

Opportunities of Machine Learning on Telecom Sector: A Case Study at BSNL

Dr.Sanjib Kumar Routray,

Faculty of Computer Science & IT, Telecom Training Center,
Bhubaneswar, Odisha, India.

ABSTRACT:

Telecom sector is the fastest growing service sector at present, where mobile service has an active role with data evolution among the people. Accordingly BSNL, a large public sector telecommunication company is offering a variety of services with latest technology while competing with no. of private operators. But customer satisfaction is poor. Hence BSNL seriously thinking regarding the importance of Customer Relationship Management (CRM) in retaining existing customer base and thinking of offering affordable quality of service through reorienting of the organization. In this context, the adoption of machine learning to CRM has also gained significance and providing many opportunities for retaining liable customers and future selection of customers according to behavior for optimizing the managerial efficiency of BSNL. This paper is aimed at exploring the determinants of CRM through machine learning techniques in data intensive computing environment, and presenting the idea that systems can learn from CRM data, identify patterns from selected attributes and make decision with minimal human intervention. Market segmentation and prediction of the customer's behavior (Churn, Choice of customer-PLAN) is determined by clustering and classifying the customers based on their attributes since grouping of customers and categorization of customers are two important factors of CRM on analysis of prediction. Machine learning. Strategies are proposed to be used for classifying the CRM datasets in BSNL under supervised learning for predicting the churn and profit of BSNL for retention of customers with making a project plan for effective decision by the marketing managers. Similarly, under unsupervised learning, market segmentation is proposed to promote the suitable customer-PLAN through clustering on same CRM datasets. Accordingly, this research has used WEKA and machine learning classification models (algorithms) for predicting output. From the experimental analysis basing upon the different performance measurement parameter, "Random Forest" is proposed for predicting churn prediction and same algorithm is for predicting selection of choice of future customer's mobile-PLAN for BSNL's Profit. For market segmentation, k-means clustering model (algorithm) is proposed to find the no. of best clustering area, where suitable affordable customer mobile-PLAN to be offered.

KEY WORDS: Customer Relationship Management, Telecommunication, Machine learning, BSNL

Date of Submission: 21-08-2021

Date of acceptance: 05-09-2021

I. INTRODUCTION

Among all the growing corporate sector of service organization, Telecommunication is the fastest service sector for utilizing the mobile with data service within the customers. The Government Telecom service sector is offering quality services through new latest technology for retaining existing customers with attracting new customers while competing with private Telecom service sectors to defend their positions in the market. But for achieving quality of service, building intimate long term relationship is highly essential. Accordingly, customer relationship management is the efficient strategy for achieving the company's goal. Telecommunications basics are integration of three systems that include service provisioning, activation and mediation along with CRM. Berry [1] recommended following strategies for practicing CRM.

- Core product/service development
- Customizing relationship
- Implementation of different attractive pricing strategy
- Augmenting VAS (Value Added Service) to customers
- Awareness on internal marketing, among employees.

Bharat Sanchar Nigam Limited (BSNL), State owned Company is a public sector Organization, providing a comprehensive range of telecom services in India [2]. But BSNL facing serious problems with severe service/product supply shortage problems and unable to provide the product/service in time during the competitive scenario of rapidly growing mobile services. Hence Churning started day by day. Further, due to Government policy, implementation of new customer attracted technology is delayed. Hence, in this research, it

is suggested in this study at BSNL, with reference to CRM to improve the customer loyalty in the light of severe competition and changing customer preference [3], examining the relationship quality in case of mobile services. Due to stiff competition by other private telecom service provider's market share is falling day by day BSNL has to concentrate on customer retention, providing quality of service, affordable customer-PLAN with maintaining customer satisfaction.

Most of the research confirmed that, need of machine learning technology is highly efficient for churn prediction has emerged as the most crucial business intelligence (BI) application and suitable customer-PLAN prediction through learning past data. Pattern of identification of possible churners and identification of customers who are about to churn is determined through commonly used supervised machine learning algorithm. Similarly prediction of suitable customer-PLAN to customers under segmented marketing area is determined through both supervised and unsupervised machine learning algorithm. Accordingly, various machine learning algorithms is proposed in this paper and it is validated on CDR (Call Detailed Record) of a BSNL. The performance of a classifier is measured by using the confusion matrix to prove the best model for churn prediction with the offering of suitable plan to retain customers by achieving high accuracy. Further for market segmentation and study the behavior of customers, we used k-means clustering algorithm by using attribute selection measures.

