# Topic Modelling Extraction of "Mann Ki Baat"

Mounika Kandukuri, and V.V.HaraGopal

***Abstract*—The purpose of this study is to give an insight about Textual Data analytics and its application in the analysis of unique public relations campaign "Mann Ki Baat" that was initiated by incumbent Prime Minister of India, honourable "Mr. Narendra Modi" which was initially aired on All India Radio Programme on Vijaya Dashami on October 3rd, 2014 followed by second on November 2nd, 2014 of the same year till December 2019. In this paper, an analytical framework is designed using a powerful technique of textual data analytics "Topic Modelling based on LDA (Latent Dirichlet Allocation)" to accomplish the study. The proposed framework is applied to the corpus of 60 episodes (October 2014 to December 2019) of Mann Ki Baat gathered from PMindia website and was analyzed in greater detail. The terms used frequently and recurrence of the topics spoken in his popular monthly radio address program were determined and analyzed from both in statistical and dynamic perspectives. In this context the present study is a first approach of application under the conventional technique "topic modelling" on Mann Ki Baat. Further, this is the principal endeavour to excerpt the themes discussed in radio programme using statistical modelling.**

***Index Terms*—** Document Term Matrix (DTM), Latent Dirichlet Allocation (LDA), Mann Ki Baat (MKB), Textual Data Analytics, Textual Mining, Topic Modelling.

## I. Introduction

With the rise of new technologies and its usage in almost every area have made available immense quantities of digital text. This text data arriving from multiple sources at an alarming speed cannot be processed by computers to extract hidden insights [1]. For this purpose, there is a need for specific pre-processing methods and algorithms to mine useful patterns from text data. Textual Data Analytics is a task used for processing text data to derive the high quality of information and to discover patterns from text [2]. Textual Data Analytics tasks include text categorization, text summarization, document summarization, and keyword extraction, etc.., [3].

When we have an extensive collection of text documents, analyzing those documents to extract essential information is a challenging task. Topic Modelling is one of the most essential text mining techniques that can be used to extract underlying topics and themes from a massive archive of documents [4]. This topic illustration is attained by assuming that each document was formed through some generative process. There exist distinct types of topic models in the literature. The Latent Dirichlet Allocation is proven to be very popular and successful technique over the years to elicit the concealed topics from a vast content of text [4].

'Mann Ki Baat,' is an Indian radio program hosted by Prime Minister Narendra Modi, a popular and ubiquitous monthly radio address through which he renders his voices about the prospects of India under his regime, shares his experiences and ideas to the general mob of India[5][6]. He has chosen radio to be the medium of the program to reach every isolated region of the country. The first Mann Ki Baat program was aired on the occasion of Vijayadashami on October 3rd, 2014, followed by second on November 2nd, 2014, has gone on for more than 60 episodes and counting[6]. A survey was conducted in 2014 to estimate the success reveals that 66.7% of the population had tuned to listen to this program. Given its stupendous success, there has been ample curiosity to know the recurrence of the topics and common terms used in his monthly address to discover the various issues Prime Minister focusing and the reasons it became a sensation in every section of society[5].

In this paper, an attempt had been made designing framework using popular unsupervised textual data analytic technique "Topic modeling based on LDA "and analyzing the performance of framework on a corpus of 60 episodes of Unique public relations campaign "Mann Ki Baat" initiated by the honorable Prime Minister. The results are indicative of the frequent terms and recurrence of the topics entailed in the Mann Ki Baat program...

## II. Related Work

Topic Modelling, in the recent past, emerged as a preferred way to sort out large and massive web content. It usually refers to the action of extracting hidden topics and annotate the documents according to themes. Previous studies show the different techniques and algorithms developed for organizing documents.

EM algorithm in learning is proposed by Hofmann [7] using the name "Probabilistic Latent Semantic Indexing" or PLSI. Some of the limitations of the PLSI were addressed by Blei, Ng, and Jordan [8], revised the model and learning framework using the Bayesian model. This framework is called Latent Dirichlet Allocation, which is based on another approximation technique called "variation learning." Dredze, Mark, et al. [9] developed an unsupervised learning framework LDA for generating summary Keywords from emails. Griffiths et al. [10]using Bayesian Model selection analyzed abstracts from Proceedings of National Academy of Sciences (PNAS)of USA and established the number of topics and showed that the topics extracted grabbed the

Published on January 27, 2021.
Mounika Kandukuri**,** Department of Statistics, University College of Science, Osmania University, India
e-mail: mounika4324@gmail.com

V.V. HaraGopal, Department of Mathematics, BITS-Pilani, India.
e-mail: haragopal_vajjha@yahoo.com, haragopal_vajjha@hyderabad.bits-pilani.ac.in

meaningful structure in the data. Griffiths and Steyvers also proposed an efficient estimation algorithm based on Gibbs Sampling [10]. Lau et al. [11] proposed a topic labeling approach via best term, selecting one of the top ten topic terms to label the overall topic. Various distinct extensions to the primary topic model have also been developed, entailing topic models for images and text [8] [12], author-topic models [13], author-role-topic models [14] and hidden-Markov topic models for segregating semantic and syntactic topics [15]. Mean While Sentiment analysis of Mann Ki Baat tweets of year 2018 is also performed to analyze sentiment of the show [5].

In the Proposed Work, Topic modeling using LDA is applied on Mann Ki Baath episodes to know the insights of the speeches delivered. It involves pre-processing of documents, document modeling using LDA, and evoking top words of the topics

## III. MATERIALS AND METHODS

This section includes techniques like Data collection, Pre-processing, topic modeling with LDA. The Detailed content of the methods is explained below:

### *A. Data Collection*

For this study, we considered the written English transcripts of the “Mann Ki Baat “show available in https://www.pmindia.gov.in/en/mann-ki-baat/.”PmIndia.gov.in “Website. We have collected written a total of 60 episodes of the show from October 2014 to December 2019. We used R programming and its packages such as tm and topic models to perform the .The Overview of Data collected is explained in the following figure1 as

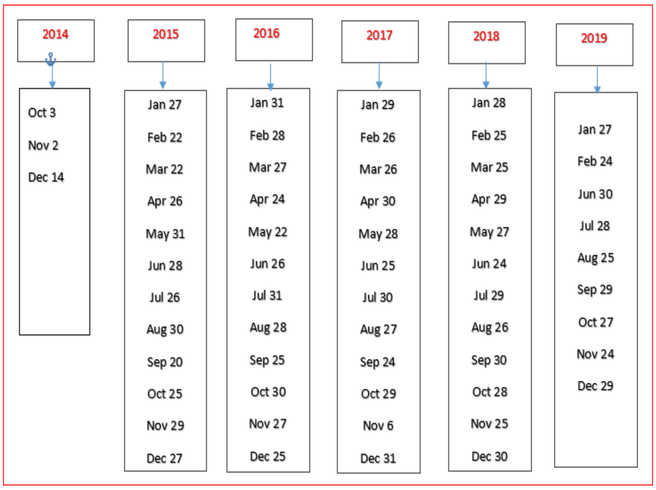


Fig.1. Mann Ki Baath 60 programme episodes detailed overview

### *B. Pre-processing*

Text documents collected may consist of a lot of noise and maybe in a variety of forms from individual words to multiple paragraphs consisting of special characters like punctuation marks and numbers etc.., hence, it is necessary to clean the data for extracting exact hidden information. Pre-processing is a method of resolving such issues. Pre-processing involves the Removal of numeric values, stop words, Lower case conversion, Stemming [16]. This is an important step as it helps in reducing the dimensionality of data by transforming raw data into an understandable format. The following code explains the process of creating text corpora of all documents in R. [16].

```
Install. Packages ("tm")
docs<-Corpus (DirSource ("path to your folder"))
```

The algorithmic approach of Pre-processing steps is explained in the following figure.

