

RESEARCH ARTICLE

 OPEN ACCESS

Received: 20-04-2022

Accepted: 18-08-2022

Published: 26-09-2022

Citation: Makwana K, Ganatra AP (2022) Textual Data Analysis of 'Mann Ki Baat' Show. Indian Journal of Science and Technology 15(37): 1859-1867. <https://doi.org/10.17485/IJST/v15i37.848>

* **Corresponding author.**

kirtimakwana.mba@charusat.ac.in

Funding: None

Competing Interests: None

Copyright: © 2022 Makwana & Ganatra. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN

Print: 0974-6846

Electronic: 0974-5645

Textual Data Analysis of 'Mann Ki Baat' Show

Kirti Makwana^{1*}, Amit P Ganatra²

¹ Assistant Professor, Faculty of Management Studies (FMS), Indukaka Ipcowala Institute of Management (I2IM), Charotar University of Science and Technology (CHARUSAT), Changa, Gujarat, India

² Vice Chancellor (Provost), Parul University, Vadodara, Gujarat, India

Abstract

Background: The present research gives insights about Textual Data analytics of "Mann Ki Baat" – a monthly radio programme hosted by Shri. Narendra Modi, the Prime Minister of India on All India Radio. Using techniques of textual data analytics, an analytical framework is designed in this research paper. **Methods:** The outline is applied to the corpus of twelve (12) episodes of Mann Ki Baat 2.0 (March 2020 – February 2021, after nationwide a lockdown in India) collected from PMIndia's official website and was studied in detail. Recurrently spoken words, diversity of topics covered, Topics Correlations, Sentiment Analysis, and Hierarchical Clustering were determined and analyzed. Further, word clouds for each episode were also generated. **Findings:** The research is a new approach of an application under the traditional "topic modeling" techniques on Mann Ki Baat 2.0. **Novelty and applications:** Additionally, this research is an attempt to extract the themes using statistical modeling deliberated in the popular radio programme – Mann Ki Baat 2.0.

Keywords: Textual Data Analytics Topics Correlations; Sentiment Analysis; Hierarchical Clustering; Word Clouds; Mann Ki Baat 2.0

1 Introduction

In the majority of the countries, radio is considered a popular medium of mass communication with the widest and highest outreach capacity. Radio is accessible by all irrespective of the economic background. The first radio station in India was established in 1927 in Mumbai. Further, it was set up in Calcutta and Delhi in 1936. All India Radio (AIR) services broadcast radio in India. Because of recent developments in science and technology, radio is playing a significant role in spreading awareness, knowledge, and information. Mann Ki Baat is an Indian radio programme, pioneered and hosted by Prime Minister, Shri Narendra Modi in which Shri. Modi addresses Indians. Till date, 74 episodes were aired till February 2021. Because of varied topics and aspects covered, spreading awareness about various important national agenda and actions taken by the government, important national and international happenings, etc. the show has gained popularity in every single segment of society within and outside India.

Because of the significant advancement of digitization, a huge volume of unstructured format data such as news stories, blogs, social networking websites like Twitter,

Facebook, etc. adds textual data in digital libraries. The computer cannot take directly the enormous quantity of data in an unstructured format for analysis. Thus, to extract the unseen patterns on, or after the text data, precise pre-processing methods and algorithms are required. Text analytics has numerous uses and applications. Textual data analytics applications are being extensively used in recent times. This can be used to identify the unseen associations, patterns, and models in qualitative data, descriptions, or unstructured data. It uncovers the novel information which is hidden in unstructured qualitative data. (Santiago M. Pinto, 2019) It can be used to recognize the emergent areas of concern or attention and trending themes in specific targeted groups. ⁽¹⁾ It derives new, earlier unknown, and thought-provoking information from text data. Text Analytic methods generally involve four steps. There are various text analytics tools. The list is as follows:

