ISSN (Print): 0974-6846 ISSN (Online): 0974-5645

### A Review on Customer Churn Prediction Data Mining Modeling Techniques

#### Nadeem Ahmad Naz, Umar Shoaib and M. Shahzad Sarfraz

Department of Computer Science, University of Gujrat, Pakistan; naz\_mphil@yahoo.com, umar.shoaib@uog.edu.pk, shahzad.sarfraz@uog.edu.pk

#### **Abstract**

**Objectives:** To find one of the best data mining techniques in telecommunication especially in customer churn prediction. **Methods/Statistical Analysis:** This paper presents a review of customer's churn prediction in the telecommunication. The study shows a large number of attributes that are used to put into practice to develop customer churn prediction model by the large number of reviewer. These attributes are segmentation, account info, billing info, call dialup types, line-info, and payment info, and complain info, service provider info, and services info. In this study appropriate modeling techniques such as LR, NNM, DT, FL, CMC, SVM and DME are discussed for the churning purpose. **Findings:** The Review shows that to find customer churn prediction depends on the objectives of decision maker's e.g. DT and SVM with a low ratio used, if interested in the true churn rate and false churn rate. The Logistic Regressions might be used if looking for the churn probability. DMEL modeling technique is impractical and ineffective for churn prediction on a large dataset with high dimension. **Application/Improvements:** The Technique proposed in this paper will overcome discussed issues and it will be applied on those customers who want to leave in near future and predict them based on some parameters.

**Keywords:** Customer Prediction, Churn Prediction, Data Mining, Data Modeling

#### 1. Introduction

The digital media is the one of the most favorite and powerful media that store a large amount of data for the last three decades1. There is an essential need of large data repositories to store and manage these types of data. The big challenge for such data repositories is that the information is bulky. The one most significant question arises now what is the method to extract meaningful information from such a huge amount of data. The answer is the Data Mining<sup>2</sup> because the data mining is the process of analyzing data from various aspects and summarizes it into valuable information<sup>3</sup>. Since the early 1960 Data Mining techniques have been considered to be as an area of applied artificial intelligence<sup>4</sup>. A large number of data mining techniques available to find out hidden information<sup>4</sup>. Data mining has the following main job practice of the daily life Anomaly detection<sup>5</sup>, Association<sup>6</sup>, Classification<sup>7</sup> Clustering<sup>8</sup>. Data mining is a multi-phased process that involved in mining of data process9.

All the leading organizations have been worked for the customer's best interest. A customer has a choice due to healthy competition among the services providers and there is no end to best services. Shortage of data, targeted sales and up-gradation of companies are the major challenges while attaining the new customers<sup>10,11</sup>. It is found that customer value and increasing revenue are the factors of current customer retention instead of the new customer acquiring<sup>11–14</sup>.

The companies know their current existing customers and come into a strong relation with them and have a huge amount of data about them which is the key point to increase the profit and customer value<sup>15–20</sup>. It has been assumed that CRM (Customer Relationship Management) is the valued topic in the telecommunication industries which is the main cause for increasing the sales as well as the customer retention<sup>21</sup>. This review paper focuses on how to conduct data mining in telecommunication especially in customer churn prediction. The technique and methodology used for churn prediction<sup>22</sup>

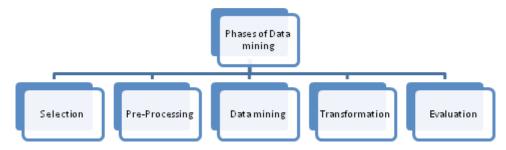


Figure 1. Phases of data mining.

by different researchers and their conclusion. Different phases of data mining are given in Figure 1.

#### 1.1 Customer Churn Prediction

When we talk the word churn the one thing which comes in mind is that customer churn is the big issue of telecommunication market<sup>23</sup>. The churn means those customers who want to leave in near future. There is essential need to predict those customers on behalf of some parameter to initiate some suitable action to minimize their leaving. The most of the mobile phone companies invest under CRM (customer relationship management) technology<sup>24</sup>. Multiple churn types are given in Figure 2.

