



An Evolutionary Approach of Machine Learning for Monitoring Churn Prediction of Broadband Customer

Imran Pathan¹, Nadeem Ahmed Kanasro², Farhan Bashir Shaikh³, Mujeeb U Rehman Maree⁴,
Aftab A. Chandio⁵

¹ M.Phil. Scholar, IMCS University of Sindh Jamshoro Sindh Pakistan Email: Imran8net@gmail.com

² SUCL and IMCS University of Sindh, Jamshoro, Sindh Pakistan Email: nadeem.kanasro@usindh.edu.pk

³ SUCD University of Sindh, Jamshoro, Sindh Pakistan Email: farhan.shaikh@usindh.edu.pk

⁴ IMCS University of Sindh, Jamshoro, Sindh Pakistan Email: mujeeb@usindh.edu.pk

⁵ IMCS University of Sindh, Jamshoro, Sindh Pakistan Email: chandio.aftab@usindh.edu.pk

ABSTRACT

Era of industrial Revaluation technologies and multiple end users emerging flavours of services can change the mind of end user at any time. This situation increases the demands of technological flavours.

Extreme traffic of streaming, high quality of multimedia systems, flavours of services, on demand access can makes the mind diversion of humans, situation increases market competition of telecom to rendering the demanding services and that's make ruin. This situation increase such type of customer those who stop the business with entire company and deal with another company which gives its demanding services. The leading situation increases the user churn or unsatisfied users. [3, 4, 5, 6]. Presented work is an Evolutionary Approach of Machine Learning for Monitoring Churn Prediction of Broadband Customer "MCPOBBC" for telecom industries.

The work ascertain the Characterization of Real world Broadband Users Data (N1=3400 and N2=6400 datasets and N3=4600 from History Log of Broadband user of Internet Service Provider: ISP) according churn prediction model to elevate the telecom industries which are providing DSL, ADSL broadband services to their customer.

The whole work has been accomplished by the implementation

C5-J48 decision tree algorithm on Weka 3.8 machine learning simulations and resulting predicted churn validated on CHAID, CART, and k-mean for Clustering algorithms

Key words: CART, CHAID, CHURN, DSL

1. INTRODUCTION

The Evolutionary approach of Machine Learning for churn prediction is a preferred churn prediction mechanism that illustrated on real-world ISP provided datasets for churn

prediction. This framework ascertain the simulation of trained dataset, on deployment of C5-J48 decision tree algorithm on Weka 3.8 simulations as a Machine Learning tool, and same also tested and validated on CHAID, CART, and k-Mean for Clustering algorithms for knowledge discovering of broadband users churn and their possible retention from predicted churn.

The study helps to detect the customers who are likely to cancel a subscription by setting up the criteria that in which condition the broadband user may not or may leave the broadband services, the MCPOBBC will have settled criteria find the user churn.

In order to retain the predicted churn user, the developed system corroborates broadband DSL registered/contractual user/customer recommended corporate/business ISP.

Intended to provide competitive service to high-speed communication technology from private to public networks, Statistics variations are not more important than ever, operational support services, for service providers and more importantly at customer's end. High speed and accuracy are customer's satisfaction and key to customer satisfaction [1, 2, 7, 8, 9, 13]. Small and big companies are good service for ideal companies and customer satisfaction is very important.

- Analysis intended Consumer market orientation; this makes campaigns to maintain retention, with support cloud base software service (SaaS), a third-party provider's hosting application.
- The domain of knowledge has been taken with the possibility of customer satisfaction with full broadband services for Digital Subscribers Line (clustered networks).
- To ascertain for monitoring and analytical judgement of customer behaviours regarding to avail BB services or let them leave means satisfaction and dissatisfaction.
- Also propose the campaign for "MCPOBBC" of predicted churn and constraint management that will be a continuous process.
- Monitoring is to be done regularly, not only once.

1.1 CHURN PREDICTION FOR UNSATISFIED USER

A client or any firm of who can stop from an unanticipated business or enters a specific company is anxious to leave the customers as described as a Churn Prediction, is important for dealings as it helps you detect customers who are expected to scratch a subscription, artefact or service depicted as in FIGURE 1. Churn prediction can be tremendously useful for customer withholding and by predicting in advance customers that are at risk of send-off.

