International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 10, ISSN No. 2455-2143, Pages 55-60
Published Online February 2018 in IJEAST (http://www.ijeast.com)

CUSTOMER CHURN PREDICTION IN MOBILE TELECOMMUNICATIONS USING FEATURE SELECTION BASED ENSEMBLE LEARNING TECHNIQUE

Maninderjeet Kaur, Priyanka
Computer Science and Engineering
Swami Vivekanand Institute of Engineering and Technology

Abstract-Customer acquisition and customer retention are one of the most competitive factors in most of the companies. Due to ever increasing competitions of customers in companies, the company owners are unable to maintain the customer satisfaction which leads to customer churn. The customer wishes to leave the service of the company causes churn. Most of the sectors are affected by churn problems. Telecommunication is one of the main industries that are affected by churn problem. Prediction of customers who are at risk of leaving a company is known as churn prediction and it is imperative for sustainable growth of a company. Supervised classification technique suits best for solving this problem. In this research work a technique is proposed using Correlation based Symmetric uncertainty feature selection and ensemble learning for customer churn prediction in telecommunication industry. Dataset has been collected from UCI Dataset repository and various other telecommunication websites. To predict whether a customer will be a churner or non-churner, there are a number of data mining techniques applied for churn prediction, such as artificial neural networks, decision trees, and support vector machines. The proposed customer churn prediction technique for telecommunication industry is tested using various parameters like accuracy, precision, recall, f-measure and error rate. Proposed technique is compared with support vector machine technique and Bayesian network. Experimental results demonstrate that proposed technique outperforms various algorithms in certain parameters and also gives fewer errors as compared to the previous techniques.

Keywords – Churn prediction, customer churn, churner, non-churner, customer acquisition, telecommunication industry**.**

I. INTRODUCTION

Data mining is the computing process of discovering patterns in tremendous data sets involving methods at the intersection of machine studying, information, and database programs. It's an predominant approach where intelligent methods are utilized to extract data patterns. The total goal of the information mining method is to extract expertise from a knowledge set and turn out to be it into an comprehensible structure for further use. Excluding the raw analysis step, it includes database and information administration facets, knowledge pre-processing, mannequin and inference issues, metrics, complexity issues, submit-processing of found out buildings, visualization, and on-line updating. Information mining is the analysis step of the "knowledge discovery in databases" procedure, or KDD.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 [1]. The discovered knowledge can be applied to process control, query processing, decision making, and information management.In recent years, management of organizations is moving from “Product-Centric” to “Customer-Centric” [2]. They are not only provides products to meet the need of customers but also improving their services to increase the loyalty and satisfaction of the customers. Intense competition in the market has increased the need for retailers to use strategies focused on retaining the right customers. Acquiring the new customers is more expensive than retaining the existing customers. To retain the customers, organizations are more concern about the customer behaviour analysis. The major factors of success include learning costumers’ purchase behavior, developing marketing strategies to discover latent loyal customers [3]. However, a strategy that is effective in acquiring new customers may not be the most effective in retaining existing customers so in order to design the effective activity to retain customers they need to use the effective strategy for this. So different marketing strategies can be devised that will target different sets of customers. Predicting those profitable customers is important to inform

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 10, ISSN No. 2455-2143, Pages 55-60
Published Online February 2018 in IJEAST (http://www.ijeast.com)

and guide the decision making to keep the products and services competitive. Customer, churn also known as customer attrition, customer turnover, or customer defection, is the loss of clients or customers. Churn administration entails making quintessential reactions with a purpose to keep shoppers who're at the threat of leaving a carrier supplier, for instance via delivering them higher services and offers.Churn prediction is process which helps you predict which of your customers is going to leave your product or service. Very important is that you want to know it before this happens. To do this you need to have data that describe your users and historical knowledge that left.

