

A Novel Approach for Providing the Customer Churn Prediction Model using Enhanced Boosted Trees Technique in Cloud Computing

Kiranjot Kaur
M.Tech Scholar
Lovely Professional University
Phagwara, Punjab, India
Pin no. 144402

Sheveta Vashisht
Assistant Professor
Lovely Professional University
Phagwara, Punjab, India
Pin no. 144402

ABSTRACT

Organizations earn huge amount of money by providing the different services to their customers. In today's world of competition, organizations need to focus on customer relationship management. Retaining the existing customers is as much important as attracting the new customers for an organization. For this purpose, organizations use data mining techniques for segmenting the churn customers and loyal customers so that special offers can be provided to churn customers to retain them as customers are the most valuable asset for organizations. The aim of this paper is to provide a customer churn prediction model using a standard CRISP-DM methodology based on RFM and Boosted Trees Technique. To enhance the performance of the technique, hybrid approach for building classifiers is used. There is also a comparison between the performances of both techniques. Results show that enhanced boosted trees technique performs better than existing boosted tree technique. Proposed approach is then implemented on the cloud environment to provide the cloud facilities for mining the data.

General Terms

Data Mining, Accuracy, Client Churn Attrition, Comparison, Algorithms.

Keywords

Data Mining, Cloud Computing, Boosted Trees Technique, Customer Churn Prediction Model, Retail Store, RFM Model, CRISP-DM Methodology.

1. INTRODUCTION

Customer is like the God for businesses. To be in the competition, organizations always want to retain their profitable customers and also attract more and more new customers. But retaining the old customers is much cheaper than attracting the new customers. Though intensified competition and increasingly complex behavior, it has become increasingly difficult for marketing practitioners to maintain customer loyalty. So they need to focus on marketing strategies using data mining techniques for discovering and retaining the churn customers.

With the technologies such as data mining, data ware housing, businesses have more concern about customer relationship management. There are many data mining techniques which are used to detect customer behavior patterns to formulate marketing, sales, and customer support strategies. Organizations need to understand the customer behavior to improve their marketing strategies. Data about the spending behavior of customers is stored in the database about which a consumer is not even aware and there are industries which use this data to get the interesting patterns about customer

purchasing behavior and take decision to beat the competition in the market [1].

Berry and Linoff (1997) said that data mining techniques can be used to retain the loyal customers, look out the right prospects, identify new markets for products and services, and recognize cross-selling opportunities on and off the web. Data mining techniques are effective tool for analyzing consumer behaviors. There are seven powerful techniques which are useful for this purpose [2] such as Cluster Detection, Memory-Based Reasoning, Market Basket Analysis, Generic Algorithms, Link Analysis, Decision Trees, and Neural Nets.

2. DATA MINING

According to Jiawei Han (2006), [3] "Data Mining is extracting or mining knowledge from large amounts of data". Data mining is the process of getting knowledge from the data to make decisions for the growth of an organization. It discovers patterns which are present in large datasets. This knowledge is stored in data warehouses, the Web, large databases, other massive information repositories, or data streams. Using this knowledge, managers can take the good decision within the time for the growth of an organization. Analysts use computers to analyze the data which will provide high level of speed, efficiency, and accuracy. Data mining tools perform data analysis and find interesting and useful data patterns which are helpful for making business strategies, fraud detection, and scientific and medical research.

Data Mining is used in areas such as machine learning, database and data warehouse technology, statistics, information retrieval, neural networks, pattern recognition, knowledge-based systems, high-performance computing, artificial intelligence, image and signal processing, spatial or temporal data analysis, and data visualization [3]. The discovered knowledge can be applied to process control, query processing, decision making, and information management. So it is very useful tool for the growth of an organization. It is an iterative process [4] which consists of six phases such as problem definition, data exploration, data preparation, modeling, evaluation, and deployment.

