International Journal of Research in Engineering, Science and Management
Volume-3, Issue-6, June-2020
www.ijresm.com | ISSN (Online): 2581-5792

Comparative Study of Machine Learning Algorithms for Hospital Rating System

Varichitty Kalyan[1*], T. Jaswanth Reddy[2], S. V. Nikhilesh[3], A. G. Mowli Prakash[4]
[1,2,3]*Student, School of Computer Science and Engineering, VIT, Vellore, India*
[4]*Student, Department of Computer Science and Engineering, CEG, Anna University, Chennai, India*
*Corresponding author: varichitty.kalyan2017@vitstudent.ac.in

***Abstract*: In the current world, everyone depends on the review system. The review system is the method to extract information from a wide variety of available data. It has many applications on the Internet, which becomes a convenient place to give feedback about a product, to watch a movie. When any user is trying to find the best service in a certain hospital, it is very difficult to do so by going through each of them manually. This makes it difficult to track and consider the sentiment of the patients, which is why manual searching is not effective. In this work, the idea is to develop and implement the Hospital Rating and Review System that will be implemented on a website using natural language processing. The extraction of subjective knowledge in source material through the application of natural language processing is an emerging area of research.**

***Keywords*: Comments, Multinomial NB, NLP, Random Forest, Support Vector Classifier.**

1. Introduction

In the current world, the opinion of people has become one of the most important sources in the ever-growing popular social networks for various services. Online opinions have, in particular, turned into a kind of virtual currency for companies seeking for marketing their products, identify new opportunities, and manage their reputations. People's opinion has become one of the most important sources in the ever-growing popular social networks for various services. Online views, in particular, have turned into a kind of virtual currency for businesses seeking for marketing their products, identify good opportunities, and manage their reputations.

Practically, the users ask our trusted sources to recommend one when we are not familiar with a specific product. Today, the Internet's popularity encourages people to search for the views of other people on the Internet before buying a product or watching a film. Many websites offer services for user ratings and feedback, and these reviews may reflect the opinions of users about a product. For example, Amazon.com's customer review page lists the number of reviews, the percentage of different ratings, and reviewer comments. Such reviews and ratings typically affect their purchasing habits when people want to buy books, CDs, or DVDs.

Today everyone depends on the reviews before doing anything like going to movies if the review is the best. But the reviewing system is not perfect in our society. The reviewing system in this present digital world is biased to make the profit for certain people. In the hospitals, the people who don't know about the new hospitals visit them and suffering more. A smaller number of hospitals are working perfectly for the people's sake.

In this paper, we came up with the idea of a Hospital rating system using sentimental analysis where the search engine is another important source for people to look for the views of others in addition to these websites. When user enters a query in a search engine, the search engine checks its database and lists the best fitting web pages according to its requirements, usually with a summary containing the name of the report and sections of the text at times.

2. Literature Survey

The main models of these methods are SVM-based machine learning on training data and K-means cluster analysis on database inquiry result sets. The system takes personal information from the users and gives the training rating to each hospital [1], [2].

These hospitals are then clustered, and questions are produced to refine the recommendation. Finally, it recommends patients to choose the best hospital for their medication [3], [4]. Through the design of the method, assessing success is not a simple job because there is no correct or wrong recommendation; it's only an opinion issue. He has earned favorable feedback from users based on the researcher's informal evaluations. A broader collection of data will make for more accurate findings for this method. Ultimately, this work wants to implement a web- based user interface with a user database, and the training template is customized to each user [5].

For other applications, the authors applied data mining to carry out the moving model score. Several hospital attributes, such as doctors, specialists, are of the hospital building, are being researched. Study-based, it is found that doctors and hospitals are the most important attributes. The technique being studied is described. The database of hospital information is created. To create a decision tree, this data is then collected, cleaned, and transformed into' Weka.' The difficult part is the

removal of terms that influence the score in the subtitles / short stories, keyword choice, and keyword classifications [6].

In the use of genre and word category properties, the generated0 model will achieve approximately 80 percent accuracy. The concept is incorporated in the software application for storing details about films and proposed the moving score. The system for collecting classification content may also be extended to take into account other characteristics and image processing techniques [7].

In the other proposed models, Rotten Tomato Reviews is gathered from one database. Then the tokens are filtered by the length of each review tokenization. After this stemming is completed, delete tokens that are not necessary for the analysis of sentiment [8,9].

Multiply operator that compares each token with the positive word dictionary and the negative word dictionary. If the given token matches any of the word dictionaries, the token will be categorized as such [10], [11]. All occurrences in both positive databases and negative databases after that sums up. Apply the join operator, which eliminates the positive and negative-sum and creates the analysis class tag and shows it to the user [12].

