



COVID-19 Pandemic: Insights of Newspaper Trends

Jasdeep Kaur, Guru Nanak Dev University, India*


Amit Chhabra, Guru Nanak Dev University, India

 <https://orcid.org/0000-0003-2056-6231>

Munish Saini, Guru Nanak Dev University, India

 <https://orcid.org/0000-0003-4129-2591>

Nebojsa Bacanin, Singidunum University, Serbia

 <https://orcid.org/0000-0002-2062-924X>

ABSTRACT

The study aims to analyze the change in coverage of health issues awareness, printed on the front page of Indian e-papers (The Hindustan Times and The Times of India) for the pre- and peri-coronavirus period. The collected news articles are examined by performing the latent dirichlet allocation algorithm. The sentiment analysis is performed to analyze the change in the emotions aroused from news articles. The outcome regarding the pre-coronavirus period reveals that the focus of the e-papers was mostly on politics, crime, and economy whereas, in the peri-coronavirus period, the e-papers are focusing more (i.e. 40% topics) on publishing the news related to disseminating the awareness about the coronavirus disease. The priority of news topics includes the active number of cases, medical facilities, and COVID-19 testing. The outcome regarding sentiment analysis reveals that negative sentiments are prominent in the peri-coronavirus period due to fear of the outbreak of the virus.

KEYWORDS

Coronavirus, Health Issues, LDA, Newspapers, Sentiment Analysis, Topic Modeling

1. INTRODUCTION

Indian mass media plays a very important role in shaping our country especially the news media (Ram, 2011). News media is very active in India as there are more than 100000 publishers registered with the Registrar of India and have the second-highest selling newspapers in the world¹. The recent trend now in India is shifting towards the online world as people are showing interest in the e-papers and also on various social media platforms such as Facebook, Twitter, etc. (Sahni and Sharma, 2020). The public health issue is among one of the most prominent topics discussed these days as we know that Indian Health care has a very contradicting landscape (Kasthuri, 2018). On one end people are served with high-tech medical facilities especially in urban areas but the scenario is different in the remote

DOI: 10.4018/JITR.299390

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

areas where people are not able to get full medical facilities as worsened conditions of the health care centers (Kasthuri, 2018). There are mainly four main problems that contribute to the worst-case scenario that can be presented in four A's which are lack of Awareness, Access, Affordability, and Absence (Kasthuri, 2018). Mass media plays a very crucial role in spreading health information among people (Liu and Lo, 2014). The awareness is spread through newspapers by publishing numerous articles on hygiene, immunization, population control, mental health, and various deadly pandemic like the H1N1 pandemic (Kasthuri, 2018).

The world is going through a deadly pandemic known as Coronavirus (Covid-19) which is one of the worst global health crises faced by the human race since World War II (Wang et al., 2020). Coronavirus causes an acute respiratory illness that was first observed in the city of Wuhan, China in December 2019 and gradually it spread across the globe (Kong et al., 2020). The pandemic has been severely affecting India as it has the second most active number of cases². India has a total of more than eleven million active cases and more than 1.5 lakh people have lost their lives³. The whole nation was under lockdown in four stages named Phase 1 (March 25- April 14), Phase 2 (April 15- May 3), Phase 3 (May 4- May 17), and Phase 4 (May 18- May 31). Eventually, in June the government announced to give relief in lockdown to put the country's economy back on track. The stages of giving relief were named Unlock1 (June 1- June 30), Unlock 2 (July 1- July 31), and Unlock 3 (Aug 1- Aug 31)⁴. After the unlock procedure started India experienced a steep rise in the Covid-19 cases. During the lockdown period, people have started using online platforms such as e-papers, social media (Twitter, Facebook, Instagram, etc.) to keep them aware of happenings in India.

The changing scenario around the world due to the pandemic our study enlightens the change in health issues coverage during the pre-and peri- Covid-19 period in India printed on the e-paper's front page. The front page of the newspaper is of utmost importance and carries very crucial news about the important events occurring around the world (Kim and Chung, 2017). So, a total of 6347 news articles are scrapped from the front pages of best-selling newspapers in India namely "The Times of India" and "The Hindustan Times"⁵ using the data scrapper tool from July 2019 to June 2020. From July 2019 to December 2019 is considered a pre-Covid-19 period as there were no coronavirus active cases at that time and January 2020 onwards is considered as peri-coronavirus period as the active cases started emerging in January in India. According to statistics, The Times of India and The Hindustan Times are mostly read English newspapers in India⁶. Further, our study deals with analyzing the change in the sentiments aroused from news articles through sentiment analysis. The sentiment analysis is performed as the comprehension of the message by newspapers affects the attitude of the people (Wyver and Shrum, 2015). Our study findings will help Government officials, researchers, practitioners, and other stakeholders to better understand the prevalence and spreading of Coronavirus disease in India. Moreover, Government officials can initiate the formulation of post-event policies to strengthen the health sector of the country to fight for any future epidemic outbreak.

The paper focuses on the following research objectives:

- Change in health issues coverage printed on the e-paper's front page during the pre-and peri-Covid-19 era using the LDA topic modeling technique.
- Examining the change in the sentiments aroused from news articles during the pre-and peri-Covid-19 era using sentiment analysis.

The rest of the paper is structured as follows. Section 2 deals with the literature review. Section 3 discusses the methodology opted to analyze the newspaper's front page. Section 4 discusses the findings. Further, section 5 and section 6 deals with the discussion and the conclusion respectively.

1.1. Background

Researchers from various fields including medical, computer science, social science, and various others are conducting researches to overcome this deadly virus. In the field of computer science,

various researchers performed the analysis of the behavior of the people towards the pandemic by using various social media platforms (Jahanbin and Rahmanian, 2020). The Covid-19 pandemic has affected the mass media world. There is a tremendous use of technology in the different fields during the pandemic time especially in the education system (Garcia-Peñalvo, 2020; Garcia-Peñalvo et al., 2020, Garcia-Peñalvo et al., 2021). This pandemic has also affected work and employment (Hodder, 2020). It has given rise to digital inequalities (Beaunoyer et al., 2020). Even the mass media was loaded with fake news regarding the pandemic (Apuke and Omar, 2021) Further, numerous researchers have performed an analysis of health communication through newspapers of various countries.

One of the studies states that the role of mass media in covering important issues has always been remarkable (Sharma and Gupta, 2017). It has been concluded in a study that the role of mass media has been crucial in spreading awareness about health education among people. It not only spreads awareness but also educates people from time to time.

Gupta and Sinha (2010) conducted a study about the health-related messages coverage that appears in print and electronic media. The results concluded that emphasis is laid more on a political subject and less stress is laid on health-related news articles. This study uses a manual method to carry out the research. On the contrary, we have conducted research using the LDA topic modeling approach which automatically extracts the frequent topic covered.

Liu et al. (2020), collected media reports on Covid-19, and an investigation was carried out to analyze the role of media in China on the ongoing crisis of Covid-19. The results depicted that mass media news reports in China lagged behind the reporting of Covid-19 as the focus of news reports was on larger society rather than individuals. The main approach use in this study is topic modeling to analyze the situation. In our study, we have inculcated topic modeling to analyze how the awareness of the health problems was dealt with during the pre-and peri- coronavirus period in Indian newspapers. Moreover, sentiment analysis is also performed to analyze the change in the emotions aroused from the news articles.

Further, a researcher carried out a study regarding the COVID-19 and performed a collection of tweets of the two Spanish newspapers to understand how Spanish news media cover public health crises on a social media platform, and analysis was carried out using two main approaches named topic modeling and network analysis (Yu et al., 2020). The main drawback of the study is that it has used tweets which are very short texts as compared to full-length articles (Zang et al., 2017). Our study has overcome this drawback by scrapping the full-length news articles from the e-papers. Barkur and Vibha (2020) conducted a study to check the attitude of the people towards the pandemic. The sentiment analysis of the tweets that originated from India was performed after the lockdown announcement. The main drawback is that it doesn't take into account the behavior of people before the outbreak. This drawback has been overcome by our study by comparing the sentiments of the pre-and peri- coronavirus period.

Also, a researcher performed the LDA topic modeling technique to extract the most discussed topics, and the study includes the sentiment analysis of the tweets (Xue et al., 2020). This study analyzed the work using tweets which is not reliant as many tweets are fake. As for the newspapers, the articles are fully confirmed before publishing. Researchers have opted for the LDA topic modeling approach in enormous fields.

2. MATERIALS AND METHODS

This section described the methodology opted for the discourse analysis of the change in health-issues awareness by the newspapers for the pre-and-peri COVID-19 period which includes data collection, data cleaning, topic modeling under which LDA and sentiment analysis is performed as illuminated in Figure 1.

