International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878, Impact Factor (2014): 3.05

A Review on Customer Churn Prediction in Telecommunication Using Data Mining Techniques

S. Babu[1], Dr. N. R. Ananthanarayanan[2]

[1]Research Scholar, Dept of Computer Science and Applications, SCSVMV University, Enathur, Kancheepuram, India
[2]Associate Professor, Dept of Computer Science and Applications, Enathur, Kancheepuram, India

Abstract: ***Customer churn is the term which indicates the customer who is in the stage to leave the company. Particularly it is happening recurrently in the telecommunication industry and the telecom industries are also in a position to retain their customer to avoid the revenue loss. Prediction of such behaviour is very vital for the present market and competition and Data mining is the one of the effective technique for the same. So the research is focused in this area and this paper reviewed about 64 research papers in the aim of identifying the data Mining techniques and models used to predict the customer churn. In addition to this, study about the lack of the existing models, there by defining the new model to predict the churn in the telecom industry.***

Keywords: Churn, Telecommunication, Retain, Data Mining Techniques

1. Introduction

The Indian telecom sector is the one of the fastest growing sector in the world. Mobile prepaid Customers are tremendously changing their service providers with in the short span of time. So the customer churn rate increases in the same way as the increase of the mobile subscribers. Irrespective of the growth of the telecom sector, customer churn is the most challenging task for the industry. Service providers also redesign their service plans with new offers to retain the customers. In this perspective it is necessary to investigate the reason for customer churn in India.

2. Review of Literature

1. [Rahul J. Jadhav, Usharani T. Pawar] Data mining can be applied for the following issues of telecommunication sector.

A) Churn Prediction:

Prediction of the customers who are in the stage to leave the company is called as Churn Prediction. The company should take the necessary steps to retain them. This issue is very important one because retaining the customer is less expensive than acquiring a new one.

B) Insolvency Prediction:

Increase of due bills is the next big issue of telecom sector. In the current competitive market, the telecom companies cannot afford the cost of insolvency. Insolvent customers, (i.e.) who refuse to pay their bills can be predicted in advance with the help Data mining techniques.

C) Fraud Detection:
Fraud is very expensive activity in the telecom sector. So the telecom companies should try to find the fraudulent users and their patterns.

2. [Gary M. Weiss, 2005]: To find the solution for the above issues using data mining, the first step is to understand the data. The following are the three main types of data which is available in telecommunications.

A) Customer Data:

Large business companies like Telecommunication have millions of customers. As a basic requirement the database of all the customers are maintained. This database contains the information like name, address, age, gender, income, service plan, contract information, credit score, and payment history, etc.

B) Call Detail Data:

Call detail data has been formed when a call is made on the network. During the call the descriptive information about the call is stored as a call details record. As a minimum the following are the information that is recorded for every call. They are originating and terminating phone numbers, Date and Time of Call, Duration of call.

C) Network Data:

Network of the telecom sectors are very complex, comprising of thousands of interrelated components. Every network components are able to generating error and status messages, which falls in a huge volume of network data. This type of error message minimally contains a timestamp, a unique string that reveals the hardware or software which generates the error message and the code which explains why the error has been generated.

3. [Li-Shang Yang, Chaochang Chiu, 2006]: This paper focused on a preliminary study about applying data mining techniques to solve real world customer churn problem in telecom sector. The data from the Taiwan mobile Service Company was used and to perform the analysis Logistic regression technique and Decision tree was used. The analysis was done on both the techniques with 5000 name list and the results are compared based on hit ratio. From the comparison it has been concluded that, the performance of logistic regression was found as poor than the decision tree.

4. [Ionut Brandusoiu, Gavril Toderean, 2013] In this paper, as an advanced methodology, Support Vector Machines algorithm with four kernel functions were used

to predict the churners and the non churners in telecommunications. To perform the analysis, the dataset of 21 attributes with 3333 records has been considered. Based on the result it has been concluded that, prediction of both churners and non churners, has been performed best by Support Vector Machine Polynomial kernel function with the accuracy of 88.56%. It is also concluded that the other three kernel functions also performed best in the prediction of customer churn with the performance accuracy around 80%.

