



A Review on: Data Mining for Telecom Customer Churn Management

Dr. Mamta MadanProfessor at VIPS, GGSIPU
India**Dr. Meenu Dave**Professor at Jagannath University, Jaipur
India**Vani Kapoor Nijhawan**Assistant Professor at VIPS, GSIPU
India

Abstract- Customer acquisition and retaining those customers is a matter of concern for all companies and the same applies for telecom sector too. So customer churn is an important area of concern. This research work aims at carrying out a literature review for the past decade reviewing around 50 research papers in the area of telecom churn with two perspectives: mining technique applied and publication year. Such a survey will be a help to the telecom service providers in determining the appropriate model to identify the prospective churners and accordingly framing the marketing strategies for different customer groups, to reduce the rate of customer turnover.

Keywords :- Churn Prediction, Data Mining Techniques, Decision Tree, Logistic Regression, Neural Network, Support Vector Machine.

I. INTRODUCTION

In the telecom industry, the biggest loss of revenue is happening because of increasing customer churns. Such customers who are not loyal to the company result in a financial burden on the company. This fact is very well known that the cost of finding new customers is far more than retaining the old ones. So, detecting the “going to be churner” customers beforehand is the objective of the telecom companies.

A. Data Mining

Data Mining can be defined as “the process of searching large stores of data to discover patterns, associations and trends to dig out useful structures from large amounts of data stored in different databases or other information repositories.” [w1]

There are many organizations which are using data mining techniques for managing their customer relationships, including getting new customers, increasing revenue from existing customers, and retaining high valued and loyal customers.

B. Data Mining In Telecom

According to [1], data mining in the field of telecommunication can be used for the following purposes:

- Churn prediction: - The process of predicting the customers who are at a risk of leaving the company is known as churn prediction in telecommunication. These customers should be focused upon, and efforts should be made to retain them. This is very important because retaining a customer is less expensive than acquire a new one.
- Insolvency prediction: - Hike in the number of due bills are becoming an important area of concern for all telecom companies. In such a competitive environment, companies cannot bear the burden of insolvency. To find such insolvent customers, data mining technique can be used. Customers who will turn defaulters, can be predicted beforehand.
- Fraud Detection: - Fraud is an expensive affair for the telecom industry, so the companies should try to predict fraudulent users by identifying their usage patterns.

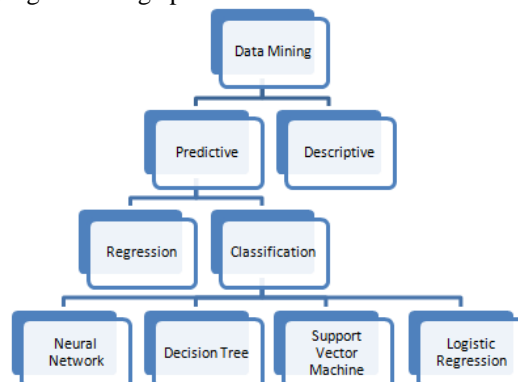


Fig.1 Research Methodology

C. Customer Churn

Subscriber churning in the mobile telecommunication industry refers to the shift of subscribers from one provider to another. It is a very common practice of many subscribers to frequently churn from one service provider to another in search of better prices or services or features.

With the help of data mining, we aim to fetch some interesting patterns, on the basis of which we may be able to identify the reasons for churn and we may also predict who are about to churn in the near future.

II. DATA MINING IN CHURN: LITERATURE REVIEW

A lot of researchers have used different data mining techniques for predicting customer churn in the field of mobile telecommunications. They have explored a number of techniques and applied various algorithms, namely neural network, decision tree, logistic regression and have also given some suggestions.

John Hadden et al.[1] have shared with the help of a tabular listing (Fig. 7) that the decision tree has come out to be the best technique out of the three may be because of its rule based architecture. And about neural networks and regression it is said that these methods have been trained to make calculations to check for churners and non churners. But for neural and regression the accuracy depends mainly on the weights for the neural networks and coefficients for the regression.

Khalida et al. [2] described and demonstrated a predictive model for customer churn using Decision Tree Analysis model. They have used a training sample set to conduct an experiment of customer churn and as a result they analyzed that area is the main factor for the customer to churn.

Another researcher paper by KhakAbi et al. [3], provides a brief review of papers from two perspectives: techniques used(model applied) and statistical reports. Through their findings, they have tried to point out the gaps and strengths in the area of churn.