Figure 1: Churn Prediction Landscape

II. BACKGROUND

Yong Liu et al. dissect client conduct information of a telecom organization. Endeavors have immense measures of client conduct information in the period of enormous information. Instructions to exploit this information to assess custom relinquish chances adequately is a typical issue looked by undertakings. The vast majority of conventional client beat anticipating models disregard client division and misclassification cost, which diminishes the objectivity of model. Managing these lacks, inquire about model of client agitate is built up in light of client division and misclassification cost.[4]

Kiran Dahiya et al.proposes a new framework for the churn prediction model and implements it using the WEKA Data Mining software.. To address the issue of getting by in the focused condition, the maintenance of existing clients has turned into a gigantic test. In the study done in the Telecom business, it is expressed that the cost of obtaining another client is much more that holding the current one. Accordingly, by gathering learning from the telecom enterprises can help in foreseeing the relationship of the clients as whether or not they will leave the company. The required action needs to be undertaken by the telecom industries in order to initiate the acquisition of their associated customers for making their market value stagnant. The efficiency and the performance of Decision tree and Logistic regression techniques have been compared. [5]

Ramakrishna Vadakattu et al. portray the way toward building an agitate forecast stage for huge scale membership based organizations and items. These organizations are continually enhancing on client maintenance strategies, one of which is to anticipate which client will stir in not so distant future. Further, this paper utilizes the choice tree to frame business rules which help in better focusing on the maintenance plans. The Churn expectation is a ceaseless procedure and it winds up noticeably basic to track clients. This portrays the novel usage of a record/score which we use to track and screen client responsiveness to maintenance plans and execution over some undefined time frame. Paper is finished up by giving new roads to dealing with the stir forecast issue better. [6]

Sindhu M E et al.employs the variants of Support Vector Machine such as Proximal Support Vector Machine, Active Support Vector Machine and Lagrangian Support Vector Machine for creating the prediction models and the best model is recognized based on predictive accuracy. The customer wishes to leave the service of the company causes churn. Most of the sectors are affected by churn problems. Telecommunication is one of the main industries that are affected by churn problem. Prediction of customers who are at risk of leaving a company is known as churn prediction and it is imperative for sustainable growth of a company. Supervised classification technique suits best for solving this problem. [7]

Ramakrishna Vadakattu et al. In this paper the author has described the procedure of building a beat forecast stage for substantial scale membership based organizations and items. The novel procedure of utilizing information division and past forecast of the client to additionally expand the exactness and review of the model is proposed. We depict the novel execution of a file/score which we use to track and screen client responsiveness to maintenance plans and execution over some stretch of time. The created stage is conveyed on a few eBay destinations and has brought about the expansion of key business measurements [8].

Ammar A Ahmed et al. Customer churn prediction has gathered greater interest in business especially in telecommunications industries. Many authors have presented different versions of the churn prediction models greatly based on the data mining concepts employing the machine learning and meta-heuristic algorithms. This aim of this paper is to study some of the most important churn prediction techniques developed over the recent years. This paper summarizes the churn prediction techniques in order to have a deeper understanding of the customer churn and it shows that most accurate churn prediction is given by the hybrid models rather than single algorithms so that telecom

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 10, ISSN No. 2455-2143, Pages 55-60
Published Online February 2018 in IJEAST (http://www.ijeast.com)

industries become aware of the needs of high risk customers and enhance their services to overturn the churn decision. [9]

RiddhimaRikhi Sharma et al. speaks to that Churn Prediction has been significant research issue with the development of market improvement as client's benefit more profitable people for development of organization. The undertaking of stir forecast is to distinguish the clients who are putting on a show to move starting with one organization then onto the next. The general goal of this paper is to look at the different calculations of Data Mining have been utilized for making recognize clients into faithful and stir, with the goal that fitting advances can be thought about so as to hold the agitate clients to the organization as clients are more important to the survival and improvement of the organization. [10]

III. PROPOSED TECHNIQUE

The process of customer churn prediction model is described in **Figure 3.** In the first step, samples are selected for better understanding and predicting customer attrition. In second step, necessary features for predicting customer attrition are collected. In third step, data pre-processing is performed. In data pre-processing, data has been collected from different sources, having some missing values in it so there is a need to integrated, cleaned and transformed 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.