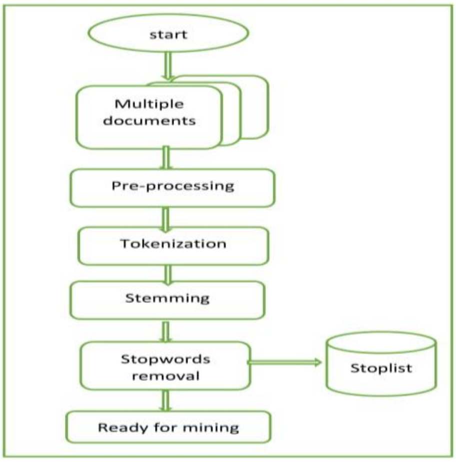


Fig. 2.Pre-processing steps in Text data

### *C. Topic Modelling with Latent Dirichlet Allocation*

Topic modeling is a probabilistic based method which is primarily used for organizing, and summarizing large electronic documents [17]. It is an unsupervised technique since there is no predefined classification; it is based on the word frequency distribution of the documents to determine the various topics. The topic can also be termed as a probability distribution over words by which new documents can be generated, there are different types of topic models such as Mixture of Multinomial, Gamma-Poisson, Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA)[4]. But among all the available topic models, LDA has proven to be a very successful and most popular topic model for organizing a large collection of archives. In LDA, the general assumption is that each document in the corpus is derived from various distinct topics with varying probabilities [18].
The general framework of topic modeling is illustrated in the following figure.

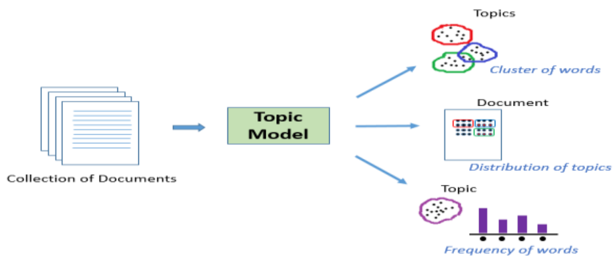


Fig. 3.Flowchart of Topic Modelling Using LDA

All topic models are based on the same basic assumption that

1. Each document consists of a mixture of topics.
2. Each topic consists of a collection of words.

Blei et al.explained Latent Dirichlet Allocation (LDA) as a Bayesian Probabilistic Latent semantic analyzing method using Dirichlet priors for the document-topic and word –topic distributions [8]. In the process for LDA, The outcomes from Dirichlet are used to allot the terms in the document to different topics It is also termed as a Bayesian Mixture model for discrete data in which themes are uncorrelated. It is also known as a three-level Bayesian model in which each item of accumulation is modelled as a finite mixture over an underlying set of topics.

LDA assumes the following generative process for a Corpus D comprising of M documents, with document d having $N_d$ words (d€1, 2...M).

a) Choose a multinomial distribution $\varphi_t$ for topic t (t ϵ (1….K)) from Dirichlet distribution with parameter β.

b) Choose a multinomial distribution θd for document d (d ϵ (1….M)) from a Dirichlet distribution with parameter α.

c) For a word Wn (n ϵ (1, 2… $N_d$)) in document d.

1. Select a topic $Z_n$ from $\theta_d$
2. Select a word $W_n$ from $\varphi_z$

The general process of latent Dirichlet Allocation using Plain model is explained in the following figure

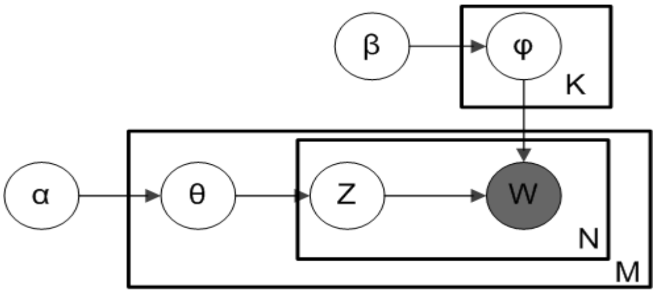


Fig. 4. Graphical Model representation of LDA

Where

K- Number of topics; N-Number of words in the document

M-Number of documents to analyze;α- Dirichlet-Prior concentration parameter of the per-document topic distribution

φ(k)-Word distribution for topic K;θ(i)-topic distribution for topic i

Z(i,j)- topic assignment for W(i,j);W(i,j)-jth word in the ith document

Φ and θ are Dirichlet distribution, Z and W are multinomial.

## IV. Experimental Results

This section highlights the experimental results obtained through designed analytical framework Topic modelling with LDA performed on Mann Ki Baat 60 episodes from October 2014 to December 2019. For performing computations, R software was utilized, specifically to perform text mining tasks such as creating a corpus, text pre-processing initially “tm” packages were loaded. Thereafter “topic models “packages and its dependencies were installed to perform topic modelling with LDA. As an initial step Corpus was created and text corpora were preprocessed. To determine the topics using topic modelling with LDA initially the optimal number of topics spoken were determined using the simple harmonic mean method from Martins work and the experimental results are illustrated in figure 5.

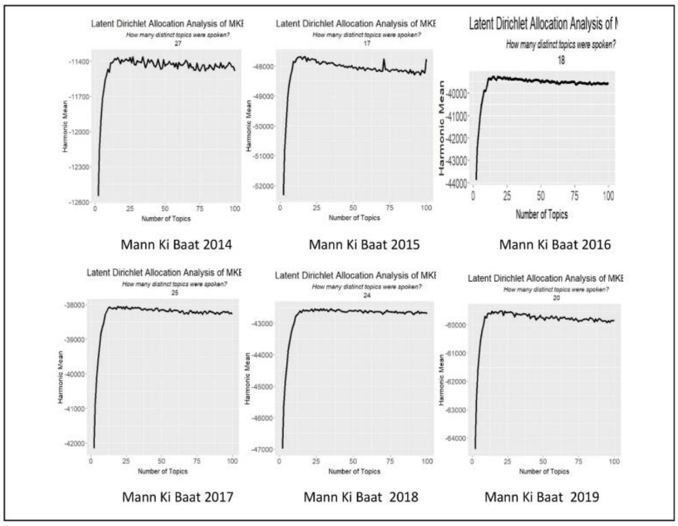


Fig. 5.Optimal number of topics spoken in Mann Ki Baat from 2014 to 2019

TABLE I: REPRESENTS THE OPTIMAL NUMBER OF TOPICS SPOKEN FROM 2014 TO 2019

| Year | Number of Topics Spoken in a Year |
|---|---|
| 2014 | 27 |
| 2015 | 17 |
| 2016 | 18 |
| 2017 | 25 |
| 2018 | 24 |
| 2019 | 20 |

From the Figure 5 and Table 1, the number of topics spoken are much varied with Minimum Topics spoke is 17 and maximum being 27 Topics. Also during 2015, 2016 the number of Topics spoke were less than 20 while for 2014 were much high and between 20-27 Topics which evident with the graph above. Hereafter LDA model is executed on text corpora and the outcome of the model is a matrix of topic probabilities. The document topic probability values are presented in below Table 2 to Table 7.

