

Comparative Analysis for Churn Prediction Model in Telecom Industry

S.P.Pund^{1*}, Dr. S .N. Deshmukh²

¹Department of CS & IT Dr.Babasaheb Ambedkar Marathwda University Aurangabad, India

² Department of CS & IT Dr.Babasaheb Ambedkar Marathwda University Aurangabad, India

*Corresponding Author: shwetasp.pund@gmail.com

Available online at: www.ijcseonline.org

Accepted: 19/May/2018, Published: 31/May/2018

Abstract- Churn prediction is the demanding field today and to stand in the market place or to capture market and for profit maximization churn prediction is very useful. Churn defines the customers switching another company, this is because the market strategy is rapidly changing. Other competitive companies give something new to the customers with low cost. Hence customers change their service provider very fast. Whereas retaining old customers is easy than gaining new customers. Retaining the customers by giving more offers is easy. The goal of this paper is to predict customer churn which will help to retain them. Many organizations feel the data base containing old customer information effectively predicts or generates the outputs. Data mining plays a vital role in churn prediction. Comparative study of the various classification algorithm can be done to give more accurate results.

Keywords— Churn, weka, decision tree, classification, telecommunication

I. INTRODUCTION

The Telecommunication sector involves many companies that make communication possible on a global scale, whether it is through the phone or the Internet. These companies created the infrastructure that allows data to be sent anywhere in the world. Much of the industry focus is on providing faster data services, especially in the area of high-resolution video. Essentially, the driving forces are toward quicker and clearer services, increased connectivity, and multi-application usage. This is the main reason behind switching to another company, which reduces the profit. Churning process plays an important role for identification of these patterns. Many organizations have large volume of data to handle. Large business applications have huge amount of data to handle. The aspect of churn prediction is profit maximization and avoid revenue loss. Customer churn prediction has emerged as an indispensable part of telecom sectors strategic decision making and planning process. Customer retention is one of the main objectives of CRM (Customer Relationship Management). Basically data depends on three types, customer data, call data and network data. This data base contains customer related information like name, address, gender, marital status, income, service plan, payment history etc. Call data contains call information or call details for every call like call duration, date and time of call etc. Whereas network data are very complex including thousands of inter-related components. It comprises handling network problems, errors, timestamps etc. Today's market

focus on maintaining customer loyalty or customer satisfaction which benefits in customer life time value whereas previous market strategy involved sales related issues for gaining new customers. Here gaining new customers is not a crucial issue but to satisfy existing customers and retaining them is the goal. Churn prediction focuses active customers and inactive customers. Basically churn involves two categories, 1] Involuntary churn 2] Voluntary churn. Involuntary churn involve fraud customers who churns for various reasons like pending payments, other crimes or those who does not use the service. These type of churners does not affect the company as much. Voluntary customers are the customers who decides to churn because company does not satisfy their needs or improper service. In this paper churn management is performed with data mining tool WEKA. Weka is a data mining tool that uses various machine learning algorithms. It is platform independent software, open source and free. Weka provides many algorithms to perform tasks like data pre-processing, classification, clustering, association, visualization etc.

II. LITERATURE REVIEW

In Now a days it is more demanding for churn management to analyse data in various aspects for telecom market. There are many fields that helps us to identify the customer who is going to churn. There is as much work in churn management. [1] Worked on profit maximization metric by applying classification algorithm on churn prediction models. They focused on cost benefit analysis for classification with binary

outcome, evaluation gives accuracy, sensitivity, and specificity for the confusion matrix. ROC curve is used for graphical representation performed on a threshold value t . They mainly focused on H measure which results in finding misclassification cost. MP (Maximum profit), EMP (Expected Maximum Profit) and H measure (Misclassification cost) are identified by various classifiers involving decision tree approaches, SVM Based methods, statistical classifiers etc. Performance is tested on 8 different telecom data sets. [13] Used Data visualization tools, Based on Tenure, Based on tech support, SVM (Support vector regression) with R, backward logistic regression. [2] Has tested decision tree algorithms to measure classification performances. Decision tree algorithms provides optimization computational efficiency as well as provides classification accuracy. The data sets were tested using the J48 decision tree-inducing algorithm.

This section, the author describes the previous research works in the form of title, problem statement, objectives, not repeat the information discussed in Introduction.

III. METHODOLOGY

Details below shows phases involved in fig: 1 prediction of churn prediction model.

IV. RESULTS AND DISCUSSION

A. DATA AQUISION:

A large amount data can be collected from a data warehouse. This raw data may contain missing values, redundant values, errors, incorrect values etc. To get correct results, data must be cleaned and errorless.

