**Aspect Extraction for evaluating hospitals and its visualization**

|  |  |
| --- | --- |
| [1]Deepsi Kumari  School of Computer Science and Engineering  Vellore Institute of Technology,Chennai  deepsi.kumari2019@vitstudent.ac.in | [2]Dr. Sathis Kumar B.  School of Computer Science and Engineering  Vellore Institute of Technology,Chennai  sathiskumar.b@vit.ac.in |

**ABSTRACT**

Healthcare consumers nowadays rely on various online resources, such as hospital review sites like Yelp and Mouthshut, as well as social media, to make informed decisions when selecting a hospital or clinic for consultation. This paper investigates the use of opinion mining on textual data extracted from hospital review sites, and proposes the use of data visualization tools like PowerBI or Tableau to aid patients in identifying good hospitals easily. The visualization techniques presented in this paper include faceted and filtered visualization, which allow for a more advanced analysis of the sentiment expressed in the reviews. Our evaluation of the proposed techniques shows a high level of accuracy in opinion mining and detection of pain points, which are crucial for understanding the patients' sentiments and reasons for changes in their opinions. Moreover, the prototype developed in this study can help identify both the strengths and weaknesses of hospitals' hospitality services.

# **Keywords**

Semantic analysis, feature extraction, temporal opinion mining

# INTRODUCTION

We live in a Golden Age of Information where, with product information and reviews available across several outlets, making informed choices is a piece of cake.. The contents on such reviews websites is user-generated, thus giving access to the opinions of many individuals. When contributing opinions to the web- sites, users typically select grades for a a number of facets, and additionally add a textual review. During subsequent search of hospitals,users get a ranked list of hospitals, where ranking is based on the grades given by the patients or their relatives.

When studying existing websites,social medias and previous research in the area, two observations can be made: 1) the visualization of the hospitals is not quite primitive that shows the declining or rising reviews trend or other benefits of hospitals with the help of charts but just shows the direction on google maps, and

2) the only use of the textual descriptions is for browsing, they are not part of the ranking process or visualized. To our knowledge, these issues have not been studied before. In this paper we also describe how to use opinion mining techniques to analyze changes in opinions about hospitals and gathering pain points. Hospitals are seen as a public-based service is an area where multiple factors may impact patients or their relatives sentiment. For instance service, doctor’s punctuality, cleanliness, staffs behavior and more such events may influence the overall sentiment at any given time, creating a dynamically changing sentiment. Managing to identify why changes occur in such a setting, may provide both patients and hospital administration some valuable information regarding the interpretation of large amounts of opinionated data.

Opinion mining tools are used to identify and extract subjective information from user reviews, and then to determine the sentiment of the text. Two different techniques are studied: feature extraction and visualization. Feature extraction is a technique to identify and extract product features and extract the pain points and visualization means to present the information graphically. Evaluation is performed by comparing the actual review scores with our sentiment scores.

In order to perform visualization experiments, a web prototype was created. This provides a way to detect "good" and "bad" aspects based on the hospital reviews in a user- friendly interface. These scores are calculated based on user opinions, and is an effective way for users to filter sentiment data. For commercial use, the prototype can help analyze the massive amount of hospital information published each day by users, and can help hospital administration analyze their services. It can also be used as a more advanced hospital search engine where users can find extra information in a map user interface that can serve as a benefit for the people who don’t want to read long passages.

# PROBLEM STATEMENT

In today's world, consumers have access to a wealth of information about products and services, including online reviews from other customers. However, with this abundance of information comes the challenge of sorting through and understanding it all. This can be particularly difficult in the healthcare industry, where hospital reviews and ratings are important sources of information for both patients and healthcare organizations.It is encouraged to create a model, powered entirely by publicly available data, that would cater to a wide range of users looking for information regarding reviews of Hospitals/Laboratories etc., The whole visualization process meant for patients choosing a certain hospital in their city for a medical procedure, or hospital administrators looking for insights into improving their services with the help of pain points generated.

# RELATED WORK

In this section, our focus is on two main topics of this paper: temporal opinion mining and visualization. The selected references covered the period from 2013 to 2022 and highlighted the diverse approaches, methodologies, and findings of the studies. The review showed that sentiment analysis has become a valuable tool for extracting insights from large amounts of unstructured data and can be applied in various fields, including healthcare, social media, customer service, and marketing.

Cobb et al. (2013) conducted sentiment analysis to evaluate the impact of online messages on smokers' choices to use varenicline. The study revealed that negative messages about the drug influenced the smokers to avoid using it. Ebrahimi et al. (2016) identified the recognition of side effects as implicit-opinion words in drug reviews. They used sentiment analysis to determine the sentiment of the reviewers towards specific side effects, which can aid in improving drug safety. Rastegar-Mojarad et al. (2015) analyzed patient experiences of healthcare from social media using sentiment analysis. The study found that social media platforms can be used to collect and analyze patient experiences of healthcare, which can help in improving healthcare services.

Abirami and Askarunisa (2017) used sentiment analysis to emphasize the impact of online reviews in the healthcare industry. The study found that positive reviews significantly influenced patients' decisions to visit healthcare facilities. Zakkar and Lizotte (2021) analyzed patient stories on social media using text analytics. The study found that text analytics can be used to extract valuable information from patient stories on social media, which can aid in improving healthcare services.

In the customer service and marketing domain, Kelly Aponté and Shiseida Sade (2020) reported that customer pain points can be identified and addressed through sentiment analysis of customer feedback. They emphasized the importance of addressing customer pain points to enhance customer satisfaction and loyalty. Tao et al. (2019) used sentiment analysis to mine pain points from hotel online comments. The study revealed that sentiment analysis can be used to identify the areas of concern for customers, which can help in improving the quality of services.

