Towards Understanding Polarization, and Its Life on Social Media

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science by Research
in
Computational Linguistics

by

Ashutosh Ranjan
201425031
ashutosh.ranjan@research.iiit.ac.in

Language Technology Research Centre
International Institute of Information Technology
Hyderabad - 500 032, INDIA
March 2021

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "**Towards Understanding Polarization, and Its Life on Social Media**" by **Ashutosh Ranjan**, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date

Adviser: Prof. Dipti Misra Sharma and Prof. Radhika Krishnan

To the randomness of life.

Acknowledgments

It's difficult to explain how amazing I feel writing this acknowledgement, because I have been struggling to finish this for a very long time.

I would like to thank Prof. Dipti Misra Sharma for bearing with me all this while, for being available whenever I have wanted her guidance, for excusing my intermittent breaks and for her encouragement and optimism in allowing me to pursue this vague idea of mine. I'd also like to thank Prof. Radhika Krishnan for evaluating and helping this work from the perspective of social sciences.

And then, I would like to thank myself, for finishing this, for not giving up everytime it felt it couldn't be done.

Abstract

This work is an attempt to study polarization on social media data, in the Indian context. We focus on multiple controversial and talked about events and topics in the context of India, namely 1) the Sabarimala Temple (located in Kerala, India) incident which became a nationwide controversy when two women under the age of 50 secretly entered the temple breaking a long standing temple rule that disallowed women of menstruating age (10-50) to enter the temple and 2) the Indian government's move to demonetise all existing 500 and 1000 denomination banknotes, comprising of 86% of the currency in circulation, in November 2016 3) The Pulwama-Balakot incidents and its effect on popular sentiment towards India's Prime Minister Narendra Modi 4) Social media activity on events surrounding multiple lynching incidents from 2017 to 2019 and 5) the Indian Economy from 2013 to 2019 and popular response towards it on social media. We gather tweets from news media, and from common people on Twitter around these events in various time periods. Then, we pre-process and annotate them with their sentiment polarity and emotional category, and analyse trends to help us understand changing polarity over time around controversial events. While this study tries to find trends and patterns to decode the aspect of polarization in India's social fabric, it's also a case study on all these events and how India reacted to each one of them. The tweets collected are in English, Hindi and code-mixed Hindi-English.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Social media, over the past few years, has become the fastest growing medium for expressing opinions for billions of people worldwide. People have always had opinions in history, some extreme and some moderate. Before the proliferation of mass media and the emergence of the Internet, the spread of word-of-mouth and media monitoring were geographically and temporally limited. The media landscape, however, changed dramatically as the majority of communication migrated to the Internet, completely eliminating regional barriers (Mancini [46]). Glean [22], for instance, a 24/7 media monitoring service, monitors over 60,000 online sources in more than 250 languages in 191 countries. Consequently what has changed today, is the way people can make their opinions reach masses (read millions) of people in seconds to minutes to hours, depending on the popularity of the person and the general controversiality (for lack of a better word) of the opinion. Today any person is able to express their ideas to millions of people, and thereby has the potential to influence masses much faster.

A famous real life example comes to mind when one thinks of the change in the speed and ease of disseminating information over the last 2-3 decades. It is the case of Theodore John Kaczynski, also known as the Unabomber. In 1995, he sent a letter to The New York Times and promised to ”desist from terrorism” and further bombings if the Times or The Washington Post published his essay *Industrial Society and Its Future*. In this day today, you don't need to become a terrorist to get your ideas to a large audience, but you only need to have a Twitter account, and something to say. And not just Twitter, today there are thousands of new websites, youtube channels, blogs etc. launched (and discontinued) daily. As reported by Aldred et al. [2] in *The World's Most Powerful Blogs, 2008*, many of the highly influential media (i.e. Mashable, Techcrunch, Endgadget, The Huffington Post) in those times had begun only a couple of years ago as personal blogs.

India is no different. Easy access to internet has brought millions of Indians to Twitter, where they express their opinions on topics that interest, affect them. Leaders use social media to make their opinions reach the masses as well. In a recent survey by Pew Research Center it was revealed that a significant percentage of people get their news from social media [15]. People have always been influenced by the opinions and attitudes of their relatives or celebrities whom they respect (Anderson [3], Goldenberg et al. [23]), with the advent of modern communication technologies like MySpace, Twitter, Facebook and Instagram opinions can be shared to a much greater extent. As the interactions on social media have been increasing multi-fold year on year, it is so much more easier to get influenced and have a change of opinion on different topics. As an example, BrightLocal [14] reports that customers and users who have been following online reviews tend to trust reviews more than personal recommendations. This activity on twitter consequently helps people form opinions or change their already existing opinions on different topics. We have chosen five relevant topics/events and this thesis attempts to study this change in opinion of the masses of people by observing and analysing their tweets on social media with respect to these topics. Tweets have been collected for various periods before and after the events, and analysed upon to study the life of polarization on social media in the Indian context.

Apart from classifying tweets in positive and negative category, we also classify tweets in eight emotional categories namely: joy, fear, anger, anticipation, trust, surprise, sadness and disgust. This helps us analyze the changing emotions of people with respect to an idea, before and after an event.

We also collect tweets from top Indian news media outlets, analyse their changing sentiment and emotional polarity around these topics and events and try to compare the changes with the changing sentiments of the common people on twitter. Hereby we look to see if we can establish any relationship between the two and say that news media influences common people or otherwise.

We decide to focus our efforts on five pressing, controversial and divisive issues - 1) the Sabarimala Temple (located in Kerala, India) incident which became a nationwide controversy when two women under the age of 50 secretly entered the temple breaking a long standing temple rule that disallowed women of menstruating age (10-50) to enter the temple and 2) the Indian government's move to demonetise all existing 500 and 1000 denomination banknotes, comprising of 86% of the currency in circulation, in November 2016 3) The Pulwama-Balakot incidents and its effect on popular sentiment towards India's Prime Minister Narendra Modi 4) Social media activity on events surrounding multiple lynching incidents from 2017 to 2019 and 5) the Indian Economy from 2013 to 2019 and popular response towards it on social media. We study how the polarity of opinions of people change with respect to these issues over time. The complete context of all these events is described later in this chapter, but before that we'd like to define polarization.

1.1 Defining Polarization

Polarization is a concept that comes from science, and it involves light, radiation, or magnetism moving in specific directions. When extended outside science, polarization usually refers to how people think, especially when two views emerge that drive people apart, kind of like two opposing magnets. Polarization involves people moving in two directions — as if they're becoming almost as separate as the North and South Pole. Meriam-Webster defines polarization as the concentration about opposing extremes of groups or interests formerly ranged on a continuum.

1.2 Know Your History

In this section we discuss the history around all these five events, and why we chose them for our analysis.

1.2.1 Sabarimala

The Sabarimala temple is a temple complex located at Sabarimala inside the Periyar Tiger Reserve in Kerala, India. It is the site of one of the largest annual pilgrimages in the world, visited every year by an estimated 40 to 50 million devotees ([9]).

According to the Memoir of the Survey of the Travancore and Cochin States, published by the Madras government in the 19th century, women of menstruating age were denied entry into the Sabarimala temple two centuries ago, as all sexual activity in that vicinity is averse to the celibate deity Lord Ayyappa. The Kerala High Court, in response to a Public Interest Litigation filed in 1991, ruled that the restriction on the entry of women between the ages of 10-50 into the temple was consistent with the use prevalent from time immemorial, and directed the board concerned to maintain the traditional practices of the temple ([47]). The Supreme Court of India, however, overturned the ban on the entry of women on 28 September 2018, ruling it unconstitutional and discriminatory ([74]). On 2nd January 2019, for the first time since the Supreme Court ruling, two women under the age of 50 entered the shrine after attempts by many others failed due to protests by devotees ([30]).

The event of two women entering caused huge uproar on social media. One side of the debate argued for the conservation of tradition, and the other side argued for gender equality (Bijukumar [12]).

We have chosen this topic because it is extremely divisive with two explicit sides with differing points of view. We have tried to understand the sentiment change regarding the topic across time periods, before, during and after the event. In the context of this topic, which includes two important events - the Supreme Court judgement and the entry of women into the temple, when we talk about negative sentiment, we refer to people being unhappy with these developments, people who would be on the side of tradition in this debate. Similarly, when we talk about positive sentiment, we refer to people who are happy with the change and in this case draw more towards gender equality than tradition.

1.2.2 Pulwama Balakot incident

On 14 February 2019, at Lethpora (near Awantipora) in the Pulwama district, Jammu and Kashmir, India, a convoy of vehicles carrying security personnel on the Jammu Srinagar National Highway was targeted by a vehicle-borne suicide bomber. The attack resulted in the deaths of the offender and 40 Central Reserve Police Force (CRPF) workers. The Pakistan-based Islamist militant group Jaish-e-Mohammed claimed responsibility for the attack ([61]).

In response to the Pulwama bomb blast, Balakot airstrike was conducted by India in the early morning hours of February 26 when Indian warplanes crossed the border in the disputed region of Kashmir (currently Pakistan-Occupied-Kashmir) ([56]), and dropped bombs in the vicinity of the town of Balakot in Khyber Pakhtunkhwa province in Pakistan in an attempt to destroy Jaish-e-Mohammed terror camps, the militant group which claimed responsibility for the Pulwama bomb blast.

The terrorist attack on Pulwama happened on 14th February and India's strike in Balakot happened on 26th February. After the Balakot airstrike, there were a lot of doubts and controversy regarding the losses suffered by the terrorists and some claims by the Indian government were unsubstantiated by international media ([80][8]).

These events were right before the Lok Sabha elections, and speculations were made by opposition parties and news media regarding the affect of these airstrikes on the outcome of the elections, and questions were raised if the incumbent Prime Minister Narendra Modi's popularity had increased after the Balakot airstrike (Quint [69], Hindu [29] Wire [86]). Questions were raised on the Indian media as well regarding their overtly celebratory portrayal of the attack ([67]).

We chose this topic because we wanted to investigate the change in sentiment of the Indian populace regarding PM Narendra Modi from before to after the event, and also understand the role of Indian media in this context.

1.2.3 Demonetisation

In November 2016, the government of India announced the demonetization of the Mahatma Gandhi Series of all 500 and 1,000 banknotes. In return for the demonetised banknotes, it also announced the issuance of new 500 and 2,000 banknotes [58]. Narendra Modi, India's Prime Minister, claimed that the action would curb the shadow economy and reduce the use of illicit and counterfeit cash to finance illegal activity and terrorism [78].

In the weeks that followed, the declaration of demonetisation was followed by prolonged cash shortages that caused major economic disruption ([17]). People trying to exchange their banknotes had to wait in long queues, and the rush to exchange cash was connected to many deaths ([37][66]).

The issue came to relevance again in August 2018 when the Reserve Bank of India released a report, according to which approximately 99.3% of the demonetised banknotes, or 15.30 lakh crore (15.3 trillion) of the 15.41 lakh crore that had been demonetised, were deposited with the banking system, leading analysts to state that the effort had failed to remove black money from the economy ([75]). By many economists, the move is blamed for reducing the country's industrial production and slowing down its GDP growth rate ([79]).

The move was initially supported by some bankers as well as by some foreign commentators. It was strongly criticised by others as poorly conceived and unjust, and was met with demonstrations, lawsuits, and strikes across India against the government in many locations ([50]). The stated failure of demonetisation and subsequent downfall of India's GDP growth rate was also a major topic of interest during the 2019 general elections and cited by many opposition parties as a failure of the presiding government.

We chose this event because it had divisive reactions, and wanted to understand change in opinion polarity over time.

1.2.4 Lynching

Over the past few years reported lynching incidents have been on a rise ([36]). The reasons have been plenty, from fake child kidnapping rumours to cow vigilante-ism.

The 2017 Alwar mob lynching was the attack and murder of Pehlu Khan, a dairy farmer from Nuh district of Haryana, allegedly by a group of 200 cow vigilantes affiliated with right-wing Hindutva groups in Alwar, Rajasthan, India ([35]).

Also, it has been reported that the number of lynching incidents in the 2-3 months after the 2019 Lok Sabha elections have been in dozens ([32]).

We have taken four different time periods (across 2016-19), including before event states (2016), to understand how Indian populace has reacted to such lynching incidents on social media.

1.2.5 Indian Economy

Indian economy has always been a point of discussion and debate in the Indian community. With India's falling GDP over the last 5-6 years and radical govt. moves such as demonetisation it has been in the limelight even more. We took this topic to analyse general trends of twitter activity on a fairly commonly discussed topic.

