SENTIMENT ANALYSIS AND TOPIC MODELLING ON TWITTER FOR CLEAN INDIA MISSION

Sangeeta Rani

Research Scholar
Department of Computer Science and Application
Maharishi Dayanand University, Rohtak, Haryana, India
Sangeeta.yogi@gmail.com

Nasib Singh Gill

Professor
Department of Computer Science and Application
Maharishi Dayanand University, Rohtak, Haryana, India
nasibsgill@gmail.com

Preeti Gulia

Assistant Professor
Department of Computer Science and Application
Maharishi Dayanand University, Rohtak, Haryana, India
Preetigulia81@gmail.com

Abstract

Twitter is an important source of information but it is challenging to analyze this data in order to recover meaningful inference. The present paper uses topic modelling and sentiment analysis to draw useful context from Twitter data set related to 'Clean India Mission'. Latent Dirichlet Allocation is used in the research to identify twenty most trending topics and top seven terms related to each of the twenty topics. Coherence and prevalence values represent model efficiency. Topic clustering is also used in the research to identify how strongly topics are related to each other. Five different clusters are created from the top trending topics reflecting different aspects in the corpus. The average silhouette width is employed to determine the optimal number of clusters. Lexicon based classification using 'nrc' sentiment directory is also used to reflect people's sentiment at ten different sentiment levels for the mission. Twitter data for the research is collected from seven different Hashtags, including the official page of the clean India campaign. The most relevant subject segments are identified after evaluating the trending topics by utilizing topic coherence value.

Keywords*: *Twitter sentiment analysis; Topic Modelling; LDA; Opinion Mining; Clean India Mission, Topic Clustering.

1. Introduction

Twitter is one of the most widely used social media platforms, with millions of users across the world to share their opinion on diverse issues related to various products and services regarding health care, politics, news, sports, education, government, pubic polices, natural disaster and many more. A huge data repository is created and it is quite challenging to find out useful context from this hybrid data. Twitter opinion mining is a way to transform semi structured twitter data to more structured from to find out sentiment attached with the tweets [Agarwal et al., (2011) and Maheshwari et al., (2019)]. It aids in determining what people think about a specific product or issue, which in turn aids the relevant company or organization in improving their service or product based on the input obtained. The perception of individuals is computed using a variety of automated machine learning and sentiment analysis techniques. Sentiment analysis is carried out using various supervised and unsupervised classification techniques [Ray (2017) and Kurnaz et al., (2019)].

Topic modelling is also a useful way to uncover the hidden semantics context in tweets and for detailed analysis. A topic is a group of words co-occurring in multiple documents, related to the same context. It is a rapidly developing branch of text mining that can be applied to twitter data for more elaborate text analysis. Topic modelling is a way to find out the group of words, which are expected to appear in the corpus and best reflects the context of the corpus. In topic modelling we are more concerned about long-range context like in the case of n-grams and local dependencies. It aids in the discovery of latent semantic structure [Kherwa et al., (2018) and

Sokolova et al., (2016) and Wisnu et al., (2020) and Shah et al., (2020)]. Different topic modelling techniques viz. LSA (Latent Semantic Analysis), LDA (Latent Dirichlet Allocation), PLSA (Probabilistic Latent Semantic Indexing), GSDMM (Gibbs Sampling algorithm for the Dirichlet Multinomial Mixture) are used by researchers for more elaborate analysis. [Rania et al., (2020)] mentioned in their study that LDA works efficiently for small and noisy text. In the present work also, LDA topic modelling technique is used to dig out the most trending topics and terms frequently used in the tweets to have an idea about people's involvement and their opinion. It uncovers the hidden semantics related to 'clean India Tweets'. 'Clean India Mission' also termed as 'Swatch Bharat Abhiyan' is an Indian government ambitious mission for making India clean and making the masses aware about cleanliness. Millions of people across India are connected with the mission and put their opinion, comments, grievances and suggestions about the mission on various social media platforms. Twitter is one of the platforms used for this purpose. The analysis of this data can be quite useful in anticipating people's attitudes.

Clustering is a term that refers to a collection of approaches used for identifying subgroups of similar observations in a data corpus. It is an unsupervised method for dividing observations into different clusters by determining the relationship or distance between variables. The technique is helpful in identifying diverse aspects from the data corpus. In the present research, clustering is also used to group the observations in different clusters to find different aspects related to the mission from the Twitter data set [Curiskis et al., (2020) and Ayo et al., (2021)].

The rest of this paper is organized as follows: The section I of the study introduces the concept of sentiment analysis, LDA topic modelling, and topic clustering. The related literature is summarized in section II. In section III, LDA (Latent Dirichlet Allocation) topic modelling technique is explained. Section IV discusses the methodology and tools employed. The results are discussed in Section V. The outcomes are concluded in section VI.