5. **[Amjad Hudaib, Reham Dannoun, Osama Harfoushi, Ruba Obiedat, Hossam Faris, 2015]** In this paper, to enhance the existing models, three Hybrid models has been developed for the efficient churn prediction over the telecommunication market. The models are based on the following two phases, i.e. the first phase is the clustering phase and the second one is the prediction phase. During the analysis, in the first phase the customer data has been filtered and in the second phase prediction of customer churn has been done. Out of this three models, the first model performs the analysis based on k_means clustering for filtering the customer data and Multi Layer Perceptron (MLP – ANN) for predicting the customer churn, the second model performs the analysis based on Hierarchical clustering for filtering the customer data and Multi Layer Perceptron (MLP – ANN) for predicting the customer churn, and the third model performs the analysis based on Self Organizing Maps (SOM) with MLP-ANN. The developed three models are validated using the real time data which is provided by Jordanian Telecommunication Company. The performance accuracy and churn rate values are calculated and the same was compared. To prove the performance of the Hybrid models, the results are compared with other models (C5.0 & MLP-ANN) also. From the comparison, it has been concluded that the Hybrid models outperformed than the single models.

3. Research Methodology

As the Journals are the most reliable source of research work, some online journals are referred to get a clear vision on the topic.

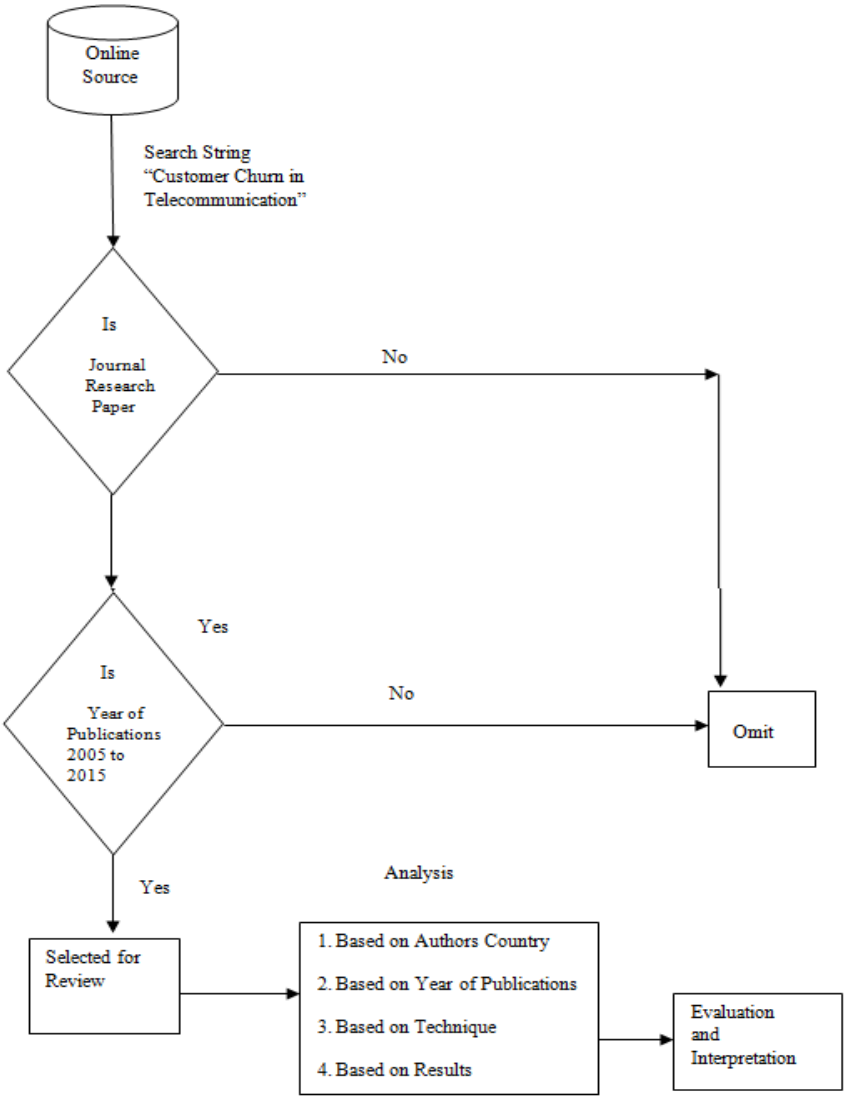

Figure 1: Overall Research Methodology

As a first step, all the journals related to Customer Churn in Telecommunication have been searched. As a next step, the journals which are published during the year 2005 to 2015 were considered for the review. Finally the selected articles were reviewed in the following four classes:

- Review of Journals published based on the country of authors and year wise (India & Others)
- Review of Journals based on Techniques used for the prediction of the customer churn.
- Review of Journals based on the year of publication.
- Review of the techniques based on the performance of the models.

