IJCTA, 9(34) 2016, pp. 661-671

Stochastic Customer Loyalty and Satisfaction Prediction using the SEM and SVM

Manpreet Kaur* **and Manjot Kaur Sidhu****

ABSTRACT

Service and product reviews play an important role in determining what kind of product is. Such reviews provide useful information about customer concern and their experience with the product. Consequently, these reviews will be helpful for a business making products for the purpose of product recommendation, better customer understanding and attracting more loyal customers. As ecommerce has become so popular, numbers of reviews are increasing day by day. It is difficult for a customer to read all the reviews manually. In this paper, an approach is developed which is used to obtain the summary from thousands or hundreds of online reviews. This approach uses extraction summarization for summarizing the reviews thereby selecting the original sentences and putting it together into a new shorter text explaining the overall opinion about the product. Although previous studies of deriving useful information from customer reviews focus on categorical or numerical data and textual data has been ignored. But textual data are of equal importance so it should not be ignored. So, this approach includes every aspect of the review in the summary so that a customer would be able to make a right decision regarding product.

Keywords: Customer loyalty, customer satisfaction, structural equation modeling, support vector machine, churn prediction.

INTRODUCTION

Customer relationship management (CRM) is a strategic approach which targets the development of profitable, long-term relationships with key customers and stakeholders [1]. Due to saturated markets and intensive competition, more and more companies have recognized the importance of CRM and have changed their product centric, mass marketing champion strategies toward customer centric, targeted marketing. Nowadays, the rapid development of digital systems and associated information technologies provide enhanced opportunities to understand customers and build reliable digital CRM systems [2]. Customer churn management, as a part of CRM, has become a major concern. In mobile telecommunications, the term 'churn' refers to the loss of subscribers who switch from one provider to another during a given period. Based on an earlier study [3], the estimated average churn rate for mobile telecommunications is about 2.2% per month. This means that one in fifty subscribers of a given company discontinues their services every month. As it is more profitable to retain existing customers than to constantly attract new customers [4-6], it is crucial to build an accurate churn prediction model for identifying those customers who are most prone to churn. Established literature on customer churn uses various data mining technologies, such as Neural Networks [7], Clustering [8], Decision Tree [7, 9], Regression [10, 11], Support Vector Machine [4, 12] and ensemble of hybrid methods [13], to provide more accurate predictions. According to a review on customer churn prediction modeling [14], Regression is the most commonly adopted technique, probably because of its high reported accuracy and interpretability for understanding key drivers, as well as for providing information to set up retention actions. As the churner usually takes only a fraction of the

* Chandigarh University, *E-mail: birdipreety@gmail.com*

** Associate Professor Chandigarh University, *E-mail: manjotsidhucu@gmail.com*

customer base, the problem of customer churn prediction is always combined with the problem of highly skewed class distribution or lack of churner data. One of the most common techniques for dealing with rarity is sampling [15]. Methods that adopt the sampling technique alter the distribution of training examples and generate balanced training set(s) for building churn prediction model(s) [8, 9, 12, 13]. However, a recent study [11] on the class imbalance issue in churn prediction reveals that advanced sampling technique does not increase predictive performance. Although the weighted Random Forests technique is suggested [11], tree ensembles, such as Random Forests, are often criticized for being hard to interpret [16, 17], i.e., it is difficult to identify risk factors which can be addressed by the retention process to prevent a customer from leaving, thus they are not the preferred methods in this study. Other research explores the power of new features for churn prediction, such as social network [18] and text information of customer complaints [19], which is beyond the scope of this paper. This paper presents a churn prediction model in the telecom industry using a boosting algorithm which is believed to be very robust [11] and has demonstrated success in churn prediction in the banking industry [20]. The established literature only uses boosting as a general method to boost accuracy, and few of researchers have ever tried to take advantage of the weight assigned by boosting algorithms. The weight also provides important information, specifically, outliers. The testing results show that boosting provides a good separation of the customer base, which also leads to a better overall performance.

SIGNIFICANCE OF CUSTOMER REVIEW CLASSIFICATION

There are various reasons that show the importance of customer reviews for an organization selling products online:

1. Whenever a company introduces a new product then customer feedback is very important for deterring customer needs and tastes.
2. Companies can better understand that how their products are better than competitive products by analysing the customer ratings of product and their reasons for selection.
3. Companies can determine whether their customers are getting satisfactory level of service by their employees.
4. Customer reviews help in deterring why consumers are no longer interested in buying products from them, if any. This will help in building up strategies that would help lose customers back into business.