2. Related Literature

Topic modelling is used by a number of researchers in a variety of application areas to find the hidden crux of the topic and for sentiment analysis of text. [Habibi et al., (2021)] used LDA algorithm for topic modelling of German Instagram post to find the most trending topic and uncover the fact that Germans are very much concerned regarding healthy lifestyle diet. The data set is partitioned in eight different topic segments by using topic coherence value. All of these topics contain terms that are associated with healthy lifestyle. [Kaila et al., (2020)] performed topic modelling and sentiment analysis on tweets related to the outburst of the corona virus. LDA topic modelling is used in research to check the reliability of information flow related to corona virus via Twitter social media. According to research, the topics identified from LDA demonstrate that the most relevant information is communicated during corona disaster. [Garcia and Berton (2021)] also performed topic modelling on COVID-19 tweets from Brazil and USA. They identified ten different subjects and evaluated the information discussed on Twitter, providing an assessment of the discourse's evolution. [Yang (2018)] implemented LDA topic modelling in their research on twitter data using R language to find the most trending topics and related terms for these topics. [Kuuliala et al., (2021)] used LDA based topic modelling for food spoilage analysis. Based on the findings of the LDA model, a comprehensive spoilage assessment protocol was proposed that can be used as an indicator to detect possible spoiling of food under different conditions. [Parveen et al., (2021)] et al. used LDA topic modelling to analyze online data for women's apparel to enhance shopping experience and get the nerve of customers. R-Spark is used in the implementation of research work. [Gangadharan and Gupta (2020)] used LDA topic modelling in the area of agriculture for named entity recognition. They implemented LDA for 3000 sentences collected from agriculture sites. The research used a hybrid approach and identified the most common crops, crop diseases, soil types, fertilizers names and pathogen names by using topic modelling. Manual evaluation and model results are compared and show an accuracy of 80%. [Bastani et al., (2019)] at al. also used LDA for topic modelling of CFPB consumer complains. A LDA based decision support system is proposed in the research that summarizes the customers complains in meaningful aspects. [Mantyla et al., (2018)] worked in their research regarding the stability of LDA model by using clustering and replicated run of LDA model. After the evaluation of multiple metrics, rank biased overlap metric was suggested that represents stability of topic inside the cluster. The LDA technique is explored in various researches in diverse domains. In the present research, LDA topic modelling is used to uncover hidden semantic structure of twitter posts related to 'clean India mission'.

3. Latent Dirichlet Allocation and Topic Modelling

Probabilistic topic modelling is used in machine learning and helps to get deep inside the topic. LDA is an unsupervised probabilistic topic modelling approach that was initially proposed by Blei, Ng and Jordan in 2003. LDA has swiftly become one of the most prominent text modelling approaches having a major influence in the domains of statistical machine learning and natural language processing. It can be used in correlation with sentiment analysis techniques for more precise predictions.

As per the LDA algorithm specification, a semantic structure is hidden in a document that can be revealed from word-document co-occurrences. Documents in the corpus have hidden abstract topics and topics have related

terms. LDA is used for extracting the most likely terms for the topics and most likely topics related to the documents [Onan et al., (2016)]. DTM (Document term matrix) created from the data set is used for the implementation of LDA model. LDA model is graphically shown in figure 1. For tuning the LDA model, several parameters must be specified, including 'α', 'β', 'M', 'N', 'z', 'θ' and number of iterations. 'α' represents document-topic density, 'β' represents topic-term density, 'M' is number of documents, 'N' represents total number of terms, 'z' represents the topic for a given term in a document and 'θ' represents topic distribution for particular document. Topic coherence value represents the quality of topics generated by the model. A higher coherence value is expected. The nature of data and quality of preprocessing also affect the outcome of the model.

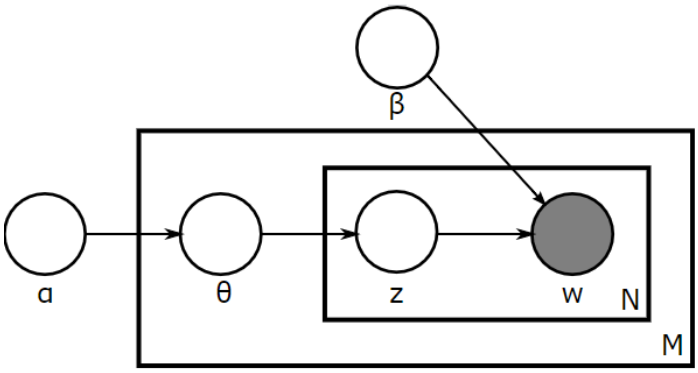

Fig. 1. Graphical Representation of LDA Model.

LDA is used in various researches and in diverse application domains for text mining and topic modelling. [Kalepalli et al., (2020)] compared LSA and LDA topic modelling techniques in their research work. As per research, LDA outperformed LSA with better divergence value. [Mohammed and Al-augby (2020)] also performed comparative analysis of LDA and LSA topic modelling techniques. They mentioned in their research that LDA performed better than LSA. LDA technique partitioned the data set into 20 optimal topics with (0.592179) coherence value and LSA partitioned the data into 10 optimal topics with (0.577302) coherence value. The LSA approach does not work with ploysemous words and is less efficient than LDA. LDA is well suited to small and noisy data like tweets. In the present research also, LDA is used for Twitter data analysis and topic modelling.

4. Methodology Design and Implementation

Tweets are collected from seven different Hashtags and handles related to 'Clean India Mission' by using the Twitter API. The Collected tweets are cleaned to remove URL's, stop words, numbers, punctuation symbols, conjunctions and re-tweets. Lemmatization and tokenization are applied on the resultant data. After cleaning and preprocessing, the resultant corpus having 2209 tweets are used in the research. The implementation work flow is shown in Figure 2.

LDA is used in the present work as a topic modelling technique to dig out useful context from twitter data. DTM (Document term matrix) is created on the cleaned and processed data and topic modelling using LDA is applied on that matrix to extract the most trending topics and related terms for the twitter corpus. The model is tuned for 500 iterations to get optimal results for retrieving twenty most trending topics and related top terms. Coherence and prevalence are used to evaluate the efficiency of LDA model. Unsupervised sentiment extraction technique is used to find opinion of common masses for the mission. For this, sentiment polarity of tweets are predicted by using get_nrc_sentiment ( ) function of the syuzhet package of R Studio. Sentiments are predicted at ten different sentiment levels.