2.1. Data Collection

For collecting the news articles from E-papers, we have used the automated scraping tool (Data Miner⁷). The automated process outperforms the manual way of copying and pasting the required

Figure 1. Data Analysis and Methodology



content from the web sources, as it takes less time and effort to do the same task. It helps in scraping the data from the web pages (here, we have scraped the news articles from E-Papers) and caters output in CSV file format. In this study, we have collected the 6347 news articles printed on the front page of two prominent and reputed Indian newspapers (*The Hindustan Times* and *The Times of India*) from July 2019 to June 2020. Data Miner uses Xpath, JQuery, and CSS Selector to identify the information on the HTML web page. The data is fetched by creating recipes or by using the already available recipes. For extraction, we have created our recipes. The recipes are created in the following steps:

- By choosing the type of the page List page or detailed page. In this extraction, we have used the List page as we want multiple rows to be extracted.
- Specifying the column names under which the data is to be stored. The data under different columns can be selected using the Find tool. In this case, we have defined two columns first one extracts the headlines and the other column extracts the detailed news from the e-paper.
- The recipe is then saved and run. After which it will automatically extract the news headlines and the detailed news into the respective columns.

The number of articles scrapped from the e-paper's front page for the pre-Covid-19 and peri-Covid-19 period is presented in Table 1.

Figure 2 and Figure 3 presents the number of articles that were printed on the front page of the e-paper during the pre-and- peri Covid-19 period. All segments of the front page are extracted for analysis purpose. The results show that number of articles printed on the front page of the Times of India was more than the Hindustan Times. The number of articles printed on the front page for the peri-Covid-19 was more as compared to the pre-Covid-19 period for both the E-papers.

2.2. Data Preprocessing

The data processing step involves removing the noise from the data. It is one of the most important tasks if not performed properly can lead to errors in the results (Garcia et al., 2015). The preprocessing involves the following main steps⁸:

- **Segmentation:** This process involves breaking down the larger strings into smaller chunks called tokens.

Table 1. Total number of news articles collected for pre-and-peri Covid-19 period

Newspaper	July-December (2019)	January-June (2020)
1. Times of India	1628	2107
2. The Hindustan Times	1300	1312

Figure 2. Month-wise articles for a Pre-Covid-19 period

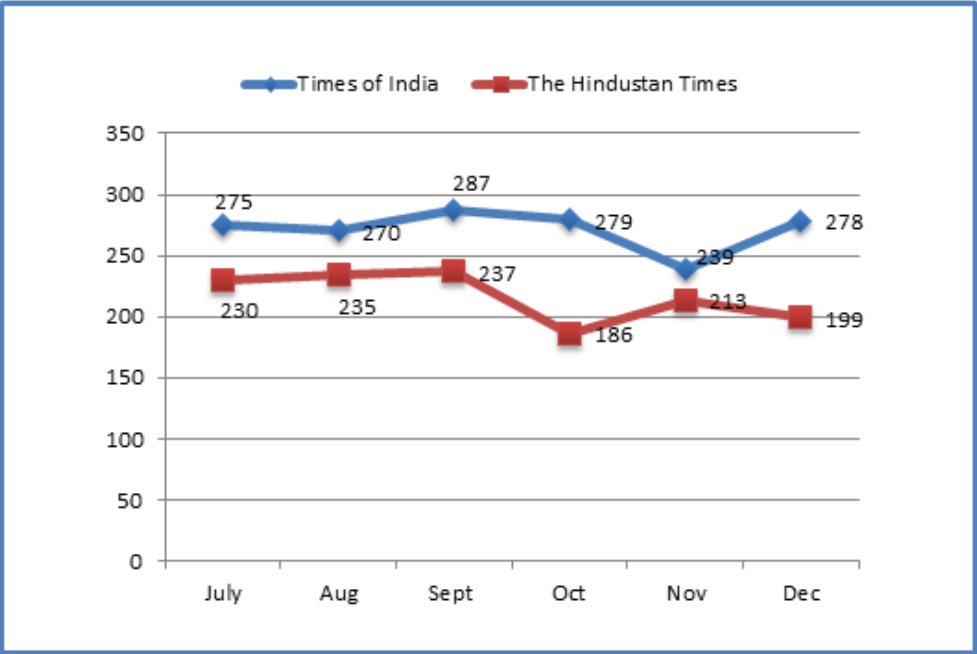
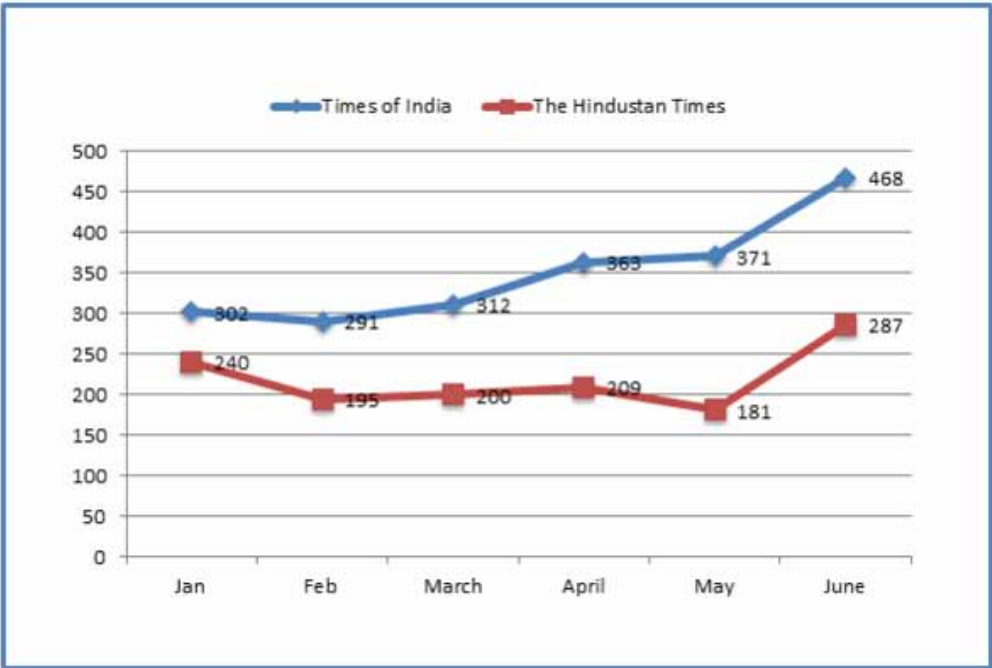


Figure 3. Month-wise articles for a Peri-Covid-19 period



- **Cleaning:** It involves removing stop words and dealing with the capital letters and various other characters.
- **Normalization:** It consists of the mapping of scheme terms or linguistic reductions via stemming, lemmatization.
- **Annotations:** It includes labeling, adding markups, or part-of-speech tagging.
- **Analysis:** It involves a generalization of the dataset for feature analysis which is further used for finding the relationship between words.

In this study, the preprocessing of the news articles is done using RStudio. For performing preprocessing “tm” package is used⁹. Firstly, the corpus is created using the function “Corpus”. After creating the corpus, the “tm_map” function is used for the preprocessing. Using the “tm_map” function the news articles are converted into lower case, punctuations, stopword and numeric values are removed. Further, stemming is performed that reduces the words to unify across the documents. Extra blank spaces which are also known as white spaces are being expunged during pre-processing. A separate CSV file is created after pre-processing for further analysis.

2.3. Topic Modeling

In Natural language processing, Topic modeling is an unsupervised machine learning approach that can scan various documents, identify the words and similarities between them, and automatically cluster together similar words that best describe the documents¹⁰. The clusters are known as the topics which provide the abstract view of the documents. There are four main methods to implement topic modeling which consist of Latent Semantic Analysis (Deerwester, 1990), Probabilistic Latent Semantic Analysis (Hofmann, 1999), Latent Dirichlet Allocation (LDA), and Correlated Topic Model. In this study, we have incorporated the LDA technique as it performs better than the other topic modeling techniques (Chehal et al., 2020).

2.3.1. Latent Dirichlet Allocations (LDA)

Latent Dirichlet allocation (LDA) is considered to be a generative probabilistic model of a corpus (Blei, 2003). It is a three-level hierarchical Bayesian mixture model in which the main goal is to map documents present in the corpus to an appropriate topic that covers plenty of words present in the document. The topics are represented as the mixture of underlying topic probabilities which provides the explicit representation of the documents. The functioning of LDA is defined in Figure 4 (Hidayatullah et al., 2019).

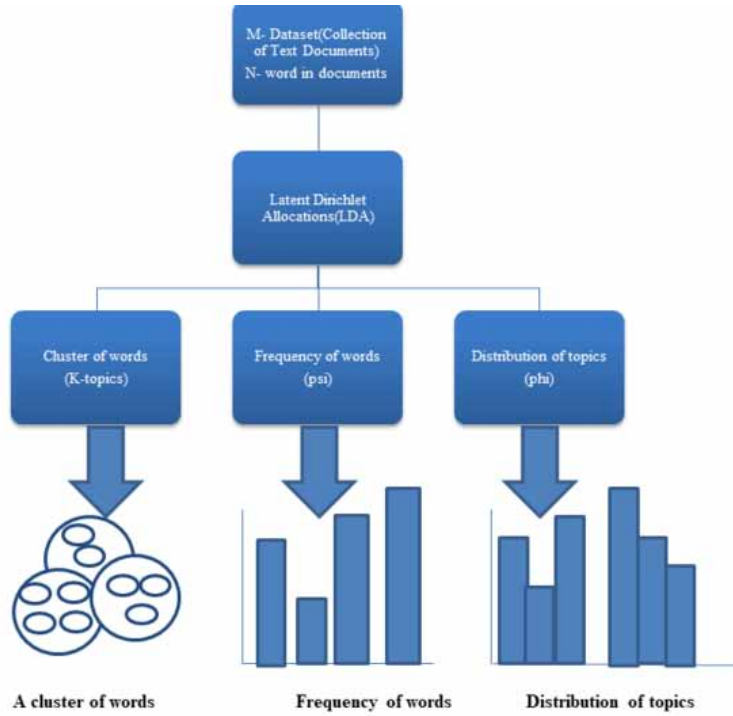
The basic notation used to represent the LDA model is given below:

- T denotes the number of documents in a given corpus.
- N is the number of words in a given document (document i has words).
- α is the parameter of the Dirichlet prior on the per-document topic distributions.
- β is the parameter of the Dirichlet prior on the per-topic word distribution.
- θ_i is the topic distribution for the document i .
- ψ_k is the word distribution for topic k .
- x_{ij} is the topic for the j th word in document i .
- w_{ij} is a specific word.