B. DATA PREPROCESSING:

It involves data filtering process. Before data classification data should be pre-processed by applying various filters to the attributes to evaluate correct and incorrect no. of instances.

C. IMPLEMENTATION OF CLASSIFICATION ALGORITHMS

Classification is a task of analysing data correctly. Classifiers used to classify the raw data with the help of classification rules for various aspects. Classification for prediction can be depending on some measures such as accuracy, speed, robustness, scalability, interpretability etc. Traditional classification algorithms and soft computing algorithms comparison provides more accuracy in churning process. J48 decision tree classifier is used to build a predictive model to be mapped with a minimal number of decisions. And outcomes are compared and analysed with Naïve Bayes classifier which is easy and fast predicting class classifier. It need small training data set and handle categorical and numeric data efficiently.

Classification algorithms generate confusion matrix which can focus on true churners and false churners. With the help of two classifiers mentioned above accuracy is calculated estimated with the time taken to build the model. ROC curves help to understand a gap between sensitivity and specificity. [1]

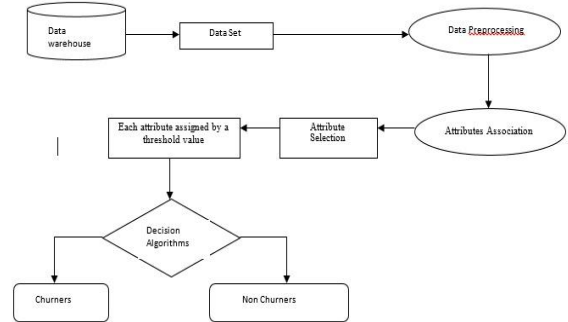


Figure1. Churn Prediction Framework

A raw data can be collected from a data ware house which arranged properly to form a dataset. This data set contain types of attributes with its feature. Data pre-processing cleans the obtain data by removing noise, incorporates missing value and apply appropriate filters to make attribute efficient to work and are useful for attribute association. Attribute selection accounts a target variable use to predict churn. Here each attribute is assigned by a thresholds value's' which help in the decision making process to predict the output "Would be churners" and "Non churners".

D. DATASET

Here data set used for telecom industry is collected from kaggle's.com. This data set contains 21 attributes of telecom field involve state, area code, phone number, total day calls, total day calls, customer service call, churn etc. Some attributes are nominal and some are numeric. Churn attribute is a Boolean attribute which contains true and false values to indicate churners and non-churners.

A. Figures and Tables

Table 1. Attribute Description of the Dataset

1	state	Nominal
2	account length	Numeric
3	area code	Nominal
4	phone number	Nominal
5	international plan	Nominal
6	voice mail plan	Nominal
7	number vmail messages	Numeric
8	total day minutes	Numeric
9	total day calls	Numeric
10	total day charge	Numeric

11	total eve minutes	Numeric
12	total eve calls	Numeric
13	total eve charge	Numeric
14	total night minutes	Numeric
15	total night calls	Numeric
16	total night charge	Numeric
17	total intl minutes	Numeric
18	total intl calls	Numeric
19	total intl charge	Numeric
20	customer service calls	Numeric
21	churn	Nominal

Figure 1 shows pre-processed data, the dataset contain numeric and nominal attributes. All numeric attributes are normalized and some filters are applied to nominal attributes to get better results. Pre-processing phase cleans the data and discretise the attributes.

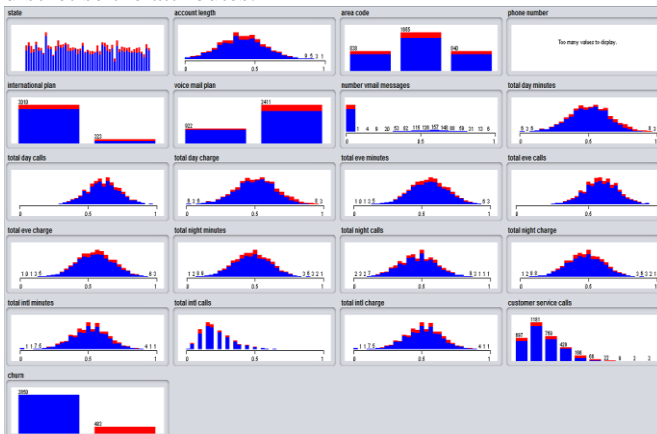


Figure: 1 Pre-processed data

Figure: 2 and Figure: 3 shows outcome of the classifiers after applying algorithms.

