining for retaining them. Refah Department Store helped them to test their method by giving dataset to them.

REFERENCES

- Aur lie, L. and C. Croux, 2006. Bagging and boosting classification trees to predict churn. *J. Market. Res.*, 43: 276-286.
- Bong-Horng, C., T. Ming-Shian and H. Cheng-Seen, 2007. Toward a hybrid data mining model for customer retention. *Knowledge-Based Syst.*, 20: 703-718.

- Chip, R.B., 1996. Customers as Partners. Berrett-Koehler, UK.
- Gerpott, T.J., W. Rams and A. Schindler, 2001. Customer retention, loyalty and satisfaction in the German mobile cellular telecommunications market. *Telecommun. Policy*, 25: 249-269.
- Greco, G., A. Guzzo and L. Pontieri, 2008. Mining taxonomies of process models. *Data Knowledge Eng.*, 67: 74-102.
- Gupta, S., D. Hanssens, B. Hardie, W. Kahn and V. Kumar, 2006. Modeling customer lifetime value. *J. Service Res.*, 9: 139-155.
- Hung, S.Y., D.C. Yen and H.Y. Wang, 2006. Applying data mining to telecom churn management. *Expert Syst. Appl.*, 31: 515-524.
- Jiafei, L., D. Liu and B. Yang, 2007. *Process Mining: Extending α -Algorithm to Mine Duplicate Tasks in Process Logs*. Springer, Berlin, Heidelberg.
- Kamber, M. and J. Han, 2000. *Data Mining: Concepts and Techniques 1st Edn.*, Morgan Kaufmann, San Mateo CA, ISBN-10: 1558604898.
- Nigel, H., B. Self and G. Roche, 2000. *Details of Book: Customer Satisfaction Measurement For Iso 9000: 2000*. Butterworth-Heinemann, UK., ISBN: 0750655135, pp: 176.
- Petersen, J.A., L. McAlister, D.J. Reibstein, R.S. Winer, V. Kumar and G. Atkinson, 2009. Choosing the right metrics to maximize profitability and shareholder value. *J. Retailing*, 85: 95-111.
- Reinartz, W. and V. Kumar, 2003. The impact of customer relationship characteristics on profitable lifetime duration. *J. Market.*, 67: 77-99.
- Quinlan, J.R., 1986. Induction of decision trees. *Machine Learn.*, 1: 81-106.
- Quinlan, R.J., 1993. *C4.5 Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Sunil, G., D.R. Lehmann and J.A. Stuart, 2004. Valuing customers. *Journal of Marketing Research*, pp: 7-18, February 2004, HBS Marketing Research Paper No. 03-08. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=459595.