The generative model of LDA can be achieved in three main stages:

1. Choose $\theta_i \sim \text{Dir}(\alpha)$ where i belongs to $\{1, \dots, T\}$ and $\text{Dir}(\alpha)$ is Dirichlet distribution with symmetric parameter α .
2. Choose $\psi_k \sim \text{Dir}(\beta)$ where k belongs to $\{1, \dots, N\}$.

Figure 4. Functioning of LDA



3. For each word position i, j where i belongs to $\{1, \dots, M\}$ and j belongs to $\{1, \dots, N\}$:
 - a. Choose topic $x_{ij} \sim \text{Multinomial}(\theta_i)$
 - b. Choose a word $w_{ij} \sim \text{Multinomial}(\psi_{z_{ij}})$

One of the researchers has specified that the LDA is an appropriate method to study news media coverage (Daud et al., 2010). Another study has used this method for the identification of 16 frames from European refugee crisis news across five countries (Heidenreich et al., 2019). One of the researchers (Poirier et al., 2020) has performed the LDA to identify six news frames (Chinese outbreak, economic crisis, health crisis, helping Canadians, social impact, Western deterioration) from 12 Canadian media sources. The LDA topic modeling method is applied in this research as it helps in identifying the prominent topics covered in the newspapers.

2.3.2. The Optimal Number of Topics

For the appropriate implication of the LDA algorithm, it is important to identify a meaningful number of topics (Arun et al., 2010). A low value of K (number of topics) can result in few or broad topics, whereas a high value of K results in uninterruptable topics. We have used the topic coherence method to find the optimal number of topics. Topic coherence measures the score of a single topic by measuring the degree of semantic similarity between the high-scoring words in the topic. In this study, the C_v coherence method (Roder et al., 2015) has been used which is considered to have the highest correlation with the human interpretation. Topic coherence is achieved through four stages.

- **Segmentation:** In this case, splitting of the top- n words into pairs. The segmentation is defined as in Eq. 1:

Algorithm 1. Working of the LDA

Input: Training Data W, number of topic K, hyperparameters α, β
Output: topic assignment matrix X, topic-document matrix Y, Topic-word matrix Z

1. Initialize Y, Z to zero.
2. for document $p \in [1, D]$ do
3. for token position s in document p do
4. $X_{ps} = k \sim \text{MUL}(1/K)$
5. $Y_{kp} += 1, Z_{wk} += 1$
6. end for
7. end for
8. repeat
9. for document $p \in [1, D]$ do
10. for token position s in document p do
11. $Y_{kp} = -1, Z_{wk} = -1$
12. $X_{ps} = k' \sim p(X_{ps} = k | \text{rest})$
13. $Y_{k'p} += 1, Z_{wk'} += 1$
14. end for
15. end for
16. Until convergence

$$S = \{(W', W^*) \mid W' = \{w_u\}; w_u \in W; W^* = W\} \quad (1)$$

- **Probability Estimation method:** This method defines how probabilities can be derived from the given data source. In this case, the Boolean sliding window methods in which words are counted using the sliding window. The windows move from documents to documents one word token per step and with each movement, a new virtual document is created of size s .
- **Confirmation Measure:** For every segmented pair confirmation measure (ϕ) is calculated that measures how W^* and W' are correlated with each other that is based on the similarity of W' and W^* concerning all words in W . The similarity is calculated using the vectors $\vec{v}(W')$ and $\vec{v}(W^*)$ that is represented in Eq. 2. Further, normalized pointwise mutual information (NPMI) is calculated to check the agreement between w_i and w_j as exemplified in Eq. 3. Finally, confirmation measure (ϕ) for all segmented pairs (S_i) is calculated via cosine vector similarity of all context vectors as equated in Eq. 4:

$$\vec{v}(W') = \left\{ \sum_{w_i \in W'} NPMI(w_u, w_v)^\gamma \right\}_{v=1..|W|} \quad (2)$$

$$NPMI(w_u, w_v) = \frac{\log \frac{P(w_u, w_v) + \varepsilon}{P(w_u) \cdot P(w_v)}}{-\log(P(w_u, w_v) + \varepsilon)} \quad (3)$$

$$\Phi_{Si}(\vec{u}, \vec{w}) = \frac{\sum_{i=1}^{|W|} u_i, w_i}{\|\vec{u}\|_i \cdot \|\vec{w}\|_i} \quad (4)$$

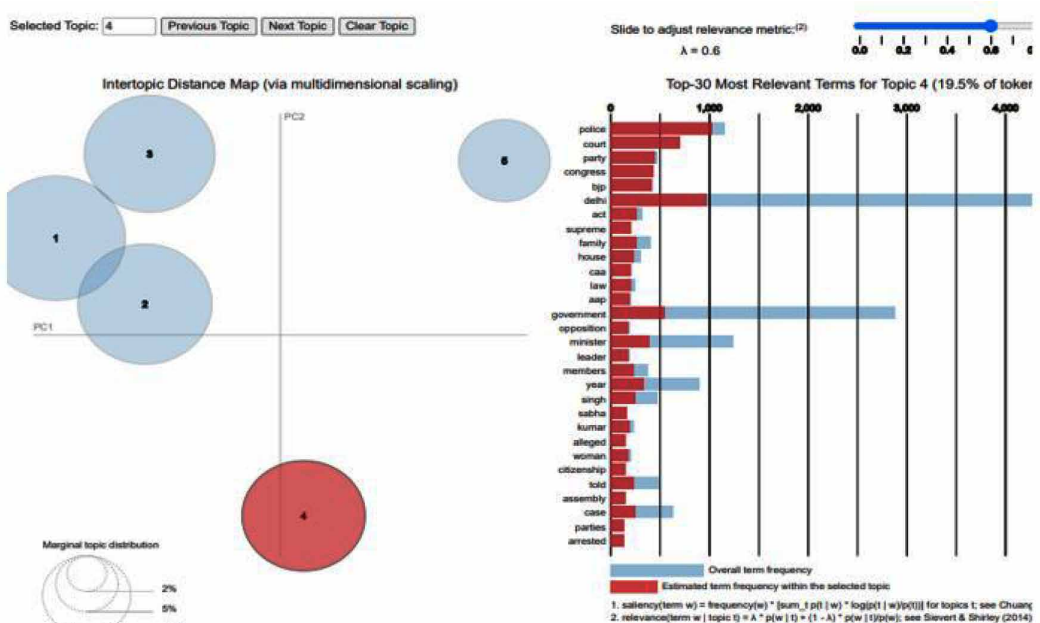
- **Aggregation:** The final coherence score is the arithmetic mean of all confirmation measures Φ .

In this study, the “genismr” library is used under which the ‘c_v’ metric is implemented for calculating the coherence score. The coherence score is observed for the different number of topics. Further, the plot is drawn for coherence score vs a total number of topics to find the optimal number of topics.

2.3.3. Visualization of Topics

The topic visualization is another important aspect of topic modeling to understand the documents. In this study, the visualization of topics is achieved using library LDAvis (Sievert and Shirley, 2014) in the R studio. This library is designed such that it helps in the interpretation of the topics and the information is extracted from the fitted LDA topic model to interactive web-based visualization¹¹ (Fig. 5). The visualization of topics has two important aspects that play an important role in interpreting the topics. The left panel of the visualization represents the global view of the topic. The topics are represented in the form of circles in a 2-D multidimensional plane in which the center is depicted by calculating the distance between the topics as presented in Figure 5. Topic importance is calculated by using the area of the circle. Topics with a larger area are the most prevalent topics. On the other hand, the right panel represents the bar chart in which bars represent the frequencies of the terms that helps in depicting the topics. The red bars depict the topic-specific frequency of the terms whereas the bluish-grey bar represents the corpus-wide frequency of the terms. The useful terms are interpreted using a metric known as relevance which is denoted by λ . Relevance (Sievert and Shirley, 2014) is defined as the degree to which a term appears in the topic in exclusion to the other terms. The “optimal” value of λ was tested for 0.6, and it resulted in an estimated 70% probability of correct identification (Sievert and Shirley, 2014). For values of λ near 0 and 1, the correct value was estimated to be near about 53% and 63% respectively. So in our study, we have taken the value of λ as 0.6.