Another model that has been built and tested by Shyam and Ravi [4], to predict future churn for wireless customers. This model obtained a 68% predictive accuracy, using the Naïve Bayes algorithm. In their model, Shyam and Ravi has suggested a very unique approach to "mine their own data", which means that the organizations should mine their existing customer information to develop models of predicting customer behavior.

Junxiang Lu[5] in his research paper, says that methods like decision tree, logistic regression, etc. are predicting customer churn but they hardly tell that when the churn will happen. He has used survival analysis techniques to predict which customer will churn and when the churn will happen, thereafter, helping the telecom companies in customizing their customer treatment programs.

The work of Chen et al.[6] is based on the use of SAS Enterprise Miner to design a predictive model for churn management, providing the foundation for predicting the data of customer churn and reduce the number of churns.

One more study by Shan Jin et al.[7] emphasizes on a predictive model to find out possible churners and provide personalized services. The researchers have explored a few data mining techniques, namely decision tree c4.5, BPN and compared the effectiveness of each for customer churn.

K. Dasgupta, R. Singh et al.[9] have shown a different perspective. They have said that chances of a subscriber to churn depend upon the number of his social ties i.e. his friends, who have already churned. Based on some findings, they have suggested a technique that is predicting potential churners by observing the current set of churners and their social network.

M. Lejeune[10] has given an interesting fact that customer churn is correlated with the life cycle of the telecom industry. If the industry is in growing phase, then more and more customers are getting associated with the company. So even if some of the customers are churning, total number of customers are still increasing.

M. Mozer, Grimes et al.[12] in their paper on wireless telecom churning, have discussed that the nonlinear neural network outperforms the logistic regression. Also the neural net appears to outperform the decision tree and boosting seems to help in at least the upper portion of the lift curve.

Reference [27] have used three hybrid models. First model uses k-means clustering and multilayer perceptron artificial neural network (MLP-ANN). Their second model applies hierarchical clustering with MLP-ANN. And the third model uses self organizing maps (SOM) with MLP-ANN. It was observed that all three models gave better results than any single model.

After reading all these research works, some of these papers have been categorized as per the techniques used by their authors and put it in table 2.

Table 2: Data Mining Techniques with corresponding count of authors following it.

DataMining Technique	Reference
Neural Networks	[20],[22],[27],[28],[30],[35],[39],[40],[42],[43],[44],[45],[47],[53]
Decision Tree	[2],[22],[29],[31],[32],[34],[35],[37],[38],[39],[40],[44],[47],[54],[55],[56],[57]
Logistic Regression	[33],[39],[41],[43],[47],[48],[49],[50],[51],[52]
Support Vector Machine	[21],[33],[39],[40],[41],[52]

Another table, Table3 is showing the same work as the Table2, but in a distributed way on the basis of year in which the paper was published. Here, in a decade's span we are trying to show the continuity of the work being done in the area of telecom churn. Also, we can see in the Table3 that the technique which is getting maximum count is Decision Tree.

Table 3: Research Papers following different mining techniques with their publication year.

	Publication Year											Total
	20 05	20 06	20 07	20 08	20 09	20 10	20 11	20 12	20 13	20 14	20 15	
Neural Network	3	3	1	2	1	1	1			1	1	14
Decision Tree	2	1	2	2	3	4		1	1	1		17
Logistic Regression	2	2		3	3							10
Support Vector Machine				3	2			1				6

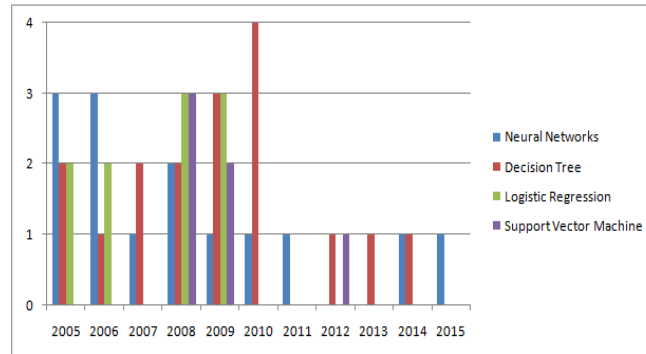


Fig.2 Application of Data mining techniques in publications from 2005 to 2015.