We'll be analysing social media data related to the India economy across 7 years from 2013 to 2019, from before the 2014 Lok Sabha elections to the 2019 Lok Sabha elections.

1.3 Contributions of this thesis

In this thesis, we try to understand polarization in India's context. We pick up five important events and find patterns and trends in changing opinions, before, during and after the event. We collect more than 2 million tweets, annotate them with a sentiment analyser and an emotional analysis lexicon to conduct the experiments. We show how a controversial event leads to a significant increase in polarization on social media even months after the event. We also analyze how Hindi and English news media in India may have different reactions to an event, possibly due to the difference in audience.

We also study the effectiveness of sentiment analysis using tools built for English language on data translated from Hindi to English and code-mixed data transliterated and translated to English.

This thesis also acts as a case study on the reaction of Indians on social media in the context of the five events mentioned above.

Most importantly we are looking for answers to the following questions:

Questions raised in this thesis:

- Is society inherently polarised?
- How does a divisive event affect the polarity?

- How does news media react to controversial events and play a role in the conversation on social media platforms such as Twitter?

As listed above we have chosen a variety of topics in an attempt to answer these questions. The topics belong to spheres ranging from tradition, religion, economy to politics.

1.4 Thesis Roadmap

The roadmap for this thesis is as follows:

- Chapter 2 discusses the Where and How of data collection, while Chapter 3 mentions the basic pre-processing techniques applied to transform the data to fit our usecase.
- Chapter 4 discusses related work.
- Chapter 5 mentions the methodology and tools used for our analysis, including the sentiment analyser, the KDE plot and NRC emotion lexicon. **It also discusses the performance of the sentiment analyser on our data, and how considering shorter tweets helped in enhancing performance**.
- Chapter 6 discusses the experiments done on news media data and common people data and shows the results obtained, simultaneously trying to interpret and rationalize them for the reader.
- Chapter 7 discusses the key observations from an event specific point of view and if we found any patterns across the events.
- Chapter 8 lists down some limitations and potential solutions to those limitations, finally concluding the thesis in Chapter 9.

Chapter 2

Data Collection

To analyse how common people reacted to these events, we crawled tweets from twitter using keyword search. We used the API provided by Henrique [27] to gather the tweets. The API allowed us to scrape tweets with specific keywords and in specific time frames in the past. The tweets crawled are collected according to the time periods of interest, as mentioned below.

2.1 Sabarimala

We crawled tweets with keyword ”sabarimala” for five 3-month periods, spread across two years (2017-2019), to study the changing sentiments of people with respect to the event. The 3-month periods are:

1. Oct-Dec 2017, a supposedly non controversial period, long before Supreme Court’s decision in September 2018. This period also coincides with the days of Mandalapooja festival (around 15 November to 26 December) which is when Sabarimala is the busiest ([83]). Hence this period saw a lot of activity on Twitter.

2. Mar-May 2018, another non controversial period. The activity on Twitter was a lot lower during this time with respect to the previous period.

3. Oct-Dec 2018, right after Supreme Court Judgement (on 29th September, 2018), but before the day (2nd Jan 2019) the two women entered the temple for the very first time.

4. Jan-Mar 2019, right after the two women entered the temple, out-breaking controversy and dialogue on news and social media. This period is one of the two event states.

5. May-July 2019, five to eight months after the controversy. This period to help us analyze the after-event state.

Table 2.1 gives the full picture regarding the number of the tweets gathered with respect to each of the time periods.

Time Period	No. of Tweets
Oct - Dec, 2017	6,445
Mar - May, 2018	770
Oct - Dec, 2018	15,373
Jan - Mar, 2019	74,498
May - July, 2019	12,115

Table 2.1 Sabarimala Twitter Data

2.2 Pulwama Balakot incident

We crawled common people Twitter data with keyword ”modi” for two 2 week periods, before the Pulwama incident which occured on 14th feb and after Balakot strikes which occured on 26th Feb. We did not do it for multiple time periods because Narendra Modi is an omnipresent figure figure on social media and it was difficult to analyze this specific event's affect on the polarity around Modi in other time periods. The aim here is to analyze change in opinion of the masses regarding PM Narendra Modi. The 2 week periods are:

- 2nd to 14th Feb, a supposedly non controversial period, just before the Pulwama bomb blast.
- 26th Feb to 9th March, the 14 days succeeding the Balakot airstrike.

2.3 Demonetisation

We crawled tweets with keyword ”demonetisation” for five 3-month periods, spread across two years, to study the changing sentiments of people with respect to the event. The 3-month periods are:

- 8th Nov 2016 - 7th Feb 2017, right after the demonetisation was announced on 8th Nov 2016.
- May-July 2018, 5-8 months after the event.
- 9th Nov 2017 - 8th Feb 2018, 1 year after the announcement.
- 29th Aug 2018 - 29th Nov 2018, right after Reserve Bank of India's review that reported that 99.3% of the cash has been deposited back, effectively stating the failure of demonetisation.
- Feb 2019 - Apr 2019, during general election campaigning days just before the elections in Apr and May 2019.

2.4 Lynching

We crawled tweets with keyword ”lynching” for four 3-month periods, spread across four years, to study the changing sentiments of people with respect to the event. The 3-month periods are:

Time Period	No. of Tweets
9th Nov'16- 8th Feb'17	711373
May - July, 2017	41461
9th Nov'17 - 8th Feb'18	58464
29th Aug - 29th Nov, 2018	45108
Feb - Apr, 2019	38534

Table 2.2 Demonetisation Twitter Data

- Mar-May 2016, 1 year before the Pehlu Khan lynching incident, assuming this to be the pre-event state.
- Apr-July 2017, just after the Pehlu Khan lynching incident.
- Apr-July 2018, 1 year after the Pehlu Khan lynching incident.
- 21 May - 21 Aug, right after election results, when apparently the number of reported lynchings increased.

2.5 Indian Economy

We crawled tweets which contained both the keywords "india" and "economy" for nine 3-month periods, spread across four years, to study the consistently changing sentiments of people with respect to the indian economy. The 3-month periods are:

- Jan-Mar 2013, 2014, 2015, 2016, 2017
- Jan-Mar 2018, Oct-Dec 2018 (right after RBI report)
- Jan-Mar 2019, May-July 2019 (after election)

2.6 News Data

We also wanted to analyze news media reactions for these events, and hence to gather the tweets from news outlets for the above mentioned events, we crawled tweets from the following news twitter accounts using the same keywords as mentioned in each of these topics above. All of these twitter accounts have followers in hundreds of thousands. Table 2.3 lists down the English News Twitter accounts while Table 2.4 lists the Hindi News Twitter accounts we referenced. The number of followers have been mentioned just to highlight the reach of these channels, and the numbers are from when the experiments were being conducted (December 2019).

News Handle	No. of Followers
$@ani$	3.3 M
$@CNNnews18$	4.3 M
$@DeccanChronicle$	72 K
$@httweets$	7 M
$@indiatoday$	5.3 M
$@ndtv$	11.9 M
$@the\_hindu$	5.7 M
$@timesnow$	9.2 M
$@timesofindia$	12.2 M
$@zeenews$	4.7 M

Table 2.3 English News Twitter Data

News Handle	No. of Followers
$@aajtak$	8.7 M
$@abpnews$	8.8 M
$@amarujalanews$	995 K
$@bbchindi$	1.5 M
$@ddnewshindi$	746 K
$@live\_hindustan$	203 K
$@jagrannews$	882 K
$@navbharattimes$	836 K
$@news18india$	1.4 M
$@zeenewshindi$	2.4 M

Table 2.4 Hindi News Twitter Data

Chapter 3

Data Preprocessing

To make preprocessing easier, the tweets were separated into three groups based on the language of content, English, Hindi and Hindi-English code-mixed. Tweets containing words from any other language were removed. This separation was done with the help of the language identifier by Bhat et al. [11] which can identify words from seven Indian languages viz. Hindi, Bengali, Kannada, Tamil, Malayalam, Telugu and Gujarati apart from English. Please find the step by step data processing flowchart in Figure 3.1.

All three groups of tweets were preprocessed separately as follows:

3.1 Preprocessing English Tweets

These tweets contain only English words, and do not contain words of any other language recognised by the language identifier.

1. Remove special characters, username mentions (beginning with @), urls.
2. We also remove hashtags because most hashtags were a concatenation of words which was not of any use to us.

3.2 Preprocessing Hindi Tweets

These tweets contain only Hindi words, and do not contain words of any other language recognised by the language identifier.

1. Remove all special characters, username mentions (beginning with @), and urls.
2. We also remove hashtags because most hashtags were a concatenation of words which was not of any use to us.
3. Translate the whole tweet to English using python freeware translate (Han [25]).

3.3 Preprocessing Code-Mixed Hinglish Tweets

These tweets were in roman script, with Hindi words written in Roman script as well.

1. Use language identifier (Bhat et al. [11]) to find all the Hindi words.
2. Use transliterator (Bhat et al. [11]) to transliterate the Hindi words from Roman script to Devanagri script.
3. Translate Hindi words in Devanagri script to English using python freeware translate (Han [25]).

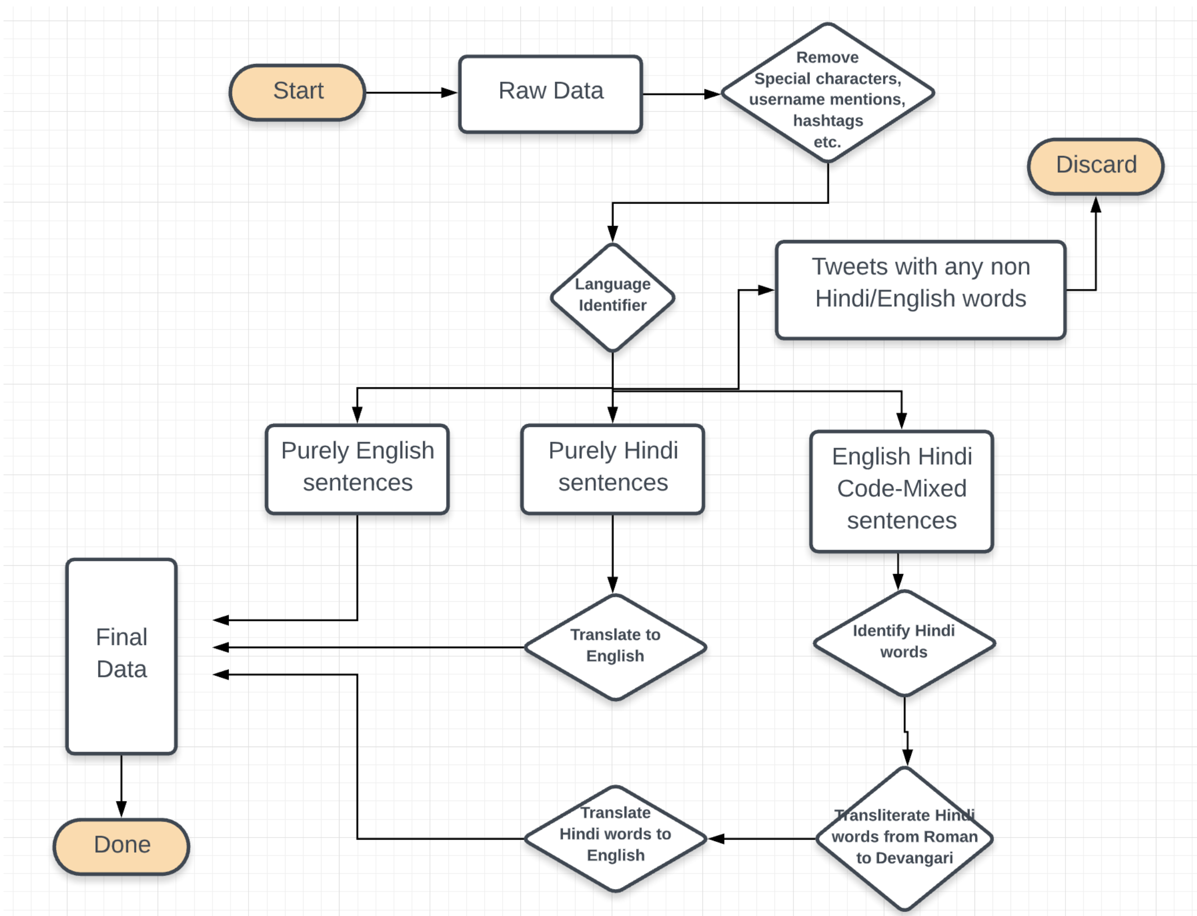

Figure 3.1 Data Preprocessing Flowchart

Chapter 4

Related Work

The Hidden Tribes of America [16], is a 2018 study on polarization in the United States of America. They survey the people of America on major issues that split the country such as, immigration, white priviledge, sexual harassment and Islamophobia. One the basis of this survey they divide the people of America into seven different groups (tribes), from extreme left to extreme right. They state that a majority of Americans, whom they've called the "Exhausted Majority," are fed up by America's polarization. These people constitute the moderate left, centre and the moderate right groups who believe that they have more in common than that which divides them. They believe that their differences aren't so great that they can't work together and are extremely tired of the tribalism attitude from the extreme sides of the spectrum, namely the extreme left and extreme right. Tribalism leaves no room for compromise and eventual middle ground solutions, it only leaves people who are not even willing to acknowledge or understand the reasoning behind the views of the opposer. The exhausted majority believes that the people belonging to the extremes no longer just disagree, they reject each other's premises and doubt each other's motives and they question each other's character and block their ears to diverse perspectives, effectively constituting a metaphorical (and isolated) tribe.