Customer reviews are also important in determining technological trends in the market.

LITERATURE REVIEW

Aktepe, Adnan *et al.* [1] has worked on the customer satisfaction and loyalty analysis with classification algorithms and structural equation modeling. Businesses can maintain their effectiveness as long as they have satisfied and loyal customers. Customer relationship management provides significant advantages for companies especially in gaining competitiveness. In order to reach these objectives primarily companies need to identify and analyze their customers. In this respect, effective communication and commitment to customers and changing market conditions is of great importance to increase the level of satisfaction and loyalty. To evaluate this situation, level of customer satisfaction and loyalty should be measured correctly with a comprehensive approach. In this study, customers are investigated in 4 main groups according to their level of satisfaction and loyalty with a criteria and group based analysis with a new method. The authors have used classification algorithms in WEKA programming software and Structural Equation Modeling (SEM) with LISREL tools together to analyze the effect of each satisfaction and loyalty criteria in a satisfaction–loyalty matrix and extend the customer satisfaction and loyalty post-analysis research bridging the gap in this field of research.

Gaiardelli, Paolo *et al.* [2] has worked towards the classification model for product-service offerings. In this paper, the authors have developed a comprehensive model for classifying traditional and green Product-Service offerings, thus combining business and green offerings in a single model. They have also described the model building process and its practical application in a case study. The model reveals the various traditional and green options available to companies and identifies how to compete between services; it allows servitisation positions to be identified such that a company may track its journey over time. Finally it fosters the introduction of innovative Product-Service Systems as promising business models to address environmental and social challenges. Lu, Ning et. al. [3] has developed the customer churn prediction model in telecom industry using boosting. This research conducts a real-world study on customer churn prediction and proposes the use of boosting to enhance a customer churn prediction model. Unlike most research, that uses boosting as a method to boost the accuracy of a given basis learner, this paper tries to separate customers into two clusters based on the weight assigned by the boosting algorithm. K. Coussement and D. V. Poel [4] has workd on the churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques. This study applies support vector machines in a newspaper subscription context in order to construct a churn model with a higher predictive performance. Moreover, a comparison is made between two parameter-selection techniques, needed to implement support vector machines. W. Verbeke et. al. [5] has focused on the new insights into churn prediction in the telecommunication sector: a profit driven data mining approach. in the first part of this paper, a novel, profit centric performance measure is developed, by calculating the maximum profit that can be generated by including the optimal fraction of customers with the highest predicted probabilities to attrite in a retention campaign. The novel measure selects the optimal model and fraction of customers to include, yielding a significant increase in profits compared to statistical measures. M. Owczarczuk et. al. [10] has worked on the churn models for prepaid customers in the cellular telecommunication industry using large data marts. In this article, the authors have tested the usefulness of the popular data mining models to predict churn of the clients of the Polish cellular telecommunication company. J. Burez and D. V. Poel [11] has worked towards handling class imbalance in customer churn prediction. In this paper, the authors have investigated the increase in performance of sampling (both random and advanced under-sampling) and two specific modeling techniques (gradient boosting and weighted random forests) compared to some standard modeling techniques. AUC and lift prove to be good evaluation metrics. AUC does not depend on a threshold, and is therefore a better overall evaluation metric compared to accuracy. Lift is very much related to accuracy, but has the advantage of being well used in marketing practice. N. Kim et. al. [12] has worked towards the development of uniformly subsampled ensemble (use) for churn management. The present paper explores the possible application of a new ensemble model. The model, which is based on multiple SVM classifiers, is employed to address churner identification problems in the mobile telecommunication industry, a sector in which the role of customer retention program becomes increasingly important due to its very competitive business environment. In particular, the current study introduces a uniformly subsampled ensemble (USE) model of SVM classifiers, not only to reduce the computational complexity of large-scale data, but also to boost the reliability and accuracy of calibrated models on data sets with highly skewed class distributions.