1.2 CUSTOMER RETENTION WITH BUSINESS ISP'S

The Enterprise, Internet Service Provider (ISP) provides the services to its valuable customers such that end-user and also a corporate customer for example Bank, Mart, Shop, Enterprise and Corporate ISP with fixed paid bandwidth for 100MB, 200MB e.t. per month.

2. PROBLEM STATEMENT

The research will address the problem is to ascertain the telecom companies. The Customer Graph has come up in such a way that it should be actively started to resolve the situation. Energy companies, creating these channels, responded to the situation by which they want to attract new customers and more retention products that are actively presented in maintaining campaigns, the churn rate calculation is also one of most legible situation in telecom companies[1,4,10,11,12].

Research on the reasons for and consequences of this shift has focused on objective measures Churn prediction, but there has been little work required subjective to economy of software industries, continuously to rendering the services for intended situations and its solutions with the help of Monitoring Churn prediction a software industry can resolve out the customer problems and their retention.

- **ENTROPY FOR CHURN RATE**

Total Number of Broadband Users in a time subtracted from total broadband users next time will become the ratio of total broadband churn users.

Finding the churn level..

$$\text{Total users in a time} = U_{(t)}$$

$$\text{Total users in a next time} = U_{(t+1)}$$

$$\text{Total churn user } CU = U_{(t)} - U_{(t+1)}$$

$$\text{Churn ratio} = CR = \frac{CU}{U_{(t)}}$$

If value become 5% or below means not high risk.

If the average become 10% and above, this intended condition for high risk

If ratio becomes negative it means that similar ratio for gain and loss.

To accomplish the task for $\text{Churn ratio} = CR = \frac{CU}{U_{(t)}}$ a complete model has been developed illustrated in Figure 1 which furnish the dataset according to simulation of algorithm.

- Setting up churn prediction criteria results for model implementation

$$\text{Total users in a time} = U_{(t)}$$

$$\text{Total users in a next time} = U_{(t+1)}$$

$$\text{Total churn user } CU = U_{(t)} - U_{(t+1)}$$

- Apply and validate best/suitable Classification data mining model for churn prediction and retention
- Evaluated the performance delves by Decision tree Algorithms and churn Models
- Validation data accuracy according to prescribed entropy and attribute relationship monitoring for registered user's activities from logs, for CART and C5-J48 Decision Tree Algorithm.
- C5-J48 predicts and classifies the set of churn subscriber user to recommend alternative business/corporate ISPs to the churn user for retention in case dissatisfaction from broadband services.
- Validate C5-J48 the resultant trained datasets to be simulated into data mining process and validating same with "Weka data mining simulation and ML software".

3. RESEARCH OBJECTIVES METHODOLOGIES

Data gathering of real-world "HOME USERS" broadband subscribers line (DSL/ADSL) i.e. for corporate telecom industry of ISP.

3.1 DEPLOYMENT OF WEB BASED CHURN PREDICTION MODEL TO SIMULATE TRAINED DATASETS ON C5J48, CART ALGORITHMS

- Monitor the user activities by user logs
- Using C5-J48 decision tree predict a set of churn broadband subscriber user.
- Determine the C5-J48 and CHAID compares the churn predicted on basis result variance that could be finalized for churn.
- Investigate the evaluated result of C5-J48 and CHAID validate the results on Weka 3.8 Simulation for more result narrow down.
- Explore by simulation on model implementation and evaluation and validating of trained datasets.
- Recommendation process for Business ISPs to the churn user for retention.

3.1.1 Web based monitor constructed PREDICTION MODEL

Web base constructed Churn prediction model in Figure 1 shows the flow of data of developed system to monitor the churn prediction through over the web services.