The remaining part of the paper provides related work, proposed methodology, experimental evaluation-results and final conclusion with future scope.

II. RELATED WORK

Many researchers defined CRM in many ways. Considering Telecom Industry, an increase in customer value, and the correct means by which to motivate valuable customers to remain loyal- indeed are important two factors of CRM. Accordingly "*A management approach which enables organizations to identify, attract, increase retention of profitable customers by managing relationships with them*" as Bras (2001) defines. The goal of Telecom Industry mainly attracts and retains the customers through their service quality and on customer satisfaction. Bhavne and Ashish [4] in their article entitled "Customer Satisfaction Measurement", where the success of the Telecom products or service in the market depends on customer's perceptions towards service quality and product quality. On 01.10.2000, Bharat Sanchar Nigam Limited (BSNL) was formed as a single large public sector Company is providing no. of Telecom service till date in competition with no. of private Telecom service providers. But the main issue is lack of issues on CRM. Then, according to article of V. Mohan Kumar [5], there is a big challenge on customization of CRM in BSNL focusing on perceptions of customers through

Marketing activities like Road Shows/MELA

By different type of Channel partners

Maintaining customer satisfaction [6]

Offering of Value Added Services and various attractive combo products [7]

Finally the effectiveness of CRM of BSNL has focused on the big management challenge in the new era of liberalization and Globalization with an understanding of customer's needs and wants on marketing aids for churn management.

A primer research area is for automatic learning through computer programs to recognize interesting, complex patterns and make intelligent decisions based on data sets [HMS01] called machine learning. The main two categories of machine learning like supervised learning, which supervises in the training data sets from labeled examples through classification models and unsupervised learning, where learning process is involved in the examples are not class labeled through clustering models. V. Jaaraj, J. Lavanya, J. Jagathesh Amairaj, M. Rajkumar (2013) analyzes the existing machine learning techniques, i.e. Clustering, Classification and described how these techniques were applied to improve the customer relationship with the company for generating considerable profit. Further the paper of Dr. T. Bhubaneswari (2013) is reviewed; where it is found that Companies need to understand their customers better and quality respond to their needs in proper time otherwise churning takes place. Accordingly Ayman Alawin, Mohammad Al. ma aithah, Al. Balqua (2014) stated to identify the profitable customers, churn one and develop two models, i.e. one is a physical model (OLAP) (Continuous mining of data sets wherever it resides) and the other is logical model, i.e. adoption of classification " on research point of view. From the customer churn prediction point of view in Telecom sector, supervised machine learning technique, i.e. classification is used to model the churn prediction with reference to Telecom sector [8] and in the computer science, application of Telecom big data is being researched from online classification from huge collected data pool [9]. Similarly for finding the targeted high value customers of Telecom sector, market segmentation is essential through clustering machine learning techniques [10] and for increasing the market share, company growth and profitability machine learning based CRM technique, i.e. classification is essential. From different research view, it is found that the classification model mainly used for churn analysis where Kiran Dahiya, Surbhi Bhatia [11] stated that "Customer Churn Analysis" is the essential factor of CRM in Telecom Industry as retention of customers depends on finding out of "Churn". Hence In this

paper author compared the efficiency and performance of Decision tree and Logistic regression techniques by using WEKA Data mining software to meet the need of surviving customers of Telecom sector in a competitive environment. Trilock Chand Sharma [12] made a Comparative Study on different classification algorithm, where the author studied the comparison of various classification techniques through WEKA tool with performance tools and quality measures. Different survey on classification methods using “WEKA”, where the author compared different classification techniques and investigated the performances of different classification methods, i.e. Bayes Network, Navie Bayes, Logistic rule, jrip, and J48. Out of which, the performance of Bayes Network is highest and fast to implement. According to Tina R. Patil, Mrs. S.S. Sherekar [13] performance analysis of Naïve Bayes and J48 Classification Algorithm for Data Classification, it is observed that the efficiency and accuracy of J48 are better than Naïve Bayes on experiments of accuracy, sensitivity, specificity of Naïve Bayes and J48 Algorithm. The finding of recent journal of Sanjeev Gour [14] on classifying the Incidence Rates of Cancers Using Data Mining Techniques (Perspective to Gas Leakage Accident of Bhopal City) is clustering techniques by K-Means. Hence Random Tree, Random forest (high value of accuracy prediction)& J48 Decision tree is good classifiers for decision making on service sectors in machine learning 48 Decision Classification Algorithm and K-Means Clustering Algorithm is proposed in the article as reviewed in this research paper for maximizing the organization’s satisfaction for increasing loyalty, retaining customer business over their lifetimes [15]