Tammina and Annareddy (2020) applied convolutional neural network (CNN) for sentiment analysis of customer reviews. The study found that CNN outperformed traditional machine learning algorithms in sentiment classification. Salminen et al. (2022) developed a machine learning model to detect pain points from user-generated social media posts. The study found that the model achieved high accuracy in identifying pain points, which can aid in improving customer satisfaction.

Ahmad et al. (2019) used sentiment analysis techniques to detect and classify social media-based extremist affiliations. The study found that sentiment analysis can be used to identify extremist affiliations and prevent radicalization.

Praphula Kumar Jain et al. (2021) conducted a systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. The study found that machine learning algorithms have been extensively used for consumer sentiment analysis and have shown promising results.

In conclusion, sentiment analysis has become an important tool for extracting valuable insights from unstructured data in different domains. The studies reviewed in this literature review demonstrated the effectiveness of sentiment analysis in healthcare, social media, customer service, marketing, and security. The findings of these studies can aid in improving the quality of services and enhancing customer satisfaction and loyalty. Future research can focus on developing more advanced sentiment analysis techniques to extract deeper insights from unstructured data.

1. THEORY

4.1. ASPECT EXTRACTION

Aspect-based sentiment analysis is a powerful tool that can be applied to various domains, including product reviews, social media sentiment analysis, and customer feedback analysis. It allows for a more granular analysis of feedback by identifying specific aspects of a product or service that are driving the overall sentiment.

In the context of hospital reviews, aspect extraction can help hospital administrators and healthcare providers to identify areas where they are performing well and areas that need improvement. For example, if the aspect "staff behavior" has a negative sentiment score, it can indicate that patients are dissatisfied with the behavior of the hospital staff. This feedback can be used to improve staff training programs or to address specific issues with staff behavior.

Overall, the Python code implements aspect-based sentiment analysis to extract key aspects and their corresponding sentiments from hospital reviews. This approach can provide valuable insights into patient experiences and help healthcare providers improve the quality of care they provide.

There are several algorithms used for feature extraction in sentiment analysis, some of which are:

Bag of Words (BoW) - BoW is a common algorithm used in sentiment analysis. It involves counting the frequency of words in a text corpus and using these counts as features for a machine learning model.

Term Frequency-Inverse Document Frequency (TF-IDF) - TF-IDF is another popular algorithm used in sentiment analysis. It involves measuring the importance of a word in a document corpus by taking into account its frequency in the document and the frequency of the word across the corpus.

Word2Vec - Word2Vec is a neural network-based algorithm that converts words into dense vectors. It can capture the semantic relationships between words and can be used as features for a sentiment analysis model.

Latent Dirichlet Allocation (LDA) - LDA is a topic modeling algorithm that can identify latent topics within a text corpus. These topics can then be used as features for a sentiment analysis model.

Non-negative Matrix Factorization (NMF): NMF is a matrix factorization technique that factorizes a non-negative matrix into two non-negative matrices. In topic modeling, one matrix represents the document-topic distribution, while the other represents the topic-word distribution. NMF has been shown to perform well for short and sparse text data.

Latent Semantic Analysis (LSA): LSA is a matrix factorization technique that uses Singular Value Decomposition (SVD) to identify latent topics in a corpus. It is similar to NMF, but it can handle negative values in the matrix. LSA has been used for applications such as information retrieval and text classification.

4.2. VISUALIZATION

In order to be able to evaluate my ideas in a controlled fashion, I developed a prototype for visualizing sentiment changes over time. With the data sets we gathered, we had all the data we needed to look at sentiment changes over time.

An interactive dashboard is created where one can visualize all aspect of the hospital he or she wants to know about.It shows the specialities of the hospitals too.As this saves time for a user by exempting them to waste a lot of time on reading reviews ,so the dashboard created will give an overview of good and bad sentiments of the hospital.

1. PROPOSED SYSTEM

Preprocessing: First, we preprocess the text data by removing stop words, punctuation, and other noise, and then represent the reviews as a bag-of-words model, where each review is represented as a vector of word frequencies.

Model Training: LDA is a generative probabilistic model that identifies the underlying topics in the data by assuming each document is a mixture of multiple topics, and each topic is a distribution over words. The LDA model is trained by iteratively estimating the topic distribution for each document and the word distribution for each topic based on the observed word frequencies in the data.

Topic Interpretation: Once the LDA model is trained, we can interpret the resulting topics to identify the aspects mentioned in the reviews. This involves looking at the most frequent words in each topic and using your domain knowledge to identify the aspects that those words represent.

Aspect-Based Extraction: After identifying the topics and corresponding aspects, we can use the topic proportions for each review to extract aspect-based features. The topic proportions represent the degree to which each review is associated with each aspect.

Sentiment Analysis: Finally, we can use sentiment analysis techniques to determine the sentiment expressed for each aspect mentioned in the reviews. This involves using lexicons or machine learning models to identify the sentiment expressed in each review, or aggregating the sentiment expressed for the relevant terms within each aspect to form aspect-level sentiment scores.

As the objective of this project is to develop a system that can categorize aspects of hospitals from user reviews,so the system should use feature extraction techniques such as Bag of Words, TF-IDF, and Topic Modeling (e.g., LDA,LSA) to identify the most important aspects of a hospital that are mentioned in user reviews. The identified aspects can then be used to know about the quality of the services and their aspects.

5.1. DATA COLLECTION

A large dataset of hospital reviews is collected containing ratings and reviews by scraping online platforms such as Yelp, Google, Healthgrades, etc.All the datasets are combined into one dataset for a large amount of reviews and further experiment.