Figure 1 depicts the overall development of the research methodology described above in more comprehensive way.

4. Classification of Journals

The total of 64 journals been selected and analysed in the above mentioned four dimensions. The objective of this analysis is to get the clear view on the following vision.

1)To investigate the level of Research focus on the concept of data mining in the telecommunication industry in India.
2)To reveal about the techniques of data mining in the telecommunication industry.
3)To study about the performance of the existing models
4)To establish the opportunities and challenges in the advanced techniques of data mining in telecommunication.

4.1 Journals by Country of Authors

Irrespective of the growth of the telecom sector, customer churn is the most challenging task for the industry. In this perspective, more insight is needed to know the reason for customer churn in India. Data mining is the one of the best technique for this. In this view, to know the awareness of data mining in telecommunication in India, the 64 journals are classified as shown in table-1.

Table 1: Year wise No of Journals based on authors country

Year of Publication	*No. Of. Research Papers Based on Country*	
	India	*Others*
2005	-	4
2006	-	6
2007	-	5
2008	-	8
2009	-	8
2010	-	4
2011	3	6
2012	1	4
2013	2	5
2014	3	1
2015	3	1
Total	**12**	**52**

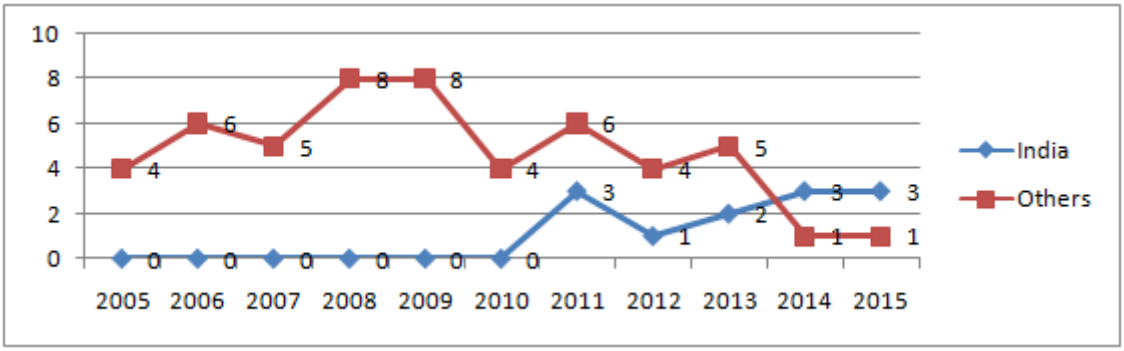

Figure 2: Year wise No of Journals based on authors country

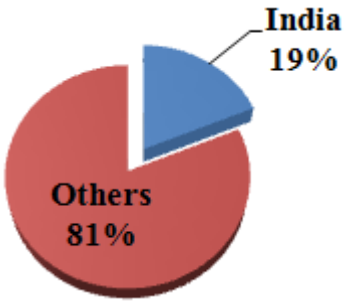

Figure 3: Comparison of Research Journals based on author's country over a decade

The statistical Table-1 clearly reveals that, the awareness of the concept of data mining in telecommunication in India is not matured that much during the last decade. It also state that only 19% (12 out of 64) of research papers only published during a decade.

4.2 Journals by Techniques

To predict the churn, different prediction algorithms used. In this, Logistic Regression, Decision Tree, Neural Networks, Support Vector Machine and Combination of above are the some of the well known techniques. To know about the usage of above data mining techniques in the area of telecommunication, the 64 journals were classified based on the techniques.

Table 2: No of Journals based on Technique

Techniques	*No. of Research Papers*
Decision Tree	27
Neural Networks	14
Support Vector Machine	11
Logistic Regression	7
Hybrid Models (Clustering + ANN)	3

From the table 2 and Figure 3, it has found that, the usage of Decision Tree was more matured than the other technique.

As next level, the research focus been turned towards Neural Networks and Support Vector Machine. The table also reveals that very limited focus was given on logistic regression and the concept of hybrid techniques has just been started, to improve the performance quality of the churn prediction model.