3. DATA MINING WITH CLOUD COMPUTING

As it is discussed earlier that data mining is the process of analyzing the data to get correlations of factors. It analysis of huge amount of data to fetch new interesting patterns such as groups of data records based on cluster analysis, dependencies based on association rule mining, and unusual records based on anomaly detection. The resultant patterns are the summary of the input data that can be easily understand and further analysis to get the good decision for the benefit of an organization. Traditionally, only large organizations were using data mining due to high cost involve in this process but

now due to the new generation of data mining tools and technologies, small organizations can also use this. But still it is costly process because every day users generates Tera-Bytes of data for a single application with a high velocity. Data mining can be used with cloud computing. Buyya have defined the term cloud computing as [5] “Cloud is a parallel and distributed computing system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements (SLA) established through negotiation between the service provider and consumers.” Cloud computing provides efficient and cheap way to store and analyze huge amount of input data and give the useful patterns in less time. Before analyze something there is a need to understand the data mining strategy based on cloud computing from the theoretical as well as practical point of view [6]. Cloud computing provides a variety of on demand computational resources which provides the parallel processing on low cost. This helps the data mining algorithm to work faster to analyze large amount of input data and give the useful patterns in less time. Using cloud computing for data mining is very helpful for analyzing the data as cloud provides many features such as speed, scalability of resources, parallel processing, just pay the used resources, choose another technology at any time to further work, 24/7 availability of services, device and location independent, provides reliability and security etc.

4. CUSTOMER CHURN

Due to grown and development of businesses, there are multiple varieties of a single product are available in the market. Now customers have many choices to buy a product. They can choose a product on the basis of price, quality, trend, design, brand, etc. When customer of one supplier leaves that supplier and goes to another supplies by choice then this phenomenon is called customer churn. There are two types of churning customers [7]

4.1 Non-Voluntary Churners

This category includes those customers who have broken the relationship with the organization.

4.2 Voluntary Churners

This category is divided into two subcategories such as

4.2.1 Incidental Churn

This type of customers unintentionally leaves the organization because of reasons such as due to migration, change in the location of their occupation.

4.2.2 Deliberate Churn

This type of customers decides to end the relationship with an organization to seek the better services and cheaper good quality products from another organization.

For customer churn management, organization identifies the group of deliberate churners to any proactive approach which is used to retain those customers [8]. For identifying churners, one thing should keep in mind that organization needs to identify the most profitable or high valued customers who want to end the relationship with the organization. For this RFM model is used. It have recency feature (time period from last transaction of client with organization), purchase frequency (number of purchases done by customer in the given time period), and monetary value (amount of money spend by client in a given time period) [12]. These RFM features are good predictors for customer lifetime value.

5. ADABOOST.M1 TREE

AdaBoost.M1 is the version of AdaBoost Tree Technique. Full form of AdaBoost is “Adaptive Boosting” [3]. It is most popular boosting algorithm in machine learning proposed by Yoav Freund and Robert Schapire [8]. In AdaBoost.M1, initialize the equal weight to each tuple in dataset ‘D’. Then generate ‘k’ classifiers in each round. In round i, choose a set of tuples as D_i and train a classifier model using tuples in D_i . Then calculate the error of classifier using mathematical equation.

$$Error(M_i) = \sum_j^d W_j * err(X_j)$$

Here $err(X_j)$ is the misclassification error of tuple X_j . If tuple was misclassified then assign $err(X_j)$ is 1 otherwise it is 0. If the error of classifier is equal to 0 or greater than equal to 0.5 then abort the classifier. Otherwise, update the weight of correctly classified tuples by

$$Weight\ of\ the\ tuple * \frac{Error(M_i)}{1 - Error(M_i)}$$

So that weight of misclassified tuples is increased weight of correctly classified tuples are decreased. Normalize the weight of each tuple in set by

$$W\ of\ the\ tuple * \frac{Sum\ of\ the\ old\ weights}{Sum\ of\ the\ new\ weights}$$

Then assign weight zero to each class and calculate the weight of each derived classifier’s vote by

$$W_i = \log \frac{1 - Error(M_i)}{Error(M_i)}$$

Then add above calculated weight in the class’s weight for classifier predicted class. The class whose total net weight is higher that is the winner class and returned as class prediction for a tuple. This technique is also represented in flowchart (see Figure 1).

