

# **Customer Churn Prediction Using a New Criterion and Data Mining; A Case Study of Iranian Banking Industry**

**Davoud Gholamiangonabadi**

Department of Industrial Engineering and Management Systems  
Amirkabir University of Technology (Tehran Polytechnic)  
Tehran, Iran  
[dgholamian@gmail.com](mailto:dgholamian@gmail.com)

**Jamal Shahrabi**

Department of Industrial Engineering and Management Systems  
Amirkabir University of Technology (Tehran Polytechnic)  
Tehran, Iran  
[jamlshahrabi@gmail.com](mailto:jamlshahrabi@gmail.com)

**Seyed Mohamad Hosseinioun**

Industrial and Systems Engineering Department  
Northern Illinois University  
Chicago, USA  
[moh.hoss@gmail.com](mailto:moh.hoss@gmail.com)

**Sanaz Nakhodchi**

Department of Computer Engineering  
Iran University of Science and Technology  
Tehran, Iran  
[S.nakhodchi@gmail.com](mailto:S.nakhodchi@gmail.com)

**Soma Gholamveisy**

Department of Industrial Engineering  
Islamic Azad University (South Tehran Branch)  
Tehran, Iran  
[soma.ghv@gmail.com](mailto:soma.ghv@gmail.com)

## **Abstract**

In this paper, in order to measure the customer churn rate in the Iranian banks, a new approach is introduced. First, using data preparation, the data is normalized. Then, using a k-medoids method, data clustering is formed. To assess the clustering performance, Davies-Bouldin index is applied. To detect the patterns available in the data, different Neural Network (NN) methods were exploited namely, Radial Basis Function (RBFNN), Generalized Regression (GRNN), and Multilayer Perceptron (MLPNN) and Support Vector Machine (SVM). To increase the precision in the model performance measurement, hybrid of first and second error type cost functions is considered. The results suggest that the MLPNN and SVM models show higher precision and lower cost function.

## **Keywords**

Customer Churn, Bank Industry, K-medoids Clustering, Artificial Neural Networks, Cost Function

## **1. Introduction**

In the preceding decade, data have been stored in the data bases. As the volume of this stored data increased the need and desire for faster, less expensive and more acceptable methods of analysis began to grow. Obviously, if efficient mechanism to deal with this huge amount of data is not implemented, and if we are unable to extract the knowledge existing in the data, it would eventually lose value. Thank to those, aware of such possible scandal, processes using statistics, mathematics, artificial intelligence, machine learning, etc. to elicit the useful information and consequently, knowledge, from data-bases, was born, named as data mining. Data mining techniques can be exploited to detect properties of the data, in huge data bases.

Data mining technics can be classified into seven types as following, Association, Classification, Clustering, Predication, Regression, Sequence discovery and Visualization (Giraud-Carrier and Povel, 2003; Mitra et al., 2002; Shaw et al., 2001; Turban et al., 2014).

One of the most noticeable credits of data mining is the vast domain of its technics, applicable in various fields.

Nowadays, organizations are in continuous struggle to retrieve the information associated with their actual or potential customers to analyze their requirements and respond to them (Mutanen et al., 2006). This strategy in marketing is called Customer Relationship Management (CRM) (Peppard, 2000). CRM is consisted of four main aspects: Customer Identification, Customer Attraction, Customer Retention and Customer Development (Parvatiyar and Sheth, 2001; Swift, 2001). Customer retention is considered as the core of the customer relationship management. Customer satisfaction level relies highly on the product expectation that is bounded by their conception of the respective product (Kracklauer et al., 2004).

In the modern competitive markets, growing number of organizations are beginning to realize the essence of customer based strategy, that can aid the system's customer retention in the competitive market, constant level of profit, minimizing customer churn, maximizing the profit (Tsai and Lu, 2009). Customer churn has been defined variously in the literature, Customer churn means to lose one's customer and their choosing the competitor's services (Chu et al., 2007), for example the reluctance of the customer to continue doing business with the organization (Xie et al., 2009). Commonly, organizations use either reactive or active approaches to treat their churning customers (Burez and Van den Poel, 2007). Reactive organization, waits for the customer to seek its desired services and once it demands to leave the business with the company, it is given incentives to persuade the stay. On the opposite, active organizations make efforts to recognize the prospective leaving customers and plan to provide incentives before it becomes imminent. Obviously, if the predictions associated with the customer churn is not as precise as it should be, any preventing efforts would be in vain. In addition, the system has paid for the customers who were not going to leave anyway (Tsai and Lu, 2009).

There have been a lot of research undertaken into the importance of customer behavior prediction and customer churn; a quick summary is as follow:

- The cost dictated for acquiring new customers is on average, ten or five times more than keeping current customers. Consequently, it seems necessary to select the right approach for assuring the customer retention. For that, it is essential to predict the customer behavior well (Chu et al., 2007).
- Should the rate of customer retention increases 5%, it is expected that the organization profit grow 25% to 85% (Feinberg and Trotter, 2001).
- Old customers, buy more, take less time, show less sensitivity to price changes and take with them new customers to the system (Ganesh et al., 2000).
- Intimate relationship with customers in long term, is one of the prerequisites of prosper in any business (Burez and Van den Poel, 2007).

