International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 08 | Aug -2017 www.irjet.net p-ISSN: 2395-0072

IMPROVED CUSTOMER CHURN BEHAVIOUR BY USING SVM

Deepika Mahajan[1], Rakesh Gangwar[2]

[1]*M.tech Scholar Department of Computer Science & Engineering, Beant College of engineering and technology Gurdaspur, Punjab*
[2]*Assistant Prof Department of Computer Science & Engineering, Beant College of engineering and technology Gurdaspur, Punjab*

Abstract - *Churn Prediction has been major research problem with the growth of market development as customers asset more valuable persons for growth of company. The proposed Hybrid approach is an integration of two techniques named random forest and Support Vector Machine (SVM) provides better and accurate results in the prediction of churn customers The proposed Hybrid approach is implemented in MATLAB with Statistics toolbox on the dataset of customers having 3333 instances and 21 attributes to evaluate the performance of proposed Hybrid approach. Various parameters are to be considering into experiment in enhancing the performance of Hybrid approach. The experiment results reveal good distinction of churn and loyal customers from the given dataset and provide more accurate and satisfactory results when the Hybrid approach is compared with various classifiers or algorithms.*

***Key Words*: Customer churn behaviour, Churn customers, CRM.**

1. INTRODUCTION

An organization contains huge volume of data and it is not possible to make prediction on huge amount of Data. There is need to extract useful data so that predictions can be made on them. Data mining is the process of preparing useful or meaningful, taken from large Databases (A.churi and R,Mahe, 2015). Converting raw data into useful data in order to make patterns is called Data Mining.

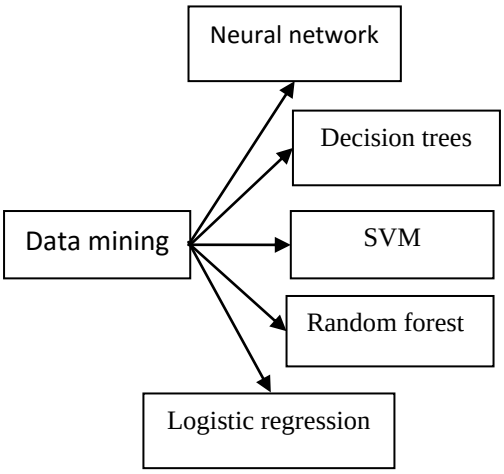

Fig 1. Data Mining Technique

2. Customers relationship management (CRM)

CRM can be defined as a set of business activities to increase business performance in customer management. Customer demands can be changed with time variation. CRM relationship is understandable by customer life cycle or customer lifetime. The goal of CRM is to ensure customer satisfaction and delight at every level of interface with the company.'CRM' refers to managing relationship with customers. It is a process or method used to learn more about customers need and behaviors in order to develop strong relationships with customers. It is type of Management, which is used to satisfy the customer needs. The loyalty of customers depends on their satisfaction of product or service.

CRM helps to manage the churn customers in the company. It helps to attract the new customers. There exist four dimensions of CRM - Customer Identification, Customer attraction, Customer Retention and Customer development (K.Rodpysh and M. Majdi, 2012).

- **Customer identification**: CRM begins with acquiring customers to the company by indentifying them. This is phase where people want to become customers or most profitable persons for the company. This is basically related to group of customers as they may lead to profit or loss of company.

- **Customer attraction**: This phase of CRM helps to make long relationship with their customer by providing them numerous offers such as discount on product, free products etc. Customer attraction depends on the satisfaction of products. Satisfied persons help to increase the retention rate by providing positive information to new employees. It becomes major role to attract the new customers as churn customers likely to move from one company to another.

- **Customer retention**: It occurs when company fulfill the needs of customers. A customer can retain themselves in a company only when their needs are fulfilled or they are satisfied with the service given by existing company. To retain the customers in the company, it becomes necessary to complete the demands of the customers.