TABLE II: DOCUMENT TOPIC PROBABILITY VALUES OF MKB 2014

| Month | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 |
|---|---|---|---|---|---|---|---|---|---|
| October | 0.005647 | 0.465053 | 0.003667 | 0.015548 | 0.025449 | 0.015548 | 0.009608 | 0.015548 | 0.007627 |
| November | 0.010679 | 0.010679 | 0.069073 | 0.012503 | 0.017978 | 0.007029 | 0.012503 | 0.025277 | 0.025277 |
| December | 0.025223 | 0.003148 | 0.26253 | 0.046194 | 0.015289 | 0.056128 | 0.016393 | 0.019704 | 0.006459 |

| Month | Topic 10 | Topic 11 | Topic 12 | Topic 13 | Topic 14 | Topic 15 | Topic 16 | Topic 17 | Topic 18 |
|---|---|---|---|---|---|---|---|---|---|
| October | 0.003667 | 0.02941 | 0.003667 | 0.011588 | 0.011588 | 0.015548 | 0.013568 | 0.02941 | 0.160103 |
| November | 0.063598 | 0.014328 | 0.023452 | 0.076372 | 0.266153 | 0.003379 | 0.008854 | 0.032576 | 0.007029 |
| December | 0.00977 | 0.048402 | 0.019704 | 0.014185 | 0.006459 | 0.077099 | 0.021912 | 0.015289 | 0.003148 |

| Month | Topic 19 | Topic 20 | Topic 21 | Topic 22 | Topic 23 | Topic 24 | Topic 25 | Topic 26 | Topic 27 |
|---|---|---|---|---|---|---|---|---|---|
| October | 0.005647 | 0.005647 | 0.011588 | 0.013568 | 0.007627 | 0.005647 | 0.063073 | 0.011588 | 0.03337 |
| November | 0.109219 | 0.010679 | 0.061773 | 0.008854 | 0.008854 | 0.049 | 0.008854 | 0.014328 | 0.0417 |
| December | 0.011978 | 0.088137 | 0.013082 | 0.061647 | 0.072684 | 0.013082 | 0.008667 | 0.052817 | 0.010874 |

TABLE III: DOCUMENT TOPIC PROBABILITY VALUES OF MKB 2015

| Month | Topic 1 | Topic 2 | Topic3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 |
|---|---|---|---|---|---|---|---|---|---|
| Jan | 0.0095 | 0.0353 | 0.0108 | 0.0054 | 0.0108 | 0.0176 | 0.0108 | 0.0067 | 0.0258 |
| Feb | 0.0130 | 0.0113 | 0.8006 | 0.0048 | 0.0081 | 0.0195 | 0.0081 | 0.0162 | 0.0081 |
| Mar | 0.0081 | 0.0116 | 0.0163 | 0.0081 | 0.0046 | 0.0081 | 0.0093 | 0.0163 | 0.0163 |
| April | 0.0339 | 0.0496 | 0.0129 | 0.0234 | 0.0391 | 0.0339 | 0.0365 | 0.0339 | 0.0129 |
| May | 0.0131 | 0.0263 | 0.0247 | 0.0131 | 0.0065 | 0.0197 | 0.0263 | 0.7338 | 0.0098 |
| June | 0.0283 | 0.0216 | 0.0166 | 0.6493 | 0.0483 | 0.0133 | 0.0216 | 0.0166 | 0.0099 |
| July | 0.0187 | 0.0130 | 0.0318 | 0.0055 | 0.0055 | 0.0751 | 0.5788 | 0.0506 | 0.0187 |
| Aug | 0.0097 | 0.0195 | 0.0136 | 0.0253 | 0.0410 | 0.0469 | 0.0332 | 0.0195 | 0.0155 |
| Sep | 0.0174 | 0.0174 | 0.0204 | 0.0160 | 0.0116 | 0.0160 | 0.0204 | 0.0204 | 0.7240 |
| Oct | 0.0128 | 0.0075 | 0.0214 | 0.0042 | 0.0128 | 0.0247 | 0.0107 | 0.0300 | 0.0225 |
| Nov | 0.0229 | 0.0297 | 0.0269 | 0.0269 | 0.0107 | 0.0067 | 0.0080 | 0.0080 | 0.0175 |
| Dec | 0.0084 | 0.0204 | 0.0187 | 0.0289 | 0.7081 | 0.0221 | 0.0255 | 0.0067 | 0.0153 |

| Month | Topic 10 | Topic 11 | Topic 12 | Topic 13 | Topic 14 | Topic 15 | Topic 16 | Topic 17 |
|---|---|---|---|---|---|---|---|---|
| Jan | 0.0081 | 0.0163 | 0.0108 | 0.0135 | 0.0122 | 0.0081 | 0.7819 | 0.0163 |
| Feb | 0.0130 | 0.0081 | 0.0081 | 0.0113 | 0.0113 | 0.0211 | 0.0113 | 0.0260 |
| Mar | 0.0151 | 0.8212 | 0.0058 | 0.0034 | 0.0081 | 0.0233 | 0.0081 | 0.0163 |
| April | 0.0156 | 0.0234 | 0.0103 | 0.0234 | 0.5915 | 0.0208 | 0.0260 | 0.0129 |
| May | 0.0065 | 0.0412 | 0.0065 | 0.0065 | 0.0164 | 0.0148 | 0.0214 | 0.0131 |
| June | 0.0183 | 0.0183 | 0.0166 | 0.0249 | 0.0099 | 0.0199 | 0.0116 | 0.0550 |
| July | 0.0206 | 0.0544 | 0.0074 | 0.0488 | 0.0074 | 0.0262 | 0.0243 | 0.0130 |
| Aug | 0.5302 | 0.0860 | 0.0155 | 0.0645 | 0.0292 | 0.0175 | 0.0273 | 0.0058 |
| Sep | 0.0160 | 0.0131 | 0.0087 | 0.0218 | 0.0320 | 0.0116 | 0.0116 | 0.0218 |
| Oct | 0.0118 | 0.0042 | 0.0107 | 0.7580 | 0.0139 | 0.0376 | 0.0075 | 0.0096 |
| Nov | 0.0161 | 0.0188 | 0.7094 | 0.0148 | 0.0432 | 0.0148 | 0.0107 | 0.0148 |
| Dec | 0.0153 | 0.0118 | 0.0118 | 0.0170 | 0.0187 | 0.0357 | 0.0187 | 0.0170 |

TABLE IV: DOCUMENT TOPIC PROBABILITY VALUES OF MKB 2016

| Month | Topic 1 | Topic 2 | Topic3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 |
|---|---|---|---|---|---|---|---|---|---|
| Jan | 0.0095 | 0.0353 | 0.0108 | 0.0054 | 0.0108 | 0.0176 | 0.0108 | 0.0067 | 0.0258 |
| Feb | 0.0130 | 0.0113 | 0.8006 | 0.0048 | 0.0081 | 0.0195 | 0.0081 | 0.0162 | 0.0081 |
| Mar | 0.0081 | 0.0116 | 0.0163 | 0.0081 | 0.0046 | 0.0081 | 0.0093 | 0.0163 | 0.0163 |
| April | 0.0339 | 0.0496 | 0.0129 | 0.0234 | 0.0391 | 0.0339 | 0.0365 | 0.0339 | 0.0129 |
| May | 0.0131 | 0.0263 | 0.0247 | 0.0131 | 0.0065 | 0.0197 | 0.0263 | 0.7338 | 0.0098 |
| June | 0.0283 | 0.0216 | 0.0166 | 0.6493 | 0.0483 | 0.0133 | 0.0216 | 0.0166 | 0.0099 |
| July | 0.0187 | 0.0130 | 0.0318 | 0.0055 | 0.0055 | 0.0751 | 0.5788 | 0.0506 | 0.0187 |
| Aug | 0.0097 | 0.0195 | 0.0136 | 0.0253 | 0.0410 | 0.0469 | 0.0332 | 0.0195 | 0.0155 |
| Sep | 0.0174 | 0.0174 | 0.0204 | 0.0160 | 0.0116 | 0.0160 | 0.0204 | 0.0204 | 0.7240 |
| Oct | 0.0128 | 0.0075 | 0.0214 | 0.0042 | 0.0128 | 0.0247 | 0.0107 | 0.0300 | 0.0225 |
| Nov | 0.0229 | 0.0297 | 0.0269 | 0.0269 | 0.0107 | 0.0067 | 0.0080 | 0.0080 | 0.0175 |
| Dec | 0.0084 | 0.0204 | 0.0187 | 0.0289 | 0.7081 | 0.0221 | 0.0255 | 0.0067 | 0.0153 |