Especially, banks and financial agencies:

- A bank can earn up to 85% of all its profit with merely keeping 5% of its leaving customers (Reicbheld and Earl Sasser, 1990).
- The profit associated with a 1% increase in the retention of the customers have been found noticeable and it is believed to positively affect the interests and the profit of the organization (Van den Poel and Lariviere, 2004).
- Generally, old customers, stay longer with a bank, possess a longer life cycle and cause more value for the bank (Benoit and Van den Poel, 2009).

Banks and financial agencies have grown in number recently in Iran, causing the competition to be more difficult than before and as expected, increasing their desire to recognize their leaving customers, and trying to prevent their

prospective churn. In this research, our aim is to detect the prospective leaving customers in an anonymous Iranian bank.

In the second part a literature review is presented on the field. Next, the methodology is explained and finally, empirical data and the conclusion are organized.

## **2. Literature Review**

Applying the new technologies such as storing data, data mining and other related software can give the organization competitive advantage. Especially data mining can extract valuable information hidden in the stored data (Rygielski et al., 2002).

In some of the fields in business, especial attention has been paid to the churning customer's data by the organizations. In premier research attempts, few features used to be considered. Then, the dependence of customer churn and organization policies was sought. In the premier research classic statistical methods were used (Chu et al., 2007).

Numerous attempts have made using different algorithms, sequential algorithms (Chiang et al., 2003) for instance, genetic models (Eiben et al., 1998), decision tree (Lemmens and Croux, 2006) neural network (Mozer et al., 2000) and support vector machine (Zhao et al., 2005) are other examples of the approaches used to predict customer churn. (Nie et al., 2011) carried out between 2006 and 2007 in Chinese banks on the churning customers using credit cards, both logistic regression and decision tree was used for the purpose of recognition and prediction. To validate the models, ten-folds validation model was applied. Results of this model show slightly better performance in logistic regression comparing the peer decision tree alternative.

(Xie et al., 2009) a new method is been deployed to predict the customer churn. This method, named as modify randomly forests. The principle is to keep the good properties in this correction process by learning them iteratively. In addition, for the purpose of more precise prediction, the data was balanced and sampling and cost-sensitive learning was exploited. The methods were examined on 2000 cases of customers and the result supported the performance of the presented method- the modify randomly forests- in comparison with neural network, decision tree and support vector machine.

(Meer, 2006) In this research conducted between 2003 and 2004, in a Dutch financial agency, customer churn was studied using click stream analysis the respective website- that helps monitoring the online customers. The results show strong connection between online activities and customer behavior.

(Benoit and Van den Poel, 2012) In this research, undertaken in 2006 in a financial agency in Europe, the advantages of considering the social networks variables in customer retention was studied. Added to the common variables in the literature, variables of kinship network information were also studied to help the precision of the model. Due to the large volume of data, egocentric network was deployed and for the purpose of prediction modify randomly forests algorithm was applied. The results show the customer churn model is indeed improved by using the social network variables and possess more precision.

## **3. Methodology**

In this research, using neural networks, a new approach has been presented to study and predict the customer churn. As is views in the Fig 1, firstly pre-process is performed; next, clustering and finally, using neural networks and considering cost function, a solution is found the problem of customer churn.

### **3.1. Data Transfer**

To discover the hidden knowledge, there are preparations to be made and these steps are called data preparation. The importance of data preparations lies upon the fact that the lack of quality data leads to the lack of quality in the data mining process. The main role of data pre-analysis is to organize data in standard shapes for data mining or other computer based operations. The principle steps in data preparation are: data clearance, data integration, data transformation, data reduction and dimensional reduction. The data pre-analysis methods applied in this research the is presented in the following:

#### **3.1.1. Min-Max method**

This method applies a linear transformation on a set of Continuous data. The goal of this, is to increase precision in the next phases. Assume that  $x_{min}$  and  $x_{max}$  are the minimum and the maximum of an attribute  $j$  respectively. Also,  $x'_{min}$  and  $x'_{max}$  are the new minimum and maximum for this attribute; then this transformation is conducted using the equation (1).

$$x'_{ij} = \frac{x_{ij} - x_{min,j}}{x_{max,j} - x_{min,j}} (x'_{max,j} - x'_{min,j}) + x'_{min,j} \quad (1)$$

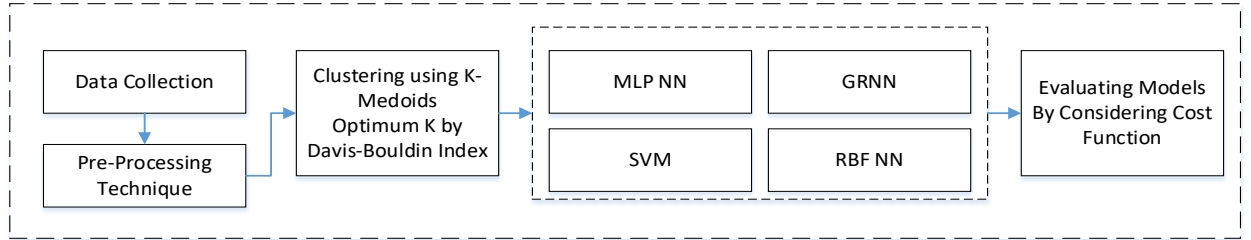


Fig 1. Proposed Methodology

### 3.2. K-medoids

The K-medoids algorithm is well-known for its partitioning around medoids, is one of the expanded algorithms form K-mean. This method was proposed in 1987 (Kaufman and Rousseeuw, 1987). The purpose of it is reducing the sensitivity of the values generated using the mean of the cases, it uses the approximation of medoids as its suggested outcome.