Clustering is also used to predict how strongly topics are correlated and identify the diverse domains in the data corpus. 'hclust' with 'ward.D2' method is used to create five clusters. Optimal number of clusters are identified by comparing Average Silhouette Width. R Studio tool is used for the implementation and analysis. R Studio is an open source tool and an IDE for the R programming language that can run on different platforms. Various functionality provided by R Studio is abstracted in different packages. TwitterR, textmineR, tm and syuzhet packages are used in the implementation. 'nrc' sentiment directory is used for mining Twitter sentiment. Package 'ggplot2' is used for the analysis and visualization of results.

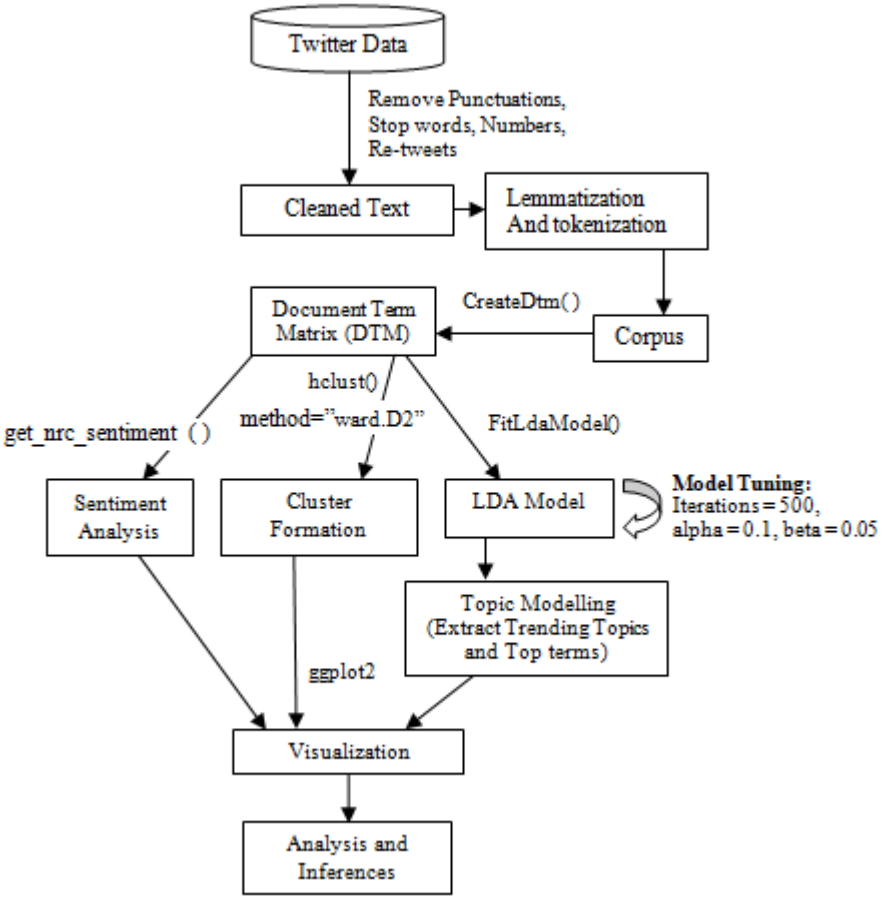

Fig. 2. Process Flow Diagram for Topic Modelling and Sentiment Analysis.

5. Results and Analysis

LDA model is used in the research to find the most trending topics and top terms related to each topic using 'textmineR' library of R Studio. The research uses 2209 real-time tweets after pre-processing, cleaning and removal of re-tweets related to 'clean India mission'. LDA model is tested for 100, 200 and 500 iterations. After evaluation and analysis, LDA model is tuned to run for 500 iterations with the value of 'k' as 20, alpha as 0.1, beta as 0.05. Figure 3 shows the twenty most frequently tweeted terms with their corresponding frequency. The LDA model is tuned to retrieve top seven terms related to each of the twenty most trending topics. Figure 4 represents these top seven terms for each of the most trending topics. The log_likelihood and Coherence are used as an objective measure to judge the quality of LDA model and topic quality. Figure 5 represents change in log_likelihood value with respect to iteration. The model's coherence is used to ensure that the resultant topic modelling is a stable model and capable of representing the whole content. It is a metric that assesses the degree of semantic similarity between high-scoring terms in a single topic. A higher value of coherence is required for an efficient model. Figure 6 shows the coherence and the corresponding topic frequency. The most common subjects in the corpus are determined by prevalence. It represents the probability of the topic's distribution throughout the documents. Figure 7 graphically represents prevalence for most trending topics corresponding to their alpha value. Alpha value is the density of topic in documents. The results show that the model starts stabilizing after 200 iterations.

