Evaluating Prediction of Customer Churn Behavior Based On Artificial Bee Colony Algorithm

Riddhima Rikhi Sharma, Rajan Sachdeva

M.Tech Scholar, Department of Computer Science & Engineering,
Guru Gobind singh college of modern technology
kharar, Punjab
2riddhima.sharma@gmail.com
Associate Professor, Department of Computer Science & Engineering,
Guru Gobind singh college of modern technology
kharar, Punjab
mp13rajan@gmail.com

Abstract— As in the competitive environment, it becomes necessary to focus on retaining churn customers as well as attracting new customers. Various algorithms of Data Mining have been used for making distinguish between customers into loyal and churn. This paper represents the ant bee colony algorithm which has been used to get accurate results. The use of ABC with two best values local best and global best makes it more effective to obtain the more pleasing and comfortable results. PSO is used to search the best solution with two best values named pbst and gbst by using iteration with initial velocity and positions.

Keywords— Customers Relationship management, Ant Colony Bee algorithm, SVM

I. 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.
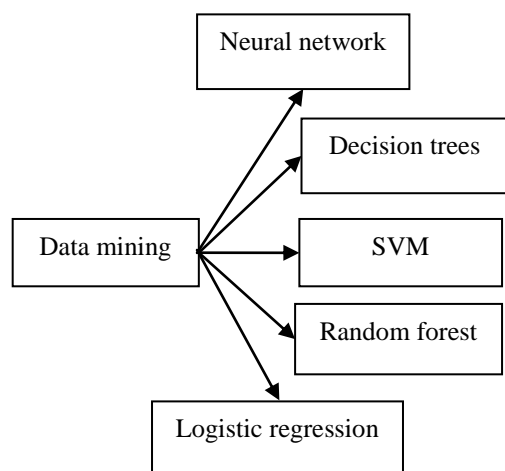

Fig1. Data Mining Technique

1.1 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).

1. 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.

2. 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.

3. 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.

II. ARTIFICIAL BEE COLONY ALGORITHM

ABC algorithm is on the premise of the smart technique for the honey bees getting alongside each other. Most loved honey bees being social creepy crawlies split their perform among themselves: Used honey bees, Onlooker honey bees and Look Bees. Their measures are ordered specifically into four main times: Initialization point, used honey bee point, Onlooker honey bee point and Look honey bee stage. In introduction point, each utilized honey bee is furnished with different nourishment assets. In utilized honey bee point, each utilized honey bee figures the nectar measure of the foodstuff source associated with it and the separation of it from the hive. In pursuit honey bees point, the utilized honey bees whose nourishment source gets to be overlooked gets to be hunt honey bee. The primary component work of hunt honey bees is to discover new nourishment assets. Mostly of pc study and reason study, ABC is primarily utilized for answer of advancement issue. At the point when identified with improvement issue, the foodstuff potential outcomes would be the pair of shifted conceivable arrangements accessible.

III. RELATED WORKS

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. 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. 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. 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. 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. 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. 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. 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.

IV. GAPS IN LITERATURE

Following are the various gaps in earlier work.

1. The use of unsupervised filtering is ignored to reduce the effect of noise in the data.
2. The use of evolutionary optimization is also neglected by the majority of existing researchers.
3. The benefits of artificial bee colony like its speed and global best properties still need to be explored to predict churn users.

V. METHODOLOGY 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
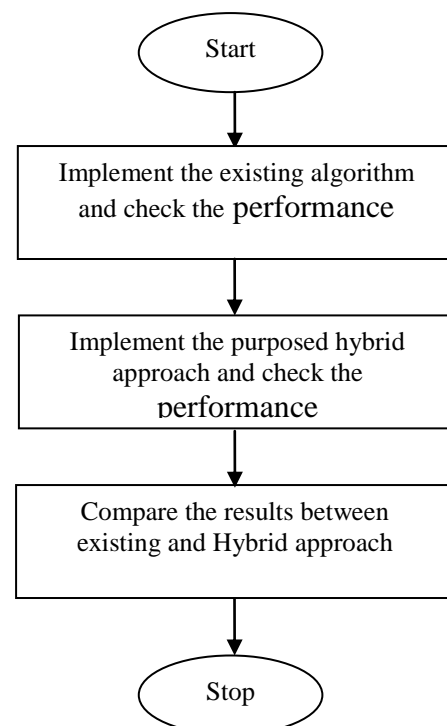Step 5: Stop the Algorithm.

Fig 2: Flowchart of the proposed technique

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 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 1, it has been clearly indicated that proposed ABC approach provides the better, results that is accuracy. 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 proposed algorithm.