Our work, for a small part, investigates the nature of opinions on social media in India. We have tried to understand if anything like "the exhausted majority" exists in India.

In the book *From Media Hype to Twitter Storm. News Explosions and Their Impact on Issues, Crises, and Public Opinion*, Vasterman [82] explores the effect of news waves on current issues and public opinion, and shows that media hype is a creation of both news media and social media, and they are extremely co-dependent on each other. In one of the very few works on the India subcontinent, *A Computational Analysis of Polarization on Indian and Pakistani Social Media*, Tyagi et al. [81] examine polarizing messaging on Twitter during the Pulwama-Balakot incident period, particularly focusing on the positions of Indian and Pakistani politicians. Their analysis reveals that politicians in the ruling political party in India (BJP) used polarized hashtags and called for escalation of conflict more so than politicians from other parties. Their work offers the first analysis of how escalating tensions between India and Pakistan manifest on Twitter. In our work we do analyse how sentiment towards Narendra Modi changed (become more positive) after the airstrike on Balakot.

In his work, *The variable nature of news media influence*, Gene Zucker [20] presents a theory of variable news media influence, which states that the less obtrusive an issue is, and the less time the issue has been prominent in the media, the greater is the news media's influence on opinion about that issue. They establish news media's influence and our work tries to understand that as well.

In other works, Garibay [19] discusses how polarization in social media assists influencers to become more influential, and thus could be used as a tool to gain popularity and Barberá [6] talks about how echo chambers help in creating polarization on social media.

Mahapatra and Plagemann [45], in their work *Polarization and Politicisation: The Social Media Strategies of Indian Political Parties* talk about how political parties in India are using social media to influence voters which brings more significance to understanding polarization on social media in India. Neyazi [57] discusses how polarization doesn't just exist in election times but continues to thrive outside of election campaign periods among certain groups and is becoming much more evident in daily conversation on social media. Our work here is an effort to establish such trends among multiple topics across periods and hopefully brings more light to the whole debate of polarization in the India context.

Chapter 5

Methodologies and Tools Used

5.1 Sentiment Analysis

Sentiment analysis is a field of study that focuses on examining the perceptions, assessments, behaviours, evaluations, and emotions of individuals towards different objects (Liu [42]). Such artefacts include, but are not limited to, goods, services, persons, organisations, events, problems, subjects, and their characteristics. Sentiment analysis became popular with the rapid growth of social media and the popularity of opinion-rich resources in the early 2000s (Feldman [18]). It helps us understand people's opinions on everything from services to political ideologies.

There is a number of approaches to automated sentiment analysis (Yusof et al. [88]). Lexicon-based approaches depend heavily on the quality of the sentiment lexicon used (Feldman [18]). Supervised machine learning approaches do not require any sentiment lexicons, but they rely on labelled training data (Pak and Paroubek [64]). From many different classification algorithms, SVM (support vector machine), Naïve Bayes, or Maximum Entropy are widely used and provide the best performance (Liu [42], Medhat et al. [48], Zhang et al. [89]). When there are no labelled training datasets available, semi-supervised or unsupervised (as in our case) approaches are used (e.g. He and Zhou [26], Xianghua et al. [87]). Other researchers have tried different approaches utilizing neural networks (Moraes et al. [54]), decision tree structures (Hu and Li [33]), or rule-based techniques. Some authors (e.g. Prabowo and Thelwall [68]) experimented with multiple classifiers used in a hybrid manner to achieve a better classification performance than any individual classifier. In recent years, deep learning approaches have gained a foothold in the context of sentiment analysis (e.g. Habimana et al. [24]).

The performance of models of sentiment analysis usually varies from about 70-80% of correctly classified instances. In real world applications/context, most of the methods fail to keep their promises and are much less precise, as users use sarcasm, irony, or metaphorical phrases pretty frequently (Poria et al. [65]), specially on social media platforms like Twitter. Some authors caution against chasing unreasonable standards of classification accuracy because it can be difficult to evaluate sentiment in the same way as human beings through automated methods, as even human raters do not agree in about 20 percent of cases (Ogneva [62]). Domain and topic dependence is another sentiment analysis problem that is to be addressed more thoroughly in research Kincl et al. [39].

There are also some major challenges to be addressed in multi-lingual sentiment analysis research. Much of the research has been conducted in English (or more recently in other major languages such as Chinese). While there has been work done in Hindi and some other Indian languages, the research is still developing and not robust enough. In case of sentiment analysis on Hindi-English code-mixed data, there's no direct tool available.

In case of scarcity of resources, several methods of approaching sentiment analysis in multiple languages have been suggested in previous studies. The most common approach is to adapt English tools and resources and translate them into new languages using bilingual dictionaries, machine translation, cross-lingual projections, or monolingual and multilingual bootstrapping (Balahur et al. [4]). English sentiment lexicons were translated to other languages in cases such as, Chinese (Ku et al. [40]) and Romanian and Spanish (Banea et al. [5]). Other studies experimented with translating the analysed text into English and then using the tools developed for English to determine sentiment. Languages like Arabic, Chinese, French, German, Italian, Japanese, and Korean were part of these studies (Bautin et al. [7]). We'll be using this exact approach where we'll be translating Hindi and code-mixed Hindi-English to English and use English tools for analysis. Joshi et al. [38] in their work implemented this approach to achieve a 65% precision, which is similar to what we achieve as well.

Since our data is across domains, multi-lingual and not manually labelled, we were looking for a widely accepted and used English language tool which we could use directly on our data.

Textblob (Loria [44]) is one of the most popular quick to use open-source natural language processing (NLP) libraries for English language available online. It is a Python (version 2 and 3) library for processing textual data. It provides a simple API for diving into common NLP tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

The sentiment analysis module in Textblob is a pre-trained Naive Bayes model. The sentiment analyzer grades the polarity of text on a continuous scale of -1 to 1. It also offers an objectivity and subjectivity score for the text, but that has not been used in our investigations.

Textblob sentiment analyzer has been used extensively towards understanding the sentiment polarity of Twitter data in the context of political events such as the Nigerian Presidential Election, 2019 (Oyebode and Orji [63]) and the Punjab Legislative Assembly Elections, 2017 (Singh et al. [71]). It has also been used to predict social media trends using Twitter data (Munjal et al. [55]) which is what our work also focuses on in the context of polarization. The use of Textblolb as a valuable resource in wide-ranging published work such as analyzing customer reviews on TripAdvisor (Laksono et al. [41]), analyzing social media content to enhance online marketing strategies (Micu et al. [49]), financial lexicon sentiment analysis (Sohangir et al. [72]) and general social media clustering analysis (Ahuja and Dubey [1]) reinforces the robustness of the tool.

5.1.1 Kernel Density Estimation

In our experiments, we have have taken the help of Kernel Density Estimation (KDE) to portray some of our results. In statistics, KDE is a non-parametric way to estimate the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample.

The fundamental problem of density estimation is that of infering the density function f, given an independent and identically distributed sample $x_1, x_2, ..., x_n$ from the corresponding probability distribution. The kernel density estimate (KDE) is a particularly simple instance-based method. The idea is to compute the density of a test point as the average of the density of test point under each point in the training data. The estimate is

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right), \tag{5.1}$$

where h is a bandwidth parameter, and the kernel is commonly a Gaussian,

$$k(z) = \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}z^2). \tag{5.2}$$

5.1.1.1 Example

Kernel density estimations try to mimic a histogram, but additionally use smoothing techniques to present the data in the form of a curve. They are helpful in analysis of continuous variables, for example, sentiment polarity in our usecase.

Let's take an example [85] of a dataset containing six data points $-2.1, -1.3, -0.4, 1.9, 5.1, 6.2$. We represent these six datapoints as a histogram and as a smoothed out kernel density curve in Fig. 5.1.

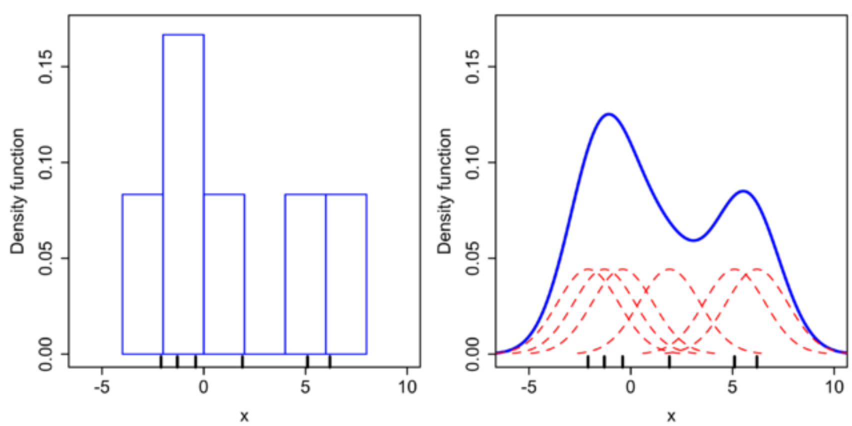

Figure 5.1 Histogram vs. Kernel Density Curve

The total area under the kernel density curve is always equal to 1, that is, if you take an integral of the curve, it'll be equal to 1. The x axis represents the data, and the y axis represents the probability density. Probability density is not the same as probability and can be greater than 1.

As you'll see, we have used KDE curves in our reports to analyze and compare the polarity of the Twitter dataset over different time periods where the x axis would range from -1 to 1. These plots help us understand the prevalence of sentiment scores in a range with respect to other ranges. Say, if

Textblob Range	Annotation Label
$>= 0.1$	Positive
$(-0.1, 0.1)$	Neutral
$<= -0.1$	Negative

Table 5.1 Textblob Range Mapping

the curve rises the highest in range 0.2 to 0.4, we can assume a prominence of tweets with sentiment score falling in this range. Similarly if the curve rises higher in the positive range (> 0.0) than in the negative range (< 0.0) we can assume a prominence of positivity, or a greater percentage of positive tweets. Another example where these curves help is when we want to compare two different datasets, for instance, say curve 1 represents the data obtained in time period 1 (T1) and curve 2 represents the data obtained in time period 2 (T2). Then, curve 1 when compared against curve 2 (a similar kdeplot of a different time period T2), would also help us compare the popular sentiment ranges across those two time periods, for example, if curve 1 rises above curve 2 in negative zone and curve 2 rises above curve 1 in positive zone, then we can assume that in the time period represented by curve 2 (T2) there was more positivity than in time period represented by curve 1 (T1). We'll see many examples of this in the Experiments and Results section.

Kernel density estimations have been used extensively in the field of Natural Language Processing as evidenced by Hernandez-Suarez et al. [28] where the authors use Twitter Data to monitor natural disaster social dynamics using a recurrent neural network approach with word embeddings and kernel density estimation, Gerber [21] where the author conducts linguistic analysis on twitter data to predict crime, Liu et al. [43] where the authors propose a small sample text classification algorithm based on kernel density and Hulden et al. [34] where the authors investigate common methods for determining the geographic point of origin of a text document by kernel density estimation.

5.1.2 Analyzing the Effectiveness of Sentiment Analysis tool TextBlob on Twitter Data

Here, we'll try to analyze the effectiveness of sentiment analysis tool, Textblob, on our data. As mentioned above, we are using Textblob's tool to find the polarity of the tweets. Since Textblob is a black box we did not have much control over the analysis, thus, we tried to analyze it's output with manual evaluation. Textblob's sentiment analyzer's polarity score is a continuous variable lying between -1 to 1. To make the job of the manual evaluators practical, we asked the annotators to rate the sentiment of a tweet in three categories - positive, neutral and negative. For the convenience of comparison, Textblob's polarity score range was divided into three parts ($<= -0.1$, -0.1 to 0.1, $>= 0.1$) to correspond to the three labels used in the manual annotation. The range division and corresponding labels can be observed in Table 5.1.