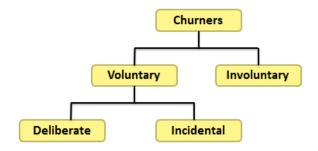


Figure 2. Types of churner.

## 1.2 CRM (Customer Relationship Management)

The CRM covers the Cross-Selling, Up-Selling, Customer Retention, New Customer Acquiring module. In Cross-selling we want to promote a specific product to those Customers who are not buying it normally. In Up-Selling we want to promote a product to that specific customer who is already its purchaser but our target is to force him for costly version to generate large amount of revenue<sup>25</sup>.

The main focus of Customer Retention is retention of current customer in an organization. The following are the reasons

- There is rare or no chance of the new customer in the telecommunication Industries due to saturation.
- Acquiring of new customer is costly for a company due to various reasons<sup>26</sup>

There is ten times increase in expenditure when we acquire new customer related to the expenses of retaining the current existing customer<sup>27</sup>.

The acquisition cost per customer was \$300 in 2004 as compared to Customer retention cost of \$25 in the US economy telecommunication industries in the same financial year. This is due to saturation in the telecommunication market<sup>28</sup>.

The Paper focal point is Customer retention or churning customer. The churn is defined as the inclination of a customer to leave a services provider. Churn prediction helps us when we identify expected churning customer, so that the companies could target only those customers instead of all customers for giving some incentives to retain them.

#### 1.3 Types of Churners

Churners are classified into two main categories that are voluntary and involuntary<sup>29</sup>. The voluntary churners are further more subdivided into deliberate and incidental churners. Involuntary churners are those customers that are remove from the list of customers due to their non-payment of bills, fraud or don't want to use telephone. The voluntary churners are difficult to find due to that customers want to terminate his services from his services providers.

The incidental churners comes due to incident because the churners have no plan to leave but this one done due to some reasons change of location, change in financial position etc. The deliberate churners occur due to customers need to change technology or price rate. The last one is the most important type of churners which are focus by mostly companies<sup>30</sup>.

## 2. Modeling Techniques of Churn Prediction

It is the basic need of the companies to develop an efficient and effective model to manage customers churn. There are so many modeling techniques that are used to predict customers churn in different organization. Ratio of papers and techniques is given in Figure 3. Here is the table of customer churn prediction models/techniques papers ratios<sup>21–31</sup>.

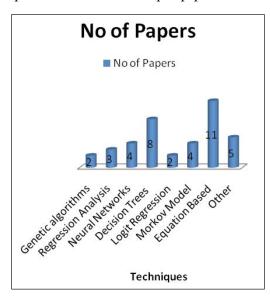


Figure 3. Paper and technique ratio.

CRT, NNM, LR, DT, SVM and fuzzy logic<sup>32</sup> are most frequently used techniques for Churn prediction. This paper describe about churn, churn prediction, and CRM, and the churn prediction techniques that used by various researchers in their research on customer churn prediction<sup>33</sup>. The paper describes the review of these techniques in short and also concludes that which one is the best technique under what condition and also a literature review of these techniques<sup>34</sup>.

# 3. Churn Prediction Techniques Methodology

Following are the techniques, algorithms and methodology that are used by different researchers in their research.

These techniques are of two types, Traditional Techniques and Soft computing Techniques. When we discussing different techniques of the customers churn prediction then first of all we want to know about the data set and the name of company.

Here is a list of those attributes that affect churning process and help us to predict the churners<sup>35</sup>.

- Customer demographics data
- Customer data
- Complaint data
- · Call detail
- Bill info
- · Payment info
- Customer age
- Fault report
- Monthly call details
- Payment type
- Consumption level rates
- Coverage of networks
- Area of customer
- Quality of services
- Price of per call or call rate
- Mobile packages
- Purchases history
- · Survey report

Hence a lot of others but the most important and suitable are as given below in a graph<sup>2</sup>. Figure 4 describe the ratio of attributes and churn effect. The data mining techniques are mostly used to predict the behavior of customer in the near future after observing the past behavior. The predicting model means predicting the customer who is likely to be churned<sup>36</sup>.

