The Effect of A Natural Catastrophe on Election's Outcomes: A Psychic Named Twitter

By Ehsan Asghari-Ghara and Abu Shonchoy*

Draft: October 22, 2021

Voter's myopia theory claims that voters form their opinion more on the latest events than earlier events; for instance, the most recent government's response to a natural disaster has a crucial role in an upcoming election. In this paper, we study how the Indian government's response to cyclone Fani affected the parliamentary election in 2019 based on the myopia theory. We built a theoretical model explaining the behavior of politicians in the presence of myopic voters. Unlike other studies, we used sentiment analysis rather than public spending to calculate voters' attitudes towards the government's response. Our result shows that (1) tweets sent in areas hit by Fani have a higher sentiment than tweets posted outside of Fani's path, (2) the probability of winning the election increases for the ruling party after Fani, (3) an increase in sentiment leads to a rise in the likelihood of getting reelected in areas that parliament members are from the ruling government's party.
Keywords: Myopic Voter, Sentiment Analysis, Cyclone Fani

Elections are the bedrock of democracies, and political parties benefit from it to gain power and serve people. The political structure, politicians' accountability, and the interaction between politicians and citizens have meaningful impacts on

* Asghari-Ghara: DePauw University, 205 Harrison Hall, 7 E. Larabee St., Greencastle, IN 46135 , ehsanasghari-ghara@depauw.edu.
Shonchoy: Florida International University,11200 SW 8th Street, Miami, Florida (33199), shonchoy@fiu.edu.

society's general welfare. Therefore, elections and voters' preferences have long been studied in political science and economics.

Levitt (1996) estimated the weight of different inputs in a senator's utility function. The result shows that only one-quarter of the total weight is assigned to voters' preference, while a significant weight is assigned to the senator's ideology. Pagano and Volpin (2005) analyzed the effect of election structure on shareholders' protection and employment protection. They predicted that proportional electoral systems provide weaker investor protection but more robust employment protection than majoritarian systems. Oswald and Powdthavee (2010) used nationally representative longitudinal data to show the effect of a child's gender on parents' voting decisions. They concluded that having daughters makes people more likely to vote for left-wing political parties, and having sons leads people to favor right-wing parties.

Studies have recommended few initiatives to improve politicians' accountability. One example is the use of citizen scorecards, whose efficiency was analyzed by conducting a multi-year field experiment involving 408 politicians in 20 Ugandan district governments between the 2011 and 2016 elections (e.g. Grossman and Michelitch 2017, Humphreys and Weinstein 2012). It was found that politicians' performance increased in competitive constituencies due to these scorecards.

Many researchers have analyzed how voters form their opinions, beliefs, and preferences during an election and studied what external factors affect their decisions. Golovchenko et al. (2020) investigated the role of propaganda and Internet Research Agency (IRA)—Russian trolls— during the 2016 US election by analyzing the hyperlinks contained in 108,781 relevant tweets. They concluded that the IRA actively and disproportionately promoted Conservatives campaigns. Furthermore, links to YouTube videos supported the Conservatives. In general, the propaganda campaign was active on both sides; however, the evidence verifies the IRA's support of the Republican campaign. Researchers have also utilized other social media for this purpose. Tella, Galiani and Schargrodsky (2019) also

explored the effect of the Argentine government's propaganda campaign against apposition before the 2015 presidential election. They showed the treatment group members an advertisement aired by the government against the opposition candidate and then asked them about their political views. This group's preference declined by 6.5 percentage points, driven mostly by women, compared to the control group. Interestingly, the opposition's efforts in response to this campaign did not affect the manipulated political view. Stephens-Davidowitz (2012) used Google search queries and found that racially charged language is a negative predictor of black candidates' vote.

One leading hypothesis is that of voters' myopia, which claims that voters form their opinion based more on the latest events than on earlier events or their future consequences. Stigler (1973) showed that the underlying economic conditions do not play an essential role in national elections. Specifically, fluctuations in real income do not affect the election's outcomes. Fournier et al. (2004) distinguished between individuals who make up their minds before the campaign launch (based on partisan identification and ideology) and individuals who decide to respond to campaign events and messages. The latter group forms a significant part of the population; for instance, they represent roughly half of the Canadian electorate. The authors argue that campaign deciders' decisions are strongly affected by campaign events such as leaders' debates and media coverage. Wlezien (2015) argues that voters are generally myopic; however, they do look reasonably far back, about two years, while evaluating the US presidents' economic sector performance.

Some of the studies on voters' myopia analyze the government's response to an event, usually a natural disaster, and its effect on the upcoming election. Healy and Malhotra (2009) show that governments have no incentive to invest in natural disaster preparedness, even though investment will substantially reduce future disaster damage. They point to myopic voters as the culprit, along with voters who do not care about preventative initiatives. This behavior discourages politicians from adopting preventive measures and incentivizes them to provide relief

funds after the natural disaster. The authors estimate that $1 spent on preparedness is worth about $15 regarding the future damage it alleviates. Bechtel and Hainmueller (2011) also investigated the government's response to a natural disaster, the 2002 Elbe flooding in Germany. Their results reject the idea of myopic voters and claim that an efficient response generates long-lasting voter gratitude. In their sample, the flood response increased vote shares for the incumbent party by seven percentage points in affected areas in the 2002 election. Twenty-five percent of this short-term reward was carried over to the 2005 election before the gains vanished in the 2009 election.

Many of these studies use government spending as a proxy for the response's efficiency and voters' satisfaction. However, this method raises many concerns. For instance, federal spending is usually not separable from states' or local governments' spending (as addressed by Healy and Malhotra 2009). High expenditure is attributed to the federal government's response; however, voters may reward state politicians for better performance. Moreover, a high level of relief funds does not imply a better response to a natural disaster. Voters react to the efficiency of the government's response and not to their level of spending. It is essential to consider how efficiently the government spends the relief funds. A better approach is to directly measure the satisfaction of voters with the government's response. Therefore, we suggest using sentiment analysis to investigate voters' sentiment toward government responses.

Some theoretical works have investigated elections, voting patterns, and the economic effects of political structures. Cho and Duggan (2009) provided game-theoretic foundations for the median voter theorem's predictions by analyzing committee decision-making with a non-cooperative, infinite horizon bargaining model based on a random recognition rule and majority voting. Caplin and Nalebuff (1991) extended the median voter model by allowing candidates to differ in more than one dimension. They reported that under specified conditions, the social choice follows the mean voter's preferred outcome.

Other studies provide empirical evidence for economic and political theories. Scervini (2012) claimed that the data do not support the median voter theorem. They showed that the link between income and redistribution is weaker for the middle class than for any other income class, and the amount of redistribution targeting the middle class is lower in more asymmetric societies. The findings of Levitt's (1996) also contradicted the empirical results supporting the median voter theorem. This is also the case for the myopic voters' theorem. Some researchers have provided evidence in favor of the theorem (e.g. Healy and Malhotra 2009, Stigler 1973), some have rejected it (e.g. Bechtel and Hainmueller 2011), and some have provided mixed evidence (e.g. Fournier et al. 2004, Wlezien 2015).

To contribute to this literature, we use a unique study setting to understand the myopic voters' phenomenon. We benefit from a rare concurrence of events in India in May 2019. There was a general election at the time coinciding with an intense hurricane hitting India (a robust but scarce, scientific and natural design). We use this setting to investigate the government's response to the natural disaster, cyclone Fani, in India and its effect on the upcoming parliamentary election of 2019. First, we find how the government's response to a natural disaster is perceived by voters and, further, we investigate how it affects the election's outcomes. Unlike other studies exploring the government's response to natural disasters, we use voters' sentiments rather than public spending to mirror actual voters' views and opinions about the government's response.

We draw on existing literature to calculate tweets' sentiment. Many campaigns utilize scientific tools such as natural language processing (NLP) to evaluate voters' sentiments and behaviors (e.g. Tumasjan et al. 2010, Williams and Gulati 2008). Researchers have also used Twitter for a better understanding of voters' opinions and sentiments. Digrazia et al. (2013) investigated the US congressional elections and concluded that as Twitter users mention Republican candidates' names more in their tweets, the candidates' vote margin increases in subsequent elections. Tumasjan et al. (2010) used Twitter data to determine

whether the sentiment reflected in tweets is in line with offline political sentiment. Their study of the German federal election shows that Twitter is extensively used to express political opinions and reflects true political sentiment.

Our dataset includes more than a million tweets posted in our sample period. We selected Twitter, as it is a primary source of information about voters' thoughts and sentiments. Twitter users post texts, photos, and videos reflecting their opinion on national discourses, their sentiment at the time of disaster or celebration, and their urgent needs at the time of emergency (e.g. Cheong and Cheong 2011, Mandel et al. 2012). Many scientists have developed tools to analyze these tweets, extracting useful information to optimize first response and disaster relief. For instance, Ashktorab et al. (2014) developed a Twitter mining tool using classification, clustering, and extraction techniques to provide first responders with relevant information. Cheong and Cheong (2011) identified meaningful clusters during the 2010–2011 floods in Australia. Moreover, Kumar et al. (2011) used geo-location analysis and tweet keywords to help first responders when a disaster transpires. Some researchers have investigated the welfare effect of social media. Other social media platforms have also been used to extract valuable information about users. For instance, Allcott et al. (2020) carried out a randomized controlled trial to evaluate the welfare effect of Facebook users in the US before the 2018 midterm election and conclude that Facebook deactivation results in socializing with family and friends, reduces political pluralization, and increases subjective well-being.

I. Election structure in India and Cyclone Fani

The Lok Sabha election is a nationwide general election of India. The election result determines who sits in India's lower house of Parliament, the Lok Sabha, or the House of the People. The first-ever general election was held in 1952, and the general election of 2019 was the 17th Lok Sabha election of India. The Lok Sabha has 545 seats; 543 of them are directly elected by different parliamentary

constituency (PS) regions across India. The president nominates two additional members of parliament from the Anglo-Indian community to be seated in the Lok Sabha. The winning party or a coalition of leading parties will select the prime minister, and in turn, the prime minister will choose the cabinet members.

India's electoral law mandates the presence of a polling station within two kilometers of every habitation. This guarantees voting opportunity for every Indian. All votes in the general election of 2019 were cast electronically, and 1.72 million machines were used to that end. The 2019 election required many efforts and planning, making it a costly election. Indeed, India has the world's second most expensive elections after the United States (India Elects 2019: The World's Largest Election (2019)).

In the general election of 2019, the two main competing political parties were the Bharatiya Janata Party (BJP), ruled by the incumbent prime minister, Narendra Modi, and the Indian National Congress (INC) led by Rahul Gandhi. Other participating parties are mostly regional or caste-based parties, campaigning in a specific region.

A. Election Timeline

The general election took place over several weeks to ensure that any of the 900 million eligible voters had a chance to cast a vote securely. The entire process of the 2019 Lok Sabha election spanned from March to May. Table 1 shows the timeline of the election's administration process, managed by the Election Commission of India. It took about ten weeks for the whole process to get accomplished, while the polling started just about the sixth week.

Each state comprises several parliamentary constituencies (PCs), and each PC represents a seat in Lok Sabha. These PCs have different polling dates, as announced by the Election Commission of India. For instance, Madhya Pradesh has 29 PCs, and they participated in 4 separate phases of the election (Figure 1).