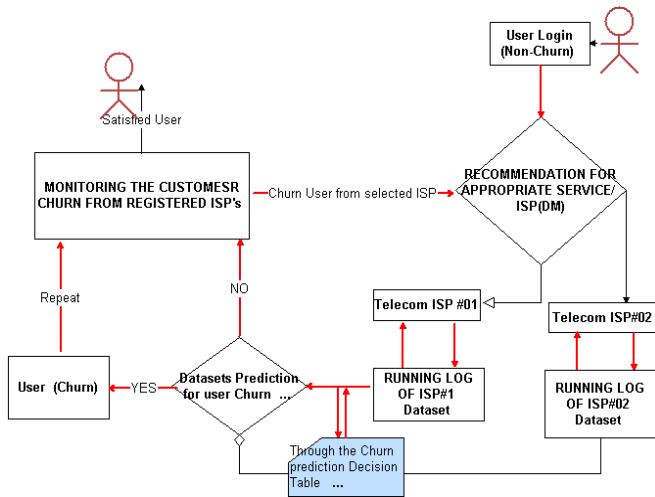


Figure 1: Preferred Churn Prediction Model for Broadband Corporate DSL Users

3.2 METHODOLOGY FOR CHURN PREDICTION MONITORING

Analysis and development of churn prediction entropy as an initial stage of deployment.

Datasets Characterization of real-world broadband users of telecom, discovering training datasets according to the churn prediction model in which work done by pre-processing and model selection, Oracle SQL, SPSS as analysis data mining tool, Classification k-mean clustering Algorithm to validate the criteria and threshold level for churn prediction[1,2, 13].

Building a model of classes from a set of records that contain class labels

Evaluating and testing Model Results to deploy the churn prediction model by using CHAID and CART Classification and Decision Tree algorithms using classes from a set of records that contain class labels

Model Deployment and Data mining simulations of data mining Algorithms and visual Simulations on Weka 3.8 Data Mining tool.

Evaluated and validated the accuracy of CHAID implemented results on Weka

Generation of classification and decision tree for data accuracy using the C5-J48 algorithm on weka to validate the accuracy of attributes relationship by regression Models

3.3 DATASETS EVALUATION FROM CHARACTERIZATION AND ANALYTICAL PICTORIAL PRESENTATION OF CHURN PREDICTION THRESHOLD LEVEL

The developed Model of “MCPOBBC” have utilized Real-World datasets of broadband ISP’s Records about to

4600 history of users log and 3400 running log of user as shown in Figure 2 while data simulation for high bandwidth and low bandwidth fluctuation of DSL is given in Figure 3 and Figure 4 shows aggregated group of initial cluster to validate pairs.

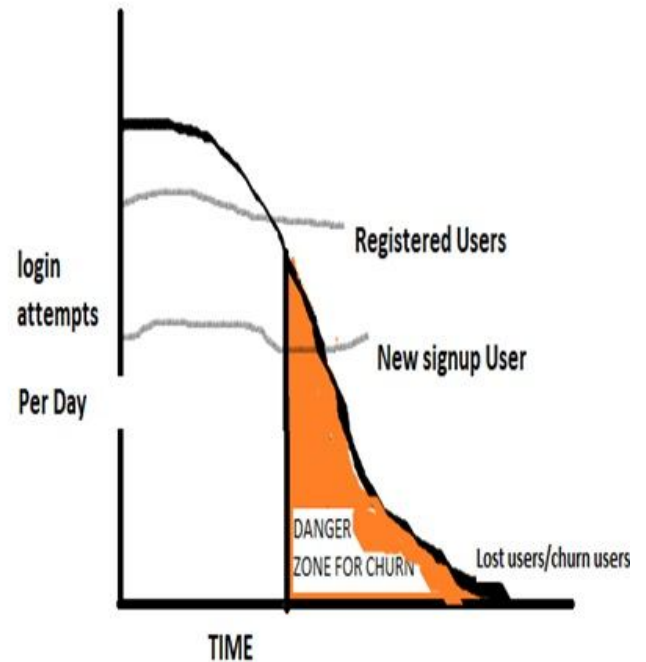


Figure 2: Indication of Churn Prediction Monitoring Graph for Registered and not to be Registered Broadband Users for MCPOBBC System, as $Churn\ ratio = CR = \frac{CU}{U_{(t)}}$