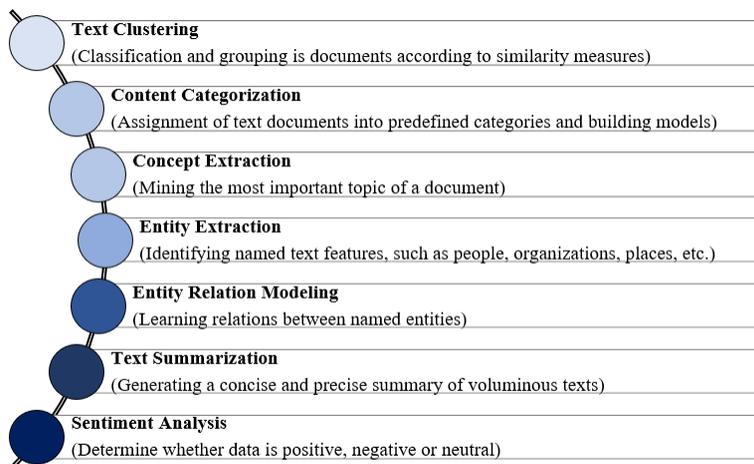


Fig 1. Text Analytics Tools

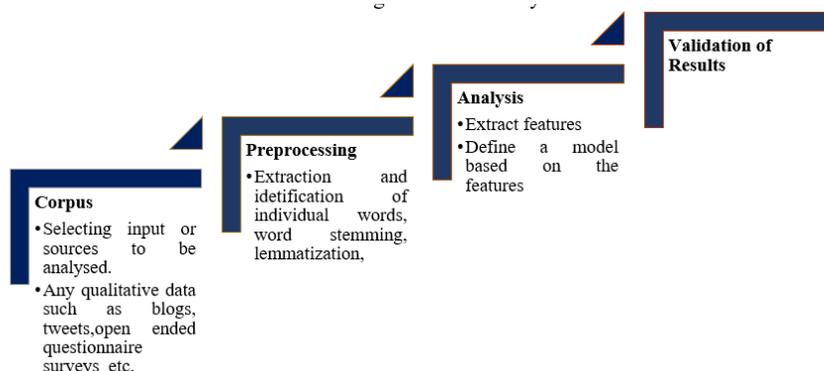


Fig 2. Steps of Text Analytics Method

Textual data analytics, in recent times, has appeared to a favored method to sort out bulky and gigantic textual content. It typically involves extracting hidden themes and explain the documents as per the themes. Earlier research studies illustrate the diverse methods and algorithms to organize documents. Hofmann (1999) proposed an EM algorithm in learning named “Probabilistic Latent Semantic Indexing” (PLSI). ⁽²⁾ The restrictions of the PLSI were inscribed by Blei, Ng, and Jordan (2003), reviewed the framework using the Bayesian model, popularly known as Latent Dirichlet Allocation (LDA), which is built on the “variation learning” approximation method. ⁽³⁾ Dredze, Mark, et al. (2008) established an unsupervised learning framework for creating summary keywords from emails. ⁽⁴⁾ Griffiths et al. (2004) analyzed synopses from the Proceedings of National Academy of Sciences (PNAS), USA, and recognized the numeral of themes in addition to disclosed the themes extracted from the data. ⁽⁵⁾

Griffiths and Steyvers (2005) similarly projected an effective approximation algorithm constructed on Gibbs sampling. ⁽⁶⁾ Lau et al. (2010) anticipated a topic labeling method in which the top ten topics are to be extracted to label the overall topic. ⁽⁷⁾

2 Materials and Method

2.1 Data Collection and Processing

For this exploratory research work, the researchers considered the English version of transcripts of the “Mann Ki Baat 2.0” show accessible on <https://www.pmindia.gov.in/en/mann-ki-baat/>.⁽⁸⁾ The researchers collected a total of twelve (12) episodes from March 2020 - February 2021. For this analysis, the researchers considered only these twelve episodes to see the response of Indian PM on Covid-19 and other key issues covered or not. Documents collected in text format possibly contain noise in various modes from distinct words to numerous sections containing distinct characters such as punctuation marks and figures etc. Henceforth, it becomes essential to perform data cleaning for further data analysis. Pre-processing is a technique to resolve such problems. For the same, the researchers pre-processed the data and data cleaning was performed.

