

# Hybrid Data Mining Models for Predicting Customer Churn

Amjad Hudaib, Reham Dannoun, Osama Harfoushi, Ruba Obiedat, Hossam Faris

King Abdullah II School of Information Technology, The University of Jordan, Amman, Jordan  
Email: [ahudaib@ju.edu.jo](mailto:ahudaib@ju.edu.jo), [r.dannoun@ju.edu.jo](mailto:r.dannoun@ju.edu.jo), [o.harfoushi@ju.edu.jo](mailto:o.harfoushi@ju.edu.jo), [r.obiedat@ju.edu.jo](mailto:r.obiedat@ju.edu.jo),  
[hossam.faris@ju.edu.jo](mailto:hossam.faris@ju.edu.jo)

Received 3 February 2015; accepted 7 April 2015; published 9 April 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The term “customer churn” is used in the industry of information and communication technology (ICT) to indicate those customers who are about to leave for a new competitor, or end their subscription. Predicting this behavior is very important for real life market and competition, and it is essential to manage it. In this paper, three hybrid models are investigated to develop an accurate and efficient churn prediction model. The three models are based on two phases; the clustering phase and the prediction phase. In the first phase, customer data is filtered. The second phase predicts the customer behavior. The first model investigates the *k*-means algorithm for data filtering, and Multilayer Perceptron Artificial Neural Networks (MLP-ANN) for prediction. The second model uses hierarchical clustering with MLP-ANN. The third one uses self organizing maps (SOM) with MLP-ANN. The three models are developed based on real data then the accuracy and churn rate values are calculated and compared. The comparison with the other models shows that the three hybrid models outperformed single common models.

## Keywords

Data Mining, *k*-Means, Hierarchical Cluster, Self Organizing Maps, Multilayer Perceptron Artificial Neural Networks, Churn Prediction

---

## 1. Introduction

Customers are the most important asset in any organization since they are considered as the main profit source. Although attracting new customers is important for a company’s growth, all businesses have agreed that retaining existing ones is more important because of many reasons: First, it is easier to sell to an existing customer than to attract and sell a new one. Second, building customer loyalty is a target for all industries. The loyal customers are

less sensitive to price changes. Statistics show that it costs five to six times more to gain a new customer than to keep the existing ones [1].

In the industry of information and communication technology (ICT), the customer life cycle contains five main stages: acquisition, build up, peak, decline and churn [2]. The term customer churn is used in telecommunication industry to define customers who change their supplier or provider to a new one offering same service [3] [4]. Therefore, a churn management process is necessary for any successful company who wants to compete in the marketplace and to keep profit at higher levels. The goal of the churn management process is to retain those about to churn before they do so and this is done by churn prediction [5]. Churn prediction helps assess the current companies' situation and setting future plans for specific, focused group or setting targeted marketing campaigns [6]. In fact, churn prediction is an important element in making an accurate and effective decision [7].

In literature, many researchers proposed different machine learning approaches for predicting customer churn in telecommunication business. Most of these approaches are based on applying single classifier algorithm. Example of these works can be found in [1] [8]-[12]. However, churn prediction is a challenging problem due to the imbalance distribution of its classes where usually number of churning customers is much less than the non-churning ones. This problem makes most of the conventional single machine learning approaches inappropriate for classifying the data [13]. For this reason, other researchers proposed hybrid machine learning approaches by combining two or more algorithms for predicting customer churn where one of the algorithms is used for data preprocessing before performing the classification task [13]-[15]. Some authors proposed clustering the data into a number of clusters then eliminating some small clusters as a way to filter the data from unrepresentative data [15]-[17].

In this paper, we propose three hybrid data mining models for predicting customer churn in a telecommunication company. All three models are composed of a clustering algorithm and a classification one. First, the clustering algorithm is applied to filter the data from outliers and unrepresentative behaviors. The resulted filtered data of the clustering algorithm become the input to the classification algorithm. Then the classification algorithm uses these data in the learning process and builds the final churn prediction model. For clustering, we use  $k$ -means clustering, hierarchical clustering and self organizing maps (SOM) while for classification we use the Multilayer Perceptron Artificial Neural Networks (MLP-ANN).