•  $X_k$  is assigned to the cluster associated with the medoid  $u_h$  and we have equation (2).

$$dist(x_i, x_k) \leq \min_{u_e \in U, e \neq h} dist(u_e, x_k) \quad (2)$$

In such condition, the observation  $x_k$  is chosen as the new medoid of  $x_i$  to represent the cluster  $C_h$ . The contribution of the substitution may be positive, negative or zero and is calculated using the equation (3).

$$R_{ihk} = dist(x_i, x_k) - dist(u_h, x_k) \quad (3)$$

•  $X_k$  is currently assigned to the cluster associated with the medoid  $u_h$  and we have equation (4).

$$dist(x_i, x_k) \leq \min_{u_e \in U, e \neq h} dist(u_e, x_k) \quad (4)$$

Now, the observation  $x_i$  is assigned to another cluster and the contribution of this substitution is equation (5).

$$R_{ihk} = \min_{u_e \in U, e \neq h} dist(u_e, x_k) - dist(u_h, x_k) \quad (5)$$

•  $X_i$  is not yet assigned to the cluster associated with the medoid  $u_h$  and we have equation (6).

$$dist(x_i, x_k) \geq \min_{u_e \in U, e \neq h} dist(u_e, x_k) \quad (6)$$

•  $X_k$  is not yet assigned to the cluster associated with the medoid  $u_h$  and we have equation (7).

$$dist(x_i, x_k) \leq \min_{u_e \in U, e \neq h} dist(u_e, x_k) \quad (7)$$

In such condition, the observation  $x_k$  is to be assigned to the cluster  $C_h$  and the substitution contribution is equation (8).

$$R_{ihk} = dist(x_i, x_k) - \min_{u_e \in U, e \neq h} dist(u_e, x_k) \quad (8)$$

$$T_{ih} = \sum_{x_k \in U} R_{ihk} \quad (9)$$

### 3.3. Davies-Bouldin Index

This criterion was presented in 1979 to assess clustering algorithms (Davies and Bouldin, 1979). If  $X_j$  is assigned to the cluster  $C_j$  and the respective cluster center is shown by  $A_i$  then,  $S_i$  calculates the scatter in the cluster using the equation (10).

$$S_i = \left( \frac{1}{T_i} \sum_{j=1}^{T_i} |X_j - A_i|^q \right)^{\frac{1}{q}} \quad (10)$$

Now, if the separation between the clusters  $i$  and  $j$  ( $C_i$  and  $C_j$ ) is called  $M$ , we have:

$$M_{i,j} = ||A_i - A_j||_p \quad (11)$$

If  $M_{i,j}$  is defined as how good the clustering scheme, then we have:

$$R_{i,j} = \frac{S_i + S_j}{M_{i,j}} \quad (12)$$

$$D_i = \max_{j; i \neq j} R_{i,j} \quad (13)$$

And if  $N$  is the number of the clusters, we have: The Davies-Bouldin criterion is the average value for all the clusters; in other words it equals:

$$DB = \frac{1}{N} \sum_{i=1}^N D_i \quad (14)$$

### 3.4. Neural Networks

Artificial neural network is an adaption from the biological neural system, trying to mimic the process the data like human brain. The key element of this method is the new structure of data processing system. This system is consisted of large number of highly interconnected processing elements which act consistently for solving a problem. Alike humans brain, artificial neural networks, can be trained with examples. An artificial neural network is adjusted for doing a special task like detecting rules and categories in a period of learning process. In biological systems, learning is associated with adjusting the connections of synapsis which is located between nerves. Artificial neural networks exploit the same method, with their significant capacity to deduce meaning of complicated and vague data can extract the rules and methods which is difficult and elaborate for humans and other computer techniques to discover. It is used in different fields such as Face Recognition (Azami et al., 2013), Diabetes Diagnosis (Fiuzy et al., 2013), Bankruptcy Prediction (Bagheri et al., 2012), prediction changes in stock (Gholamiangonabadi et al., 2014), and prediction quality in devices (Gholamiangonabadi et al., 2015).

#### 3.4.1. General Regression Neural Network

GRNN is one of the neural network type presented in 1991 (Specht, 1991). This is one of the radial basis neural networks. The advantage of this method is it can be used to train the network when we lack data. In addition, to train a network using this method, an iterative training procedure is followed instead of back propagation neural network. This network is able to approximate any given function between the input and the output. As can be seen in the Fig 2 a GRNN is consisted of four layers: the input layer, pattern layer, summation layer and the output layer. Assuming there are  $q$  neurons as the input layer- that equal the number of the input parameters, the output for this layer is considered as the input for the pattern layer where  $p$  neurons are designed. The output of the pattern layer is entered to the summation layer where two neurons named Denominator and Numerator are considered. Clearly, each neuron in the pattern layer is connected to the two abovementioned neurons (S, D). The neuron S calculates the summation of weighted response associated with the pattern layer and the neuron D does the same for the un-weighted outputs. The output layer and the summation layer, together normalize the output set. To train such network, Radial Basis Function or Linear Basis Function can be used.