Additionally, Python programming and its packages on Orange Canvas 16 to perform the textual data analytics. The synopsis of data collection is described in Table 1 :

Table 1. Summary of Mann Ki Baat 2.0 Episodes

Year	Episode No.:	Date	Broad Area of Talk
2020	10 th	29 March	• COVID-19 Pandemic in India
	11 th	26 April	• Contributions of arrangements by Indians to fight against the COVID-19 pandemic. • Admired farmers for continuing food demands. • The new normal - Practicing of wearing masks
	12 th	31 May	• Nationwide Lockdown • Crisis of Migrants • Relevance of Yoga during Pandemic
	13 th	28 June	• Self-reliance • Migrant workers • Pandemic amongst other things
	14 th	26 July	• Kargil Vijay Diwas • Pandemic • Madhubani masks • Sports and entrepreneurship
	15 th	30 August	• Onam • Made in India (MiI) Applications and toys • Teachers' Day • Dogs working in the security forces
	16 th	27 September	• Storytelling in India • Hitopadesha and Panchatantra • The surgical strikes' anniversary
	17 th	25 October	• Festivals • 'Vocal for Local' • Khadi • Mallakhamba • Pencil making in Pulwama
	18 th	29 November	• Guru Nanak Ji 's 550 th Prakash Parv
	19 th	27 December	• Tributes to Sikh Saints • Heritage Week • Indian scriptures
2021	20 th	31 January	• Active alumni network for educational institutions (colleges, universities, schools, and villages) through innovative approaches to engage the alumni
	21 st	28 February	• Farm protests • COVID-19 Vaccination drive • Performance of Indian cricket team in Australia
			• Be a warrior, not a worrier • Pariksha Pe Charcha • Nature Conservation • Aatmanirbhar Bharat Abhiyaan

Further, a general framework for textual data analysis was prepared on Orange Canvas 16.

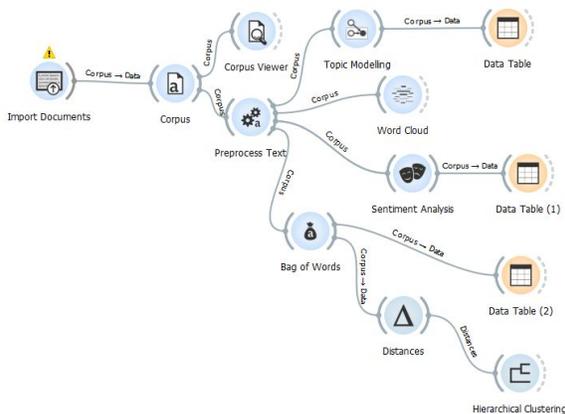


Fig 3. General Framework of Textual Data Analytics

2.2 Objectives

- To explore the recurrence of topics and words and extract the themes using statistical modeling deliberated in 12 episodes of the popular monthly addressing radio program Mann Ki Baat 2.0. (March 2020 to February 2021) by using textual data analysis techniques.
- To capture the sentiments of Mann Ki Baat episodes i.e. “positive”, “negative”, and “neutral”.
- To apply other textual data analysis techniques such as Topic correlation, Hierarchy clustering, and preparing word clouds from the text.

3 Results and Discussion

3.1 Top Ten Spoken Words

Using the Figure 3, the researchers extracted the top ten words spoken (in terms of frequency) in each Mann Ki Baat 2.0 episode. The list is illustrated in Table 2 :