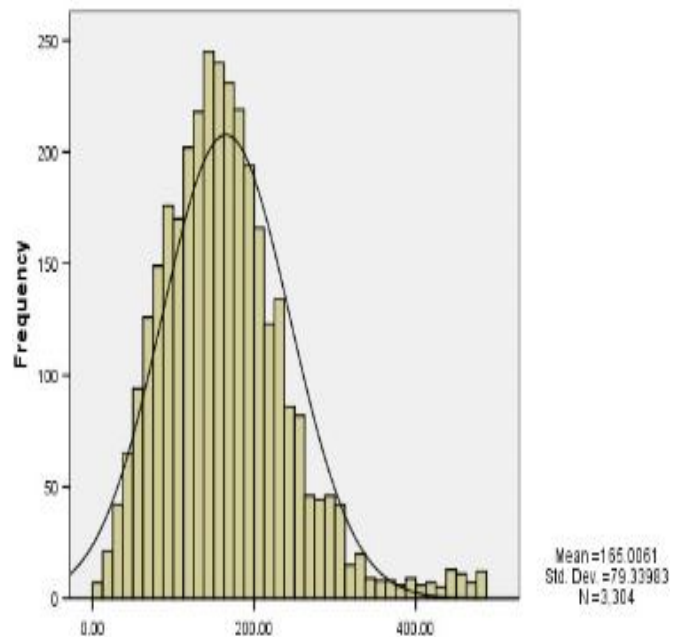


Figure 3: Data Simulation for High Bandwidth and Low Bandwidth Fluctuation of DSL Users.

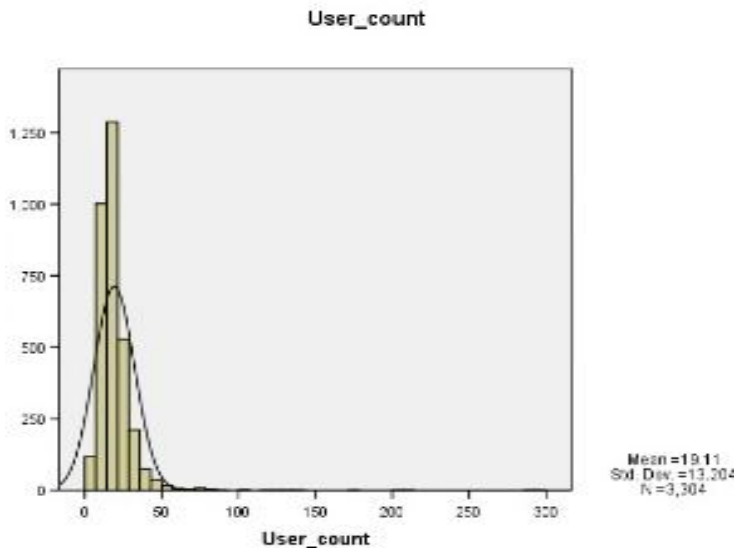


Figure 4: Aggregated Group of Initial Cluster to validate the pair's e.t.

Figure 5 shows Evaluation report of Broadband user's maximum and minimum bandwidth for three days session to ascertain the Logins attempts in every next 30 Mints intervals

- Maximum Logins attempts observed >1800 users
- Minimum Logins attempts observed <200 users
- Minimum range of logins attempts users are indicated the Alarm of User-churn

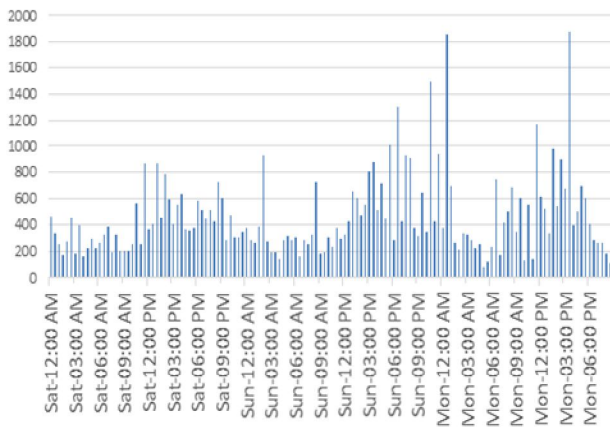


Figure 5 show time interval analysis for churn indicators Characterization and their Analytical Evaluation for churn threshold level by logins.