III. METHODOLOGY

3.1 Existing System

Existing system of BSNL performed churn prediction and profit prediction through various data mining, machine learning and statistical OLAP analysis. Those techniques enable and support Company solve the issues of CRM in some extent. But it is not appropriate for complex problem like on big data set environment. The present strategy will be more accurate for less no. of data but not on big data sets.

3.2 Proposed System

In this system, we use various advance tree algorithms like j48, Random Forest, Logistic Regression, OnetoR , Naive-Bayes to find accurate values for predicting the customer churn and profitable customer-PLAN. Maximum correct value is determined through the training and testing BSNL-CRM data sets for above prediction. Initially pre-processing of data sets and attribute selection is done through greedy heuristic methods. Finally, prediction and classification is done using different classifier Fig.1 shows methodology for churn prediction and profitable PLAN prediction of customers with strategy for market segmentation on the retention policy under different machine learning techniques. Testing methods used is cross validation of 10 folds for data partitioning. We observe the behavior and trend through analysis based on the results obtained through different algorithm of classification for predicting the churn and profitable PLAN for the customers for future decision of BSNL management.

K-means clustering algorithm is also used for creating retention policies by decision makers considering between of churn and no-churn customers with profit and no-profit customers of different groups.

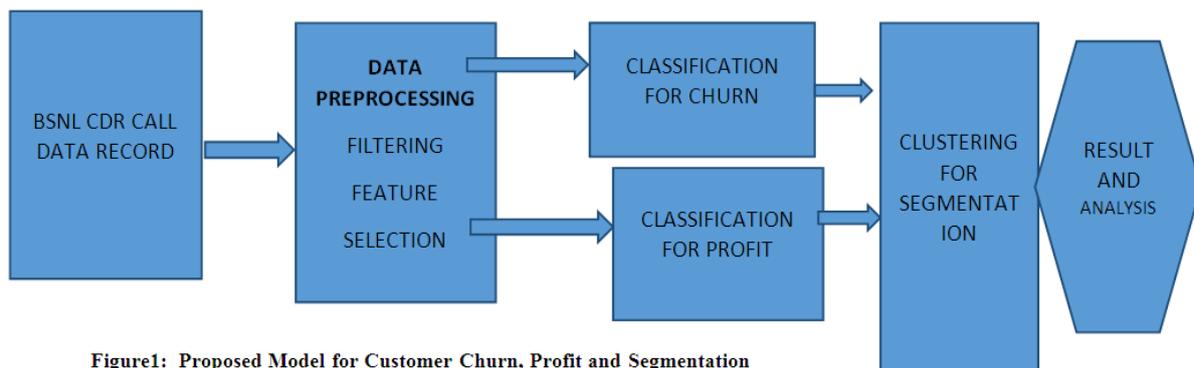


Figure1: Proposed Model for Customer Churn, Profit and Segmentation

IV. EXPERIMENTALEVALUATION AND RESULTS

We performed several experiments by using WEKA-3.9.5 on the proposed churn model and proposed profitable PLAN prediction model using machine learning classification model through no. of different algorithms. Finally experimented proposed clustering model to find out the retention strategy through k means algorithm. Experimental results evaluated through various performance factors of models confusion matrix.

4.1 DATASETS

In this paper, two types of datasets Table-1, are used for BSNL real CRM data set. One is extracted from Call Data Records for studying the customer churn prediction and other is extracted from customer profile-PLAN allotment for studying the profitable customer-PLAN prediction problem. The first data sets where data are labelled as “T” (true customers) that represents churning and labelled as “F” (false customers) that represent non-churners. It has important types of attributes selected on feature selection techniques that allows identifying the most relevant, useful and effective attributes for customer churn prediction. Another second data sets are transactional datasets are also taken with three main attributes like lower income group (LIG), Middle Income Group (MIG) and higher income groups (HIG) selecting suitable PLAN among Individual (IND), Business (BUS) and Others (OTH) groups for selecting two categories of customers like profitable and not profitable.