In Table3 and Fig.1, we have presented a statistical report of the research of a decade (i.e. from 2005 to 2015) being done in the field of telecom customer churn. From this, it can be observed that *Neural Network* and *Decision Tree* are the two most popular mining techniques being used in telecom customer churn. Both of these have their own advantages. Decision trees are easy to understand and interpret. They don't quite robust and do not require much of the data preparation. Also, they can perform well with large volumes of data. And Neural networks have the biggest advantage that they can recall information based on incomplete or noisy data. So they perform well with data intensive applications.

III. CONCLUSION AND FUTURE SCOPE

This paper aims at reviewing the recent literature in the area of telecom customer churn mainly with two perspectives, i.e. technique being applied to telecom churn prediction and the publication year. Also the aim is to help the researchers to gain insight into the recent trends in this area, which will guide them in finding the possible reasons for the churn and consequently reducing them and helping the telecom sector to reduce their financial loss.

As per data presented in Table2 and Table3, it can be seen that the current flow of the research is using either Neural Network or Decision Tree as the mining techniques.

It should be noticed that, the review does not include all the research in this area, but the results are only on the basis of the papers studied by the authors.

REFERENCES

- [1] Hadden J. , Tiwari A. , Roy R. , Ruta D. "Churn prediction: Does technology matter?" International Science Index, Engineering and Technology, Vol.2, No.4 , 2008.
- [2] K. binti,O. binti,Mohd Shukor,Norazrina,A. Haris,F. Bakar, "Data Mining in Churn Analysis Model for Telecommunication Industry", Journal of Statistical Modeling and Analytics, pp 19-27, vol.1,2010.
- [3] S. KhakAbi, M. Namvar, Mohd R. Gholamian, "Data Mining Applications in Customer Churn Management", Proc. Of IEEE international conference on Intelligent Systems, Modelling and Simulation,2010.
- [4] S. V. Nath, Dr. R. S. Behara," Customer Churn Analysis in the Wireless Industry: A Data Mining Approach ", <http://download.oracle.com>.
- [5] Junxiang Lu, " Predicting Customer Churn in the Telecommunications Industry – An Application of Survival Analysis Modeling Using SAS", <http://www2.sas.com/proceedings/sugi27/p114-27>.
- [6] Chen Yu-bao, Li Bao-sheng,Ge Xin-quan, "Study on Predictive Model of Customer Churn of Mobile Telecommunication Company",Proceeding IEEE Fourth International Conference on Business intelligence and Financial Engineering, 2011.
- [7] S. Jin,Y. Meng, C. Fan,F. Peng,Q. Chen," The Research on Applying Data Mining to Telecom Churn Management",Proc.2nd International Conference on Materials, Mechatronics and Automation,2012, Vol. 15, 2012, pp.39-45.
- [8] "Predictive Data Mining",Introduction to Data Mining and Knowledge Discovery,Two crows, 3rd Edition.
- [9] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi,"Social ties and their relevance to churn in mobile telecom networks", Proceedings of the International conference on Extending database technology, pp. 668-677, New York, NY, USA, 2008. ACM.
- [10] M. Lejeune, "Measuring the impact of data mining on Churn Management," Journal of Electronic Network Applications and Policy, pp. 375-387, vol. 11 (5), 2001.