Table 2. Top Ten Words Spoken (Frequency) in Each Episode of Mann Ki Baat 2.0

Word	Count	Word	Count	Word	Count	Word	Count
10th Episode		11th Episode		12th Episode		13th Episode	
Doctor	35	People	24	People	60	Country	39
Family	31	Country	24	Country	58	India	33
People	30	India	20	Corona	40	People	13
Corona	29	Corona	17	Yoga	30	Narasimha	05
						Rao Ji	
Home	23	World	16	Scheme	26	Calamity	05
Quarantine	17	Countryman	11	Friend	25	Corona Pan- demic	04
						Reliant India	04
Patients	16	Fight	10	Ayushman Bharat	22	Lockdown	03
Battle	12	Mask	09	Beneficiaries	12	Defense Sec- tor	03
Nurse	09	Pandemic	07	Eastern India/Region	12		

Continued on next page

Table 2 continued

Social Media	04	Government / Administration	06	Honest Tax Payer	06	Family Member	03
14th Episode		15th Episode		16th Episode		17th Episode	
Country	24	Toys	47	Story	44	People	19
People	22	Country	30	Farmers	28	Khadi	18
Corona	15	Children	20	Family	21	Unity	15
Countryman	15	Games	16	Life	21	Farmers	14
Mantra	08	Students	11	Vegetables	20	Library	13
Atal Ji	04	Teachers	11	Today	16	Manzoor	05
						Bhai	
Dragon Fruit	04	Indian Breed	09	King	16	Maharishi	05
Gandhi Ji	03	Nutrition Month	04	Bhagat Singh	10	Valmiki	
						Pencil Slat	04
Armed Force Battle of Kargil Kargil War	02	Innovation Challenge Bharat App Innovation	03	Fruit	09	Khadi Mask	04
Farmers of Kutch	02	Cooperation / Freedom Movement	03	Tenali Rama	07	Local Products	04
18th Episode		19th Episode		20th Episode		21st Episode	
Farmers	22	Country	21	Vegetables Market	06	People	23
Institutions	12	People	21	Rice Mill	05	Friend	19
Alumni	11	Kashmiri Saffron	21	Jhansi	05	Pride	07
Opportunities	10	India	16	Namo App	04	Chia Seed	04
Gurudwara	12	Shri Guru Gobind	11	Road Safety	04	Water Source	04
Guru Sahib / Darbaar Sahib	09	Leopard	05	Toll Plaza	03	Reliant India	04
Sri Aurobindo	08	Social Media	04	Strawberry Festival	03	Pariksha Pe Charcha	03
Guru Nanak Dev	08	Someshwar Beach	03	Incredible India Weekend	02	Drum Stick	03
Bird Watching	04	Srinivasacharya Swami	02	Dr. Rajendra Prasad	02	Bank of River	02
Agricultural Reform	03	Guru Tegh Bahadurji	02	National Yoga Day	02	Corona	02

Table 2 depicts that all the episodes of the Mann Ki Baat program are different from March 2020 to February 2021. A variety of generalized concepts and topics are discussed by Prime Minister Modi on the show.

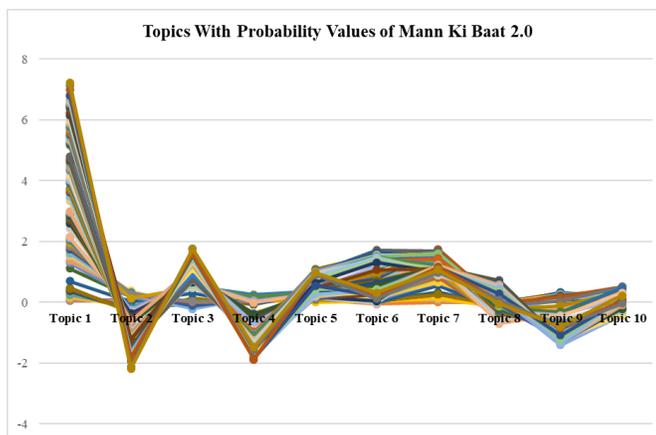


Fig 4. Topic Extraction of Mann Ki Baat 2.0

3.2 Topic Correlation Analysis

As shown in Figure 5, Topic 1 concerning Topic 2 and Topic 4 has a moderate positive significant relationship, whereas, $r = -0.2$ suggests a weak or negative association among topics, further, correlation close to zero suggests no linear association between the two topics. This specifies that, during all twelve episodes of Mann Ki Baat, PM Modi covered diverse and wide-ranging topics.