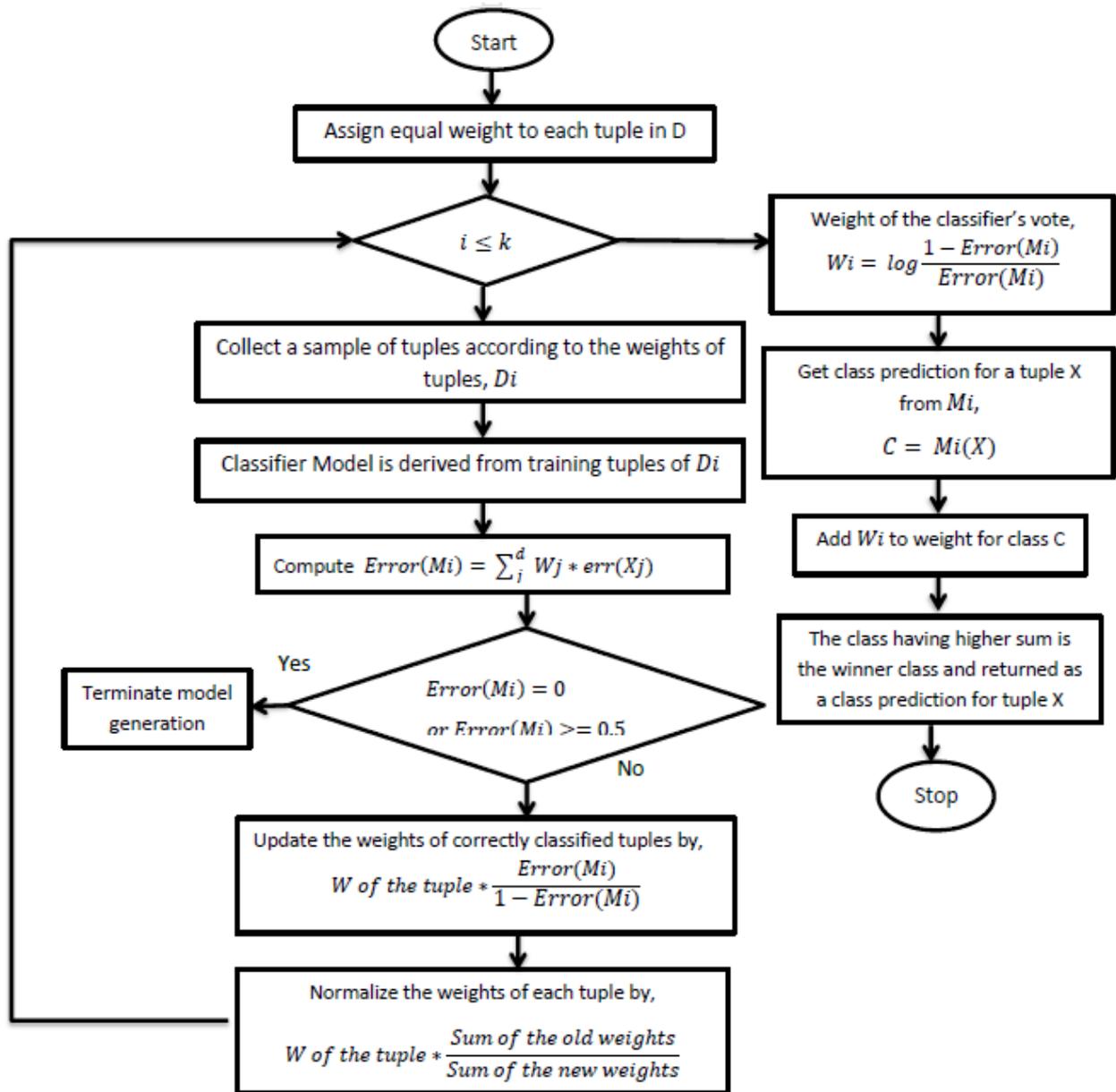


Figure 1: Flow Diagram of Existing AdaBoost.M1 Technique

6. DECISIONSTUMP

It is a one-level decision tree with single node which decides how to classify input instances on the basis of single feature [13]. One-level decision tree means it has one root node which is immediately connected to the terminal nodes or leaves or decision labels. It makes prediction on the basis of single input feature's value that's why it is also called as 1-rules [14]. For nominal features, DecisionStump algorithm build a stump which gives one leaf for each possible value of a feature and then assign the class label to that feature value. It is used as a component (called as weak learner or base learners) in the boosting technique.

7. AODE

It is Averaged one-dependence estimators (AODE) a classification learning technique. It supports incremental

learning in which classifier can be efficiently updated. There is an ensemble of classifiers are generated and then the attrition is produced by averaging the probability estimates of all the classifiers [15]. For each classifier a single attribute is chosen as the parent attribute of all other attributes. Most important feature is that it does not simply predicting a class, it predicts class probabilities. This classification technique also handles the situation if some data are missing in the given dataset.

8. PROPOSED WORK

In model generation process of proposed approach, single classifier of the boosted tree is replaced by hybrid approach of two classification learning techniques. One is AdaBoost.M1 having DecisionStump classifier and another is AODE. In classification process, it is proposed that if error rate of

