

# Mining Patient Health Care Service Opinions for Hospital Recommendations

G.Sabarmathi<sup>1</sup>, Dr.R.Chinnaiyan<sup>2</sup>

Research Scholar, VTU-RC, Department of MCA, CMR Institute of Technology, Bengaluru-560037

<sup>1</sup>Assistant Professor, <sup>2</sup>Associate Professor

<sup>1</sup>Computer Science and Applications, Christ Academy Institute for Advanced Studies, Bangalore

<sup>2</sup>Department of Information Science and Engineering, CMR Institute of Technology, Bengaluru-560037

<sup>1</sup>vijayachinns@gmail.com, <sup>1</sup>chinnaiyan.r@cmrit.ac.in, <sup>2</sup>sabarganesh@gmail.com

**Abstract** - Patient's experience within the hospital environment is of paramount importance for the health care sector. Online reviews are recognized as the significant yardstick to scale the hospital's performance. This research proposes a novel machine learning-based ensemble classifier model to interpret the reviews in terms of patient's experience and for the hospital recommendation system. The outcomes were compared with various machine learning classifications using a cross-validation approach to predict the most accurate model. The predicted result brings out an interesting viewpoint that avails the healthcare sector an opportunity to look into their service offerings in improving patient's experience and hospital recommendation systems.

**Keywords** - Patient Reviews, Machine Learning, Classification, ensemble, recommendation

## I. INTRODUCTION

A rapid growth in technology necessitated a pool of smart devices usage. In the present scenario, the active smartphone user amounting to nearly 3 billion across the globe [1]. Nowadays, the majority of the customers are accustomed to express their opinions, reviews, tweets, and feedback on the social platform. Since it has various hidden components of fruitful information, it is worth looking into the opinions that carry meaningful patterns of sentiments that relate to customer's experience in different sectors [2-3]. The emotions of the patients are imitated based on many factors, and it can be the approach of the doctor, nurse, administrative staff, moral support, room facility, billing services, etc. These factors echo the opinion of the hospital as positive, negative, neutral. In this research work, we are going to analyze the social media reviews of various multi-specialty hospitals situated in and around Coimbatore, India, and aimed at proposing the best machine learning approach model for analyzing the data using classification models[39-47] to organize the sentiments of the patient as positive, negative and neutral.

## A. Problem Statement

Patient satisfaction is the key factor that impacts the healthcare business globally. In a study, it was seen that a US-based chain of hospitals faced a problem related to patient satisfaction, which impacts the growth of hospitals towards down. To solve this problem, the machine learning technique is used in analyzing the patient's opinions related to the services offered. The main challenges in handling these issues lie with a) mining the varied unstructured data from social media networks, b) removing duplication of reviews, and c) identifying forged reviews.

## B. Proposed Methodology

This section discusses data collection, data description, along pre-processing. It also describes the feature extraction methods used for the pre-processed data like tokenization, Bag-of words, constructing N-grams. Again, it talks about the proposed SCSP approaches and model implementation techniques. Let us discuss each division as mentioned below.

## II. SENTIMENT ANALYSIS

The sentiment analysis research work on hospital reviews will use the sequence of steps like data pre-processing, selecting relevant features using feature extraction, classifying the models based on test and train data set, checking the polarity of those reviews, and finally predicting the result based on accuracy and selecting the best model for this research along with recommendation percentage for recommending the hospitals based on the polarity of the reviews on the particular hospital. The objective of this paper is to evaluate the strength of the dependencies between the user rating of a review and the different text representations of the review given by the user.

## A. Data collection and pre-processing

The online reviews of various multi-specialty hospitals located in and around Coimbatore, India, have been considered for this study. The data retrieval has been accomplished with the help of an online API developed from the python script.



The criteria for this study are outlined in Table 1 and Table 2.

**Table 1. Inclusion condition**

| ID  | Condition  |
|-----|--|
| IN1 | Healthcare-related reviews were taken as input text  |
| IN2 | The input text captured from feelings and emotions of the affected individual (e.g., patient or carer) |
| IN3 | Natural language processing is used to analyze sentiment   |

**Table 2. Omission condition**

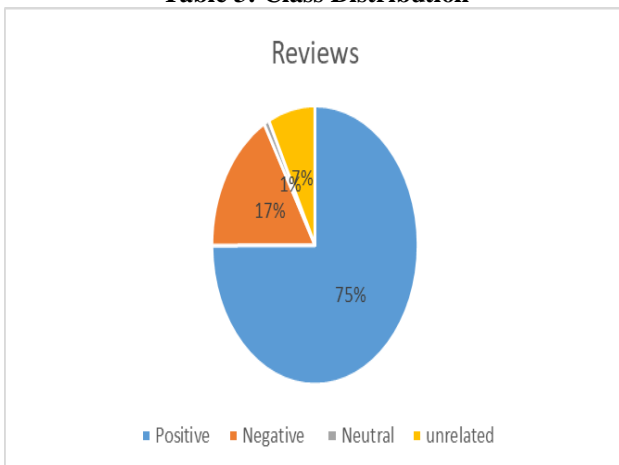
| ID  | Condition  |
|-----|--|
| OX1 | Sentiment analysis is executed in a language excluding English |
| OX2 | Reviews published before January 1, 2020                       |