The rest of this paper is organized as follows: Section 2 covers the preliminaries of the algorithms applied in the models proposed in this work for churn prediction. In Section 3, the proposed models are presented. A description of the dataset used is given in Section 4. Evaluation criteria used to evaluate the proposed models are listed in Section 5. Finally, the experiments and results are discussed in Section 6.

## 2. Preliminaries

### 2.1. $k$ -Means Clustering Algorithms

The  $k$ -means is a clustering algorithm to group objects based on their attributes into a  $k$  number of groups; where  $k$  is a positive integer number [18]. The grouping is based on minimizing some convergence criterion usually by minimizing the sum of squares of distances between the data points and the centroid of the cluster. The algorithm works as follows: First, we determine the number of  $k$  clusters we want to create. Then we randomly select the centers of these clusters (called centroids) by taking any random data points as the initial centroids. Then, determine the distance between each data point and the centroid. Euclidean distance or the Manhattan distance is usually selected to measure the distance. In case Manhattan distance is selected to measure the distance, then new centroids are computed as the component-wise median otherwise the mean computes the new centroids [19]. Data points are assigned to the closest centroid. The process is repeated until the minimum criterion is met or the assignment of data points to clusters becomes unchanged.

### 2.2. Hierarchical Clustering

Hierarchical clustering algorithm builds a cluster hierarchy or a tree of clusters. It assumes that every point is its own cluster, finds the most similar pair of clusters and start merging them into a parent pair of clusters [20]. There are two types of hierarchical clustering techniques either top down or bottom up. Bottom up is an agglomerative method that starts with one point (singleton). Then recursively add two or more appropriate clusters. The process is repeated till  $k$  number of clusters is achieved. The second technique is called Divisive or top down where it starts

with a big cluster then recursively divided into smaller clusters. The algorithm stops when  $k$  number of clusters is achieved.

### 2.3. Self Organizing Maps

The SOM clustering algorithm provided by Teuvo Kohonen, is an unsupervised learning neural network that comprises a one, two, or three-dimensional lattice of units, connected by weighted links to a layer of dummy input units. Lattice maps should be trained and strategy for training is simply based on that a unit or cluster of units will fire (generate output) when a certain kind of input is presented [21].

Training SOMs is based on vector quantization. We have a set of input vectors  $x_1, x_2, \dots$ . With each  $x$  being a vector using the concept of an input space, we can visualize the input vectors lying within an input space of the same dimension as the vectors. The idea of vector quantization is to divide up the entire space into regions. SOMs are based on competitive, or winner-takes-all, learning. The idea of competitive learning is that when a pattern represented by the input vector  $x$  is applied to the input units, one of the responding units wins the competition by virtue of its weight vector being the one with the smallest Euclidean distance from the input vector  $x$ . Then the weight vector is changed by moving it closer to input vector  $x$ . This is repeated for all patterns in the set [21].

### 2.4. Multilayer Perceptron Neural Networks

Artificial Neural Networks are mathematical models inspired by the biological nervous systems. Many tasks that humans perform naturally fast, such as recognition proves to be a very complicated task for a computer when traditional programming methods are used. By applying Neural Network techniques a model can learn by examples, and create an internal and complex structure of rules to classify different inputs, such as predicting customer churn. Neural networks are useful for pattern recognition or data classification, through a learning process. They simulate biological systems, where learning involves adjustments to the synaptic connections (weights) between neurons. They map a set of input-nodes to a set of output-nodes. The number of inputs/outputs is variable. And the Network itself is composed of an arbitrary number of nodes with an arbitrary topology. Artificial Neural Networks have been widely used to solve data mining problems since they adapt to unknown situations, robustness, fault tolerance, autonomous learning and generalization. In this work, we use the Multilayer Perceptron Neural Network (MLP-ANN) which is one of the most commonly applied neural networks models in the literature. We train our MLP-ANN model classifier by means of the famous back-propagation learning algorithm [22] [23].