classifier model is from 0.3 to 0.4 then it means the classifier is more accurate than those classifiers whose error rate is from 0.1 to 0.2. So a small value that is 0.1 is added to the weight of the vote for more accurate classifiers. In this way the weightage of the votes of these correct classifiers is increased as compared to others. It provides more accuracy for predicting a class for instance in the dataset. In this research work, proposed approach is implemented in cloud environment so that features of cloud computing can easily make mining the huge amount of data.

8.1 PROPOSED ENHANCED ADABOOST.M1 ALGORITHM

8.1.1 Model Generation

- Assign equal weight to each training tuple in dataset D
- For i iterations:
 - Collect a sample of tuples according to the weights of tuples, D_i
 - Hybrid Classifier Model is derived using DecisionStump and AODE from training tuples of D_i
 - Compute error of model on weighted dataset D_i :
$$Error(M_i) = \sum_j^d W_j * err(X_j)$$
 - If Error = 0 or Error ≥ 0.5 :
 - Terminate model generation
 - For each tuple in D_i that was correctly classified:

$$W \text{ of the tuple} * \frac{Error(M_i)}{1 - Error(M_i)}$$

- Normalize the weights of each tuple by,

$$Weight \text{ of the tuple} * \frac{Sum \text{ of the old weights}}{Sum \text{ of the new weights}}$$

8.1.2 Classification

Assign weight = 0 to all classes

For each i classifier model:

For the class this model predicts, add weight of the classifier's vote to this class's weight

If $Error(M_i) \geq 0.1$ and $Error(M_i) \leq 0.2$

$$Weight \text{ of classifier's vote} = \log \frac{1 - Error(M_i)}{Error(M_i)}$$

Else if $Error(M_i) \geq 0.3$ and $Error(M_i) \leq 0.4$

$$Weight \text{ of classifier's vote} = \log \frac{1 - Error(M_i)}{Error(M_i)} + 0.1$$

- Return class with highest weight

The proposed enhanced technique is also representation in flowchart (see Figure 2)