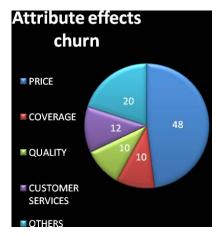


Figure 4. Attribute and churn effect.

#### 3.1 Neural Networks (NNM)

The Neural Networks Model (NNM) is used to elaborate functionality like non-linear. The model holds the capability to learn due to its comparable data processing structure. These techniques provide successful results after applying on many problems like classification, control and prediction due to biological brain<sup>37</sup>. The model is dissimilar to classification model as well as decision tree due to its likely hood prediction. The neural network has several techniques having merits and demerits. The researcher suggests neural network is better than decision tree and regression analysis model of churn prediction<sup>38</sup>.

#### 3.2 Linear Regression Model (LRM)

To predict customer satisfaction, regression analysis model is another popular technique that is based on supervised learning model. In this model a data set of past observation is used to see future values of explanatory and numerical targeted variable<sup>39</sup>

The Formula of LRM is given<sup>40</sup>.

$$prob(y=1) = \frac{e^{\beta 0} + \sum_{k=1}^{k} \beta_k X_k}{1 - e^{\beta 0} + \sum_{k=1}^{k} \beta_k X_k}$$

y is a binary variable. This shows event. If y = 1 the event occur else not occur.

X1, X2, ....., X Kbe the self-determining inputs. B0, B1, ...., BK be the failure

#### 3.3 Naive Bayes Model (NBM)

In this model the probabilities of a given input sample are calculated that belongs to a particular class y. The set of variable is given (X1---Xn).

The given formula is used to calculate probabilities.

$$(y_j|X)=p(y_j|X)p(y_j)=p(x_{1,}x_{2,...}X_n|y_j) p(y_j)$$

The y<sub>i</sub> is the probability of the previous calculations. The probability of independent variable is independent.

#### 3.4 Decision Tree (DT)

The decision tree is the most prominent predictive model that is used for the purpose of classification of upcoming trial<sup>41,42</sup>. The decision tree consists of two steps, tree building and tree pruning. In tree building the training set data is recursively partitioned in accordance with the values of the attributes. This process goes on until there is no one partition is left to have identical values. In this process some values may be removed from the data due to noisy data.

The largest estimated error rate branches are selected and then removed in pruning. To predict accuracy and reducing complexity of the decision tree is called tree pruning<sup>38,39</sup>.

#### 3.5 Support Vector Machine (SVM)

The SVM classifier deals with linear permutation of subset of the training set by finding a maximum edge over energized plane. The SVM plots the data into high dimensional features space closing to infinite with the help of most important part if vectors are nonlinearly divisible input features<sup>43</sup> and then categorize the data by the highest scope hyper-plane

$$f(\overline{x}) = \operatorname{sgn}\left(\sum_{i}^{M} y_{i} \alpha_{i} \Phi(\overline{x}_{i}, \overline{x}) + \delta\right)$$

where, as

M = the number of samples in training data set.

Xi = shows vector support when ai > 0

/ = shows a core function

X = unidentified sample feature vector

d = is a doorstep.

(ai) is a parameter that is the result of curved quadratic programming problem with respect to linear constraint<sup>44</sup>. In this technique it shows that Polynomial kernel & Gaussian radial basis functions (RBF) are frequently put into practice in favor of kernel functions.

The (d) is another parameter that is the result of picking any i where ai> 0 and the condition is Karush-Kuhn-Tucker condition44.