## 3. Proposed Models

The churn prediction model proposed in the study is based mainly on two processes. In the first process a clustering algorithm is applied on the collected customers data in order to separate customers to a number of groups that represent different behaviours. Largest two clusters that represent churners and non-churners are merged again and considered as an input to the next process. On the other hand, the small clusters are neglected as they represent outliers and unrepresentative behaviours. In this research, three different clustering approaches are applied and compared;  $k$ -means, Hierarchical clustering and SOM. In the second process, MLP-ANN is applied on the clustered and filtered data obtained from the first process in order to learn how to classify customer behavior into two classes; churner customer and non churner customer. **Figure 1**, represents the two steps model applied in this work. The performance of the final developed MLP-ANN classifier model is assessed and evaluated using different evaluation criteria based on new testing data which was unrepresented in the classifier learning phase.

## 4. Dataset

Data used in this research was provided by Jordanian Telecommunication Company; one of the major cellular telecommunication company in Jordan. The data set contains 11 attributes of randomly selected 5000 customers subscribed to a prepaid service for a time interval of three months. The attributes cover outgoing/incoming calls statistics. The data were provided with an indicator for each customer whether the customer churned (left the company) or still active. The total number of churners is 381 (7.6% of total customers).

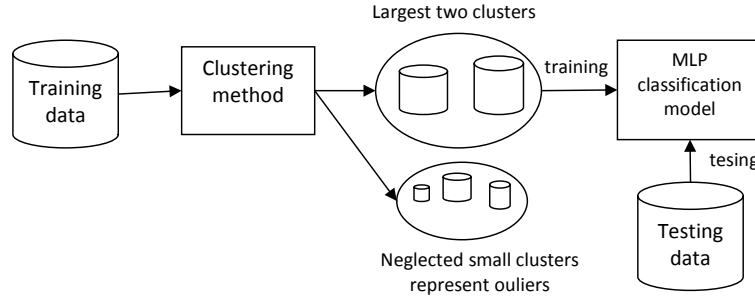


Figure 1. Proposed model for predicting customer churn.

## 5. Evaluation Methods

In order to evaluate the developed hybrid models we measure the model prediction accuracy and churn rate based on the confusion matrix shown in [Table 1](#). Accuracy and churn rate are calculated by Equation (1) and Equation (2) respectively.

$$\text{Accuracy} = \frac{(TP + TN)}{TP + FP + FN + TN} \quad (1)$$

$$\text{Churn Rate} = \frac{TN}{FP + TN} \quad (2)$$

Accuracy measures the rate of correctly classified instances of both classes and churn rate measures the rate of predicted churn in actual churn. The goal is to achieve high accuracy and churn rate values, high values mean well established model.

## 6. Experiments and Results

First, the dataset described earlier is split equally into training and testing parts. Then a clustering algorithm is applied on the training data in order to filter it from outliers and unrepresentative data. In the next stage, an MLP neural network model is developed using the filtered data and evaluated using the other testing data.

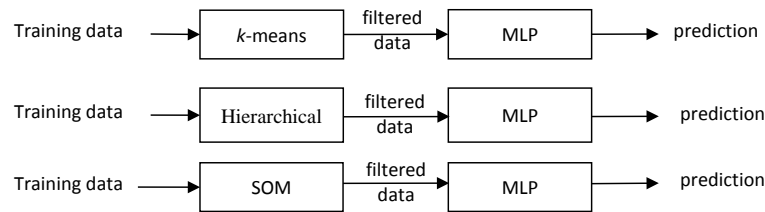
In our experiments we tried and compared three different clustering algorithms: *k*-means, hirarical clustering and SOM. *k*-means algorithm and hierarchical clustering are applied several times with different *k* value each time, starting from 4 up to 10 clusters. For SOM, different sizes are also applied as follows  $(2 \times 2)$ ,  $(2 \times 3)$ ,  $(3 \times 3)$ , and  $(4 \times 4)$ . The starting learning rate was set as 0.20 with variance distance normalization.