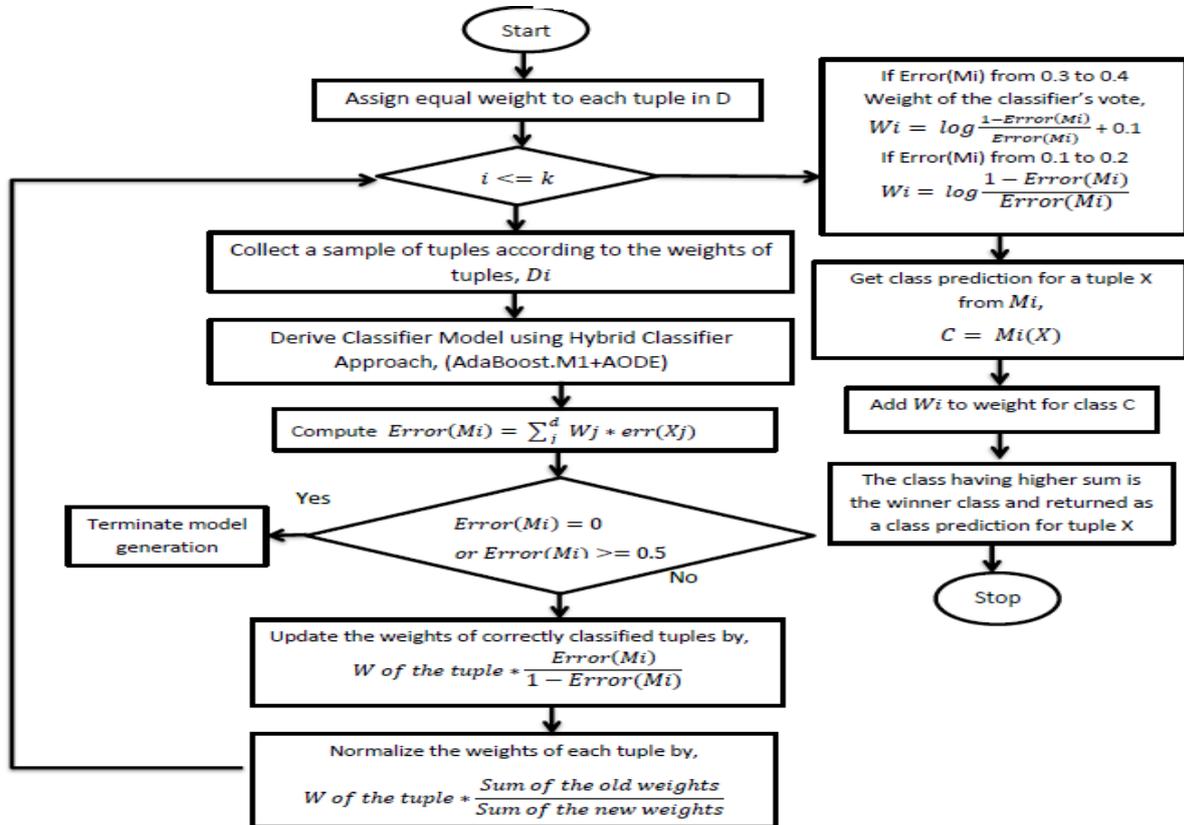


Figure 2: Flow Diagram of Proposed Approach

9. METHODOLOGY

Standard CRISP-DM algorithm is used to mine the data to get useful knowledge. This algorithm consists of six phases as business understanding, data understanding, data preparation, modeling, evaluation and deployment [11]. There are five phases for customer behavior analysis such as selecting samples of data for analyzing, select churn factors or variables, process data, build customer churn prediction model, and evaluate trained model. In the first step, samples are selected for better understanding and predicting customer attrition [9]. In second step, necessary features for predicting customer attrition are collected. In third step, data preprocessing is performed. In data preprocessing, data may be collected from different sources, having some missing values in it so there is a need to integrate, clean and transform the data before supplied to the data mining algorithm. As the quality of output is directly affected by the quality of input values. In fourth and fifth step, there is training and evaluation of predictive models to choose the one best model for predicting customer attrition [10]. For customer churn prediction model, five major steps are there (see Figure 3).