| Month | Topic 10 | Topic 11 | Topic 12 | Topic 13 | Topic 14 | Topic 15 | Topic 16 | Topic 17 | Topic 18 |
|---|---|---|---|---|---|---|---|---|---|
| Jan | 0.01316 | 0.01088 | 0.02911 | 0.01316 | 0.01316 | 0.00861 | 0.02227 | 0.71703 | 0.01772 |
| Feb | 0.00610 | 0.01418 | 0.00772 | 0.00933 | 0.01095 | 0.00933 | 0.79932 | 0.00933 | 0.02387 |
| Mar | 0.00710 | 0.72891 | 0.00898 | 0.02590 | 0.00522 | 0.00898 | 0.02214 | 0.00522 | 0.02214 |
| Apr | 0.01548 | 0.03488 | 0.00843 | 0.01724 | 0.01372 | 0.00666 | 0.01548 | 0.00666 | 0.02959 |
| May | 0.00631 | 0.01632 | 0.65405 | 0.04470 | 0.01966 | 0.03302 | 0.02300 | 0.00965 | 0.00631 |
| Jun | 0.00750 | 0.02734 | 0.02932 | 0.01345 | 0.00551 | 0.00750 | 0.01742 | 0.00551 | 0.01345 |
| Jul | 0.01179 | 0.00483 | 0.00657 | 0.00831 | 0.02222 | 0.00831 | 0.01005 | 0.00831 | 0.00831 |
| Aug | 0.01703 | 0.00439 | 0.00913 | 0.03598 | 0.00913 | 0.01071 | 0.00597 | 0.00755 | 0.01229 |
| Sep | 0.04724 | 0.01153 | 0.01153 | 0.02683 | 0.00813 | 0.00813 | 0.01153 | 0.01833 | 0.01663 |
| Oct | 0.86131 | 0.00670 | 0.00670 | 0.01024 | 0.00847 | 0.01556 | 0.00847 | 0.00670 | 0.00493 |
| Nov | 0.01745 | 0.01149 | 0.00751 | 0.00552 | 0.00751 | 0.84250 | 0.01149 | 0.00751 | 0.00950 |
| Dec | 0.00588 | 0.01679 | 0.00744 | 0.02146 | 0.63205 | 0.14140 | 0.00744 | 0.00900 | 0.01367 |

TABLE V: DOCUMENT TOPIC PROBABILITY VALUES OF MKB 2017

| Month | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 | Topic 10 | Topic 11 | Topic 12 | Topic 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan | 0.010 | 0.008 | 0.004 | 0.008 | 0.004 | 0.004 | 0.865 | 0.004 | 0.004 | 0.008 | 0.006 | 0.006 | 0.004 |
| Feb | 0.005 | 0.003 | 0.015 | 0.007 | 0.027 | 0.013 | 0.013 | 0.020 | 0.007 | 0.022 | 0.015 | 0.003 | 0.015 |
| Mar | 0.013 | 0.023 | 0.006 | 0.006 | 0.010 | 0.010 | 0.017 | 0.006 | 0.027 | 0.012 | 0.013 | 0.709 | 0.006 |
| Apr | 0.008 | 0.008 | 0.018 | 0.036 | 0.006 | 0.026 | 0.024 | 0.016 | 0.006 | 0.004 | 0.012 | 0.016 | 0.018 |
| May | 0.658 | 0.020 | 0.025 | 0.016 | 0.007 | 0.007 | 0.025 | 0.004 | 0.009 | 0.004 | 0.009 | 0.011 | 0.025 |
| Jun | 0.026 | 0.020 | 0.006 | 0.014 | 0.008 | 0.012 | 0.012 | 0.010 | 0.650 | 0.008 | 0.010 | 0.006 | 0.032 |
| Jul | 0.010 | 0.010 | 0.004 | 0.006 | 0.018 | 0.012 | 0.010 | 0.021 | 0.016 | 0.004 | 0.014 | 0.010 | 0.012 |
| Aug | 0.009 | 0.009 | 0.007 | 0.007 | 0.009 | 0.026 | 0.011 | 0.020 | 0.006 | 0.013 | 0.009 | 0.007 | 0.028 |
| Sep | 0.009 | 0.028 | 0.006 | 0.021 | 0.006 | 0.006 | 0.009 | 0.017 | 0.009 | 0.011 | 0.011 | 0.004 | 0.006 |
| Oct | 0.009 | 0.018 | 0.009 | 0.701 | 0.018 | 0.018 | 0.024 | 0.018 | 0.004 | 0.007 | 0.004 | 0.004 | 0.018 |
| Nov | 0.007 | 0.012 | 0.019 | 0.012 | 0.012 | 0.014 | 0.007 | 0.009 | 0.012 | 0.751 | 0.007 | 0.012 | 0.012 |
| Dec | 0.007 | 0.013 | 0.041 | 0.004 | 0.009 | 0.015 | 0.007 | 0.020 | 0.026 | 0.009 | 0.037 | 0.017 | 0.007 |

| Month | Topic 14 | Topic 15 | Topic 16 | Topic 17 | Topic 18 | Topic 19 | Topic 20 | Topic 21 | Topic 22 | Topic 23 | Topic 24 | Topic 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan | 0.004 | 0.008 | 0.008 | 0.004 | 0.004 | 0.004 | 0.004 | 0.010 | 0.004 | 0.004 | 0.004 | 0.008 |
| Feb | 0.035 | 0.013 | 0.010 | 0.005 | 0.013 | 0.005 | 0.705 | 0.007 | 0.007 | 0.017 | 0.003 | 0.015 |
| Mar | 0.025 | 0.008 | 0.006 | 0.010 | 0.004 | 0.010 | 0.015 | 0.010 | 0.008 | 0.015 | 0.004 | 0.029 |
| Apr | 0.014 | 0.042 | 0.010 | 0.012 | 0.014 | 0.008 | 0.010 | 0.055 | 0.606 | 0.018 | 0.008 | 0.008 |
| May | 0.004 | 0.016 | 0.027 | 0.011 | 0.011 | 0.016 | 0.004 | 0.013 | 0.011 | 0.011 | 0.038 | 0.018 |
| Jun | 0.016 | 0.004 | 0.068 | 0.004 | 0.010 | 0.006 | 0.026 | 0.028 | 0.004 | 0.006 | 0.010 | 0.006 |
| Jul | 0.006 | 0.029 | 0.010 | 0.012 | 0.033 | 0.021 | 0.012 | 0.016 | 0.004 | 0.008 | 0.008 | 0.698 |
| Aug | 0.011 | 0.665 | 0.017 | 0.020 | 0.020 | 0.009 | 0.022 | 0.007 | 0.009 | 0.013 | 0.033 | 0.009 |
| Sep | 0.019 | 0.006 | 0.013 | 0.011 | 0.006 | 0.028 | 0.006 | 0.021 | 0.011 | 0.011 | 0.718 | 0.006 |
| Oct | 0.016 | 0.015 | 0.011 | 0.015 | 0.011 | 0.027 | 0.005 | 0.005 | 0.015 | 0.013 | 0.005 | 0.007 |
| Nov | 0.016 | 0.007 | 0.016 | 0.010 | 0.017 | 0.003 | 0.005 | 0.010 | 0.009 | 0.003 | 0.009 | 0.007 |
| Dec | 0.009 | 0.004 | 0.020 | 0.653 | 0.013 | 0.013 | 0.013 | 0.009 | 0.004 | 0.028 | 0.015 | 0.009 |