EXPERIMENTAL DESIGN

The prediction and rating algorithm has been used to fetch the important terms out of the input historical data, which are further used to evaluate the class of the customer loyalty and customer satisfaction (CL-CS) data. The historical data undergoes the rating index building using the term list processing. The dataset terms contain the rating data for different rating for customer loyalty and customer satisfaction. The proposed algorithm predicts the customer loyalty and customer satisfaction indices to predict the overall business of the firms,

Algorithm 1: CL-CS Prediction algorithm

1. Input historical data
2. Fetch keyword array from historical data
3. Match keyword array with all terms lists of the given category
4. Build historical indices with the pre-defined rate values in the category specific rating arrays
5. Repeat the step 3 and 4 for all categories

In the Rating algorithm, all training samples were used for training, that is, whenever a new test sample need to be classified, it is necessary to calculate similarities between that sample and all documents in the training sets, and then choose Rating with word samples which have largest similarities. Due to numbers of calculation taken between the test sample and all the training samples, the traditional method of Rating has less computational complexity. To overcome the complexity, this paper introduced combination Rating algorithm with a clustering method.

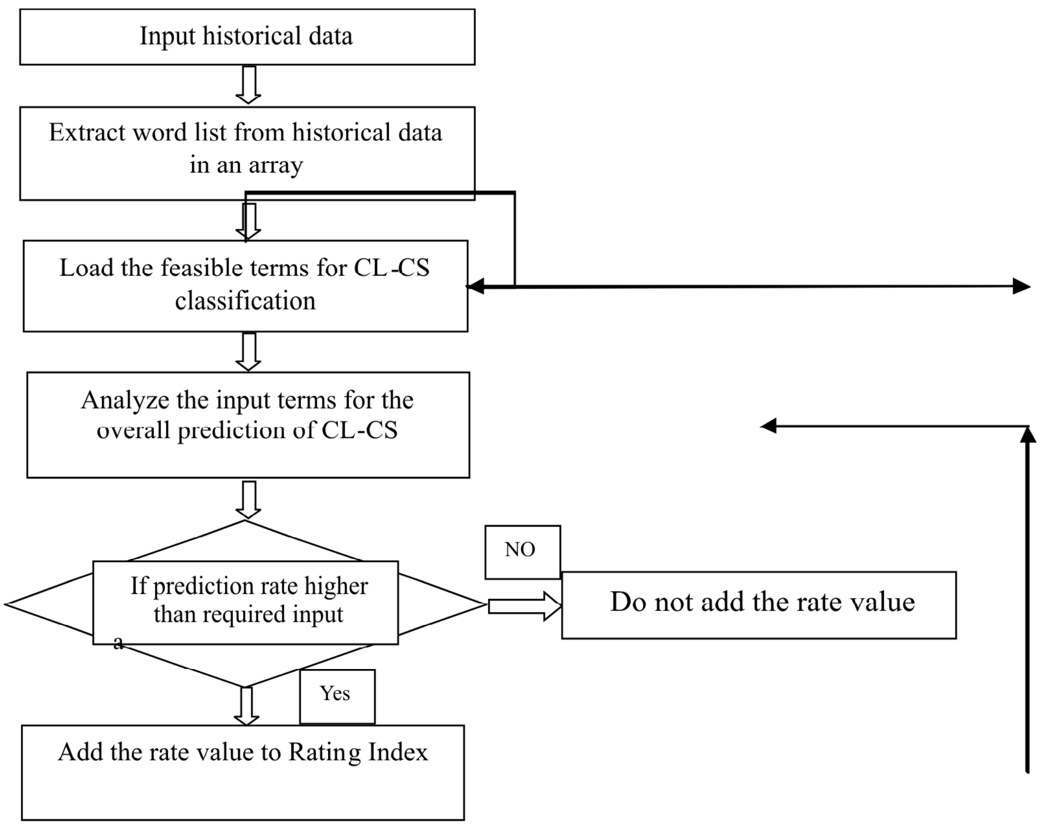

Figure 5.1a: The Rating algorithm for populating Rating Index on the basis of word list matching

First, calculate the weight of each document by summing all term weight and divide it with total term in the document. Second, each category of training sample is clustered by CL-CS customer grouping algorithm:

Algorithm 2: Weighted CL-CS categorization algorithm

1. Initialize the value of for number of categories of document to be created.
2. Get the centroid data from pre-defined centroid set
3. Assign each object to the group that has be closest centroid
5. Update the centroid by calculating the average value of the existing data on the cluster; C_i = 1 n dj n j = 1 (5) Ci : centroid to-i from the cluster n : number of documents in a cluster dj : document vector to-j
5. Repeat step 3 and 4 until the centroids no longer move (convergent).

