J. Mountain Res. P-ISSN: 0974-3030, E-ISSN: 2582-5011
Vol. 19(2), (2024), 203-211

DOI: https://doi.org/10.51220/jmr.v19-i2.21

Sentiment Analysis, Emotion Detection and Topic Modeling of Covid-19 Twitter Posts in Uttarakhand

Mahendra Singh Aswal[1] • Vishnu Vishwakarma[1] • Krishan Kumar[1] • Durgesh Pant[2]

[1]*Department of Computer Science, Gurukula Kangri (Deemed to be University), Haridwar*

[2]*Uttarakhand Council of Science and Technology, Dehradun*

*Corresponding author email id: msa@gkv.ac.in

Received: 22.06.2024 Revised: 08.11.2024 Accepted: 17.11.2024

Abstract: The COVID-19 pandemic, which began in 2019, caused significant global upheaval, with severe impacts across countries like the United States, Brazil, Russia, and India. Due to the lack of effective treatments initially, governments and citizens faced serious challenges. Social distancing and lockdowns prompted many to express their views on platforms like Twitter (now X). This research analyzes sentiments from COVID-19 tweets of Uttarakhand residents, posted between March 2020 and December 2022. Using a dataset of these tweets, NLP tools TextBlob and VADER were applied for sentiment analysis, emotion detection, and topic modeling. Findings indicate that most users displayed a positive outlook during the pandemic. This study aims to help authorities in Uttarakhand understand public sentiment, enabling them to make informed decisions for crisis management.

Keywords: COVID-19 • Sentiment Analysis • Tweets • Social Media • NLP • Uttarakhand

Introduction

Nowadays, Internet is easily accessible on most devices and has become an integral part of our lives. It provides a cost-effective platform for information propagation through the widespread use of social media. Various social media platforms like Facebook, blogs, and Twitter are being sought for deriving people's opinions or sentiments related to a specific item, institution, or event. Perception and emotions constitute a crucial part of assessing the behavior of a person that is termed as sentiments. The analysis of these sentiments about an object or situation is known as sentiment analysis or opinion mining. Sentimental analysis enables us to understand the sentiments or emotions of users and categorize them into different labels that can further help an organization or authority to be aware of the opinions of the people and take decisions accordingly (Singh et al 2021).

The emergence of the novel coronavirus, officially named as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), in late 2019, marked the onset of global health crisis COVID-19 that has since reshaped societies, economies and healthcare systems worldwide. COVID-19, initially identified in Wuhan China swiftly evolved into a global pandemic, spreading across the countries and promoting widespread concern. The first case of the COVID-19 pandemic in India was reported on 30 January 2020. Gradually, the pandemic spread to various states and union territories, including the state of Uttarakhand. The first case of COVID-19 in this area was registered on 15 March 2020. (Kumar 2022)

The COVID-19 pandemic has brought a multitude of challenges for people around the world. These challenges span various aspects of life, including health, economy, education and mental well-being (Jain and Yadav 2021). Some of the main challenges faced by people during the COVID-19 period are risk of infection, healthcare access, job loss and income reduction, school closures, educational inequities, social isolation, mental stress, travel restrictions and remote-work challenges and so on (Archna 2021). This unforeseen disaster witnessed an unprecedented increase in the use of social networking sites. People took up the social media platforms like Twitter, Facebook etc. to express their opinions about the prevailing situations. Throughout the different waves of the COVID-19 pandemic and

DOI: https://doi.org/10.51220/jmr.v19-i2.21

periods of stringent lockdowns, individuals have been conveying their feelings and opinions on platforms like Twitter. Numerous studies were conducted on COVID-19 tweets by researchers to investigate the sentiments of the people using various artificial intelligence (AI) enabled methods, such as machine learning (ML), deep learning (DL) and natural language processing (NLP) tools. The majority of research was focused on examining tweets exclusively in the English language originating from different countries or a particular region (Nemes and Kiss 2020). In India, a number of such studies were also carried out to analyze the mood of the different state's people during the pandemic. However, no such study was conducted in Uttarakhand state of India.

In this paper, we have proposed a framework to classify the sentiments or opinions of the people of the Uttarakhand state by analyzing their COVID-19 related tweets. The COVID-19 Twitter posts originating from Uttarakhand from March 2020 to December 2022 were analyzed to discover the polarity of sentiments and emotional state of the people using the NLP libraries of python. We have also derived the topics of discussion from the Twitter posts using topic modeling technique. COVID-19 Hindi tweets are also analyzed in the proposed study by converting them into English tweets using the Google translator service.