TABLE VI: DOCUMENT TOPIC PROBABILITY VALUES OF MKB 2018

| Month | Topic1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 | Topic 10 | Topic 11 | Topic 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan | 0.006 | 0.017 | 0.021 | 0.014 | 0.012 | 0.006 | 0.006 | 0.004 | 0.021 | 0.008 | 0.021 | 0.008 |
| Feb | 0.006 | 0.011 | 0.008 | 0.011 | 0.006 | 0.017 | 0.025 | 0.004 | 0.028 | 0.01 | 0.687 | 0.023 |
| Mar | 0.013 | 0.008 | 0.037 | 0.008 | 0.011 | 0.011 | 0.617 | 0.028 | 0.011 | 0.021 | 0.013 | 0.015 |
| Apr | 0.027 | 0.02 | 0.006 | 0.006 | 0.003 | 0.016 | 0.014 | 0.011 | 0.01 | 0.022 | 0.008 | 0.019 |
| May | 0.604 | 0.028 | 0.006 | 0.004 | 0.013 | 0.013 | 0.015 | 0.004 | 0.028 | 0.017 | 0.011 | 0.023 |
| Jun | 0.014 | 0.041 | 0.012 | 0.009 | 0.005 | 0.009 | 0.01 | 0.007 | 0.031 | 0.012 | 0.01 | 0.009 |
| Jul | 0.026 | 0.016 | 0.007 | 0.026 | 0.643 | 0.007 | 0.005 | 0.005 | 0.008 | 0.011 | 0.015 | 0.016 |
| Aug | 0.009 | 0.008 | 0.019 | 0.006 | 0.006 | 0.748 | 0.008 | 0.006 | 0.011 | 0.019 | 0.019 | 0.009 |
| Sep | 0.007 | 0.014 | 0.66 | 0.016 | 0.019 | 0.009 | 0.035 | 0.021 | 0.012 | 0.033 | 0.005 | 0.009 |
| Oct | 0.007 | 0.011 | 0.031 | 0.031 | 0.006 | 0.013 | 0.009 | 0.02 | 0.031 | 0.009 | 0.007 | 0.006 |
| Nov | 0.017 | 0.017 | 0.009 | 0.015 | 0.006 | 0.006 | 0.011 | 0.004 | 0.031 | 0.646 | 0.019 | 0.017 |
| Dec | 0.005 | 0.009 | 0.018 | 0.005 | 0.023 | 0.007 | 0.007 | 0.706 | 0.014 | 0.018 | 0.005 | 0.018 |

| Month | Topic 13 | Topic 14 | Topic 15 | Topic 16 | Topic 17 | Topic 18 | Topic 19 | Topic 20 | Topic 21 | Topic 22 | Topic 23 | Topic 24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan | 0.019 | 0.054 | 0.012 | 0.015 | 0.025 | 0.012 | 0.006 | 0.023 | 0.008 | 0.669 | 0.006 | 0.008 |
| Feb | 0.006 | 0.004 | 0.019 | 0.026 | 0.015 | 0.023 | 0.008 | 0.011 | 0.015 | 0.019 | 0.011 | 0.008 |
| Mar | 0.004 | 0.008 | 0.019 | 0.013 | 0.009 | 0.028 | 0.006 | 0.011 | 0.034 | 0.004 | 0.065 | 0.006 |
| Apr | 0.005 | 0.013 | 0.005 | 0.025 | 0.01 | 0.005 | 0.008 | 0.02 | 0.017 | 0.01 | 0.706 | 0.014 |
| May | 0.025 | 0.055 | 0.011 | 0.025 | 0.013 | 0.004 | 0.01 | 0.013 | 0.032 | 0.008 | 0.019 | 0.019 |
| Jun | 0.01 | 0.015 | 0.004 | 0.004 | 0.007 | 0.019 | 0.721 | 0.009 | 0.009 | 0.012 | 0.014 | 0.01 |
| Jul | 0.028 | 0.01 | 0.026 | 0.015 | 0.021 | 0.008 | 0.005 | 0.037 | 0.01 | 0.003 | 0.016 | 0.036 |
| Aug | 0.031 | 0.006 | 0.004 | 0.011 | 0.011 | 0.02 | 0.013 | 0.004 | 0.013 | 0.004 | 0.008 | 0.009 |
| Sep | 0.016 | 0.014 | 0.007 | 0.012 | 0.019 | 0.016 | 0.005 | 0.014 | 0.012 | 0.014 | 0.012 | 0.019 |
| Oct | 0.006 | 0.011 | 0.02 | 0.006 | 0.007 | 0.013 | 0.004 | 0.011 | 0.004 | 0.013 | 0.022 | 0.705 |
| Nov | 0.017 | 0.013 | 0.009 | 0.011 | 0.011 | 0.019 | 0.009 | 0.036 | 0.04 | 0.017 | 0.013 | 0.011 |
| Dec | 0.004 | 0.011 | 0.018 | 0.016 | 0.013 | 0.004 | 0.014 | 0.009 | 0.029 | 0.009 | 0.013 | 0.025 |

TABLE VII:DOCUMENT TOPIC PROBABILITY VALUES OF MKB 2019

| Month | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 | Topic 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Jan | 0.015 | 0.015 | 0.011 | 0.014 | 0.022 | 0.01 | 0.031 | 0.015 | 0.024 | 0.625 |
| feb | 0.036 | 0.016 | 0.007 | 0.014 | 0.011 | 0.022 | 0.017 | 0.01 | 0.005 | 0.014 |
| Jun | 0.01 | 0.008 | 0.012 | 0.003 | 0.013 | 0.036 | 0.022 | 0.024 | 0.006 | 0.009 |
| Jul | 0.022 | 0.032 | 0.021 | 0.027 | 0.02 | 0.022 | 0.017 | 0.033 | 0.005 | 0.008 |
| Aug | 0.619 | 0.037 | 0.002 | 0.016 | 0.014 | 0.008 | 0.021 | 0.048 | 0.017 | 0.013 |
| Sep | 0.012 | 0.018 | 0.008 | 0.025 | 0.051 | 0.023 | 0.024 | 0.018 | 0.014 | 0.008 |
| Oct | 0.011 | 0.044 | 0.035 | 0.012 | 0.581 | 0.062 | 0.028 | 0.012 | 0.016 | 0.006 |
| Nov | 0.008 | 0.014 | 0.008 | 0.007 | 0.009 | 0.009 | 0.029 | 0.009 | 0.737 | 0.003 |
| Dec | 0.021 | 0.023 | 0.611 | 0.033 | 0.021 | 0.017 | 0.064 | 0.013 | 0.005 | 0.017 |

| Month | Topic 11 | Topic 12 | Topic 13 | Topic 14 | Topic 15 | Topic 16 | Topic 17 | Topic 18 | Topic 19 | Topic 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Jan | 0.004 | 0.04 | 0.026 | 0.023 | 0.009 | 0.063 | 0.014 | 0.017 | 0.017 | 0.01 |
| feb | 0.031 | 0.054 | 0.006 | 0.036 | 0.07 | 0.559 | 0.021 | 0.041 | 0.016 | 0.014 |
| Jun | 0.026 | 0.578 | 0.03 | 0.024 | 0.029 | 0.027 | 0.051 | 0.019 | 0.047 | 0.02 |
| Jul | 0.022 | 0.07 | 0.52 | 0.012 | 0.024 | 0.014 | 0.033 | 0.005 | 0.015 | 0.079 |
| Aug | 0.003 | 0.056 | 0.014 | 0.031 | 0.036 | 0.004 | 0.038 | 0.008 | 0.003 | 0.003 |
| Sep | 0.629 | 0.012 | 0.018 | 0.03 | 0.026 | 0.003 | 0.032 | 0.012 | 0.011 | 0.024 |
| Oct | 0.018 | 0.038 | 0.015 | 0.005 | 0.021 | 0.011 | 0.015 | 0.046 | 0.02 | 0.006 |
| Nov | 0.014 | 0.011 | 0.019 | 0.051 | 0.017 | 0.015 | 0.012 | 0.002 | 0.007 | 0.018 |
| Dec | 0.013 | 0.021 | 0.018 | 0.018 | 0.003 | 0.011 | 0.026 | 0.019 | 0.037 | 0.009 |

Plotting all these Document topic probability values in a graph of each year separately we can determine the topic that is most probably spoken in each month in each year by the top node in the graph that usually helps in determining the hidden topic. Document probability graph of each year are plotted, and respective figures are illustrated in below Figure 6.