In the existing work, PCA is used to feature selection which is based on a Gaussian process (assumption), to measure the variance and sort the eigen values which are proportional to the variances in descending order. The assumption is that the main Eigen values (EVs) contains most of the information and therefore, we use the main components (EVs) for data reduction. So far so good. But applying this approach for feature reduction is risky because this assumes that the feature in itself is stable (invariant) AND the feature's variance contains all information for classification.

Real-world features contain noise, etc. and therefore, PCA generates in its main components EVs with the highest variance which are related to noise. Hence, it will generate "new" features which are not stable. To solve this problem, the symmetric uncertainty based feature selection with hybrid data classification is used to enhance the results.

The proposed approach comprises of using Correlation based Symmetric uncertainty feature selection and ensemble learning for customer churn prediction in telecommunication industry.

The methodology for proposed technique is as follows:

1.Collection of dataset

Dataset has been collected from UCI Dataset repository and various other telecommunication websites.

2.Preprocessing and Filtering

Collected raw data can be further preprocessed and filtered using various filters. Replace missing value filter can be used to replace the missing values from the dataset.

3.Features Selection

To reduce the dimensionality of dataset, Feature selection approach can be used. A component choice is the procedure that can be utilized to expel terms in the preparation records that are factually uncorrelated with the class names. This will decrease the arrangement of terms to be utilized as a part of characterization, in this manner enhancing both proficiency and precision.

With a specific end goal to choose the best highlights Symmetric vulnerability can be utilized that will compute the wellness of highlights for include choice by ascertaining amongst highlight and the objective class. The element which has high estimation of SU gets high significance. We observe an element to be great in the event that it is more significant to the class and not excess to some other highlights of the class. So in short a feature should be highly correlated to the class and not much correlated to any other feature of the class. For this we have used information theory based on entropy. Entropy is a measure of uncertainty of a random variable. It can be defined by the following equation 1 as

$$H(X) = - \sum P(xi) \log2 (P(xi) \quad \ldots\ldots (1)$$

And the entropy of X after observing values of another variable Y is defined in equation 2 as

$$H(X/Y) = - \sum P(yj) \sum P(xi/yj) \log2 (P(xi/yj)) \quad \ldots\ldots (2)$$

Here, P (xi) is the prior probabilities for all values of X, and P (xi/yj) is the posterior probabilities of X when values of Y are given. The amount by which the entropy of X decreases reflects additional information about X provided by Y is called information gain given the equation 3 as

$$IG(X/Y) = H(X)-H(X/Y) \quad \ldots\ldots (3)$$

We can conclude that feature Y is regarded to be more correlated to feature X than to feature Z, if IG(X/Y) > IG (Z/Y). We have one more measure symmetrical uncertainty which shows correlation between features defined by equation 4 as

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 10, ISSN No. 2455-2143, Pages 55-60
Published Online February 2018 in IJEAST (http://www.ijeast.com)

SU (X, Y) = 2 [IG(X/Y) / H(X) + H(Y)] …… (4)

SU compensates information gain's bias toward features with more values and normalizes its value to range of [0,1] with 1 showing that knowledge of either one completely predicts the value of other and 0 shows that X and Y are independent. It considers pair of features symmetrically. Entropy based measures require nominal features, but they can be applied to measure correlations between continuous features as well if they are discretized properly.