The accuracy of the sentiment analyzer stood at a meagre **56%**. This could be because of multiple reasons, such as:

1. The tool used is generic and has been trained on pure English data

	Event	**Tweet**	**Textblob Polarity Score**
T1	Pulwama-Balakot	*"BJP President Amit Shah - In Surat earlier our jawans used to be beheaded and insulted but today situation is such that when our jawan fell in Pakistan while shooting down a fighter plane, within hours he was back. This change is because of Narendra Modi's will power"*	-0.04
T2	Sabarimala	*"Absolutely correct sir these are leftist DK women probably funded by anti Hindu parties, aim to desecrate Sabarimala period. We need to get bjp back to power in and for that we need Hindus to unite and stay united please help Jai Hind"*	0.05
T3	Pulwama-Balakot	*"West Bengal CM - We are against Modi and BJP Modi has turned the BJP into a private organisation. Whenever someone says something against Modi babu then that person is branded a Pakistan supporter. My father was a freedom fighter. I will not learn patriotism from them."*	0.0
T4	Pulwama-Balakot	*"Derek O Brien, TMC - Mr. Narendra Modi, are you sending our soldiers to die without plan or is your purpose is only to win polls. Mr Modi, you shamelessly reduce precious pictures of martyred jawans to use as a backdrop to your political rally Mr Modi you are quite shameless"*	0.325

Table 5.2 Tweets with no limit on word count

2. Our data is a collection of pure English, transliterated and translated Hindi and English-Hindi code mixed data

3. The data has been labelled into three classes making it a multi-class classification problem

At the time of this analysis, there were no better tools, mechanisms and options in our knowledge that could work better on such complex data. We did not need a state of the art sentiment analyzer, but a good enough one because the subject at hand and the volume of the tweets that we had allowed us some freedom in this context.

We decided to delve deep into the data and figure out the reasons for the disagreements between Textblob's polarity scores and the manual annotation. We noticed a pattern. Let's look at some examples:

The tweets listed in Table 5.2 are labelled incorrectly. In $T1$, as we can see, the tweet talks positively about Narendra Modi, but is still labelled negative, though very close to neutral. Similarly $T3$ and $T4$ are negative towards Modi, but are either labelled positive or neutral. In $T2$, the tweet is against the event happening at Sabarimala, but is labelled as positive (close to neutral). We hypothesise that this is because of the length of the tweets. Even though the tweets contains very explicit adjectives against/for Modi or Sabarimala, the content in the rest of the tweet confuses the sentiment analyzer to give wrong results.

	Event	Tweet	Textblob Polarity Score
T1	Pulwama-Balakot	*"Good news for commuters, PM Modi launches OneNationOneCard for Pan India transport mobility."*	0.7
T2	Pulwama-Balakot	*"PM Modi has misplaced priorities desperate only for re election : Congress"*	-0.26
T3	Pulwama-Balakot	*"India is safe under the leadership of PM Modi. Only Modi can give a sense of security to the Indians - BJP President"*	0.0
T4	Sabarimala	*"incredible BBC News, Sabarimala temple Indian women form 620 km human chain for equality"*	0.25

Table 5.3 Tweets with number of words $<=$ 30

Tweets Type	Accuracy
No condition	56%
$<=$ 30 words	65%

Table 5.4 Accuracy Improvement

Let's look at the tweets listed in 5.3. All these tweets are labelled correctly. The other difference is that all these tweets have much lesser word count than the other table. The explicit emotion/polarity expressed is not lost in the noise of the rest of the tweet. In an ideal world, this problem can be solved by aspect based sentiment analysis, and shortening the length of the tweets helps us come closer to aspect based sentiment analysis.

Looking at the data and these results, we decided to evaluate the performance of our sentiment analyzer on tweets whose word count is lesser, taking word count less than equal to 20 and 30 in different sets. We realized the data was reducing by a lot in the case when the word count was less than 20. Hence, to keep a balance we decided to do the manual evaluation of our sentiment analyzer on tweets with word count less than 30. The resulting accuracy stood at **65%**, which is an increase of **9%** from the dataset of tweets with no condition on word count, as shown in Table 5.4.

In some cases, like Pulwama-Balakot incident, the tweets with lower word count even helped us in gaining more clarity from our results.

As demonstrated in Fig. 5.2, this was the initial Kernel Density comparison obtained with no limit on number of words in the tweet. Here we want to understand the difference between the "before" and "after" scenarios of an event, Pulwama-Balakot in this case. While it seems clear that the "after"/"green" curve is higher on the positive side (covers more area) of the x axis and lower on the negative side of the axis, the differences are minimal.

We did the same experiments by considering the tweets with 20 and 30 word limit respectively. As you can see in Figures 5.3 and 5.4, the difference between the before and after curves increases (becomes more explicit) as we reduce the word limit on the tweets. In Figure 5.4, the green curve can be seen to

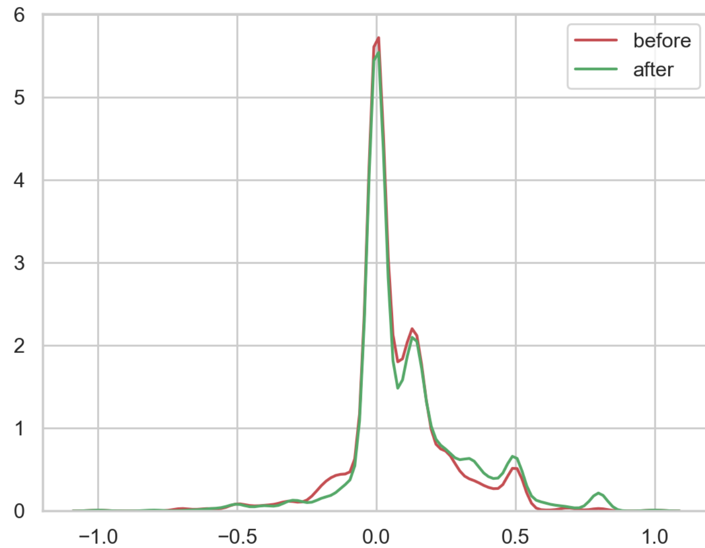

Figure 5.2 Comparison using tweets with no condition on word length - Context: Pulwama-Balakot

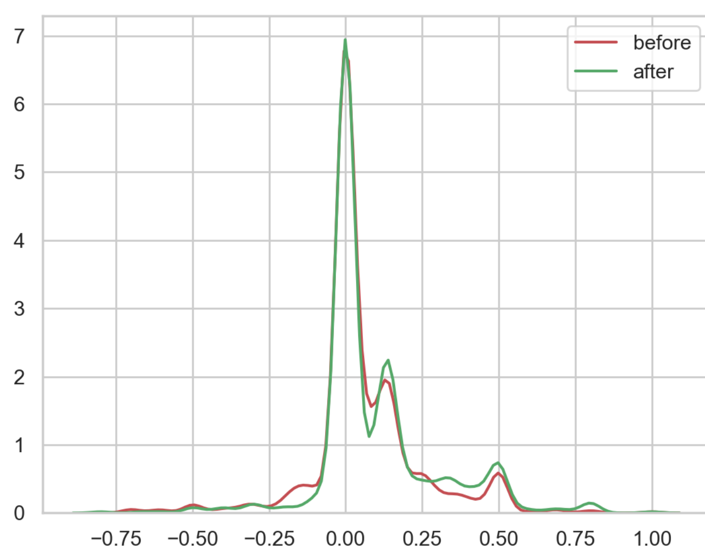

Figure 5.3 Comparison using tweets with word limit: 30 - Context: Pulwama-Balakot

cover much more area on the positive side wrt. the red curve and much less area on the negative side, thereby providing more clarity and margin of error to our interpretation in Figure 5.2.

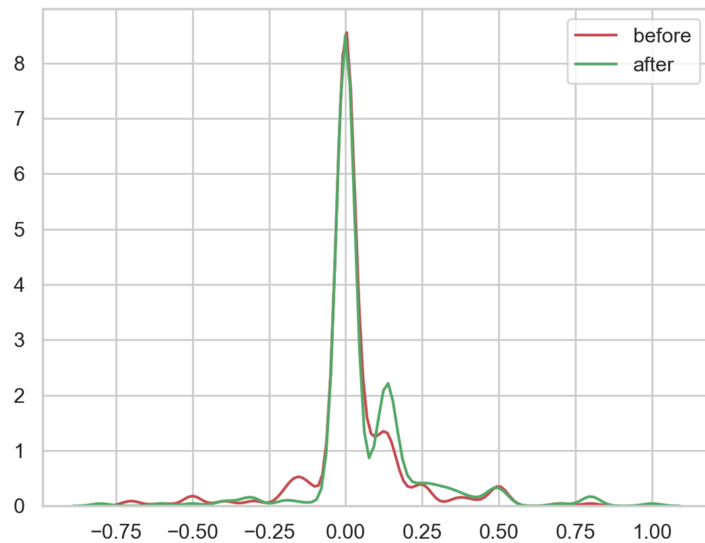

Figure 5.4 Comparison using tweets with word limit: 20 - Context: Pulwama-Balakot

Given that we are working with unsupervised data, the 65% accuracy seemed good enough to move forward. While this is not state of the art accuracy, but sentiment analysis for Hindi and sentiment analysis for code mixed data are still very much in ongoing research, and deemed out of the scope of this project.

5.2 Emotional change analysis through the different periods

We have used the NRC Emotion Lexicon *(version 0.92)* by Mohammad and Turney [53] to find the emotions expressed and their change through the different periods. The NRC Emotion Lexicon is a labelled dataset that associates words with emotions, which can be of eight types - anger, sadness, disgust, joy, surprise, trust, anticipation and fear. As a bonus, it also associates each word with either positive or negative connotation. Each word is associated with one or more of the eight emotions. The analysis counts the total number of words with each emotion over the whole corpus.

The authors have used *Roget's Thesaurus* [70] as the source of the terms. With the help of five annotators, the authors have annotated a lexicon containing $24,200$ word sense pairs. The information from different senses of a word is combined by taking the union of all emotions associated with the different senses of the word. This resulted in a word-level emotion association lexicon for about $14,200$ word types. Each word could have more than one associated emotion.

The NRC lexicon is a heavily used resource having been referenced extensively in the domain of NLP and social analysis as evidenced by Vosoughi et al. [84] where the authors study the spread of false news online, Mohammad and Kiritchenko [52] where the authors try to understand emotions expressed in the tweets on Twitter using the hashtags used, Tang et al. [73] where the authors develop a deep learning Twitter sentiment classifier, Mohammad [51] where the author works on detecting valence, emotions and other affectual states from text, Bravo-Marquez et al. [13] where the authors develop meta-level sentiment models for big social data analysis and so on.

The experiments and analysis using the lexicon can be found in Chapter 6.

Chapter 6

Experiments and Results

6.1 Scrutinizing News Media

Analyzing print media was one of the first steps of this study. We wanted to understand the role media plays in any kind of opinion formation, or opinion polarization of the masses. One of the directions that we looked at was understanding if there was any difference in the content of disseminated news if the languages were different. Does news media change the tone of their news articles depending on the language of the audience, deliberately or just by chance? Let's look at a few examples.

We first looked at articles from the The Times of India - an English newspaper, and Navbharat Times - a Hindi one. Both of these newspapers belong to the same media house. This helped us minimise the variables that might play a role in the tone of reporting. Being from the same organization, we assumed that the source of their information would be similar or same, and any deliberate or non-deliberate difference of ideology would be minimal.

Let's look at Times of India's reporting to the two Sabarimala events - the Supreme Court decision, and the event when two women entered the temple, right on the day after the events happened.

The Times of India's headlines were:

- 29th Sept, 2018 - "*SC smashes gender barrier, opens Sabarimala for all women*" [59]
- 3rd Jan, 2019 - "*Women scale one more summit; taken to Sabarimala in commando-raid style*" [60]

The Navbharat Times' headlines were:

- 28th Sept, 2018 - "*sabarimala mandir: tooti varshon purani parampara, isliye mandir me mahilaon ke pravesh pe thi rok*". A rough tanslation would be: "Sabarimala temple: Centuries old tradition broken, let's see why women weren't allowed inside the temple" [76]
- 3rd Jan, 2019 - "*Sabarimala: Ayappa ke darshan kar do mahilaon ne yu racha itihaas, kuch logon ko hi thi jaankari*" Sabarimala: See how two women created history visiting Ayappa, only few people knew about this. [77]

Times of India's reporting uses extremely positive phrases such as "smashes barrier" and "scale summit". On the other hand Navbharat Times had a mixed response. In one case they do mention "history being made" but in the other case they address the situation as "breaking of *parampara*" which is an extremely negative connotation. While we couldn't find any such pairs of different languages newspapers in the same media house with data from last years available online, we looked at a few others for whom data was available online such as The Hindu and Dainik Bhaskar. On 3rd Jan, The Hindu, an English newspaper, headlined with "*Two women make it to Sabarimala*" [31] while Dainik Bhaskar, a Hindi newspaper, addressed the issue with a "*800 saal purani parampara tooti*",[10] which roughly translates to "800 years old tradition broken".