Initially, reviews were collected in the form of excel format, and later it was converted as .csv format. Reviews collected have original reviews and repeated ones. We excluded the unrelated and named each review as positive, negative, and neutral classes manually that include original tweets and retweets. The reviews are collected from six multi-specialty hospitals in and around Coimbatore. The total number of reviews amounts to 1350 were collected and pre-processed as stated above, and it is represented in Table 3 and Figure 1. After pre-processing and removing duplications, 1250 reviews were forwarded to further processing by taking the reviews of classes positive, negative and neutral.

**Table 3 Class distribution**

| Class   | Positive | Negative | Neutral | Unrelated |
|---------|----------|----------|---------|-----------|
| Reviews | 1012     | 227      | 11      | 100       |

**Table 3: Class Distribution**



**Fig 1: Class Distribution on Reviews**

To train the model and classification, the distribution of the class should be balanced, and it is good to check the prior possibilities of unbiased that lead to an imbalanced distribution of class. We randomly resampled the reviews with the same number of reviews for each class. We got 1012, 227, and 11 reviews for each sentiment and 1250 reviews in total. Later we removed the unwanted information like url, punctuation, special characters from the reviews as these features do not provide any valid information in sentiment classification.

**III. PROPOSED MODEL**

- Step1: Data set is taken by exporting review from API using Python
- Step 2: Apply parsing to tokenize the input data
- Step 3: Apply N-gram algorithm to transfer the input data to feature vector
- Step 4: Create Bag of words and sampling
- Step 5: Split dataset into train and test dataset in the ratio of 80:20
- Step 6: Apply Machine Learning Classification Model
- Step 7: Generate the classification report on Precision, Recall, F1-score
- Step 8: Check for accuracy on models using Cross-Validation

**A. Feature Extraction**

Feature extraction is implemented in sentiment analysis after pre-processing the data. In social reviews, the texts are not written by following proper writing skills like correct spelling, grammar, etc., and many symbols, abbreviations may appear in the reviews. To address this issue, a technique called stemming has been implemented in this work where the various replications of words are stemmed together. For example, “likes” and “liked” can be stemmed to an identical stem word “like”. It also helps in reducing the feature sparsity.

In-text sentiment classifier, the features can be unigrams, bigrams, and many more. The idea of considering the N-gram is its ability to give more indication on various information on sentiment than unigrams. Here we have used N-gram and Corpus Approach using stop words and lemmatization for representing as weighted featured vectors. N-gram is a valuable approach in review analysis as it provides a sequence of words within a fixed window size ‘n’. The below Figure-2 shows the posted reviews of various healthcare organizations.

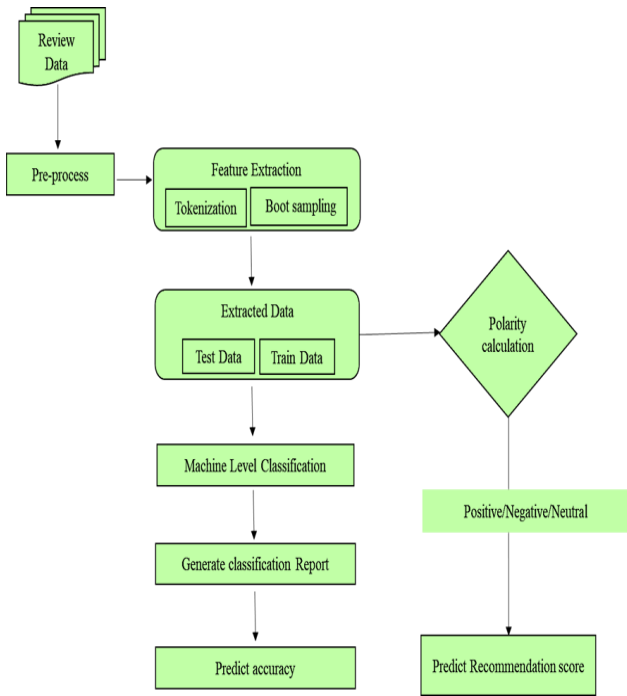


Fig 2: Proposed Architecture

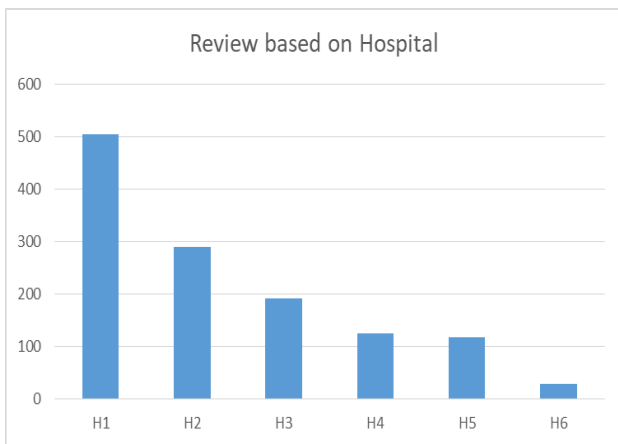


Fig 3: Reviews based on Hospital

Diagram Figure-4 depicts before and after the procedure of lemmatization and bag words using stop words.