Moreover, each PC comprises several Assembly constituencies (ACs), but all

Table 1—Timeline of the election's administration process.

	Administration Process
Week 1	The Election Commission of India announces election dates.
Week 2	The Election Commission of India formally notifies the elections.
Week 3	Nominations of candidates are accepted.
Week 4	Official campaigning commences.
Week 6	Polling starts for the first phase.
	The process repeats on a staggered basis for subsequent phases.
Week 10	Votes are counted.
	The results are usually announced within hours.

Note: The table represents an approximation for administration process timeline announced by the election commission.
Source: India Elects 2019: The World's Largest Election (2019).

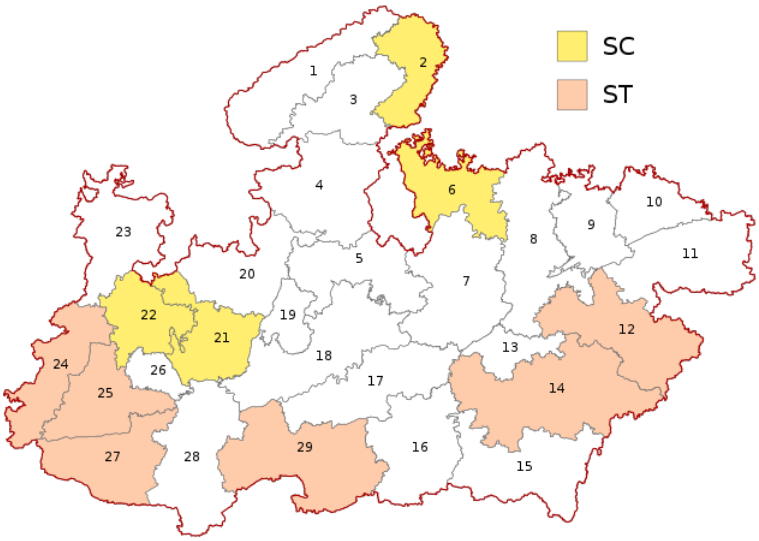

Figure 1. Lok Sabha constituencies in Madhya Pradesh.

Note: SC means seat is reserved for scheduled castes, and ST means seat is reserved for scheduled tribes.
Source: wiki/User:Furfur (2014*a*).

the ACs located in the same PC have the same election dates. For instance, the state of Madhya Pradesh has 230 ACs (Figure 2). The election results of these ACs will get aggregated at the PC level to determine the winning party

representing that PC.

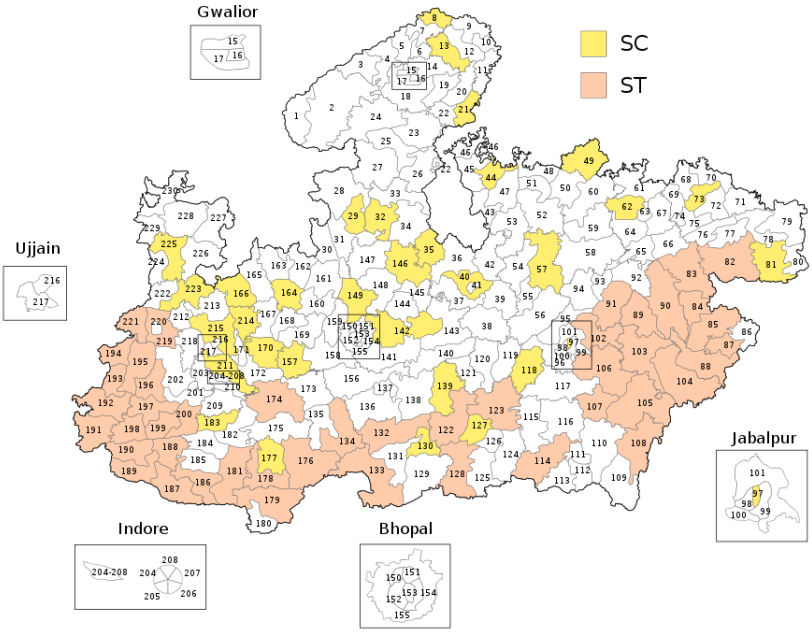

Figure 2. Assembly constituencies in Madhya Pradesh.

Note: SC means seat is reserved for scheduled castes, and ST means seat is reserved for scheduled tribes.
Source: wiki/User:Furfur (2014*b*).

The general election comprises seven phases (Table 2 and Figure 3) spanning from Weeks 6 to 10. Each parliamentary constituency had a specific polling date, and hence all assembly constituencies in a specific parliamentary constituency shared the same polling date. Table 2 provides information about polling dates, and the number of constituencies participated in each phase. A comprehensive list of constituencies and their polling dates is available in the appendix.

Figure 3 visualizes the seven phases of the 2019 Lok Sabha election for different states and their parliamentary constituencies.

Table 2—Seven phases of the election.

Phase	Polling Date
Phase 1	April 11, 91 parliamentary constituencies participated.
Phase 2	April 18, 95 parliamentary constituencies participated.
Phase 3	April 23, 116 1/3 parliamentary constituencies participated.*
Phase 4	April 29, 71 1/3 parliamentary constituencies participated.*
Phase 5	May 6, 50 1/3 parliamentary constituencies participated.*
Phase 6	May 12, 59 parliamentary constituencies participated.
Phase 7	May 19, 59 parliamentary constituencies participated.

Note: * Polling in Anantnag was scheduled over three days.
Source: India Elects 2019: The World's Largest Election (2019).

B. Cyclone Fani

Cyclone Fani originated from a tropical depression that formed in the west of Sumatra in the Indian Ocean on April 26. The storm escalated to a deep depression on April 27. The Joint Typhoon Warning Center (JTWC) upgraded Fani to a Category 1-equivalent cyclone late on April 29. On April 30, Fani was upgraded to a very severe cyclonic storm by the Indian Meteorological Department (IMD). Moreover, on this date, the JTWC upgraded the storm to a Category 3-equivalent cyclone.

Fani rapidly intensified into a severe cyclonic storm and reached its peak intensity on May 2 as a high-end Category 4 major hurricane. Shortly after, Fani started another rapid intensification period and almost received a Category 5-equivalent tropical cyclone intensity, according to the JTWC.

On May 3, Fani made landfall near Puri, Odisha, as a severe cyclonic storm. Land interaction quickly degraded Fani's convective structure, and it weakened to a Category 1-equivalent tropical cyclone soon after landfall.

On May 4, Fani weakened to a deep depression and moved into Bangladesh, and its convective structure rapidly degraded after that, degenerating into a remnant low on May 4. On May 5, Fani's remnant low dissipated over Bhutan.

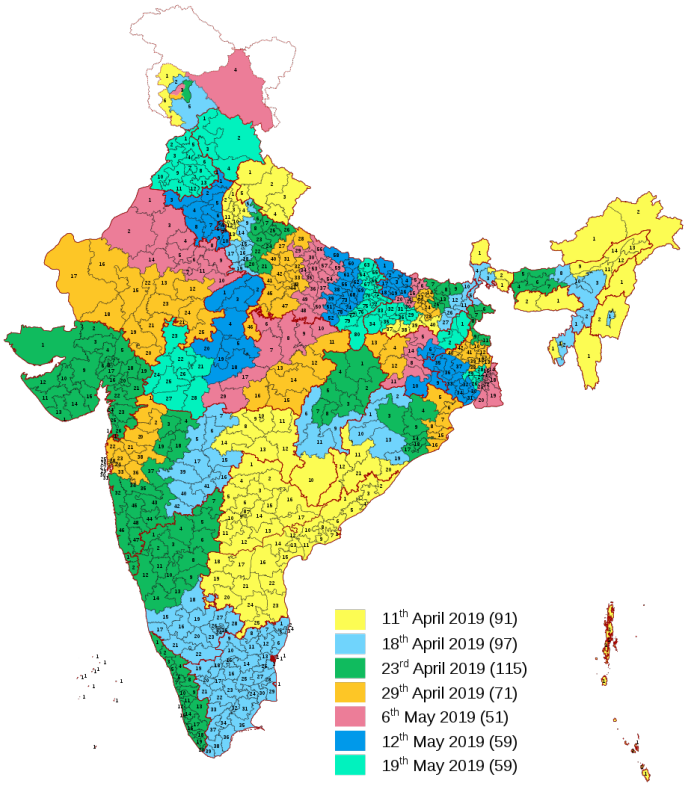

Figure 3. Schedule of Lok Sabha Election, 2019.

Source: wiki/user:furfur (2019).

C. *Timeline of Events*

Figure 4 shows the event timeline, plotting the election's polling dates along with the events related to Fani. The first phase of India's general election started on April 11, and in the following weeks, two other phases took place on April 18 and April 23. However, before the fourth phase of the election, cyclone Fani originated in the Indian Ocean. The fourth phase of the election occurred on April 29, while cyclone Fani continued to become more severe and finally reached its peak on May 2.

On May 3, Fani made its landfall causing damage to infrastructure and human lives. It caused at least 89 deaths and approximately 8.1 billion dollars of loss

in both India and Bangladesh (Global Catastrophe Recap: First Half of 2019, 2019). Cyclone Fani continued its path through Bangladesh and dissipated over Bhutan on May 4. After the dissipation of Fani, the remaining phases of the general election took place on the polling dates of May 6, May 12, and May 19.

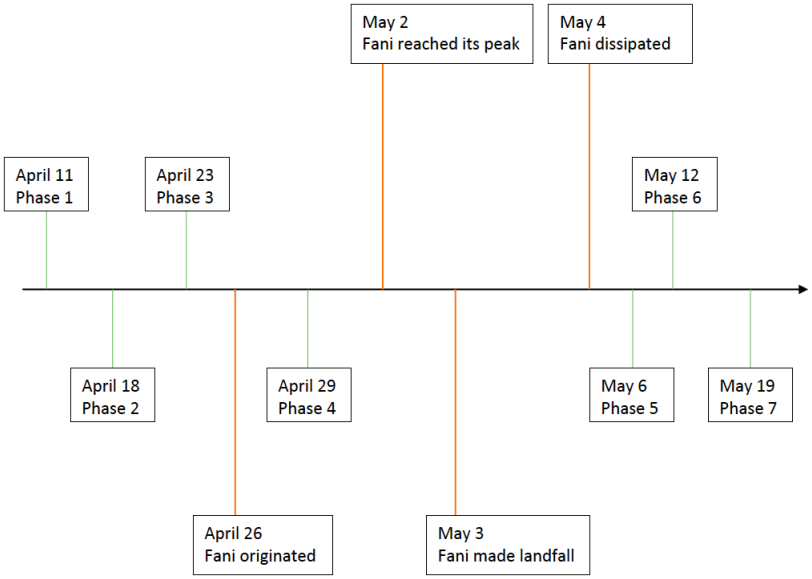

FIGURE 4. TIMELINE OF EVENTS.

Figure 5 shows Cyclone Fani's path and regions with the same polling dates. The polling phases of all PC units located around the path of Fani seems reasonably diversified.