There is indeed a difference of connotation in these cases. The English newspapers sound celebratory, while the Hindi newspapers mixed and sometimes with negative connotation. Since these are too few examples to base our judgements on, we decided to do it on a bigger scale where we could look for patterns and base our judgements with more evidence. To improve the scale we decided to scrape the Twitter accounts of popular English and Hindi media channels and newspapers. This allowed us to have quick and easy access to a large amount of data and was in line with out purpose of understanding cross-lingual change of intonation by news media.

We collected all tweets with keyword "Sabarimala" from the 10 English and 10 Hindi news Twitter accounts, as mentioned in Chapter 2, for the period of Jan to March 2019. The results can be seen in Figure 6.1 and Table 6.1.

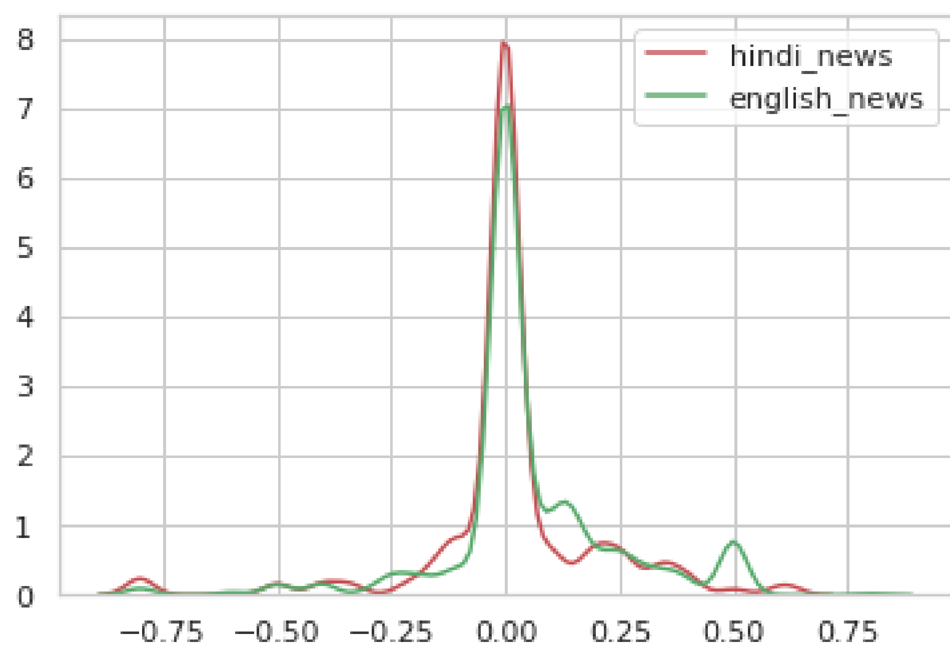

Figure 6.1 Hindi news vs. English news during the period of Jan to Mar 2019 containing keyword Sabarimala

	positive %	neutral %	negative %
English News	26	66	8
Hindi News	17	73	10

Table 6.1 English news vs. Hindi news for Jan-Mar 2019 (Sabarimala)

As you can see, the curve representing the English news data (green) in Figure 6.1 covers more area than the red curve on the positive side and lesser area than the red curve on the negative side. Table 6.1 represents the same data in tabular format, and English new data has 26% positive tweets compared to 17% for Hindi. The negative tweets stand at 10% for Hindi news while 8% for the English news.

These results do point towards a contrasting reaction to the event by different language news outlets. Are news outlets catering to the audience? Do English news outlets tend to serve their customer who might be more liberal in nature, and stands for gender equality, than the Hindi news customers who might be more conservative in nature, being more traditional and religious? We don't think we can definitively answer such questions, but these results do point in a certain direction.

A further analysis shown in Figure 6.2 compares the news twitter polarity with the polarity of tweets tweeted by common people during the same period. The blue line that shows the polarity of common people tweets has the least percentage of neutral tweets. For most of the non-neutral part the curve lies between the hindi news and english news twitter polarity. Almost seems like the Hindi and English news media fall on opposite sides to what an average common person thinks.

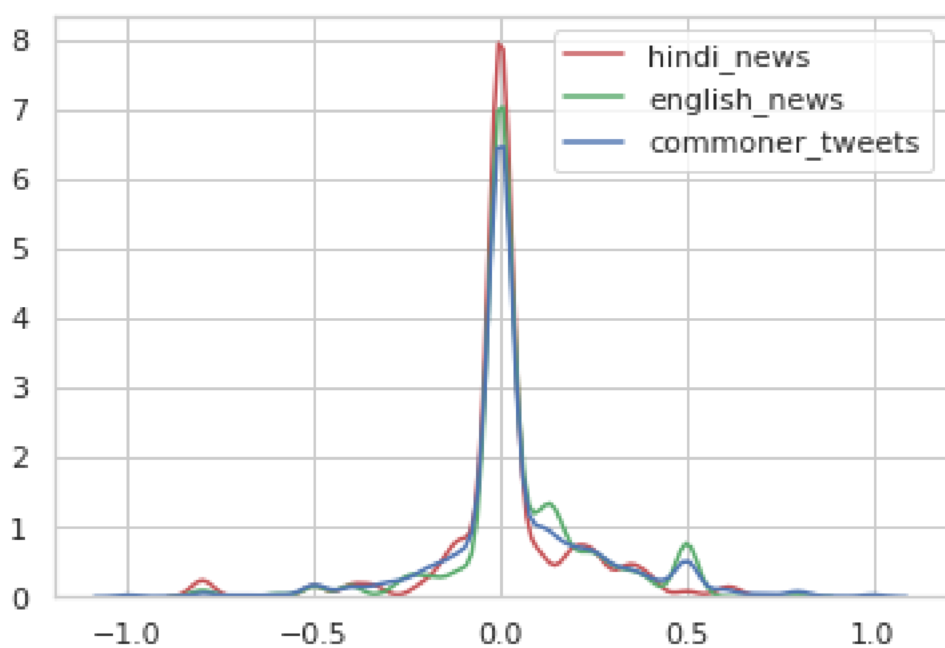

Figure 6.2 Hindi news vs. English news vs. Commoner Tweets

The topic of Sabarimala, which becomes more of a question of tradition vs. gender equality is suited for such an analysis because of the clearly distinct conservative vs. liberal values involved. The same analysis is difficult to do with the other topics, such as demonetisation, economy, lynchings etc., that we have covered in this thesis. The experiments carried out for those events didn't yield relevant results, and probably rightly so.

6.2 Scrutinizing Common People Data

When we say "Common People Data", we refer to the twitter activity of everyone except news media. We'll explore our experiments and results around the five topics that we have collected the data for.

6.2.1 Sabarimala

Figures 6.3, 6.4, 6.5, 6.6 and Table 2.1 are some of the graphs and figures obtained from experiments done on Sabarimala twitter data spreading over five time periods. Let's analyse them.

1. The activity on Twitter increased multifold after the two women entered the temple early morning 2nd Jan, as seen in Figure 6.3. The activity on twitter reduced drastically in the period of May to July 2019, i.e. 5-8 months after the incident, similar to how it was before the event. The proportion of positive, negative and neutral tweets are fairly similar in the first four periods, where the neutral percentage is higher than the percentage of positive and negative tweets individually, except the last period of May-July 2019. The neutral tweet count decreases disproportionately in the last period, after the event, where the neutral tweets are lesser than the positive tweets and just more than the negative tweets.

2. Figure 6.4 shows the kernel density estimations of tweets for different time periods showing polarity from -1 to 1. Table 6.2 breaks down the kernel density estimations in tabular format. As a reminder, we are considering tweets with sentiment $>= 0.1$ as positive, $<= -0.1$ as negative and neutral tweets between -0.1 to 0.1. The neutral tweets are around 65-70% for the two non-controvertial periods Oct-Dec 2017 and Mar-May 2018. The percentage of tweets with neutral polarity decrease to around 61% in the controvertial periods i.e. Oct-Dec 2018 and Jan-Mar 2019. The percentage of neutral tweets decreases to just around 55% in the last period of May-July 2019. This drop in neutral tweets was also witnessed in Figure 6.3 as well. It is also noteworthy that the blue curve, denoting tweet polarity percentage in the last period (after event) rises the highest above all other curves as we move away from the neutral zone. This is not a surprise as the area under the KDE plot is always 1, so if the neutral part has decreased it'd have increased towards the polar sides and thus the non-neutral part of the curve rises the highest for that period. That is to say, more percentage of tweets are polar than before, also shown in Table 6.2.

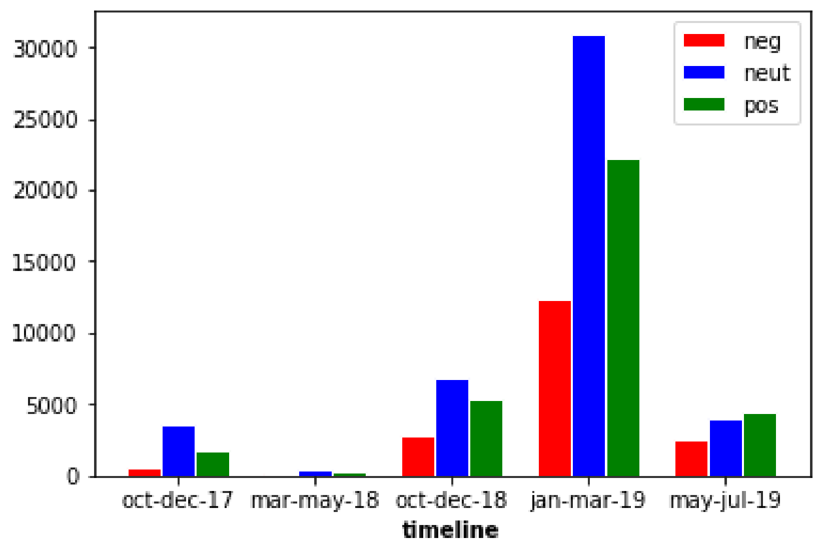

Figure 6.3 Sabarimala Tweet Count

	positive %	**neutral %**	**negative %**
Oct-Dec, 2017	25	68	6
Mar-May, 2018	27	65	7
Oct-Dec, 2018	26	61	12
Jan-Mar, 2019	25	62	12
May-July, 2019	29	55	14

Table 6.2 Tabular breakdown of the polarity across different time periods (Sabarimala) - extension of Figure 6.4

3. Figure 6.5 compares the sentiments of people w.r.t., Sabarimala in a non-controversial period with controversial period. As we can see the neutrality decreases significantly and opinions increase on both sides.

4. Figure 6.6 shows the common people tweets categorised by emotion.The last three periods show increased anger, sadness and negativity.

Rationalising the above results for Sabarimala

- Overall negative sentiment increases after the Jan 1 incident, also seen in the grouped emotion bars (6.6) which show increased anger, sadness, negativity.

- Opinions are polarized due to the event (on both sides), in comparison to a non-event (non-controversial) normal state. The graph shows a significant decrease in the neutral population in the subsequent periods.