Table 1: Data set description of BSNL

CRM Data sets	Instances	Selected Attributes	Models Used
1	49	5	Classification for Churning of Customers Clustering for Market Segmentation
2	120	4	Classification for Profitable PLAN Prediction

4.2 MODELS USED

(I) Classification

It is the process of data analysis for finding a model or functions that distinguish and describes data classes or predicts categorical labels.

- (a) **J48:** It is an open source Java implementation of C 4.5 decision tree that mainly uses a machine learning model for prediction purpose. On the various values of different attributes of available data, it calculates the resultant value of new sample. The external nodes of J48 represent a class attribute, internal nodes represents all other different attributes and branches of trees represents possible outcomes.
- (b) **Random Forest:** It is an ensemble learning method for classification models that operating through no. of decision trees at training time by inputting of random sampled records and features of dataset to predict the high value of accuracy and low variance.
- (c) **One-R:** It is a very effective rule based like IF-Then simple machine learning classifier algorithm that does not rely the frequency of target, but based on the value of a single predictor according to classification rules.
- (d) **Naive Bayes:** It is an intuitive method based on conditional probabilities of each individual attribute belong to each class to make a prediction. It uses a small percentage of training data sets as a prime advantage.
- (e) **Logistic Regression:** It is a function based supervised learning algorithm used for classification. Based on the threshold variable value as previously set, we made the classification.

(II) Clustering

Clustering is the common data analysis technique for identifying subgroups in the data such that data points in the same sub-group called clusters are very similar while data points in different clusters are very different means: It is an algorithm, where data sets will be partitioned into k pre-defined overlapping subgroups called clusters on which each data point belongs to one group.

4.3 PERFORMANCE EVALUATION

The performance of different classifiers is measured for predicting of churn and profit on the basis of some parameters, i.e. (1) True positive rate (2) False positive rate (3) Precision (4) Recall (5) F-Measure (6) ROC (Receiver Operating Characteristics). Confusion matrix is used to evaluate the classifier quality for a two class problem, i.e. True Positive, True Negative, and False Positive & False Negative. It is a useful tool for analyzing the performance of classifiers by recognizing tuples of different classes. Accordingly Similarity measurement of the classifier’s performance as mentioned in Table-2, includes TPR, TFR, sensitivity (recall), F-measure and precision etc. Where further construction and evaluation requires partitioning of training set and test set. Holdout, Random Sampling and Cross validation are important steps of partitioning.

- (a) Confusion Matrix can be represented as
- (b) **Predicted Class**
- (c) YES NO
- (d) **Actual Class** YES TP FN Positive
- (e) NO FP TN Negative
- (f) Where TP (True Positive) -> Refers to positive tuples that were correctly labeled by the classifier. TN (True Negative) -> Refers to negative tuples that were correctly labeled by the classifier. FP (False Positive) ->

Refers to negative tuples that were incorrectly labeled as positive. FN (False Negative) -> Refers to positive tuples that were mislabeled to negative.

Table 2: Performance Measurement Parameters

Measure	Formulae
Accuracy	$TP + TN / P + N$
Error Rate	$FP + FN / P + N$
Precision	$TP / TP + FP$
TP Rate	$TP / TP + FN$
FP Rate (yes)	$FN / FN + TN$
FP Rate (no)	$FN / TP + FP$
Precision	$TP / TP + FP$
Recall	$TP / TP + FN$
F-measure	$2 \times (\text{Precision} * \text{Recall} / \text{Precision} + \text{Recall})$

In addition to above performance measurement parameters, ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. The curve plots between two parameters

- (g) **True Positive Rate:** $TP / TP + FN$
- (h) **False Positive Rate:** $FP / FP + TN$
- (i) **AUC (Area under the ROC Curve)** measures the entire two dimensional area that provides an aggregate measure of performance across all possible classification thresholds.