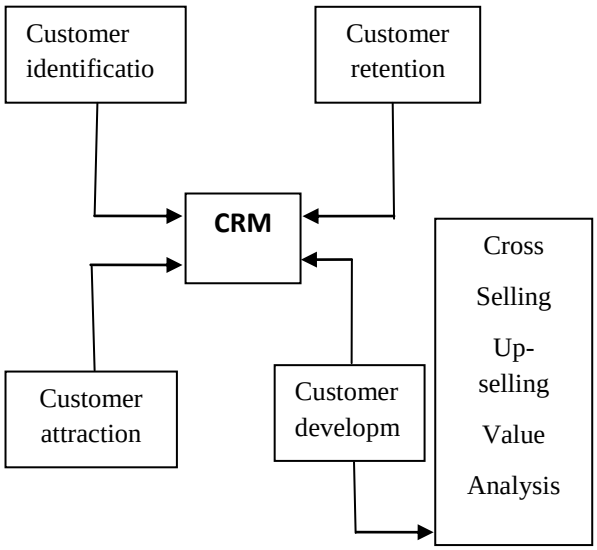

Fig2 : CRM Framework

CRM process can be defined as activities to manage the customer relations by grouping them according to behavior. It helps to make perceptions by identifying customers, customer knowledge preparation and by building customer relationships. Basically CRM process can be grouped into different parts- Vertical & Horizontal process, Front & Back office and Primary & Secondary process and three levels- Customer Facing level, Functional level and Companywide level into three levels- Customer Facing level, Functional level and Companywide level.

3. Literature survey

Dr. M. Balasubramanian and M. Selvarani (2014) used KDD (Knowledge Discovery in Data Mining) to hold the churn members in the company, as competitor's increases at high rate. KDD is performed with using two categories hypothesis and Discovery Oriented. The limitation arise with the model is that it only helps to predict churn customers, not to determine the appropriate solution to retain them in company. The methodologies used by them were Data Acquisition, Data Preparation, Derived Variable and Variable extraction. With the selection of appropriate attributes and fixing threshold values, accurate results may be produced.

M. Lapczynski (2014) developed a hybrid model of C& RT logit model by integration of decision tree and logistic model.Hybrid model produce improved results than basic logistic model as it used decision tree with it. It provider better results when compared to single DT. The hybrid approach also helps to obtain different probabilities of each test case. Decision trees help to detect lack of data and logistic regression extend the interpretation. At each test, unique predicted probabilities are obtained. The limitation that arises during research is that Hybrid C& RT -logit can be applicable only to Single Decision tree.

S.Sonia and Dr. C. Nalini(2014) used Mapreduce to predict the churn customers in telecom industry, Mapreduce and HDFS(Hadoop Distributed File system) helped to mine the large dataset. The use of hadoop MapReduce resolved the problems of data mining. MapReduce used to provide the good performance in the form of reliability; scalabity and efficiency. Mapreduce helps to reduce data size, hence in reducing complexity. NameNode, and the DataNodes for HDFS JobTracker and the TaskTracker nodes For MapReduce nodes are used for analysis. Hadoop NameNode is converted into Hadoop Distributed File System. HMapReduce help to predict churn customers that lead to produce customized approach for retention methods.

Manjari Anand et al. (2014) implemented ART (Artificial Resonance Theory) algorithm to perform the customer classification based on the choices. The dataset was taken from the company having vehicles on sale. Classification of customers can be implemented using ART algorithm and was compared with back propagation algorithm. ART algorithm was proved to as better algorithm for the classification customers and found to use less time for customer classification. This algorithm used less time complexity than propagation algorithm. ART algorithm provides the best time complexity and was implemented in MATLAB.

Manjeet kaur and Dr.Kawaljeet Singh (2013) elaborated guidepost on exchanging unnecessary client information of a bank into effective and useful information with DM techniques like naive bayes, decision trees and SVM to pick out important client features to predict churn . The methodology is made of data sampling, data preprocessing, model construction, and model evaluation phases . Churn rate success is greater than loyal class when prediction is made. The prediction of loyal customers is more than churn customers when all algorithms are analyzed.

N.Kamalraj and Dr.A.Malathi(2013) determine the possible churners using the predictive data mining model. The main goal of the research is to get the complete investigation about the data analysis in the critical process to precede the successful data mining application. It is used to investigate the data analysis; robust predictive model can be built by discovering the significant churn factors. It also examines in keeping the predictive models to make the mobile operators in order to perform them accurately. The techniques are used for the large data sets of the telecommunications industry.