4. PROB RESULTS OUTPUT DEPLOYED BY CART DECISION TREE ALGORITHM TO “PERTAIN THE CRITERIA FOR PREDICTED CHURN

Figure 6 shows the CART decision tree model is to get the criteria using function Criteria in Between (Lower session to higher session). In lower session at Node 0 10.707 mean is found while at higher session at Node 1 and at Node 2 it varies.

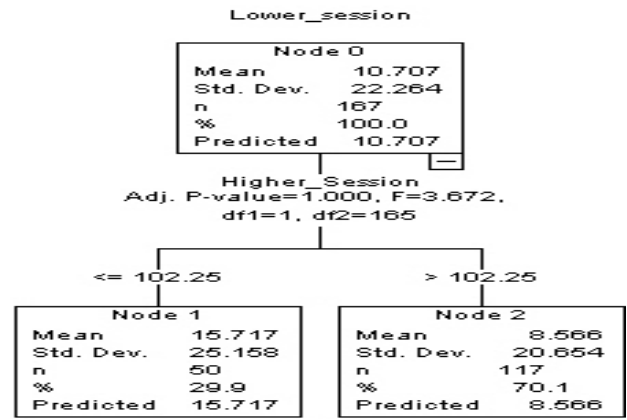


Figure 6: Shows the CART decision tree model is to get the criteria using function Criteria in Between (Lower session to higher session).

Making Similarities in bandwidth usage by Minimum and Maximum Using K-mean Clustering Algorithm Results. Data mining techniques has been applied between to variables Maximum Bandwidth and Minimum Bandwidth through each clustered we have find range between Max and Min Data usage.

Initial Cluster Centers to validate the pairs e.t. (Min (Max_Bandwith), Min (Min_Bandwidth)) ~Churn parameters

Table 1 show the bifurcation of cluster maximum and minimum bandwidth of different Churn position while table 2 shows the bifurcation of cluster max and min (bandwidth) of broadband users.

Table 1: Shows The Bifurcation Of Cluster Maximum And Minimum Bandwidth Broadband Users If Minimum Is .02 And Maximum Is 1.98 So Far Identified Churn Cluster.

	Cluster				
	1(Churn)	2	3	4	5(Churn)
Maximum Bandwidth h	1.98	480.02	480	480	151.58
Minimum Bandwidth h	.02	.02	480	225	151.58

Table 2: Shows the bifurcation of cluster maximum and minimum bandwidth broadband users.

USERS	1	209.000
	2	2445.000
	3	11.000
	4	37.000
	5	298.000
VALID		3000.000
MISSING		.000

4.1 A COMPREHENSIVE ANALYSIS DECISION TREE ALSO VALIDATING PREDICTED CHURNS

Figure 7 particularizes Clustered CHAID (Decision Tree) Deployment of 3400 and 4600 rec for Churn prediction to Evaluations the predicted values of churn accuracy Validation.