GRNN is widely used in detecting cancer, diabetes and heart disease, also in fraud detection. A short review of the calculations done in GRNN is presented below:

$$Y(x) = \frac{\sum_{i=1}^n y_i \cdot \exp(-D(x, x_i))}{\sum_{i=1}^n \exp(-D(x, x_i))} \quad (15)$$

In which we have:

$$D(x, x_i) = \sum_{k=1}^m \left( \frac{x_i - x_{ik}}{\sigma} \right)^2 \quad (16)$$

Where  $y_i$  is the weight connection between the  $i$ th neuron in the pattern layer and the S-summation neuron,  $n$  is the number of training patterns,  $D$  is the Gaussian function,  $m$  is the number of elements of the input vector,  $x_k$  and  $x_{ik}$  are the  $j$ -th element of  $x$  and  $x_i$ , respectively,  $r$  is the spread parameter, whose optimal value is determined experimentally.

#### 3.4.2. Radial Basis Function

Radial Basis Neural networks were presented by different researchers. The Fig 2 shows the structure of this network. The input neurons bear no weight, thus the first hidden layer receives the exact same values as the first layer. The

function designed in the hidden layer are the Radial Basis type. The transfer function for the neurons of the hidden layer are non-monotonic. Then the output of these neurons are sent to the output layer by weights. The neurons of the output layers are actually, simple summations. Let us assume that there are H neurons in the hidden layer. The transfer function are mostly like Gaussian Density Functions. If this function is Gaussian:

$$a_{h,k} = \exp\left(-\frac{\|\hat{x}_h - x_k\|^2}{\sigma_h^2}\right) \quad (17)$$

In which  $a_{h,k}$  is the output of the hth neuron in the hidden layer. Also,  $\hat{x}_h$  is the center of the radial function and is the distance scaling parameter which determines over what distance in the input space the unit will have a significant influence. Finally, the weighted average of the outputs associated with the hidden layer determines the output. In other words, the equation (18) shows this output value.

$$y_i = \sum_{i=1}^n w_i \times a_{h,i} \quad (18)$$

In which the  $w_i$ ,  $k$  is the weight assigned to the neuron  $i$  –th in the hidden layer and the  $k$  –th neuron in the output layer. As this method is an observer learning method, the exact values for  $x_i$  and  $y_i$  are predetermined. Thus to have the weights in the second layer, in this research, the pseudo-inverse method is used, in which:

$$G = \{g_{i,j}\} \quad (19)$$

In which:

$$g_{i,j} = \exp\left(\frac{-\|x_i - v_j\|}{2\sigma_j^2}\right) \quad i = 1,2, \dots, n; j = 1,2, \dots, p \quad (20)$$

And we have:

$$D = GW \quad (21)$$

Where  $D$  is the desired output for the trained data. If  $G^{-1}$  exists, then we have:

$$W = G^{-1}D \quad (22)$$

If  $G$  is ill-conditioned (close to singularity) or is a non-square matrix, then:

$$W = G^+D \quad (23)$$

Where:

$$G^+ = (G^T G)^{-1} \times G^T \quad (24)$$

### 3.4.3. Support Vector Machine

Support Vector Machine selects a number of observations as the representative of a certain class. These observations, determine the separation process in the classification of the feature space. If the space is linear, there are infinite numbers of lines and planes that can separate different classes. The optimized separation line, is a line that bear the best level of expansion, and the amount of error. The separation margin, here is twice the size of the distance between the trained data and separating hyper-plane. In addition, they are the support vectors that have the least distance from the separating hyper-plane. To determine these hyper-planes the pattern below is followed:

- If  $w$  is the coefficient vector associated with the hyper-plane and  $b$  is the bias, then the separating hyper-plane is:

$$w'x = b \quad (25)$$

- The two supporting focal hyper-planes are:

$$w'x - b - 1 = 0, \quad w'x - b + 1 = 0 \quad (26)$$

In which the separation margin is:

$$\delta = \frac{2}{\|w\|}, \quad \|w\| = \sqrt{\sum_{j \in N} w_j^2} \quad (27)$$

- To determine  $w$  and  $b$ , a quadratic optimization problem with linear constraints is to be solved:

$$\max_{w,b} \frac{1}{2} \|w\|^2 \text{ s.t. } y_i(w'x_i - b) \geq 1, \quad i \in M \quad (28)$$

- The objective function seeks to maximize the separation margin by minimizing the inverse, and the constraints have each  $x_i$  stay at the associated class  $y_i$ .

- Thus the objective function and the constrains are:

$$\min_{w,b,d} \frac{1}{2} \|w\|^2 + \lambda \sum_{i=1}^m d_i \text{ s.t. } y_i(w'x_i - b) \geq 1 - d_i, \quad i \in M, d_i \geq 0, i \in M \quad (29)$$

- The abovementioned optimization problem can be solved using Lagrangian duality:

$$L(w, b, d, \alpha, \mu) = \frac{1}{2} \|w\|^2 + \lambda \sum_{i=1}^m d_i - \sum_{i=1}^m \alpha_i [y_i(w'x_i - b) - 1 + d_i] - \sum_{i=1}^m \mu_i d_i y_i (w'x_i - b) \geq 1 - d_i, \quad i \in M d_i \geq 0, i \in M \alpha_i, \mu_i \geq 0 \quad (30)$$

To find the optimized solution, the partial deviations to b, d and w must equal zero. By the placement of the calculated values in the dual objective function and by applying the Kuhn-Tucker's conditions on the problem and the dual, we have:

$$\alpha_i [y_i(w' - b) - 1 + d_i] = 0, \quad i \in M \mu_i (\alpha_i - \lambda) = 0, \quad i \in M \quad (31)$$

Where (31) is to detect the support vectors.