4.3 Journals by Techniques and Year of Publication

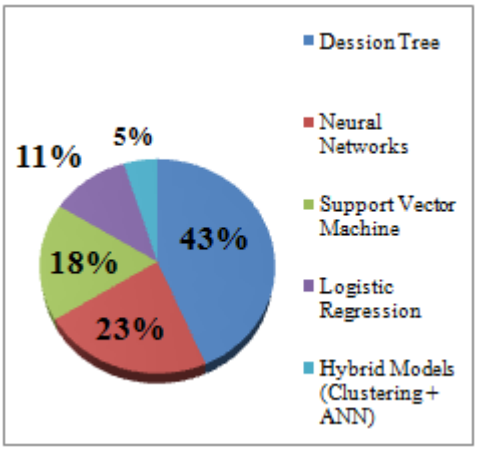

Figure 4: Research Focus Based on Techniques

Table 3: Distribution of Journals based on Techniques and Year of publication

Year of Publication	*No. of Research Papers*				
	Decision Tree	*Neural Networks*	*Support Vector Machine*	*Logistic Regression*	*Hybrid Models (Clustering + ANN)*
2005	1	1	1	1	
2006	1	3		2	
2007	4		1		
2008	1	2	4	1	
2009	3	2	1	2	
2010	4				
2011	5	1	2	1	
2012	2	2	1		
2013	1	3	1		1
2014	3				1
2015	2				1

To know the recent techniques of data Mining in telecommunication and to know about the effective techniques which perform best in all the way, the distribution of the journals has done as shown in table 3.

The above table clearly states that, the recent trend also focuses on the decision tree to predict the customer churn. Even though the focus over the neural networks and Support Vector Machine has been slightly reduced, reviews show that they are used in the hybrid models.

Apart from that from the analysis it was proved that, research focus has slightly turned towards the Hybrid models such as **cluster + MLP** and **cluster + SOM** to improve the accuracy over the churn prediction.

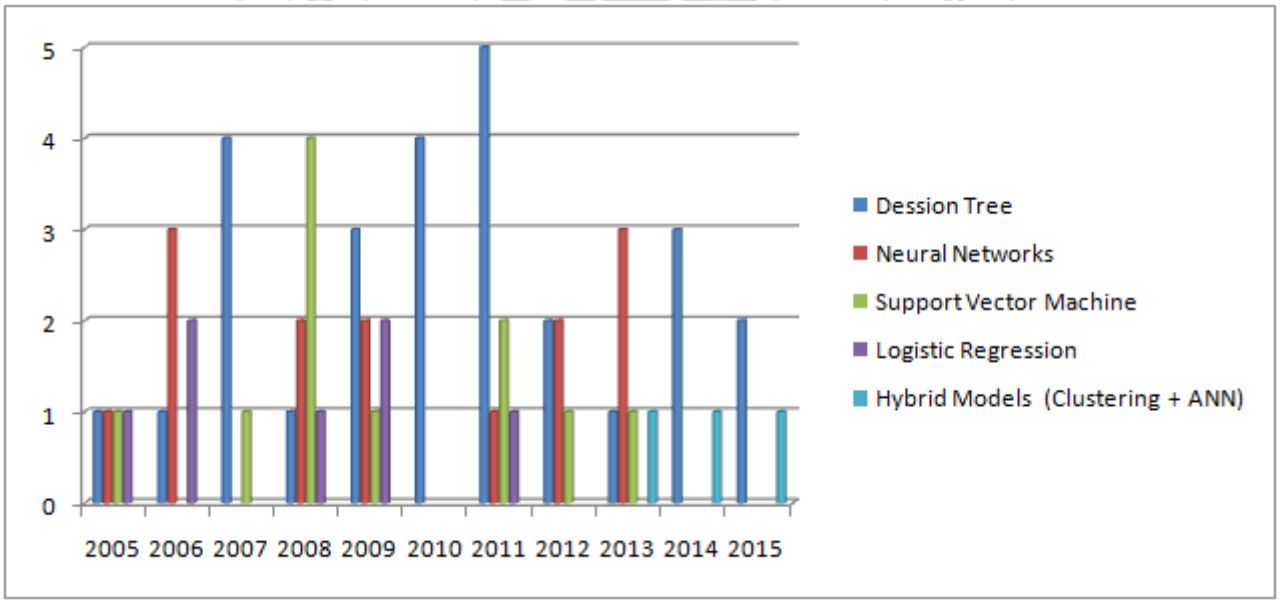

Figure 5: Comprehensive representation of journal distribution based Techniques and year of Publication

4.4 Journals by Performance

To establish and to identify the opportunities and the challenges in the advanced techniques of data mining in telecommunication, the comparative study has been done based on the accuracy of results produced by various models.