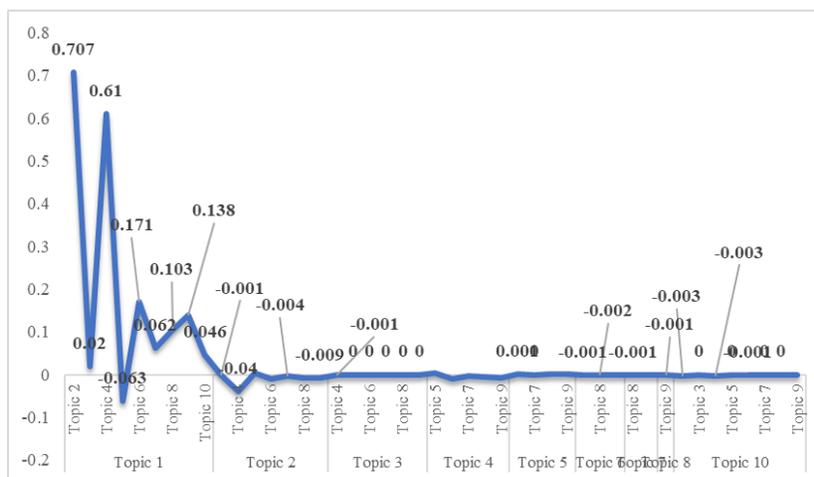


Fig 5. Topic Correlation Analysis

3.3 Word Clouds of Mann Ki Baat 2.0

The figure 6 demonstrates the word clouds of Mann Ki Baat. As seen, “Corona”, “Corona Virus”, “Pandemic”, “Quarantine” and Health-care related words are the most highlighted words starting from March – June 2020. It is worthy to note that, COVID -19 / Corona virus pandemic was at pick during this time. Further, “Country”, “Countrymen”, “People”, and “India / Nation” are the words which were spoken repetitively in each episode of MKB 2.0 which shows the ideology where the country, and people are the primary important stakeholders and PM conveys that, he is working for their better and eager to learn about their ideas, issues, and concerns. Shri. Modi makes sure that every citizen of the nation is actively involved in the development path of the nation. Further, “Friend (s) word frequently used during all the episodes indicates, his rapport with the audience. Using this word PM sounds very friendly, free, and open-minded.

Table 3. Sentiment Analysis of Mann Ki Baat 2.0

Sr. No.:	Episode No.:	Positive Sentiment	Negative Sentiment	Neutral Sentiment
1	10 th Episode	0.22	0.083	0.698
2	11 th Episode	0.149	0.065	0.787
3	12 th Episode	0.12	0.059	0.822
4	13 th Episode	0.172	0.044	0.784
5	14 th Episode	0.159	0.042	0.799
6	15 th Episode	0.171	0.023	0.806
7	16 th Episode	0.149	0.031	0.82
8	17 th Episode	0.187	0.014	0.799
9	18 th Episode	0.177	0.021	0.803
10	19 th Episode	0.131	0.025	0.844
11	20 th Episode	0.149	0.025	0.826
12	21 st Episode	0.158	0.014	0.828

3.5 Hierarchical Clustering

Hierarchical clustering obtains the sibling-sibling associations among topics and organizes the topics into a hierarchical tree. Figure 8 shows the topic hierarchy from the textual data of twelve episodes of Mann Ki Baat.