This produces a separation of the objects into groups from which the metric to be minimized can be calculated. After clustering for each category, the cluster centers were chosen to represent the category and they become the new training sets for structural equation modeling (SEM) with support vector machine (SVM) algorithm. By this method, the number of samples for training is reduced efficiently, so the time for calculating similarities in SEM-SVM algorithm also reduced.

Algorithm 3: SEM-SVM algorithms

1. Input dataset with news rate values
2. Number of categories is a pre-defined number
3. Algorithm determines the pre-defined data points equal to the cluster number
4. The algorithm evaluates the distance of each data point from all of the pre-defined initial data points.
5. The point is added to the cluster with the lowest distance
6. Evaluate all points with method from 4 to 5 until last point.
7. Return the clustered (classified) data
8. The input data acquisition is done on the customer query data.
9. The customer query data contains the customer messages about the various products or services of a company.
10. The individual messages from the users or customers are individually inspected and classified according to their inter-sentence similarity.
11. The similarity analysis depicts the mode of operation to divide a message written in multiple sentences according to the sentence to message relationship.
12. After the classification the intensity of the individual messages is inquired and calculated, which prepares and classifies the customer loyalty (CL) and customer satisfaction (CS) matrix.
13. The CL-CS matrix is further evaluated for the structural equation modeling (SEM).
14. The structural equation modeling (SEM) is the operation performed on the given data matrix obtained after the CL-CS analysis.
15. The structural equation modeling is the product to customer review matrix building and the product relation building, which predicts the customer behavior towards the product and its probability to select or reject the other product in the list.

RESULTS AND DISCUSSION

The statistical parameters to measure the statistical errors (Type 1 and Type 2) are measured in order to evaluate the overall performance of the proposed model by evaluating the samples by the means of the programming or the manual binary classification. The proposed model evaluation is entirely based upon this statistical analysis. The following table explains the significance of the type 1 and type 2 statistical errors for the evaluation of the hypothesis.

Table 5.1
Hypothesis decision parameter entities in type 1 and 2 errors

	Doesn't Have The Condition (Satisfies Null Hypothesis)	*Has The Condition (Does Not Satisfy Null Hypothesis)*
Tests Negative (Null Accepted)	True Negative TN or n_{00}	False Negative FN or n_{10}
Tests Positive (Null Rejected)	False Positive FP or n_{01}	True Positive TP or n_{11}

5.1.1. True Positive

The individual has the condition and tests positive for the condition. The individual does not satisfy the null hypothesis and the test rejects the null hypothesis. The matching samples correctly identified as matching and defined as the following:

$$TP = n_{11} = \text{number of such individuals}$$

And symbolically defined as A = True Positive

5.1.2. True Negative

The individual does not have the condition and tests negative for the condition. The individual satisfies the null hypothesis and the test accepts the null hypothesis.

$$TN = n_{00} = \text{number of such individuals}$$

And symbolically defined as B = True Negative

5.1.3. False Positive

The individual does not have the condition but tests positive for the condition. The individual satisfies the null hypothesis but the test rejects the null hypothesis.

$$FP = n_{01} = \text{number of such individuals}$$

And symbolically defined as C = False Negative

5.1.4. False Negative

The individual has the condition but tests negative for the condition The individual does not satisfy the null hypothesis but the test accepts the null hypothesis

$$FN = n_{10} = \text{number of such individuals}$$

And symbolically defined as D = False Positive

Table 5.1.1
The statistical errors obtained from the simulation

Error Type	*Round 1*	*Round 2*	*Round 3*	*Round 4*
True Positive	16	16	17	19
False Positive	4	4	3	1
True Negative	0	0	0	0
False Negative	0	0	0	0