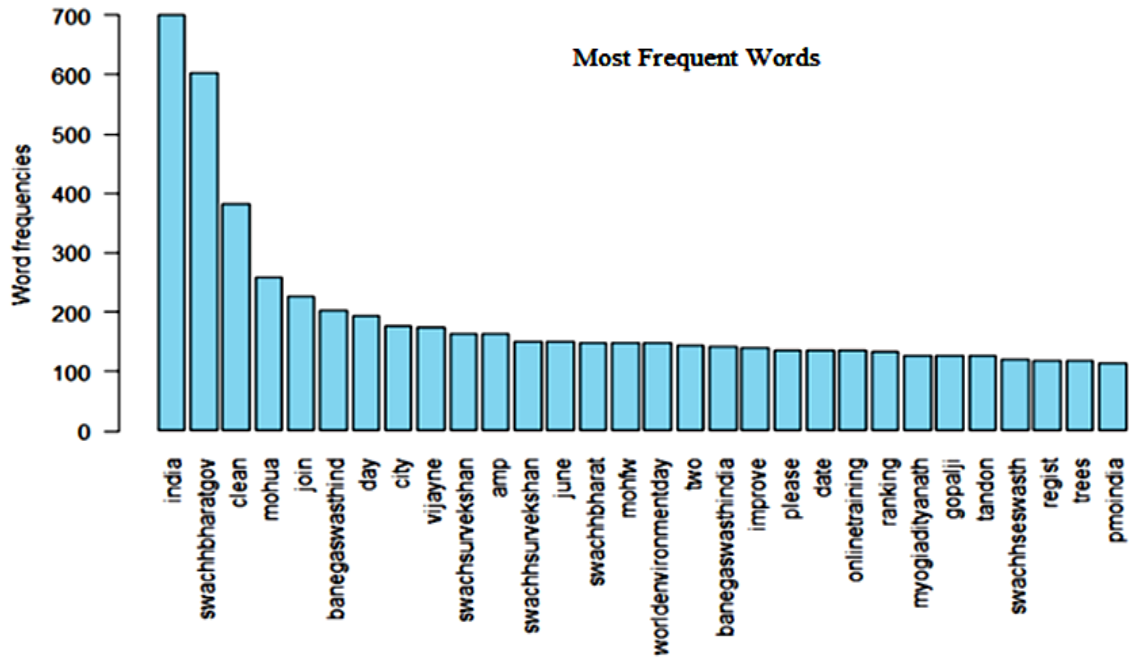

Fig. 3. Visualization Results of Top Twenty Trending Terms vs. Tweet Frequency.

```
> model$top_terms <- GetTopTerms(phi=model$phi, M=7)
> model$top_terms
     t_1              t_2              t_3              t_4              t_5          t_6
[1,] "living"         "swachhbharatgov" "swachhbharatgov" "noida"          "clean"      "city"
[2,] "healthy"        "odisha"         "door"           "narendramodi"   "india"      "join"
[3,] "reducing"       "cmo"            "dmkoraput"      "swachhbharatgov" "water"      "day"
[4,] "reducing_weight" "mybmc"         "sir"            "ceonoida"       "covid"      "improve"
[5,] "weight"         "cmo_odisha"     "swachh"         "authority"      "waste"      "june"
[6,] "healthy_living" "mumbai"         "matter"         "noida_authority" "amp"       "swachhsurvekshan"
[7,] "weight_healthy" "mosarkar"       "garbage"        "energy"         "karimnagar" "date"
     t_7          t_8              t_9              t_10             t_11             t_12
[1,] "trees"      "amp"            "unite"          "swachsurvekshan" "swachhbharatgov" "india"
[2,] "clean"      "mygovindia"     "fightcorona"    "leader"         "killing"        "clean"
[3,] "air"        "options"        "unite_fightcorona" "team"        "killing_trees"  "environment"
[4,] "clean_air"  "pl"             "join"           "position_required" "drive"       "green"
[5,] "india"      "amp_showcase"   "fight"          "swachhbharatmission" "cleanliness" "world"
[6,] "forest"     "film_swachhata" "corona_virus"   "assistance"     "members"        "make"
[7,] "provide"    "ready_shoot"    "virus"          "consultant"     "carried"        "day"
     t_13             t_14                                  t_15             t_16
[1,] "india"          "banegaswasthind"                     "swachhbharatgov" "swachhbharatgov"
[2,] "mohua"          "banegaswasthindia"                   "northdmc"       "swachhbharat"
[3,] "mohfw"          "swachhseswasth"                      "official"       "swachhbharatgov_swachhbharat"
[4,] "mohfw_india"    "swachhseswasth_banegaswasthindia"    "cpcb"           "india"
[5,] "mohua_india"    "worldenvironmentday"                 "cpcb_official"  "india_swachhbharatgov"
[6,] "secretary"      "banegaswasthindia_worldenvironmentday" "area"         "mohfw"
[7,] "secretary_mohua" "ndtv"                               "cmohry"         "mohfw_india"
     t_17                      t_18                  t_19             t_20
[1,] "swachhbharat"            "swachhbharatgov"     "gopalji"        "swachhbharatgov"
[2,] "swachhbharatgov"         "amp"                 "gopalji_tandon" "bhaindarmira"
[3,] "swachhbharatgov_swachhbharat" "cleanliness"    "tandon"         "mncdfbombay"
[4,] "narendramodi"            "drive"               "myogiadityanath" "roadsofmumbai"
[5,] "india"                   "municipal"           "naali"          "authackeray"
[6,] "swachhbharat_swachhbharatgov" "corporation"    "nawagarhi"      "authackeray_mncdfbombay"
[7,] "pib"                     "municipal_corporation" "nawagarhi_naali" "bhaindarmira_swachhbharatgov"
>
```

Fig. 4. Experimental Results Screen Shot Showing Topic Wise Top Terms Extracted by Using LDA.