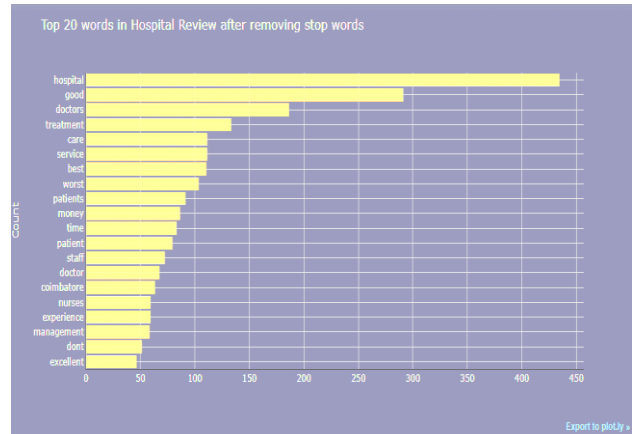
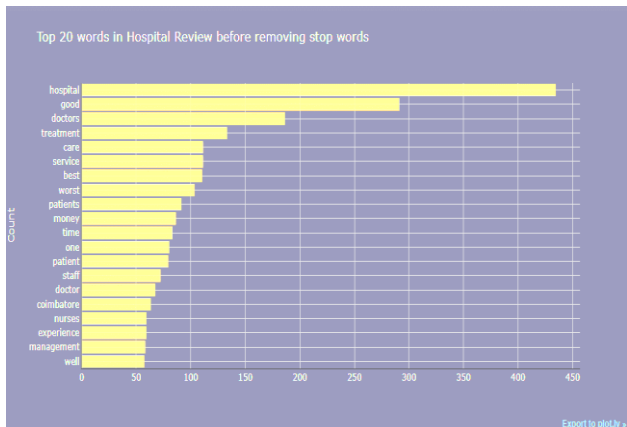


Figure-4: Top 20 words in-hospital review before and after removing stop words

Below Figure 5 show the usage of bigrams before and after stop words. Where two words are joined together to form bigrams

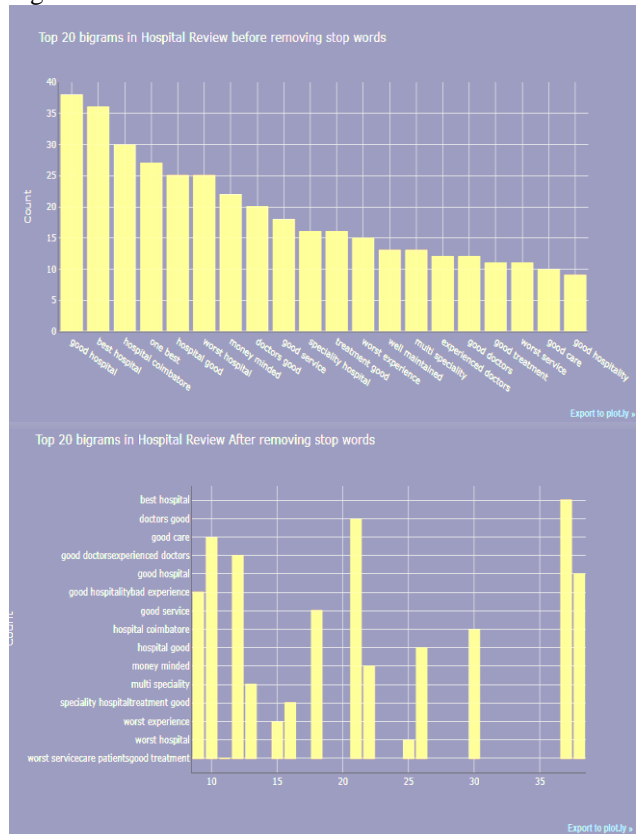


Figure-6: Top 20 bigrams in-hospital review before and after removing stop words

**B. Ensemble classification.**

The research is conducted with 7 classification models. 10-fold validation in cross-validation is used to evaluate the machine learning classification approaches, including SVM, LR, NB, RF, Decision Tree (DT), and KNN classifier. Test results for the 5 class and 3 class classification experiment are shown in Table 4 and Table 5. Since the decision tree gave the lowest accuracy compared to others and the bagging classifier gave the same accuracy as RF, both have been removed from comparison. The ensemble classifier model is working with the six sentiment classifiers to predict the

accuracy of the overall training dataset. It has been divided into many subsets and trained separately, as illustrated in the below Figure 8.

| Classifier | SVM | LR | NB | RF | KNN | DT |
|------------|-----|----|----|----|-----|----|
| Accuracy   | 63  | 64 | 65 | 60 | 59  | 60 |

**Table 4: Accuracy in 5 class and 3 class data set**

The ensemble classifier model adopts the method of majority voting for Bootstrap sampling for classifying each review. The six classifiers of the same weights have been considered in this process. While the classification of the sentiment of each review is classified separately by six classifiers, the final result predicts the outcome corresponds to each class which gets the maximum vote from six classifier output.