II. Data

We began collecting data from May 2 to June 9, 2019, and collected over a million tweets after filtering for English tweets. We restricted our dataset to tweets posted in the English language as NLP tools are not available in many other languages. Furthermore, English is an official language in India. The data were collected based on geolocation criteria, that is, we collected tweets sent

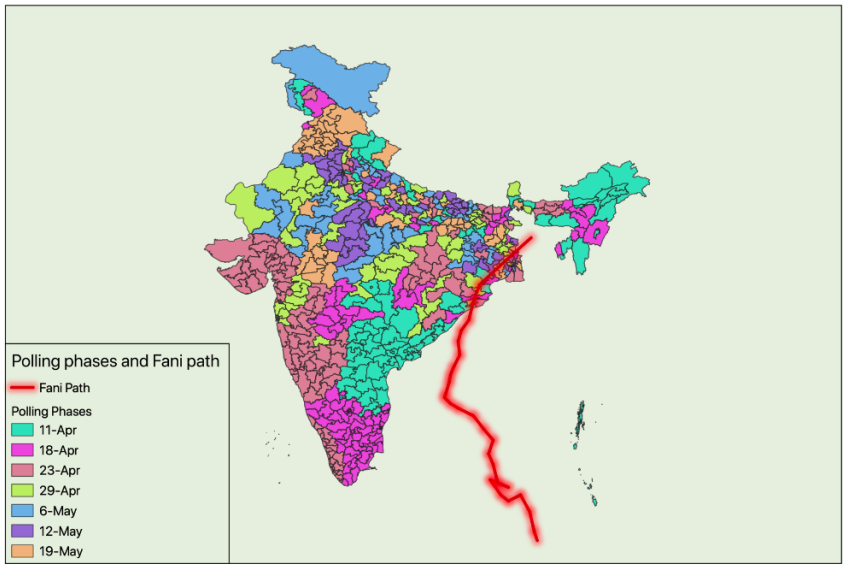

Figure 5. Polling phases and cyclone Fani's path.

within a geographical rectangular area encompassing India's boundaries. In the next step, we narrowed the data to those tweets sent within the country.

There are some limitations to tracking a user's location due to privacy protection policies enforced by Twitter. One can track the location only if the user opts for location tracking or if the user has manually added the location to the tweet. The former provides precise location tracking in the form of a GEOJson object called a point, while the latter offers a GEOJson object called a polygon representing a wider area, such as a city. We assigned a point object to the tweets with general geolocation data by calculating the polygon's centroid. Figure 6 depicts the geographical diversity of tweets in our dataset. Tweets are posted all around India, and hence applying the sentiment analysis with our sample will accurately measure Indians' general sentiment.

Figure 7 shows the percentage of tweets posted in the top ten states with the most frequencies after applying the aforementioned filters. The state of Maharashtra has the most-posted tweets in our sample, followed by Delhi and Karnataka.

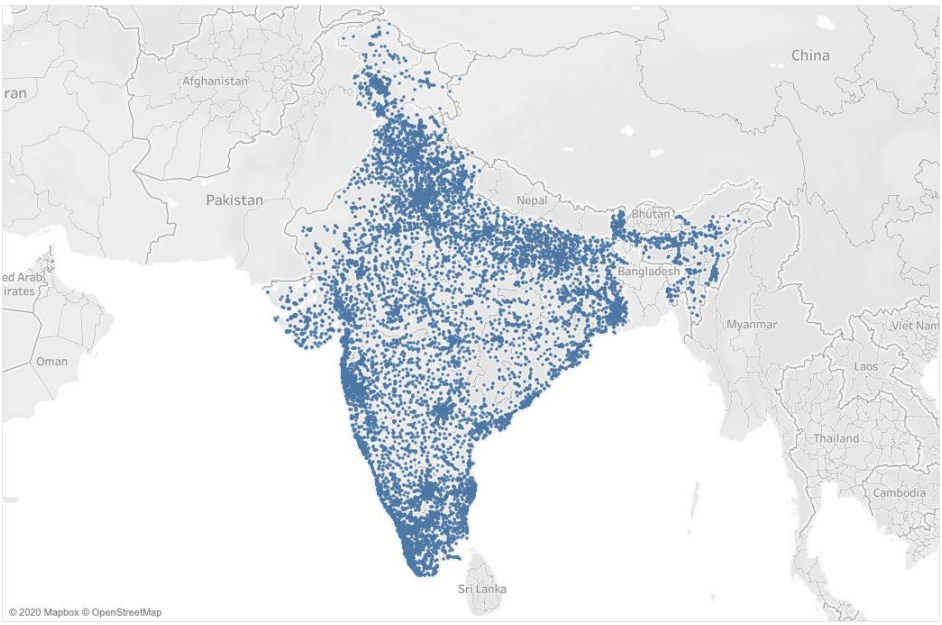

Figure 6. The geographical diversity of tweets.

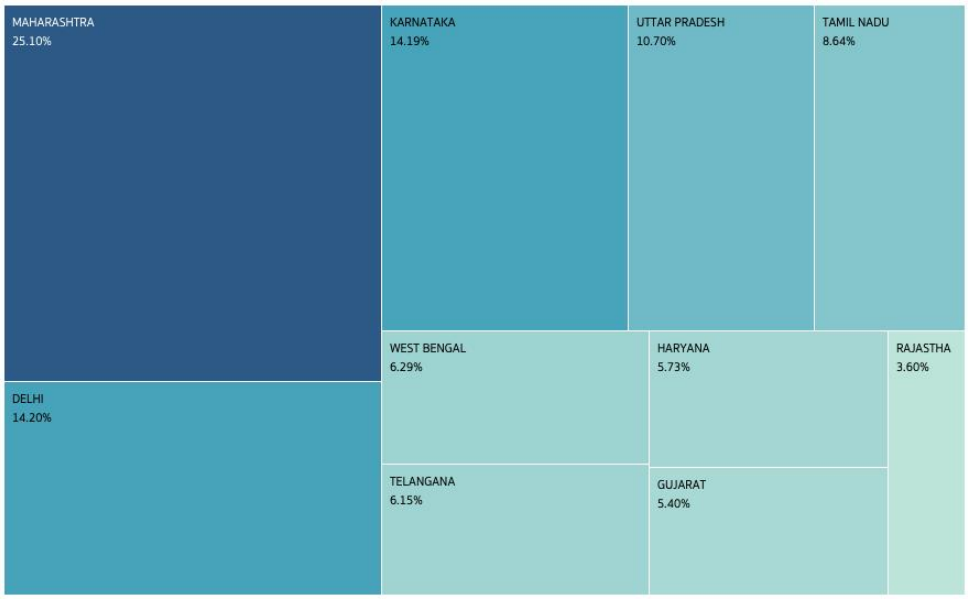

Figure 7. Percentage of tweets in different states (Top 10 states with most frequency).

Figure 8 shows the percentage of tweets posted in the top ten PCs with the most frequencies after applying the filters. The parliamentary constituency of Mumbai North-West in the state of Maharashtra has the most-posted tweets in our sample, followed by West Delhi in Delhi and Bangalore South in Karnataka.

Figure 8. Percentage of tweets posted in the top ten PC areas with the most frequencies.

Figure 9 shows the percentage of tweets posted in the top ten AC areas with the most frequencies after applying the filters. It follows the same pattern as Figure 7 and 8. The most posted tweets are from the states of Maharashtra, Delhi, and Karnataka.

We use word clouds to illustrate the most frequently used words among our sample of twitter users. This will give us a general idea about what Indian Twitter users in our sample talked about. The word cloud for 300 frequently used words is shown in Figure 10. According to the word cloud, some of the most used words by Twitter users are Fani, cyclone, Odisha (landfall place), India, time, Modi (Prime Minister), and thank. In the next section, we use the bag of

Figure 9. Percentage of tweets posted in the top ten AC areas with the most frequencies.

words to analyze the most discussed topics and tweets' sentiment in the sample.

Figure 10. Word cloud of 300 frequently used words.

III. Text Mining and NLP

Traditionally, people use cable news and newspapers to collect information about their surrounding areas. National discourses were also triggered by these means of information propagation. However, after the advent of the Internet and specifically in recent years, social media has become the most critical communication tool. These social networks provide eyewitness accounts of local events and make them a subject of discussion in different areas worldwide. They reflect both opinions and sentiments and can mobilize people in response to various events and crises.

Social networks can be used to analyze what people think and feel. They can provide insights into different communities and cultures. Twitter is an exceptional case among all the social media platforms. Twitter has been used to propagate real-time information at the time of crises and political upheavals. For instance, it played a significant role in the aftermath of Iran's 2009 presidential election and its Green Movement. However, Twitter is a microblogging platform that offers limited use of characters to communicate. Each tweet is a very concise piece of information, and this feature distinguishes it from other conventional social media platforms.

In this section, we use the main methods in text mining to analyze our sample data. We use topic modeling to find relevant topics in the data and sentiment analysis to explore people's emotions while struggling with a natural disaster like Cyclone Fani.

A. Topic Modeling

Topic modeling is a combination of methods to find the latent patterns in text content. Many kinds of research benefited from topic modeling in their work. Weng et al. (2010) identified the influential users on Twitter by considering both the topic similarity between users (with the help of topic modeling) and the link structure.

Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan (2003)) is one of the most utilized topic modeling methods. The extended forms of LDA have specifically become the primary standard tools for modeling the latent text patterns in social media.

In this study, we use LDA to extract the main topics discussed in our dataset. However, some limitations exist when using LDA. Twitter is a microblogging platform that offers limited use of characters to communicate. The platform initially offered 140 characters per tweet, recently extending it to 280 characters. This feature makes it difficult to discuss many topics in a tweet. Users circumvent this limitation by improvisation, including using emoticons and shortened URLs. They also present their ideas very briefly through the use of hashtags. These limitations may question the efficiency of applying LDA to the Twitter dataset. However, Hong and Davison (2010) show that the topic modeling approaches can be helpful for short text and can be applied to platforms like Twitter or other similar structures like blog comments. They suggest that the effectiveness of these methods can be improved by aggregating short messages.

Latent Dirichlet Allocation (LDA)

LDA is an unsupervised machine learning technique to identify predefined numbers of latent text patterns (topics) in a document. The LDA method assumes that there exist fixed numbers of topics, and each topic is a specific distribution of words. Hence, each document (in our case, each tweet) is described as a distribution of topics, and each topic is defined with a distribution of words. However, we only observe the documents and the words while the topics are initially latent.

LDA does not consider the order of words in a document. It uses the "bag of words" approach to represent a document, that is, each document is a vector of word counts. This means that LDA assumes that one can identify the same topics in both ordered and unordered sets of words.

Formally, LDA assumes that the document generation process follows four steps

(Lettier (2018)).

- Each document in the corpus has a multinomial distribution over T topics.
- Each topic also has a multinomial distribution over words.
- The Bayesian prior is assumed to be the Dirichlet distribution. That is, both document and topic distributions have Dirichlet prior with hyperparameters α and β, respectively.

  Each document is a sampling mixture of topics drawn from Dirichlet distribution, and each topic is a sampling mixture of words drawn from the Dirichlet distribution. The sampling is initialized with two predefined hyperparameters: alpha and beta. Dirichlet distribution takes alpha for each topic (every topic in the document is assigned with the same alpha) and takes beta for each word (every word in the topic is set with the same beta). The alpha hyperparameter controls the distribution of topics for each document. The lower alpha implies fewer topics discussed in each document. Moreover, the beta hyperparameter controls the distribution of words for each topic. The lower the beta, the fewer words are likely to appear in each topic.