On the two CDR data sets, we performed a number of experiments on the proposed machine learning model of by using a WEKA-3.9.5 toolkit for providing the factors behind customer churn. Different techniques like Random Forest, J48 Decision Tree, OneR, Function Logistics, and Naïve Bayes tested. Out of which, Random forest is a useful technique for classification and shows better accuracy (87.75%) and less error in comparison to other on correctly classifying of churn and non-churn customers with 10 fold cross validation as mentioned in TABLE 3 with corresponding graph Figure-2. Similarly for selecting the suitable PLAN, for future decision making by the management with profit, above machine learning techniques also used. The Random Forest technique shows excellent (100%) classification accuracy with no error with 10-fold cross validation, as mentioned in TABLE 5 with corresponding graphFigure-4 among individual, business and other groups of people under Lower income group (LIG), Middle Income Group (MIG) and higher income group (HIG) selecting suitable PLAN. The performance vector output of Weka-3.9.5 tool-kit in case of above two best techniques is mentioned in terms of graph as per Figure-3 and Figure-5 on Table-4 and Table-6. Comparison of performance measurement parameter between churning and profit clearly displayed as per Figure-6 on Table-7. Further, in this study, Attribute Selected Classifier algorithm is used for identification of factors, which clearly indicates the finding of churn customers. Also in this paper, k-means clustering algorithm is used for creating retention policies by decision makers considering between of churn and no-churn customers with profit and no-profit customers of different groups. Analysis of Training data after pre-processing, the discovery of Cluster by clustering data mining techniques to predict the BSNL market segments. The basic purpose of clustering in the BSNL real data set is market segmentation [16]. By testing through K-means Algorithm, it is found that segmentation of the market, according to profit is predicted and suitable plan for specific segmented market is to be chosen by BSNL.

Table 3: Classification details for Churning

Algorithm Used	Correctly Classified	Incorrectly Classified
Random Forest	87.7551	12.2449
J48 Decision Tree	79.5918	20.4082
OneR	79.5918	20.4082
Function Logistics	79.5918	20.4082
Naïve Bayes	67.3460	32.6531

Table 4: Performance Measurement Parameter details for Churning

	Algorithm Used	TPR	FPR	Precision	Recall	F-Measure	ROC Area
Wtd. Average	Random Forest	0.878	0.148	0.878	0.878	0.878	0.835
	J48 Decision Tree	0.796	0.301	0.793	0.796	0.788	0.676
	OneR	0.796	0.329	0.799	0.796	0.783	0.733
	Function	0.796	0.246	0.796	0.796	0.796	0.818