The authors have shown their research that the extraction of people opinions on components of an entity is the important task of opinion gathering. To generate a valuable summary, it is important to combine such terms and phrases, which are domain synonyms, into the same feature category. There is also a problem in the features and opinions perception relationship [13], [14]. A novel approach is introduced to cope with feature-level opinion mining problems. The specific features and implied features are included in the introduced study. The classified divisions of opinion words are ambiguous opinion words and simple opinion words, which seek to discover the hidden features and clutch the features [15].

The catching of features depends on three aspects: associated opinion terms, feature similarity, and feature formation [16]. In fact, background knowledge is used to improve the clutching in the process, which is seen to be helpful in clustering or clutching. Opinion mining at function level through three

phases (1) Extract the features of the comments and the corresponding opinions (2) Cluster the features of the comments (3) Orient the opinions of the feature [17], [18].

3. Comparative Study

In these work three machine learning algorithms are compared to get the best out of it. The best algorithm out of these will be implemented in the hospital rating system for the sake of sentimental analysis.

A. Support Vector Classifier

Linear svc is a method of support vector machine. Support Vector Machine is a linear model for problems with classification and regression. It can do both linear and non-linear problems and function well on other functional problems. SVM's idea is simple: The algorithm generates a line or hyperplane that divides the data into sections.

The main motto of SVC is to suit the data you have, returning the best fit hyperplane that separates the data or categorizes them. From there, you can then feed some features to your classifier after having the hyperplane to see what is predicted class is. This makes the algorithm very ideal for our needs, though this can be used in other circumstances.

Linear SVC is one of the algorithms that perform relatively well on the spectrum of text classification tasks dependent on NLP. However, if the prerequisite is to have a probability distribution for all groups, Linear SVC does not have a function like predict probability out of the box in scikit-learn. Linear SVC provides a method of decision function. The confidence score for a sample is the distance signed to the hyperplane by that sample.

Applications of this algorithm are text analysis, data classification, face detection, etc.

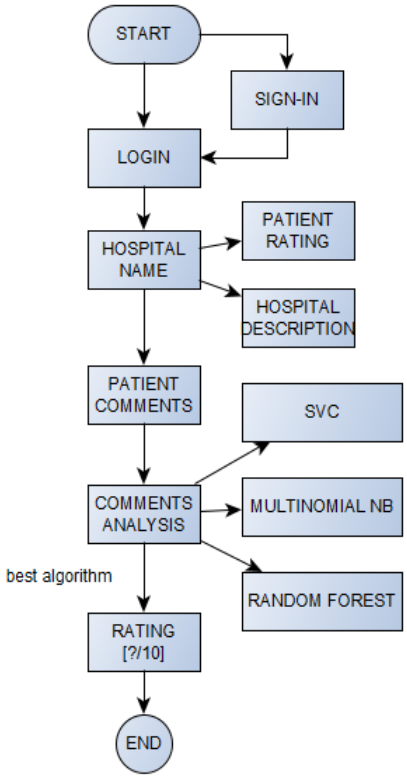

Fig. 1. Flowchart of the website to give rating by comment analysis

B. Multinomial NB

Multi-naïve Bayes is a type of naive Bayes algorithm. Naïve Bayes Classifier Algorithm is a group of probabilistic algorithms focused on applying Bayes' theorem with the "naïve" assumption that each pair of a function is conditionally independent. Bayes theorem calculates probability P(a) where a is the class of potential outcomes, and b is the defined instance to be categorized, representing some of the characteristics.

$$P(a)=P(b)*P(a)/P(b) \quad (1)$$

Naive Bayes is often used in the analysis of natural language (NLP) problems. Naive Bayes foresees a text name. They measure each tag's likelihood for a given document and then output the tag to the maximum. Multinomial Naive Bayes is a generalized variant of Naive Bayes, which is more narrowly designed for text documents, whereas basic naive Bayes will model a text as the inclusion and absence of different terms, multinomial naive Bayes directly model the word counts and change the underlying formulas to be discussed in. In documents belonging to class(c), it calculates the conditional likelihood of a particular word given to a class as the relative frequency of term d. The variance takes into account the number of terms d occurrences in class (c) instruction materials, including multiple occurrences

The Naive Bayes multinomial classifier is ideal for classifying with discrete features (e.g., word counts for text classification). Normally, the multinomial distribution includes the count of integer functions. However, fractional counts, such as tf-idf, will work in practice as well.

Applications of multi-naive bayes algorithm are real-time predictions, multi-class predictions, text classifications, text analysis, spam filtering, recommendation system etc.