Figure 5. Sample LDAvis output



2.4. Sentiment Analysis

Sentiment Analysis in the field of Natural Processing Language (NLP) is used to extract the sentiments aroused from the opinions or reviews of a particular matter¹². Opinion mining, sentiment extraction, and subjectivity analysis (Chandni et al., 2015) are the terms used interchangeably for sentiment analysis. Sentiment analysis is classified into seven dimensions which include subjectivity classification, sentiment classification, review usefulness measurement, opinion spam detection, lexicon creation, aspect extraction, and polarity determination (Ravi and Ravi, 2015). Polarity determination deals with finding the polarity of the sentence whether a sentence is positive, negative, or neutral which is carried out in our study and is of the utmost importance as it helps in analyzing people's behavior. Polarity analysis is performed using the “*sentiment*” library in Rstudio¹³. The “*sentimentr*” library provides better results than a simple dictionary lookup approach. The other advantage of using this library is that it is very simple to implement as it can be achieved with few lines of code. Moreover, It compensates for inversions¹⁴. It relies on the list of words and phrases with positive and negative connotations. Sentimentr library works on valence shifters that include negators, amplifiers, deamplifiers, and adversative conjunctions. A sentiment dictionary is used by the algorithm which uses an equation to assign value to the polarity of each sentence first to tag polarized words¹⁵. The notations used in this algorithm are shown as follows in Table 2.

Each paragraph P_i consists of the sentences as equated in Eq. 5:

$$P = \{S_1, S_2, \dots, S_n\} \quad (5)$$

Every sentence is further broken down into words represented by W as in Eq. 6:

$$S_{i,j} = \{W_1, W_2, \dots, W_n\} \quad (6)$$

The polarity of the sentence is calculated using Eq.7:

Table 2. Notations and meaning

Notations	Meaning
P	Paragraph
S	Sentence
Δ	Unbounded polarity score
$C'_{i,j}$	Cluster of words
W_{amp}	Amplified words
W_{deamp}	Deamplified words
$w_{adversative\ conjunction}$	Adversative conjugate

$$\delta = \frac{C'_{i,j}}{\sqrt{w_{i,jn}}} \quad (7)$$

where:

$$\begin{aligned} C'_{i,j} &= \sum (1 + W_{amp} + W_{deamp}) \cdot w_{i,j,k}^p \cdot (-1)^{2+W_{neg}} \\ W_{amp} &= (W_b > 1) + \sum W_{neg} \cdot (z \cdot w_{i,j,k}^p) \\ W_{deamp} &= \max(W_{deamp'}, -1) \\ W_{deamp'} &= (W_b < 1) + \sum z (-W_{neg} \cdot w_{i,j,k}^a + w_{i,j,k}^d) \\ W_b &= 1 + z_2 \cdot w_{b'} \\ w_{b'} &= \sum \left(\left| w_{adversative \ conjunction} \right|, \dots, w_{i,j,k}^p, w_{i,j,k}^p, \dots, \left| w_{adversative \ conjunction} \right| * -1 \right) \\ w_{neg} &= \left(\sum w_{i,j,k}^n \right) \bmod 2 \end{aligned}$$

The main reason to opt for polarity analysis is to observe the change in the sentiments aroused from the news articles for the pre-and- peri Covid-19 period.

3. RESULTS

3.1 Change in Health Issues Coverage Printed on the E-Paper's Front Page During the Pre- and-Peri Coronavirus Period Using the LDA Topic Modeling Technique

The issues which are prominent on the front page of the newspaper during the pre-and-peri coronavirus period are illustrating using the Latent Dirichlet Allocation. The analysis of the newspaper is divided into two parts which include the pre-coronavirus period (July 2019 – Dec 2019) and the peri-coronavirus period (Jan 2020 - June 2020).

- **Pre-Coronavirus Period:**
 - **The optimal number of the topic:** For the pre-Covid-19 dataset, the best number of topics predicted using topic coherence is 15 as illuminated in Figure 6. Higher coherence will result in the best optimal number of topics. From the graph, it can be observed that there is a drastic increase in the coherence at K=15 and after that, the value of coherence decreases constantly.
 - **Topics Visualization:** The visualization of the topics is exhibited in Figure 7 in form of circles and the prevalent terms in bar graphs are illustrating the overall picture of the pre-coronavirus matters. The overall topics from 1 to 15 covers the following matters labeled as “Political parties”, “Criminal Cases”, “Economy”, “Maharashtra election”, “Court Cases”, “External matters”, “Health Education”, “Chidambaram case”, “Weather”, “Accidents”, “Terrorism”, “Election”, “Chandryaan mission”, “CAA bill”, “Ram mandir” respectively. The labels of the topics are decided according to the frequent terms occurring in the topics (refer to Table 3). During the pre-COVID-19 period, less emphasis was laid on health-related issues on the front page of the newspapers as the results predicted only one topic related to the field of health field which is labeled as “Health Education”. The leading topics during the pre-coronavirus period were politics, crime, and court cases.