The main contributions of this study can be cited as follows:

- Proposes a framework that analyzes the tweets from social media for understanding the public sentiments during the COVID-19 Pandemic
- Predicts the emotional tone of the people during the crisis of COVID-19 by extracting features from the tweets
- Identifies the important topics of discussion from the COVID-19 tweets.
- Presents a comparison between the results obtained using sentiment classification tools VADER and TextBlob

The rest of the paper is organized as follows: In the next section, the review of the related work is presented. In the section 3 the proposed methodology is described. In section 4 the results obtained using the NLP tools TextBlob and Vader are compared and analyzed. Section 6 concludes the paper.

Review of Literature

The existing approaches for sentiment analysis of COVID-19 tweets in India are reviewed here. Majority of the works use the NLP libraries, ML and DL based models for sentiment and emotional analysis of the social media data.

The authors developed a model to analyze people's opinions about COVID-19 situations around the world including India (Singh et al 2021). In the proposed model, researchers scrapped the data from twitter using the existing Twitter APIs, and maintained two data sets, i.e., world-specific and India-specific data sets. The sentiment analysis is then performed using different matrices like Average Likes and Re-tweets of a post, Intensity Analysis, Polarity & Subjectivity, and Wordcloud. Along with this, the Bidirectional Encoder Representations from Transformers (BERT) model is also used for the classification of public opinions on the corona virus.

Twitter (Now X) was popularly used by people to express their views and opinions related to these events in diverse temporal and geographical settings. Authors represented these diversities of people's reactions by predicting and examining the sentiments of geotagged tweets during the lockdown and un-lockdown phases of the pandemic (Kumar 2022). For sentiment predictions, a hybrid DL-based model was proposed which utilizes the powers of Bidirectional Long Short-term Memory (BiLSTM) and CNN (Convolutional Neural Network) model. The model was trained on a free data-set and was tested over manually labeled COVID-19 tweets from India. The model categorized the tweets with high accuracy of around 90%, and investigation of geotagged tweets unfolds vital geographical diversification. Analyzing reactions of people towards the situations and events amidst an existent pandemic can help in decision making.

DOI: https://doi.org/10.51220/jmr.v19-i2.21

Researchers also focused on the challenges and problem faced by the people and their mental health during COVID-19 in India (Jain and Yadav 2021, Archna 2021). A structured survey was undertaken involving 250 participants spanning various age groups. The Covid-19 sentiments and manifestations based on social media platform Twitter were analyzed (Nemes and Kiss 2020). The primary goal of this article is to develop a model for sentiments analysis. Researchers collected the twitter's data between 24 April 2020 to 25 April 2020. They used the Recurrent Neural Networks (RNN) and NLP Python library TextBlob for sentiment analysis and compared the results. The accuracy of TextBlob is found 30% higher than that of RNN model. A web portal-Mood of India During Covid-19, which displayed visualizations of emotions of people across various states in the country, is developed in (Venigalla et al 2020). The tweets posted on twitter pertaining to COVID-19 are investigated and categorized into one of the seven classes of emotions - Anger, Sadness, Happiness, Surprise, Fear, Disgust and Neutral category, which are depicted on India Map indicating the place from which the tweets have been posted.

A research work was carried out to analyze the sentiments of political leaders of state and union territories during Covid-19 period in India while they took up social media platforms like twitter for communicating with peoples (Kaur et al 2021). Researchers has obtained the total 12,128 tweets from the official Twitter account of 29 political leaders from 11 March 2020 to 03 May 2020. Before the analysis, they preprocessed the data and removed all irrelevant content i.e. URLs, punctuation, @ mentions, stop words etc. and translated some non-English tweets into English tweets using translator. The frequently discussed topics by each leader were identified and visually presented using the 'ggplot' tool. Furthermore, they carried out subjectivity and polarity analysis of sentiments using TextBlob library. Subjectivity scores showed that more than half of the leaders shared information based on facts. Polarity scores revealed that almost 90% of leaders shared the information in a positive or neutral manner. Researchers investigated and analyzed the sentiments of Indian peoples regarding COVID-19 vaccine (Mir & Sevukan 2022).. The study involved the utilization of the two random chosen hashtags '#Covid19Vaccine' and '#CoronaVirusVaccine' as search keywords to download tweets. The twitter archiver was used to fetch the tweets against the hashtags. They have collected 11,815 tweets during 4 January 2021 to 22 March 2021. Researchers selected only those tweets which are related to India and keywords. They used standard pre-processing techniques including transformation, tokenization and filtering using orange software and removed 1269 duplicates tweets, resulting in only 1431 tweets left for analyzing. VADER detected 639 tweets (44.65%) as positive, 551 tweets (38.50%) as neutral and the smallest proportion (241, 16.84%) of tweets were detected having negative viewpoint regarding the vaccines. The top 76 terms were frequently used by the Indians to convey their opinions on the COVID-19 vaccination. The terms 'Delhi Heart', 'Lung Institute', 'Gift', 'Unite 2 Fight Corona' and 'Covid19Vaccine' were the most discussed terms with the highest relevance scores between 2.91 and 1.82.