In order to develop the classification model, MLP neural network is trained using the filtered data obtained by each of the clustering methods in the first stage and tested using testing data. The parameters of the MLP are tuned as listed in [Table 2](#). The three different approaches shown in [Figure 2](#) are evaluated then compared with the C4.5 decision tree algorithm and the baseline MLP model without using any clustering technique. The results are shown in [Table 3](#). Imperically, the best results by means of churn rate were obtained by setting *k* to 7 for the *k*-means algorithm and 6 for the hierarchical clustering and  $3 \times 3$  for SOM.

In general, all the hybrid approaches showed some improvement in the accuracy and churn rate compared to the common approaches; C4.5 and MLP-ANN. The first hybrid model which combines *k*-means clustering and MLP neural network outperformed the other models in terms of prediction accuracy rate while the SOM clustering with MLP provided the best churn rate value.

## 7. Conclusion

One of the data mining goals is to improve the way we understand existing data and predict future behavior. This paper presented three hybrid models to help telecommunication companies to analyze and predict the future behavior of their customers. The three hybrid models are investigated to develop an accurate and efficient churn prediction models in order to help telecommunication companies to predict and analyze the future behavior of their customers. The three models are based on two phases; the clustering phase and the prediction phase. In the first phase, customer data is filtered. The second phase predicts the customer behavior. The first model investigates the *k*-means algorithm for data filtering, and Multilayer Perceptron Artificial Neural Networks (MLP-ANN)



**Figure 2.** Proposed hybrid models.

**Table 1.** Confusion matrix.

		Actual	
		Non Churners	Churners
Predicted	Non Churners	True Positive (TP)	False Negative (FN)
	Churners	False Positive (FP)	True Negative (TN)

**Table 2.** Neural Network settings.

Parameters	Value
Number of Neurons in the Hidden Layer	6
Momentum	0.2
Learning Rate	0.3
Epochs	500
Activation Function	$f(x) = \frac{1}{1 + e^{-x}}$

**Table 3.** Comparison between Three developed models and previous work.

	C4.5	MLP-ANN	k-means + MLP-ANN	Hierarchical + MLP-ANN	SOM + MLP-ANN
Accuracy	94.3%	94.3%	97.2%	94.8%	95.9%
Churn Rate	72.3%	68.3%	73.0%	74.2%	74.6%

for prediction. The second model uses hierarchical clustering with MLP-ANN. The third one uses self organizing maps (SOM) with MLP-ANN. The three models are developed and validated using real data provided by Jordanian Telecommunication Company, and the accuracy and churn rate values are calculated and compared. The comparison with the other models shows that the three hybrid models outperformed single model.

## References

- [1] Xia, G.E. and Jin, W.D. (2008) Model of Customer Churn Prediction on Support Vector Machine. *Systems Engineering—Theory & Practice*, **28**, 71-77.
- [2] Yabas, U., Cankaya, H. and Ince, T. (2012) Customer Churn Prediction for Telecom Services. *Computer Software and Applications Conference (COMPSAC), IEEE 36th Annual*, Izmir, 16-20 July 2012, 358-359. <http://dx.doi.org/10.1109/COMPSAC.2012.54>
- [3] Yeshwanth, V., Raj, V.V. and Saravanan, M. (2011) Evolutionary Churn Prediction in Mobile Networks Using Hybrid Learning. In: Murray, R.C. and McCarthy, P.M., Eds., *Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference*, AAAI Press, 471-476.
- [4] Kim, M.-J., Koh, S.-J. and Park, Y.-J. (2006) A Study on Retaining Existing Customers in the Korean High-Speed Internet Service Market. *Technology Management for the Global Future*, 2006. *PICMET*, **4**, 1970-1976.
- [5] Kang, C. and Pei-Ji, S. (2008) Customer Churn Prediction Based on Svmrfe. *ISBIM'08. IEEE International Seminar on*