Figure 6. Optimal number of topics for the PRE COVID-19

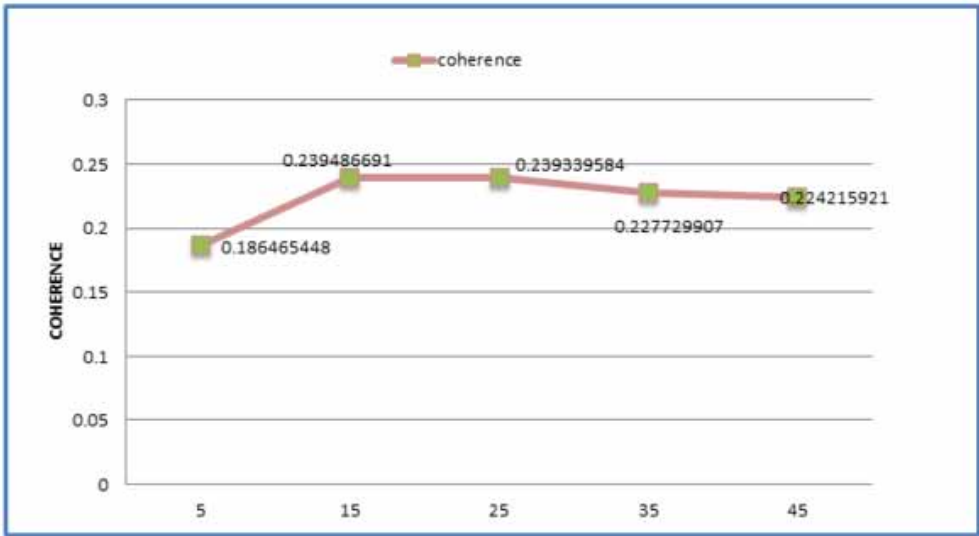


Figure 7. Visualization of PRE COVID-19 topics

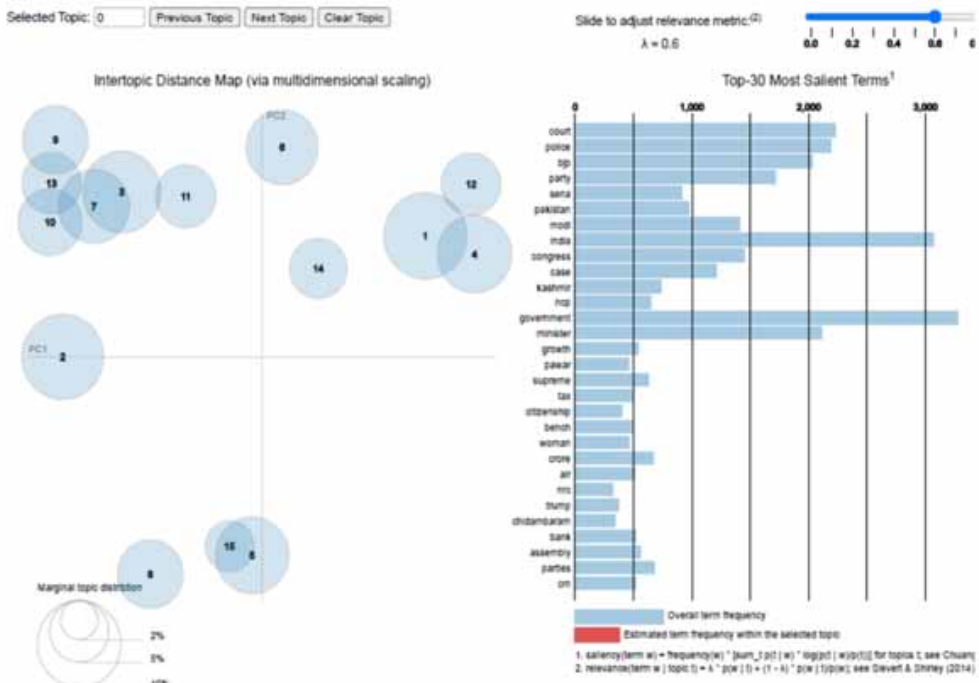


Table 3. Topics and frequent terms for PRE COVID-19 period

Topics	Top 10 frequent words
Topic 1: Political parties	Modi, government, congress, Gandhi, opposition, political, session, shah, Rahul, lok
Topic 2: Criminal Cases	Police, man, woman, hospital, murder, incident, allegedly, found, Delhi, arrested
Topic 3: Economy	Growth, tax, crore, bank, economy, GST, rs, rbi, GDP, lakhs
Topic 4: Maharashtra election	Sena, NCP, BJP, Pawar, congress, Fadnavis, Maharashtra, MLAs, Ajit, governor
Topic 5: Court Cases	Court, bench, justice, supreme, sc, petition, order, act, hearing, judges
Topic 6: External matters	India, trump, trade, Modi, China, summit, countries, world, deal, prime
Topic 7: Health Education	Heath, education, students, medical, number, million, colleges, hospital, children, cancer
Topic 8: Chidambaram case	Case, Chidambaram, court, bail, CBI, ed, accused, custody, jail, arrest
Topic 9: Weather	Pollution, monsoon, imd, rainfall, air, temperature, cold, rain, weather, climate
Topic 10: Accidents	Fire, building, people, road, Mumbai, BMC, traffic, trees, rescue, city
Topic 11: Terrorism	Pakistan, Kashmir, Jammu, terror, army, khan, Abdullah, Islamabad, valley, security
Topic 12: Election	BJP, party, congress, Jharkhand, AAP, polls, election, won, campaign, voters,
Topic 13: Chandryan mission	Flight, aircraft, space, lander, ISRO, mission, Chandrayaan, lunar, aviation, air
Topic 14: CAA bill	Citizenship, NRC, CAA, bill, protests, Assam, NPR, amendment, law, Jamia
Topic 15: Ram mandir	Ram, Ayodhya, temple, Muslim, court, waqf, disputed, board, mediation, the verdict

- **Peri-Coronavirus period:**
 - **Optimal Number of Topics:** The optimal number of topics is represented in Figure 8. From the visualization, it is clear that $K=25$ is the best number of topics as there is a drastic increase in the topic coherence at $K=25$. After that value of topic coherence increases but the change is very minor. If the number of topics will be considered higher than 25 then the terms will repeat which will lead to similar topics.
 - **Visualizing Topics:** Visualization of topics are illuminated in Figure 9 in the form of circles on the left-hand side and the most prevalent terms on the right-hand side. The prevalence of the topic is shown by the area of the circle. Topic 1 has the largest area so it is a highly prevalent topic and the least contribution is of topic 25 as it has the least area of a circle. The topics labels are decided according to the frequent terms occurring in it (refer to Table 4). "Government", "Covid-19 cases", "Economy", "Lockdown", "Medical Facilities", "India-China Conflict", "Criminal Cases", "Indian Cities", "PM Modi", "Political parties", "Migrant Workers", "Tablighi Jamaat", "Court Cases", "COVID-19 outbreak", "Covid-19 Testing", "Delhi government", "International News", "Weather", "Covid-19 treatment", "Technology", "Protestors", "Airlines", "Terrorism", "Defense", "Parliament session" these are the labels of topics 1 to 25 respectively. Topic 2, Topic 4, Topic 5, Topic 9, Topic 11, Topic 13, Topic 14, Topic 15, Topic 19 these all represent the Coronavirus related matter which consists of testing, cases, lockdown, vaccine development, plasma treatment, migrant workers during the lockdown, Tablighi jamaat COVID-19 case, health care facilities. Even some part of Topic 1 and Topic 3 discusses the coronavirus situation as it includes the steps taken by the government and the effect on the economy during a lockdown. So from the visualization in which nearly 40% of the topics are covered about COVID-19 and its impacts that indicate the

Figure 8. Optimal number of topics for PERI Covid-19

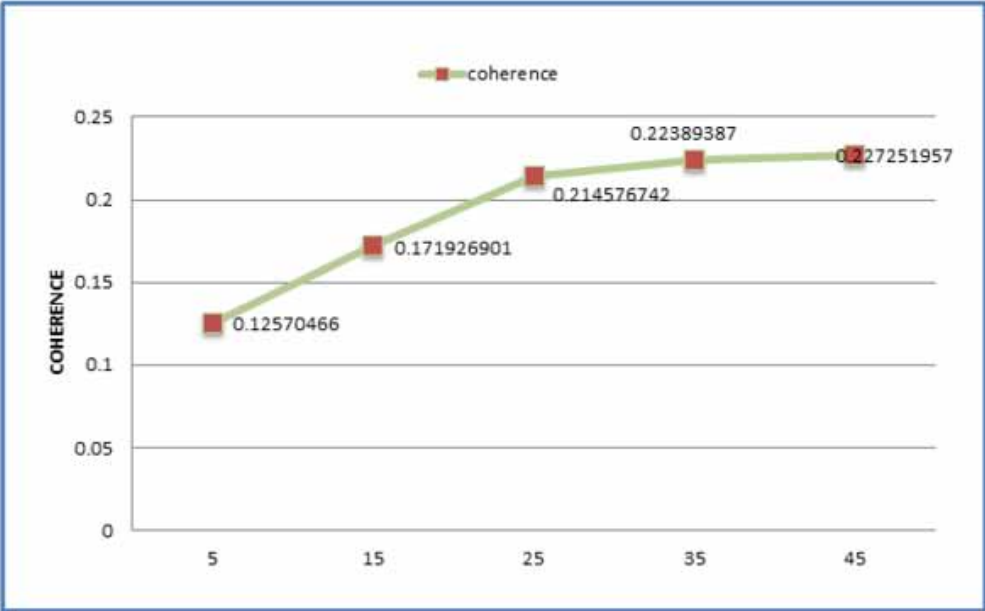


Figure 9. Visualization of PERI COVID-19 topics

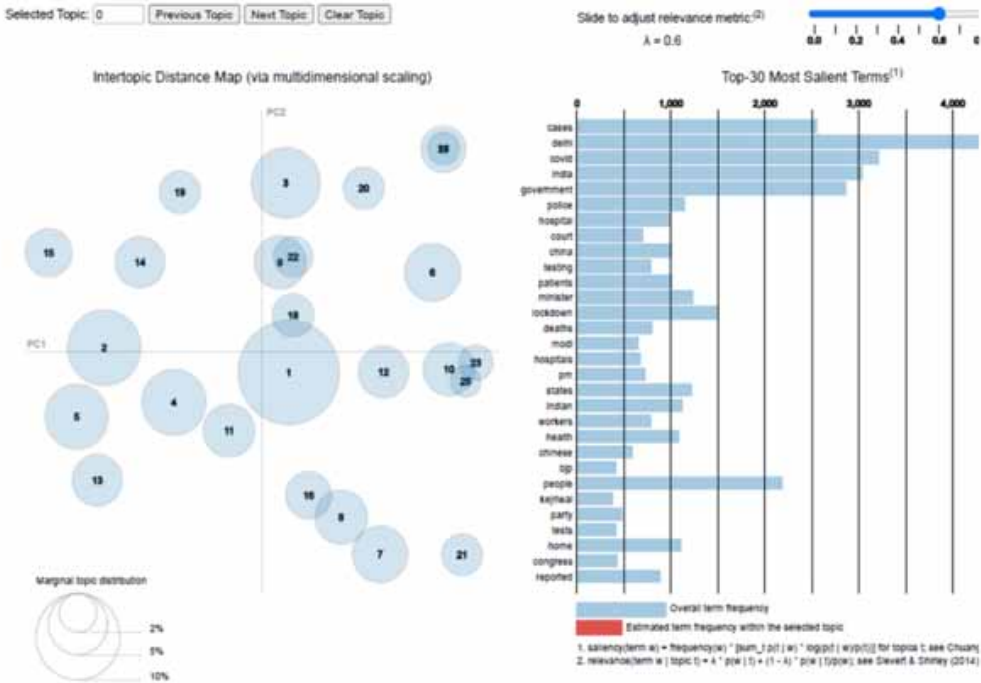


Table 4. Topics and frequent terms for PERI COVID-19 period

Topics	Top 10 frequent words
Topic 1: Government	Government, time, official, secretary, minister, COVID, center, people, report, ensure
Topic 2: COVID-19 Cases	Cases, death, COVID, reported, number, day, Delhi, Maharashtra, India, total
Topic 3: Economy	Rs, crore, India, lakh, economy, growth, lockdown, financial, rate, RBI
Topic 4: Lockdown	Lockdown, allowed, Delhi, guidelines, zones, essential, open, shops, restrictions, distancing
Topic 5: Medical Facilities	COVID, hospitals, patients, beds, health, private, doctors, positive, treatment, facilities
Topic 6: India-China Conflict	Chinese, Indian, border, LAC, Ladakh, troops, army, military, Galwan, valley
Topic 7: Criminal Cases	Police, year, man, death, officer, district, fire, body, allegedly, jail
Topic 8: Indian Cities	Noida, family, home, Delhi, lockdown, border, district, time, traffic, Ghaziabad
Topic 9: PM Modi	PM, Modi, minister, COVID, lockdown, India, Narendra, country, pandemic, people
Topic 10: Political parties	BJP, party, Congress, Government, minister, state, leader, chief, CAA, opposition
Topic 11: Migrant Workers	workers, states, migrant, trains, home, special, back, train, bihar, railways
Topic 12: Court Cases	Court, supreme, bench, justice, exams, hearing, judge, SC, HC, petition
Topic 13: Tablighi Jamaat	Containment, jamaat, zones, Tablighi, Nizamuddin, people, positive, districts, areas, tested
Topic 14: COVID-19 outbreak	Virus, outbreak, Wuhan, medical, coronavirus, China, people, Italy, infected, disease
Topic 15: Covid-19 Testing	Testing, kits, ICMR, rapid, samples, PCR, antibody, results, labs, health
Topic 16: Delhi government	Delhi, Kejriwal, Arvind, minister, government, CM, Sisodia, capital, chief, LG
Topic 17: International News	Trump, India, Nepal, China, president, countries, trade, map, oil, Washington
Topic 18: Weather	Cyclone, IMD, monsoon, temperature, Normal, ran, storm, wind, Bengal, maximum
Topic 19: Covid-19 treatment	Vaccine, drug, plasma, trials, therapy, clinical, study, COVID, drugs, research
Topic 20: Technology	Users, mobile, company, Facebook, platform, JIO, Google, RIL, products, app
Topic 21: Protestors	Police, violence, students, protestors, bagh, shaheen, Jamia, videos, riots, alleged
Topic 22: Airlines	Flights, passenger, aviation, air, airlines, Indians, aircraft, domestic, Indigo, International
Topic 23: Terrorism	Kashmir, Pakistan, terrorists, Jammu, Abdullah, detention, encounter, security, Pulwama, Jaish
Topic 24: Defense	Defense, air, missile, military, IAF, nuclear, weapons, mission, fighters, drones
Topic 25: Parliament Session	Sabha, parliament, speakers, MPs, sessions, temple, Lok, Rajya, caste, budget