A framework that utilizes DL based language model via long short-term memory (LSTM) recurrent neural networks for multi-label sentiment analysis of tweets posted during the peak time of novel COVID-19 cases in India, is proposed in (Chandra & Krishna 2021). Researchers compared the results of their LSTM and BD-LSTM models with that of state-of-art BERT model and then selected the best model for multi-label COVID-19 sentiment analysis. The study investigated three data sets featuring tweets from India and from two other states Maharashtra and Delhi collected between March to September 2020. Researchers used a sentiment dataset named as Senwave COVID-19 consisting of 10,000 tweets collected worldwide between March to May 2020 and manually labeled by a group of 50 experts for training deep learning models. They used 90% tweet data for training and 10% for testing. In the

DOI: https://doi.org/10.51220/jmr.v19-i2.21

case of India (Panel a), it is observed that sentiments such as "optimistic", "annoyed" and "Joking" were most communicated which also shows in the case of Maharashtra and Delhi (Panel b and c). It is discovered that the BERT model seems to extract more sentiments exhibited in respect to the LSTM model, specifically the "optimistic", "anxious", and "annoyed" sentiments. Authors noticed that negative sentiments such as "pessimistic", "anxious" and "sad" have been least demonstrated.

A deep-learning framework, Cov-hindia is proposed that captures the sentiment polarity of Hindi tweets related to COVID-19 posted on the Twitter platform (Singh 2020). The proposed framework performs machine translation on Hindi tweets and conveys the converted tweets as input to a deep learning model which utilizes the COVID-19 dataset available on Kaggle called "Covid19 Indian Sentiments on Covid19 and Lockdown" containing 3090 English tweets posted from India for training. The tweets have been gathered during 23rd March 2020 and 15th July 2020 and are tagged into four emotion categories fear, sad, anger, and joy. Researchers explored the performance of nine deep learning models in capturing the polarity of sentiments on an English tweet dataset and found that BERT model is having the highest accuracy in polarity detection. By exploiting Google Translate and the pre-trained BERT model, the proposed framework for Covhindia carries out sentiment classification on Hindi tweets.

Methodology

The proposed methodology for sentiment analysis of COVID-19 tweets follows a sequence of following phases: Data Collection, Data Pre-Processing, Sentiment Analysis of tweets, Emotion Classification and Topic Modelling of COVID-19 data (Fig 1).

Data Collection: COVID-19 tweets originating from Uttarakhand and spanning the critical period of the pandemic from March 2020 to December 2022 were systematically gathered from the twitter using a Python application which leverages open-source API. The relevant tweets containing the keywords and hashtags such as #COVID-19, #covid, #CoronaVirus, #StayHomeChallange and #lockdown etc. were extracted through the application. Additionally, we also used online repositories such as Kaggle and GitHub, which provide valuable resources for research and analysis in order to fetch the relevant data related to COVID-19. The extracted tweets consist of fields like user name, date of tweet, tweet text and location. A total of 679 such tweets belonging to Uttarakhand state were collected for further processing.