In this case, each new observation of x is classified as equation (32):

$$f(x) = \text{sgn}(\sum_{i=1}^m \alpha_i y_i x'_i x_i + b) \quad (32)$$

If the data are not linearly separable, the features can be transferred to a new space to make them separable by a line. In other words, if the function is the transfer function for the data from the non-linear space to a linear one, then in the Lagrangian duality would be:

$$L(w, b, d, \alpha, \mu) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{h=1}^m y_i y_h \alpha_i \alpha_h K(x_i, x_h) s. t. \sum_i \alpha_i y_i = 0 \leq \alpha_i \leq C \quad (33)$$

Where  $K(x_i, x_h) = \phi(x_i)^T \phi(x_h)$ .

Thus, to classify each new observation:

$$f(x) = \text{sgn}(\sum_{i=1}^m \alpha_i y_i K(x_i, x + b)) b = \frac{1}{|S|} \sum_i [y_i - \sum_h \alpha_j y_h K(x_i, x_h)] \quad (34)$$

### 3.4.4. Multilayer Perception

A multilayer perceptron (MLP) is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron with a activation function. MLP utilizes a supervised learning technique called back propagation for training the network. MLP is a modification of the standard linear perceptron and can distinguish data that are not linearly separable. To train the network, Levenberg-Marquardt method is used that will be explained in the following:

The training algorithm includes five steps.

- step 1: Initialize weights and thresholds to small random values.
- step 2: Choose an input-output pattern  $(x^{(k)}, t^{(k)})$  from the training data.
- step 3: Compute the network's actual output  $o^{(k)} = f(\sum_{i=1}^l w_i x_i^{(k)} - \theta)$ . ( $l$  is size of input vector or the size on input neurons). Adjust the weights and bias according to the Levenberg-Marquardt algorithm.
- step 4: If whole epoch is complete, then pass to the following step; otherwise go to step 2.
- step 5: If the weights (and bias) reached steady state ( $\Delta w_i \approx 0$ ) through the whole epoch, then stop the learning; otherwise go to through one more epoch starting from.

### 3.5. Cost Function

Based on (Nie et al., 2011) it is assumed that precision does not suffice to solve this problem and an economic cost function is required to analyze the model as well. Two types of error can be considered. The first type occurs when to assume that a customer is leaving while customer churn is not to happen, and the second type is when a customer is going to leave, but the prediction is otherwise. Obviously, the second type causes higher amount of cost. So to improve the model, the error is needed to be transferred, for that the economic cost function in this research follows the equation (35).

$$Cost = B e_s P_{ave} + \frac{M}{G e_f + B(1 - e_s)} G e_f \quad (35)$$

In which  $B$  is the number of cases associated with customer churn,  $G$  is the number of loyal customers,  $M$  is the budget available for marketing,  $P_{ave}$  is the average of earned profit from each customer,  $e_f$  and  $e_s$  represent the first and second type of error respectively.

In the first part of the cost function,  $B e_s P_{ave}$ , is the loss caused by the second type of error, representing the inability to detect a leaving customer and losing the chance to incentivize them to stay.

The second part of the cost function,  $\frac{M}{G e_f + B(1 - e_s)} G e_f$  is the loss caused by the first type of error.  $\frac{M}{G e_f + B(1 - e_s)}$  is the average cost of marketing for each customer.  $G e_f$  is the mis-identified loyal customers.

With the aid of the statistics and the experts available in the bank, the amount of marketing budget and the average of profit per customer were made available. Hence, the model chosen minimizes the cost function based on the available data.

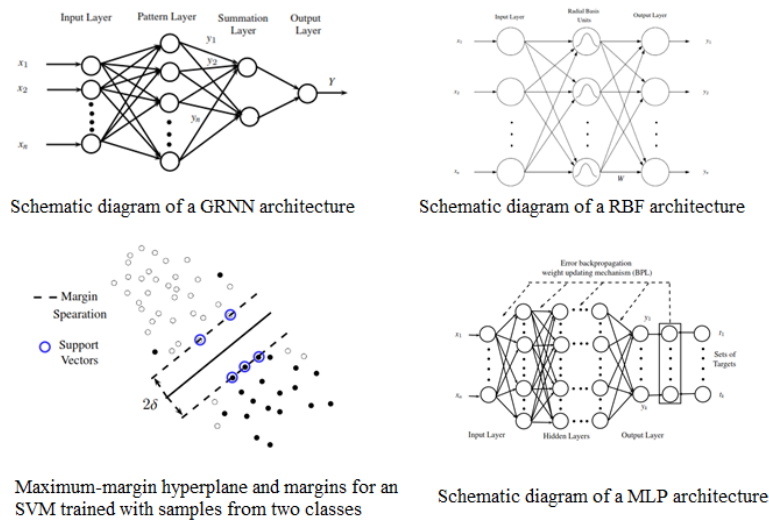


Fig 2. Schematic diagram of NNs and SVM architecture

## 4. Empirical Analysis

### 4.1. Data

Data selection has been made base on the literature review and the accessibility. To measure the customer churn, the data associated with 860 customers is gathered from Iranian banks. The input variables in this model are age, gender, residence, income, marital status, number of children, ownership of a car, ownership of a saving account, ownership of a current account and the mortgage status. The output variable is defined as retention of credit card. The data used in this research is gathered from an anonymous Iranian bank between February 2013 and June 2013. A short review of the data is shown in the Table 1.