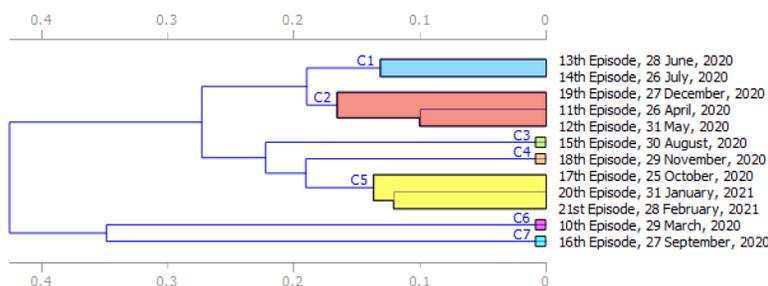


Fig 8. Hierarchical Clustering of Mann Ki Baat 2.0

As seen from figure 8, twelve episodes of Mann Ki Baat can be divided into seven different clusters. The vertical axis shows the clusters, whereas the horizontal scale on the dendrogram represents the distance among the clusters. Since the x-axis shows how close the observations were when they were merged into clusters. Looking at the dendrogram of the Mann Ki Baat data, there are clearly two very distinct groups. The right-hand side group contains two more dissimilar clusters. Whereas, most of the observations in the left group are clustering together at about the same height.

4 Conclusion and Future Research Scope

Textual Data Analytics plays a powerful technique for qualitative data analysis to extract meaning from the text. This research paper explored and extracted the topics and themes from twelve episodes of Mann Ki Baat 2.0. The work briefly conveys how topic modeling derived topics models from the corpus. Further, the result shows that Shri. Narendra Modi, PM of India addressed the citizens of India covering a wide range of topics, various initiatives, and schemes undertaken by the government. “Corona”, “Corona Virus”, “Pandemic”, “Quarantine” and Health-care related words are the most highlighted words starting from March – June 2020 as seen in the word clouds of Mann Ki Baat. Topics such as Thanking Frontline workers for Covid – 19 Pandemic, Pariksha Pe Charcha, Farmers of India, Innovation Challenge, etc. spread positivity among Indians and gave a ray of hope during a hard time of the pandemic. In almost all the episodes PM mentioned some names of famous personalities and upcoming special days/festivals/events in India which makes a rapport with the common man and builds a positive image of the ruling government and acts as a public relation tool for the PM. Further, there seem to be more neutral Mann Ki Baat episodes

compared to positive and negative sentiments. Negative sentiments are hardly observed in all the selected episodes of Mann Ki Baat. This research paper is limited to Mann Ki Baat episodes aired from March 2020 to February 2021. The researchers have avoided discussion on the topic on various social media platforms which can potentially influence the results and sentiment associated with textual information. Nevertheless, these restrictions similarly open up prospects for more research into the left-out knowledge domains. Further, the research can be extended to take related content from Twitter, Facebook, any other social media platforms, and more challenging textual data analysis can be done.

References

- 1) Pinto SM. Sentiment Analysis of the Fifth District Manufacturing and Service Surveys, *Economic Quarterly*. 2019;105:164–164.
- 2) Hofmann T. Probabilistic latent semantic indexing. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '99*. 1999.
- 3) Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res*. 2003. Available from: <https://doi.org/10.1016/b978-0-12-411519-4.00006-9>.
- 4) Dredze M, Wallach HM, Puller D, Pereira F. Generating summary keywords for emails using topics. *Proceedings of the 13th international conference on Intelligent user interfaces - IUI '08*. 2008.
- 5) Griffiths TL, Steyvers M, Blei DM, Tenenbaum JB. Integrating topics and syntax. *Advances in Neural Information Processing Systems*. 2005.
- 6) Griffiths TL, Steyvers M. Finding scientific topics. *Proceedings of the National Academy of Sciences*. 2004;101(suppl_1):5228–5235. Available from: <https://doi.org/10.1073/pnas.0307752101>.
- 7) Lau JH, Newman D, Karimi S, Baldwin T. Latent Dirichlet allocation. *Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference*. 2010.
- 8) Available from: <https://www.pmindia.gov.in/en/mann-ki-baat/>.
- 9) Merriam-Webster. Sentiment. 2016. Available from: <https://www.merriam-webster.com/dictionary/sentiment>.
- 10) Yi J, Nasukawa T, Bunescu R, Niblack W. Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. *Third IEEE International Conference on Data Mining*. 2003;p. 427–434.