- For each word in one document d, a topic "z" is sampled from the document's multinomial distribution associated with the document d, and a word "w" is sampled from the multinomial distribution associated with topic "z." This generative process is repeated Nd times, where Nd is the total number of words in document d.

Using LDA, we identified three major topics in our sample: *"Infrastructure," "Religion,"* and *"Cyclone Fani."* These topics are closely related to the occurrence of events in our sample period. Note that we label these topics based on the distribution of words in each topic.

Figure 11 shows the distribution of words for "Infrastructure." Some of the most frequent words in this topic are service, #property, #residential, #forsale, road, water, and bank.

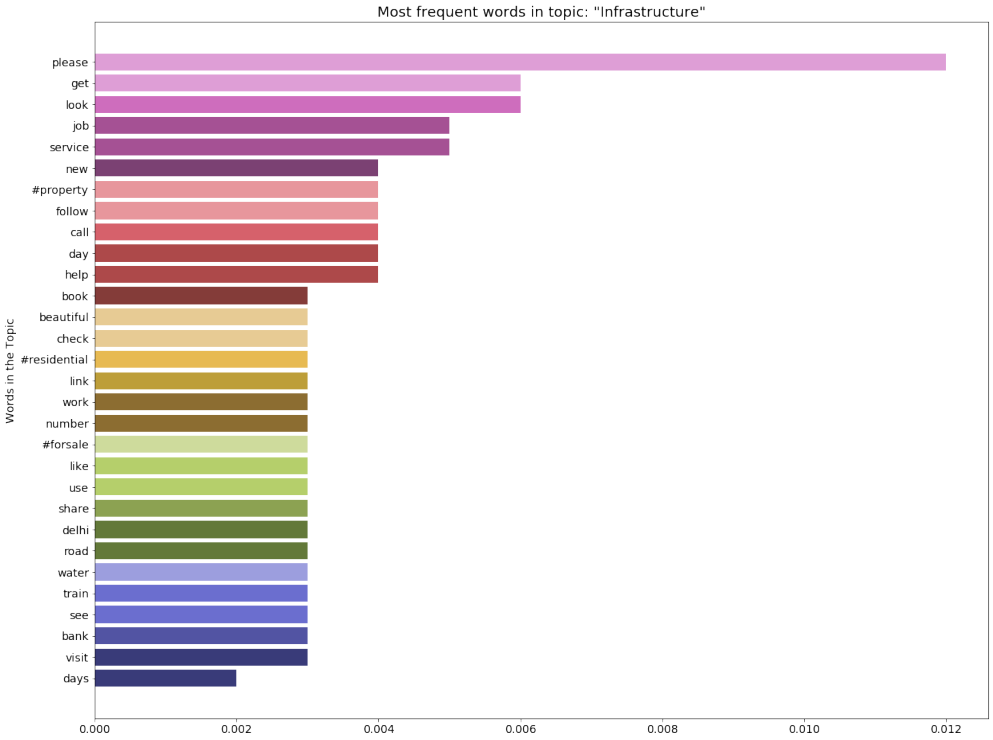

Figure 11. Topic of "Infrastructure" and its most frequent words.

Figure 12presents the distribution of words for "Religion." It shows that people were also discussing religion in our sample period, and some of the most frequent words in this topic are God, Kabir, #lordkabirji, and love. This might be expected given the incidence of major events in our sample period, including the election and hurricane. It is also in line with the findings of (Sinding Bentzen (2019)), who found that individuals become more religious when an earthquake hits, specifically in districts rarely hit by earthquakes.

The next discussed topic is "Cyclone Fani." Figure 13 demonstrates the distribution of words on this topic. Some of the most frequent words in this topic

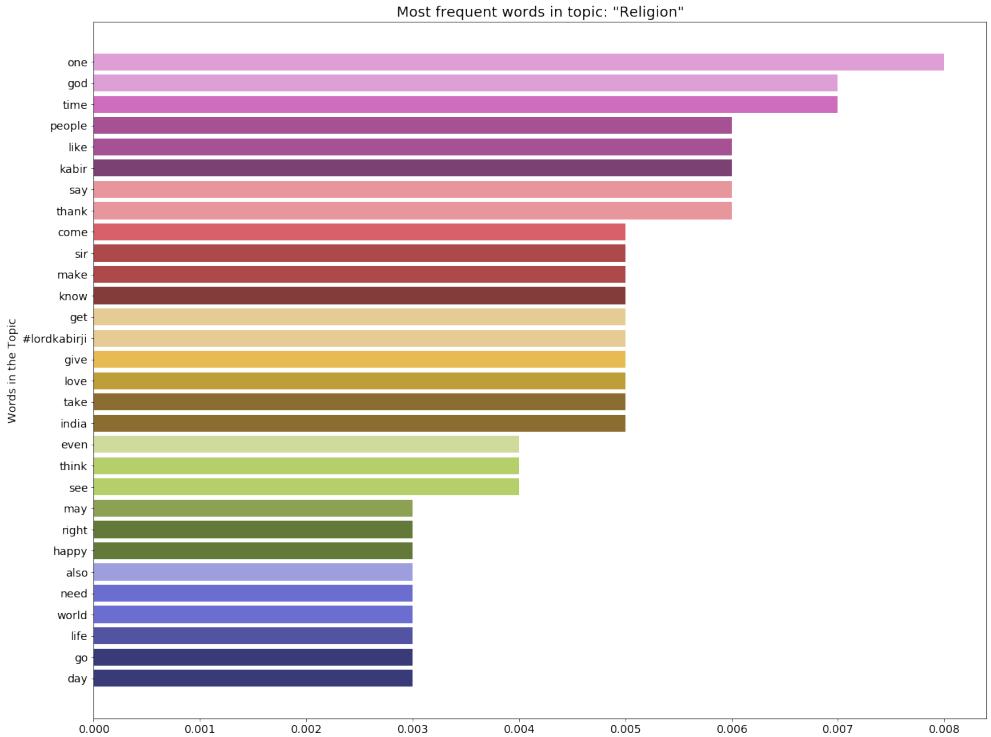

FIGURE 12. TOPIC OF "RELIGION" AND ITS MOST FREQUENT WORDS.

are India, Fani, government, Odisha, Modi, hit, and cyclone. This reflects the idea that people use social media, in this case, Twitter, to articulate their opinion and to show their sentiment. Thus, the building block of our hypothesis is that the documents (in our case, the tweets) extracted from social media at the time of a catastrophe may reveal the local government's competency and voting preferences.

In the next section, we analyze tweets' sentiment to determine whether there is a change in sentiment when the cyclone Fani hit India.

B. Sentiment Analysis

Many studies use sentiment analysis to identify the sentiment of social media users during a disaster. Nagy and Stamberger (2012) improved the existing

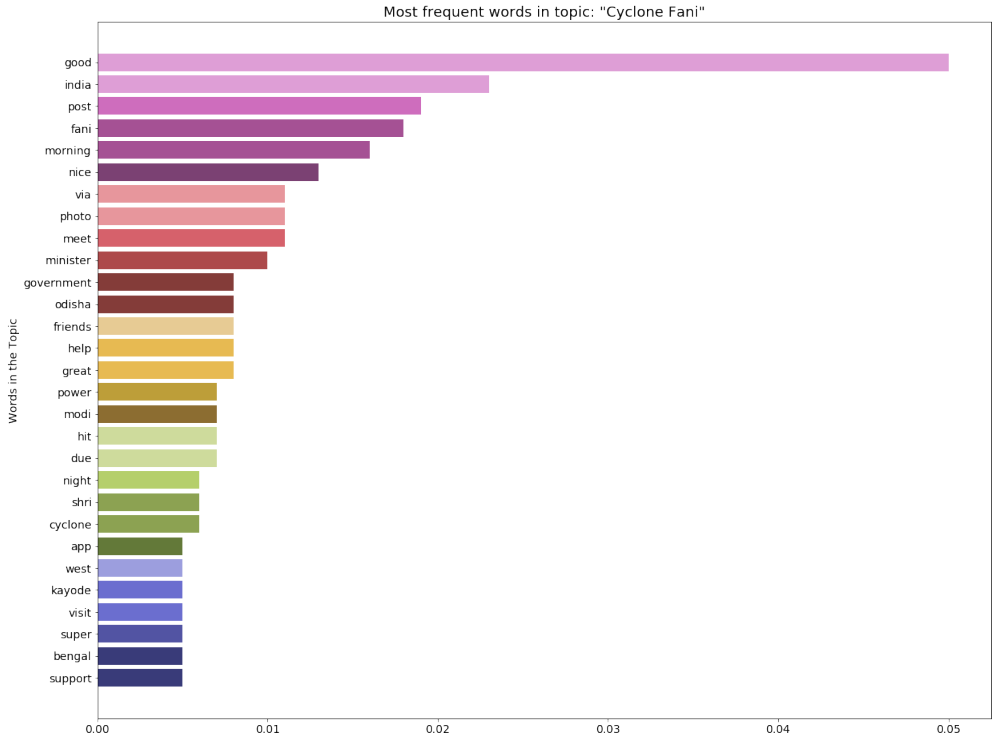

FIGURE 13. TOPIC OF "CYCLONE FANI" AND ITS MOST FREQUENT WORDS.

sentiment analysis technique to identify Twitter users' sentiment during the San Bruno gas explosion. Neppalli et al. (2017) studied the sentiment of Twitter users during hurricane Sandy and showed that users' sentiments change based on the distance from the disaster. They also suggested that sentiment analysis may help emergency responders develop situational awareness of the disaster zone. Ragini, Anand and Bhaskar (2018) presented a big data–driven approach through sentiment analysis to analyze the needs of people during disasters. In this study, we use the VADER algorithm to detect sentiment during the cyclone Fani disaster in India.

VADER algorithm

The Valence Aware Dictionary for Sentiment Reasoning (VADER) is a specific model of sentiment analysis that is useful for identifying both polarity (positive or negative) and intensity of sentiment (. Gilbert and Hutto 2014) The method is explicitly tailored for analyzing social media sentiments, although it is also applicable to other domains.

VADER uses a dictionary that assigns a sentiment score (sentiment intensity) to each lexical feature. Furthermore, the sentiment score of a text is calculated by summing up the sentiment score of its components. The dictionary is constructed by the following process:

Over 9,000 token features were given to 10 independent human raters who were pre-screened and trained for optimal inter-rater reliability. These tokens are rated on a scale from -4 (extremely negative) to +4 (extremely positive) with an allowance for 0 (neutral). We retained any token feature with a non-zero mean rating and a standard deviation of less than 2.5, as determined by the aggregate of those ten independent human raters. Applying these filters, they were left with 7,500 lexical features that reflect both polarity (negativity/positivity) and intensity (on the scale of -4 to +4).

The final dictionary includes lexicon sentiment for these types of features:

- a full list of Western-style emoticons, for example, :-) denotes a smiling face and has a positive sentiment.
- Sentiment-related acronyms, for example, "LOL."
- Commonly used slang with sentiment value, for example, "nah" and "meh."