#### 3.6 Fuzzy Logic Algorithm

A fuzzy logic technique is very simple to understand due to its very simple mathematical concepts and fuzzy reasons. Fuzzy logic has the property of flexibility, tolerant of indefinite data. The function of random data can be implemented in this model. In most of the cases the FL

system spends the idea of the predictable managed techniques and streamlines the operations. In telecommunication industries no work has been achieved related to churn prediction with fuzzy logic techniques<sup>45</sup>.

## 3.7 Evolutionary Learning Data Mining Techniques (DMEL)

Data mining by evolutionary learning techniques is inherited classification technique. Such type of genetic algorithms has some set of rules for DMEL technique.

The DMEL applies these rules on some given dataset that provide decision making results<sup>46</sup>.

#### 3.8 K-means clustering

The mostly well-known and relevant technique of clustering is K-Means presented by Mc. Queen in 1967. Following are the main steps in K means clustering.

In K mean cluster in approach, in the first step we select k objects that have their centre (mean). In this method the remaining objects are not selected yet are assigned to cluster with respect to the similarity of the object with cluster. These similarities are measured on the behalf of the distance between cluster mean and object and after this calculation, the new center point is calculated on the behalf of above fact and we repeat these steps until the required function is achieved.

In k mean clustering, the most important point is to find the numbers of clusters that is optimum as well as the distance between cluster mean and objects.

The algorithm works until no new cluster element leave a cluster and enter into other cluster and no new center point is set for any cluster. When this target is achieved the algorithm is stopped.

#### 3.9 Ant Colony Optimization

The Ant Colony Optimization (ACO) met empirical motivated seeking behavior of real ant colonies<sup>47</sup>. The algorithm is the practical behavior of actually living ant that is an insect having some rules /skills used by them to find the food from his nest through shortest path first to food source. The first Ant colony algorithm was designed as ant system<sup>48</sup>. Ant colony optimization has workforce artificial ants works like biological ant to find the optimum solution<sup>47</sup>.

In Ant system in first step an ant selects a path to reach a point we set pheromone value but in case of problem, a

heuristic value is set. The pheromone value shows the trail and heuristic value shows the problems.

ACO is applied on large collection of problems<sup>47,50</sup>, Like vehicle routing problem, scheduling<sup>49</sup> and routing in packet-switched networks<sup>51,52</sup> in recent times, ACO has applied under data mining field<sup>53,54</sup>.

#### 4. Conclusion

This paper presents a review of customers churn prediction in the telecommunication. The paper shows a large number of attributes that are used to put into practice to develop customer churn prediction model by the large number of paper reviewer. These attributes are segmentation, account info, billing info, call dialup types, line-info, and payment info, and complain info, service provider info, and services info. In this paper approximately about 10 modeling techniques LR, NNM, DT, FL, CMC, SVM and DME are discussed for the churning purpose. The Review shows that to find customer churn prediction depends on the objectives of decision maker's .e.g. DT and SVM with a low ratio used, if interested in the true churn rate and false churn rate. The Logistic Regressions might be used if looking for the churn probability. DMEL modeling technique is impractical and ineffective for churn prediction on a large dataset with high dimension. The high dimensional data for NB modeling technique is necessarily transformed into the low dimension.

#### 5. References

- Howard R, Idriss S, Amanat O. Research base for the high-level group report analysis on media. Alliance of Civilizations Secretariat, United Nations, New York; p. 1–35.
- 2. Priya PI, Ghosh DK. A survey on different clustering algorithms in data mining technique. 2013; 3(1):267–74.
- 3. Rey TD. The dow chemical company chip wells, SAS Institute Inc. Justin Kauhl, Tata Consultancy Services Using Data Mining in Forecasting Problems; 2013. p.1–85.
- 4. Liao SH, Chu PH, Hsiao PY. Data mining techniques and applications. A decade review. Expert Systems with Applications. 2012; 39(12):11303–11. Crossref
- 5. Yun L, Hangzhou. Fast Co-clustering Using Matrix Decomposition. IEEE. 2009; 2:201–4.
- Yang F, Sun F, Zhang C. An efficient hybrid data clustering method based on K-harmonic means and particle swarm optimization. Expert Systems with Applications. 2009; 36(6):9847–52. Crossref