Logistics							
Naïve Bayes	0.673	0.284	0.725	0.673	0.681	0.785	

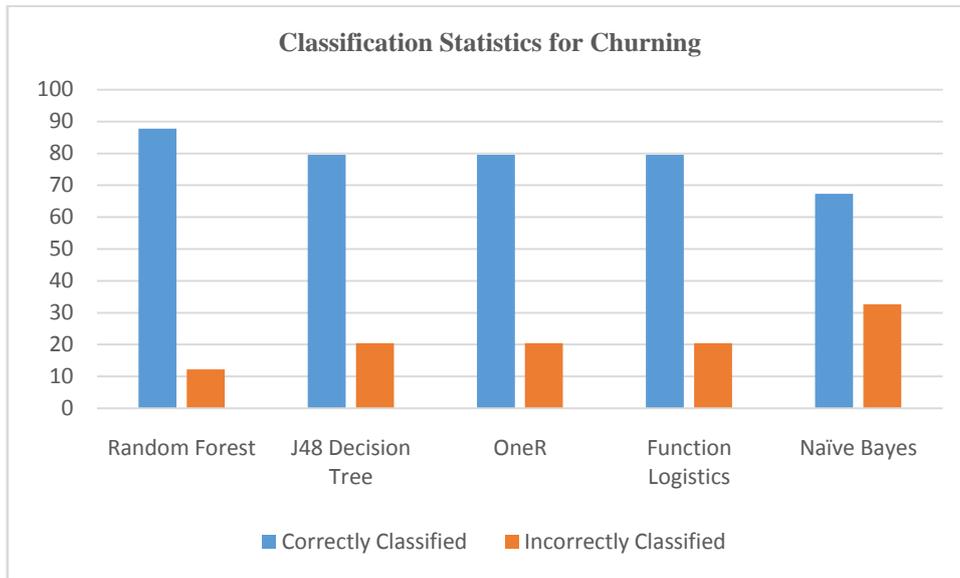


Figure 2: Classification Graph for Churning

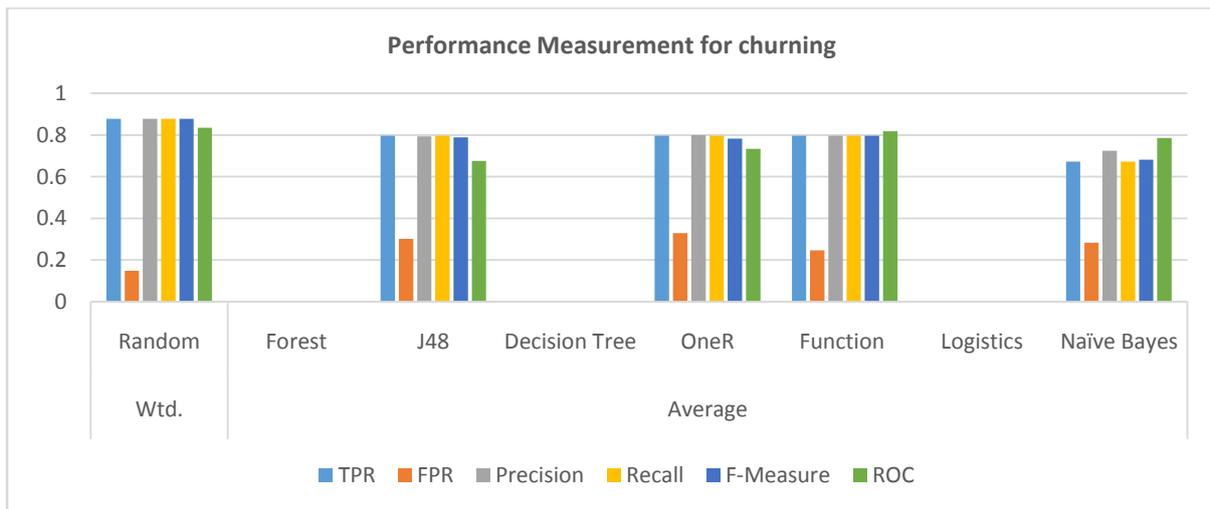


Figure 3: Performance Measurement Graph for Churning

Table 5: Classification details for Profit

Algorithm Used	Correctly Classified	Incorrectly Classified
Random Forest	100	0
Function Logistics	100	0
OneR	97.5	2.5
Function Logistics	97.5	2.5
Naïve Bayes	97.5	2.5

Table 6: Performance Measurement Parameter details for Profit

	Algorithm Used	TPR	FPR	Precision	Recall	F-Measure	ROC Area
Wtd. Average	Random Forest	1.000	0.000	1.000	1.000	1.000	1.000
	Function Logistics	1.000	0.0	1.000	1.000	1.000	1.000
	OneR	0.975	0.075		0.975		0.500
	J48	0.975	0.975		0.975		0.141
	Naïve Bayes	0.975	0.975		0.975		0.975