R .Obiedat and O. Harfoushi (2013) implemented Hybrid approach of K-mean clustering and Genetic Programming to predict churn customers. K-mean clustering is used to filter the dataset and Genetic Programming helps to classify the customers into churners and non-churners. Four clusters are to be used out of two clusters are discarded. Selected classifiers are loaded into model and results are compared with C4.5, ANN and GP with accuracy and churn rate. The accuracy rate does not classify exact churn and loyal customers, which is main limitation of this hybrid approach.

S.Nabavi and S.Jafari(2013) implemented CRISP-DM (Cross Industry Standard Process for Data Mining) on RFM (Recency, Frequency and Monetary) with two different technologies named RandomForest and Boosted technique on the dataset of Solico food industry. The churn customers are to be predicted and effective measures are to be planned in retaining them. It helps to identify churn customers using customer behavior analysis and variables. Among all variables, LOR, Relative Frequency and average inter purchase time are best predictors for churn prediction.

Dr.U. Devi and S.Madhavi (2012) used CART (Classification and Regression Tree) and C5.0 on the Dataset of Bank customers. Trees are grown and then pruned back. With CART, it becomes easy to split data into binary and make patterns for remaining data. Data Mining is used to convert raw Data into useful information. CART provides high success rate of churn class and C5.0 gives high success rate of active class. Effective churn prediction model help to attain benefits from its efforts. Holding 5 percent of old customers can increase the profit of company by 25 percent.

4.Experimentation and Results

A. METHODOLOGY

Step 1: Start the Algorithm.
Step 2: Now SVM is implemented and its performance to be checked
Step 3: New Hybrid approach for the churn prediction is to be implemented and comparisons
Step 4: Now comparisons are to be made in order to predict the accurate model from existing and purposed Hybrid
Step 5: Stop the Algorithm.

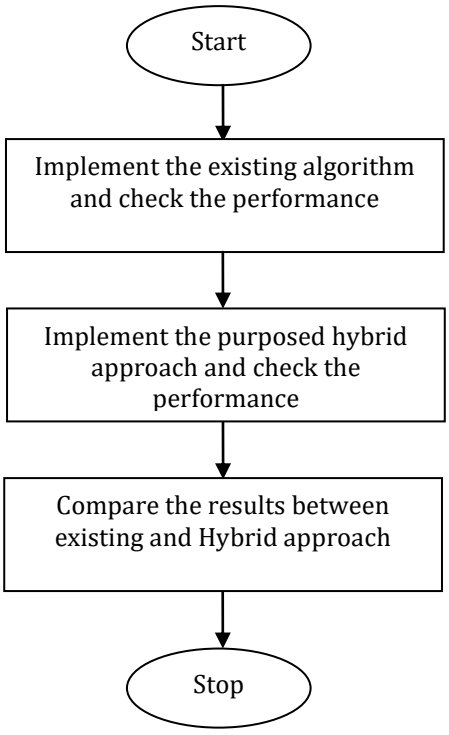

Fig 3: Proposed methodology

B. PERFORMANCE ANALYSIS

This paper has designed and implemented the proposed technique in MATLAB tool u2013a. The evaluation of proposed technique is done on the basis of following metrics i.e. Accuracy, F-measure, true positive rate, kappa statistics, error rate. A comparison is drawn between all the parameters with existing and proposed algorithm and figures shows all the results.

1. Accuracy, Error rate and RAE

It is defined as number of instance per classes that have been correctly identified. It relates to those instances, which are being identified as correct or positive while making predictions.

It can be defined as number of instances that have been negatively classified. It is also called Error rate. It is basically related to the wrong predictions or incorrect predictions. RAE is proportional to simple forecaster. RAE accepts total absolute error and anneals it by dividing by total absolute error of simple predictor.

Table 5.5, it has been clearly indicated that proposed Hybrid approach provides the better, results that is accuracy than existing. Higher the accuracy rate, higher will be the outcome produced. Accuracy rate is linked with the improvement of purposed algorithm. Highest value in Accuracy and lower values in Error rate and RAE shows the improved results obtained by purposed algorithm against existing algorithms.