4.Classification

In order to improve the performance of individual classifiers Keeping in mind the end goal to enhance the execution of individual classifiers utilized as a part of the paper is to the utilization of meta classifiers, for example, stowing. While the benefit of utilizing gathering strategies is the change of the execution, the hindrance is about the time it takes to complete the preparation stage. In any case, the fundamental concern was to fabricate a model which has a superior execution contrasted with the individual classifiers. This model is novel in light of the fact that it utilizes troupe strategy (Voting), as well as it applies a meta classifier (Bagging) as one of its classifier segment. Additionally, parameter enhancement approach was utilized on its individual classifier (SVM). Every segment in our model adapts a few sections of the characterization issue and we join these speculations to choose the likelihood level.

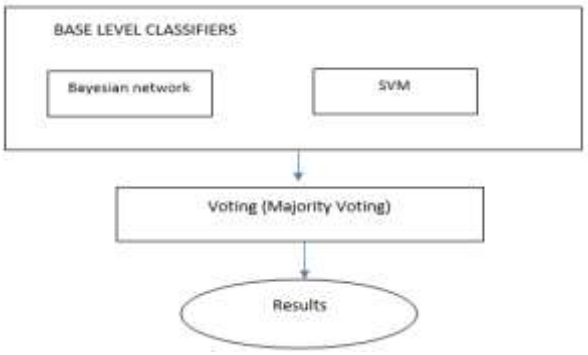

Figure 2: Ensemble learning model

IV. EXPERIMENTAL RESULTS

The simulation has been done in Java Net Beans. NetBeans is an open-source project dedicated to providing rock solid software development products (the NetBeans IDE and the NetBeans Platform) that address the needs of developers, users and the businesses who rely on NetBeans as a basis for their products; particularly, to enable them to develop these products quickly, efficiently and easily by leveraging the strengths of the Java platform and other relevant industry standards.

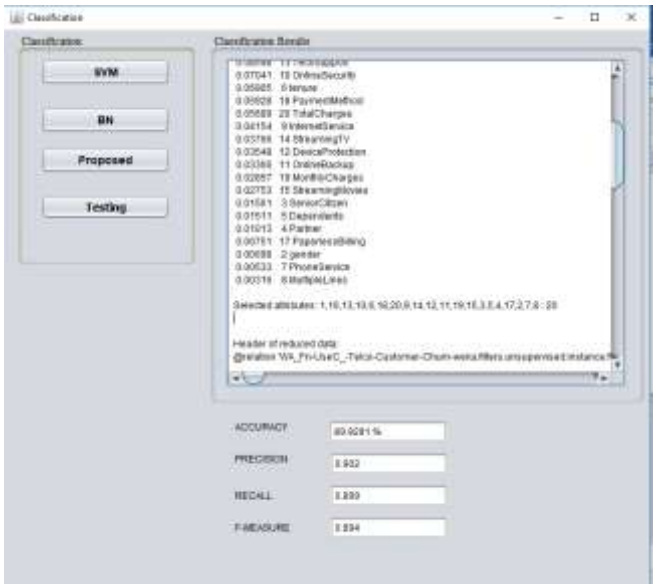

Figure 3: Showing the features ranking using Symmetric uncertainty feature selection algorithm and classification accuracy of proposed hybrid model.

The figure above shows the feature selection and classification results of proposed hybrid classification algorithm. The results show the accuracy of 89.92% i.e. 634 instances are correctly classified out of 705 instances. A Hybrid classification algorithm is an ensemble learning technique that Bayesian network and SVM algorithm and the prediction of both the algorithm is combined using majority voting. The kappa statistics for SVM algorithm is 0.7182. Class details parameters are also shown like precision which is 0.902, recall 0.899, F Measure 0.894, TP Rate 0.899 and FP rate 0.234.

i. **Accuracy:** Accuracy is the basic measure for arrangement execution. It can be measured as accurately ordered occasions to the aggregate number of examples, while blunder rate utilizes mistakenly grouped cases rather than effectively arranged cases.

$$\text{Accuracy} = \frac{\text{True Positive+True Negative}}{\text{True Positive+False Positive+True Negative+False Negative}}$$

Table 1: Representing the Accuracy of proposed method with respect to previous methods