The statistical data of table 4, depicts that model based on Logistics Regression performs poor than Decision Tree [28]. According to the analysis done by [44], it was

concluded that, Support Vector Machine has also produces best accuracy in prediction of customer churn. In addition to the above, through the table-4 it is also identified that, the recent challenges has focused on hybrid models to improve the performance quality on the prediction of churn. Based on the result comparison done by [62], hybrid models perform and produce best accuracy than the single models.

5. Conclusion

Today telecommunication industry is facing a critical issue of customer churn. Ultimately revenue loss is the cause of the issue. The only way to avoid revenue loss, which happens due to churn, was the prediction of customer churn well in advance. Data mining techniques aids the telecommunication industry to pick out such customers, so that retention activity can be taken against them.

Table 4: Comparison of Hybrid model and Single model based on results

	Other Models				*Hybrid Models*		
Techniques	Logistic Regression	Support Vector Machine	Decision Tree (c4.5)	ANN	K-Means + MLP	Hierarchical Clustering + MLP	SOM + MLP
Reference	4	5	6	6	6	6	6
Accuracy (%)	Poor Than Decision Tree	88.56	94.3	94.3	**97.2**	**94.8**	**95.9**

With this idea, the research has been started in this area. This paper focuses on reviewing the literature over a decade in the area of customer churn in telecommunication. On the process of review 64 papers has been selected and analyzed in four different dimensions, to know about maturity of data mining in Indian telecom and to know about the recent techniques in churn prediction.

From the analysis, it is concluded that till 2010, there was no focus on Indian telecom and during the last five years only there was a slight focus on Indian telecom. So the researchers can initiate their research on Indian telecom to predict the customer churn.

As next, it has been concluded that current research is focusing on decision tree and neural networks. Apart from that, recent trend initiated Hybrid models to improve the performance of the existing model and it is proved and outperformed than the single models. So the researchers can focus more on the hybrid model to enhance the existing models, to achieve more accuracy than the other models.

6. Limitations of Research

As stated earlier that, not all the research journal are considered for the review. But, only research journals which are reliable in all the perspective are considered. So this review has some limitations.

- As first, 64 journals from the last one decade has been considered for the study. More journals can be selected by extending the duration.
- As Second, the journals were searched based on the string "Customer churn in telecommunication". But churn is common activity which happens in other sector also. So search can be done based on the string "churn analysis" or "churn management" etc. to find other recent models for churn prediction.

References

[1] Dr. Mamta Madan Dr. Meenu Dave Vani Kapoor Nijhawan, "A Review on: Data Mining for Telecom Customer Churn Management", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 9, September 2015.
[2] Nabgha Hashmi, Naveed Anwer Butt and Dr. Muddesar Iqbal, "Customer Churn Prediction in Telecommunication A Decade Review and Classification", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013.
[3] Rahul J. Jadhav, Usharani T. Pawar, "Churn Prediction in Telecommunication Using Data Mining Technology", International Journal of Advanced Computer Science and Applications, Vol. 2, No.2, February 2011
[4] LI-SHANG YANG, CHAOCHANG CHIU,"Subscriber Churn Prediction in Telecommunications", 2006.
[5] Ionut BRANDUSOIU, Gavril TODEREAN, CHURN PREDICTION IN THE TELECOMMUNICATIONS SECTOR USING SUPPORT VECTOR MACHINES Issue #1, May 2013.
[6] Amjad Hudaib, Reham Dannoun, Osama Harfoushi, Ruba Obiedat, Hossam Faris "Hybrid Data Mining Models for Predicting Customer Churn", J. Communications, Network and System Sciences, May 2015 , 8, 91-96.
[7] Manpreet Kaur, Dr. Prerna Mahajan, "Churn Prediction in Telecom Industry Using R", International Journal of Engineering and Technical Research (IJETR) ISSN: 2321-0869, Volume-3, Issue-5, May 2015.
[8] 8.Aishwarya churi, Mayuri Divekar, Sonal Dashpute, Prajakta Kamble, " Analysis of Customer Churn in Mobile Industry using Data Mining", International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 5, Issue 3, March 2015).
[9] 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.
[10] S. Babu, Dr. N. R. Ananthanarayanan, V.Ramesh, "A Survey on Factors Impacting Churn in Telecommunication using Datamininig Techniques", International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 3, March – 2014.
[11] N. Kamalraj, Dr. A. Malathi, " Applying Data Mining Techniques in Telecom Churn Prediction", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 10, October 2013.
[12] 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, No. 1, February 2014.
[13] Gary M. Weiss, DATA MINING IN TELECOMMUNICATIONS, Department of Computer and Information Science Fordham University, 2005