**IV. RESULTS AND DISCUSSION**

First, the reviews are converted into pre-processed data using N-gram and stop words, as shown in Figures 4 and 5. The data set performance is subjected to sentiment analysis that is categorized into the positive, negative, and neutral categories. In general, over-fitting happens when we train a small data set. To overcome this hurdle, cross-validation is employed in our study. It is a model validation method where samples are divided into two data subsets, one for model training and another for validating the model. As we used 10-fold validation, the data subsets are partitioned into 10 data subsets. To train the model, 9 data subsets were used, and the rest 1 is used for validating the model. The overall result obtained from this validation method is the average of 10 models. Testing is performed on both the 5-class data set and the 3-class data set. The 5-class data set includes outstanding, positive, neutral, negative, poor, and the 3-class data set consists only positive, neutral, and negative. The 3-class is formed from 5-class by merging outstanding with positive and poor with negative. Table 4 represents the accuracy scores of the 5-class and 3-class data sets. The model accuracy has been verified using various metrics, namely precision, recall, and F-measure.

The best model performance has been selected by using cross-validation of accuracy metrics of varied machine learning models. It is observed from Table 4 that the accuracy obtained from the present study is low when compared to the previous works in sentiment analysis [38]. The below Table 6 classification report of each sentiment classifier and Figure 7 shows the comparison of varied classifier models using cross-validation before ensemble and class binding methods. To improve the performance accuracy of reviews, we used the ensemble technique and class binding. In this case, the size of the data set is increased using the random sample method and found a relatively high accuracy score, which is shown in Table 7. Figure 8 shows the comparison of varied classifier models using cross-validation after ensemble and class binding methods.

| Classifier | SVM | LR | NB | RF | KNN | DT |
|------------|-----|----|----|----|-----|----|
| Accuracy   | 90  | 90 | 87 | 96 | 93  | 65 |

**Table 6: Model Classification before Ensemble**

| <b>SVM ACCURACY: 0.6370</b>          |           |        |         |
|--------------------------------------|-----------|--------|---------|
| Sentiments                           | Precision | Recall | F-Score |
| Positive                             | 0.83      | 0.83   | 0.83    |
| Negative                             | 0.45      | 0.47   | 0.46    |
| Neutral                              | 0.60      | 0.57   | 0.59    |
| <b>Logistic Reg ACCURACY: 0.6451</b> |           |        |         |
| Positive                             | 0.83      | 0.85   | 0.84    |
| Negative                             | 0.45      | 0.42   | 0.43    |
| Neutral                              | 0.59      | 0.62   | 0.60    |
| <b>Naïve Bayes ACCURACY: 0.8714</b>  |           |        |         |
| Positive                             | 0.91      | 0.95   | 0.93    |
| Negative                             | 0.84      | 0.77   | 0.80    |
| Neutral                              | 0.85      | 0.88   | 0.86    |
| <b>KNN ACCURACY: 0.5967</b>          |           |        |         |
| Positive                             | 0.70      | 0.83   | 0.76    |
| Negative                             | 0.47      | 0.53   | 0.50    |
| Neutral                              | 0.57      | 0.40   | 0.47    |
| <b>rainforest ACCURACY: 0.6048</b>   |           |        |         |
| Positive                             | 0.81      | 0.83   | 0.82    |
| Negative                             | 0.36      | 0.25   | 0.30    |
| Neutral                              | 0.54      | 0.67   | 0.60    |

**Table 7: Model Classification after Ensemble**

| <b>SVM ACCURACY: 0.9047</b>         |           |        |         |
|-------------------------------------|-----------|--------|---------|
| Sentiments                          | Precision | Recall | F-Score |
| Positive                            | 0.99      | 0.96   | 0.97    |
| Negative                            | 0.80      | 0.90   | 0.85    |
| Neutral                             | 0.92      | 0.85   | 0.88    |
| <b>Logistic Reg ACCURACY: 0.90</b>  |           |        |         |
| Positive                            | 0.99      | 0.96   | 0.97    |
| Negative                            | 0.79      | 0.89   | 0.84    |
| Neutral                             | 0.91      | 0.85   | 0.88    |
| <b>Naïve Bayes ACCURACY: 0.8714</b> |           |        |         |
| Positive                            | 0.91      | 0.95   | 0.93    |
| Negative                            | 0.84      | 0.77   | 0.80    |
| Neutral                             | 0.85      | 0.88   | 0.86    |
| <b>KNN ACCURACY: 0.7619</b>         |           |        |         |
| Positive                            | 0.82      | 0.91   | 0.86    |
| Negative                            | 0.64      | 0.77   | 0.70    |
| Neutral                             | 0.83      | 0.60   | 0.69    |