Table 1. Descriptive Data

Variables	Min	Max	Mean	standard deviation
Input				
Age	18	67	42.395	14.424
Gender	0	1	0.5	0.500
Region	0	3	1.37	1.008
Income*	830	63130	3215	1372
Married	0	1	0.66	0.474
Children	0	3	1.01	1.05
Car	0	1	0.49	0.5
Save-account	0	1	0.69	0.462
Current-account	0	1	0.758	0.482
Mortgage	0	1	0.348	0.476
Output				
usage	0	1	0.46	0.49

\*: 1000 Rials



As can be seen in the Table 1, most of the observation is consisted of the young people. In addition, the income is not considered high for most of the cases. Most of these customers own current and saving account and nearly one third were in debt to the bank in form of mortgages.

## 4.2. Results

According to the methodology in the second part, as it did not consist any missing data, and as the data were balanced- 45 percent were loyal, and 54 percent churning- first the data is normalized. After that part, the clustering has been done by k-medoids method. As can be seen in the Fig 3, one of the best clustering conducted was with the number of cluster 9. Thus this is chosen as the best number of clusters for the given set of data. Moreover, it's used to define the center and the width of the RBF neural network. Once the optimized clustering is done, using GRNN, SVM, MLPNN and RBF prediction models of the customer churn has been built. To monitor the performance of the built models, sensitivity and specificity criteria were calculated using the equations (39) and (40), respectively. To analyze the precision of the classification the equation (41) was used:

$$Sensitivity = \frac{TP}{TP+FN} (\%) \quad (39)$$

$$Specificity = \frac{TN}{TN+FP} (\%) \quad (40)$$

$$Accuracy = \frac{\sum_{k=1}^{|C|} assess(c_k)}{|C|} \quad c_k \in C \quad (41)$$

In which:

True positive (TP): Correctly identified

False Positive (FP): Incorrectly identified

True Negative (TN): Correctly rejected

False Negative (FN): Incorrectly rejected

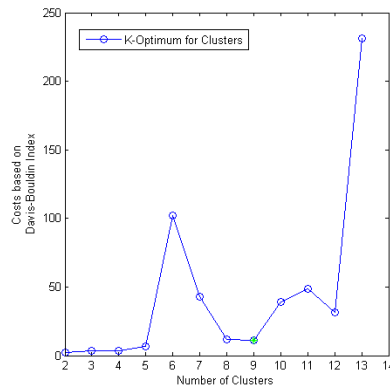


Fig 3. Determine best K by Consider Davies-Bouldin Index

## 4.3. Validation

To validate the model, 10-fold cross validation method is applied. In this method, firstly, the data is divided into ten equal parts, then the network is trained and tested for ten times. For example, for the first training of the network, the first nine part of the data is considered as the training and the last part as the test. For the second run, the first eight part and the 10th part of the data is considered as the training set and the ninth part for the test and so on.

As can be seen in the Table 2. associated with the GRNN, the network provide 74.2 percent precision. Using this method, the prediction of customer churn, compared to the prediction of the loyal customers show insignificant higher level of precision. Considering the Table 2 associated with the RBF Neural Network, show 83 percent precision. Again, the customer churn prediction is more precise compared to loyal customer prediction. The SVM results is reflected in Table 2 showing 85 percent precision in the predictions. Again the value associated with the prediction of customer churn is higher than the loyal customers. As can be observed in the Table 2 associated with the MLP Network, the average of the prediction related to customer churn and loyal customers. And again the precision is higher is case of customer churn prediction.

Table 2. Confusion Matrix For All Methods

Method	Observed	Predicted				
		Good/Bad		Sum	Percentage Correct (%)	Error (%)
		Good	Bad			
GRNN	Good	293	103	396	73.98	26.02
	Bad	118	346	464	74.57	25.43
	Sum	411	449	860	74.3	25.7
RBFNN	Good	317	79	396	80	20
	Bad	65	399	464	85.99	14.01
	Sum	382	478	860	83.25	16.74
SVM	Good	325	71	396	82.07	17.93
	Bad	53	411	464	88.57	11.43
	Sum	378	482	860	85.58	14.4
MLPNN	Good	337	59	396	85.1	14.9
	Bad	40	424	464	91.38	8.62
	Sum	377	483	860	88.48	11.5

Based on the Table 3 it can be observed that the MLP and SVM Networks have shown better performance respectively than RBF and GRNN. Considering the cost function, the MLP cost function has lower value than the other models. Also, GRNN gained lower cost function compared to the RBF.