focus of newspapers have now shifted to warning people about the health-related problems as the situation is very alarming around the globe with the outbreak.

3.2. Examining the Change in the Sentiments Aroused From News Articles During the Pre- and Peri-Covid-19 Era Using Sentiment Analysis

Sentiment analysis is examined using the polarity score of the news articles of the pre- (Figure 10) and peri- (Figure 11) coronavirus period. The sentiments are quantified into three categories named positive, negative, or neutral values. The total polarity score of more than 0 is considered positive, with less than 0 being considered negative, while the polarity score of 0 is considered neutral (Kawathekar

Figure 10. PRE COVID-19 Polarity Analysis



and Kshirsagar, 2012). For the pre-coronavirus period, negative polarity is represented by red color diamonds whereas positive polarity is shown by a green-colored rectangle. Neutral polarity is specified by the blue triangle. The main reasons for the positive polarity among the news articles were due to the topics Chandryan mission, Economy, Health Education, Political Parties, and Ram mandir establishment in Ayodhya whereas negative sentiment due to Terrorism, Protests against the CAA bill, Crime, Accidents. The neutral sentiment is invoked by topics labeled as Weather, Maharashtra, and Jharkhand election, and external matters.

On the other hand, the negative polarity for the peri-coronavirus period is highlighted with a diamond of mute red color. The positive polarity is pointed by the rectangles of the middle green. The zero polarity is represented using a purple-colored triangle. During the peri- coronavirus period, the negative sentiment is due to the topics related to COVID-19 which include an increase in the tally of active cases, the effect of the lockdown on the economy of India, effects on the daily wage workers. Other factors for the arousal of negative sentiments are the India-China conflict, criminal cases, and protestors. The positive polarity is indicated by the various topics which consist of PM Modi, Medical facilities, Technology, Defense, International news, Delhi government. The neutral sentiment is the result of the topics named Weather, Parliament session.

Further, the total change in the sentiment is carefully observed for the topics of the pre- and-peri COVID-19 periods to analyze the change in sentiments (Figure 12). About 54% of the polarity score is greater than zero for the pre-COVID-19 topics. On the other hand, the graph shows a decrease in the positive polarity score to 46% in the peri-coronavirus period. Similarly, the negative polarity shows a higher value for the peri-coronavirus topics that is near about 53% whereas the pre-coronavirus topics experience a downfall of up to 44%. The neutral polarity score contributes the least for the

Figure 11. PERI COVID-19 Polarity Analysis

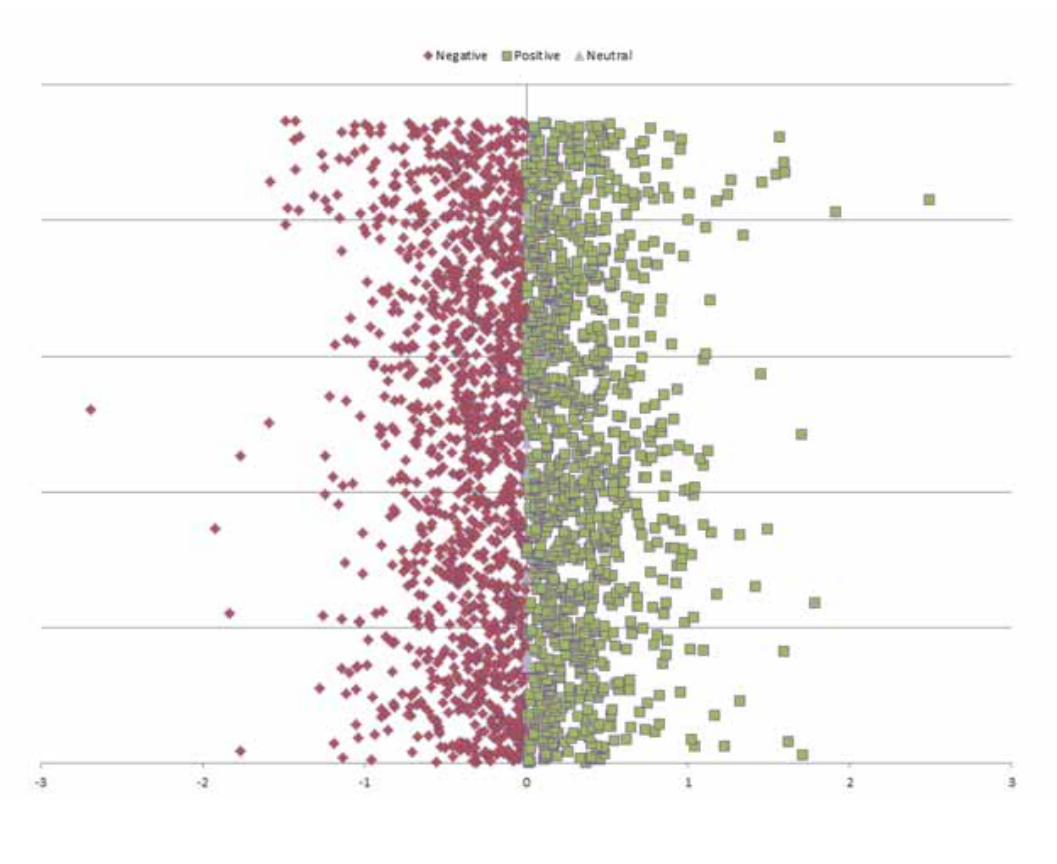
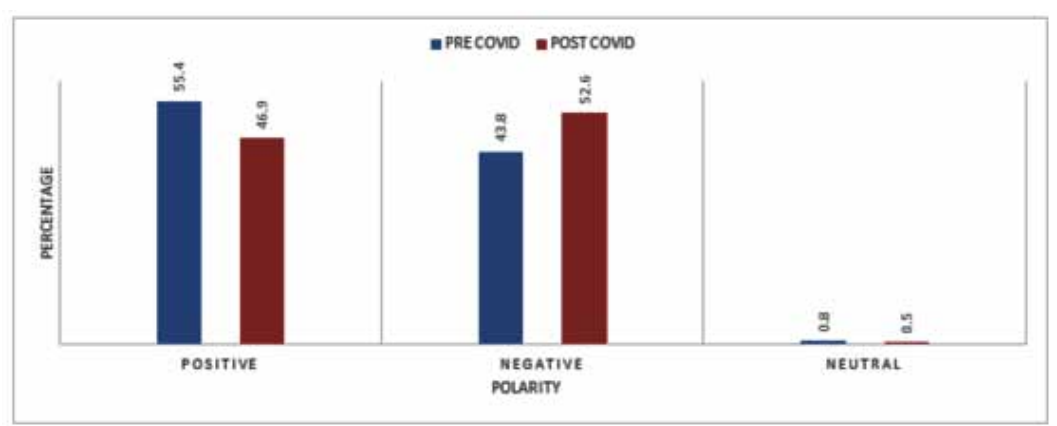


Figure 12. Polarity score comparison



pre-and-peri coronavirus topics having values of 0.8% and 0.5% respectively. The main cause of the increase in the negative sentiment during the coronavirus period is due to sudden breakdown, increase in deaths, and the restriction imposed during lockdown (refer to Table 5) such as social distancing (Mishra and Majumdar, 2020). From Table 5 it is clear that most negative polarity is due to the