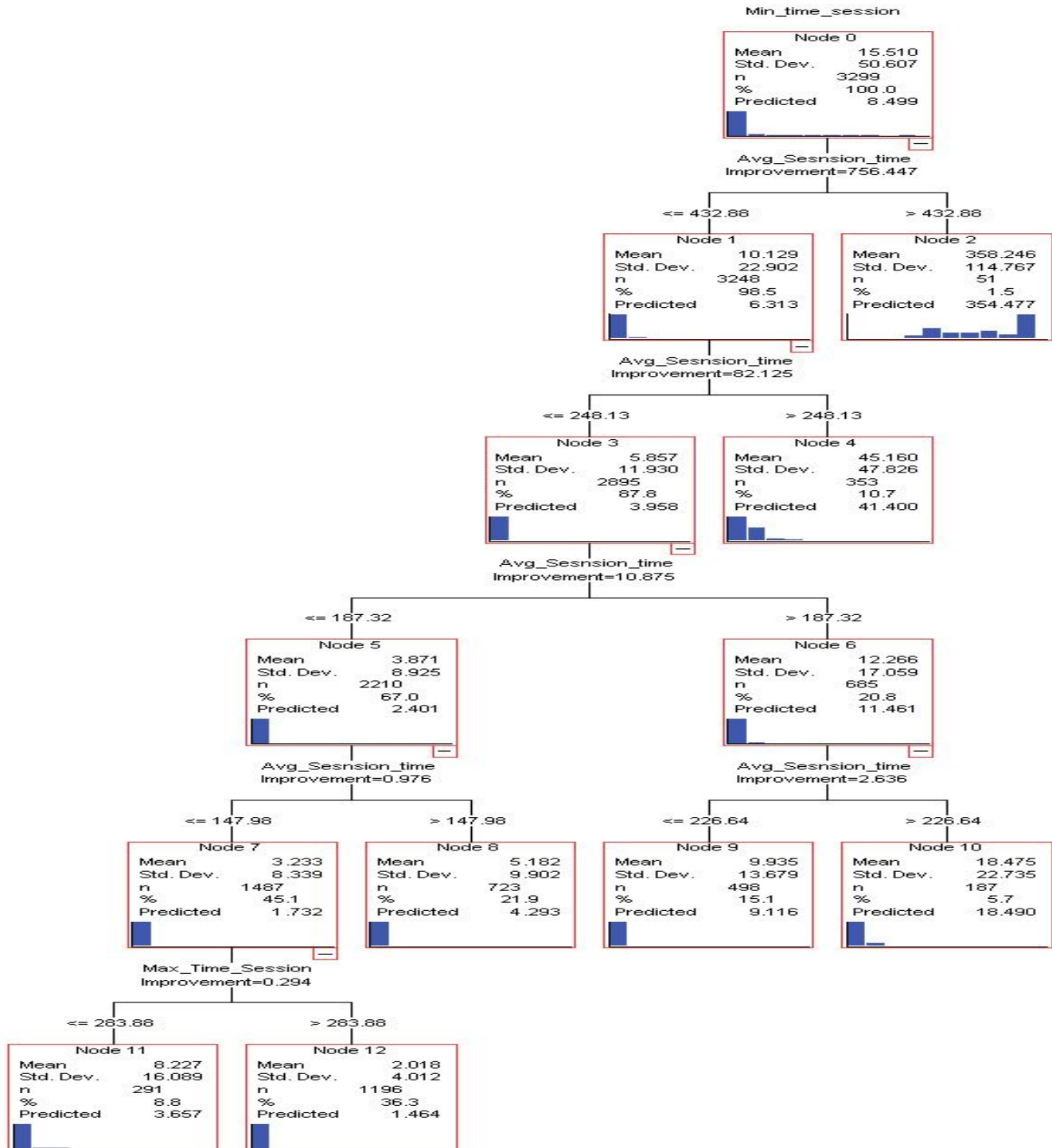


Figure 7: Clustered CHAID (Decision Tree) Deployment of 3400 and 4600 rec for Churn prediction to Evaluations the predicted values of churn accuracy Validation.

4.2 A COMPREHENSIVE ANALYSIS FOR CRITERIA’S AND VALIDATED PREDICTED CHURNS BY REGRESSION

The holistic comparison all model particularly shows the Criteria (Higher_Session, Lower_session) as given in Figure 8 approximately 100~150 is achievable targeted data analysis for churn.