| rainforest ACCURACY: 0.9285 |      |      |      |
|-----------------------------|------|------|------|
| Positive                    | 1.00 | 0.97 | 0.99 |
| Negative                    | 0.88 | 0.92 | 0.90 |
| Neutral                     | 0.90 | 0.89 | 0.90 |

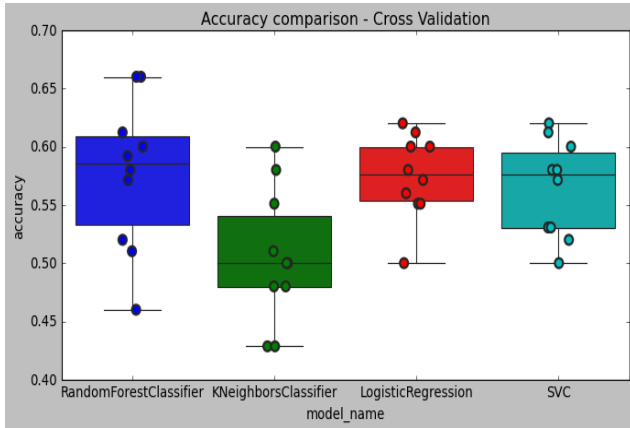


Figure 7: Comparison of Model classification based on Accuracy before the ensemble

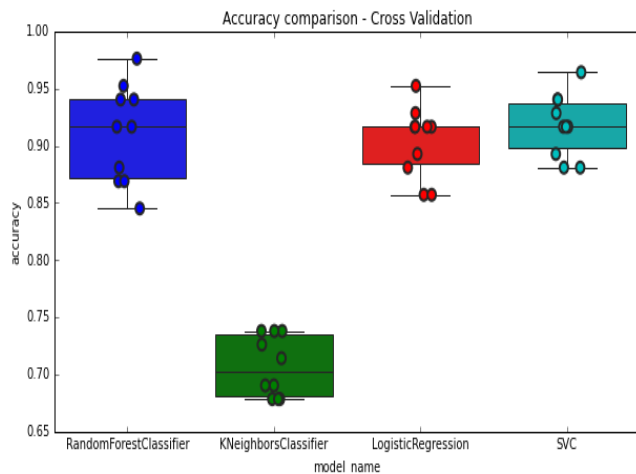


Figure 8: Comparison of Model classification based on Accuracy after ensemble

By comparing the results before and after ensemble, it observed that the RF classifier yields 93 % that outperforms the accuracy of the previous study in sentiment analysis of reviews in healthcare [38]. It assures our proposed method can be utilized to review health care sentiments with the highest level of accuracy.

**A. Model Validation**

The performance metrics of the model have been validated as per the below-described formula.

$$Accuracy = (TP + TN) / (TP + FP + TN + FN),$$

$$Precision (P) = TP / (TP + FP),$$

$$Recall (R) = TP / (TP + FN),$$

$$F-measure (F) = 2PR / (P + R)$$

With reference to a confusion matrix, TP refers to True Positive, TN refers to True Negative, FP refers to False positive, FN refers to False Negative based on the reviews posted. Below table 8 shows the different classification perform

ances of models using various ML algorithms. We used the indicators of precision, recall, and F-measure to measure the performance of sentiment classification. Cross-validation on accuracy has been used for comparing various models. The results of the sentiment classifier model after the ensemble technique are shown in above table 7.

The results are validated using the ROC curve based on various model prediction scores. The classifier that shows the closeness towards the top-left corner indicates a better performance of the model. As represented in Figure 9, the various sentiment classification models are given varied color codes. It is predicted that as per our calculation RF model demonstrates better performance compared to another model, which is substantiated by showing the AUC-ROC accuracy score of 95 % and the model accuracy of 93 %.

**B. ROC Curve**

The AUC-ROC accuracy is also engaged to measure the performance of classification models at different threshold settings. If AUC-ROC is higher, then the better the model prediction is. The various scores obtained from sentiment classifier model prediction metrics are depicted in Table 8 and Figure 10.