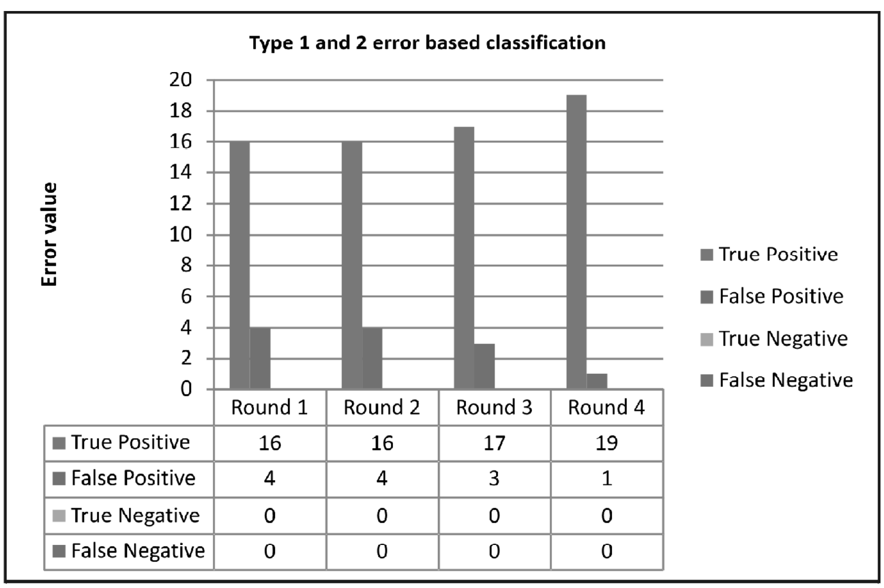

Figure 5.1.1: Type 1 and 2 errors recorded from the proposed model simulation

5.1.5. Precision

Precision can be defined as the ratio of relevant retrieved documents and the information needed by the users. High precision defines that the algorithm returns results that are relevant as compared to irrelevant results. It also defines a predictive value that is positive and this is defined in terms of the binary classification. This classification defines the documents that are retrieved. It is defined in terms of the results that the system returns at some cut-off rank. Precision is also known as sensitivity.

$$Precision = A/ (A+D)$$

Where, A= True Positive

B= True Negative

C= False Negative

D= False Positive

5.1.6. Sensitivity or Recall

Recall is the probability that a test will indicate 'test' among those with the matching sample. Sensitivity is the probability that a test will indicate 'test' among those with the matching sample.

Sensitivity: A/(A+C) × 100

5.1.7. Specificity

Positive predictive values are influenced by the prevalence of correct results in the population that is being tested. If we test in a high prevalence setting, it is more likely that persons who test positive truly have matching probability than if the test is performed in a population with low prevalence. Specificity is the fraction of those without matching sample who will have a negative test result.

Specificity: D/(D+B) × 100

5.1.8. Accuracy

The percentage of the result success out of the whole results is called accuracy. Accuracy is also known as success rate.

$$Accuracy = (Correct\ Results/\ Total\ Results) * 100$$

5.1.9. Positive Predictive Value (PPV): PPV is the fraction of the documents retrieved that are relevant to the user's information need. In binary classification, precision is analogous to positive predictive value. Precision takes all retrieved documents into account. It can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at n or P@n.

$$PPV = TP/(TP+FN)$$

5.1.10. Negative Predictive Value: *Negative predictive values are influenced by the prevalence* of disease in the population that is being tested. If we test in a high prevalence setting, it is more likely that persons who test negative truly have matching probability than if the test is performed in a population with low prevalence.

$$Negative\ Predictive\ Value: D/(D+C) \times 100$$

5.1.11. F1-Measure: In statistical analysis of binary classification, the F1 score (also F-score or F-measure) is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score: p is the number of correct positive results divided by the number of all positive results, and r is the number of correct positive results divided by the number of positive results that should have been returned. The F1 score can be interpreted as a weighted average of the precision and recall, where an

F1 score reaches its best value at 1 and worst at 0. The traditional F-measure or balanced F-score (F1 score) is the harmonic mean of precision and recall:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Table 5.1.2
Performance measures obtained from the four rounds of experiments

Performance Measures	*Round 1*	*Round 2*	*Round 3*	*Round 4*
Accuracy	80	80	80	95
Precision	80	80	85	95
Recall	100	100	100	100
F1-Measure	88.88889	88.88889	91.89189	97.4359