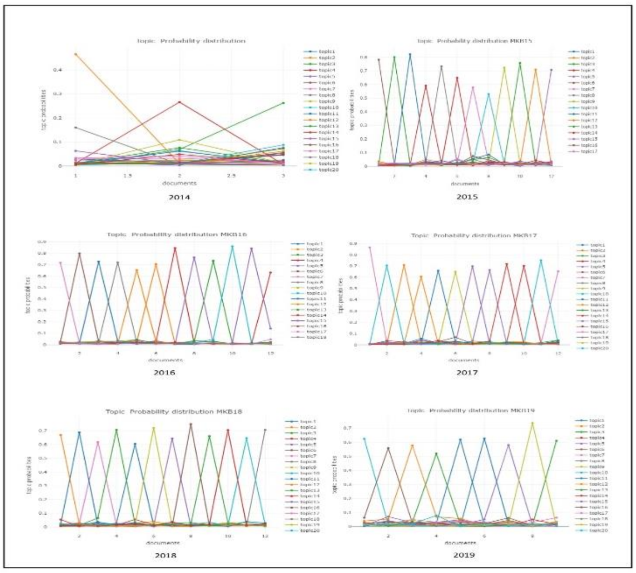


Fig. 6.Document topic Probability Graph of Mann Ki Baat from 2014 to 2019

From the figure 6 the topic probabilities are changing from 2014,2015,2016 while 2017,2018 and 2019 it look that the topic coverage seem to be almost same .

Hence, from the Document probability graph of each year from 2014 to 2019, we can observe that the top nodes represent the topic spoken in that particular month with high probability. The following tables illustrate the topics spoken with their respective probabilities (H-Highest Probability, L-Least Probability).

TABLE VIII: TOPICS WITH PROBABILITY VALUES OF MKB IN 2014

| 2014 | |
|---|---|
| **Month** | **(Topic, Probability)** |
| October | (Topic 2,0.46505)-H |
| November | (Topic 14,0.26615)-L |
| December | (Topic 3,0.26253)-L |

TABLE IX: TOPICS WITH PROBABILITY VALUES OF MKB FROM 2015 TO 2017.

| 2015 | | 2016 | | 2017 | |
|---|---|---|---|---|---|
| **Month** | **(Topic, Probability)** | **Month** | **(Topic, Probability)** | **Month** | **(Topic, Probability)** |
| Jan | (Topic 16,0.7819) | Jan | (Topic 17,0.7970) | Jan | (Topic 7,0.8653)-H |
| Feb | (Topic 3,0.8006) | Feb | (Topic 16,0.7990) | Feb | (Topic 20,0.7050) |
| Mar | (Topic 11,0.8212)-H | Mar | (Topic 11,0.7289) | Mar | (Topic 12,0.7091) |
| Apr | (Topic 14,0.5915) | Apr | (Topic 8,0.7200) | Apr | (Topic 22,0.6059)-L |
| May | (Topic 8,0.7338) | May | (Topic 12,0.6540) | May | (Topic 1,0.6585) |
| Jun | (Topic 4,0.6493) | Jun | (Topic 2,0.7050) | Jun | (Topic 9,0.6501) |
| Jul | (Topic 7,0.5788) | Jul | (Topic 4,0.8465) | Jul | (Topic 25,0.6979) |
| Aug | (Topic 10,0.5302)-L | Aug | (Topic 5,0.7640) | Aug | (Topic 15,0.6649) |
| Sep | (Topic 9,0.7240) | Sep | (Topic 3,0.7340) | Sep | (Topic 24,0.7179) |
| Oct | (Topic 13,0.7580) | Oct | (Topic 10,0.8613)-H | Oct | (Topic 4,0.7015) |
| Nov | (Topic 12,0.7094) | Nov | (Topic 15,0.8420) | Nov | (Topic 10,0.7513) |
| Dec | (Topic 5,0.7081) | Dec | (Topic 14,0.6320)-L | Dec | (Topic 17,0.6529) |

TABLE X: TOPICS WITH PROBABILITY VALUES OF MKB IN 2018 AND 2019

| 2018 | | 2019 | |
|---|---|---|---|
| **Month** | **(Topic, Probability)** | **Month** | **(Topic, Probability)** |
| Jan | (Topic 22,0.6687) | Jan | (Topic 10,0.6250) |
| Feb | (Topic 11,0.6874) | Feb | (Topic 16,0.5580) |
| Mar | (Topic 7,0.6172) | Jun | (Topic 12,0.5781) |
| Apr | (Topic 23,0.7059) | Jul | (Topic 13,0.5201)-L |
| May | (Topic 1,0.6043)-L | Aug | (Topic 1,0.6192) |
| Jun | (Topic 19,0.7214) | Sep | (Topic 11,0.6288) |
| Jul | (Topic 5,0.6425) | Oct | (Topic 5,0.5801) |
| Aug | (Topic 6,0.7478)-H | Nov | (Topic 9,0.7370)-H |
| Sep | (Topic 3,0.6599) | Dec | (Topic 3,0.6111) |
| Oct | (Topic 24,0.7040) | | |
| Nov | (Topic 10,0.6460) | | |
| Dec | (Topic 8,0.7058) | | |

Using the above tables, we extracted top terms that related to each topic as below:

TABLE XI: TOPIC AND TERMS OF MANN KI BAAT 2014

| **October** | **November** | **December** |
|---|---|---|
| ***Topic 2*** | ***Topic 14*** | ***Topic 3*** |
| strength | chang | drug |
| khaadi | poor | concern |
| sheep | commit | addict |
| poor | experi | prime |
| brother | follow | sometim |
| crore | money | topic |
| vijay | assur | joy |
| cub | month | baat |
| occas | special | interact |
| forward | baat | mention |
| help | public | apt |
| lion | call | discuss |
| nine | festiv | parent |
| product | discuss | wrong |
| sit | facil | blame |
| trust | litter | blind |
| carri | mann | cosponsor |
| dashami | sent | terrorist |
| destin | compar | enjoy |
| flock | congratul | live |

TABLE XII: TOPIC AND TOP 20 TERMS OF MANN KI BAAT 2015

| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Topic 16* | *Topic 3* | *Topic 11* | *Topic 14* | *Topic 8* | *Topic 4* | *Topic 7* | *Topic 10* | *Topic 9* | *Topic 13* | *Topic 12* | *Topic 5* |
| question | exam | farmer | india | friend | yoga | villag | bank | baat | india | organ | scheme |
| barack | friend | villag | ambedkar | channel | water | road | maharashtra | india | organ | stump | india |
| india | success | law | strength | yoga | india | accid | job | mann | gold | soil | januari |
| unit | challeng | road | nepal | failur | rain | safeti | free | kid | villag | mudra | duti |
| shri | faith | water | confid | success | pictur | soldier | invit | cleanli | donat | crop | villag |
| hon'bl | competit | speak | daughter | act | scheme | august | exploit | khadi | cleanli | scheme | effort |
| obama | topic | employ | forc | anim | launch | railway | level | octob | clean | lakh | account |
| presid | hour | farm | aliv | travel | villag | offici | ordin | radio | diwali | light | friend |
| daughter | resolut | compens | amidst | farmer | haryana | immens | memori | hour | effort | burn | term |
| modi | subject | acquir | babasaheb | choos | season | death | attent | lakh | program | energi | clean |
| health | confid | acquisit | earthquak | india | monsoon | hour | bandhan | elect | region | worker | organ |
| look | desir | consent | memori | care | daughter | toilet | raksha | messag | african | enterpris | birth |
| white | oneself | previous | soldier | support | selfi | job | dengu | baat | africa | bharat | stori |
| narendra | result | proper | war | measur | bachao | oper | scientist | mann | race | donat | labour |
| sri | test | assur | support | armi | beti | war | jawan | listen | gram | fertil | tourist |
| american | pass | better | centenari | heat | movement | launch | span | daughter | mantra | asha | constitut |
| job | height | project | pride | cultur | padhao | northeast | account | look | scheme | climat | beneficiari |
| milk | mark | requir | crisi | pass | beti | vijay | erect | movement | uniti | decemb | right |
| yeswecan | nervous | effort | defeat | polit | photo | like | interview | resid | letter | exercis | servic |
| like | paper | question | damag | speak | august | servic | reform | immens | sri | lantern | bank |