Table 8: Class Distribution on Reviews

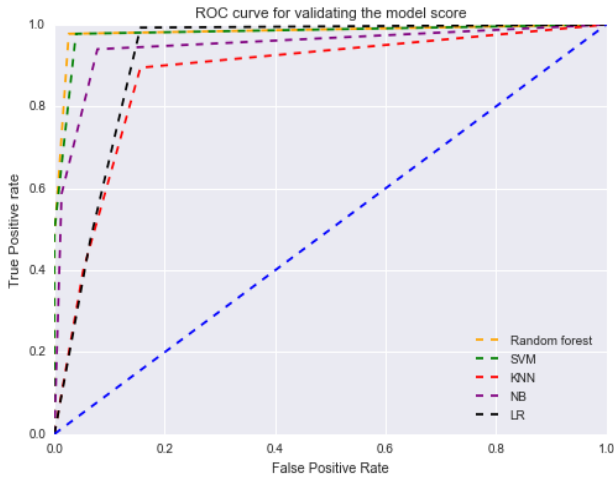
It is evident that the RF sentiment classifier model

| No   | Hospital Name | Positive | Negative | Neutral | Recommend % |
|--|---------------|----------|----------|---------|-------------|
| Class 1 Hospitals with highest percentage        |               |          |          |         |             |
| 1  | Hospital 1    | 424      | 80       | 1       | 83.96       |
| 5  | Hospital 5    | 98       | 18       | 2       | 83.05       |
| 3  | Hospital 3    | 155      | 35       | 1       | 81.15       |
| 4  | Hospital 4    | 102      | 21       | 3       | 80.95       |
| Class 2 Hospitals with second highest percentage |               |          |          |         |             |
| 6  | Hospital 6    | 23       | 3        | 3       | 79.31       |
| 2  | Hospital 2    | 219      | 70       | 1       | 75.51       |

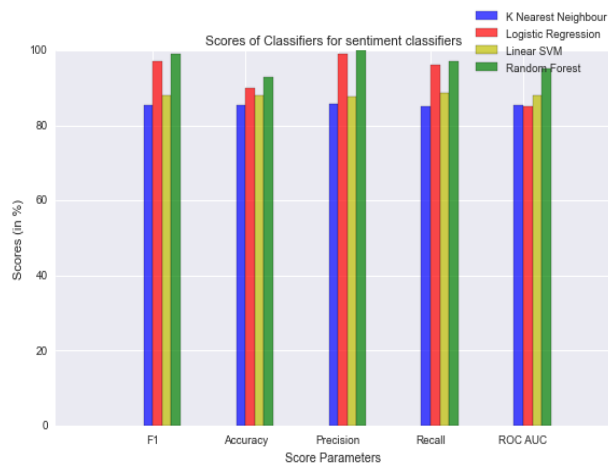
outstands with the highest performance than SVM as it scores 95 % accuracy on AUC-ROC compared to the previous analysis of the social media review published under the health care sector.

The results are analyzed by using the positive, negative, neutral ratings, and the recommendation percentage is calculated and shown in Table 9. The positive rating indicates it as excellent, while the negative represents a good hospital based on review polarity. The recommendation percentage is calculated by using below said formula





**Figure 9: ROC curve for validating the sentiment classifier model**



**Figure 10: Prediction metric scores for sentiment classifier models**

$$\text{Recommendation percentage} = \frac{\text{Total Positive reviews}}{\text{Total no. of reviews}} * 100$$

As mentioned earlier, 1250 reviews were collected from six multi-specialty hospitals in and around Coimbatore from google reviews using social sites. Considering all the reviews of six hospitals, we classified the reviews as shown in Table3. To suggest the recommendation among six hospitals, we categorize the hospital into two classes as mentioned below.

**Class 1:** The recommendation percentage greater than 80 is considered as first recommended class hospital

**Class 2:** The recommendation percentage lesser than 80 and greater than 75 is considered as the second recommended class hospital. The percentage below 75 is considered as the next third level recommendation; however, our study predicts only two classes based on the recommendation percentage. Using our model, we ranked the hospital and also provided a recommendation as stated. If we add more reviews to the dataset, the scope and reliability of the methodology can be tuned to higher accuracy.

**Table 9: Recommendation table for Hospital**

| Classifier Model | F1-Measure | Model Accuracy | Precision | Recall | AUC-ROC Accuracy |
|------------------|------------|----------------|-----------|--------|------------------|
| SVM              | 97         | 90             | 99        | 96     | 93               |
| LR               | 97         | 90             | 99        | 96     | 85               |
| NB               | 93         | 87             | 91        | 95     | 88               |
| KNN              | 86         | 76             | 82        | 91     | 81               |
| RF               | 99         | 93             | 100       | 97     | 95               |

### V. CONCLUSION

In this proposed research, a machine learning-based ensemble classifier model to analyze patient health care reviews is studied extensively. The recommendation percentage is measured by analyzing the sentiment polarity from the reviews. From the results and observations, it is obvious that proper choice of ensemble technique will definitely lift up the performance of the classification model. The review polarity recommendation augments the hope of patients in finding the best hospital within the vicinity and also provides an opportunity to the hospital management team to improvise their services that meets the patient requirements.