- [11] Wei C., Chiu I., "Turning telecommunications call details to churn prediction: a data mining approach", Expert Systems with Applications, Elsevier, pp. 103-112, vol. 23, no. 2, Aug.2002.
- [12] M. Mozer, R. Wolniewicz, D. B. Grimes, H. Kaushanky, E. Johnson,"Predicting subscriber dissatisfaction and improving retention in wireless telecommunications industry." IEEE Transactions on Neural Networks,2000.
- [13] "Customer Care For High Value Customers: Key Strategies", white paper, Infosys, Aug 2010.
- [14] Arthur M Hughes.(2003).The customer loyalty solution.[on-line]. http://www.books.mec.biz/downloads/...Customer_Loyalty_Solution.../ [Dec 12, 2013].
- [15] Gregory Piatetsky-Shapiro."What analytics, Big data, Data mining, Data science software you used in the past 12 months for a real project?"KD Nuggets(May 2013). <http://www.kdnuggets.com/polls/2013/analytics-big-data-mining-data-science-software.html>
- [16] R. Bouckaert, E. Frank, M. Hall, G. Holmes, B. Pfahringer, P. Reutemann, I. Witten, "WEKA-experiences with a java open-source project" Journal of Machine Learning Research 11. 2010.
- [17] W. H. Au, K. Chan, X Yao,"A novel evolutionary data mining algorithm with applications to churn prediction", IEEE Transactions On Evolutionary Computation, VOL. 7, NO. 6. Dec 2003.
- [18] Kirui C.,Hong L.,Cheruiyot W., Kirui H., "Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining", Intl. Journal of Computer Science Issues, Vol. 10, Issue 2, No 1,Mar 2013.
- [19] Richter Y.,Tov E.,Slonim N., "Predicting Customer Churn in mobile networks through analysis of social groups", Proc. 10th SIAM International Conference on data mining,2010.
- [20] Song G.,Yang D.,Wu L.,Wang T.,Tang S., "A mixed process neural network and its application to churn prediction in Mobile Communications", Proc. 6th IEEE International Conference on Data Mining, 2006.
- [21] E. Shaaban, Y. Helmy, A. Khedr, Mona Nasr, "A Proposed Churn Prediction Model", International Journal of Engineering Research and Applications, vol. 2, Issue 4, 2012, pp.693-697.
- [22] Hung S. ,Yen D. ,Wang H., "Applying data mining to telecom churn management", Expert Systems with Applications, Elsevier, Vol. 31,2006, pp.515-524.
- [23] C. Rygielski, J. Wang and D.C. Yen, "Data mining techniques for customer relationship management", Technology in Society, Vol. 24,2002.
- [24] Georges D. Olle Olle and Shuqin Cai, "A Hybrid Churn Prediction Model in Mobile Telecommunication Industry", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 4, 2014.
- [25] Rahul J. Jadhav and Usharani T. Pawar, "Churn prediction in telecommunication using data mining technology ", Intl. Journal of Advanced Computer Science and Applications,vol. 2, no.2,2011.
- [26] R Mattison,(2005, June 1). The Telco Churn Management Handbook (USA, XiT Press, Illinois, 2005), pp. 35-42.
- [27] A. Hudaib, R. Dannoun, O. Harfoushi, R. Obiedat, H. Faris, "Hybrid data mining models for predicting customer churn", International Journal of Communications, Network and System Sciences,volume 8, pp. 91-96, 2015.
- [28] Alam Khan A.,Jamwal S. , Sepehri M. , "Applying data mining to customer churn prediction in an internet service provider ",Intl. Journal of Computer Applications, Vol. 9, No.7, 2010.
- [29] V. Umayaparvathi, K. Iyakutti, " Applications of Data Mining Techniques in Telecom Churn Prediction", International Journal of Computer Applications, Vol. 42, No.20, 2012.
- [30] A. M. Almaná, M. S. Aksoy, R. Alzahrani, " A Survey On Data Mining Techniques In Customer Churn Analysis For Telecom Industry", Journal of Engineering Research and Applications, volume 4, Issue 5, 2014, pp.165-171.
- [31] N. Hashmi, N. A. Butt, Dr.M. Iqbal, " Customer churn prediction in telecommunication a decade review and classification" International Journal of Computer Science, Vol. 10, No 2, September 2013.
- [32] Dr. M.Balasubramanian, M.Selvarani, "Churn Prediction In Mobile Telecom System Using Data Mining Techniques", International Journal of Scientific and Research Publications, Vol. 4, April 2014.
- [33] Coussement K. , D. V. D. Poel, "Churn Prediction in Subscription Services: An Application of Support Vector Machines While Comparing Two Parameter-Selection Techniques," Expert Systems with Applications, vol. 34, 2008, pp. 313–327.
- [34] Chu Bong-Horng, Tsai Ming-Shian, Ho Cheng-Seen, "Toward a Hybrid Data Mining Model for Customer Retention," Knowledge Based Systems, volume 20, 2007, pp. 703–718.
- [35] Hu X., "A data mining approach for retailing bank customer attrition analysis," Applied Intelligence, vol. 22, 2005, pp. 47–60, Springer.
- [36] Song H. S. , Kim J. K. , Cho Y. B. , Kim S. H. , "A personalized defection detection and prevention procedure based on the Self-Organizing Map and Association Rule Mining: Applied to online game site," Artificial Intelligence Review, vol. 21, 2004, pp. 161–184.
- [37] Zan M., Shan Z., Li L., Ai-jun L. , "A predictive model of churn in telecommunications based on data mining," Proc. IEEE International Conference on Control and Automation, IEEE Press, 2007.
- [38] Wang Yi-Fan, Chiang Ding-An, Hsu Mei-Hua, Lin Cheng-Jung, Lin ILong, "A Recommender System to Avoid Customer Churn: A Case Study," Expert Systems with Applications, vol. 36, 2009, pp. 8071– 8075.