In the next step, a compound score is calculated by adding each token's intensity score, adjusting it according to pre-specified rules, and then normalizing it to a score between -1 (extremely negative) and +1 (extremely positive). For classification purposes, researchers use the following scheme:

- If the compound score is equal to or greater than +0.05, the text is classified as "positive sentiment."
- If the compound score is equal or less than -0.05, the text is classified as "negative sentiment."
- If the compound score is between -0.05 and +0.05, the text is classified as "neutral sentiment."

Figure 14 shows a pie chart of the sentiment in our entire dataset. Based on our sentiment analysis, about 48 percent of the tweets reflect a positive sentiment, while 21 percent demonstrate a negative sentiment.

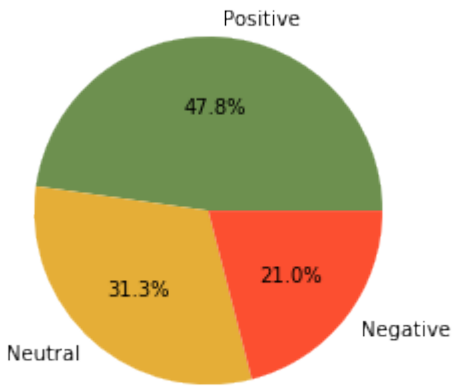

Figure 14. General sentiment.

A comparison of the sentiment before and after the cyclone Fani would be more informative. Figure 15 shows the polarity pattern of users' sentiment before and after the Fani cyclone in both the control and treatment groups. We also filtered our sample to include tweets posted by users expecting an upcoming election. Thus, the new sample is formed by users that may affect election outcomes.

The pie chart shows that the sentiment pattern is almost the same in the control group. However, in the treatment group, the positive sentiment decreases after

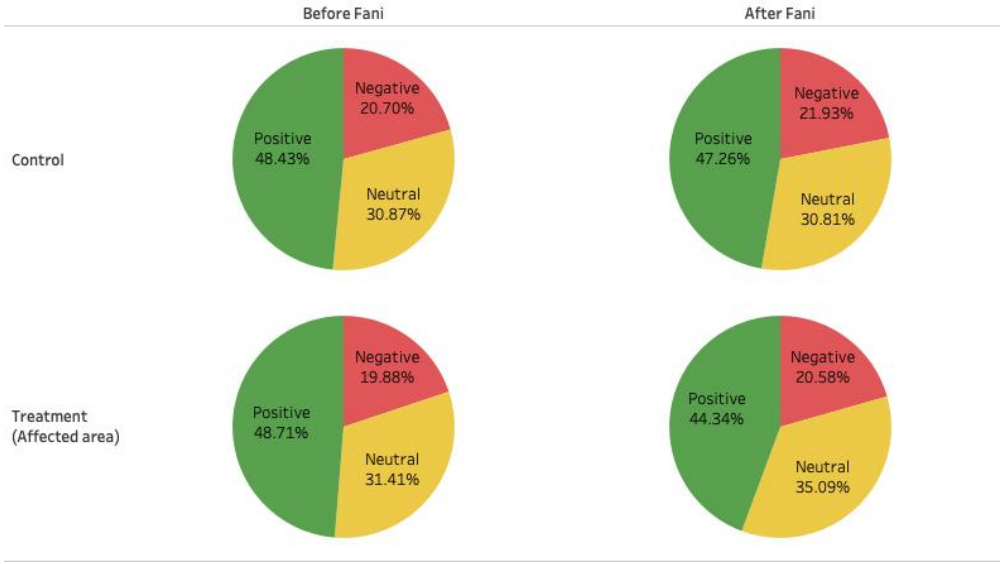

Figure 15. Sentiment before and after the Fani in both the control and treatment groups.

Fani, while the neutral and negative sentiment increases.

Figure 16 demonstrates the change in average sentiment over time. Even though the sentiment fluctuates, the trendline does not change.

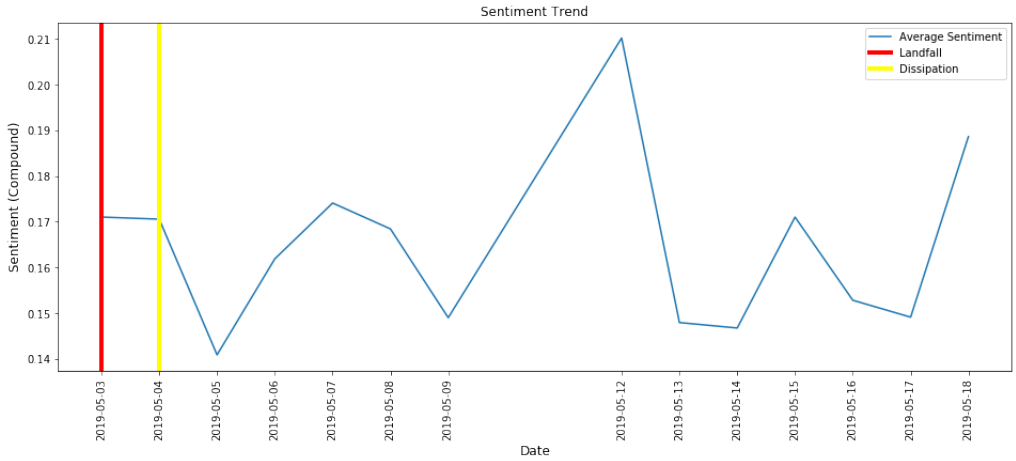

Figure 16. Sentiment trend.

We are also interested in comparing the sentiment trend in areas affected by cyclone Fani with that of the unaffected regions. We have divided the dataset into two groups. The treatment group includes those tweets sent within a specified distance from cyclone Fani's path, and the control group consists of all other tweets. Figure 17, Figure 18, and Figure 19 illustrate the control group's sentiment trend versus the treatment group for different distance thresholds.

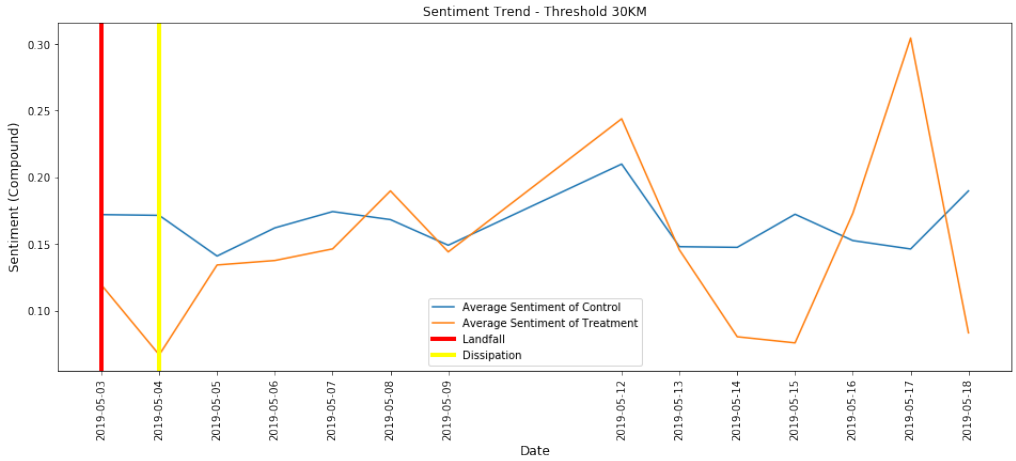

Figure 17. Sentiment trend for the treatment and control groups - Threshold of 30Km.

Figure reffig:trend30 compares the trends for the distance threshold of 30 kilometers. The data show that when cyclone Fani made landfall, the sentiment of the treatment group declined. In the days right after the landfall, the treatment group's sentiment starts to catch up with the control group's sentiment.

Figure 18 compares the trends for a distance threshold of 50 kilometers. The sentiment of the treatment group is lower as cyclone Fani makes landfall. However, the treatment group's sentiment starts to catch up and even surpasses the control group's sentiment days after the landfall.

Figure 19 compares the trends for a distance threshold of 100 kilometers. One can observe the same pattern in the trend lines for this threshold. The sentiment of the treatment group is slightly lower as cyclone Fani makes landfall. In the

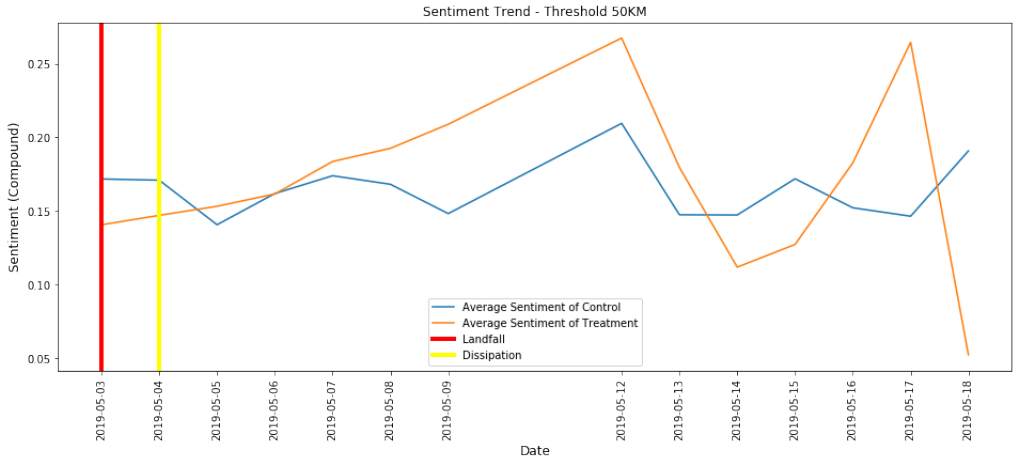

FIGURE 18. SENTIMENT TREND FOR THE TREATMENT AND CONTROL GROUPS - THRESHOLD OF 50KM.

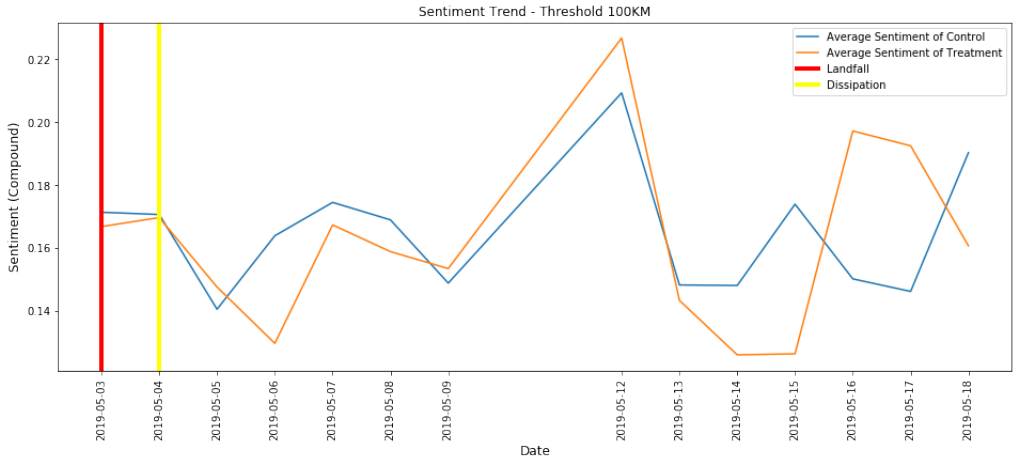

FIGURE 19. SENTIMENT TREND FOR THE TREATMENT AND CONTROL GROUPS - THRESHOLD OF 100KM.

days after Fani's landfall, the treatment group's sentiment still follows the control group's sentiment.