### REFERENCES

- [1] S.N. Manke, N. Shivale, A review on sentiment analysis mining and sentiment analysis based on natural language processing. *Int. J. Comput. Appl.* 109(4) (2015)
- [2] H. Mulki, et al., Tunisian dialect sentiment analysis: a natural language processing-based approach. *Computación y Sistemas* 22(4) (2018).
- [3] Abel F, Gao Q, Houben G-J, Tao K., Twitter-based user modeling for news recommendations. In: Rossi F (ed) *IJCAI 2013*, proceedings of the 23rd international joint conference on artificial intelligence, Beijing, China, (2013) 3–9., *IJCAI/AAAI*.
- [4] Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, et al. How should we define health? *Br Med J* 2011 Jul 26;343:d4163. [CrossRef] [Medline]
- [5] Berg O. Health and quality of life. *Acta Sociologica.*, 18(1) (1975) 3-22. [CrossRef]
- [6] A. Shoukry, A. Rafea, Sentence-level Arabic sentiment analysis, in 2012 International Conference on Collaboration Technologies and Systems (CTS), IEEE (2012) 546–550
- [7] Zomaya AY, Sakr S. *Handbook of big data technologies*. Berlin: Springer; (2017). <https://doi.org/10.1007/978-3-319-49340-4>
- [8] H. Iyer, M. Gandhi, S. Nair, Sentiment analysis for visuals using natural language processing. *Int. J. Comput. Appl.* 128(6) (2015) 31–35.
- [9] M.T. Khan, S. Khalid, Sentiment analysis for health care, in *Big Data: Concepts, Methodologies, Tools, and Applications*, IGI Global (2016) 676–689
- [10] M.M. Mostafa, N.R. Nebot, Sentiment analysis of Spanish words of Arabic origin related to Islam: a social network analysis. *J. Lang. Teach. Res.* 8(6) (2017) 1041–1049.
- [11] Denecke K, Deng Y. Sentiment analysis in medical settings: new opportunities and challenges. *Artif Intell Med.* 64(1) (2015) 17–27. doi: 10.1016/j.artmed.2015.03.006. [PubMed] [CrossRef] [Google Scholar].
- [12] Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: a review of the methods used. *JMIR Public Health Surveill.* 23 4(2) (2018) e43. doi: 10.196/publichealth.5789. <https://publichealth.jmir.org/2018/2/e43/> [PMC free article] [PubMed] [CrossRef] [Google Scholar].