C. *Random Forest*

Random forest is a supervised learning algorithm. It can be used for the classification and also regression. It's also the algorithm with the most versatility and ease of use. One forest is made up of decision trees. The more trees she has, the healthier a forest is, it is said. Random forests build trees on randomly selected samples of data, get predictions from each tree, and choose the best solution by means of voting. A big benefit of random forest is that it can be used for problems of classification as well as regression, which form the majority of existing machine learning systems. Let's look at the classification of random trees, as the classification is often called the building block of machine learning.

Random forests have exactly the same hyper parameters as a decision tree or a classifier for bagging. Luckily there is no need to pair a decision tree with a bagging classifier, because you can simply use the random forest classifier. You can also perform regression tasks with random trees, using the regressor algorithm. random forest algorithm is used in applications like healthcare, e- commerce, finance etc.,

Random forests have exactly the same hyper parameters as a decision tree or a classifier for bagging. Luckily there is no need to pair a decision tree with a bagging classifier because you can simply use the random forest classifier. You can also perform regression tasks with random trees, using the regressor algorithm. Random forest algorithm is used in applications like healthcare, e-commerce, finance, etc.

These three algorithms are compared in terms of accuracy, precision, etc., The best algorithm out of these will be used in the implementation of sentimental analysis for the hospital rating system.

4. Results

After giving comments to the three algorithms it produces the below results and from the results, we can choose the best algorithm for the rating the comments in the hospital rating system.

A. *Support Vector Classifier*

In fig. 2. It shows the results produced by the Support Vector classifier after giving comments to the algorithm

	precision	recall	f1-score	support
1	0.52	0.68	0.59	5022
10	0.49	0.64	0.55	4999
2	0.18	0.13	0.15	2302
3	0.22	0.17	0.19	2541
4	0.28	0.27	0.28	2635
7	0.25	0.23	0.24	2307
8	0.23	0.18	0.20	2850
9	0.20	0.12	0.15	2344
accuracy			0.38	25000
macro avg	0.30	0.30	0.29	25000
weighted avg	0.34	0.38	0.35	25000

```
[[3434  255  530  349  310   56   54   34]
 [ 242 3216   48   94  120  312  502  465]
 [1105  145  294  334  309   56   38   21]
 [ 858  150  348  431  473  150   94   37]
 [ 581  201  262  443  715  230  149   54]
 [ 154  554   62  151  282  542  388  174]
 [ 150 1007   45   98  233  524  499  294]
 [ 112 1075   29   59  111  274  406  278]]
```

Fig. 2. Results produced by SVC algorithm

B. *Multinomial NB*

	precision	recall	f1-score	support
1	0.40	0.87	0.54	5022
10	0.35	0.88	0.50	4999
2	0.32	0.01	0.01	2302
3	0.41	0.02	0.03	2541
4	0.32	0.05	0.08	2635
7	0.28	0.03	0.05	2307
8	0.23	0.05	0.08	2850
9	0.34	0.01	0.01	2344
accuracy			0.37	25000
macro avg	0.33	0.24	0.16	25000
weighted avg	0.34	0.37	0.24	25000

```
[[4378  592    4    6   22    2   16    2]
 [ 441 4424    2    2   19   16   89    6]
 [1794  434   15    9   29    5   14    2]
 [1714  637    4   43   92   16   33    2]
 [1515  854   11   26  127   39   60    3]
 [ 529 1488    5   10   61   62  146    6]
 [ 439 2188    3    5   33   49  129    4]
 [ 271 1921    3    4   20   30   82   13]]

In [53]:
```

Fig. 3. Results produced by Multinomial NB algorithm

In fig. 3. It shows the results produced by Multinomial NB algorithm after giving comments to the algorithm

C. *Random forest*

In fig. 4. It shows the results produced by Random forest algorithm after giving comments to the algorithm.

	precision	recall	f1-score	support
1	0.56	0.66	0.61	5022
10	0.55	0.56	0.56	4999
2	0.21	0.16	0.18	2302
3	0.23	0.17	0.19	2541
4	0.26	0.31	0.28	2635
7	0.26	0.25	0.25	2307
8	0.25	0.24	0.24	2850
9	0.21	0.21	0.21	2344
accuracy			0.38	25000
macro avg	0.32	0.32	0.32	25000
weighted avg	0.36	0.38	0.37	25000

```
[[3311  142  535  386  415   68   83   82]
 [ 278 2811   74   80  157  255  592  752]
 [ 880   97  375  352  417   65   63   53]
 [ 636  103  342  420  669  169  122   80]
 [ 399  103  292  355  820  326  240  100]
 [ 128  318   76  100  333  566  510  276]
 [ 132  693   51   81  256  476  679  482]
 [ 108  850   41   44  117  222  480  482]]
```

Fig. 4. Results produced by multinomial NB algorithm

From the above results, it is proved that SVC (Support Vector Classifier) is best algorithm in terms of precision, accuracy and other extra parameters for the analysis and SVC algorithm is used in the website for the comments analysis from the patients.