Table 5. News and its sentiment score

News	Sentiment score
Prisoners freed to decongest Tihar return to crime	-2.69191
Delhi: Woman kills spouse, says he died of Covid-19	-1.83037
Pilot on his way to Delhi airport looted near IIT	-1.77076
Voice sampling leads cops clad in PPE to blind woman's rapist	-1.76637
Coronavirus death toll hits 400,000 worldwide	-1.5906
Corporate bosses in Delhi fall prey to honeytraps on gay dating app	-1.58245
Two killed in bomb attack on anti –caa rally in W Bengal	-1.48681
Parliament opens today, turbulent session on cards	-1.4865
10th Covid death in Bhopal, all gas victims	-1.24146
Delhi declares coronavirus epidemic as first Indian dies in Karnataka	-1.22454
Coronavirus killed thousands in secret too, countries now revising Covid-19 death toll	-1.19445
Covid-19: 295 deaths, highest in single day, across India	-1.18851
Destruction in Odisha, West Bengal as cyclone Amphan strikes	-1.18015
Why death rates vary widely in states: Testing could be key	-1.16163
Donald Trump calls slain Iran general Qassem Soleimani No.1 terrorist	-1.1567
7 Congress MPs suspended amid Parliament ruckus	-1.15239
Coronavirus: India crosses 50k cases; hits new peak of 3600 cases in a day	-1.14375
Mystery virus toll rises, mutation fears emerge	-1.14217
Moody's downgrades India rating for 1st time in 22 yrs	-1.13198
Newlywed fireman dies in factory blaze in Delhi's Peeragarhi	-1.13371

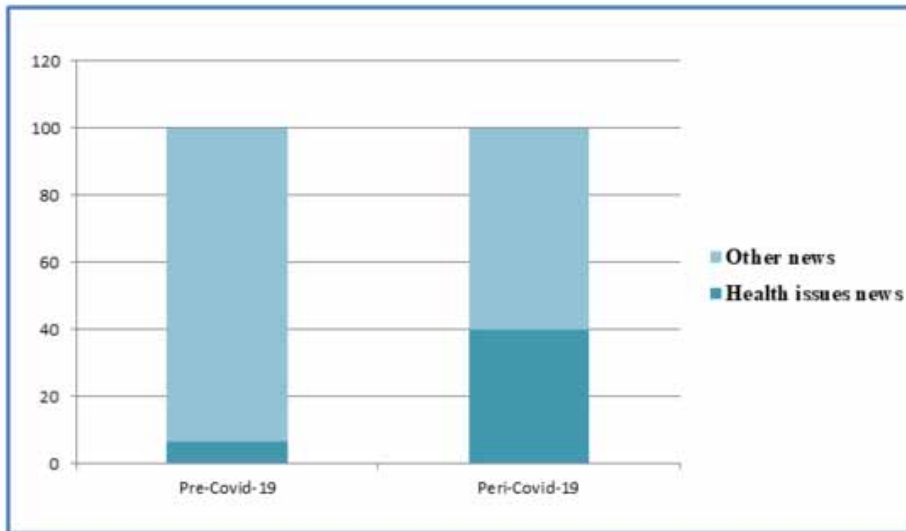
Covid-related news. These results implicate the need for formulation of the policies for improving the health sector of India as a lack of awareness can be observed.

4. DISCUSSION

In this work, we have performed the analysis of the change in health issues coverage through two Indian newspapers for the pre and peri coronavirus period. The front-page news articles for July 2019- June 2020 are collected and the LDA topic modeling technique is applied. The topics interpreted by the LDA topic modeling technique show that less stress was laid on awaring people about the various health issues during the pre-coronavirus period (Fig. 13). News covered during this period was of politics, court cases, crimes, economy, etc. Only 6% of news articles were laying stress on health education. On the other hand, the peri-coronavirus period covers nearly 40% of health-related issues which shows that the emphasis is now laid on warning people. As most of the topics were about the COVID-19 pandemic which includes COVID-19 outbreak, COVID-19 cases, COVID-19 testing, Medical facilities, Lockdown, and also includes various guidelines imposed by the government to overcome the pandemic. These observations implicate that now the health of the people is given more priority as the awareness is being spread and proper actions are also taken to control the pandemic.

The results indicate that health issues are in limelight during critical times on the front page. In both the periods' politics, crime is covered to a greater extent as it is covered under various topics named "Political parties", "Maharashtra election", "Election", "Government", "Delhi Government",

Figure 13. Change in health-issues coverage for pre-and-peri COVID-19



“Parliament session”, “CAA Bill”, “PM Modi”. Even for the crime “Criminal cases” is a common topic in both periods (refer to Table 3 and Table 4). Moreover, terrorism-related news is also given priority as it can be found in both the periods under the topic “Terrorism”. The media should also give equal priority to the health issues coverage which will ultimately lead to the improvisation of the Indian healthcare system as people will be more aware. The government should formulate some initiatives to improve the health care system of India. Further, the change in sentiments for both periods is observed using the polarity score. The results revealed that the negative sentiment is now prominent in the peri-coronavirus period. About 52% of the polarity score is negative in the peri-coronavirus period whereas it was 46% during the pre-coronavirus time. The main cause of the increase in negative polarity score is due to the restrictions are imposed on the people such as a sudden breakdown of the virus, social distancing, wearing masks, and lockdown period when people were made to live at their homes. The outbreak of the virus is sudden and people are in a state of fear and get mentally disturbed. Since the negative sentiment has increased in the peri-Covid-19 period which implicates that the media should decipher some suitable headlines for covering the spread of the Covid-19 which will lay less mental stress on the people.

In the previous work, Gupta and Sinha (2010) inspected health coverage in mass media. The analysis is done manually which takes a lot of time. The limitation of this study is overcome by our study as we have used the LDA topic modeling to carry out the analysis which automatically extracts the topics covered. The approach used by our study is comparatively faster.

Liu et al. (2020) analyzed the role of media in the on-going COVID-19 crisis is analyzed. The main drawback of this study was that it covers news related to the COVID-19 and only contains Chinese news articles that were scraped from the WiseSearch mass media. Also, it lacks the sentiment analysis of news articles. The drawbacks of this study are overcome by our research as we have collected news articles from the front page of the e-papers which contains national as well as international news. Further, the news articles collected for our study are from different areas that include politics, business, government, court-cases, etc. Moreover, we have analyzed the change in the sentiments aroused from the news articles.

5. CONCLUSION

The healthcare sector of India is considered unfit when compared with other country's medical facilities due to the lack of awareness among the public. In this study, we have analyzed the change in health issues coverage through the front page of two Indian newspapers viz. The Hindustan Times and the Times of India for pre-and- peri coronavirus periods. The results show that the newspapers have now focused on warning the people about the pandemic by providing the total number of active cases, providing prevention guidelines, and various other articles are also covered to aware people of the alarming situation that has devastated the whole world. Further, the sentiment analysis of the news articles of the pre-and-peri COVID-19 period shows an increase in the negative sentiment during the peri- coronavirus period which indicates fear due to the sudden breakdown of the pandemic. The theoretical implication of this research includes: the study fills in the gaps in empirical knowledge about the factors that trigger sentiments of the people. It also indicates the psychological effects which implicate how news can bring behavioral change. Moreover, the study proposes practical implications for researchers, health workers, the government who are interested in using the news media for disseminating knowledge about the various health-issues. Further, it will help the government in improving the Indian health sector and to initiate different programs for people to help them come out of this pandemic situation as people are at high risk of getting mental illness. In the future, we would like to combine the various social media platforms such as Facebook, Instagram, Twitter posts, and news articles from different e-papers to analyze the situation. Also, we would like to analyze the state-wise pandemic situation in India. Further, we would like to consider news articles for the whole newspaper rather than focusing on the front page only.

FUNDING AGENCY

The publisher has waived the Open Access Processing fee for this article.

REFERENCES

- Apuke, O. D., & Omar, B. (2021). Fake news and COVID-19: Modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, 56, 101475. doi:10.1016/j.tele.2020.101475 PMID:34887612
- Arun, R., Suresh, V., Madhavan, C. V., & Murthy, M. N. (2010, June). On finding the natural number of topics with latent dirichlet allocation: Some observations. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 391–402). Springer. doi:10.1007/978-3-642-13657-3_43
- Barkur, G., Vibha, G. B. K., & Kamath, G. B. (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian Journal of Psychiatry*, 51, 102089. doi:10.1016/j.ajp.2020.102089 PMID:32305035
- Beauvoys, E., Dupéré, S., & Guitton, M. J. (2020). COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. *Computers in Human Behavior*, 111, 106424. doi:10.1016/j.chb.2020.106424 PMID:32398890
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993–1022.
- Chandni, N. C., Gupta, S., & Pahade, R. (2015). Sentiment analysis and its challenges. *International Journal of Engineering Research & Technology (Ahmedabad)*, 4(3), 968–970.
- Chehal, D., Gupta, P., & Gulati, P. (2020). Implementation and comparison of topic modeling techniques based on user reviews in e-commerce recommendations. *Journal of Ambient Intelligence and Humanized Computing*, 1–16.
- Daud, A., Li, J., Zhou, L., & Muhammad, F. (2010). Knowledge discovery through directed probabilistic topic models: A survey. *Frontiers of Computer Science in China*, 4(2), 280–301. doi:10.1007/s11704-009-0062-y
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407. doi:10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9
- García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72). Springer International Publishing.
- García-Peñalvo, F. J. Transformación digital en las universidades: Implicaciones de la pandemia de la COVID-19. *Education in the Knowledge Society*, 22, e25465-e25465.
- García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande, M. (2020). Online assessment in higher education in the time of COVID-19. *Education in the Knowledge Society*, 21.
- García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande-de-Prado, M. (2021). Recommendations for Mandatory Online Assessment in Higher Education During the COVID-19 Pandemic. In *Radical Solutions for Education in a Crisis Context* (pp. 85–98). Springer. doi:10.1007/978-981-15-7869-4_6
- Gupta, A., & Sinha, A. K. (2010). Health coverage in mass media: A content analysis. *Journal of Communication*, 1(1), 19–25. doi:10.1080/0976691X.2010.11884766
- Heidenreich, T., Lind, F., Eberl, J. M., & Boomgaarden, H. G. (2019). Media framing dynamics of the ‘European refugee crisis’: A comparative topic modelling approach. *Journal of Refugee Studies*, 32(1), i172–i182.
- Hidayatullah, A. F., Aditya, S. K., & Gardini, S. T. (2019, February). Topic modeling of weather and climate condition on twitter using latent dirichlet allocation (LDA). *IOP Conference Series. Materials Science and Engineering*, 482(1), 012033. doi:10.1088/1757-899X/482/1/012033
- Hodder, A. (2020). New Technology, Work and Employment in the era of COVID-19: Reflecting on legacies of research. *New Technology, Work and Employment*, 35(3), 262–275. doi:10.1111/ntwe.12173 PMID:32836624
- Hofmann, T. (1999, August). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 50–57). ACM.
- Jahanbin, K., & Rahmadian, V. (2020). Using Twitter and web news mining to predict COVID-19 outbreak. *Asian Pacific Journal of Tropical Medicine*, 13(8), 378. doi:10.4103/1995-7645.279651