Table 3. Investigating Different Criteria for Different Methods

Methods	Criteria Performance			
	Specificity	Sensitivity	Accuracy	Cost Function*
GRNN	74.56	73.98	74.3	742
RBFNN	85.99	80.05	83.25	1345
SVM	88.57	82.07	<b>85.58</b>	<b>605</b>
MLPNN	91.38	85.1	<b>88.48</b>	<b>456</b>

\*: million Rilas

## 5. Conclusion

Nowadays, Customer Relationship Management is one of the most important managerial concepts. The core of CRM is customer churn, and customer retention. Using data mining it is possible to recognize the hidden patterns of the data. The customer churn, also prediction the leaving customers pose huge amount of cost on the organization. In this research a new approach is presented to analyze the customer churn in Iranian bank. To form a more precise model, a cost function was suggested. And the results indicate the well performance of MLP compared to SVM, RBF and GRNN in making the customer churn prediction.

## References

- Azami, Hamed, Milad Malekzadeh, and Saeid Sanei. "A new neural network approach for face recognition based on conjugate gradient algorithms and principal component analysis." *Journal of mathematics and computer Science* 6, no. 3 (2013): 166-175.
- Bagheri, Mahnaz, M. Valipour, and V. Amin. "The Bankruptcy Prediction in Tehran share holding using Neural Network and it's Comparison with Logistic Regression." *The Journal of mathematics and computer science* 5, no. 3 (2012): 219-228.
- Benoit, Dries F., and Dirk Van den Poel. "Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services." *Expert Systems with Applications* 36, no. 7 (2009): 10475-10484.
- Benoit, Dries F., and Dirk Van den Poel. "Improving customer retention in financial services using kinship network information." *Expert Systems with Applications* 39, no. 13 (2012): 11435-11442.

- Burez, Jonathan, and 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* 32, no. 2 (2007): 277-288.
- Chiang, Ding-An, Yi-Fan Wang, Shao-Lun Lee, and Cheng-Jung Lin. "Goal-oriented sequential pattern for network banking churn analysis." *Expert Systems with Applications* 25, no. 3 (2003): 293-302.
- Chu, Bong-Horng, Ming-Shian Tsai, and Cheng-Seen Ho. "Toward a hybrid data mining model for customer retention." *Knowledge-Based Systems* 20, no. 8 (2007): 703-718.
- Davies, David L., and Donald W. Bouldin. "A cluster separation measure." *IEEE transactions on pattern analysis and machine intelligence* 2 (1979): 224-227.
- Eiben, A. E., A. E. Koudijs, and Frank Slisser. "Genetic modelling of customer retention." In *European Conference on Genetic Programming*, pp. 178-186. Springer, Berlin, Heidelberg, 1998.
- Feinberg, Richard, and Mike Trotter. "Immaculate deception: the unintended negative effects of the CRM revolution: maybe we would be better off without customer relations management." *Defying the limits* 2 (2001).
- Fiuzu, Javad Haddadnia Hadi Varharam Mohammad, Azam Qarehkani, J. Haddadnia, J. Vahidi, and H. Varharam. "Introduction of a method to diabetes diagnosis according to optimum rules in fuzzy systems based on combination of data mining algorithm (dt), evolutionary algorithms (aco) and artificial neural networks (nn)." *The Journal of Mathematics and Computer Science (JMCS)* 6, no. 4 (2013): 272-285.
- Ganesh, Jaishankar, Mark J. Arnold, and Kristy E. Reynolds. "Understanding the customer base of service providers: an examination of the differences between switchers and stayers." *Journal of marketing* 64, no. 3 (2000): 65-87.
- Gholamiangonabadi, Davoud, Seyed Danial Mohseni Taheri, Afshin Mohammadi, and Mohammad Bagher Menhaj. "Investigating the performance of technical indicators in electrical industry in Tehran's Stock Exchange using hybrid methods of SRA, PCA and Neural Networks." In *Thermal Power Plants (CTPP), 2014 5th Conference on*, pp. 75-82. IEEE, 2014.
- GholamianGonabadi, Davoud, Seyed Mohamad Hosseinioun, Jamal Shahrabi, and Mohammad AliMoradi. "Investigating Performance and Quality in Electronic Industry via Data Mining Techniques." *International Journal of Data Mining Techniques and Applications*, Volume: 04 Issue: 02 December 2015, Page No. 20-24.
- Giraud-Carrier, C., and Olivier Povel. "Characterising data mining software." *Intelligent Data Analysis* 7, no. 3 (2003): 181-192.
- Kaufman, Leonard, and Peter Rousseeuw. *Clustering by means of medoids*. North-Holland, 1987.
- Kracklauer, Alexander H., D. Quinn Mills, and Dirk Seifert. "Customer management as the origin of collaborative customer relationship management." In *Collaborative Customer Relationship Management*, pp. 3-6. Springer, Berlin, Heidelberg, 2004.
- Lemmens, Aurélie, and Christophe Croux. "Bagging and boosting classification trees to predict churn." *Journal of Marketing Research* 43, no. 2 (2006): 276-286.
- Meer, Geoffrey Van. "Customer development and retention on a Web-banking site." *Journal of Interactive Marketing* 20, no. 1 (2006): 58-64.
- Mitra, Sushmita, Sankar K. Pal, and Pabitra Mitra. "Data mining in soft computing framework: a survey." *IEEE transactions on neural networks* 13, no. 1 (2002): 3-14.
- Mozer, Michael C., Richard Wolniewicz, David B. Grimes, Eric Johnson, and Howard Kaushansky. "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry." *IEEE Transactions on neural networks* 11, no. 3 (2000): 690-696.
- Mutanen, Teemu, Jussi Ahola, and Sami Nousiainen. "Customer churn prediction-a case study in retail banking." In *Proc. of ECML/PKDD Workshop on Practical Data Mining*, pp. 13-19. 2006.
- Nie, Guangli, Wei Rowe, Lingling Zhang, Yingjie Tian, and Yong Shi. "Credit card churn forecasting by logistic regression and decision tree." *Expert Systems with Applications* 38, no. 12 (2011): 15273-15285.
- Parvatiyar, Atul, and Jagdish N. Sheth. "Customer relationship management: Emerging practice, process, and discipline." *Journal of Economic & Social Research* 3, no. 2 (2001).
- Peppard, Joe. "Customer relationship management (CRM) in financial services." *European Management Journal* 18, no. 3 (2000): 312-327.
- Reichheld, Frederick P., and W. Earl Sasser. "Zero Defections: Quotify Comes to Services." *Harvard business review* (1990).