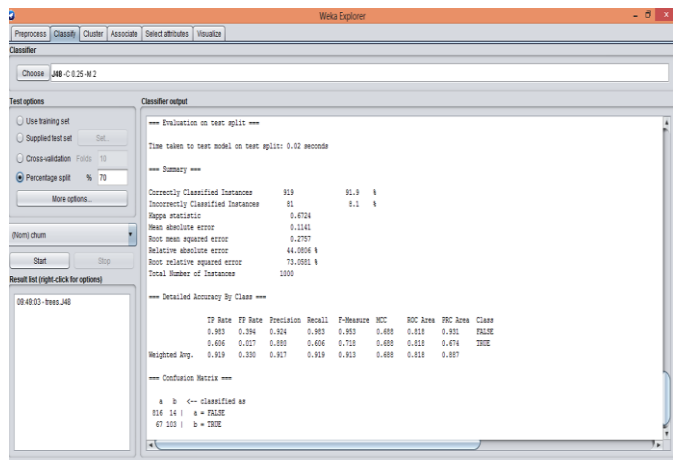


Figure: 2 Classification using decision Tree

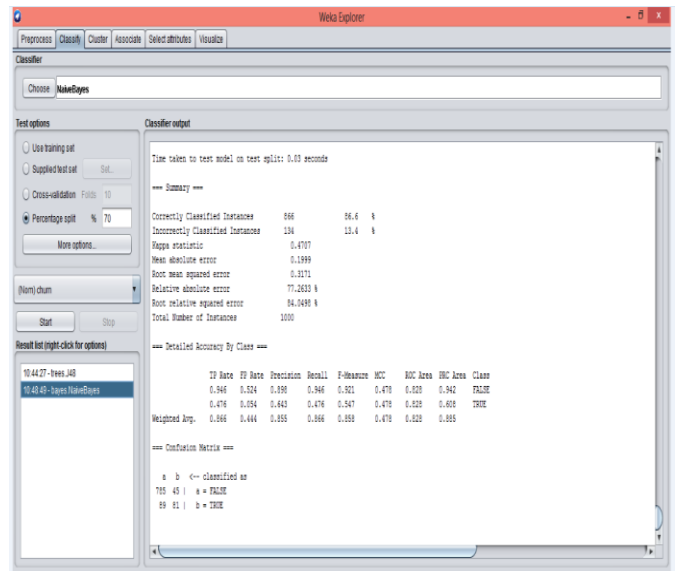


Fig 3. Classification using Naïve Bayes classifier

Figure 4 and 5 shows the decision tree created by j48 algorithm which classifies instances either true or false under other attribute considerations like total day minutes, customer service calls, internet plans etc.