IV. The Model

We assume two periods ($t = 1, 2$) with an election scheduled between the two periods. A natural disaster may transpire in any of these periods. d_t gets a value

of -1 if a natural disaster occurs in time t and 0 otherwise. d_t has a Bernoulli distribution with parameter of λ representing the probability of natural disaster's occurrence at $t = i$.

$$d_t = \begin{cases} -1 & \lambda \\ 0 & 1-\lambda \end{cases}$$

The incumbent governor needs to decide how to allocate the available endowment between two policies at the outset of $t = 1$: (a) investing in infrastructures and preemptive measures against the natural disaster, (b) and spending it in different forms of cash transfer and relief funds.

$$endowment = i + I = 1 \ , \ 0 \leq i, \ I \leq 1$$

i represents the cash transfer, and I represents the Investment. These policies together exhaust all the endowments; hence, no resources are left for the second period.

Moreover, each politician has a competency level unknown to voters. Voters' belief about politicians' competency, θ, follows a uniform distribution.

$$\theta \sim unif(0, 1)$$

Voters will update their information about an incumbent's competency at the end of $t = 1$. We will elaborate more on this later on in this section.

The output in each year follows this scheme:

$$y_1 = \bar{\theta} + f(I)d_1 + i + \epsilon_1$$

$$y_2 = \theta_2 + f(I)d_2 + \epsilon_2$$

$\bar{\theta}$ is the true competency of incumbent governor and θ_2 is the true competency of the politician who wins the election. The function f represents the preemptive

effect of investment. If the incumbent decides to invest I at the begining of $t = 1$ then this will reduce the damage caused by natural disaster by $f(I)$.

$$f : I \to [0, 1],$$

$$f(0) = 1,\ f(1) = 0,$$

$$f' < 0 \text{ and it is continuous on } (0,\ 1),$$

$$\lim_{I \to 0^+} f'(I) = -\infty,\ \lim_{I \to 1^-} f'(I) = 0$$

Finally, the ϵ_t represents any shocks at time t except the natural disaster.

Voters will try to maxamize their benefit function by selecting between candidates. Their benefit function is:

$$E(y_1 + \beta y_2)$$

And incumbent tries to maximise their benefit function by selecting i, and I. Her benefit function is:

$$E(y_1 + \beta y_2) + \nu + \beta\rho\nu$$

$0 \leq \nu \leq 1$ represents the monetary incentive of a govornor in each period. For instance, ν may be thought as the governor's salary for each term. Furthermore, ρ gets a value of 1 if the incumbent wins the election, and 0 otherwise. β is also the discount term.

Before $t = 2$ voters have to decide between the incumbent and a challenger. The challenger's competency, θ_c is unknown; however, voters have got some information about the incumbent during $t = 1$ and now they can update their expectation of incumbent's competency. They will assign a value to it, θ_{inc}, by following mechanism:

$$E(\theta_{inc}|i, I) = mp.i\ +\ (1 - mp).I,\quad 0 \leq mp \leq 1$$

We assume that voters are homogeneous regarding how myopic they are. mp is the myopia rate, a measure between zero and one. A higher mp implies that voters assign more weight to the cash transfer, i, rather than investment against natural disaster, I, and hence they value short-term benefits more than the lonf-term benefits. We call voters "myopic" if $mp \geq \frac{1}{2}$.

As decisions and outcomes in $t = 1$ are sunk cost, voters just need to compare $E(\theta_c)$ with $E(\theta_{inc})$ to decide on whom they want to vote for. Moreover, note that $E(\theta_c) = \frac{1}{2}$. For the sake of simplicity, we assume that the tie breaker is from the incumbent's party.

The incumbent knows about voters' behaviors and tries to select i, and I in a way that will maximize their benefit function.

$$E(y_1 + \beta y_2) + \nu + \beta\rho\nu$$

$$= (\bar{\theta} + \beta E(\theta_2)) - (1+\beta)\lambda f(I) + i + (1+\beta\rho)\nu$$

$$= \begin{cases} (1+\beta)\bar{\theta} - (1+\beta)\lambda f(I) + i + (1+\beta)\nu & \text{if } mp.i \ + \ (1-mp).I \geq \frac{1}{2} \\ (\bar{\theta} + \frac{\beta}{2}) - (1+\beta)\lambda f(I) + i + \nu & \text{if } mp.i \ + \ (1-mp).I < \frac{1}{2} \end{cases}$$

The difference between two scenarios is $\beta(\nu + \bar{\theta} - \frac{1}{2})$. This implies that if $\bar{\theta} > \frac{1}{2} - \nu$ then the incumbent has incentive to win the election irrespective of i and I, if $\bar{\theta} < \frac{1}{2} - \nu$ then the incumbent has incentive to lose, and she is indifferent otherwise. In other words, if the incumbent's competency is high enough then she has incentive to win, and if the competency is low then she only has incentive to win if the salary is large enough.

Hence, if $\bar{\theta} > \frac{1}{2} - \nu$ then incumbent solves:

$$(1+\beta)\bar{\theta} - (1+\beta)\lambda f(I) + i + (1+\beta)\nu$$

$$s.t. \ \ mp.i \ + \ (1-mp).I \geq \frac{1}{2}$$

$$i + I = 1, \quad i, I \in [0, 1]$$

We denote the solutin to $f'(I) = -\frac{1}{\lambda(1+\beta)}$ with I^*. Then the solution to optimization problem will be:

$$= \begin{cases} (i = 1 - I^*, I = I^*) & \text{if } I^* \geq \frac{1}{2} \; and \; mp < \frac{1}{2} \\ (i = \frac{1}{2}, I = \frac{1}{2}) & \text{if } I^* < \frac{1}{2} \; and \; mp < \frac{1}{2} \\ (i = \frac{1}{2}, I = \frac{1}{2}) & \text{if } I^* \geq \frac{1}{2} \; and \; mp > \frac{1}{2} \\ (i = 1 - I^*, I = I^*) & \text{if } I^* < \frac{1}{2} \; and \; mp > \frac{1}{2} \\ (i = 1 - I^*, I = I^*) & \text{if } mp = \frac{1}{2} \end{cases}$$

If $\bar{\theta} < \frac{1}{2} - \nu$ then incumbent solves:

$$Max_{i,I} : \; (\bar{\theta} + \frac{\beta}{2}) - (1 + \beta)\lambda f(I) + i + \nu$$

$$s.t. \;\; mp.i \; + \; (1 - mp).I < \frac{1}{2}$$

$$i + I = 1, \quad i, I \in [0, 1]$$

The solution will be:

$$= \begin{cases} (i = \frac{1}{2}, I = \frac{1}{2}) & \text{if } I^* \geq \frac{1}{2} \; and \; mp < \frac{1}{2} \\ (i = 1 - I^*, I = I^*) & \text{if } I^* < \frac{1}{2} \; and \; mp < \frac{1}{2} \\ (i = 1 - I^*, I = I^*) & \text{if } I^* \geq \frac{1}{2} \; and \; mp > \frac{1}{2} \\ (i = \frac{1}{2}, I = \frac{1}{2}) & \text{if } I^* < \frac{1}{2} \; and \; mp > \frac{1}{2} \\ (i = 1 - I^*, I = I^*) & \text{if } mp = \frac{1}{2} \end{cases}$$

One should note that this case is to some extent unrealistic. It implies that the incumbent believes that she is incompetent and tries to increase the preemptive investment to decrese her chance of winning and increase the expected future outcome. This is theoretically a rational deed but realistically politicians do not consider themselves incompetent and do not try to loose the election, i.e., the $\bar{\theta}$ is naturaly high.

If $\bar{\theta} = \frac{1}{2} - \nu$ then incumbent solves:

$$Max_{i,I}: \ (\bar{\theta} + \frac{\beta}{2}) - (1+\beta)\lambda f(I) + i + \nu$$
$$i + I = 1, \quad i, I \in [0,1]$$

Then the solution will be $(i = 1 - I^*, I = I^*)$.

We assume that $\bar{\theta}$ is inherently high per our previous argument. It means that politicians believe in themselves and consider themselves competent. In this case, as voters become more myopic (a switch from $mp < \frac{1}{2}$ to $mp > \frac{1}{2}$) then the level of preemptive investment decreases. Moreover, if the chance of natural disaster increases (an increase in λ) then the incumbent increases the level of preemptive investment.

We also assumed that $E(\theta_{inc}|i, I) = mp.i \ + \ (1 - mp).I$; however, voters do not solely rely on the data about the investment and relief funds to update their belief about politicians' competency. They refer to many sources and experiences. However, we believe that this update in the belief is mirrored in voters sentiment in social media. Hence, a more accurate presentation of the belief update would be $E(\theta_{inc}|\text{Information Set}) = S(\text{short-term actions, long-term policies}), \ S : D_S \rightarrow [0,1]$ where $S(.)$ represents a sentiment function. As long as the incumbent politician considers herself competent she will try to increase $S(.)$ above $E(\theta_c) = \frac{1}{2}$. Her action depends on the responsiveness of voters' sentiment, $S(.)$, to short-term actions and long-term policies. The sentiment function is more responsive to short-term actions for myopic voters. Hence, as sentiment increases, the incumbent's response to a recent natural disaster gets a higher weight in the election's outcomes.

One important implication of the model is the difference in election outcomes based on the type of voters.

- Voters are myopic ($mp \geq \frac{1}{2}$):

  If voters are myopic, then the incumbent allocated the most of the endow-

ment to the short-term actions ($I < \frac{1}{2}$). The sentiment function is also highly responsive to short-term actions. Hence, after a natural disaster, we expect a higher increase in sentiment and a more favorable election result for the incumbent.

- Voters are not myopic ($mp < \frac{1}{2}$):

  If voters are not myopic, then the incumbent allocated the most of the endowment to the long-term policies ($I > \frac{1}{2}$). The sentiment function is also less responsive to the sort-term actions and most responsive to the long-term policies. Hence, after a natural disaster, we expect a lower increase in sentiment (due to a lower level of cash transfer or relief funds) and a lower chance of winning for the incumbent.

V. Results

In the first step, we analyzed the effect of cyclone Fani on sentiment (Table 3, 4, and 5). We follow the specification model as given below:

$$y_{it} = \beta_0 + \beta_1\ Time_{it} + \beta_2\ Treatment_{it} + \beta_3\ Time_{it} * Treatment_{it} + \epsilon_{it}$$

where t represents the days in our sample. We filtered our sample for tweets posted on or after May 3, 2019, and tweets posted on or before May 12, 2019. We constrained our time frame because the relevance of Fani vanishes over time. The new sample includes tweets posted up to one week after Fani's landfall.

We use our data at the AC level; therefore, i represents the AC. y_it is the general sentiment of a particular AC on a specific day. It is calculated by taking the average of tweets' sentiment (compound score) in each AC per day.