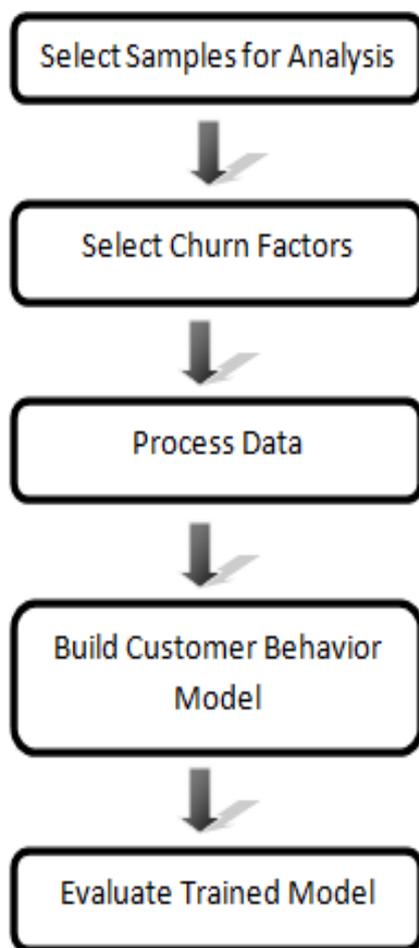


Figure 3: Customer Behavior Analysis Model

9.1 Data Sample

Collect the dataset for customers to analyze the customer purchasing behavior to predict that whether customer is churn or loyal. This research work used dataset for analyzing the customers of a retail store.

9.2 Churn Factors

There are many predictive factors which gives more specific information regarding the behavior of customers [11], [9]. Top most predictive factors are Purchase Frequency, Length of Relationship (LOR), Relative Purchase Frequency, Average Inter-purchase Time, Variance of Inter-purchase Time, Last Inter-purchase Time Difference, Monetary value, Relative Monetary value, Weight of Sold Items, Customer Behavior in Product Category, Number of Categories, Customer Type, etc.

9.3 Process Data

Data preparation, data cleaning are the most important tasks before doing analysis because incorrect or wrong input will lead to wrong output and from this wrong decision will be concluded. After doing all these tasks dataset processing is initialize.

9.4 Classification Technique

There are some active and inactive customers which might be loyal. In some situation customers who have less recency it doesn't mean that it is more loyal. So for correctly identifying churners this research work used enhanced boosted trees data mining technique for segmentation the churn customers from the loyal customers. It uses hybrid approach for classification as mentioned in above section VII.

9.5 Evaluation

The proposed approach and existing boosted trees technique both are implemented in WEKA tool. The proposed approach shows better results than existing technique. (see Table 1, Figure 4 and Figure 5). Then proposed approach is implemented in CloudSim tool using single virtual machine to check that whether it supports cloud environment or not. It also works same there. Following table shows the difference between the proposed approach and existing boosted trees technique.

Table 1: Comparison results between existing technique and proposed technique

Model	Correctly Classified Instances	Precision	Sensitivity
Existing AdaBoost.M1	75.5%	78.6%	94.9%
Proposed Approach	88.7%	89.1%	88.8%