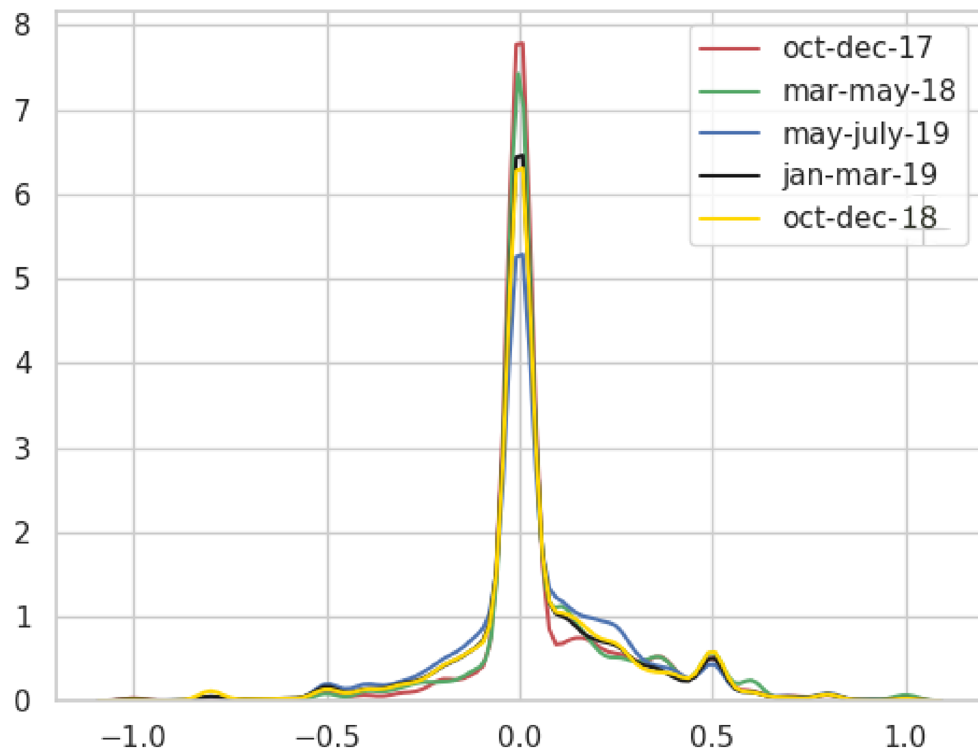

Figure 6.4 KDE plots comparing polarity across different time periods (Sabarimala)

- 5-8 months after the polarizing event, the after-event state, is very different from the pre-event state. The number of tweets with neutral opinions have died down and the tweets with polar opinions are proportionately higher. While it was intuitive to see that the event state evoked polar opinions, but it was not intuitive that in the after-event state the polar opinions remained.
- Comparing Hindi vs English news channels for Jan-Mar 2019 shows that english channels are more positive and Hindi news channels more negative.

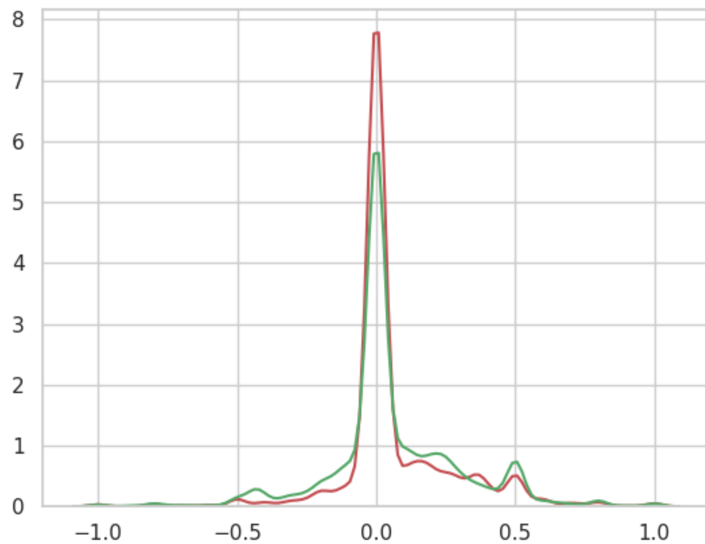

Figure 6.5 Comparing the sentiments of a non-event state (Oct-Dec 2017, red) with an event state (Jan-Mar 2019, green)

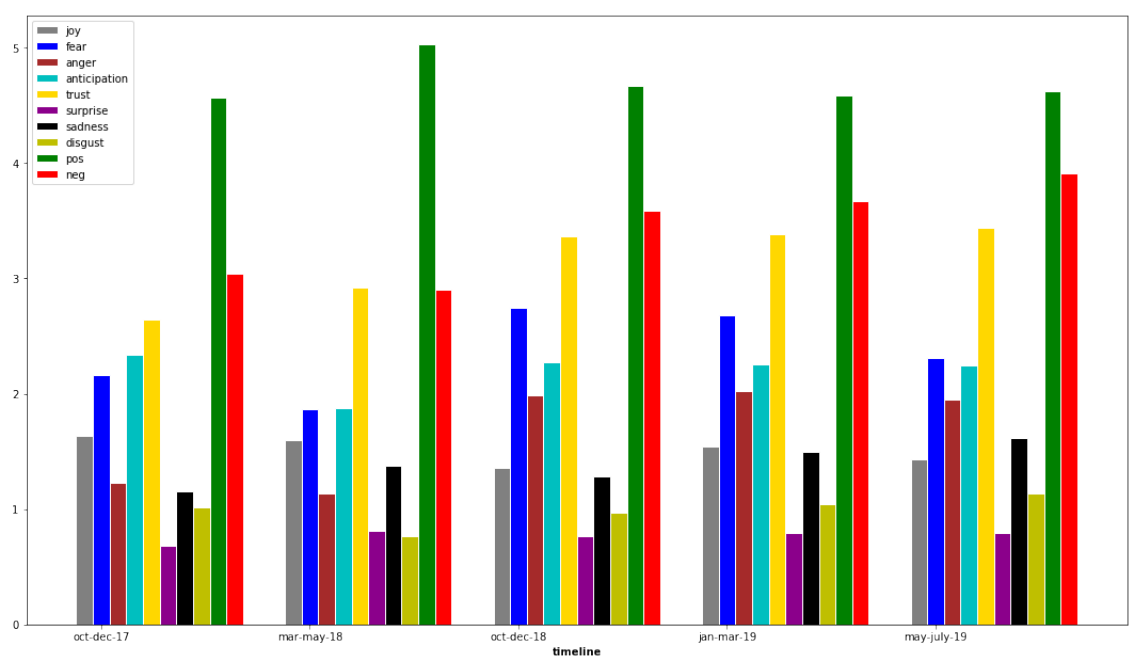

Figure 6.6 Emotion category Sabarimala

6.3 Demonetisation

Figures 6.7, 6.8, 6.9, 6.10 and Table 6.3 are some of the graphs and figures obtained from experiments done on Demonetisation twitter data spreading over five time periods. Let's analyse them.

1. As seen in Figure 6.7, the activity on Twitter was maximum in the first period, right after the announcement of the demonetisation. The activity decreased significantly in the next period and remained constantly similar in the remaining periods. Interesting to note is that, like Sabarimala, the neutral tweets have decreased disproportionately in the last two periods.

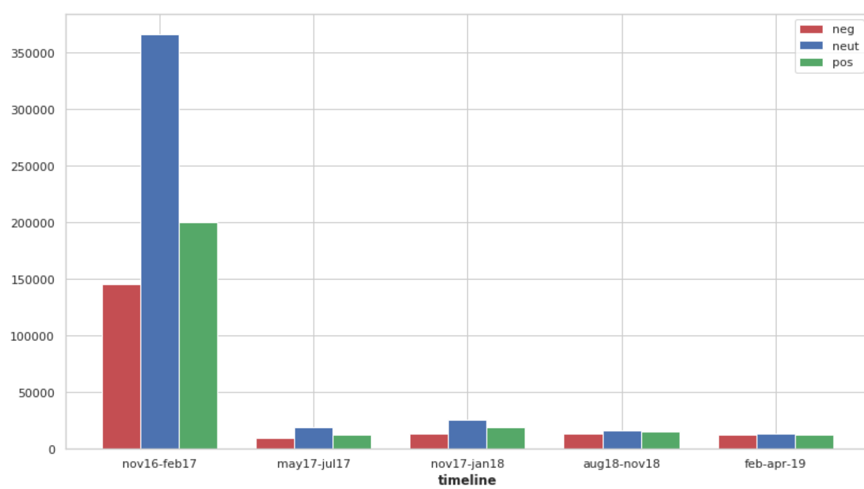

Figure 6.7 Demonetisation Tweet Count

2. In Figure 6.8 and the corresponding Table 6.3 we can see that the percentage of tweets with neutral sentiment decrease in the later periods from 62% in the period Nov 2016 - Feb 2017 to 53% in the last period, i.e. Feb to Apr 2019. As the percentage of neutral tweets decrease, the curve for the final timeline (black) rises highest in the non-neutral zones towards both sides.

	positive %	**neutral %**	**negative %**
Nov'16-Feb'17	22	62	15
May'17-Jul'17	23	58	17
Nov'17-Jan'18	24	58	16
Aug'18-Nov'18	23	54	21
Feb'19-Apr-19	23	53	22

Table 6.3 Tabular breakdown of the polarity across different time periods (Demonetisation)

3. Figure 6.9 shows increased sadness, anger and negativity in the last two periods wrt. the tweets of common people. It's fairly intuitive because the negative effects of demonetisation came to

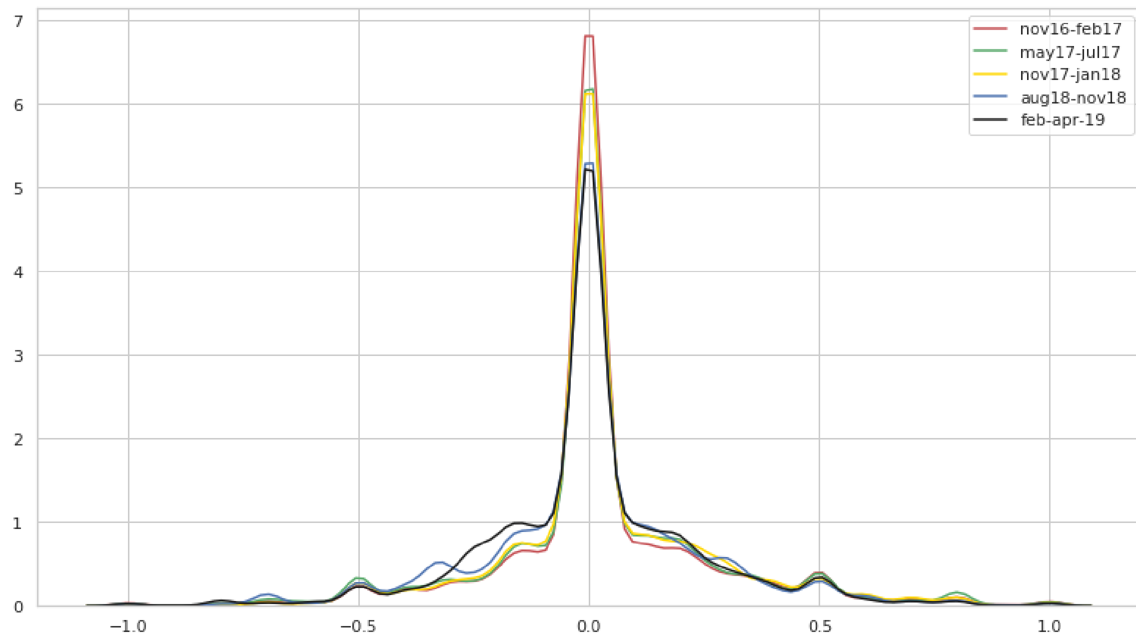

Figure 6.8 Demonetisation Gauss curve

light more and were talked about more after the RBI announcement and during the Lok Sabha elections.

4. Figure 6.10 shows the number of tweets for news media outlets in various timeperiods. The positive portion has almost died down in the last four periods, unsurprisingly.

Rationalising the above results for Demonetisation

- The tweets in the first period, i.e. right after when the event happens, are huge in number and decrease drastically in the subsequent periods.
- The disproportionate decrease in neutral tweets in the last two periods catches our attention, as it is similar to what happened in Sabarimala.
- The graph shows an increase in percentage of tweets towards negative and positive sides, signalling increase in percentage of polarized reactions in after-event states.
- Negativity, sadness emotions have increased through the timelines with the maximum negativity and sadness being in the election period.
- News media's reaction is fairly similar to how common people reacted in this instance.