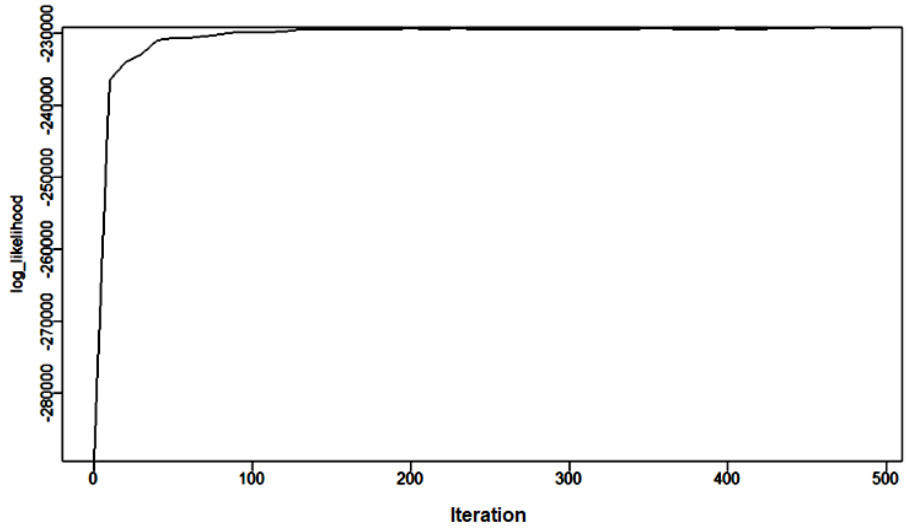

Fig. 5. Log Likelyhood vs. Iteration.

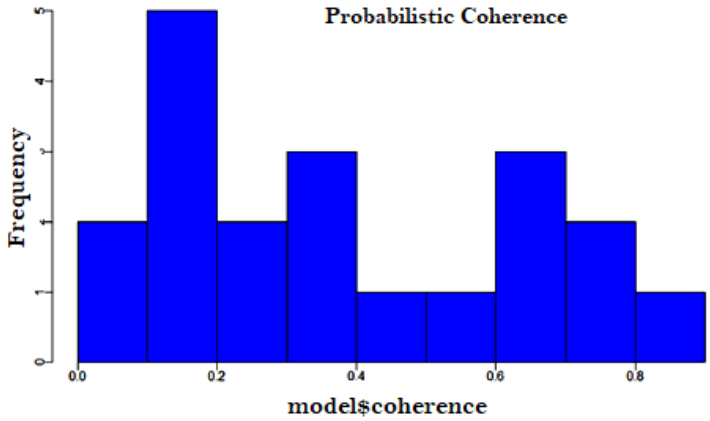

Fig. 6. Coherence vs. Frequency.

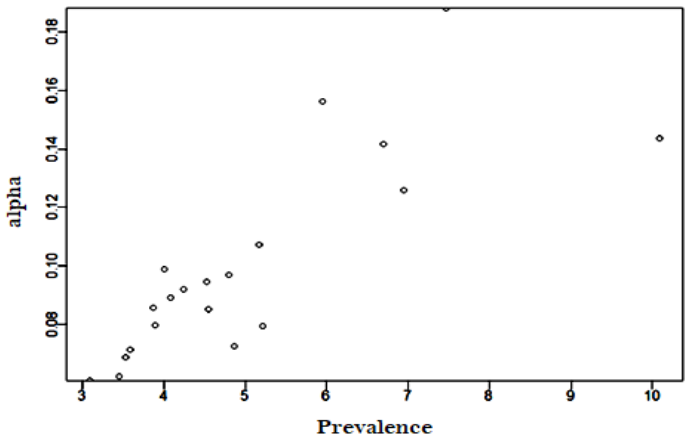

Fig. 7. Prevalence vs. Alpha Value.

Clustering can be used to find how strongly topics are related in case of unlabelled data. ‘dist ()’ function calculates distance between the samples to distinguish different clusters. ‘Ward.D2’ clustering method is used in the present research to provide spherical and compact clusters. Hierarchical clustering is accomplished by using the hclust () function of R Studio to group the samples. The resultant five clusters and related topics to the clusters are

represented in form of dendogram as show in Figure 8. Average silhouette approach is used to represent the quality of clusters. A higher average silhouette width indicates good clustering with an optimal number of clusters. Figure 9 represents highest silhouette width for five clusters and Figure 10 represents resultant five clusters in the form of scatter plot.