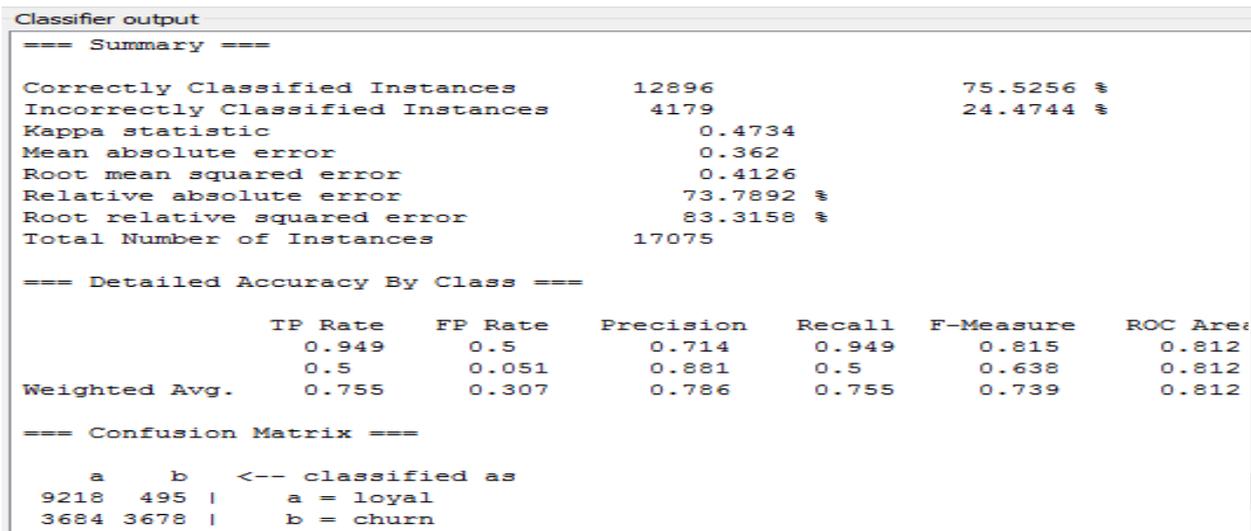


Figure 4: Results of Existing AdaBoost.M1

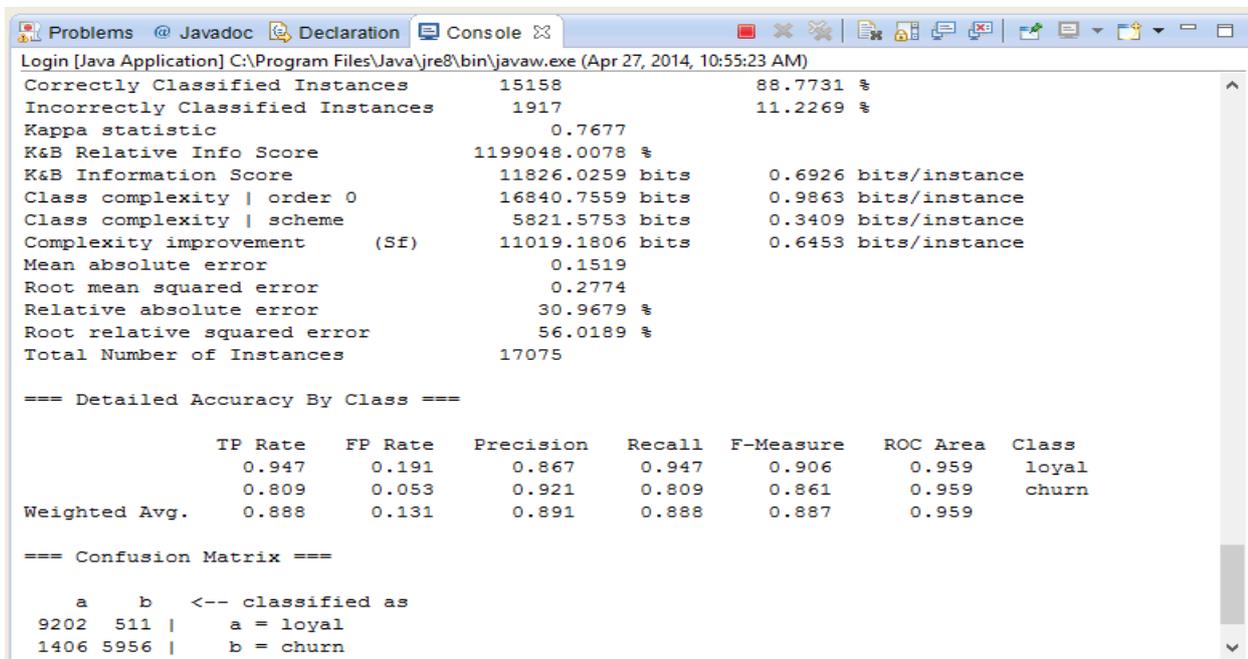


Figure 5: Results of Proposed Approach

10. CONCLUSION

In today's world of competition, organizations need to be more customers-oriented instead of product-oriented and always work for Customer Relationship Management. Customers are the most valuable asset for an organization. So organization must need to retain their profitable customers. For this purpose data mining techniques are used to identify the deliberate churn customers. This paper proposed a new approach to enhance the performance of existing boosted trees technique. This approach is implemented on a cloud environment where many features of cloud computing such as low cost, dynamic provisioning of virtualized computers, pay for used resources, and security etc., are there for efficiently mining huge amount of data. The proposed approach is used for client churn attrition. So that proactive measures could be taken by company for churn prevention. Results shows that proposed technique perform with more prediction accuracy than existing technique. For the future work, the proposed technique can be design in such a way that it work parallel for

multiple machines in a cloud environment to reduce the processing time.

11. REFERENCES

- [1] Wang, C. and Wang, Y. 2012. Discovering Consumer's Behavior Changes Based on Purchase Sequences. In Proceedings of the 9th IEEE-International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012), 642-645
- [2] Berry, M. J. and Linoff, G. 1997. Data Mining Techniques: For Marketing, Sales, and Customer Support, John Wiley & Sons, New York, NY, USA.
- [3] Han, J. and Kamber, M. 2006. Data Mining: Concepts and Techniques, Morgan Kaufmann, India