- [13] L. Igual, S. Seguí, Statistical natural language processing for sentiment analysis, in *Introduction to Data Science* Springer, Cham, (2017) 181–197.
- [14] Nikfarjam A, Emadzadeh E, Gonzalez G. A hybrid system for emotion extraction from suicide notes. *Biomed Inform Insights* 5(Suppl 1) (2012) 165-174 [FREE Full text] [CrossRef] [Medline]
- [15] Wang W, Chen L, Tan M, Wang S, Sheth AP. Discovering fine-grained sentiment in suicide notes. *Biomed Inform Insights* 5(Suppl 1) (2012) 137-145 [FREE Full text] [CrossRef] [Medline]
- [16] Liakata M, Kim J, Saha S, Hastings J, Reholz-Schuhmann D. Three hybrid classifiers for the detection of emotions in suicide notes. *Biomed Inform Insights*;5(Suppl 1) (2012) 175-184 [FREE Full text] [CrossRef] [Medline].
- [17] Pedersen T. Rule-based and lightly supervised methods to predict emotions in suicide notes. *Biomed Inform Insights* 5(Suppl 1) (2012) 185-193 [FREE Full text] [CrossRef] [Medline].
- [18] Sohn S, Torii M, Li D, Wagholikar K, Wu S, Liu H. A hybrid approach to sentiment sentence classification in suicide notes. *Biomed Inform Insights*;5(Suppl 1) (2012) 43-50 [FREE Full text] [CrossRef] [Medline]
- [19] Wang W, Chen L, Tan M, Wang S, Sheth AP. Discovering fine-grained sentiment in suicide notes. *Biomed Inform Insights*;5(Suppl 1) (2012) 137-145 [FREE Full text] [CrossRef] [Medline]
- [20] Yu N, Kübler S, Herring J, Hsu Y, Israel R, Smiley C. LASSA: emotion detection via information fusion. *Biomed Inform Insights*;5(Suppl. 1) (2012) 71-76 [FREE Full text] [CrossRef] [Medline]
- [21] Chen L, Gong T, Kosinski M, Stillwell D, Davidson RL. Building a profile of subjective well-being for social media users. *PLoS One*;12(11) (2017) e0187278 [FREE Full text] [CrossRef] [Medline]
- [22] Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: a review of the methods used. *JMIR Public Health Surveill* Apr 23; 4(2) (2018) e43 [FREE Full text] [CrossRef] [Medline]
- [23] Kitchenham B. Procedures for performing systematic reviews. *Keele University, Keele*;33(2004) 1-26 [FREE Full text]
- [24] *Cochrane Library: Cochrane Reviews.* URL: <https://www.cochranelibrary.com/> [accessed 2019-11-12]
- [25] Paltoglou G, Thelwall M. A study of information retrieval weighting schemes for sentiment analysis. In: *Proceedings of the 48th annual meeting of the association for computational linguistics (ACL '10)*. (2010) 1386–1395.
- [26] Yessenalina YY, Cardie C. Multi-level structured models for document-level sentiment classification. In: *Proceedings of the 2010 conference on empirical methods in natural language processing (EMNLP '10)*. (2010) 1046–56.
- [27] Wilson T, Wiebe J, Hofmann P. Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of the conference on human language technology and empirical methods in natural language processing (HLT 05)*, Association for Computational Linguistics. Morristown, NJ, USA. (2005) 347–54.
- [28] Serrano-Guerrero J, Olivas JA, Romero FP, Herrera-Viedma E. Sentiment analysis: a review and comparative analysis of web services. *Inf Sci.* (2015) 311:18–38.
- [29] Rushdi-Saleh M, Martín-Valdivia M, Montejó-Ráez A, Ureña López L. Experiments with SVM to classify opinions in different domains. *Expert Syst Appl.* 38(12) (2011) 14799–804.
- [30] Ye Q, Zhang Z, Law R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Syst Appl.* 36(3) (2009) 6527–35.
- [31] He Y, Zhou D. Self-training from labeled features for sentiment analysis. *Inf Process Manag.*, 47(4) (2011)606–16.
- [32] Xianghua F, Guo L, Yanyan G, Zhiqiang W. Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowl Based Syst.* 37(2013) 186–95.
- [33] Kim K, Lee J. Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recogn.*, 47(2) (2014) 758–68.
- [34] König AC, Brill E. Reducing the human overhead in text categorization. In: *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining—KDD.*, 06. New York: ACM Press; (2006) 598–03.
- [35] Abualigah, Laith & Alfari, Hamza & Shehab, Mohammad & Hussein, Alhareth., *Sentiment Analysis in Healthcare: A Brief Review.* 10.1007/978-3-030-34614-0\_7., (2020).
- [36] Zunic A, Corcoran P, Spasic I *Sentiment Analysis in Health and Well-Being: Systematic Review JMIR Med Inform.*, 8(1) (2020) e16023 doi: 10.2196/16023PMID: 32012057PMCID: 7013658
- [37] Balachandar S., Chinnaiyan R., *Reliable Digital Twin for Connected Footballer.* In: Smys S., Bestak R., Chen JZ., Kotuliak I. (eds) *International Conference on Computer Networks and Communication Technologies. Lecture Notes on Data Engineering and Communications Technologies*, 15(2019) Springer, Singapore
- [38] Balachandar S., Chinnaiyan R., *Centralized Reliability and Security Management of Data in the Internet of Things (IoT) with Rule Builder.* In: Smys S., Bestak R., Chen JZ., Kotuliak I. (eds) *International Conference on Computer Networks and Communication Technologies. Lecture Notes on Data Engineering and Communications Technologies*, 15(2019) Springer, Singapore
- [39] G. Sabarmathi and R. Chinnaiyan., *Reliable Machine Learning Approach to Predict Patient Satisfaction for Optimal Decision Making and Quality Health Care*, *International Conference on Communication and Electronics Systems (ICES)*, Coimbatore, India, (2019) 1489-1493
- [40] G. Sabarmathi and R. Chinnaiyan., *Investigations on big data features research challenges and applications*, *International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, (2017) 782-786
- [41] G. Sabarmathi and R. Chinnaiyan., *Big Data Analytics Framework for Opinion Mining of Patient Health Care Experience*, *Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, (2020) 352-357
- [42] Sabarmathi G., Chinnaiyan R., *Envisagation and Analysis of Mosquito-Borne Fevers: A Health Monitoring System by Envisagative Computing Using Big Data Analytics.* In: Pandian A., Senju T., Islam S., Wang H. (eds) *Proceeding of the International Conference on Computer Networks, Big Data and IoT (CCBI - 2018)*. ICCBI 2018. *Lecture Notes on Data Engineering and Communications Technologies*, 31 (2020) . Springer, Cham.
- [43] R Chinnaiyan, S Somasundaram., *Monte Carlo Simulation For Reliability Assessment of Component-Based Software Systems*, *i-Manager's Journal on Software Engineering*, (2010).
- [44] R Chinnaiyan, S Somasundaram., *RELIABILITY ESTIMATION OF COMPONENT BASED SOFTWARE SYSTEMS THROUGH MARKOV PROCESS*, *International Journal of Mathematics, Computer Sciences and Information Technology*.
- [45] R.Chinnaiyan, S.Somasundaram., *Evaluating the Reliability of Component-Based Software Systems*, *International Journal of Quality and Reliability Management.*, 27(1) (2010) 78-88.