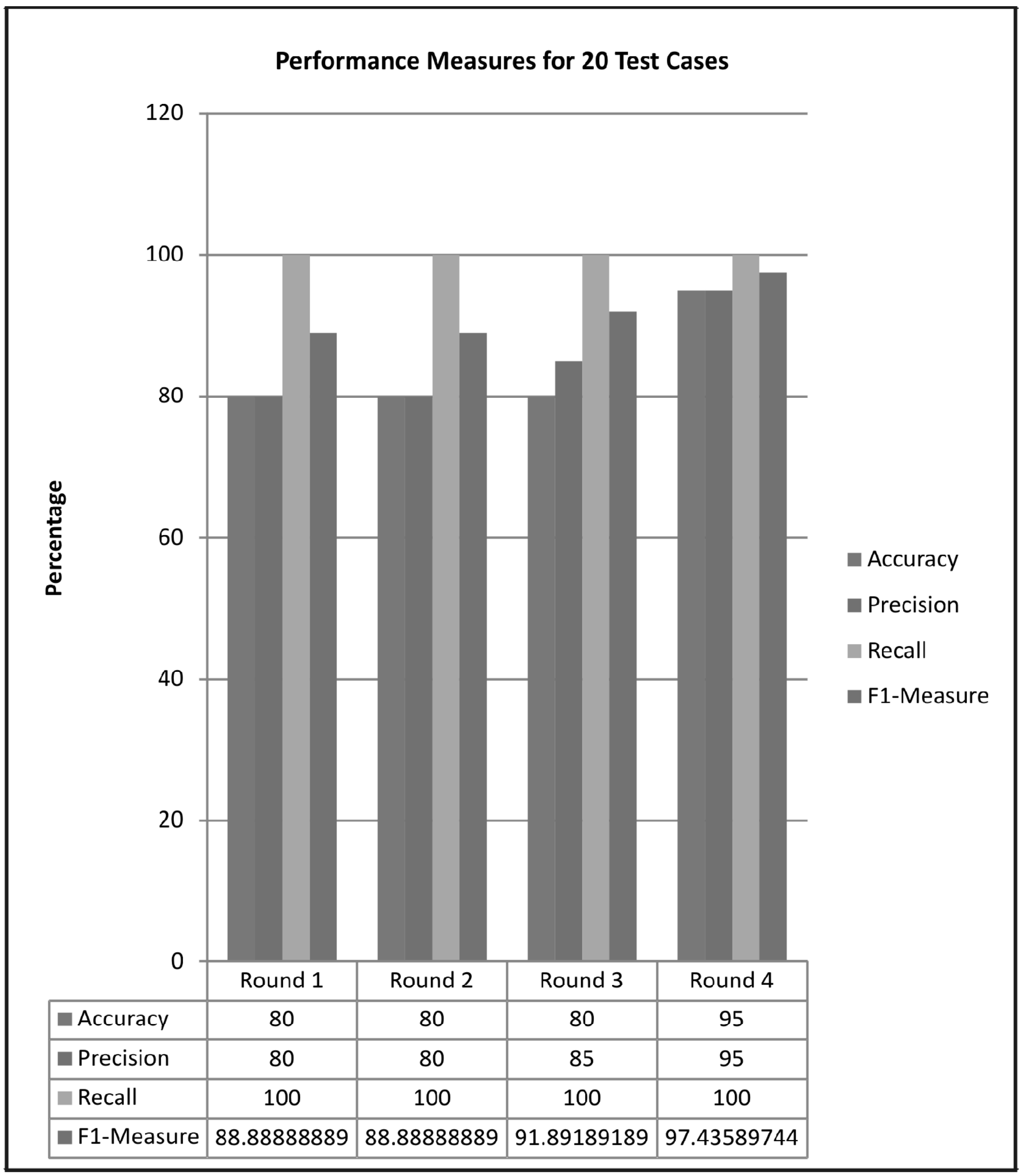

Figure 5.1.2: Graphical representation of the performance measures obtained from the four rounds of experiments