- Rygielski, Chris, Jyun-Cheng Wang, and David C. Yen. "Data mining techniques for customer relationship management." *Technology in society* 24, no. 4 (2002): 483-502.
- Shaw, Michael J., Chandrasekar Subramaniam, Gek Woo Tan, and Michael E. Welge. "Knowledge management and data mining for marketing." *Decision support systems* 31, no. 1 (2001): 127-137.
- Specht, Donald F. "A general regression neural network." *IEEE transactions on neural networks* 2, no. 6 (1991): 568-576.
- Swift, Ronald S. *Accelerating customer relationships: Using CRM and relationship technologies*. Prentice Hall Professional, 2001.
- Tsai, Chih-Fong, and Yu-Hsin Lu. "Customer churn prediction by hybrid neural networks." *Expert Systems with Applications* 36, no. 10 (2009): 12547-12553.
- Turban, Efraim, Ramesh Sharda, Dursun Delen, and Turban Efraim. *Decision support and business intelligence systems*. Vol. 9. Pearson, 2014.
- Van den Poel, Dirk, and Bart Lariviere. "Customer attrition analysis for financial services using proportional hazard models." *European journal of operational research* 157, no. 1 (2004): 196-217.
- Xie, Yaya, Xiu Li, E. W. T. Ngai, and Weiyun Ying. "Customer churn prediction using improved balanced random forests." *Expert Systems with Applications* 36, no. 3 (2009): 5445-5449.
- Zhao, Yu, Bing Li, Xiu Li, Wenhua Liu, and Shouju Ren. "Customer churn prediction using improved one-class support vector machine." In *International Conference on Advanced Data Mining and Applications*, pp. 300-306. Springer, Berlin, Heidelberg, 2005.

**Davoud Gholamiangonabadi** received his Master degree in Industrial Engineering from Amirkabir University of Technology (aka Tehran Polytechnique). He has broad interdisciplinary research interests covering topics in Big Data, Data Mining, Machine Learning, Meta-Heuristic Algorithms, Image Processing, and Artificial Neural Networks. He is considered an active member of the data mining research community and has published more than ten articles and conference posters.

**Dr. Jamal Shahrabi** received his PhD from Dalhousie University. He is a faculty member of Industrial Engineering and Management Systems faculty at Amirkabir University of Technology. His research interests lie at the intersection of data science and Management and Marketing. He has been succeed to make a synergy of university and industry to solve the industry management problems and provide the opportunity of involvement of students in real industry issues. He has managed 9 big size industry projects so far. He has published two book chapters, 13 books in Persian, 136 conference papers & 34 outstanding ISI journal papers with 898 Citations So far.

**Mohamad Hosseinioun** is pursuing a Ph.D. in Management Information Systems at the Liautaud Graduate School of Business at the University of Illinois at Chicago. He has received his bachelor's and master's degrees in Industrial and Systems Engineering and his research spans Causal Inference and Prediction in Complex and Dynamic Systems. Using Statistical and Fuzzy Machine Learning methods, Network Analysis, Game Theory, and Econometrics he investigates the behavior of social and economic entities. His primary research focuses on the impact of IT on the labor markets. More broadly, he is interested in the collective realities and collective decision making, where the outcome of the "whole" is significantly altered by the interconnected behavior of the "individual"s.

**Sanaz Nakhodchi** received her B.Sc degree in Information technology from Islamic Azad University, Mashhad, Iran, 2010. She also has 4 years experience (2014-2018) at CERT (computer emergency response team) Lab in Ferdowsi University of Mashhad as a research assistant. She has joined the University of Guelph as a master student in computer science and a member of CyberScience Lab in 2018. Her research interests are Cloud Computing, Artificial Intelligence, IOT, Security and threat hunting. Her e-mail address is: nakhodcs@uoguelph.ca

**Soma Gholamveisy** received her BSc in Applied Mathematics from Kurdistan University, in 2010. She got her MSc in industrial engineering from Islamic Azad University (South Tehran Branch) in 2015. She works as an independent researcher in a wide variety of fields. Besides, she worked in 3 industrial companies as a control project engineer. Her research interests include data mining, healthcare system, customer relationship management, statistical method, productivity, business model, and machine learning.