Kasthuri, A. (2018). Challenges to healthcare in India-The five A's. *Indian Journal of Community Medicine*, 43(3), 141.

Kawathekar, S. A., & Kshirsagar, D. M. M. (2012). Movie Review analysis using Rule-Based & Support Vector Machines methods. *IOSR Journal of Engineering*, 2(3), 389–391. doi:10.9790/3021-0203389391

Kim, Y. S., & Chung, D. S. (2017). Anatomy of front pages: Comparison between The New York Times and other US major metropolitan newspapers. *International Journal of Communication*, 11, 18.

Kong, W. H., Li, Y., Peng, M. W., Kong, D. G., Yang, X. B., Wang, L., & Liu, M. Q. (2020). SARS-CoV-2 detection in patients with influenza-like illness. *Nature Microbiology*, 5(5), 675–678. doi:10.1038/s41564-020-0713-1 PMID:32265517

Liu, Q., Zheng, Z., Zheng, J., Chen, Q., Liu, G., Chen, S., Chu, B., Zhu, H., Akinwunmi, B., Huang, J., Zhang, C. J. P., & Ming, W. K. (2020). Health communication through news media during the early stage of the COVID-19 outbreak in China: Digital topic modeling approach. *Journal of Medical Internet Research*, 22(4), e19118. doi:10.2196/19118 PMID:32302966

Liu, X., & Lo, V. H. (2014). Media exposure, perceived personal impact, and third-person effect. *Media Psychology*, 17(4), 378–396. doi:10.1080/15213269.2013.826587

Mishra, M., & Majumdar, P. (2020). Social Distancing During COVID-19: Will it Change the Indian Society? *Journal of Health Management*, 22(2), 224–235. doi:10.1177/0972063420935547

Poirier, W., Ouellet, C., Rancourt, M. A., Béchard, J., & Dufresne, Y. (2020). (Un) Covering the COVID-19 Pandemic: Framing Analysis of the Crisis in Canada. *Canadian Journal of Political Science/Revue canadienne de science politique*, 53(2), 365–371.

Ram, N. (2011). The changing role of the news media in contemporary India. *The Indian Historian*.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14–46. doi:10.1016/j.knosys.2015.06.015

Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399–408). doi:10.1145/2684822.2685324

Sahni, H., & Sharma, H. (2020). Role of social media during the COVID-19 pandemic: Beneficial, destructive, or reconstructive? *International Journal of Academic Medicine*, 6(2), 70.

Sharma, M. S. K., & Gupta, D. Y. K. (2017). Mass Media for Health Education-A Study in Rajasthan (A Study in the State of Rajasthan). *GRA's Multidisciplinary International (GRAM i) Journal*, 1(1).

Sievert, C., & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63–70). doi:10.3115/v1/W14-3110

Wang, H., Wang, Z., Dong, Y., Chang, R., Xu, C., Yu, X., Zhang, S., Tsamlag, L., Shang, M., Huang, J., Wang, Y., Xu, G., Shen, T., Zhang, X., & Cai, Y. (2020). Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China. *Cell Discovery*, 6(1), 1–8. doi:10.1038/s41421-020-0148-0 PMID:32133152

Wyer, R. S. Jr, & Shrum, L. J. (2015). The role of comprehension processes in communication and persuasion. *Media Psychology*, 18(2), 163–195. doi:10.1080/15213269.2014.912584

Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLoS One*, 15(9), e0239441. doi:10.1371/journal.pone.0239441 PMID:32976519

Yu, J., Lu, Y., & Muñoz-Justicia, J. (2020). Analyzing Spanish News Frames on Twitter during COVID-19—A Network Study of El País and El Mundo. *International Journal of Environmental Research and Public Health*, 17(15), 5414. doi:10.3390/ijerph17155414 PMID:32731359

Zhang, X., Bie, B., & Billings, A. C. (2017). Newspaper Ebola articles differ from Twitter updates. *Newspaper Research Journal*, 38(4), 497–511. doi:10.1177/0739532917739883

ENDNOTES

- ¹ https://en.wikipedia.org/w/index.php?title=Media_of_India
- ² https://covid19.who.int/?gclid=CjwKCAiA_Kz-BRAJEiw_AhJNY7_EncMYsfYey_EtTKMgnLH9uhDkCKjxTUJ7T4PA_cVD92Y_8Hl00hEU_JxoC4B_EQAvD
- ³ <https://www.worldometers.info/coronavirus/country/india/>
- ⁴ https://en.wikipedia.org/wiki/COVID-19_lockdown_in_India
- ⁵ <https://www.w3newspapers.com/india/>
- ⁶ <https://thingsinindia.in/10-best-english-newspapers-in-india/>
- ⁷ <https://data-miner.io/>
- ⁸ <https://medium.com/sciforce/text-preprocessing-for-nlp-and-machine-learning-tasks-3e077aa4946e>
- ⁹ <https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf>
- ¹⁰ <https://monkeylearn.com/blog/introduction-to-topic-modeling/>
- ¹¹ <http://github.com/cpsievert/LDAvis#:~:text=R%20package%20for%20interactive%20topic,an%20interactive%20web-based%20visualization.>
- ¹² <https://monkeylearn.com/sentiment-analysis/>
- ¹³ <https://cran.r-project.org/web/packages/sentimentr/readme/README.html>
- ¹⁴ <https://towardsdatascience.com/doing-your-first-sentiment-analysis-in-r-with-sentimentr-167855445132>
- ¹⁵ <https://CRAN.R-project.org/package=syuzhet>

Jasdeep Kaur is a research student in the Department of Computer Science and Engineering, Guru Nanak Dev University, Amritsar. Her research interests are in Machine Learning and Data Analytics.

Amit Chhabra is working as an Assistant Professor in the department of Computer Engineering & Technology, Guru Nanak Dev University, Amritsar.

Munish Saini is an Assistant Professor in the Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar, India. He received his Ph.D. in Computer Science and Engineering from Guru Nanak Dev University, Amritsar, India. His research interest are in Open Source Software, Software Engineering, Data Mining, and Machine Learning.

Nebojsa Bacanin received his first Ph.D. degree in 2014 from the domain of applied computer science, and second Ph.D. degree from Faculty of Mathematics, University of Belgrade in 2015 (study program Computer Science, average grade 10,00). He started University career in Serbia 15 years ago at Graduate School of Computer Science in Belgrade. He currently works as an associate professor and as a Vice-Rector for Scientific Research at Singidunum University, Belgrade, Serbia. He teaches 16 courses on bachelor, master and Ph.D. studies from the domain of computer science. He is involved in scientific research in the field of computer science and his specialty includes stochastic optimization algorithms, swarm intelligence, soft-computing and optimization and modeling, as well as artificial intelligence algorithms, swarm intelligence, machine learning, image processing and cloud and distributed computing. He has published more than 100 scientific papers in high quality journals and international conferences indexed in Clarivate Analytics JCR, Scopus, WoS, IEEEExplore, and other scientific databases, as well as in Springer Lecture Notes in Computer Science and Procedia Computer Science book chapters. He has also published 2 books in domains of Cloud Computing and Advanced Java Spring Programming. He is a member of numerous editorial boards, scientific and advisory committees of international conferences and journals. He is a regular reviewer for international journals with high Clarivate Analytics and WoS impact factor such as Journal of Ambient Intelligence & Humanized Computing, Soft Computing, Applied Soft Computing, Information Sciences, Journal of Cloud Computing, IEEE Transactions on Computers, IEEE Review, Swarm and Evolutionary Computation, Knowledge-based Systems, Future Generation Computer Systems, Computer and Information Sciences, SoftwareX, Neurocomputing, Operations Research Perspectives, etc. In 2020 he was designated by prestigious Stanford University list as top 2% researchers in the world.