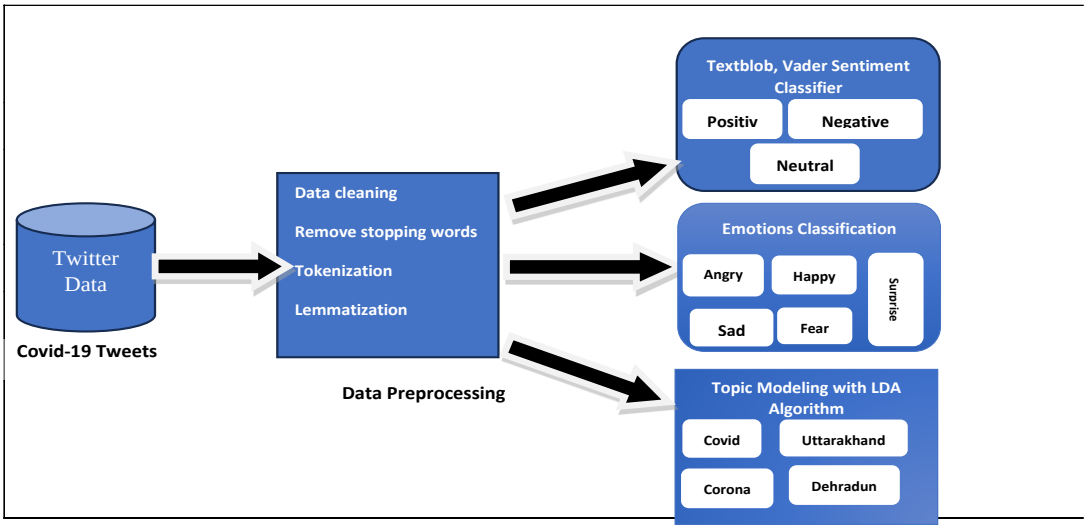

Fig. 1: Block Diagram of Methodology for Sentiment Analysis

Data Pre-processing: Data pre-processing is a crucial stage in natural language processing (NLP) where unstructured text data is cleaned and modified to make it suitable for analysis. The first

DOI: https://doi.org/10.51220/jmr.v19-i2.21

step involved the removing the URLs, emails using regular expression and text converted into lower case. URLs and emails addresses in tweets can introduce irrelevant information and potentially compromise the privacy thus making their elimination important. Next, mention (represented by the '@' symbol before the user name) and hashtag (recognized by the '#' symbol) were extracted using regular expression and subsequently deleted. Also, we implemented tokenization, lemmatization and stop-words removal using NLTK's. These elements often add noise and may not contribute to the core content of tweet.

Noisy words, such as 'RT' (Re-tweet) and '&' identified through manual inspection were eliminated from the text to enhance the semantic clarity of the tweets. Additionally, we removed special characters, numbers non-ASCII characters were removed from the tweets to standardize the character set. By applying these steps, the research aims to provide a cleaner and more structured dataset for further investigation, this enables us to extract meaningful patterns and insights from the text with greater accuracy. This methodology contributes to advancing the field of social media analytics and improving the reliability of findings the from the Twitter content.

Sentiment Analysis: Sentiment analysis is a valuable tool in natural language processing (NLP), allowing us to understand the emotional tone of a piece of text. Textblob and VADER (Valance Aware Dictionary and Sentiment Reasoner) are two popular tools employed for sentiment analysis. In this research, we conducted a comprehensive sentiment analysis of Twitter data in Uttarakhand during the COVID-19 pandemic, leveraging Natural Language Tool Kit (NLTK) with VADER and TextBlob. NLTK's VADER, is a lexicon and rule-based sentiment analysis tool, and TextBlob is an NLP library, which are used to evaluate the emotional tone of Tweets. These tools enable us to measure the polarity of sentiments such as positivity, negativity and neutrality within the text. The objective of our research work is to understand the public sentiment towards the unprecedented crisis COVID-19 in Uttarakhand.

This analysis revealed dynamic shifts in public sentiment over time, reflecting the evolving nature of the pandemic. NLTK's VADER and TextBlob provided finer insights into the emotional responses of Uttarakhand's Twitter users and a deeper understanding of how individuals expressed themselves during this challenging period.

TextBlob is a Python library used for Natural Language Processing (NLP). It relies on NLTK (Natural Language Toolkit). When you give it a sentence, it gives back two things: polarity and subjectivity.

The polarity score ranges from -1 to 1. A score of -1 means the words are super negative, like "disgusting" or "awful." A score of 1 means the words are super positive, like "excellent" or "best."

Subjectivity score, on the other hand, goes from 0 to 1. If it's close to 1, it means the sentence has a lot of personal opinions instead of just facts. For our work, we were mostly interested in the polarity score because we were not focusing on facts.

Emotion Classification: Traditional sentiment analysis sometimes falls short in capturing the subtle spectrum of human emotions. Unlike simple sentiment analysis, which only tells if something is positive or negative, emotion analysis dives deeper into understanding the real feelings people have. To address this limitation, our research also delves into emotion analysis. Using this approach, especially focusing on tweets and posts related to the COVID-19 pandemic for Uttarakhand, we have used a special tool called Text2Emotion, which acts like an emotion detective in text. It helped us find out just whether people were 'Happy' or 'Sad' but also identified more specific emotions like 'Anger', 'Surprise' and 'Fear'. It facilitated us to gain a better and more detailed insight into the true emotions people were experiencing during this challenging time in Uttarakhand.