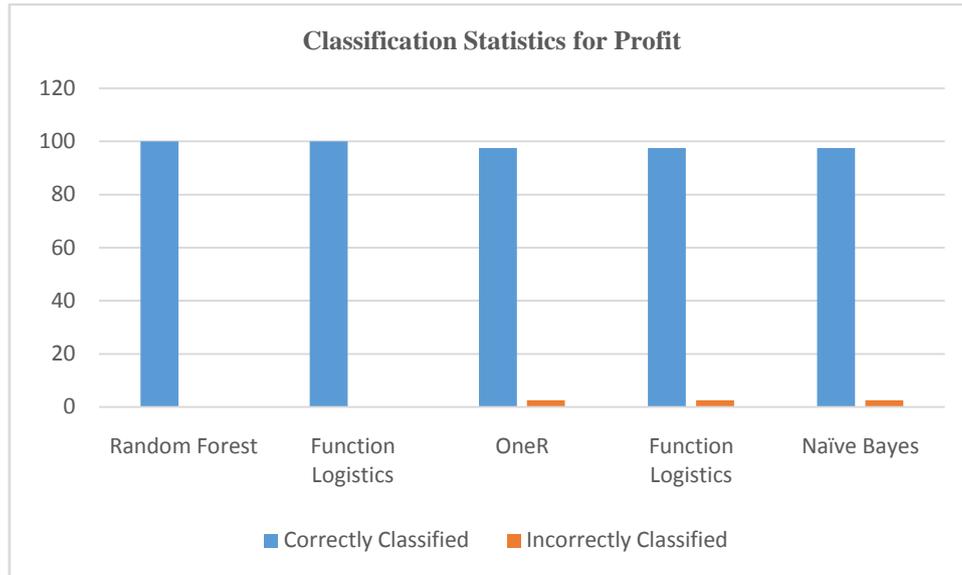


Figure 4: Classification Graph for Profit

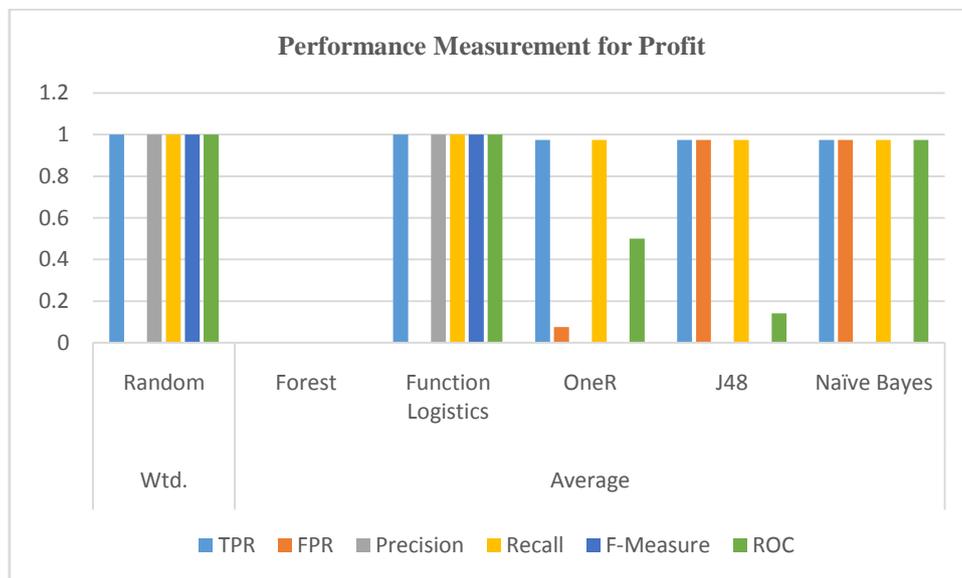


Figure 5: Performance Measurement Graph for Profit

Table 7: Performance Measurement Comparison Statistics between Churning and Profit

	Random Forest		J48		OneR		Function Logistics		Naïve Bayes	
	Accuracy (%)	Model Time (sec)	Accuracy (%)	Model Time (sec)	Accuracy (%)	Model Time (sec)	Accuracy (%)	Model Time (sec)	Accuracy (%)	Model Time (sec)
Churn	87.75	0.043	79.59	0.02	79.59	0	79.59	0.02	67.34	0
Profit	100	0.1	97.5	0	07.5	0	100	0.01	97.5	0