5.2. PRE – PROCESSING

As the datset scraped from different websites contains stop words,punctuations and other regular expression,and other noise,so the dataset is preprocessed by removing irrelevant information, data redundancy, and performing text normalization techniques such as removing punctuation marks, stop-word removal, tokenization, stemming/lemmatization. Lemmatization is the process of reducing a word to its base or dictionary form. Lemmatization can be performed using various techniques, including rule-based approaches, lookup tables, and machine learning algorithms.While preprocessing,five star ratings and 1 star ratings are encoded as 1 and 0 respectively so that the best and the worst aspect can be known and other common words used in the best and worst reviews to know the pain points.The bag-of-words model is a simple yet effective way to represent text data in a machine-readable format. It counts the frequency of each word in a piece of text and represents the text as a vector of word frequencies. This can help to identify the important words or aspects of a given text by comparing the frequency of words across different texts. In aspect extraction, the bag-of-words model can be used as a starting point for identifying the aspects of a product or service that are being discussed in customer reviews or other types of text data. By analyzing the frequency of different words and phrases in the text, it is possible to identify the aspects that are being discussed and gain insights into the opinions and attitudes of customers.

5.3. ASPECT EXTRACTION

5.3.1 DOC2VEC

The objective of this code is to create doc2vec vector columns for a set of text data, using the gensim library in Python. It does this by initializing and training a Doc2Vec model with the text data, and then transforming each document in the dataset into a vector representation using the trained model. The resulting vectors are then added as new columns to the dataset. The code uses the TaggedDocument class to prepare the text data for training, and sets various parameters for the Doc2Vec model such as the vector size, window size, and minimum count. Finally, the code prints the type of the resulting dataset "final".

The code is useful in aspect extraction from hospital reviews as it helps to convert the textual data of the reviews into a numerical representation (vector) that can be processed by machine learning algorithms. The Doc2Vec model used in the code is a neural network-based technique that represents each document (i.e. hospital review) as a dense vector, which captures the semantic meaning of the text. By using this model, we can learn vector representations for each hospital review that are similar for reviews that discuss the same aspects and different for reviews that discuss different aspects.

5.3.2. TF-IDF

The code adds tf-idf columns to a dataset of hospital reviews, which can be useful in aspect extraction. The tf-idf score represents the importance of each word in the document relative to its frequency in the entire collection of documents. Words with high tf-idf scores are those that are more relevant to the specific document and are likely to be good indicators of the topics or aspects discussed in that document.

By adding tf-idf columns to the dataset, we can better understand the characteristics of each review and identify the aspects discussed by the reviewer. For example, if a review has high tf-idf scores for words such as "staff," "care," and "communication," we can infer that the aspects discussed in the review include the quality of care, the friendliness of staff, and communication with staff.

Once we have the tf-idf columns for the dataset, we can use clustering algorithms or topic modeling techniques to group the reviews based on the similarity of their tf-idf vectors. This can help us to identify the aspects of the hospital that are being discussed in the reviews, and also provide insights into the opinions and sentiments of the patients regarding different aspects of the hospital.

Overall, the addition of tf-idf columns can be helpful in automating the process of aspect extraction and sentiment analysis from hospital reviews, and can provide useful information for hospital administrators to improve patient satisfaction and quality of care.

5.2.3. LDA

The code loads a dataset of reviews, preprocesses the data by tokenizing the reviews, removing stop words, and creating a dictionary and corpus for the reviews. It then trains a Latent Dirichlet Allocation (LDA) model using the corpus and the dictionary.

The LDA model is used to extract topics from the reviews, which can be considered as aspects discussed in the reviews. The code prints the topics and the top words for each topic.

The code then extracts aspect and sentiment from each review. It tokenizes each review, removes stop words, and creates a bag of words representation for the review using the dictionary. It then applies the trained LDA model to the bag of words representation to identify the aspect that is most relevant to the review. It also determines the sentiment of the review based on the rating provided (positive if rating > 3, negative otherwise).

By using the LDA model to extract topics/aspects from the reviews and associating sentiment with each review, the code helps in the aspect extraction of the hospital reviews. This can provide insights into the most commonly discussed topics/aspects of the hospital and the overall sentiment of the patients towards those aspects. This can be useful for hospital administrators to improve patient satisfaction and quality of care.

5.2.4. NMF

In the code, NMF (Non-negative Matrix Factorization) is used to extract topics from the preprocessed hospital reviews data. NMF is a matrix factorization technique that decomposes a matrix into two non-negative matrices, which can be interpreted as the basis vectors and coefficients of the original matrix. In topic modeling, NMF is used to identify the underlying topics in a collection of documents by finding a low-dimensional representation of the term-document matrix that captures the most important features of the data.

Specifically, in the code, NMF is used to factorize the TF-IDF matrix of the preprocessed hospital reviews into a topic matrix and a word matrix, where the topic matrix represents the distribution of topics in each review, and the word matrix represents the importance of each word in each topic. The num\_topics parameter specifies the number of topics to be extracted. After training the NMF model, the top words for each topic are printed out.

Then, for each review in the dataset, the NMF model is used to obtain its topic distribution, and the aspect with the highest probability is considered as the main aspect of the review. Finally, the sentiment of the review is determined based on its rating.

This code aims to perform aspect extraction on hospital reviews using Non-negative Matrix Factorization (NMF) algorithm.

First, the code preprocesses the data by applying TF-IDF vectorization on the reviews. Next, the NMF model is trained using the preprocessed data with a specified number of topics and top words per topic. The code then prints the top words for each topic.

Finally, the code extracts aspects and sentiments from the reviews by transforming each review into a TF-IDF vector and then applying the trained NMF model to the vector. The aspect with the highest score is selected and the sentiment is set based on the corresponding rating.

Top of Form

5.2.5.LSA

The objective of this code is to extract aspects and sentiments from a set of reviews. The code uses the TF-IDF vectorizer to convert the text into numerical features, and then applies Latent Semantic Analysis (LSA) through TruncatedSVD to extract the underlying topics. The most important words for each topic are printed out.