Topic Modeling: Topic Modeling is a powerful NLP technique designed to uncover latent thematic

DOI: https://doi.org/10.51220/jmr.v19-i2.21

structures within large collections of texts. In our study, we used the Latent Dirichlet Allocation (LDA) algorithm for topic modeling. The primary goal of topic modeling is to reveal the underlying topics in a diverse set of documents. This technique proves useful in different areas, helping with tasks such as summarizing COVID-19-related texts, categorizing documents about the pandemic, and suggesting relevant content for research and understanding of the Corona virus.

Result Analysis

In order to perform sentiment analysis of COVID-19 tweets posted on the Twitter platform, we gathered a dataset comprising 679 tweets from Uttarakhand spanning from March 2020 to December 2022. To ensure data quality, we performed thorough data preprocessing, which included removing unreadable tweets, retweets, and implementing tokenization, lemmatization, and stop-word removal. Additionally, we translated Hindi tweets into English tweets using Google Translator, which has varying accuracy rates between languages, ranging from 55% to 94%. A study conducted in 2021 by the UCLA Medical Centre found that the overall accuracy of Google Translate was 82.5%. Following the aforementioned pre-processing steps, we were left with only 512 tweets and we analyzed these for our research.

Sentiment Analysis with TextBlob and VADER

In our comprehensive performance study, we utilized the effectiveness of sentiment analysis toolkits, called TextBlob and VADER, in understanding the sentiments and emotional states of people of Uttarakhand during COVID-19 period in India. TextBlob assigned sentiment labels to the tweets, categorizing 282 as positive, 135 as neutral, and 94 as negative as shown in Figure 1. On the other hand, NLTK's VADER, another sentiment analysis tool, labeled the sentiments as 245 positive, 162 negative, and 104 neutral as depicted in Figure 2.

The variation in the sentiment distribution by TextBlob and VADER is observed in the results. An experimental study of sentiment analysis shows that VADER outperforms TextBlob in detecting negative polarity; while VADER classified certain sentences as negative, TextBlob identified them mostly as positive. What makes VADER distinct is its ability to not only classify words as positive, negative, or neutral but also evaluate the overall sentiment of a sentence.

The comparative analysis of these tools unveiled notable difference in their performance. We employed a world-wide COVID-19 dataset consisting of 5000 labeled tweets to assess the accuracy of both TextBlob and VADER sentiment tools. TextBlob demonstrated an accuracy of 64.43%, while VADER exhibited a substantially higher accuracy of 76.28%. The low accuracy from the TextBlob suggests potential limitations in capturing the emotional states of peoples, possibly due to its general-purpose nature. Conversely, VADER which is designed for social media applications, has a higher accuracy comparatively. It indicates the proficiency of VADER in deciphering the diverse emotional expression prevalent in the tweets. Moreover, VADER is optimized for analyzing the social media data and can generate good results comparatively when employed with data from twitter, Facebook, etc.

Emotion Classification using Text2Emotion

In the exploration of emotional response to the COVID-19 pandemic in Uttarakhand, text2emotion emerged as a valuable tool. Through its analysis of COVID-19 tweets, people's emotions were systematically categorized, revealing a nuanced distribution. Notably, a substantial 39.4% of the responses were identified as happy, reflecting a positive sentiment amid the challenging circumstances. Conversely, 4.1% responses were detected as anger, underscoring the diverse range of reactions. However, 30.1% of respondents expressed fear, indicating the prevalence of apprehension within the community. Other findings included 11.0% exhibiting surprise, while 15.4% expressed sadness, as shown in figure 3.