Filtered Classifier	94.2394%	5.7606%	40.4962%
Bayes Network Classifier	94.3894%	5.6106%	42.9417%
Purposed Hybrid approach	97.2997%	2.7003%	17.9552%

Table 1

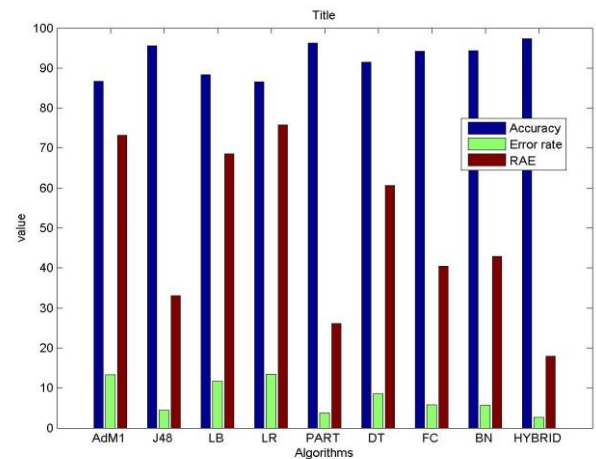| Algorithm Name	Accuracy	Error rate	RootAbsolute Error(RAE)
AdaBoostM1	86.6787%	13.3213%	73.1989%
random forest	95.5596%	4.4404%	33.0881%
SVM	88.3588%	11.6412%	68.59%
Logistic Regression	86.5887%	13.4113%	75.7602%
PART	96.2496%	3.7504%	26.0667%
Decision Table	91.4791%	8.5209%	60.6888%

Fig 3: Represents Accuracy, Error rate and RAE

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 2 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

Algorithm	RRSE	Coverage Cases	Mean rel. size

Name			
AdaBoostM1	85.2074%	99.4899%	91.2091%
random forest	58.505%	95.5596%	50%
SVM	81.0929%	98.7999%	76.2976%
Logistic Regression	87.1518%	99.0099%	81.0681%
PART	51.0711%	99.0099%	60.5911%
Decision Table	76.7574%	98.9449%	85.2835%
Filtered Classifier	63.656%	98.1998%	63.9664%
Bayes Network Classifier	59.7792%	99.4899%	68.5119%
Purposed Hybrid approach	42.3864%	98.9499%	54.4554%
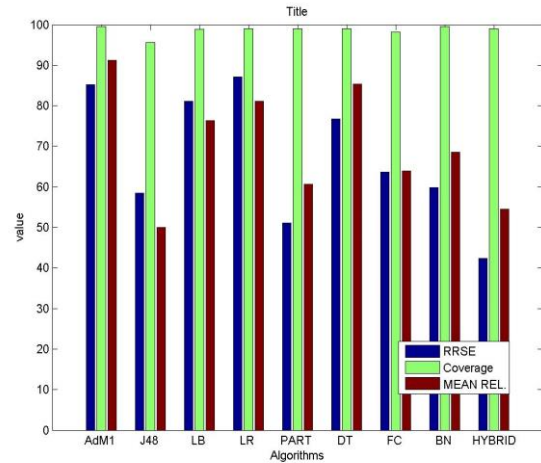
VI. CONCLUSION

This paper have represented ABC algorithm based on Data Mining and Hybrid approach of random forest and SVM provides better and accurate results when comparisons are made on various algorithms. The use of ABC with two best values local best and global best makes it more effective to obtain the more pleasing and comfortable results. PSO is used to search the best solution with two best values named pbst and gbst by using iteration with initial velocity and positions PSO is used to search the best solution with two best values named pbst and gbst by using iteration with initial velocity and positions PSO is used to search the best solution with two best values named pbst and gbst by using iteration with initial velocity and positions ABC is used to search the best solution by using iteration with initial velocity and positions. Different kinds of parameters used in order to enhance the purposed algorithm. The purposed Hybrid approach is implemented in MATLAB with statistics toolbox. The experiment results reveal that purposed Hybrid approach outperforms better results when comparisons are made over eight different 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. 508-523Burez 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 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. 497-514Burez 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. 497-514Chih-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. 1-10Chris Rygielski , Jyun-Cheng Wang   and   David C. Yen(2002) , " Data mining techniques for customer

relationship management, " Technology in Society, Vol. 24, Pp. 483–502Dr. M.Balasubramanian, M.Selvarani(2014), "Churn prediction in mobile telecom system", International Journal of Scientific and Research Publications", Vol. 4, No. 4, Pp.1-5Dr. 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. 96-101H. 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. 599-606Emtiyaz, 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. 229-236E. 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.693-697Georges 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.55-62Indranil 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 KongJ. Basiri, F. Taghiyareh, B. Moshiri(2010), "A Hybrid Approach to Predict Churn", Proceedings of Asia-Pacific Services Computing ," Conference IEEE, Pp.485-491J. Burez and D. Van den Poel (2009), "Handling class imbalance in customer churns prediction", Expert Systems with Applications, Vol.36,

Pp. 4626–4636J. 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. 815-821Kiranjot 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.1-7K. 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 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. 43-47Manjeet 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 − 725Mariusz Lapczynski(2014) , "Hybrid C&Rt-Logit Models In Churn Analysis, " Folia Oeconomica Stetinensia, Pp.37-52Md. 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.