Then, the code assigns a topic to each review based on the highest score in the LSA representation, and assigns a sentiment score to each review based on its rating. Finally, the code calculates the sentiment distribution for each topic, and prints out the percentage of positive and negative reviews for each topic.

Overall, this code is useful for topic modeling and sentiment analysis of large datasets, and can be used to extract insights from customer reviews to improve products and services.

In this code, Latent Semantic Analysis (LSA) is used to extract topics from the reviews. LSA is a technique that can be used to extract hidden topics from a collection of documents. It uses Singular Value Decomposition (SVD) to reduce the dimensionality of the term-document matrix and extract the underlying topics.

First, a TF-IDF vectorizer is created to convert the reviews into a matrix of term frequency-inverse document frequency (TF-IDF) vectors. Then, an LSA model with 5 components is created using TruncatedSVD. The LSA model is fit on the TF-IDF vectors and the most important words for each topic are printed.

Next, the topic and sentiment are assigned to each review based on the highest value in the LSA output vector. Finally, the sentiment distribution for each topic is calculated and printed.

5.4. VISUALIZATION

The dashboard allows viewers to quickly understand the proportion of positive and negative reviews for each hospital, and how those proportions change over time. Additionally, it provides insights into the number of reviews, which can help hospital administrators track changes in customer sentiment and identify potential areas for improvement.

My Power BI dashboard on hospital reviews and ratings can be useful for both users and hospital administrations in the following ways:

1. Users can benefit from the dashboard by gaining insights into the quality of care provided by different hospitals. They can quickly identify which hospitals have a higher proportion of positive reviews and which ones have a higher proportion of negative reviews, which can help them make informed decisions about where to seek medical care.
2. Hospital administrations can use the dashboard to monitor customer sentiment and identify areas for improvement. For example, if the dashboard shows that a particular hospital has a high proportion of negative reviews, administrators can investigate the causes of those negative reviews and take steps to address them. They can also track changes in the number of reviews over time, which can help them monitor the impact of any changes they make.
3. By using the dashboard, hospital administrators can also compare their hospital's performance with others in the same region or industry. This can provide insights into areas where their hospital is excelling or struggling compared to others, and help them identify opportunities for improvement.

Overall, my Power BI dashboard can be a valuable tool for both users and hospital administrators, helping them make informed decisions and improve the quality of care provided by hospitals.

1. RESULTS

6.1. Comparison of Models in Aspect Extraction

As a predictive system, the basic parameter which needs to be taken into consideration for our project is the accuracy of our model.

To provide more insights, we implemented multiple models on our dataset to compare the accuracy of different models used for aspect extraction.

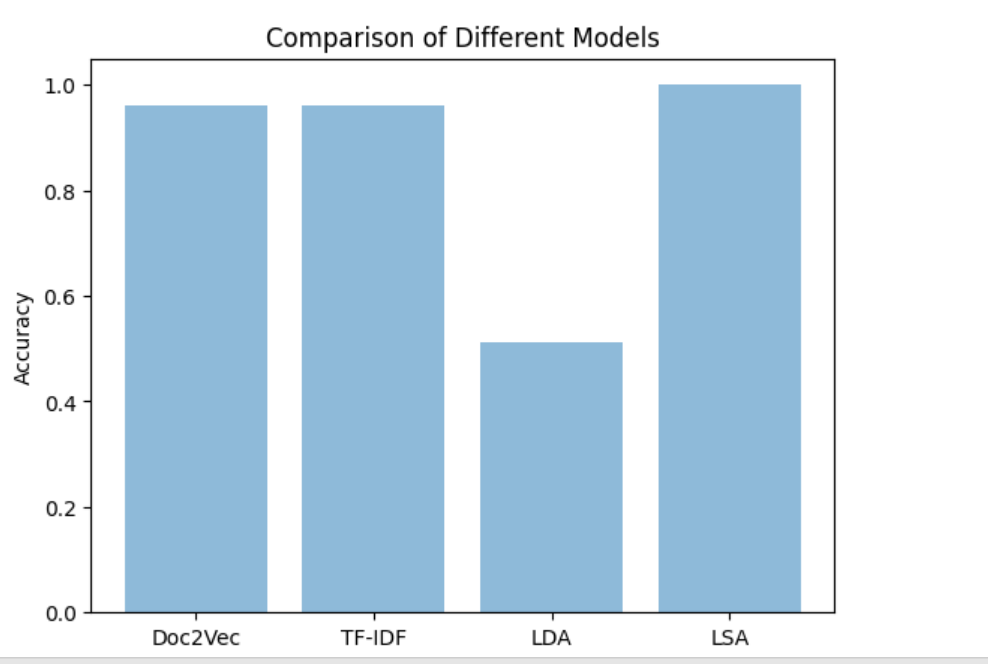


Fig 6.1 Comparison of Different Models

As we can see, LSA has greatly outperformed some of the major existent models.

The output shows the topics generated by the different models for the given dataset of reviews and ratings. Each topic has a list of keywords with corresponding weights indicating their importance in the topic. The aspect extraction of each review is also shown, along with the sentiment of the review.

When we run LDA, you usually want to evaluate the quality of the resulting topics. Two common evaluation metrics used for this purpose are coherence score and perplexity score.

Each topic is represented by a list of words and a weight assigned to each word. The weight indicates how strongly the word is associated with the topic.