DOI: https://doi.org/10.51220/jmr.v19-i2.21

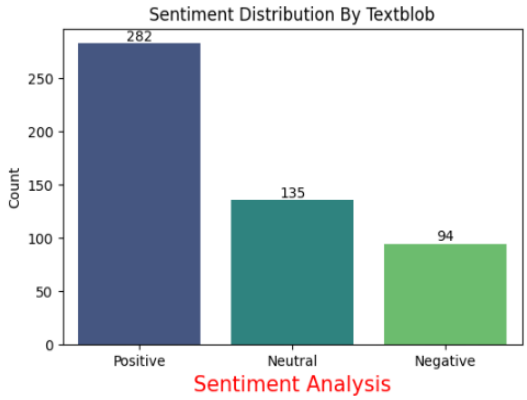

Fig.1: Sentiment Distribution with TextBlob

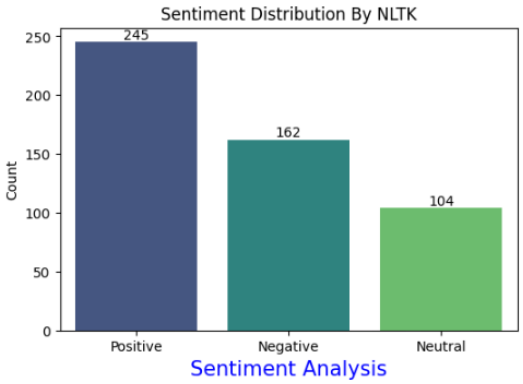

Fig. 2.: Sentiment Distribution with VADER

These insights illuminate the complex emotional landscape surrounding the pandemic in Uttarakhand, providing a valuable data for a comprehensive understanding of public sentiment. The utilization of text2emotion proved instrumental in unveiling intricacies of emotional response, contributing to a more minute comprehension of the social impact of the COVID-19 crisis.

Topic Modeling with LDA

The application of LDA provided us interesting theme which makes good sense to a considerable extent. When LDA is applied on our tweet datasets, it detected themes with the notable keywords for each theme. The results obtained are listed in figure 4. The result comprises ten distinct theme topics (0-9), each characterized by a set of ten keywords. The keywords are shown along with their associated weights. These keywords play a pivotal role in defining the content and focus of each topic.

The key elements influencing the topic 0 include the following top 10 keywords: “covid”, “uttarakhand”, “dehradun”, “curfew”, “pandemic”, “people”, “state”, “till”, “due” and “pm” and so on. These weights serve as indicator of the relevance and importance of each keyword within the context of the topic.

Visualization of Topic Keywords

The interactive visualization of the LDA model is the most effective way to understand the topics and their main words. The visualization consists of two major views, the bar chart on the right and the inter-topic distance map on the left. As depicted in figure 5, each circle in the plot on the left side represents a topic. The area of these topic circles corresponds to the number of words pertaining to each topic. The circles are drawn by applying a multidimensional scaling algorithm on the basis of number of words they encompass, so topics that are close to each other have more words in common.

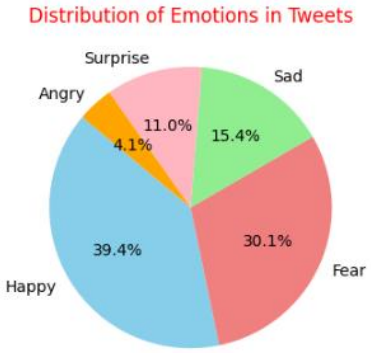

Fig.3: Distribution of emotion in Tweets

The bar chart by default depicts the 30 key terms. The bars represent the gross frequency of the term across the entire collection of texts. Saliency and Relevance, are the two particular metrics defined at the bottom of the bar chart. Saliency can be used to specify most relevant or useful terms for identifying topics in the corpus. Greater values of saliency suggest that a term is more valuable for recognizing a specific topic.

Relevance is used for grading terms within topics. It assists in detecting the most relevant terms within a given topic. Relevance utilizes a parameter called λ (lambda), which can be used to modify the words displayed in the bar chart for a topic by tuning in the slider in the visualization. Setting the values of lambda near to 0 underlines

DOI: https://doi.org/10.51220/jmr.v19-i2.21

probably unusual but more exclusive terms for the chosen topic. Higher lambda values around 1 point out more common terms in the corpus which might not be exclusive to the topic. A user study reveals that a λ value close to 0.6 was found optimal for interpreting the topics, however this value is supposed to vary based on the data and particular topics.