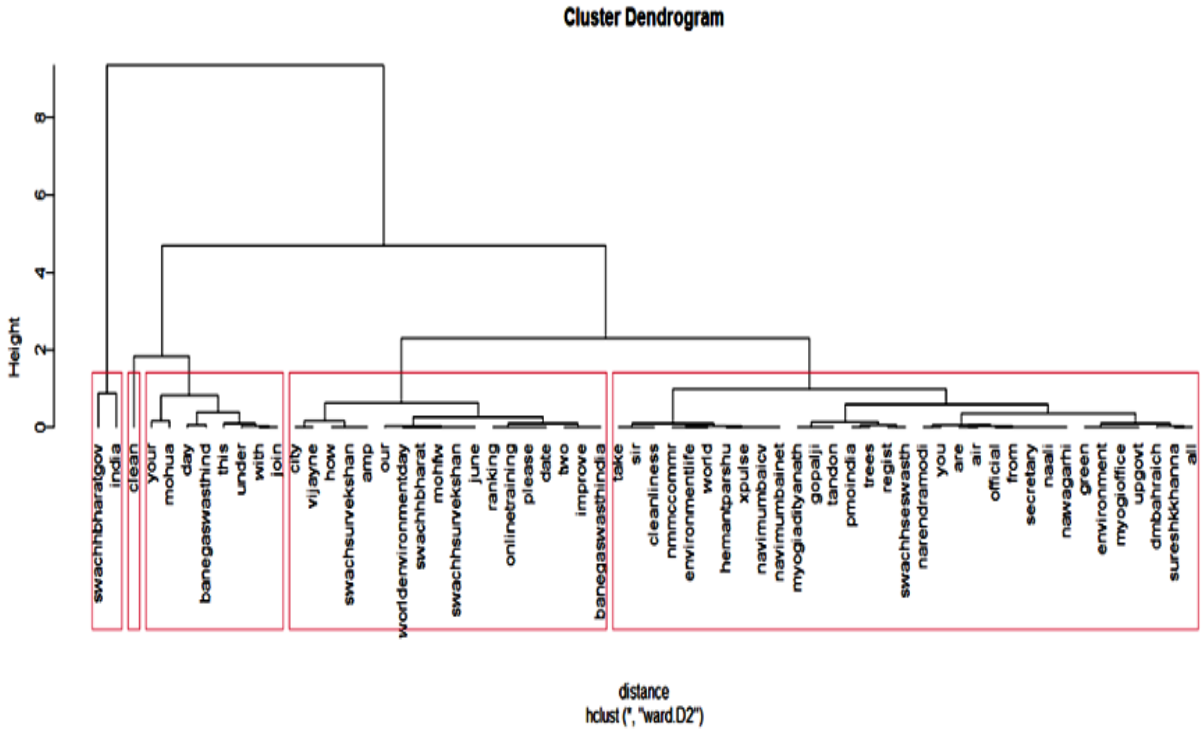

Fig. 8. Cluster Dendogram Representing Five Clusters Using Ward.D2 Method.

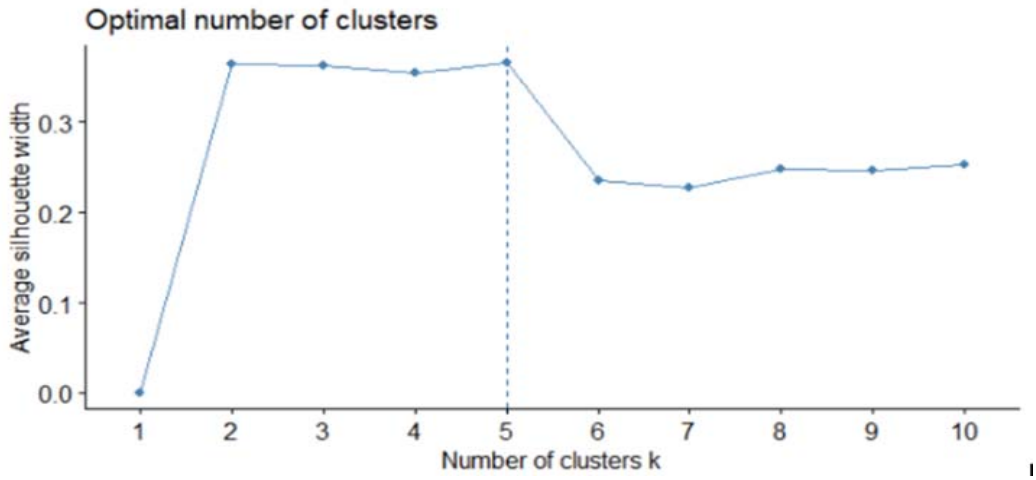

Fig. 9. Average Silhouette Width Vs. Number of Clusters.

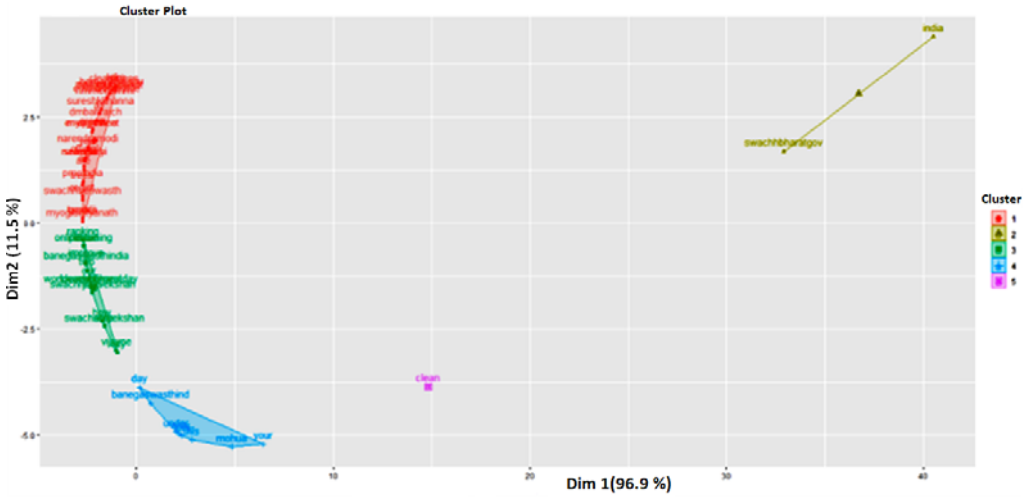

Fig. 10. Clusters in Form of Scatter Plot.

For the identification of people sentiment at different satisfaction levels, lexicon-based classification technique is applied by using 'syuzhet' package in R Studio. Sentiment directory 'nrc' developed by Saif Mohammad and Peter Turney is used in extraction of tweet sentiment [Mohammad and Turney (2013)]. Figure 11 shows sentiments related to twitter data at ten different sentiment levels. The sentiment score of individual tweets is represented in Figure 12. The results reflect the positive sentiment of people, social organization and government for the mission. The frequently coined terms and interest areas of twitter users regarding the mission is visually represented in form of wordcloud as shown in Figure 13. Wordcloud2 package of R Studio is used for this.