5.2. Comparative Analysis

5.2.1. F1-measure calculated over 20 test cases

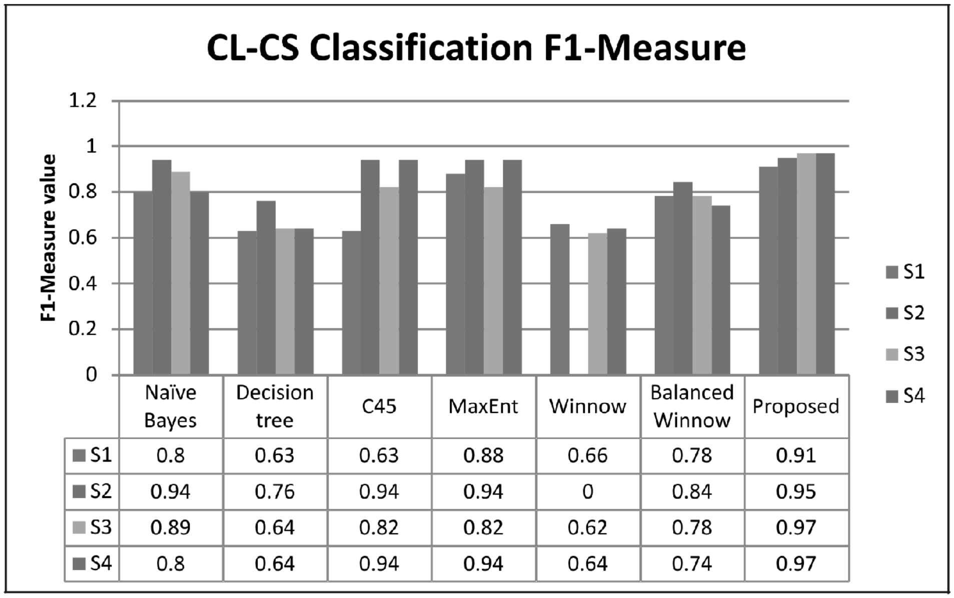

Figure 5.2.1: F1-Measure based evaluation of the proposed model against the other classification models

Table 5.2.1
The table containing the obtained values of F1-measure from the experiment

Test Cases	*Naïve Bayes*	*Decision tree*	*C45*	*MaxEnt*	*Winnow*	*Balanced Winnow*	*Proposed*
S1	0.8	0.63	0.63	0.88	0.66	0.78	0.91
S2	0.94	0.76	0.94	0.94	0	0.84	0.95
S3	0.89	0.64	0.82	0.82	0.62	0.78	0.97
S4	0.8	0.64	0.94	0.94	0.64	0.74	0.97

5.2.2. Average Prediction Accuracy (APA) with 20 test cases

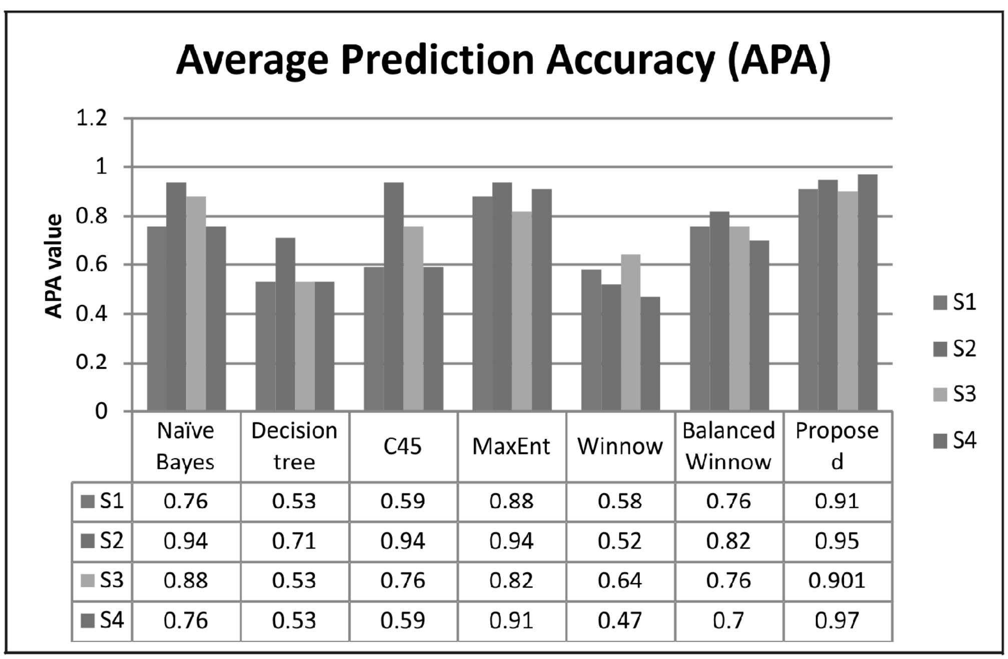

Figure 5.2.2: Average Prediction Accuracy based evaluation of the proposed model against the other classification models

Table 5.2.2
Average Prediction Accuracy values based comparison

Test Cases	*Naïve Bayes*	*Decision tree*	*C45*	*MaxEnt*	*Winnow*	*Balanced Winnow*	*Proposed*
S1	0.76	0.53	0.59	0.88	0.58	0.76	0.91
S2	0.94	0.71	0.94	0.94	0.52	0.82	0.95
S3	0.88	0.53	0.76	0.82	0.64	0.76	0.901
S4	0.76	0.53	0.59	0.91	0.47	0.7	0.97
Average	0.835	0.575	0.72	0.8875	0.5525	0.76	0.93275