When we choose a topic in the inter-topic distance map, or mention a topic in the top panel, the bar chart adjusts to display the main keywords or terms related to that particular topic. A second darker bar exhibits the topic-linked frequency of terms related to the chosen topic. If the dark bar over shadows the light bar, it conveys that term nearly exclusively belongs to the chosen topic

```
[(0,
  '0.054*"covid" + 0.027*"uttarakhand" + 0.015*"dehradun" + 0.012*"curfew" + 0.010*"pandemic" + 0.008*"people" + 0.007*"state" + 0.007*"till" + 0.006*"due" + 0.005*"pm"'),
 (1,
  '0.034*"covid" + 0.023*"uttarakhand" + 0.008*"coronavirus" + 0.008*"haridwar" + 0.008*"school" + 0.007*"test" + 0.007*"sir" + 0.007*"dont" + 0.006*"close" + 0.006*"dehradu
n"'),
 (2,
  '0.026*"covid" + 0.014*"uttarakhand" + 0.011*"last" + 0.010*"please" + 0.010*"itbp" + 0.010*"days" + 0.010*"dehradun" + 0.008*"nurse" + 0.007*"recover" + 0.006*"modi"'),
 (3,
  '0.039*"covid" + 0.017*"uttarakhand" + 0.011*"case" + 0.010*"dehradun" + 0.008*"kumbh" + 0.007*"state" + 0.007*"school" + 0.007*"india" + 0.006*"new" + 0.006*"haridwar"'),
 (4,
  '0.048*"covid" + 0.024*"uttarakhand" + 0.012*"dehradun" + 0.012*"test" + 0.009*"positive" + 0.008*"children" + 0.007*"school" + 0.006*"due" + 0.006*"days" + 0.006*"indi
a"'),
 (5,
  '0.081*"covid" + 0.066*"uttarakhand" + 0.032*"corona" + 0.028*"dehradun" + 0.018*"new" + 0.017*"find" + 0.016*"case" + 0.015*"coronavirus" + 0.014*"infect" + 0.011*"di
e"'),
 (6,
  '0.063*"covid" + 0.039*"corona" + 0.039*"case" + 0.028*"country" + 0.025*"coronavirus" + 0.024*"new" + 0.021*"coronainindia" + 0.017*"hours" + 0.015*"find" + 0.012*"patien
ts"'),
 (7,
  '0.040*"covid" + 0.027*"uttarakhand" + 0.018*"positive" + 0.014*"rise" + 0.013*"test" + 0.011*"dehradun" + 0.010*"tally" + 0.006*"news" + 0.006*"toll" + 0.005*"due"'),
 (8,
  '0.050*"covid" + 0.018*"uttarakhand" + 0.013*"india" + 0.009*"ambulance" + 0.008*"school" + 0.008*"service" + 0.007*"case" + 0.006*"dehradun" + 0.006*"haridwar" + 0.006*"y
ear"'),
 (9,
  '0.062*"covid" + 0.048*"case" + 0.046*"uttarakhand" + 0.029*"state" + 0.018*"total" + 0.016*"new" + 0.015*"today" + 0.014*"positive" + 0.013*"report" + 0.010*"number"')]
```