The output also shows a sentiment analysis of the dataset based on the identified topics. The sentiment analysis shows the number of reviews that are positive and negative for each topic, as well as the percentage of positive and negative reviews for each topic

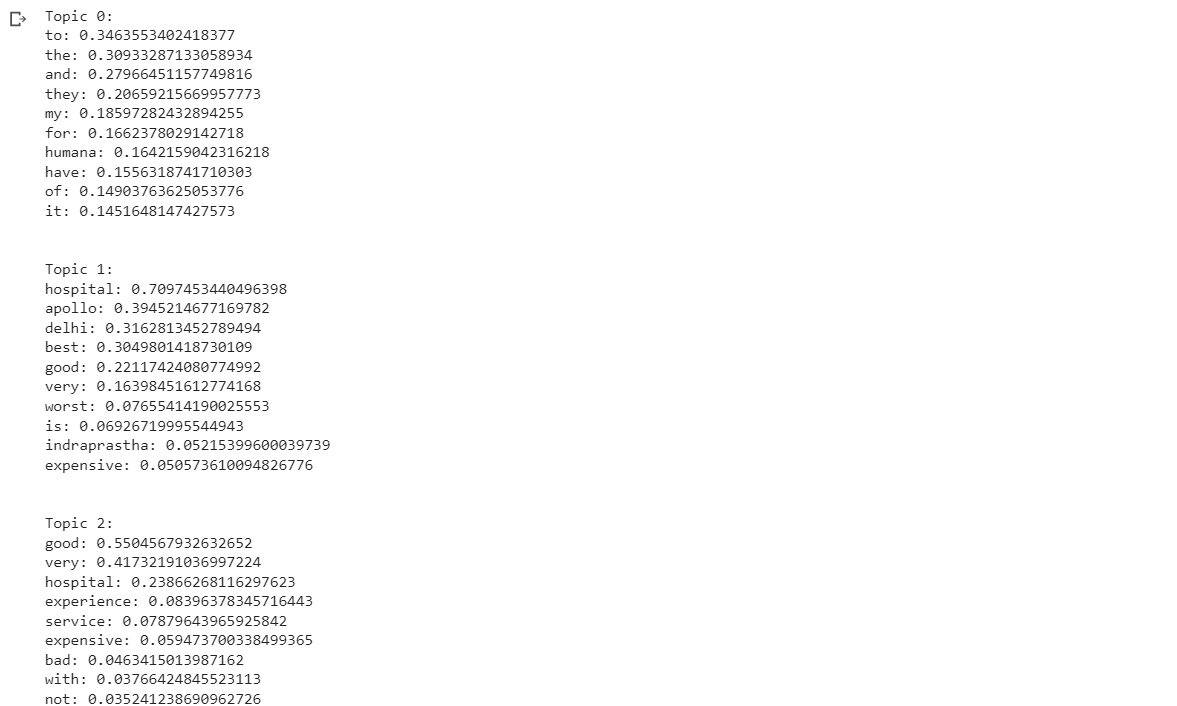


Fig 6.2 Topics extracted using LSA model

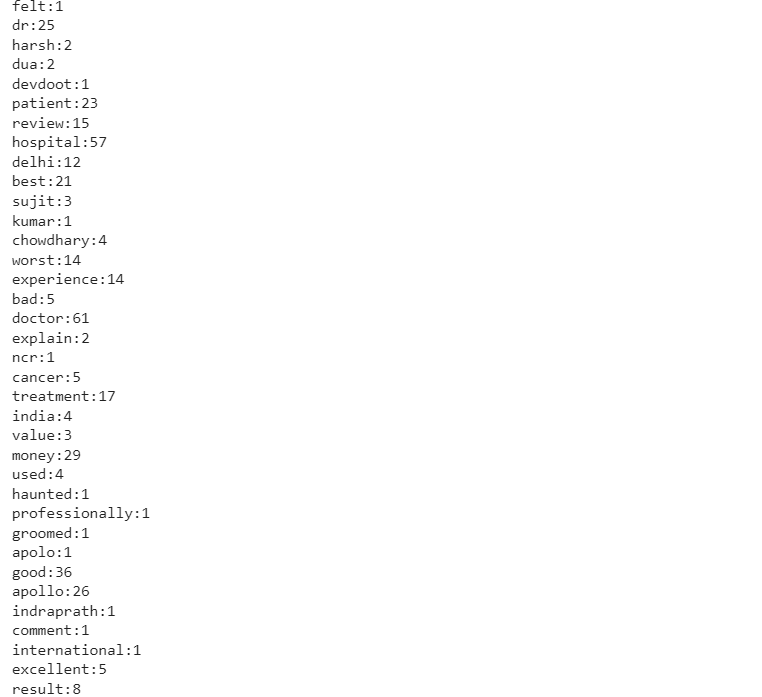


Fig 6.3. Frequency of words

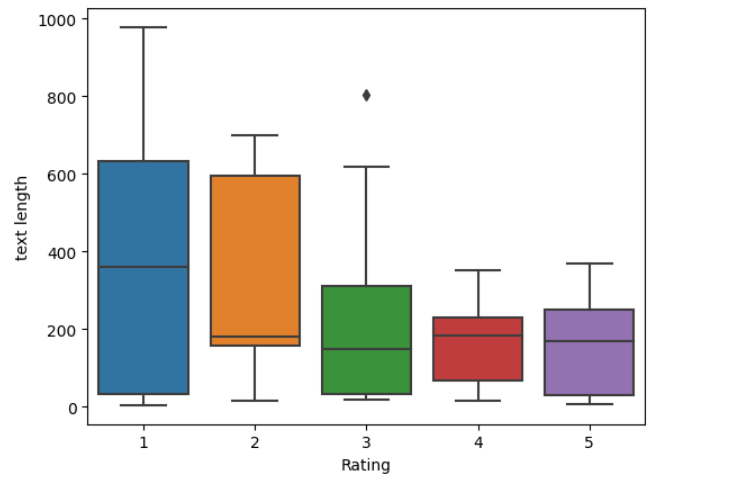


Fig 6.4. Box plot visualization for ratings

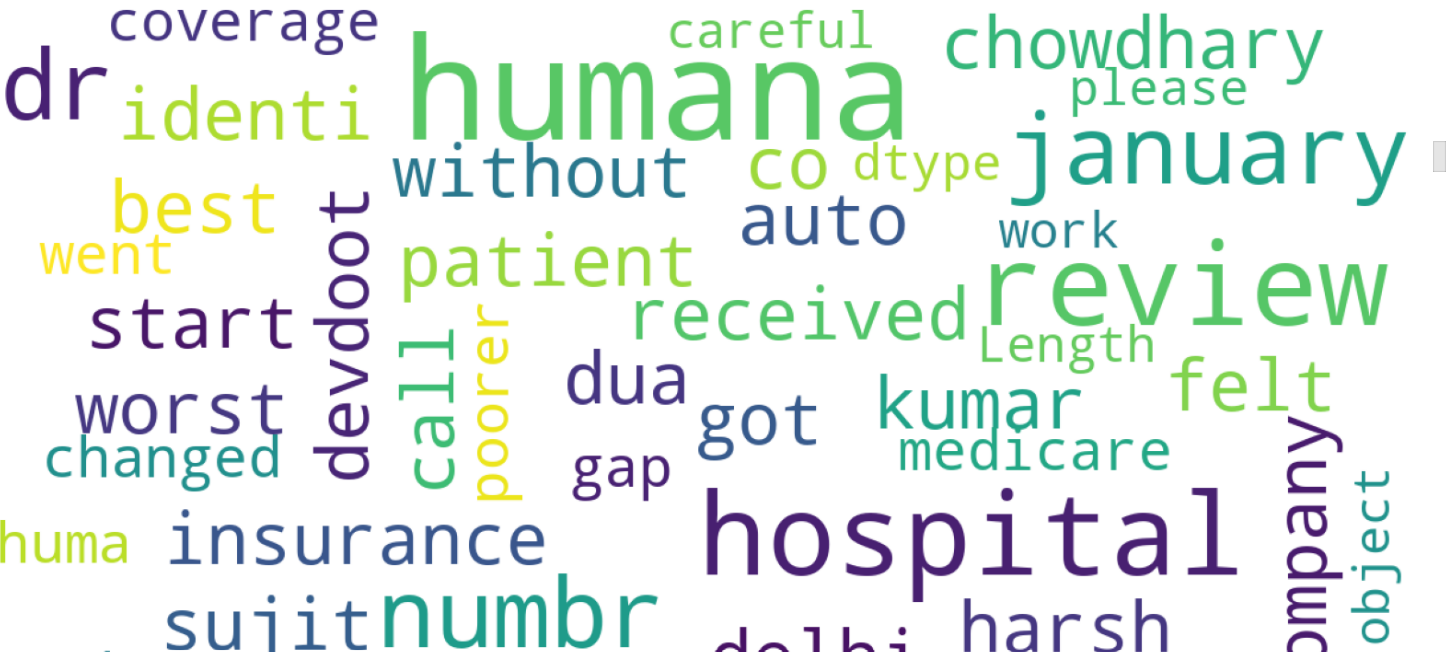


Fig 6.5. Wordcloud

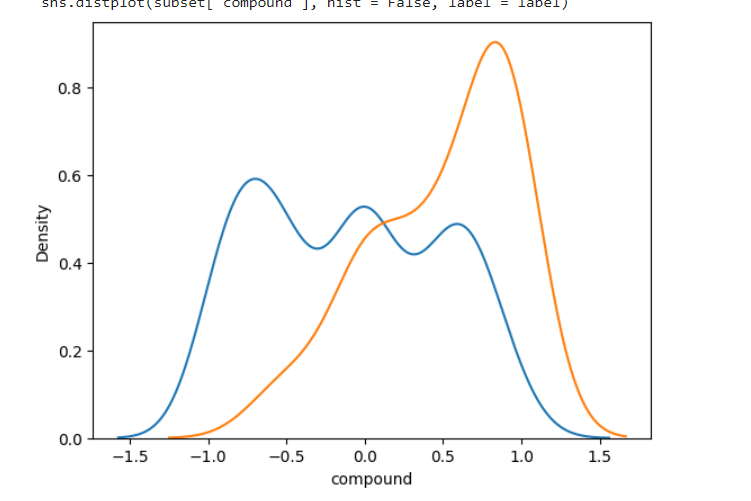