Algorithms	Accuracy
SVM	74.75
BN	74.61
Proposed	89.93

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 10, ISSN No. 2455-2143, Pages 55-60
Published Online February 2018 in IJEAST (http://www.ijeast.com)

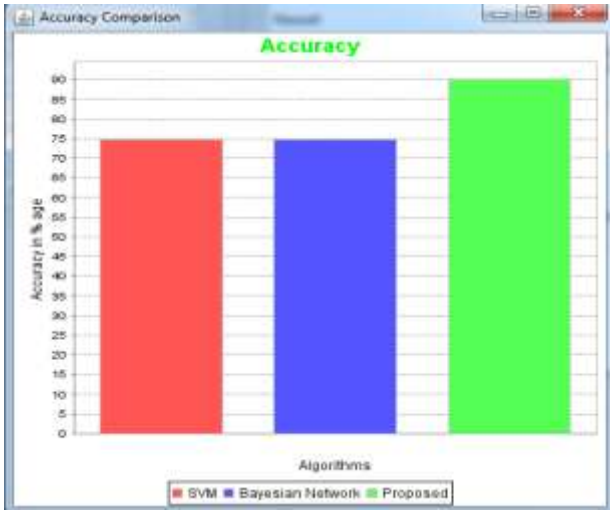

Figure 4: Graph representing accuracy variation

This is the graph representing accuracy of various techniques with respect to the accuracy percentage. It clearly depicts that proposed technique performs better than the previous methods.

i) Precision and Recall

Precision and recall are the two measurements that are broadly utilized to evaluate execution in content mining, and in content investigation field like data recovery. These parameters are utilized for calculating precision and culmination separately.

$$\text{Precision}=\frac{\text{True Positive}}{\text{True Positive}+\text{False Positive}}$$

$$\text{Recall}=\frac{\text{True Positive}}{\text{True Positive}+\text{False Negative}}$$

ii) Fmeasure

The harmonic mean of precision and recall is called F-Measure. The esteem ascertained utilizing F-measure is a harmony amongst precision and recall.

$$\text{F measure} = \frac{2*\text{recall}*\text{precision}}{\text{precision}+\text{recall}}$$

Table 2: Representing the various parameters of proposed method with respect to previous methods

rithms	Precision	Recall	F Measure
SVM	0.681	0.748	0.665
BN	0.805	0.746	0.761
Proposed	0.902	0.899	0.894

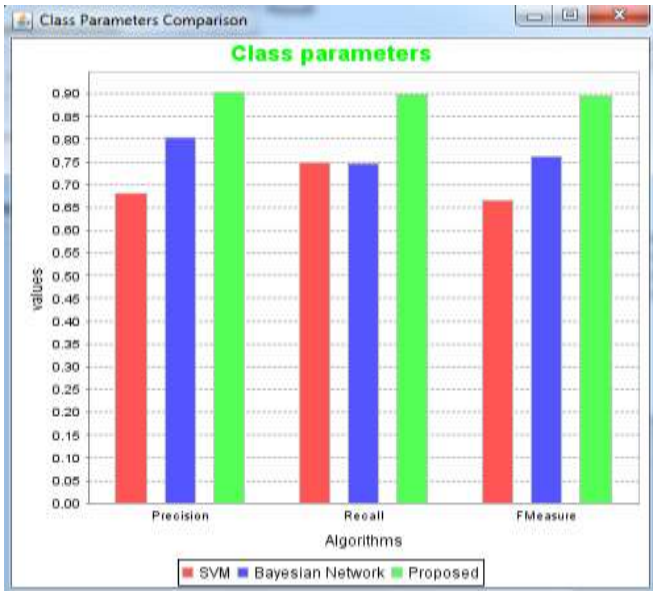

Figure 5: Graph representing class parameters of various algorithms

This graph represents various class parameters comparison like precision, recall, Fmeasure of different techniques along with the proposed technique. It is clear from this graph that proposed technique out performs the existing methods in all of the three parameters.