- Business and Information Management*, Wuhan, 19 December 2008, 306-309.
- [6] Cao, J.T., Zhang, H. and Zheng, Q.S. (2010) Retaining Customers by Data Mining: A Telecommunication Carrier's Case Study in China. 2010 *International Conference on E-Business and E-Government (ICEE)*, Guangzhou, 7-9 May 2010, 3141-3144.
  - [7] Sharma, A. and Panigrahi, D.P.K. (2011) A Neural Network Based Approach for Predicting Customer Churn in Cellular Network Services. *International Journal of Computer Applications*, **27**, 26-31. <http://dx.doi.org/10.5120/3344-4605>
  - [8] Kim, S., Shin, K.-S. and Park, K. (2005) An Application of Support Vector Machines for Customer Churn Analysis: Credit Card Case. In: Wang, L., Chen, K. and Ong, Y., Eds., *Advances in Natural Computation*, Vol. 3611 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 636-647.
  - [9] Rodan, A., Faris, H., Alsakran, J. and Al-Kadi, O. (2014) A Support Vector Machine Approach for Churn Prediction in Telecom Industry. *International Journal on Information*, **17**, 3961-3970.
  - [10] Adwan, O., Faris, H., Jaradat, K., Harfoushi, O. and Ghatasheh, N. (2014) Predicting Customer Churn in Telecom Industry Using Multilayer Preceptron Neural Networks: Modeling and Analysis. *Life Science Journal*, **11**, 75-81.
  - [11] Rodan, A., Fayyumi, A., Faris, H., Alsakran, J. and Al-Kadi, O. (2014) Negative Correlation Learning for Customer Churn Prediction: A Comparison Study. *The Scientific World Journal*, **2015**, 1-7.
  - [12] Kasiran, Z., Ibrahim, Z. and Syahir Mohd Ribuan, M. (2012) Mobile Phone Customers Churn Prediction Using Elman and Jordan Recurrent Neural Network. *7th International Conference on Computing and Convergence Technology (ICCTT)*, December 2012, Seoul, 3-5 December 2012, 673-678.
  - [13] Burez, J. and Van den Poel, D. (2009) Handling Class Imbalance in Customer Churn Prediction. *Expert Systems with Applications*, **36**, 4626-4636. <http://dx.doi.org/10.1016/j.eswa.2008.05.027>
  - [14] Faris, H. (2014) Neighborhood Cleaning Rules and Particle Swarm Optimization for Predicting Customer Churn Behavior in Telecom Industry. *International Journal of Advanced Science and Technology*, **68**, 11-22. <http://dx.doi.org/10.14257/ijast.2014.68.02>
  - [15] Tsai, C.-F. and Lu, Y.-H. (2009) Customer Churn Prediction by Hybrid Neural Networks. *Expert Systems with Applications*, **36**, 12547-12553. <http://dx.doi.org/10.1016/j.eswa.2009.05.032>
  - [16] Obiedat, R., Alkasassbeh, M., Faris, H. and Harfoushi, O. (2013) Customer Churn Prediction Using a Hybrid Genetic Programming Approach. *Scientific Research and Essays*, **8**, 1289-1295.
  - [17] Faris, H., Al-Shboul, B. and Ghatasheh, N. (2014) A Genetic Programming Based Framework for Churn Prediction in Telecommunication Industry. *Lecture Notes of Computer Science*, **8733**, 253-362. [http://dx.doi.org/10.1007/978-3-319-11289-3\\_36](http://dx.doi.org/10.1007/978-3-319-11289-3_36)
  - [18] Mitchell, T.M. (1997) *Machine Learning*. McGraw-Hill, Inc., New York.
  - [19] MacQueen, J., et al. (1967) Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Oakland, 21 June-18 July 1965, 281-297.
  - [20] Johnson, S. (1967) Hierarchical Clustering Schemes. *Psychometrika*, **32**, 241-254. <http://dx.doi.org/10.1007/BF02289588>
  - [21] Kohonen, T. (1982) Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, **43**, 59-69. <http://dx.doi.org/10.1007/BF00337288>
  - [22] Dobbyn, C. (2007) *Intelligent Machines. M366: Natural and Artificial Intelligence*. Open University, UK.
  - [23] Zhang, Y., Liang, R., Li, Y., Zheng, Y. and Berry, M. (2011) Behavior-Based Telecommunication Churn Prediction with Neural Network Approach. 2011 *IEEE International Symposium on Computer Science and Society (ISCCS)*, Kota Kinabalu, 16-17 July 2011, 307-310.