- [4] http://publib.boulder.ibm.com/infocenter/db2luw/v9r5/index.jsp?topic=%2Fcom.ibm.im.easy.doc%2Fc_dm_proc_ess.html

- [5] Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., and Brandic, I. 2009. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility, *Future Generation Computer Systems*, 25, 599-616.
- [6] Lingjuan, L. and Zhang, M. 2011. The Strategy of Mining Association Rule Based on Cloud Computing. In *Proceedings of the IEEE-International Conference on Business Computing and Global Informatization*, 475 – 478
- [7] Hadden, J., Tiwari, A., Roy, R., and Ruta, D. 2007. “Computer assisted customer churn management: State-of-the-art and future trends”, *Computers and Operations Research*, Vol. 34, Issue 10, 2902-2917.
- [8] <http://en.wikipedia.org/wiki/AdaBoost>
- [9] Kaur, K. and Vashisht, S. “Enhanced Boosted Tree Technique for Customer Churn Prediction Model”, *IOSR Journal of Engineering (IOSRJEN)*, 2014, Vol. 4, Issue 3, 41-45
- [10] Basiri, J., Taghiyareh, F., and Moshiri, B. 2010. A Hybrid Approach to Predict Churn. In *Proceedings of IEEE Asia-Pacific Services Computing Conference*, Hangzhou, China, 485-491.
- [11] Nabavi, S. and Jafari, S. 2013. Providing a Customer Churn Prediction Model using Random Forest Technique. In *proceedings of 5th IEEE-Conference on Information and Knowledge Technology (IKT)*, 202-207
- [12] Buckinx, W. and Van Den Poel, D. “Customer Base Analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting”. *European Journal of Operational Research*, 2005, Vol. 164, 252-268.
- [13] http://en.wikipedia.org/wiki/Decision_stump
- [14] <http://nltk.googlecode.com/svn/trunk/doc/book/ch06.html>
- [15] <http://en.wikipedia.org/wiki/AODE>