Table 3: Representing Error of proposed algorithm with respect to various methods

Error	SVM	BN	PROPOSED
Mean absolute error	0.2525	0.2596	0.1538
Root mean square error	0.5025	0.4596	0.2813

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 10, ISSN No. 2455-2143, Pages 55-60
Published Online February 2018 in IJEAST (http://www.ijeast.com)

Figure 6: Graph representing error of proposed algorithm with respect to previous algorithms.

The graph shows the mean absolute error and root mean square error comparison of various algorithms with the proposed. This comparison shows that proposed algorithm gives lesser errors than the existing.

V. CONCLUSION

This research work proposes a technique using Correlation based Symmetric uncertainty feature selection and ensemble learning for customer churn prediction in telecommunication industry. To predict whether a customer will be a churner or non-churner, there are a number of data mining techniques applied for churn prediction, such as artificial neural networks, decision trees, and support vector machines. Dataset has been collected from UCI Dataset repository and various other telecommunication websites. The simulation has been done in Java Net Beans. The proposed customer churn prediction technique for telecommunication industry is tested using various parameters like accuracy, precision, recall, f-measure and error rate. Proposed technique is compared with support vector machine technique and Bayesian network. The proposed technique provides 89.93 accuracy, 0.902 precision, 0.899 recall, 0.894 Fmeasure. The proposed technique out performs the existing methods in all parameters. The error rate of proposed technique is only 0.1538. The comparison shows that proposed algorithm gives lesser errors than the existing.

In future, this technique can be applied on the data set of any other industry. The proposed technique uses two algorithms called as Bayesian and SVM, in future, combination of multiple algorithms can be used to predict customer churn behaviour. It is also not necessary to take majority voting, we can also use mean, median, average methods.

VI. REFERENCES

[1] [1] Han, J., and Kamber, M. "Data Mining: Concepts and Techniques". Morgan Kaufmann, India, 2006.

[2] S.Y. Hung and D. Yen and H.Y. Wang "Applying Data Mining to telecom churn management". Expert System with Applications, Elsevier,Vol. 31, No. 3, pp. 515-524, 2006.

[3] Wang Chong and Wang Yanqing. "Discovering Consumer's Behavior Changes Based on Purchase Sequences". 9th IEEE-International Conference on Fuzzy Systems and Knowledge Discovery, pp. 642-645, 2012

[4] Yong Liu, Yongrui Zhuang, "Research Model of Churn Prediction Based on Customer Segmentation and Misclassification Cost in the Context of Big Data", Journal of Computer and Communications, pp. 87-93,2015.

[5] Kiran Dahiya, Surbhi Bhatia, "Customer Churn Analysis in Telecom Industry", IEEE, 2015.

[6] Ramakrishna Vadakattu, Bibek Panda, Swarnim Narayan, HarshalGodhia, "Enterprise Subscription Churn Prediction", IEEE International Conference on Big Data, pp. 1317-1321, 2015.

[7] Sindhu M E, Vijaya M S, "Predicting Churners in Telecommunication Using Variants of Support Vector Machine", American Journal of Engineering Research, Volume-4, Issue-3,pp-11-18, 2015

[8] Ramakrishna Vadakattu, Bibek Panda, Swarnim Narayan, HarshalGodhia, "Enterprise subscription churn prediction", IEEE International Conference on Big Data (Big Data), 2015.

[9] Ammar A Ahmed, Dr. D. Maheswari linen, "A Review And Analysis Of Churn Prediction Methods For Customer Retention In Telecom Industries", IEEE International Conference on Advanced Computing and Communication Systems, 2017.

[10]RiddhimaRikhi Sharma, Rajan Sachdeva, "Review on Prediction of Churn Customer Behavior", International Journal of Engineering Research & Technology, Vol. 6 Issue 01, pp. 257-261,2017