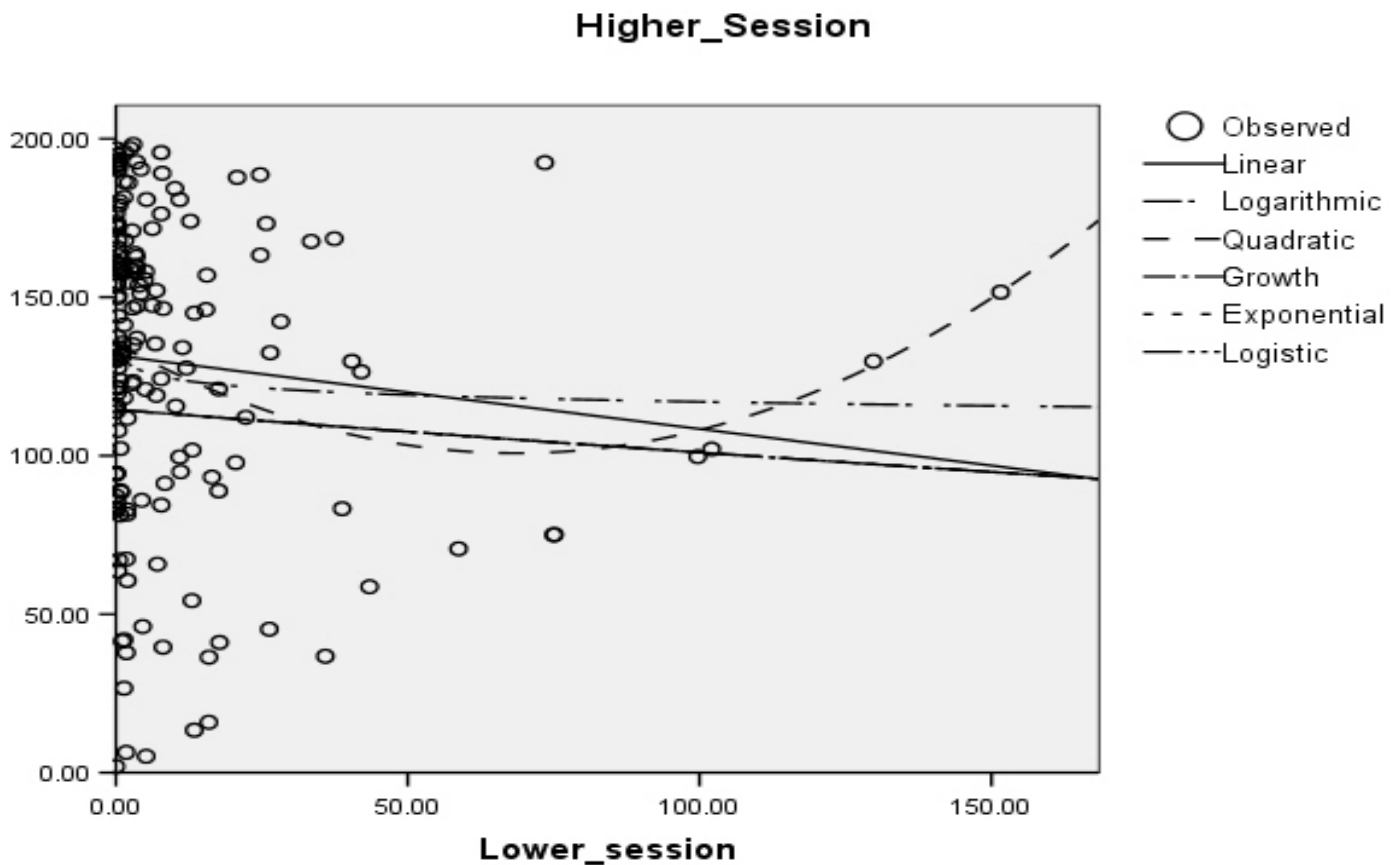


Figure 8: Holistic Evaluation of Churn Prediction over the Decision Trees projections CART, CHAID and C5-J48 with Outcome Validation on Regression Analysis.

4.3 RETENTION FOR PREDICTED CHURN

- Retention policy of real-world ISP’s still not have automated provisioning in scenario of “Predicted churn users”.
- The designed system is a preferred churn prediction system will have provisioning for predicted churn users will be recommended for corporate (Business ISP’s).
- With the help of WEKA3.8 it has been classified by dataset simulation predicted churn are valid or not.

5. CONCLUSION

Churn prediction is one of the machine learning and databases of the most popular applications in the Management of Customer Relations (MCR) and in the field of marketing. A churn simply is a user or a purchaser, who doesn't want to keep using the company's product or service [1].