Table 1. Analysis of different parameters

Algorithm Name	Accuracy	Error rate	Root Absolu Error(RAE)
AdaBoostM1	86.4787%	13.1213%	73.0989%
random forest	95.3596%	4.2404%	33.0781%
SVM	88.2588%	11.4412%	68.39%
Logistic Regression	86.3887%	13.3113%	75.5602%
PART	96.1496%	3.5504%	26.0567%
Decision Table	91.2791%	8.3209%	60.5888%
Filtered Classifier	94.1394%	5.5606%	40.2962%
Bayes Network Classifier	94.2894%	5.4106%	42.6417%
Purposed Hybrid approach	97.1997%	2.5003%	17.6552%

We have studied various algorithms based on Data Mining and Hybrid approach of random forest and SVM provides better and accurate results when comparisons are made on various algorithms.

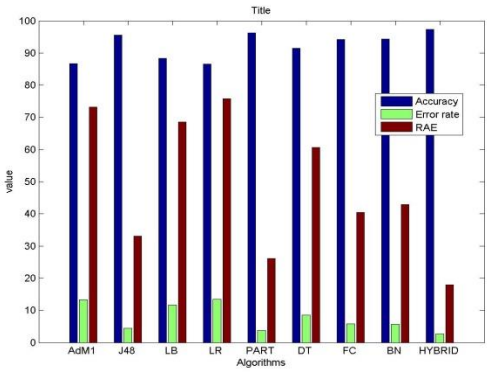

Fig 4. Analysis of different parameters

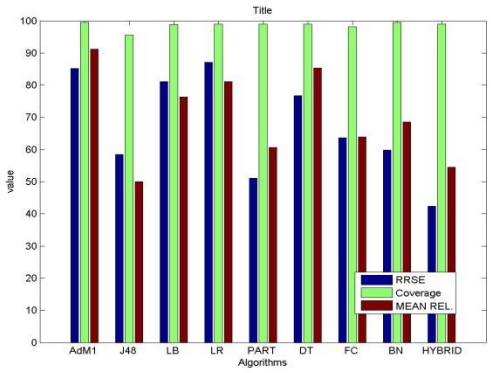

Fig 5. Analysis of different parameters

2. RRSE, Coverage Cases, Mean rel. region size

Root relative Squared Error (RRSE) is mean of genuine values. It takes the total squared error and shortens the fault of same attributes as amount being forecasted.

Number of coverage cases refers to the steps in which calculations are made. Number of coverage cases may be high or low based on the accuracy results.

Mean Absolute Error (MAE) is used to appraise that how close the anticipations are to the actual values.

Table 5.6 shows the improved and enhanced results that are made with the help of new proposed hybrid algorithm by using RRSE, coverage of cases and Mean Rel. region size. Coverage of cases shows the number of cases that has been covered with the algorithm during the implementation of the code.

Table 2. Analysis of different parameters

Algorithm Name	RRSE	Coverage o Cases	Mean rel. region size
AdaBoostM1	85.1074%	99.2899%	91.1091%
random forest	58.305%	95.3596%	50%
SVM	81.0629%	98.5999%	76.1976%
Logistic Regression	87.0518%	99.0089%	81.0581%
PART	51.0611%	99.0089%	60.3911%
Decision Table	76.5574%	98.7449%	85.1835%
Filtered Classifier	63.456%	98.1798%	63.6664%
Bayes Network Classifier	59.5792%	99.2899%	68.3119%
Purposed Hybrid approach	42.2864%	98.7499%	54.3554%

5. CONCLUSION

The occurrence of churn customers puts adverse impact on the profit of company. Therefore, it becomes necessary to predict the accurate churn customers and loyal customers so that proactive measures could be taken in consideration in order to retain the customers into the company as customers are valuable persons to the growth and survival of company. We have studied various algorithms based on Data Mining and Hybrid approach of random forest provides better and accurate results when comparisons are made on various algorithms.