5. Conclusion

This work implementation was carried out on Hospital reviews data. Due to the availability of a huge volume of user-generated content in review sites, forums and blogs, Sentiment Analysis, and Opinion mining has become a fascinating research area. Sentiment Analysis has applications from market research to decision making and advertising in a variety of fields. With the aid of Sentiment Analysis, businesses can determine the degree of brand recognition and can develop strategies to boost their product. Before, there was no online platform for reviewing hospitals with the comments given by the patients. From this website, the patients can share their comments about the hospitals they visited. These comments are used for analysis, and the rating is given to the respective hospitals. Patients can utilize this website to make the right choice of hospitals. The development of more effective algorithms focused on subject models in many other opinion mining systems and for large-scale data sets is essential for potential research.

References

[1] M. Yan, T. Qu, C. Li and S. Xu, "Impacts of health information technology on health care quality in hospital-related settings: A systematic review," 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), Zhuhai, 2018, pp. 1-4.

[2] L. Li and X. Zhou, "Analysis on archetype generating set of hospital competitiveness in Jiangxi province based on index evaluation system," 2010 IEEE 17Th International Conference on Industrial Engineering and Engineering Management, Xiamen, 2010, pp. 1155-1160.

[3] D. Sato and A. Freitas, "Understanding quality in data from hospital diagnosis and procedures: A statistical characterization," 5th Iberian Conference on Information Systems and Technologies, Santiago de Compostela, 2010, pp. 1-4.

[4] S. Swarnalatha, I. Kesavarthini, S. Poornima and N. Sripriya, "Med-Recommender System for Predictive Analysis of Hospitals and Doctors," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-5.

[5] Y. Shih, "User satisfaction with HIS outsourcing," 2010 7th International Conference on Service Systems and Service Management, Tokyo, 2010, pp. 1-4.

[6] P. Frisch, "What Is an Intelligent Hospital?: A place where technology and design converge to enhance patient care.," in IEEE Pulse, vol. 5, no. 6, pp. 10-15, Nov.-Dec. 2014.

[7] Nsamzinshuti, C. Van Elslande and B. A. Ndiaye, "Managing hospital efficiency and effectiveness: Designing a hospital performance measurement system," 2014 International Conference on Logistics Operations Management, Rabat, 2014, pp. 188-200.

[8] K. Jenoui and A. Abouabdellah, "Implementation of a decision support system heuristic for selecting suppliers in the hospital sector," 2015 International Conference on Industrial Engineering and Systems Management (IESM), Seville, 2015, pp. 625-632.

[9] T. Liu, Y. Ma and X. Yang, "Service Quality Improvement of Hospital Reservation System Based on Text Sentiment Analysis," 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, 2018, pp. 289-293.

[10] M. E. Valentinuzzi and R. Leder, "The modern hospital in historical context," 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, 2009, pp. 1089-1091.

[11] J. Wills and K. Hurley, "Testing usability and measuring task-completion time in hospital-based health information systems: A systematic review," 2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Montreal, QC, 2012, pp. 1-4.

[12] Q. Zeng, J. Du and K. Tang, "Medical Records Service Management in the Hospitals," 2010 International Conference on Internet Technology and Applications, Wuhan, 2010, pp. 1-4.

[13] Edward Kai-Ning Yung and I. C. W. Leung, "An active RFID system for hospitals," 2009 Applied Electromagnetics Conference (AEMC), Kolkata, 2009, pp. 1-4.

[14] S. Wang et al., "A New Smart Mobile System for Chronic Wound Care Management," in IEEE Access, vol. 6, pp. 52355-52365, 2018.

[15] Mangal Sain, HoonJae Lee and Wan-Young Chung, "Middleware for ubiquitous Healthcare Information system," 2009 11th International Conference on Advanced Communication Technology, Phoenix Park, 2009, pp. 2325-2328.

[16] J. F. Kolzer and R. Garcia, "Clinical engineering in medical technology assessment," Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439), Cancun, 2003, pp. 3640-3643 Vol.4.

[17] C. Meng, H. Zhang, L. Zeng, Z. Li, J. Huang and Z. Liang, "Evidence-based decision support for the clinical practice of acupuncture: Data mining approaches," 2013 IEEE International Conference on Bioinformatics and Biomedicine, Shanghai, 2013, pp. 180-181.

[18] S. Kumar and B. K. Singh, "A Review of Digital Watermarking in HealthCare Domain," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 156-159.