Whereas business companies related to the field of telephone and internet subscription face the highest potential of churn at the end of their subscribers. This professional work is known through the project which must be carried out within the software industry. Web-based third-party promotion has proved through the study analysis and network monitor for predictions. Eventually, two types of prediction models are designed which predict a wide variety of broadband subscribers including whole data and other types of standards based on certain the appropriate criteria. for Instance : Range= Min (Max_Session), Min((Min_Session)) =Min (Login_range) i.e. For prediction values, all the quality CHAID and CART algorithms are used to stream users based on the selected variables of database, and all examinations are done, such as Weka 3.8 Java Runtime Time C5-J48 rated tree results for engine, authentication and accuracy, resulting in the results of mining process of decision-making, Kmean clustering, timeline forecasting and multi-regulation data,

namely, in all this chapter. All done in addition, a concept has stated that computer programmers can create simple web scripts through data mining process so that they can play the role of third-party promotion. Note: While using of business ISP's (subsidiary ISP's) for retention, in this case broadband users will remain with under the umbrella of Enterprise ISP. This research project is taught that professional work in the software industry for predictions is necessary. It is ready to develop a web-based third-party promotion / customer support through the study of evolutionary approach of machine learning for churn prediction studies, for network monitoring for prediction Character was predicted "implemented by CHAID and CART algorithms to be implemented by using the" Real World "Broadband users' data. Moreover the authentication and accuracy of the results by " Weka "on the Java Runtime Time Engine" The rating data processing process is described as "C5.0-J48". Successful development and deployment will be satisfied to Upset or disappointed user and satisfy the service providers

The holistic comparison all model particularly shows the Criteria= (Higher_Session, Lower_session) approximately 100~150 is achievable targeted data analysis for churn.

ACKNOWLEDGEMENT

With the Praise to Almighty Allah, this research has been made possible by the great support and interest of Guide, Co-Guide, classmates and friends.

REFERENCES

1. Nath, Shyam V., and Ravi S. Behara. "Customer churn analysis in the wireless industry: A data mining approach." Proceedings-annual meeting of the decision sciences institute. Vol. 561. 2003.
2. Hung, Shin-Yuan, David C. Yen, and Hsiu-Yu Wang. "Applying data mining to telecom churn management." Expert Systems with Applications 31.3 (2006): 515-524.
3. Kim, YongSeog. "Toward a successful CRM: variable selection, sampling, and ensemble." Decision Support Systems 41.2 (2006): 542-553.
4. Zhao, Y., Li, B., Li, X., Liu, W. and Ren, S. (2005). Customer churn prediction using improved one-class support vector machine. Int. Conf. Advanced Data Mining and Applications, Springer, pp. 300–306.
5. Xia, G.-E. and Jin, W.-D. (2008). Model of customer churn prediction on support vector machine. Syst. Eng. Theory Pract., 28: 71–77.
6. Kisioglu, Pinar, and Y. Ilker Topcu. "Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey." Expert Systems with Applications 38.6 (2011): 7151-7157.
7. Chu, Bong-Horng, Ming-Shian Tsai, and Cheng-Seen Ho. "Toward a hybrid data mining model for customer retention." KnowledgeBased Systems 20.8 (2007): 703-718.
8. Tsai, Chih-Fong, and Yu-Hsin Lu. "Customer churn prediction by hybrid neural networks." Expert Systems with Applications 36.10 (2009): 12547-12553.
9. Verbeke, Wouter, et al. "Building comprehensible customer churn prediction models with advanced rule induction techniques." Expert systems with applications 38.3 (2011): 2354-2364
10. Qi, Jiayin, et al. "ADTreesLogit model for customer churn prediction." Annals of Operations Research 168.1 (2009): 247.
11. Karahoca, Adem, and Dilek Karahoca. "GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system." Expert Systems with Applications 38.3 (2011): 1814-1822
12. Keramati, Abbas, and Seyed MS Ardabili. "Churn analysis for an Iranian mobile operator." Telecommunications Policy 35.4 (2011): 344-356.
13. Lee, Hyeseon, et al. "Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model." Decision Support Systems 52.1 (2011): 207-216.