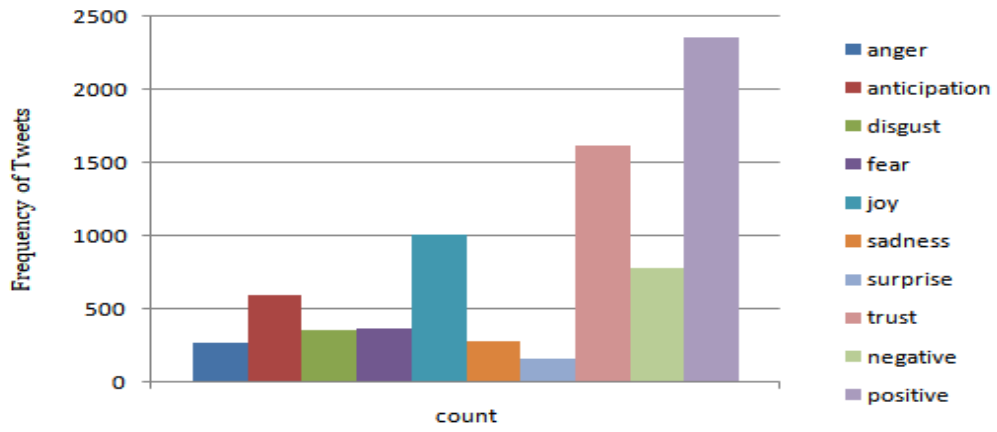

Fig. 11. Visualization of Various Levels of Emotions in Tweets.

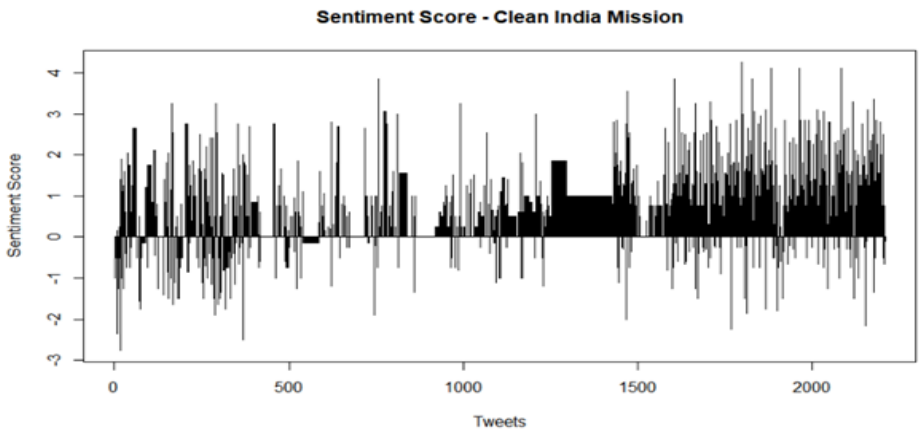

Fig. 12. Sentiment Score of Individual Tweet.

Fig. 13. Most Frequent Terms in Form of World Cloud for Tweets Related to 'Clean India Mission'

LDA topic modelling and clustering provides analysis that may be used to delve into the minds of people and dig out useful inferences. The trending topics and top terms reflect the people's positive and negative sentiments, their reactions, grievances and areas of concern for 'clean India Mission'. Several terms related to corona also link the mission to COVID-19 pandemic.

6. Conclusion

Twitter is a popular social media platform where people put their opinion, suggestion, grievances related to various topics. Analysis of this data can be very useful but not manually possible as a large-size data repository. In the present research, LDA topic modelling and clustering technique is used for tweets related to 'Clean India Mission'. The most trending topics and related terms are extracted by using LDA. The results show the connectivity and interest of the Indian masses to the mission. Sentiment analysis performed by lexicon-based approach shows that both negative and positive sentiments are reflected in the tweets posted related to the mission. Clustering further helps to partition the observations in diverse clusters that identify different aspects related to the corpus. It demonstrates and helps in finding what people are tweeting about and also discover the abstract topics. The analysis can be helpful in policy enhancement for the mission. A more elaborate analysis can be possible by retrieving tweets from other related handles and hashtags. Analysis of similar content posted on other social media platforms is also possible as opinion can be expressed there as well.

References

[1] Agarwal, A.; Xie, B.; Vovsha, I.; Rambow, O.; Passonneau, R. (2011): Sentiment Analysis of Twitter Data. In Proceedings of the workshop on Languages in Social Media, Columbia University, New York. 96(1), pp. 30-38.

[2] Ayo, F. E.; Folorunso, O; Ibharalu, F. T.; Osinuga, I. Ad.; A.-Alli, A. (2021): A probabilistic clustering model for hate speech classification in twitter. Expert Systems with Applications, vol. 173.

[3] Bastani, K.; Namavari, H.; Shaffer, J. (2019): Latent Dirichlet allocation (LDA) for topic modelling of the CFPB consumer complaints. Expert Systems with Applications, vol. 127, pp. 256-271.

[4] Blei, D. M.; Ng, A. Y.; Jordan, M. I. (2003): Latent Dirichlet Allocation. Journal of Machine Learning Research, pp. 993–1022.

[5] Curiskis, S. A.; Drake, B.; Thomas R. Osborn; Kennedy, P. J. (2020): An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. in Information Processing & Management, 57(2).

[6] Garcia, K.; Berton, L. (2021); Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. Applied Soft Computing, vol. 101.

[7] Gangadharan, V.; Gupta, D. (2020): Recognizing Named Entities in Agriculture Documents using LDA based Topic Modelling Techniques. Procedia Computer Science, vol. 171, pp. 1337-1345.

[8] Habibi, M.; Priadana, A.; Saputra, A. B.; Cahyo, P. W. (2021): Topic Modelling of Germans Related Content on Instagram Using Latent Dirichlet Allocation (LDA). Advances in Health Sciences Research, vol. 34.