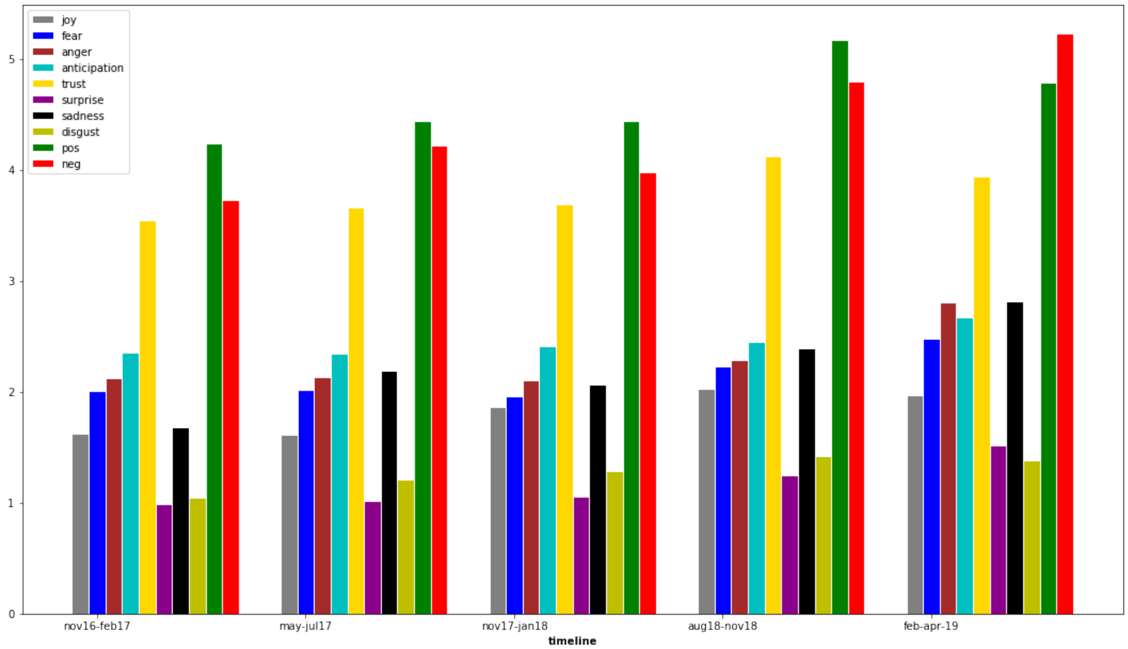

Figure 6.9 Demonetisation Emotion Categorization common people twitter

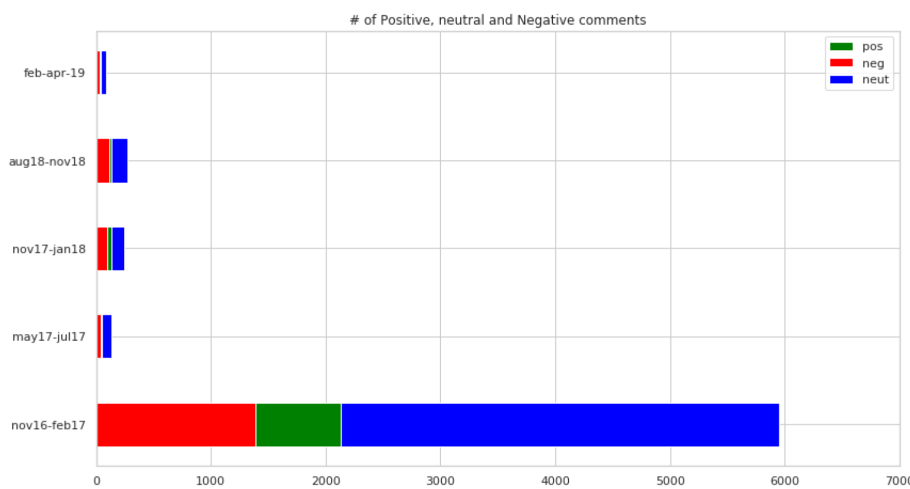

Figure 6.10 Demonetisation Tweet Count for news media outlets

6.4 Pulwama Balakot incident

Figures 6.11, 6.12, 6.13 and Tables 6.4, 6.5 are some of the graphs and figures obtained from experiments done on the Pulwama-Balakot twitter data spreading over two time periods. Let's analyse them.

1. In Figures 6.11 and 6.12 (zoomed version of Figure 6.11), red curve shows the sentiment of common people before the Pulwama attack, and green curve shows the sentiments after the Balakot airstrike. The Table 6.12 breaks down the above kde plots into numerical figures. As you can see in the figures the green curve is lower towards the negative side and higher towards the neutral and positive sides, also confirmed by the table where the negative % has decreased and the positive and neutral % have both increased. It seems as if the the polarity has shifted from the negative side to the neutral and positive sides.

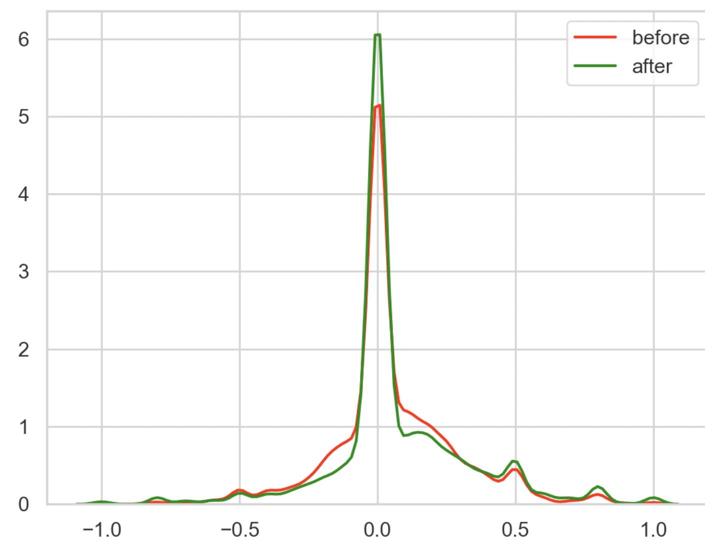

Figure 6.11 Pulwama-Balakot incident before (red) vs. after (green) commoner tweet sentiment

	positive %	**neutral %**	**negative %**
before Pulwama	30	53	15
after Balakot	31	57	11

Table 6.4 Tabular breakdown of the polarity before Pulwama and after Balakot

2. In Figure 6.13, we compare news outlets' twitter sentiment and observe that after the Balakot strikes the positive sentiment has increased. The Table 6.5 shows the % wise breakdown, which clearly shows that news media twitter sentiment towards Modi has increased towards the positive direction, which is much more prominent compared the change in people's opinions captured in Table 6.12.

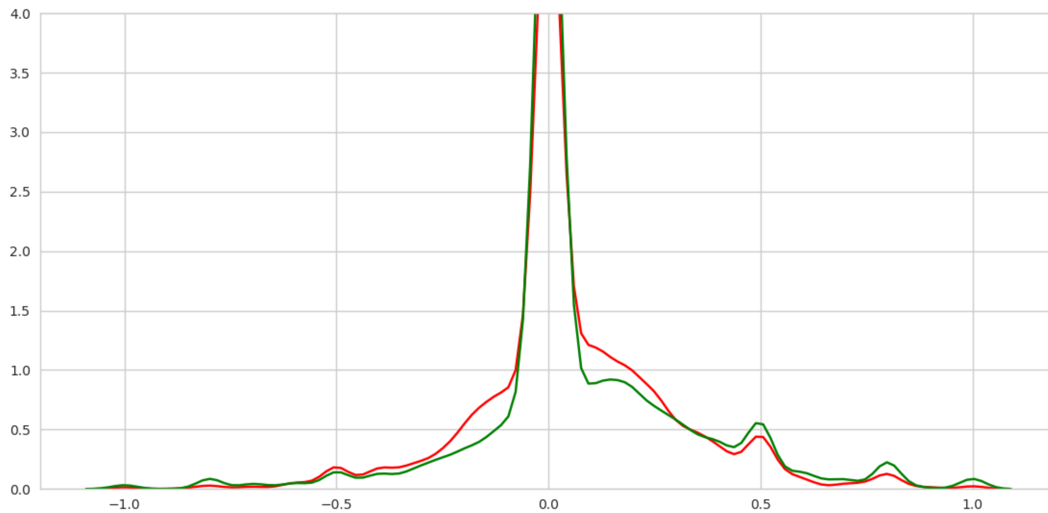

Figure 6.12 Pulwama-Balakot incident before (red) vs. after (green) commoner tweet sentiment

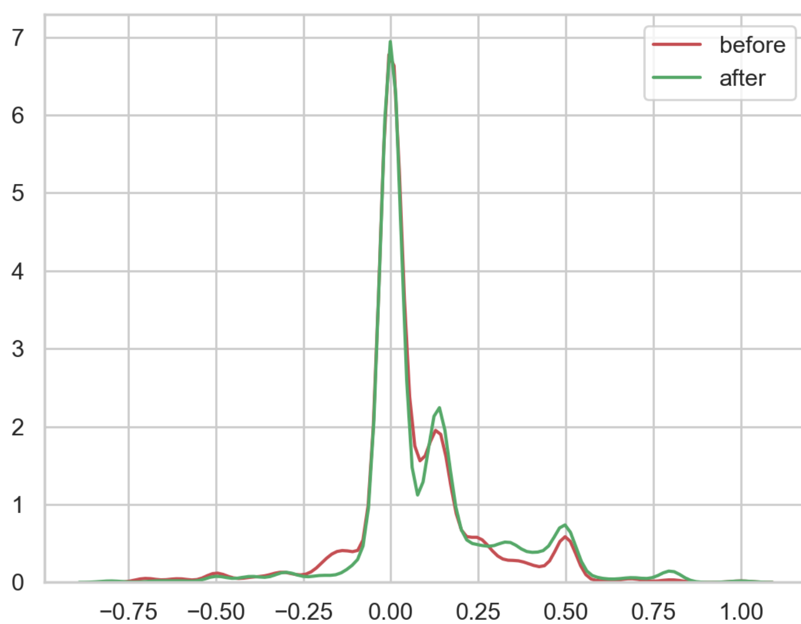

Figure 6.13 Pulwama-Balakot incident before (red) vs. after (green) NEWS tweet sentiment

	positive %	neutral %	negative %
before Pulwama	29	63	8
after Balakot	37	59	4

Table 6.5 Tabular breakdown of the polarity before Pulwama and after Balakot (NEWS data)

Rationalising the above results for the Pulwama-Balakot incident

- Positivity wrt. Modi increases after Balakot. Observed in both common people tweets and news tweets, albeit a lot more in the news tweets.

6.5 Lynching

Figures 6.14, 6.15, 6.16, 6.17, 6.18 and Table 6.6 are some of the graphs and figures obtained from experiments done on the Lynching twitter data spreading over four time periods. Let's analyse them.

1. As shown in Figure 6.14 the number of tweets in 2017 after the Pehlu Khan incident was significantly higher than 2016, 2018 - years either side of that. 2018 still had some activity, because people were now more conscious of such incidents after Pehlu Khan. The issue rose again after the General elections in 2019 when the number of reported lynchings increased.

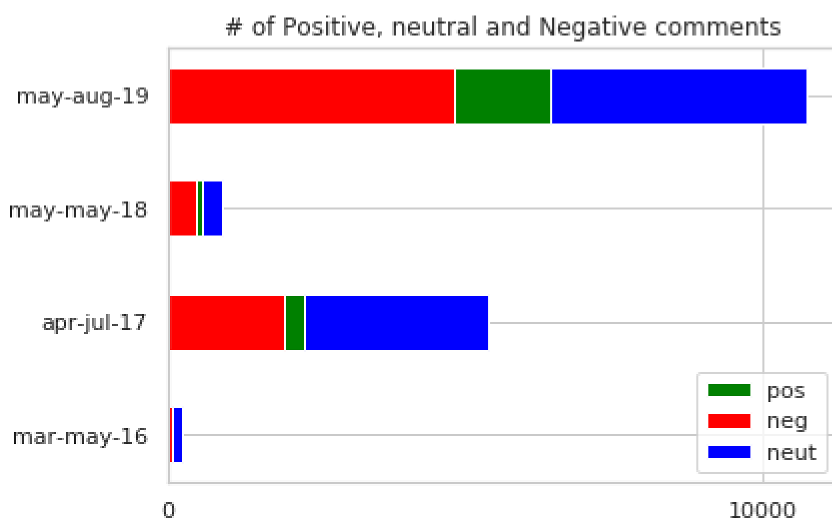

Figure 6.14 Lynching tweet count

2. As seen in Figure 6.15 and Table 6.6, about 68% of the tweets lie around the neutral region in 2016, when almost no such lynching incidents were reported. The neutrality decreased and polarity increased towards both sides in the following years as reported lynching incidents increaed.