The time variable is a dummy variable that takes a value of 1 if the tweet is posted after the dissipation of cyclone Fani, that is, on or after May 5, and receives a value of 0 otherwise. The treatment variable is also a dummy variable

that takes a value of 1 if the tweet is sent from a location hit by cyclone Fani, that is, within a specified distance threshold from the cyclone's path. It obtains a value of 0 otherwise. We also include the interaction term to capture the effect of cyclone Fani on sentiment.

The model follows the difference-in-difference (DID method. Time and treatment variables control for the effect of time and location, while the interaction term represents the effect of cyclone Fani. We assume that any systematic difference between the control and treatment groups remains constant throughout this short time frame. This assumption is especially plausible considering the fact that the treatment and control tags are assigned by nature.

Table 3, 4, and 5 present the results of this regression for different thresholds of 30km, 50km, and 100km, respectively. Each of these tables includes two models. The second model extends the former by including the day fixed effect and PC fixed effect.

The results are consistent in all three tables. Table 3 shows that the time coefficient is not significant, while the treatment coefficient is strongly significant. We are mostly interested in the interaction term's coefficient as it represents the effect of cyclone Fani on the sentiment. This coefficient is positive and strongly significant. It means that tweets sent in areas hit by Fani after its dissemination have a higher sentiment compared to tweets posted outside of Fani's path.

One possible explanation is the government's response to the cyclone Fani after its landfall. This may outweigh the effect of physical and emotional damages caused by Fani on the general sentiment. We specifically expect this when voters are myopic, as the incumbent is then more inclined to allocate the budget to short-term actions such as the first-response budget for natural disasters. Voters are also more responsive to these short-term actions; therefore, an increase in sentiment is expected based on our model.

We can observe the same results for the distance threshold of 50 km in Table 4. The coefficient of the interaction term is positive and strongly significant.

Table 3—The Effect of Fani on Sentiment - Threshold of 30km.

	Model 1	Model 2
Time	-0.0057	-0.0071
	(0.007)	(0.007)
Treatment	-0.1367***	-0.1433***
	(0.041)	(0.042)
Time # Treatment	0.1494***	0.1478 ***
	(0.049)	(0.050)
Constant	0.2244***	0.2280***
	(0.006)	(0.012)
PC FE	NO	YES
Day FE	NO	YES
Level of observations	AC / Day	AC / Day
Number of Observations	12130	12130

Standard Errors are heteroscedasticity robust.

* p<0.05, ** p<0.01, *** p<0.001

It implies that tweets sent in areas within 50 Km of Fani's path have a higher sentiment after dissemination than tweets posted outside of this threshold.

Table 4—The Effect of Fani on Sentiment - Threshold of 50km.

	Model 1	Model 2
Time	-0.0068	-0.0082
	(0.007)	(0.007)
Treatment	-0.0968***	-0.0952***
	(0.030)	(0.031)
Time # Treatment	0.1205***	0.1202***
	(0.036)	(0.036)
Constant	0.2250***	0.2290***
	(0.006)	(0.012)
PC FE	NO	YES
Day FE	NO	YES
Level of observations	AC / Day	AC / Day
Number of Observations	12130	12130

Standard Errors are heteroscedasticity robust.
* p<0.05, ** p<0.01, *** p<0.001

The significance of results decreases when we apply the distance threshold of 100km (Table 5). The sign of coefficients is consistent but the effect of cyclone Fani on sentiment is now significant at the 90% confidence level.

In the next step, we analyze the ruling party's (BJP's) chance to win the election before and after cyclone Fani. We use the election reports provided by the Election Commission of India. We use a logit model for this analysis.

$$y_i = \beta_0 + \beta_1 \, Time_i + \beta_2 \, Margin_i + \beta_3 \, Turnout_i + \epsilon_i$$

where i represents the AC regions in the sample. y_i is a dummy variable

Table 5—The Effect of Fani on Sentiment - Threshold of 100km

	Model 1	Model 2
Time	-0.0061	-0.0077
	(0.007)	(0.007)
Treatment	-0.0433*	-0.0424*
	(0.024)	(0.024)
Time # Treatment	0.0500*	0.0508*
	(0.027)	(0.027)
Constant	0.2246***	0.2286***
	(0.006)	(0.012)
PC FE	NO	YES
Day FE	NO	YES
Level of observations	AC / Day	AC / Day
Number of Observations	12130	12130

Standard Errors are heteroscedasticity robust.

* p<0.05, ** p<0.01, *** p<0.001

indicating whether the BJP party won the 2019 parliamentary election in the AC region of i. $Time_i$ is our variable of interest, taking a value of 1 if the polling date for the AC region of i was after cyclone Fani. It takes a value of 0 if the polling date was before Fani. $Margin_i$ is the margin votes of the BJP party in the 2014 election in the AC region of i divided by the total votes cast in that AC. This variable's sign is positive if BJP was the winner in 2014 and negative otherwise. We also include the turnout rate in our regression.

Table 6—Fani's effect on the winning chance of BJP candidates

	Estimated Coefficient	Odd Ratio
Time-Fani	0 .929***	2.532***
	(.108)	(.274)
BJP Margin 2014	4.852***	128.016***
	(.330)	(42.269)
Turnout	1.405 **	4.077**
	(.579)	(2.362)
Constant	-0.811 **	0.444**
	(.396)	(.176)
Level of observations	AC	AC
Number of Observations	2,396	2,396

Standard Errors are heteroscedasticity robust.

* p<0.05, ** p<0.01, *** p<0.001

The chances of BJP winning the election increases after cyclone Fani. This observation might result from many factors, including the government's performance in response to Fani. The margin of votes in 2014 increases the BJP's chance to win the election in 2019, in addition to the turnout rate. This is in line with our model's prediction assuming that voters are myopic. The increase in myopic voters' sentiment (reported in previous tables) leads to an increase in winning probability as it increases the sentiment function and hence its likelihood

to surpass $E(\theta_2 = \frac{1}{2})$.

In the next analysis, we want to capture the causal effect of sentiment on election outcomes using a 2SLS model. We build our sample by using tweets that expect a forthcoming election. Hence, tweets posted after the election are filtered out. The sentiment variable is aggregated at the AC level, and all results are reported at this level. Table 7 and 8 show the results of the 2SLS model. The former uses the turnout rate as an outcome, and the latter uses the winning status of the incumbent party as an outcome. We use the distance to cyclone Fani as an instrument for sentiment.

Table 7 shows that an increase in sentiment leads to an increase in the turnout rate in areas where the incumbent parliament member is from opposition parties. However, we cannot make a causal interpretation for areas where the incumbent parliament member is from the government's party as the result of the first stage regression does not support a strong correlation between the sentiment and the instrument.

Table 8 shows that an increase in sentiment causes a significant increase in the probability of getting reelected in areas where the incumbent member of the parliament is from the ruling government's party. However, the sentiment does not have a significant effect on getting reelected in areas where the opposition parties are ruling.

We interpret these results based on the difference between parties ruling at the local and federal levels. Voters try to link the increase in their sentiment to the government's performance, either the local or the federal one. The task is easy when voters are located in areas that the parliament member was from the federal ruling party, i.e., BJP, as both local and federal governments are from the same party. Hence, they consider the BJP party's response as the cause of the increase in their sentiment. This consensus implies no competition in the election; hence, we expect to observe a lower turnout rate and a higher chance of winning for BJP.

However, confusion arises when voters are located in areas that the parliament

Table 7—2SLS Model: The Effect of Sentiment on Election outcomes

	Turnout	
	Opposition	BJP
Sentiment	2.8559***	-2.414201
	(0.613)	(11.47019)
	First Stage: Instrument: located within 100KM of Cyclone's path	
First Step's Coefficient	.102861 ***	-.0159577
	(.0250248)	(.0811864)
Level of Observations	AC	AC
Number of Observations	204	347
PC Fixed Effect	Yes	Yes

Standard Errors are adjusted for the state-level clusters

* p<0.05, ** p<0.01, *** p<0.001

Table 8—2SLS Model: The Effect of Sentiment on Election outcomes

	Incumbent Wins	
	Opposition	BJP
Sentiment	1.205741	4.453711 ***
	(1.496505)	(1.135534)
	First Stage: Instrument: located within 100KM of Cyclone's path	
First Step's Coefficient	.0995961 ***	-.1251653 ***
	(.0267885)	(.0407249)
Level of Observations	AC	AC
Number of Observations	159	149
PC Fixed Effect	Yes	Yes

Standard Errors are adjusted for the state-level clusters

* p<0.05, ** p<0.01, *** p<0.001

member was from opposition parties since the ruling party at the local level is different from the ruling party at the federal level. Voters need time to update their information set by evaluating the performance of authorities of these parties. If the election date is very close or there is a friction of access to the information, voters will decide based on the already available information set or stick to their decision in the last election. In this case, we have a competition between campaigns, and we expect to observe a higher turnout rate but no change in the winning rate as long as the voters' information set is not updated.

VI. Conclusion

The political structure and the interaction between politicians and citizens have meaningful impacts on society's general welfare. Therefore, election and voters' preferences have long been studied by researchers. Many theories, including the median voter model and mean voter model, try to explain how voters' preferences form and the underlying mechanisms of elections.

One leading theory is voter's myopia. It claims that voters form their opinion based more on the latest events than on earlier events or their future consequences. Some studies have analyzed the government's response to a very recent natural disaster and its effect on the upcoming elections, with the aim to explore whether the most current government's response has a crucial rule on an upcoming election's outcome.

However, these studies use government spending as a proxy for the response's efficiency and voters' satisfaction. This is problematic because federal spending is usually not separable from states' or local governments' spending, and a high level of relief funds does not imply a better response to a natural disaster..

A better approach is to directly measure the satisfaction of voters with the government's response. We use sentiment analysis to measure the general attitude toward the government. This study incorporates various machine learning and NLP techniques with a 2SLS model to capture the effect of sentiment on election

outcomes.

We collected our dataset from May 2 to June 9, 2019, using Twitter's API, including more than one million tweets after filtering English tweets. We use the topic modeling approach, the LDA model, to find relevant topics in the data. We identified three topics for our sample, namely "Infrastructure," "Religion," and "Cyclone Fani." Next, we use sentiment analysis, the VADER algorithm, to analyze people's sentiment while dealing with cyclone Fani.

Our results show that the tweets sent in areas hit by Fani after its dissemination have a higher sentiment compared to tweets posted outside Fani's path. We specifically expect this when voters are myopic, as the incumbent is then more inclined to allocate the budget to short-term actions such as the first-response budget for natural disasters. Voters are also more responsive to these short-term actions; therefore, an increase in sentiment is expected based on our model.

We also explored the election outcome and concluded that the probability of winning the election increases for the ruling party (BJP) after cyclone Fani. Moreover, the margin of votes in 2014 increases the BJP's chance to win the election in 2019, in addition to the turnout rate. This is also in line with our model's prediction assuming that voters are myopic as it predicts that an increase in myopic voters' sentiment leads to an increase in winning probability.

We also studied the causal effect of sentiment on election outcomes. We used the distance to cyclone Fani as an instrument for sentiment in a 2SLS model. The results show that an increase in sentiment leads to a rise in the probability of being reelected in areas where parliament members are from the ruling government's party. The ruling party can direct the budget towards short-term actions and increase myopic voters' sentiment in favor of the ruling party.

