Prediction of Customer Churn in Telecom Industries

Shivam Sharm , Suryavamsi Saripudi, Sushasukhanya

*Abstract***:** *Churn Examination is one of the widespread used study on Subscription Oriented Businesses for analyzingthebehavior and activities of customers in order to predict beforehand which customer is likely to exit the service agreement.Built on Machine Learning procedures and algorithms it has become very significant for companies in today's marketas securing of another client is more costlier than their maintenance.The paper analyses the relevant studies on Customer Churn Analysis in Telecom Business to present overall information to readers about the commonly used data mining means, and performance of the methods. Initially, we present the details about the availability of public datasets and various customer details in each dataset for predicting customer churn. Then, we compare and contrast various analytical modeling systems and compare their performances and results. Conclusively, we review what kinds of performance metrics have been used to gauge the current churn prediction approaches. Examining all these three viewpoints is very critical for developing a more well-organized churn prediction structure.*

*Index Terms***:** *EDA – exploratory data analysisCRM- customer relationship managementLRM- logistic regression model SVM- support vector machines*

I. INTRODUCTION

Mass customer data have been collected   through out the functional procedure of the mobile telephone compan y, it is quite important for a competitive company to collect effective information in excessive data resources,the n to create a combined information platform, but it seems im possible to deal with the useful data using the conventional method of database management   telecommute   market competition coming into rough competition at times, severaldomestic telecommunication corporation commence to use several systems to solve the problem. Customer churn is a crucial tool during foundation and system set up predictive model based on customer conduct.

With sometimes harsh competition from the telecommuni cations market, the telecommunications company starts usin g several systems to solve the problem. Customer churn is a crucial tool for establishing a predictive model based on cust omer behavior during the foundation and system.In   this paper we propose extrapolative models using machine

Shivamsharm, SRM institute of science and technology, shivamsharma_sa@srmuniv.edu.in

Suryavamsisaripudi, SRM institute of science and technology, Surya.saripudi@gmail.com.

Mr Sushasukhanya, SRM institute of science and technology

learning to predict whether the customers in Communications/telecomfirm will churn or not. We propose the machine learning models with different algorithms such as Naïve Bayes , Random forest. Prediction performance of each algorithms is estimated using accuracy matrix, More challenging is to seta model for Telecommunication sector as there are no contracts between a customer and Telecommunication concerning the duration of facilities/services.The telecom industry enduresrising valuing burden globally. Studies to be done on customer churns is more critical for the Telecom companies nowadays.Classification problem–classification task comes under supervised learning in machine learning where the main goal is to establish models which are supervised by an external agent where the classifiers to training samples are known well in advance. The created models identify the class labels of an unseen sample using this technique.Feature Selection-Feature selection is the process of identifying and selecting the relevant features from the environment. The selected features are extracted using various feature extraction techniques .a feature vector is made which represents the set of features such that each feature vector is mapped to a class label. This helps in identifying class label . Moreover only relevant features are extracted as any unnecessary feature adds more to computational cost and random errors.

II. LITERATURE SURVEY

This paper proposes a model for predicting employee churn in an organization. Employees are an important part of organization further, hiring a new employee turns out expensive for any organization and thus retaining current employees is the optimal solution. Linear support vector machine,c.5 decision tree random forest ,k-nearest neighbor and naïve bayes classifiers are used for classification.   This study requires further exploration to minimize the prediction rate.[2] This paper builds model to predict customer behavior in fitness industries it is found that a yearly gym membership allows the client to terminate their membership with hardly any advanced notice.[3]model based on logistic regression, decision trees and neural networks is built. This study finds it hard to find false positive rate for leavers at certain situations. Moreover it assumes that customers that have sporadic transactions i.e irregular gym attendance are also classified as churned customers. [3] A comprehensible customer churn prediction model is built for analyzing client behavior for

determining which customer is likely to churn in the future. In order to prevent A model based on multiple kernel vector support machine approach is built .the drawback of MK-SVM is that while feature selection they can reduce some of the relevant features .This study has left the application of this framework to financial etc institutions as a future work.[4]This paper focuses on analyzing customer churn in telecom industry using logistic regression, neural network and decision trees, although neural networks perform very well for classification and prediction tasks it only does so for very large datasets. Moreover it takes the same amount of time for processing much smaller data sets.they left this work for future studies for making this model to handle large datasets.[5] This paper uses the technique of decision trees to develop a model for telecommunication subscribers churn prediction. This study can't work on diverse data, which is considered as a drawback. .Its future work is test the approach on bigger data sets containing data over a longer period of time.[6] This work made use of four different rule generation algorithms (i.e. Exhaustive, genetic ,covering and LEM2 to predict customer churn in telecom industry with and without counter. The most fou problem of which classification technique could use to tactic the churn prediction in a more appropriate manner ,remains an open exploration problem. At the same time, the black-box model generated by SVM is also considered as one of its main drawback. It is necessary to increase the number of home appliances to be controlled.