TABLE XIII: TOPIC AND TOP 20 TERMS OF MANN KI BAAT 2016

| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Topic 17* | *Topic 16* | *Topic 11* | *Topic 8* | *Topic 12* | *Topic 2* | *Topic 4* | *Topic 5* | *Topic 3* | *Topic 10* | *Topic 15* | *Topic 14* |
| khadi | exam | water | water | water | yoga | tree | teacher | countrymen | festiv | money | countrymen |
| startup | don | holiday | educ | yog | countrymen | countrymen | ganesh | medal | sardar | bank | rupe |
| organ | scienc | cup | ganga | don | democraci | innov | countrymen | octob | diwali | note | reward |
| railway | sleep | fifa | subsidi | sportsperson | satellit | rupe | wrote | athlet | deepawali | busi | digit |
| station | yoga | diabet | qualiti | money | incom | plant | olymp | divyang | dark | payment | black |
| insur | target | footbal | gas | mark | june | rio | ganga | paralymp | uniti | rupe | corrupt |
| haryana | paper | summer | organ | environ | septemb | sapl | toilet | construct | jawan | card | parti |
| januari | scientist | tourism | april | conserv | kulkarni | doctor | chand | gandhi | countrymen | countrymen | law |
| bapu | answer | bird | panchayati | june | tax | antibiot | pradhan | mahatma | defec | soldier | divyaang |
| sea | teacher | don | polit | irrig | chandrak | festiv | festiv | depart | armi | wage | player |
| beema | inner | tourist | raj | bank | pension | station | medal | won | saheb | rupay | sector |
| charkha | discoveri | host | mela | music | intern | sportsperson | idol | kashmir | toilet | worker | hockey |
| eight | examin | abhi | kumbh | drop | scientist | research | player | toilet | dev | pradhan | fertil |
| fasal | shri | april | don' | flow | tax | rain | public | uniti | haryana | tax | cashless |
| host | tomorrow | conserv | industri | cultiv | diabet | africa | channel | centenari | construct | black | transact |
| mantri | calm | mine | cylind | drought | date | incub | shri | dayal | soldier | cashless | rumour |
| market | gain | travel | pond | card | undisclos | pledg | kashmir | deen | guard | currenc | busi |
| saarc | failur | treatment | trust | forest | shri | railway | demand | anger | guru | kashmir | payment |
| statu | pressur | coal | laid | gaurav | fli | medicin | score | pandit | border | hardship | fight |
| beti | innov | diseas | teacher | balanc | rainwat | abdul | hockey | upadhyay | freedom | society' | categori |

TABLE XIV: TOPIC AND TOP 20 TERMS OF MANN KI BAAT 2017

| **Jan** | **Feb** | **Mar** | **Apr** | **May** | **Jun** | **Jul** | **Aug** | **Sep** | **Oct** | **Nov** | **Dec** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Topic 7*** | ***Topic 20*** | ***Topic 12*** | ***Topic 22*** | ***Topic 1*** | ***Topic 9*** | ***Topic 25*** | ***Topic 15*** | ***Topic 24*** | ***Topic 4*** | ***Topic 10*** | ***Topic 17*** |
| exam | farmer | resolv | vacat | yoga | yoga | resolv | mantri | khadi | khadi | soil | januari |
| mark | scienc | depress | travel | freedom | democraci | august | pradhan | baat | peacekeep | navi | resolv |
| parent | pit | bhagat | class | june | lord | gst | ganesh | mann | nivedita | constitut | survey |
| question | prize | champaran | summer | bin | jagannath | freedom | teacher | seven | secur | farmer | build |
| examin | technolog | freedom | ramanujacharya | jail | book | organis | yojna | octob | diwali | resolv | centuri |
| relax | toilet | satyagraha | technolog | modi | ramzan | quit | stand | tourism | guru | arm | devote |
| coast | isro | april | beacon | book | regist | attain | octob | travel | oper | terror | enthusiasm |
| test | satellit | digit | comfort | drive | handkerchief | ganesh | forgiv | incid | dedic | decemb | name |
| guard | ministri | british | worker | read | read | struggl | jan | tourist | diseas | ship | baat |
| knowledg | space | singh | twenti | block | rain | ‘quit | name | presid | yoga | earth | singh |
| sleep | brother | struggl | babasaheb | litter | bouquet | speech | resolv | materi | sardar | fertil | haj |
| answer | digit | bangladesh | obtain | garbag | egem | decis | money | bilal | handloom | sea | mann |
| minut | puls | gandhi’ | satellit | type | expect | movement’ | sea | deen | nanak | secur | drive |
| unfair | crop | rajguru | offer | struggl | depart | flood | jayanti | srinagar | mega | brother | guru |
| appear | flower | sukhdev | advic | afroz | money | billion | communiti | resolv | dev | divyang | muslim |
| burden | product | yoga | budha | beach | satellit | rakhi | stop | ambassador | appear | humanitarian | voter |
| amongst | yojana’ | river | sabka | cellular | space | tax | symbol | bharat | carv | product | garbag |
| bodi | vasant | decis | saint | hon’bl | toilet | personnel | veget | destin | won | flag | popul |
| breath | baba | examin | vip | liquid | june | realis | dhanyojna | throughout | ideal | maratha | asean |
| januari | demonstr | baba | book | ordinari | dark | economi | law | sardar | sahab | naval | christma |

TABLE XV: TOPIC AND TOP 20 TERMS OF MANN KI BAAT 2018

| **Jan** | **Feb** | **Mar** | **Apr** | **May** | **Jun** | **Jul** | **Aug** | **Sep** | **Oct** | **Nov** | **Dec** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Topic 22*** | ***Topic 11*** | ***Topic 7*** | ***Topic 23*** | ***Topic 1*** | ***Topic 19*** | ***Topic 5*** | ***Topic 6*** | ***Topic 3*** | ***Topic 24*** | ***Topic 10*** | ***Topic 8*** |
| medicin | safeti | yoga | water | game | yoga | sardar | engin | forc | singh | radio | kumbh |
| woman | disast | cost | buddha | savarkar | doctor | lok | atalji | right | patel | constitut | singh |
| avail | wast | ambedkar | lord | everest | gst | manya | medal | soldier | communiti | idea | rajani |
| morna | scienc | saheb | conserv | prakash | prasad | statu | kerala | sanit | game | polit | mela |
| clean | intellig | baba | medal | june | shyama | tilak | sanskrit | mantra | tribal | question | guru |
| padma | artifici | ministri | fit | team | industri | azad | pass | freedom | sardar | comment | gobind |
| femal | dung | anim | sportsperson | fit | guru | father | session | octob | hockey | baba | bapu |
| jan | elephanta | prevent | rabindra | yoga | mukherje | ahmedabad | player | bapu | team | saheb | calendar |
| lot | jharkhand | avail | video | rao | tax | cave | polit | purchas | medal | assembl | mahakumbh |
| tribal | light | industri | biscuit | exercis | match | ekta | parliament | labour | gold | dev | endeavour |
| acknowledg | march | medic | prophet | atal | singl | ganesh | sabha | gandhiji | player | expect | fight |
| administr | question | care | april | slum | afghanistan | neeraj | disast | commiss | tiger | tea | lot |
| empower | rural | extens | athlet | adventur | cast | pandharpur | sentenc | nhrc | cup | guru | south |
| fighter | garbag | paid | baodi | eid | kabir | patel | monsoon | fight | tree | communic | patient |
| hospit | idea | afford | buddhist | tea | profession | fight | bill | organis | chanc | constitu | box |
| origin | alert | water | purnima | veer | das | juli | august | bapu’ | match | travel | devote |
| station | bio | backward | templ | board | maghar | august | guru | convent | para | channel | medal |
| afford | cattl | feder | lot | neighbourhood | nanak | saint | lok | offer | solut | nanak | renew |
| akola | colour | jayanti | rupe | rain | histor | tilakji | guilti | cloth | octob | teenag | solar |
| aushadhi | convert | poverti | donat | cloth | saint | centr | rape | jawan | statu | right | mandela |