REFERENCES

[1] Bart Baesens, Geert Verstraeten, Dirk Van den Poel, Michael Egmont (2004), " Bayesian network classifiers for identifying the slope of the customer lifecycle of long life customers", European Journal of Operational Research, Vol. 156, Pp. 508-523

[2] Burez J. and Van D. (2008), "Separating Financial from Commercial Customer Churn: A Modeling Step towards Resolving the Conflict between the Sales and Credit Department," Expert Systems with Applications, Vol. 35, Issue 1, pp. 497-514

[3] Chih-Fong Tsai and Mao-Yuan Chen (2009), "Variable selection by association rules for customer churn prediction of multimedia on demand," Expert Systems with Applications, Vol.30, Pp. 1-10

[4] Chris Rygielski , Jyun-Cheng Wang and David C. Yen(2002) , " Data mining techniques for customer relationship management, " Technology in Society, Vol. 24, Pp. 483–502

[5] Dr. M.Balasubramanian, M.Selvarani(2014), "Churn prediction in mobile telecom system", International Journal of Scientific and Research Publications", Vol. 4, No. 4, Pp.1-5

[6] Dr. U. Devi Prasad and S. Madhavi (2012)," Prediction Of Churn Behavior Of Bank Customers Using Data Mining Tools, " Business Intelligence Journal, Vol.5, No.1, Pp. 96-101

[7] H. Abbasimehr, M. Setak, and M. Tarokh (2014), "A Comparative Assessment of the Performance of Ensemble Learning in Customer Churn Prediction," The International Arab Journal of Information Technology, Vol. 11, No. 6, Pp. 599-606

[8] Emtiyaz, Mohammad Reza Keyvanpour(2011), "Customers Behavior Modeling by Semi-Supervised Learning in Customer Relationship management ", Advances in information sciences and Service Science, Vol.3, No. 9, Pp. 229-236

[9] E. Shaaban, Y. Helmy, A. Khedr and M. Nasr(2012) , " A Proposed Churn Prediction Model, " International Journal of Engineering Research and Applications, Vol. 2, No. 4, Pp.693-697

[10] Georges D. Olle Olle and Shuqin Cai(2014), "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,Pp.55-62

[11] Indranil Bose and Xi Chen (2009), "Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn" Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol. I, IMECS, Hong Kong

[12] J. Basiri, F. Taghiyareh, B. Moshiri(2010), "A Hybrid Approach to Predict Churn", Proceedings of Asia-Pacific Services Computing ," Conference IEEE, Pp.485-491

[13] J. Burez and D. Van den Poel (2009), "Handling class imbalance in customer churns prediction", Expert Systems with Applications, Vol.36, Pp. 4626–4636

[14] J. Hadden, A.Tiwari, R. Roy, and D. Ruta(2008), " Churn Prediction: Does Technology Matter?, " World Academy of Science, Engineering and Technology,Vol. 2,No.4, Pp. 815-821

[15] Kiranjot Kaur and Sheveta Vashisht, " A Novel Approach for Providing the Customer Churn Prediction Model using Enhanced SVMs Technique in Cloud Computing, " Vol. 114 , No. 7 Pp.1-7

[16] K. Rababah, H. Mohd, and H. Ibrahim (2011), "Customer Relationship Management (CRM) Processes from Theory to Practice: The Pre-implementation Plan of CRM System, " International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 1, No. 1, Pp. 22-27

[17] Manjari Anand, Zubair Khan, Ravi S. Shukla(2013), "Customer Relationship Management using Adaptive Resonance theory", International journal of Computer Applications, Vol. 76, No. 6, Pp. 43-47

[18] Manjeet kaur, Dr.kawaljeet Singh and Dr.Neeraj Sharma(2013), " Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers," International Journal on Recent and Innovation Trends in Computing and Communication ,Vol.1, No. 9 , Pp. 720 – 725

[19] Mariusz Lapczynski(2014) , "Hybrid C&Rt-Logit Models In Churn Analysis, " Folia Oeconomica Stetinensia, Pp.37-52

[20] Md. Faisal Kabir, Alamgir Hossain, Keshav Dahal(2011), " Enhanced Classification Accuracy on Naïve Bayes Data Mining Models", International Journal of Computer Applications, Vol. 28, No. 3,Pp. 9-16.