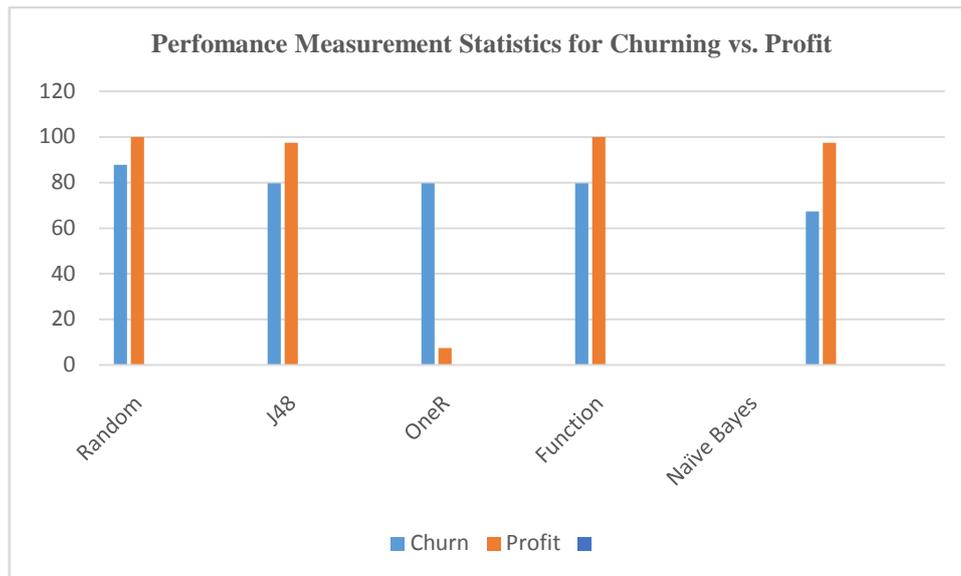


Figure 6: Performance Measurement Comparison Graph for Churning vs. Profit

V. CONCLUSION

Many Telecommunication Company depends on machine learning technology, due to proper classification of different type of customers and market segmentation. Accordingly, It is proposed to avail the opportunities of machine learning algorithm for predicting customer churning, customer profit and market segmentation by BSNL as a Government of India Telecommunication Company for enhancing managerial efficiency and making proper decisions.

REFERENCE

- [1]. Berry I, G. Shostack and G.Upah (Eds), (1983) "Emerging perspectives on service marketing",Chacago,IL, American marketing Association,pp.25-28.
- [2]. Amulya, M. and Anand, D,(2011) "Market Competence of BSNL in the Dynamic Telecom World", Indian Journal of Marketing, Vol.41, No.6, pp.13-21.
- [3]. Radha Krishna, G. and. Shylaiian, C.S, (2007) "Determinants of Habitual Buying Behaviour: A Study of Branded Apparels", The ICFAI Journal of Marketing Management, Vol. 6, No.3, pp.6- 21.
- [4]. Bhave and Ashish, (2002) "Customer Satisfaction Measurement", Quality and Productive journal, pp.562-572.
- [5]. V Mohan Kumar, (2013) "A Study of Customer Relationship Management (CRM) in Bharat Sanchar Nigam Limited (BSNL) AP Telecom Circle",AphDThesis in R&D Cell, JNTU, Hyderabad-5.
- [6]. NematSheereen, S, (2014) "A Study on Customer Satisfaction of BSNL Services in Kerala", Intercontinental Journal of Marketing Research Review, ISSN: 2321-0346 – online, ISSN: 2347-1670,Vol. 2, Issue 9 pp.137- 140.
- [7]. Sreeekumar D Menon, (2014) "Customer Perception towards A Public Sector Telecom Company (BSNL) in Mobile Services", International Journal of Management and Social Sciences Research, (IJMSSR), ISSN: 2319-4421, Vol.3, No.3, pp.29.
- [8]. Kiran Dahiya, KanikaTalwar, (2015) "Customer Churn Prediction in Telecommunication Industry using Data mining technique-A Review", International Journal of Advanced Research in Computer Science and Software Engineering, Vol-5, Issue-4.
- [9]. Jiang Heling, An Yang, Fengyun Yan, Hang Miao (2016) "Research on pattern Analysis and data Classification methodology of Data mining and knowledge Recovery", International journal of Hybrid Information Technology, IEEE, Vol-9, pp.179-186.
- [10]. Saravaman Suba, T Christopher, (2016) "An efficient Data Mining methods to find frequent item sets in large data base using TR-FCTM", ICTACT Journal of Soft Computing,Vol-6.
- [11]. Kiran Dahiya, Surbhi Bhatia, (2015) "Customer Churn Analysis in Telecom Industry", IEEE.
- [12]. Trilok Chand Sharma, (2013) "WEKA Approach for Comparative Study of Classification Algorithm", International Journal of Advanced Research in Computer and Communication Engineering, Vol-2, Issue-4, pp.1925-1931.
- [13]. Tina R. Patil, Mrs. S.S. Shrekar (2013), "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification", International Journal of Computer Science and Applications, Vol-6, pp.256-260, 2013.
- [14]. Sanjeev Gour, (2018) "Classifying the Incidence Rates of Cancers Using Data Mining Techniques (Perspective to Gas Leakage Accident of Bhopal City)", International Journal of Computer Science and Engineering, Vol-6, Issue-1, pp.95-100.
- [15]. SwetaVerma, VivekBadhe, (2016) "Survey on Big Data and mining Algorithm",IJSRSET,IEEE Papers, Vol-3, Issue-2.
- [16]. Meenakshi, Geetila, (2014) "Survey on classification methods using WEKA", International Journal of Computer Applications, Vol-86, pp.16-19.