TABLE XVI: TOPIC AND TOP 20 TERMS OF MANN KI BAAT 2019

| Jan | Feb | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|
| ***Topic 10*** | ***Topic 16*** | ***Topic 12*** | ***Topic 13*** | ***Topic 1*** | ***Topic 11*** | ***Topic 5*** | ***Topic 9*** | ***Topic 3*** |
| netaji | soldier | water | particip | gandhi | ecigarett | guru | sir | eclips |
| museum | martyr | yoga | water | tiger | sister | unity | minist | product |
| program | month | stori | baat | bear | parent | sardar | yes | himayat |
| ravida | institut | elect | chandrayaan | mahatma | benefit | diwali | ncc | decad |
| elect | son | mann | mann | food | societi | dev | river | train |
| space | brave | read | book | serv | cigarett | patel | languag | women |
| subhash | sacrific | democraci | month | plastic | laxmi | nanak | khan | solar |
| vote | secur | conserv | offici | forest | daughter | home | fit | star |
| paint | student | express | panchayat | octob | exam | octob | tarannum | meet |
| shri | award | lakh | space | mohan | letter | tourist | akhil | programme |
| voter | tata | relat | read | societi | book | holi | hari | local |
| commiss | mann | societi | student | art | whatev | local | hind | sky |
| radio | soldiers | journey | conserv | bose | ripudaman | run | cadet | studi |
| sant | women | vote | season | human | home | uniti | vinol | alumni |
| swamiji | british | meet | wait | perform | read | lakshadweep | jai | presid |
| bose | examin | movement | people | inner | singl | lakshmi | camp | astronomi |
| democraci | scheme | women | tournament | environ | continu | rever | exam | till |
| januari | tree | emot | yatra | septemb | defeat | run | student | centr |
| particip | valour | subject | avail | fit | look | statu | examin | jammu |
| student | courag | bless | perform | associ | tourism | attract | program | kashmir |

Tables 11 to 16 clearly shows that the talk in the radio programs are different from 2014 to 2019 with a variety of generalised notions of his speeches and not any of the fundamental problems being discussed.

## V. CONCLUSION AND FUTURE WORK

Topic modelling with Latent Dirichlet Allocation plays a significant role in text mining to explore large set of documents and extract insights from the text data. The principal commitment of our study is to explore documents containing more than one topic with the designed framework. From this analysis, it is also evident that the designed framework Topic modelling with Latent Dirichlet Allocation greatly helped in deriving topics from corpus. The results also demonstrate that honorable PM in his radio program Mann Ki Baat addressed the nation covering the generalised notions of various topics of the country like achievements and works of historical leaders and freedom fighters who have nourished the country with their commitment and love. Also, various important initiatives like Fit India, Hum Fit Toh India Fit insisting the importance of yoga, wellness awareness for enhancing the health and quality of life, exam warriors for living a stress-free and healthy life are addressed and themes such as social life, public life, lifestyle, cleanliness, environmental conversation had been addressed which helped to spread positivity among the people and left an impact on them. But when asked about the various issues that could be covered in 'Mann Ki Baat' program. Our analysis shows that there could be more focus on issues such as Employment opportunities for the youth, Economy and details about the GDP, Energy saving Initiatives, Irrigation schemes, Jan Dan Yojana scheme details, and development Schemes. But overall analysis of program shows that Mann Ki Baat has changed societal discourse, brought a new communication revolution and a Positive change in such a way that now even a common man can express his thoughts and views directly with the Honorable Prime Minister of India and also brought a revolution in the way government connects with people. In the future, we would like to apply this technique to more challenging textual data. Also we will try to analyze the semantic pattern structure and discover the association between the words that represent topics at a granular level.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the financial and fellowship support granted by the Department of Science and Technology (DST-INSPIRE (Innovation in Science and Pursuit for Inspired Research)), New Delhi, India.For carrying out the study. The authors also wish to thank Dr.S.A.JyothiRani from Department of Statistics, Osmania University and Mokalla Thirupathi Reddy from National Institute of Nutrition (NIN) for encouragement and support.

## REFERENCES

[1] Kumar A, Dabas V, Hooda P, Text classification algorithms for mining unstructured data: a SWOT analysis. *Int. J. Inf. Technol*. 1–11 (2018). https://doi.org/10.1007/s41870-017-0072-1.

[2] Tong Z., Zhang H.(2016). A Text Mining Research Based on LDA Topic Modelling.

[3] Gentzkow M, Kelly B, Taddy M(2019). Text as data. *J. Econ. Lit*.

[4] Mazarura J, Waal A De, Kanfer F, Millard S. Topic Modelling for Short Text. PrasaOrg 2014.

[5] Garg K. Sentiment analysis of Indian PM's "Mann Ki Baat." *Int J Inf Technol [Internet]*. Springer Science and Business Media LLC; 2020 [cited 2020 Feb 24];12:37–48. Available from: http://link.springer.com/10.1007/s41870-019-00324-8.

[6] Upadhyay S, Upadhyay N. Investigating Prime Minister Narendra Modi's Usage of Pathos in the Cyber-Physical Society – A Case of Public Relations Campaign. *Procedia Comput Sci*. 2019.

[7] Hofmann, T.: Probabilistic latent semantic indexing. In: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR 1999 (1999).

[8] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. *J. Mach. Learn. Res.* (2003). https://doi.org/10.1016/b978-0-12-411519-4.00006-9.

[9] Dredze, M., Wallach, H.M., Puller, D., Pereira, F.: Generating summary keywords for emails using topics. In: *International Conference on Intelligent User Interfaces, Proceedings IUI* (2008).

[10] Griffiths, T.L., Steyvers, M.: Finding scientific topics. *Proc. Natl. Acad. Sci. U. S. A.* (2004). https://doi.org/10.1073/pnas.0307752101.

[11] Lau, J.H., Newman, D., Karimi, S., Baldwin, T.: Latent Dirichlet allocation. In: Coling 2010 - 23rd *International Conference on Computational Linguistics, Proceedings of the Conference* (2010).

[12] Blei, D.M., Jordan, M.I.: Modeling Annotated Data. In: *SIGIR Forum (ACM Special Interest Group on Information Retrieval)* (2003).

[13] Rosen-Zvi, M., Chemudugunta, C., Griffiths, T., Smyth, P., Steyvers, M.: Learning author-topic models from text corpora. *ACM Trans. Inf. Syst*. (2010). https://doi.org/10.1145/1658377.1658381.

[14] McCallum, A., Corrada-Emmanuel, A., Wang, X.: Topic and role discovery in social networks. In: *IJCAI International Joint Conference on Artificial Intelligence* (2005).

[15] Griffiths, T.L., Steyvers, M., Blei, D.M., Tenenbaum, J.B.: Integrating topics and syntax. In*: Advances in Neural Information Processing Systems* (2005).

[16] Kulkarni, A., Shivananda, A.: Natural Language Processing Recipes. (2019).

[17] Gupta, M., Gupta, P.: Research and implementation of event extraction from twitter using LDA and scoring function. *Int. J. Inf. Technol*. 11, 365–371 (2019). https://doi.org/10.1007/s41870-018-0206-0.

[18] Anupriya, P., Karpagavalli, S.: LDA based topic modeling of journal abstracts. In: ICACCS 2015 - *Proceedings of the 2nd International Conference on Advanced Computing and Communication Systems* (2015).