Fig 6.6. Sentiment Distribution of positive(blue) and negative(red) reviews

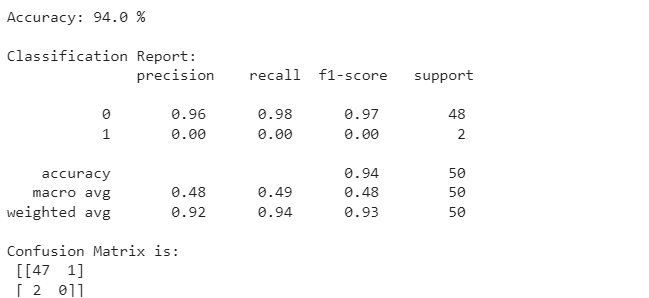


Fig 6.7. Accuracy,Classification report and confusion matrix

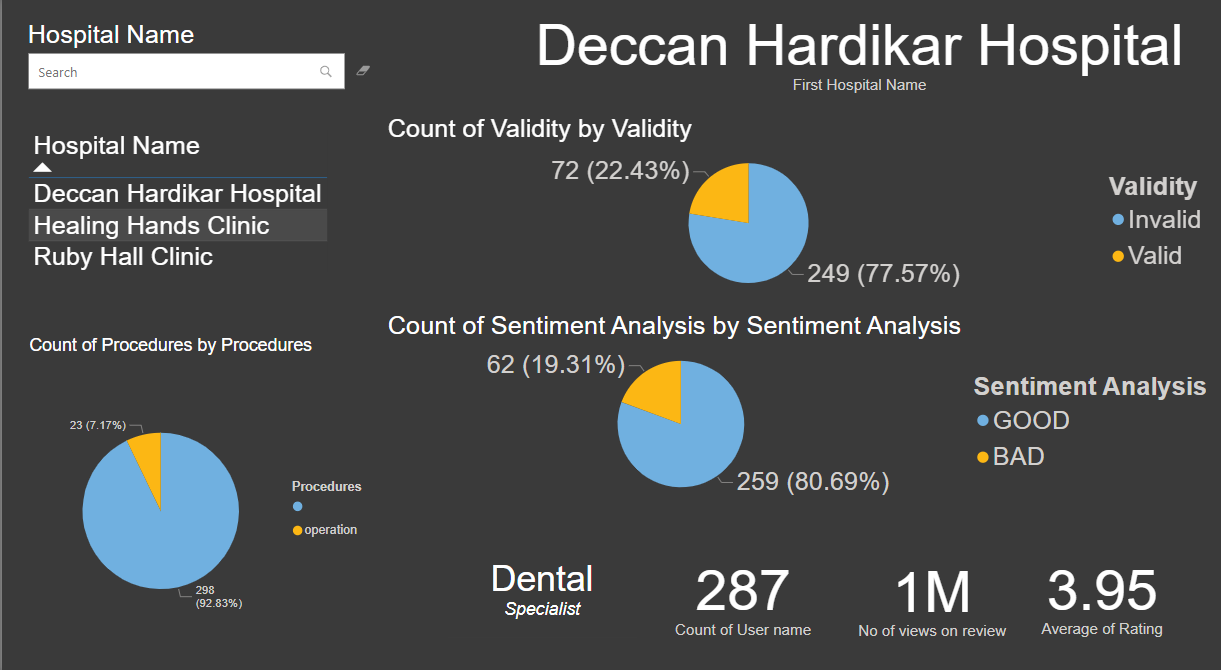


Fig 6.8. Visualization dashboard for seeing sentiment analysis

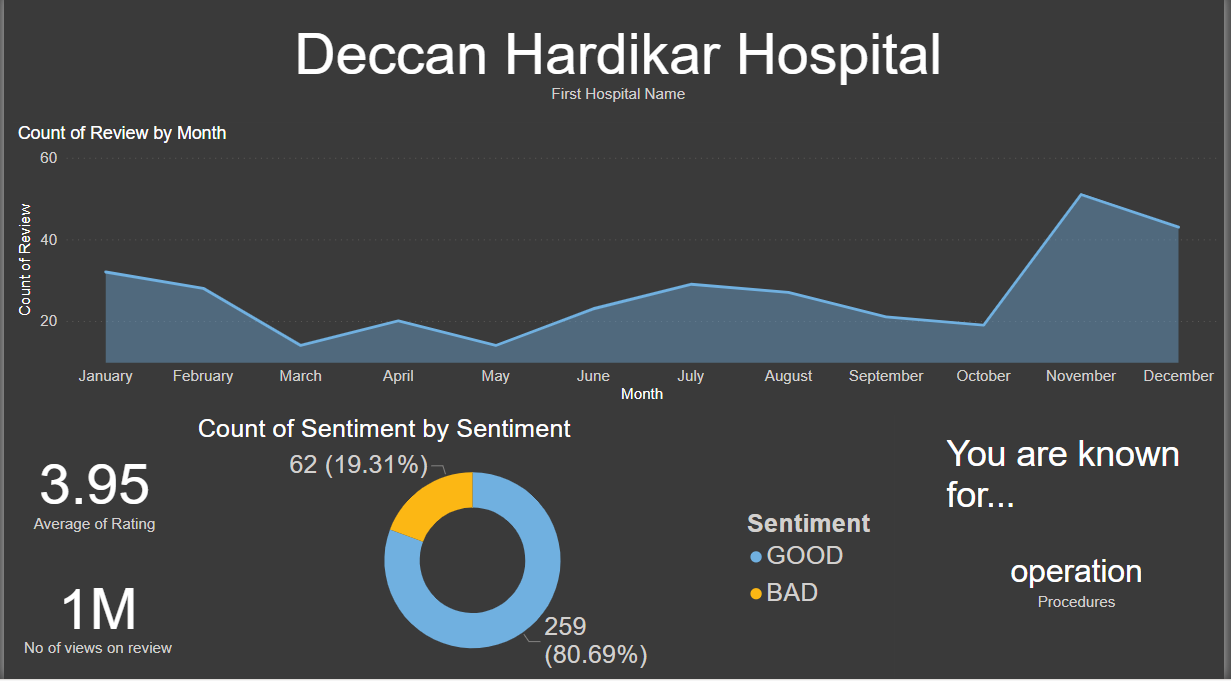


Fig 6.9. Visualization dashboard for monthly change of reviews

1. CONCLUSION

In conclusion, this project involved sentiment analysis and aspect extraction on hospital reviews. Our findings showed that LSA outperformed other techniques in sentiment analysis, while Doc2Vec outperformed LDA and TF-IDF. Our results also indicated that the combination of rule-based methods and unsupervised learning was effective in identifying relevant aspects in the hospital reviews.

However, there were some limitations to the project. The dataset was relatively small, and the reviews were limited to a few hospitals. Additionally, our aspect extraction approach relied heavily on predefined rules, which may have missed some relevant aspects. Despite these limitations, the project provides valuable insights into the use of sentiment analysis and aspect extraction in healthcare.