[9] Kaila, R. P.; Krishna Prasad, A. V. (2020): Informational Flow on Twitter - Corona Virus Outbreak – Topic Modelling Approach. International Journal of Advanced Research in Engineering and Technology (IJARET), 11(3), pp. 128-134.

[10] Kuuliala, L.; Pérez-Fernández, R.; Tang, M.; Vanderroost, M.; Baets, B. De; Devlieghere, F. (2021): Probabilistic topic modelling in food spoilage analysis: A case study with Atlantic salmon (Salmo salar). International Journal of Food Microbiology, vol. 337.

[11] Kherwa, P.; Bansal, P. (2018): Topic Modelling: A comprehensive review, ICST Trans. Scalable Inf. Syst. 159623, 2018. https://doi.org/10.4108/eai.13-7-2018.159623.

[12] Kalepalli, Y.; Tasneem, S.; Phani Teja, P. D.; Manne, S. (2020): Effective Comparison of LDA with LSA for Topic Modelling. 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 1245-1250.

[13] Kurnaz, S.; Mahmood, M. A. (2019). Sentiment Analysis in Data of Twitter using Machine Learning Algorithms. In International Journal of Computer science and Mobile Computing. 8(3), pp .31 – 35.

[14] Maheshwari, S.; Shukla, S.; Kumari, D. (2019): Twitter Opinion Mining Using Sentiment Analysis. World Scientific News. An International Scientific Journal, pp. 78-87.

[15] Mohammed, S. H.; Al-augby, S. (2020): LSA & LDA Topic Modeling Classification: Comparison study on E-books. Indonesian Journal of Electrical Engineering and Computer Science, 19(1).
[16] Mohammad, Saif & Turney, Peter. (2013). NRC emotion lexicon. DOI: 10.4224/21270984.
[17] Mika V. Mantyla; Claes, M.; Farooq, U. (2018): Measuring LDA topic stability from clusters of replicated runs. In Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '18), Association for Computing Machinery, New York, NY, USA, Article 49, pp. 1–4.
[18] Onan, A.; Korukoğlu, S.; Bulut, H. (2016): LDA-based Topic Modelling in Text Sentiment Classification: An Empirical Analysis. International Journal of Computational Linguistics and Applications, 7(1), pp. 101–119.
[19] Parveen, N.; Santhi, M.V.B.T.; Burra, L. R.; Pellakuri, V.; Pellakuri, H. (2021): Women's e-commerce clothing sentiment analysis by probabilistic model LDA using R-SPARK. Materials Today: Proceedings 2021.
[20] Rania, A.; Yeap Tet Hin; Morad, B. (2020): Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis. Frontiers in Artificial Intelligence, vol. 3, pp. 42.
[21] Ray, D. (2017): Lexicon Based Sentiment Analysis of Twitter Data. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 5 (X).
[22] Sokolova, M.; Huang, K.; Matwin, S.; Ramisch, J.; Sazonova, V.; Black, R.; Orwa, C.; Ochieng, S.; Sambuli, N. (2016): Topic Modelling and Event Identification from Twitter Textual Data. In Social and Information Networks.
[23] Shah, C. S.; Sebastian, M. P. (2020): Sentiment Analysis and Topic Modelling of Indian Government's Twitter Handle #IndiaFightsCorona. In: Sharma S.K., Dwivedi Y. K., Metri B., Rana N.P. (eds) Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation. TDIT 2020. IFIP Advances in Information and Communication Technology, vol. 618. Springer, Cham.
[24] Wisnu, G. R. G.; Muttaqi, A. R.; Santoso, A. B.; Putra, P. K.; Budi, I. (2020): Sentiment Analysis and Topic Modelling of 2018 Central Java Gubernatorial Election using Twitter Data. *2020 International Workshop on Big Data and Information Security (IWBIS)*, Depok, Indonesia, pp. 35-40. DOI: 10.1109/IWBIS50925.2020.9255583.
[25] Yang, S.; Zhang, H. (2018): Text Mining of Twitter Data Using a Latent Dirichlet Allocation Topic Model and Sentiment Analysis. International Journal of Computer and Information Engineering, 12(7).

Authors Profile

Sangeeta Rani : She received B.Tech in Computer Science & Engineering in 2003 and M.Tech in 2008 from Maharishi Dayanand University of Rohtak (India). At present she is a research scholar at Maharishi Dayanand University, Rohtak in Department of Computer Science and Applications. Her area of research is "Twitter data Sentiment Classification". Her research interests include Data Mining, Text Mining, Pattern Recognition, Character Recognition, Natural Language Processing, Artificial Intelligence, and Big Data Analysis.

Email:sangeeta.yogi@gmail.com

Nasib Singh Gill: He is Professor and Head of department of Computer Science & Application at Maharishi Dayanand University, Rohtak. He is Director of University Computer Centre and MDU Alumni at M. D. University, Rohtak. He is having Post Doctoral Research (Computer Sc.) from Brunel University (UK), Ph.D. (Computer Sc.), Master's Degree in Science and MBA. His research interests includes Software Metrics, Component -based Metrics, Testing, Reusability, Data Mining and Data Warehousing, NLP, AOSD, Information and Network Security. He has written 6 books and Author of more than 150 publications.

Email:nasibsgill@gmail.com

Preeti Gulia: She is Assisatnt Professor in department of Computer Science & Application at Maharishi Dayanand University, Rohtak, India. She is Doctor of Philosophy in Computer Science from Maharishi Dayanand University, Rohtak. Her research interests include Machine Learning, Artificial Intelligence, Deep Learning, Neural Network, Big Data Mining, and Wireless Networks. She is an Author of several books and publications and research guide of Ph.D. Scholars.

Email:preetigulia81@gmail.com