CONCLUSION

The proposed model is based upon the customer loyalty and customer satisfaction prediction mechanism has been proposed, which utilizes the various mechanisms for the prediction of the customer trends. The quality based assessment cellular service operations and the major assessments are assessed and predicted by using the multiple factors in order to assess the customer satisfaction and customer loyalty indexes, which signifies the trends for the cellular operator's business in the given span of time. The historical has been analyzed under the proposed model to assess the customer loyalty and satisfaction index assessment. The experimental results are showing the higher accuracy and F1-measure than the other classifiers. Approximately, 10-15% improvement has been recorded in the proposed model in comparison with the existing models based upon the various classifiers.

REFERENCES

[1] Aktepe, Adnan, Süleyman Ersöz, and Bilal Toklu. "Customer satisfaction and loyalty analysis with classification algorithms and structural equation modeling." *Computers & Industrial Engineering* 86 (2015): 95-106.

[2] Gaiardelli, Paolo, Barbara Resta, Veronica Martinez, Roberto Pinto, and Pavel Albores. "A classification model for product-service offerings." *Journal of cleaner production* 66 (2014): 507-519.

[3] Lu, Ning, Hua Lin, Jie Lu, and Guangquan Zhang. "A customer churn prediction model in telecom industry using boosting." *Industrial Informatics, IEEE Transactions on* 10, no. 2 (2014): 1659-1665.

[4] K. Coussement and D. V. Poel, "Churn Prediction in Subscription Services: An Application of Support Vector Machines while Comparing Two Parameter-Selection Techniques," *Expert Systems with Applications*, Vol. 34, No. 1, Jan. 2008, pp. 313-327.

[5] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach," *European Journal of Operational Research*, Vol. 218, No. 1, Apr. 2012, pp. 211-229.

[6] W. J. Reinartz and V. Kumar, "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, Vol. 67, No. 1, Jan. 2003, pp. 77-99.

[7] P. Datta, B. Masand, D. R. Mani, and B. Li, "Automated Cellular Modeling and Prediction on a Large Scale," Artificial Intelligence Review, Vol. 14, No. 6, Dec. 2000, pp. 485-502.

[8] D. Popoviæ and B. D. Bašiæ, "Churn Prediction Model in Retail Banking Using Fuzzy C-Means Algorithm," *Informatica*, Vol. 33, No. 2, May. 2009, pp. 235-239.

[9] C.-P. Wei and I. T. Chiu, "Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach," *Expert Systems with Applications*, Vol. 23, No. 2, Aug. 2002, pp. 103-112.

[10] M. Owczarczuk, "Churn Models for Prepaid Customers in the Cellular Telecommunication Industry Using Large Data Marts," *Expert Systems with Applications*, Vol. 37, No. 6, Jun. 2010, pp. 4710-4712.

[11] J. Burez and D. V. Poel, "Handling Class Imbalance in Customer Churn Prediction," *Expert Systems with Applications*, Vol. 36, No. 3, Apr. 2009, pp. 4626-4636.

[12] N. Kim, K.-H. Jung, Y. S. Kim, and J. Lee, "Uniformly Subsampled Ensemble (USE) for Churn Management: Theory and Implementation," *Expert Systems with Applications*, Vol. 39, No. 15, Nov. 2012, pp. 11839-11845.

[13] D. A. Kumar and V. Ravi, "Predicting Credit Card Customer Churn in Banks Using Data Mining," *International Journal of Data Analysis Techniques and Strategies*, Vol. 1, No. 1, Aug. 2008, pp. 4-28.

[14] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, “Computer Assisted Customer Churn Management: State-of-the-Art and Future Trends,” *Computers and Operations Research*, Vol. 34, No. 10, Oct. 2007, pp. 2902-2917.

[15] G. M. Weiss, “Mining With Rarity: A Unifying Framework,” *ACM SIGKDD Explorations Newsletter*, Vol. 6, No. 1, Jun. 2004, pp. 7-19.

[16] A. Payne and P. Frow, “A Strategic Framework for Customer Relationship Management,” *Journal of Marketing*, Vol. 69, No. 4, Oct. 2005, pp. 167-176.