Our study also has some limitations. Our dataset is filtered based on English tweets as our NLP tool is well-designed for this language. Using other languages will increase the sample size and also will capture the sentiment more accurately. We also filtered the data based on available geolocation information. This is a

restriction imposed by Twitter for users' privacy. Hence, we used tweets posted by users who shared their exact or approximate locations. Moreover, our sample's time frame begins just two days before Fani's dissipation. An extended time frame provides an opportunity for a placebo test and a more accurate estimation of the sentiment.

Our work draws on a large dataset of tweets to analyze people's sentiment before and after cyclone Fani. The NLP methods are also incorporated with a causal inference method, 2SLS, to capture the effect of sentiment on election outcomes. This is empirically significant in the literature as the previous studies in the strand of myopic voter literature used government spending to reflect voters' sentiment and opinion. However, We utilize NLP methods to measure sentiment directly. Furthermore, our study also focuses on the election outcomes in a developing country, while many other studies focus mostly on the USA and Europe. Our framework will be a useful tool for future elections and may work as a good predictor of electoral outcomes.

REFERENCES

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The welfare effects of social media." *American Economic Review*, 110(3): 629–676.

Ashktorab, Zahra, Christopher Brown, Manojit Nandi, and Aron Culotta. 2014. "Tweedr: Mining twitter to inform disaster response." 269–272. amulyayadav.com.

Bechtel, M M, and J Hainmueller. 2011. "How lasting is voter gratitude? An analysis of the short-and long-term electoral returns to beneficial policy." *American Journal of Political Science*.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent Dirichlet Allocation." *Journal of machine Learning research*, 3(Jan): 993–1022.

Caplin, Andrew, and Barry Nalebuff. 1991. "Aggregation and Social Choice: A Mean Voter Theorem." *Econometrica*, 59(1): 1–23.

Cheong, France, and Christopher Cheong. 2011. "Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods." *PACIS*, 11: 46–46.

Cho, Seok-Ju, and John Duggan. 2009. "Bargaining foundations of the median voter theorem." *Journal of Economic Theory*, 144(2): 851–868.

Digrazia, Joseph, Karissa McKelvey, Johan Bollen, and Fabio Rojas. 2013. "More tweets, more votes: social media as a quantitative indicator of political behavior." *PLoS One*, 8(11): e79449.

Fournier, Patrick, Richard Nadeau, André Blais, Elisabeth Gidengil, and Neil Nevitte. 2004. "Time-of-voting decision and susceptibility to campaign effects." *Electoral Studies*, 23(4): 661–681.

Gilbert, C H E, and Erric Hutto. 2014. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." Vol. 81, 82. eegilbert.org.

Golovchenko, Y, C Buntain, G Eady, and others. 2020. "Cross-Platform

State Propaganda: Russian Trolls on Twitter and YouTube During the 2016 US Presidential Election." *The International Journal of Press/Politics.*

Grossman, Guy, and Kristin Michelitch. 2017. "Information Dissemination, Competitive Pressure, and Politician Performance between Elections: A Field Experiment in Uganda."

Healy, Andrew, and Neil Malhotra. 2009. "Myopic Voters and Natural Disaster Policy." *American Political Science Review*, 103(3): 387–406.

Hong, Liangjie, and Brian D Davison. 2010. "Empirical study of topic modeling in Twitter." *SOMA '10*, 80–88. New York, NY, USA:Association for Computing Machinery.

Humphreys, Macartan, and Jeremy Weinstein. 2012. "Policing politicians: citizen empowerment and political accountability in Uganda preliminary analysis." *Columbia Universities. Unpublished manuscript.*

India Elects 2019: The World's Largest Election, Explained. 2019. *https://carnegieendowment.org/publications/interactive/india-elects-2019*, Accessed: 2020-9-14.

Kumar, Shamanth, Geoffrey Barbier, Mohammad Ali Abbasi, and Huan Liu. 2011. "Tweettracker: An analysis tool for humanitarian and disaster relief." *ICwSM*, 11: 78–82.

Lettier. 2018. "Your Guide to Latent Dirichlet Allocation." *https://medium.com/@lettier/how-does-lda-work-ill-explain-using-emoji-108abf40fa7d*, Accessed: 2020-9-14.

Levitt, Steven D. 1996. "How Do Senators Vote? Disentangling the Role of Voter Preferences, Party Affiliation, and Senator Ideology." *The American Economic Review*, 86(3): 425–441.

Mandel, Benjamin, Aron Culotta, John Boulahanis, Danielle Stark, Bonnie Lewis, and Jeremy Rodrigue. 2012. "A demographic analysis of

online sentiment during hurricane irene." 27–36. aclweb.org.

Nagy, Ahmed, and Jeannie A Stamberger. 2012. "Crowd sentiment detection during disasters and crises." pdfs.semanticscholar.org.

Neppalli, Venkata K, Cornelia Caragea, Anna Squicciarini, Andrea Tapia, and Sam Stehle. 2017. "Sentiment analysis during Hurricane Sandy in emergency response." *International Journal of Disaster Risk Reduction*, 21: 213–222.

Oswald, Andrew J, and Nattavudh Powdthavee. 2010. "Daughters and Left-Wing Voting." *The Review of Economics and Statistics*, 92(2): 213–227.

Pagano, Marco, and Paolo F Volpin. 2005. "The Political Economy of Corporate Governance." *American economic review*, 95(4): 1005–1030.

Ragini, J Rexiline, P M Rubesh Anand, and Vidhyacharan Bhaskar. 2018. "Big data analytics for disaster response and recovery through sentiment analysis." *International Journal of Information Management*, 42: 13–24.

Scervini, Francesco. 2012. "Empirics of the median voter: democracy, redistribution and the role of the middle class." *The Journal of Economic Inequality*, 10(4): 529–550.

Sinding Bentzen, Jeanet. 2019. "Acts of God? Religiosity and natural disasters across subnational world districts." *The Economic Journal*, 129(622): 2295–2321.

Stephens-Davidowitz, Seth I. 2012. "The Effects of Racial Animus on a Black Presidential Candidate: Using Google Search Data to Find What Surveys Miss."

Stigler, George J. 1973. "General Economic Conditions and National Elections." *The American Economic Review*, 63(2): 160–167.

Tella, R D, S Galiani, and E Schargrodsky. 2019. "Persuasive Propaganda during the 2015 Argentine Ballotage."

Tumasjan, Andranik, Timm O Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. "Predicting elections with twitter: What 140 characters

reveal about political sentiment." Citeseer.

Weng, Jianshu, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. "TwitterRank: finding topic-sensitive influential twitterers." *WSDM '10*, 261–270. New York, NY, USA:Association for Computing Machinery.

wiki/User:Furfur. 2014*a*. "Lok Sabha constituencies in Madhya Pradesh after 2008." *https://commons.wikimedia.org/wiki/{\}File:Madhya_Pradesh_Wahlkreise_Lok_Sabha.svg*.

wiki/User:Furfur. 2014*b*. "Vidhan Sabha constituencies in Madhya Pradesh after 2008." *https://commons.wikimedia.org/wiki/{\}File:Wahlkreise_zur_Vidhan_Sabha_von_Madhya_Pradesh.svg*.

wiki/user:furfur. 2019. "Time schedule of the Indian Lok Sabha election, 2019." *https://commons.wikimedia.org/w/index.php?curid=77491634*.

Williams, Christine, and Girish Gulati. 2008. "What is a social network worth? Facebook and vote share in the 2008 presidential primaries." Citeseer.

Wlezien, Christopher. 2015. "The myopic voter? The economy and US presidential elections." *Electoral Studies*, 39: 195–204.

Appendix

Here is a comprehensive list of states with phase distribution of their constituencies:

- Phase one - 11 April:

  Andhra Pradesh (25 constituencies), Arunachal Pradesh (2 constituencies), Assam (5 out of 14 constituencies), Bihar (4 out of 40 constituencies), Chhattisgarh (1 out of 11 constituencies), Jammu and Kashmir (2 out of 6 constituencies), Maharashtra (7 out of 48 constituencies), Manipur (1 out of 2 constituencies), Meghalaya (2 constituencies), Mizoram (1 constituency), Nagaland (1 constituency), Odisha (4 out of 21 constituencies), Sikkim (1 constituency), Telangana (17 constituencies), Tripura (1 out of 2 constituencies), Uttar Pradesh (8 out of 80 constituencies), Uttarakhand (5 constituencies), West Bengal (2 out of 42 constituencies), Andaman and Nicobar Islands (1 constituency), Lakshadweep (1 constituency)

- Phase two - 18 April:

  Assam (5 out of 14 constituencies), Bihar (5 out of 40 constituencies), Chhattisgarh (3 out of 11 constituencies), Jammu and Kashmir (2 out of 6 constituencies), Karnataka (14 out of 28 constituencies), Maharashtra (10 out of 48 constituencies), Manipur (1 out of 2 constituencies), Odisha (5 out of 21 constituencies), Tamil Nadu (38 out of 39 constituencies), Uttar Pradesh (8 out of 80 constituencies), West Bengal (3 out of 42 constituencies), Puducherry (1 constituency)

- Phase three - 23 April:

  Assam (3 out of 14 constituencies), Bihar (5 out of 40 constituencies), Chhattisgarh (7 out of 11 constituencies), Goa (2 constituencies), Gujarat (26 constituencies), Jammu and Kashmir (1/3 out of 6 constituencies),

Karnataka (14 out of 28 constituencies), Kerala (20 constituencies), Maharashtra (14 out of 48 constituencies), Odisha (6 out of 21 constituencies), Tripura (1 out of 2 constituencies), Uttar Pradesh (10 out of 80 constituencies), West Bengal (5 out of 42 constituencies), Dadra and Nagar Haveli (1 constituency), Daman and Diu (1 constituency)

- Phase four - 29 April:

  Bihar (5 out of 40 constituencies), Jammu and Kashmir (1/3 out of 6 constituencies), Jharkhand (3 out of 14 constituencies), Madhya Pradesh (6 out of 29 constituencies), Maharashtra (17 out of 48 constituencies), Odisha (6 out of 21 constituencies), Rajasthan (13 out of 25 constituencies), Uttar Pradesh (13 out of 80 constituencies), West Bengal (8 out of 42 constituencies)

- Phase five - 6 May:

  Bihar (5 out of 40 constituencies), Jammu and Kashmir (1 1/3 out of 6 constituencies), Jharkhand (4 out of 14 constituencies), Madhya Pradesh (7 out of 29 constituencies), Rajasthan (12 out of 25 constituencies), Uttar Pradesh (14 out of 80 constituencies), West Bengal (7 out of 42 constituencies)

- Phase six - 12 May:

  Bihar (8 out of 40 constituencies), Haryana (10 constituencies), Jharkhand (4 out of 14 constituencies), Madhya Pradesh (8 out of 29 constituencies), Uttar Pradesh (14 out of 80 constituencies), West Bengal (8 out of 42 constituencies), Delhi (7 constituencies)

- Phase seven - 19 May:

  Bihar (8 out of 40 constituencies), Himachal Pradesh (4 constituencies), Jharkhand (3 out of 14 constituencies), Madhya Pradesh (8 out of 29 constituencies), Punjab (13 constituencies), Uttar Pradesh (13 out of 80 con-

stituencies), West Bengal (9 out of 42 constituencies), Chandigarh (1 constituency)