- 7. Cheng F, Tsai, Huang TW. QIDBSCAN a quick density based clustering technique idea. International Symposium on Computer. 2012; 3(1):1–322.
- 8. Esfandani G, Mohsen S, Amin N. GDCLU: A new grid density based clustering algorithm; 2012. p. 102–7.
- 9. Kaur AM, Kaur N. Review paper on clustering techniques. Global Journals Inc. (USA) Online. 2013; 13(5):1–7.
- Bhattacharya M. Telecom sector in India: Vision, IES (Deputy Director General (Finance), Department of Telecommunications, Ministry of Communications & IT, Government of India; 2020. p. 1–33.
- 11. Rasmusson E. Complaints can build relationships. Sales and Marketing Management. 1999; 151(9):89–90.
- 12. Athanassopoulos A. Customer satisfaction cues to support market segmentation and explain switching behavior. Journal of Business Research. 2000; 47(3):191–207. Crossref
- 13. Bhattacharya C. When customers are members: Customer retention in paid membership contexts. Journal of the Academy of Marketing Science. 1998; 26(1):31–44. Crossref
- 14. Colgate M, Stewart K, Kinsella R. Customer defection: A study of the student market in Ireland. International Journal of Bank Marketing. 1996; 14(3):23–29. Crossref
- 15. Mizerski R. An attribution explanation of the disproportionate influence of unfavourable information. Journal of Consumer Research. 1982; 9:301–10. Crossref
- 16. Stum D, Thiry A. Building customer loyalty. Training and Development Journal. 1991; 45(4):34–6.
- 17. Reichheld F. Learning from customer defections. Harvard Business Review. 1996; 74(2):56–69.
- 18. Zeithaml V, Berry L, Parasuraman A. The behavioural consequences of service quality. Journal of Marketing. 1996; 60(2):31–46. Crossref
- Paulin M, Perrien J, Ferguson R, Salazar A, Seruya L. Relational norms and client retention: External effectiveness of commercial banking in Canada and Mexico. International Journal of Bank Marketing. 1998; 16(1):24–31. Crossref
- Ganesh J, Arnold M, Reynolds K. Understanding the customer base of service providers: An examination of the differences between switchers and stayers. Journal of Marketing. 2000; 64(3):65–87. Crossref
- 21. Idris A, Rizwan M, Khan A. Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. 2012; 38(6):1808–19.
- 22. Verbeke W, Martens D, Mues C, Baesens B. Building comprehensible customer churn prediction models with advanced rule induction techniques. 2011; 38(3):2354–64.
- 23. Chandrasekhar S. Predicting the Churn in Telecom Industry; 2015.