III. EXISTING SYSTEM

Studies shows that acquiring new customers is about 5 to 10 times expensive than retaining their existing customers and moreover keeping the customers loyal in today's competitive conditions has become priority for any organizations , according to reports an average business loses around25-35% of their customers every year. Many companies, realizing this situation,are strongly focused in satisfying and retaining their customers in order to prevent churn. Particularly in the subscription oriented businesses, such as telecommunications, banking sectors, insurance companies, andin general inany particular field where customer relationship management(CRM) is crucial for the organization .The revenue generated and overall profits of the companies are provided by the payments/investments made by the customers periodically. Therefore the need of hour is to be able to keep customers gratified in order to be able to sustain this profits and revenue with the least expenses and minimize loss.In today's technological conditions, large volume of data is being produced from different sources in various sector .It is very important the data extractedfor large chunks of data repositories is pre-processed properly because the useful information hidden in these data sets can't be put into use, unless they are processed properly. In order to find out this hidden information and features, data science comes in handy for information extraction using severaldata mining methods and machine learning algorithms.Wereview the existing works on churn prediction in three different perspectives:

datasets, methods, and metrics for banking sectors. Initially, we present the particulars about the availability of public datasets and customer detailsavailable in each dataset for predicting customer churn. Secondly, we compare and contrast the various predictive modeling methods that have been used in the literature for predicting the churners using different categories of customer records, and then quantitatively compare their performances. Finally, we summarize what kinds of performance metrics have been used to evaluate the existing churn prediction methods. Analyzing all these three perspectives is very crucial for developing a more efficient churn prediction system for telecom industries.
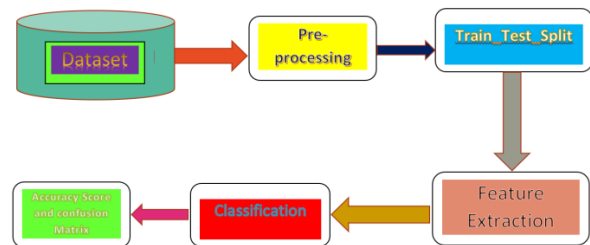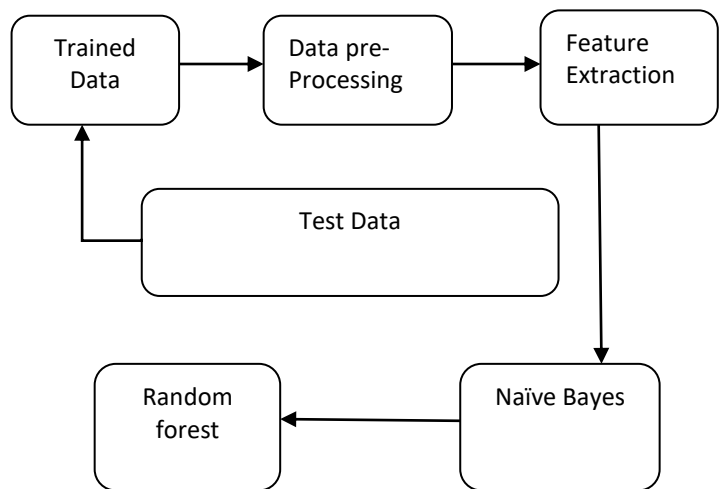

Fig-1.2system architecture and flowchart

Predicting Customer churn is a business scenario in which a company is trying to retain a customer which is more likely to leave the services. For reducing churn rate, we have to classify which customers are most probably going to churn and which will not. Also we have some data to train our model which makes our problem as Supervised Classification problem. EDA It includes looking into the data analyzing various variables, visualization, missing value analysis, correlation analysis

Dataset analysisThe data set is obtained from KaggleWebsite, is used in this

370

paper for the churn analysis and prediction.This data set comprises 18 attributes and 7044records or tuples. The bar graph (fig1.4) and pie chart (fig1.5) respectively shown below gives the number of customers that will churn.
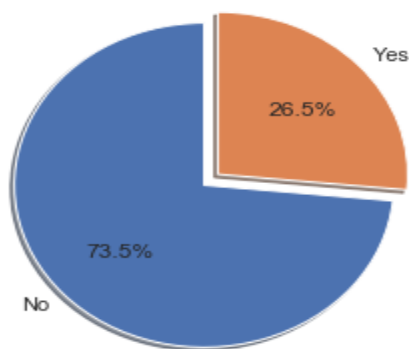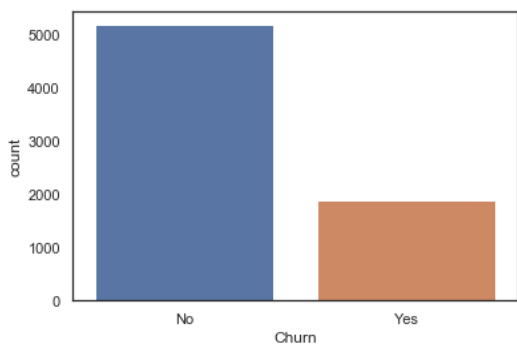