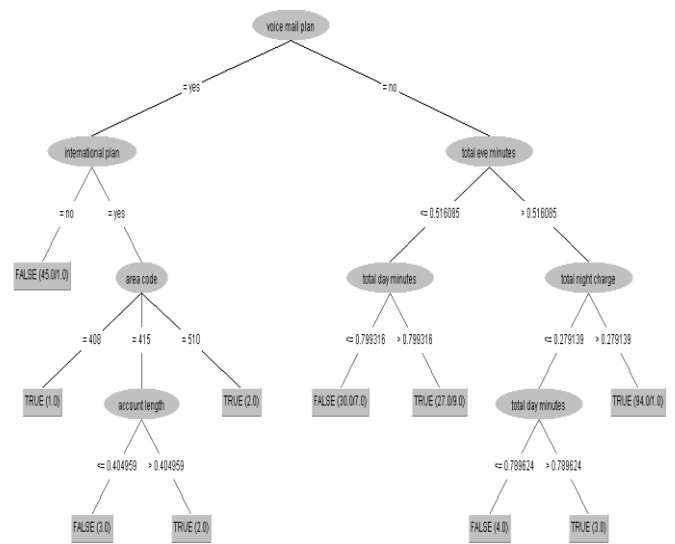


Figure4. Right Sub tree of Total Day Minutes

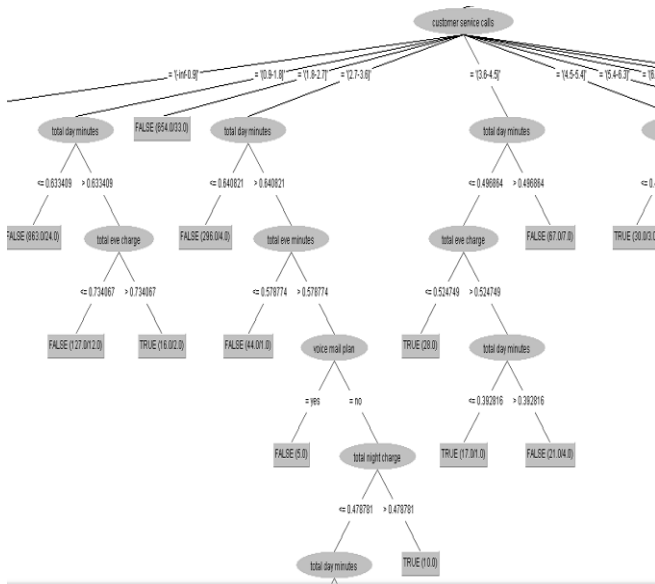


Figure 5. Left Sub tree of Total Day Minutes

Classification Algorithms	Class value	Accuracy	Error Rate	TP Rate	FP Rate	Precision	Recall	Time taken
Decision Tree (J48)	FALSE	93.0993	41.8884	0.981	0.366	0.94	0.981	0.09 Seconds
	TRUE			0.634	0.019	0.852	0.981	
Naive Bayes	FALSE	87.6088	0.199	0.94	0.499	0.917	0.94	0.03 Seconds
	TRUE			0.501	0.06	0.585	0.501	

Table 2. Results of the Classifiers

V. CONCLUSION

Telecommunication industry is a rapidly changing field hence it requires churn prediction effectively to prevent their customers from churn. By churn management one can help the business to avoid revenue loss. In this paper comparative analysis made, studied and analyzed the outputs of two classifiers J48 and Naïve Bayes. We found that accuracy of obtained by decision tree is much higher than naïve Bayes. Hence decision trees are more efficient technique. So customers churn to other service providers. This may lead to revenue loss. Customer churn prediction is the need of telecommunication industry to gain new customers as well as hold loyal customers. This can be archive by examining behavioral patterns of customer. This paper reviews various methods and algorithms for customer churn prediction which will give more accuracy. Comparison of traditional classifiers verses hybrid classifiers provides more efficiency for churning.

REFERENCES

- [1] Thomas Verbraken, Student Member, IEEE, Wouter Verbeke, and Bart Baesens “A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models” MAY 2013, pp 961
- [2] Sam Drazin and Matt Montag “Decision Tree Analysis using Weka”
- [3] Georges D. Olle Olle and Shuqin Cai “A Hybrid Churn Prediction Model in Mobile Telecommunication Industry” February 2014, PP 55-62
- [4] V. Umayaparvathi, K. Iyakutti “Applications of Data Mining Techniques in Telecom Churn Prediction” March 2012, PP Page 1065-1070
- [5] Ms Nisha Saini “Churn Prediction in Telecommunication Industry Using Decision Tree” June 2016
- [6] Dr. Mamta Madan, Dr. Meenu Dave, Vani Kapoor Nijhawan “Data Mining for Telecom Customer Churn Management” September 2015, pp. 813-817
- [7] Vani Kapoor Nijhawan, Mamta Madan, Meenu Dave ” The Analytical Comparison of ID3 and C4.5 using WEKA” June 2017
- [8] Kiran Dahiya, Surbhi Bhatia “Customer Churn Analysis in Telecom Industry” 2015 IEEE
- [9] Yong Liu, Yongrui Zhuang “Research Model of Churn Prediction Based on Customer Segmentation and Misclassification Cost in the Context of Big Data” June 2015
- [10] Clement Kirui, Li Hong, Wilson Cheruiyot and Hillary Kirui “Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining” March 2013
- [11] Ammar A.Q. Ahmed, Maheswari “Churn prediction on huge telecom data using hybrid firefly based classification” March 2017
- [12] Adnan Anjum, Adnan Zeb, Imran Uddin Afridi, Pir Masoom Shah, Saeeda Usman, “ Optimizing Coverage of Churn Prediction in Telecommunication Industry” 2017
- [13] HelenTreasaSebastian* and RupaliWagh “ Churn Analysis in Telecommunication using Logistic Regression” March 16, 2017, Pgs. 207-212
- [14] Gaganjot Kaur, Amit Chhabra “Improved J48 Classification Algorithm for the Prediction of Diabetes” Volume 98 – No.22, July 2014
- [15] P.Rutravigneshwaran” A Study of Intrusion Detection System using Efficient Data Mining Techniques” IJCSE Journal Vol.5 , Issue.6 , pp.5-8, Dec-2017
- [16] U.Kaur 1 , M. Mahajan2 , D. Singh” A Comparative Analysis of Trust Models in Cloud Computing”IJCSE Journal Vol.6,Issue.2,pp.19-23, Apr-2018

Authors Profile

MISS S.P.Pund Pursuing Master of philosophy in Computer Science from University of Dr.B.A.M.University Aurangabad INDIA in year 2018.She is currently working as Assistant Professor in Department of Computer Science, Deogiri College Aurangabad.