- [39] J. Zhaoa and Xing-Hua Dang, "Bank customer churn prediction based on support vector machine: Taking a commercial Bank's VIP customer churn as the example," Proceedings International Conference on Wireless Communications, Networking and Mobile Computing, 2008, pp. 1-4.
- [40] Y. Xie and X. Li, "Churn Prediction with Linear Discriminant Boosting Algorithm," Proceedings International Conference on Machine Learning and Cybernetics, Kunming, 2008.
- [41] J. Burez, Poel D. Van den, "Handling Class Imbalance In Customer Churn Prediction," Expert Systems with Applications, vol. 36, 2009, 4626–4636.
- [42] Zhang G. , "Customer retention based on BP ANN and survival analysis," Proc. International Conference on Wireless Communications, Networking and Mobile Computing, 2007, pp. 3406-3411.
- [43] Buckinx W., Poel D. Van den, "Customer Base Analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting," European Journal of Operational Research, vol. 164, 2005, pp. 252–268.
- [44] L. Yan, M. Fassino, P. Baldasare, "Predicting Customer Behavior via Calling Links," Proceedings International Joint Conference on Neural Networks, Montreal, Canada, 2005.
- [45] E Xu, S. Liangshan, G. Xuedong, Z. Baofeng, "An algorithm for predicting customer churn via BP Neural Network Based on Rough Set," Proceedings The 2006 IEEE Asia-Pacific Conference on Services Computing, 2006.
- [46] "Predicting Customer Behavior in Telecommunications," Yan L. , Wolniewicz R. H. , Dodier R. , IEEE Intelligent Systems, IEEE Computer Society.
- [47] Glady N., Baesens B. , Croux C. , "Modeling churn using customer lifetime value," European Journal of Operational Research, vol. 197, 2009, pp. 402–411.
- [48] J. H. Ahna, S.P. Hana, Y.S. Lee, "Customer churn analysis: Churn Determinants and Mediation Effects of Partial Defection in the Korean Mobile Telecommunications Service Industry," Telecommunications Policy, vol. 30, 2006, pp. 552–568.
- [49] Coussement K. , Poel D. Van den , "Integrating the Voice of Customers Through Call Center Emails into a Decision Support System for Churn Prediction," Information & Management, vol. 45, 2008, pp. 164–174.
- [50] B. Larivie`re, Poel D. Van den, "Predicting Customer Retention and Profitability By Using Random Forests And Regression Forests Techniques," Expert Systems with Applications, vol. 29, 2005, pp.472–484.
- [51] A. Prinzie T, Poel D. Van den, "Incorporating Sequential Information Into Traditional Classification Models By Using An Element/Position-Sensitive SAM," Decision Support Systems, vol. 42, 2006, pp. 508–526.
- [52] Coussement K. , Peol D. Van den, "Improving Customer Attrition Prediction By Integrating Emotions From Client/Company Interaction Emails and Evaluating Multiple Classifiers," Expert Systems with Applications, vol. 36, 2009, pp. 6127–6134.
- [53] A. Sharma, Dr. Panigrahi P. K. , " A Neural Network based approach for predicting customer churn in cellular network services", International Journal of Computer Applications, vol. 27, no. 11, 2011.
- [54] Goran , Sven, "Modeling Data Mining Applications for Prediction of Prepaid Churn in Telecommunication Services",ISSN 0005-1144, ATKAFF 51(3),pp. 275-283, 2010.
- [55] C. F. Tsai, M. Y. Chen, "Variable selection by association rules for customer churn prediction of multimedia on demand", Expert Systems with Applications, 2009, doi: 10.1016/j.eswa.2009.06.076.
- [56] R. Mansouri, Mohd. Saraee, R. Amirfattahi, "Application of Data mining in predicting cell phones Subscribers Behavior Employing the Contact pattern", International Conference on Data Storage and Data Engineering, 2010, doi: 10.1109/DSDE.2010.57.
- [57] Ding Pan, "On Customer Churn and Early Warning Model of Telecom Broadband" Third International Symposium on Information Processing, China, 2010.
- [58] Han, Kamber, *Data Mining: Concepts and Techniques*. Berlington, US: The Morgan Kaufmann Series in Data Management Systems, 2006, pp. 1-30.

Web References

- [w1] http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm. [September, 2015]
- [w2] <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.2563&rep=rep1&type=pdf>.