Fig. 4: Visualization illustrating the LDA model with 10 topics

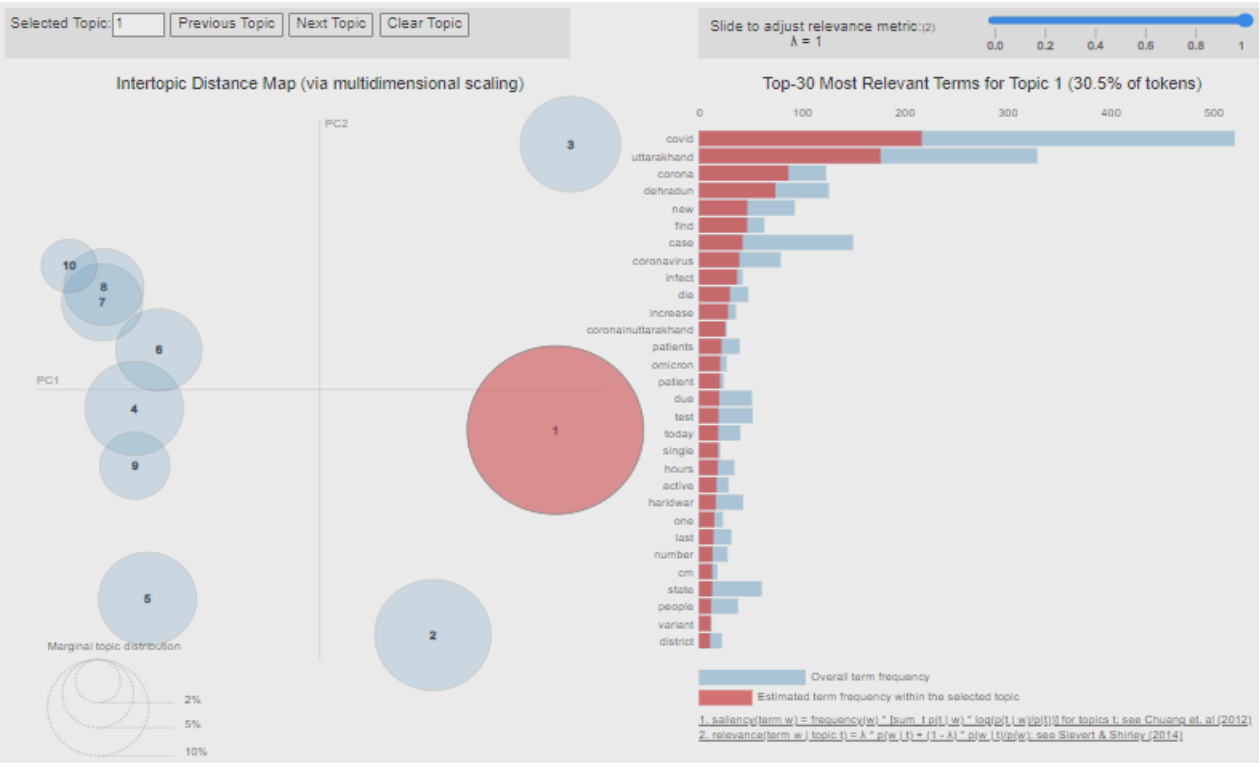

Fig. 5: Interactive Visualization of topic model

DOI: https://doi.org/10.51220/jmr.v19-i2.21

Conclusion

This research analyzes public sentiment towards the COVID-19 pandemic in Uttarakhand, India, using Twitter data. With social media growing as a platform for opinion-sharing, this study leverages tweets tagged with relevant COVID-19 hashtags from Uttarakhand. Sentiment analysis techniques, specifically TextBlob and NLTK's VADER, were used to classify sentiments as positive, negative, or neutral. Results indicate that both tools identified a predominance of positive sentiments, though negative and neutral views were also present. VADER achieved higher accuracy at 76.28% compared to TextBlob's 64.43%, highlighting a difference in effectiveness between the two tools for COVID-19 data. Additionally, emotion classification with the text2emotion tool provided insights into individual emotional states, showing a generally positive outlook among the population during the pandemic. Topic modeling using Latent Dirichlet Allocation (LDA) further enriched the study, uncovering underlying themes within the dataset. This multi-dimensional approach not only reveals sentiment patterns but also captures the nuanced topics shaping public perspectives on the pandemic in Uttarakhand. Understanding the sentiments and mental state of people during the pandemic will help the state authorities to initiate appropriate actions in order to manage the situation and creating awareness among the public towards adopting appropriate safety measures to combat the spread of pandemics.

Acknowledgement

The authors acknowledge the UCOST for providing the fund for this research. We are thankful to all faculty members, staff and research scholars of the Department of Computer Science, Gurukula Kangri (Deemed to be University) Haridwar for their timely support and valuable suggestions.

References

Archana BR (2021) Challenges faced during COVID-19 pandemic. *IP International Journal of Medical Microbiology and Tropical Diseases*. 7(1): 1-2

Chandra R and Krishna A (2021) COVID-19 sentiment analysis via deep learning during the rise of novel cases. PLoS ONE 16(8): e0255615.

Jain Sarika & Yadav M (2021) COVID-19 in India: Problems, Challenges and Strategies (Psychology Aspects). In: Baddour MM(Ed.) *Fighting the COVID-19 Pandemic*. DOI: .

Kaur Manpreet, Verma Rajesh & Ranjan Sandeep (2021) Political Leaders' Communication: A Twitter Sentiment Analysis During Covid-19 Pandemics. *Jurnal The Messenger.* 13(1): 45-62.

Kumar Vaibhav (2022). Spatiotemporal sentiment variation analysis of geotagged COVID-19 tweets from India using a hybrid deep learning model, *Scientific Reports*, 12:1849 | ,

Mir AA, & Sevukan R (2022). Sentiment analysis of Indian Tweets about Covid-19 vaccines. Journal of Information Science, 0(0). DOI:

Nemes L and Kiss A (2020). Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication*, 5(1): 1–15

Singh Mrityunjay, Amit Kumar Jakhar, Shivam Pandey (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Social Network Analysis and Mining*, 11:33,

Singh P (2020) Covhindia: Deep Learning Framework for Sentiment Polarity Detection of COVID-19 Tweets in Hindi. *International Journal on Natural Language Computing* (IJNLC). 9(5): 23-34.

Venigalla Akhila Sri Manasa, Vagavolub Dheeraj, & Chimalakondac Sridhar (2020) Mood of India During Covid-19 - An Interactive Web Portal Based on Emotion Analysis of Twitter Data. arXiv:2005.02955.