Fig-1.4 count vs churn and Fig-1.5 pie chart

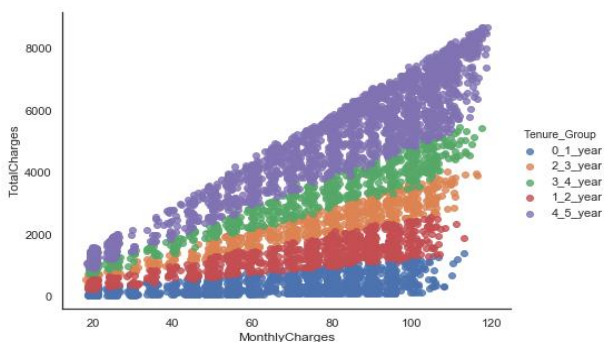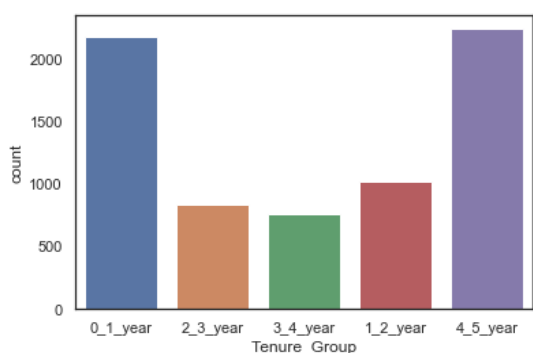The categorical data values are converted into numerical values.

Fig-1.7 tenure_groupvs count

This bar chart fig(1.7) demonstrates the comparision of the services that are being provided by a particular company in the particular years (tenure_group).The no. of services (Count) being provided by the company vary from year to year tenure_group).By this comparison we can understand what is lagging and predict the churn by the means of changes in the amount of services that are being provided by the company to the customers.
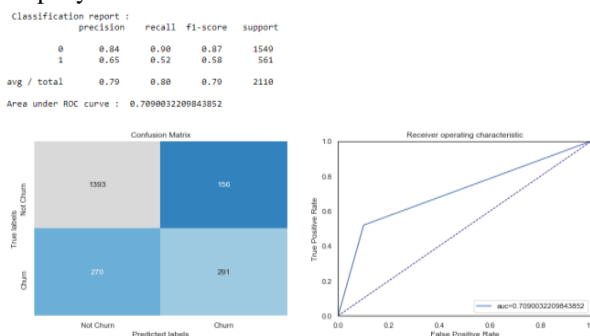

Fig 1.8 classification performance

Factors considered in measuring performance are accuracy , sensitivity, specificity and precision .accuracy in the sense how good the classifier is performing sensitivity means how perfect the classifier is with respect to positive entries. There are 2 kinds tuples ,positive and negative tuples. positive tuples obey some specific rules whereas negative tuples do not . 4 factors are taken into consideration .true positive(TP) tuples under consideration is perfectly positively classified i.e expected and observed tuple were positive.True negative(TN):if tuple was observed negative and expectation was same .False positive(FP): data tuples were mistakenly classified as positive however expected outcome was the opposite.False negative(FN)were actually supposed to be classified as positive but were shown to be negative.We can calculate the performance measures of the created classifiers by using the following equations representing the relationship b/w the tuples wrt to the performance measuring factors.

Accuracy=TP+TN/P+N---------------(1)
Sensitivity=TP/P-----------------------(2)
Specificity=TN/N----------------------(3)
Precision=TP/TP+FP------------------(4)

Confusion matrix:simple tool to check whether tuples belonging to a class are perfectly classified.

IV. CONCLUSIONS

This paper find out which machine learning model is efficient and performs wellin order to predict and analyse on the customer data with the help of the predictive algorithm and is used to find whether the customer will churn or not. it got better prediction result with each algorithms.Finally, we summarize what kinds of performance metrics have been used to evaluate the existing churn prediction methods. This study shall help telecom companies what are the factors that

causes churn in customers and can take the necessary steps to minimize that.

REFERENCES

1. Dilip Singh Sisodia, SomduttaVishwakarma, AbinashPujahari. "Evaluation of machine learning models for employee churn prediction" , 2017 International Conference on Inventive Computing and Informatics (ICICI), 2017
2. Jas Semrl, AlexandruMatei. "Churn prediction model for effective gym customer retention" , 2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC), 2017
3. zhenyuchen ;zhiping fan(2011)building comprehensible customer churn prediction model.
4. Preeti K. Dalvi; Siddhi K.Khandge ;AshishDeomore; AdityaBankar ;V. A. Kanade(2016) Analysis of customer churn prediction in telecom industry
5. Saad Ahmed Qureshi ;AmmarSaleemRehman ; Ali Mustafa Qamar; Aatif Kamal(2013) . Telecommunication subscribers' churn prediction model using machine learning
6. Adnan Amin ;Changez Khan ;Imtiaz Ali ;Sajid Anwar(2014) Customer Churn Prediction in Telecommunication Industry: With and without Counter
7. Alpaydin, E. (2010). Introduction to Machine Learning. London, England: The MIT Press.
8. Archaux, C., Laanaya, H., Martin, A., &Khenchaf, A. (2004). An SVM based Churn Detector in Prepaid Mobile Telephony. IEEE .