- 24. Ondrus J, Pigneur Y. Coupling mobile payments and CRM in the retail industry; 2004. p. 1–8.
- 25. Hung S, Yen DC, Wang H. Applying data mining to telecom churn management. Expert Systems with Applications. 2006; 31:515–24. Crossref
- 26. den Poel VD, Lariviere B. Customer attrition analysis for financial services using proportional hazard models. European Journal of Operational Research. 2004; 157(1):196–217. Crossref
- 27. Jhon H. A customer profiling methodology for churn prediction; 2008. p. 1–313.
- 28. Seo D, Ranganathan C, Babad Y. Two-level model of customer retention in the US mobile telecommunications service market. Telecommunications Policy. In Press, Corrected Proof. 2008; 32(3):182–96.
- 29. Gotovac S. Modeling data mining applications for prediction of prepaid churn in telecommunication services. 2010; 51(3):275–83.
- 30. Kim, Yoon C. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. Telecommunications Policy. 2004; 28(9):751–65. Crossref
- Huang B, Kechadi MT, Buckley B. Customer churn prediction in telecommunications School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland. 2012; 39:1414–25.
- 32. lazarove V, Churn MC. prediction business analytics course. Tum Computer Science; 2007. p. 104–212.
- 33. Coussement K, Van D, Poel D. Churn prediction in subscription services an application of support vector machines while comparing two parameter selection techniques. Expert Systems with Applications. 2008; 34:313–27. Crossref
- 34. Kaur M, Mahajan P, Churn Prediction in telecom industry using R. International Journal of Engineering and Technical Research. 2015; 3(5):46–53.
- 35. Georges D, Olle O, Cai S. A hybrid churn prediction model in mobile telecommunication industry. International Journal of e-Education, e-Business, e-Management and e-Learning. 2014; 4(1):55–62.
- 36. Hassouna M. Agent based modelling and simulation: An examination of customer retention in the UK mobile market; 2012. p. 1–173.
- 37. Behara R, Fisher W, Lemmink J. Modelling and evaluating service quality measurement using neural networks. International Journal of Operations and Production Management. 2002; 22(10):1162–185. Crossref
- 38. Chan C, Yao X. A novel evolutionary data mining algorithm with applications to churn prediction. 2003; 7(6):532–45.
- 39. Vercellis C. Business Intelligence Data Mining and Optimization for Decision Making. Wiley; 2009. p. 1–420.

- 40. Langley P, Iba W, Thompson K. An analysis of Bayesian classifiers. AAAI'92 Proceedings of the tenth national conference on Artificial intelligence; 1992. p. 223–8.
- 41. Muata K, Bryson O. Evaluation of decision trees a multi criteria approach. Computers and Operational Research. 2004; 31:1933–45. Crossref
- 42. Boser B, Guyon I, Vapnik V. A training algorithm for optimal margin classifiers. In Proceedings the 5th annual ACM workshop on computational learning theory Pittsburgh, PA: ACM Press; 1992. p. 144–52. Crossref
- 43. Burges CJC. A tutorial on support vector machines for pattern recognition.1998; 2:121–67.
- 44. Data Mining and Knowledge Discovery. Springer. 1997; 2(2):121–67.
- 45. Selvi O. Traffic accident predictions based on fuzzy logic approach for safer urban environments. Case Study: Izmir Metropolitan Area; 2009. p. 1–131.
- Au W, Chan C, Yao X. A novel evolutionary data mining algorithm with applications to churn prediction. IEEE Transactions on Evolutionary Computation. 2003; 7:532–45. Crossref
- 47. Dorigo M, Stützle T. Ant colony optimization. Cambridg. MA: MIT Press; 2004. p. 1–321.
- 48. Dorigo M, Maniezzo V, Colorni A. Ant system optimization by a colony of cooperating agents. IEEE Transactions

- on Systems, Man, and Cybernetics Part B Cybernetics. 1996; 26(1):29–41. Crossref. PMid:18263004
- Larivière B, Van D, Poel, D. Predicting customer retention and profitability by using random forest and regression forest techniques. Expert Systems with Applications. 2005; 29(2):472–84. Crossref
- 50. Vanthienen J, Mues C, Wets G, Delaere K. A tool-supported approach to inter-tabular verification. Expert Systems with Application.1998; 15(3):277–85. Crossref
- Babu S, Ananthanarayanan NR, Ramesh V. A survey on factors impacting churn in telecommunication using data mining techniques. International Journal of Engineering Research and Technology. 2014; 3(3):1745–8.
- Abraham A, Ramos V. Web usage mining using artificial ant colony clustering. Congress on Evolutionary Computation; 2003. p. 1384–91. Crossref
- 53. Handl J, Knowles J, Dorigo M. Ant-based clustering and topographic mapping. Artificial Life. 2006; 12(1)35–61. Crossref. PMid:16393450
- 54. Churi A, Mayuri D, Sonal D, Prajakta K. Analysis of customer churn in mobile industry using data mining. International Journal of Emerging Technology and Advanced Engineering. 2015; 5(3):2250–459.