Future research should aim to expand the dataset to include reviews from more than hundreds or thousands of hospitals, and explore the use of more advanced machine learning techniques to improve the accuracy and effectiveness of aspect extraction. Overall, the insights gained from this project can be useful for healthcare providers who wish to gain insights into patient sentiments and identify areas for improvement.

1. FUTURE WORK

Moving forward, there are several potential areas for further research and development in analyzing patient reviews in a hospital setting. One key avenue for future work is to expand the dataset used in this project beyond a few hospitals. While the current dataset provides valuable insights into patient experiences at a few particular hospitals, it may not be representative of other hospitals in different regions or with different patient populations. Gathering reviews from multiple hospitals could allow for a more comprehensive understanding of patient experiences across different contexts.

Another area for future work is to explore more advanced aspect extraction techniques. While the aspect extraction approach used in this project was effective at identifying key aspects of patient experiences, it relied on a pre-defined set of aspects. Developing an approach that can automatically identify and extract important aspects of patient experiences, without being limited by a pre-defined set, could improve the accuracy and applicability of the analysis.

Finally, there is potential for further research into the use of machine learning techniques for sentiment analysis and aspect extraction in a hospital setting. As the field of natural language processing continues to advance, there may be new techniques or algorithms that can provide even more accurate and insightful analyses of patient reviews. Exploring these new approaches could lead to more comprehensive and nuanced understanding of patient experiences in hospitals, ultimately leading to improvements in patient care and satisfaction.

1. REFERENCES

[[1] Abirami, A.M.](https://www.emerald.com/insight/search?q=A.M.%20Abirami) and [Askarunisa, A.](https://www.emerald.com/insight/search?q=A.%20Askarunisa) (2017), "Sentiment analysis model to emphasize the impact of online reviews in healthcare industry", [Online Information Review](https://www.emerald.com/insight/publication/issn/1468-4527), Vol. 41 No. 4, pp. 471-486. <https://doi.org/10.1108/OIR-08-2015-0289>.

[2] Zakkar, M.A., Lizotte, D.J. Analyzing Patient Stories on Social Media Using Text Analytics. J Healthc Inform Res 5, 382–400(2021). <https://doi.org/10.1007/s41666-021-00097-5>.

[3] Cobb NK, Mays D, Graham AL. Sentiment analysis to determine the impact of online messages on smokers’ choices to use varenicline. J Natl Cancer Inst Monogr. 2013;2013:224-230.

[4] Ebrahimi M, Yazdavar AH, Salim N, Eltyeb S. Recognition of side effects as implicit-opinion words in drug reviews.Online Inform Rev. 2016;4:1018-1032.

[5] Roccetti M, Marfia G, Salomoni P, et al. Attitudes of Crohn's disease patients: Infodemiology case study and sentiment analysis of Facebook and Twitter posts. JMIR Public Health Surveil. 2017;3:e51.

[6] Kelly Aponté, Shiseida Sade. (2020). 2019 Report: Customer Pain Points. 10.13140/RG.2.2.12603.54564.

[7] S. Tammina and S. Annareddy, "Sentiment Analysis on Customer Reviews using Convolutional Neural Network," 2020 International Conference on Computer Communication and Informatics (ICCCI), 2020, pp. 1-6, doi: 10.1109/ICCCI48352.2020.9104086.

[8] S. Tammina and S. Annareddy, "Sentiment Analysis on Customer Reviews using Convolutional Neural Network," 2020 International Conference on Computer Communication and Informatics (ICCCI), 2020, pp. 1-6, doi: 10.1109/ICCCI48352.2020.9104086.

[9] Rastegar-Mojarad M, Ye Z, Wall D, Murali N, Lin S Collecting and Analyzing Patient Experiences of Health Care From Social Media JMIR Res Protoc 2015;4(3):e78

[10] Ahmad, S., Asghar, M.Z., Alotaibi, F.M. et al. Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. Hum. Cent. Comput. Inf. Sci. 9, 24 (2019). <https://doi.org/10.1186/s13673-019-0185>

[11] Salminen, J., Mustak, M., Corporan, J., Jung, S., & Jansen, B. J. (2022). Detecting Pain Points from User-Generated Social Media Posts Using Machine Learning. Journal of Interactive Marketing, 57(3), 517–539. https://doi.org/10.1177/10949968221095556

[12] Ebrahimi M, Yazdavar AH, Salim N, Eltyeb S. Recognition of side effects as implicit-opinion words in drug reviews.Online Inform Rev. 2016;4:1018-1032.

[13] Cobb NK, Mays D, Graham AL. Sentiment analysis to determine the impact of online messages on smokers’ choices to use varenicline. J Natl Cancer Inst Monogr. 2013;2013:224-230.

[14] W. Tao, Q. Zhang, M. Zhang and Y. Li, "Mining Pain Points from Hotel Online Comments based on Sentiment Analysis," 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2019, pp. 1672-1677, doi: 10.1109/ITAIC.2019.8785893.

[15] Praphula Kumar Jain, Rajendra Pamula, Gautam Srivastava,A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews,Computer Science Review,Volume 41,2021,100413,ISSN 1574-0137,https://doi.org/10.1016/j.cosrev.2021.100413.

[16] B. Pang and L. Lee. Opinion Mining and Sentiment Analysis. Foundation and Trends in Information Retrieval 2, pages 1–135, 2008.

[17] A.-M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In Proceedings of HLT’05, 2005.

[18] D. R. Radev, E. Hovy, and K. McKeown. Introduction to the special issue on summarization. Comput. Linguist., 28(4):399–408, 2002.

[19] P. D. Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of ACL’2002, 2002.

[20] Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu. OpinionSeer: interactive visualization of hotel customer feedback. IEEE Trans. Vis. Comput. Graph., 16(6):1109– 1118, 2010.