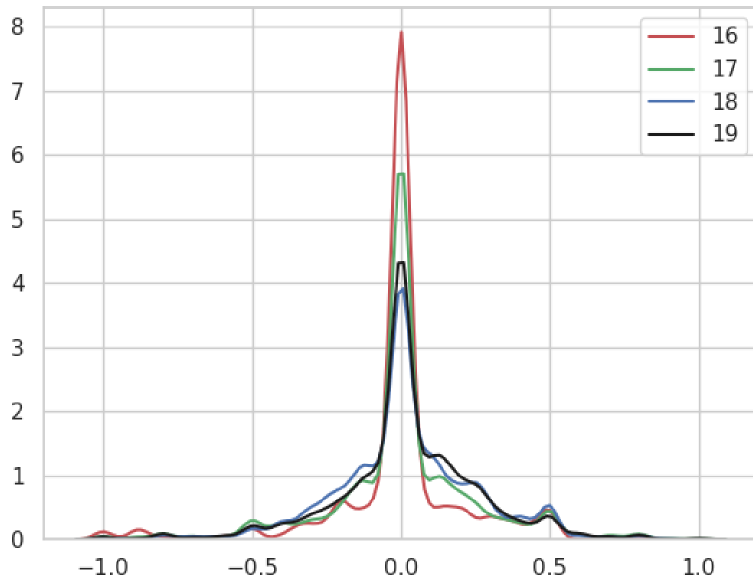

Figure 6.15 Lynching sentiment KDE curves

	positive %	neutral %	negative %
Mar-May, 2016	17	68	15
Apr-Jul, 2017	23	56	21
Mar-May, 2018	28	46	25
May-Aug, 2019	30	49	21

Table 6.6 Tabular breakdown of the polarity for Lynching topic (common people data)

3. Figure 6.16 shows the twitter activity of news media. It is more in 2017 than in 2019, unlike the twitter activity of common people.

4. Figure 6.17 shows the categorization of common people tweets in 8 emotions. Anger and sadness emotions are very high.

5. Figure 6.18 shows the categorization of news tweets in 8 emotions. Anger and sadness emotions are very high.

Rationalising the above results for lynching

- Unsurprisingly, a large percentage of tweets are negative.

- Fear emotion in news tweets is high in all timeperiods, the highest being in the period after the election. Same for negativity. Fear emotion is high in common people tweets as well, but not as high as in news. In common people tweets the anger emotion is also unusually high.

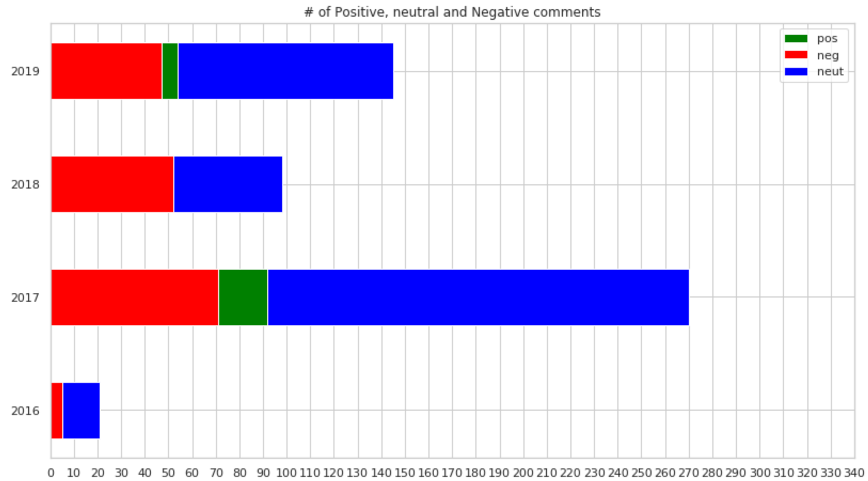

Figure 6.16 Lynching News Media Twitter Activity

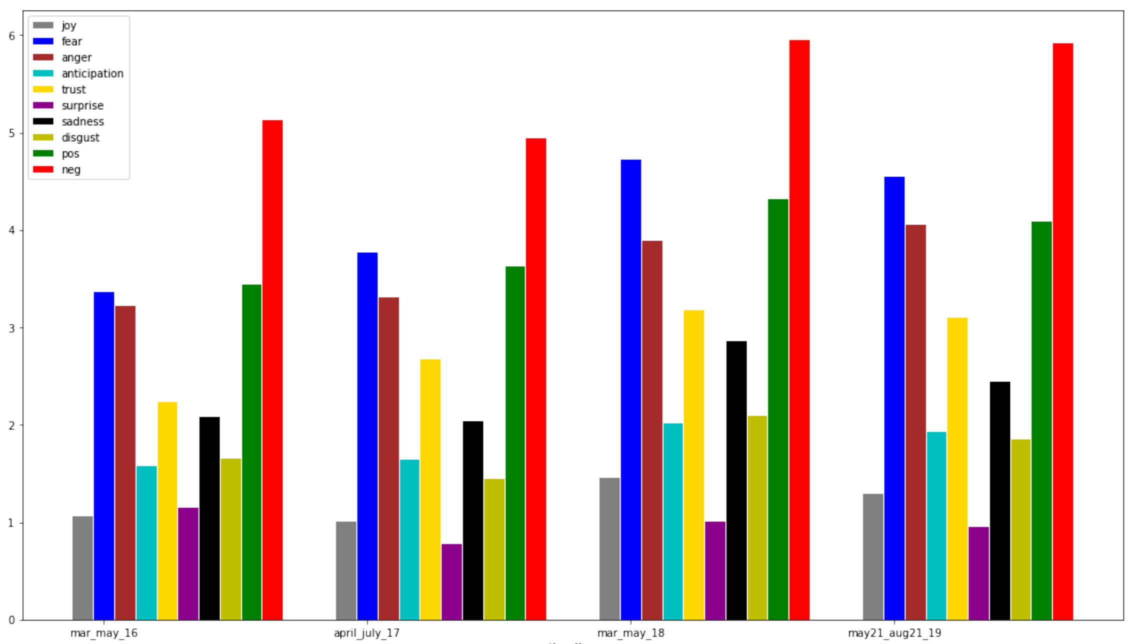

Figure 6.17 Lynching common people twitter emotion categorization

- Activity in news tweets most after pehlu khan's lynching (2017), followed by the period after elections (2019). Though for common people, most activity is found after elections, followed by the pehlu khan incident period (2017).

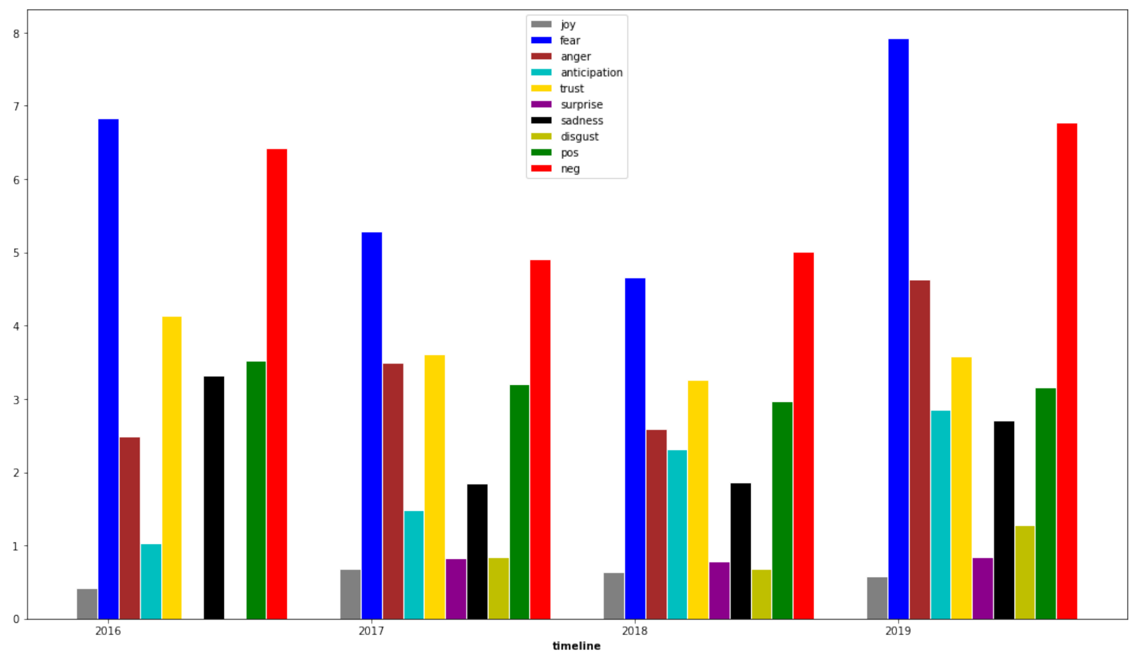

Figure 6.18 Lynching news media twitter emotion categorization

6.6 Indian Economy

Figures 6.19, 6.20, 6.21 and Table 6.7 are some of the graphs and figures obtained from experiments done on the Indian Economy twitter data spreading over multiple time periods. Let's analyse them.

1. Figure 6.19 tells us about the number of tweets related to Indian economy made in the first 3 months each year. The activity has increased year by year mostly because of the increasing availability and usage of internet in India.

2. Figure 6.20 and Table 6.7 shows the KDE plots and corresponding % change across years with polarity ranging from -1 to 1. As we can see the percentage of neutral tweets has decreased significantly over the years from 64% in 2013 to 49% in 2019. The curves have become higher towards both sides, more towards the positive side.

3. Figure 6.21 shows the emotion categorization of tweets made on the Indian Economy. The trust has decreased quite a bit after 2017.

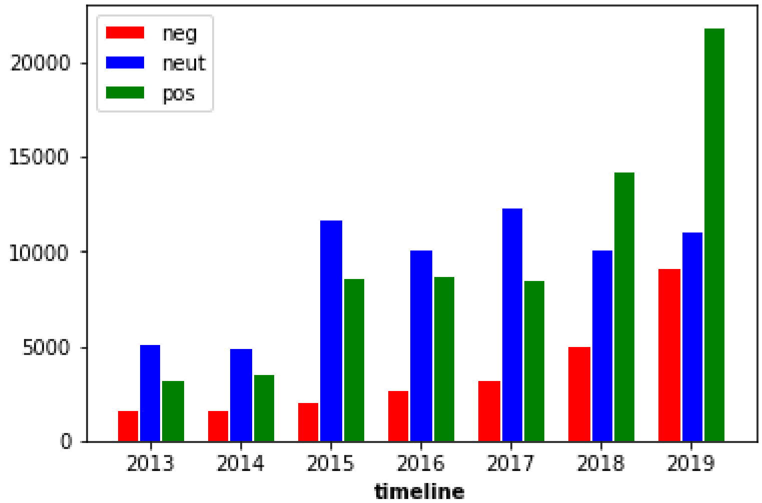

Figure 6.19 Indian Economy Tweet Count

	positive %	neutral %	negative %
Jan-Mar, 2013	24	64	12
Jan-Mar, 2014	27	61	11
Jan-Mar, 2015	31	62	7
Jan-Mar, 2016	31	60	8
Jan-Mar, 2017	26	63	9
Jan-Mar, 2018	35	54	10
Jan-Mar, 2019	37	49	13

Table 6.7 Tabular breakdown of the polarity for Indian Economy topic (common people data)

Rationalising the above results for Indian Economy

- Trust has decreased after 2017.
- Words expressing negativity have increased more in proportion wrt words expressing positivity.
- Sadness has increased gradually after 2015.
- Over the years the opinions are more polarized.

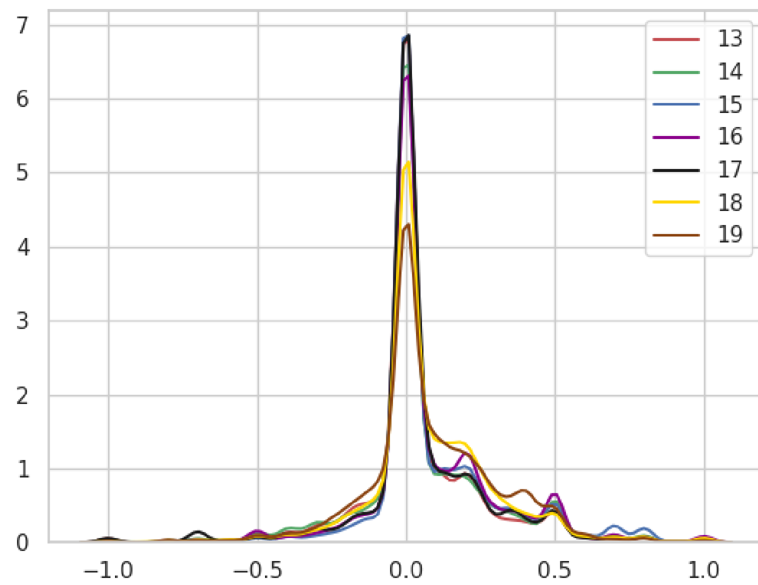

Figure 6.20 Indian economy sentiment percentage curves

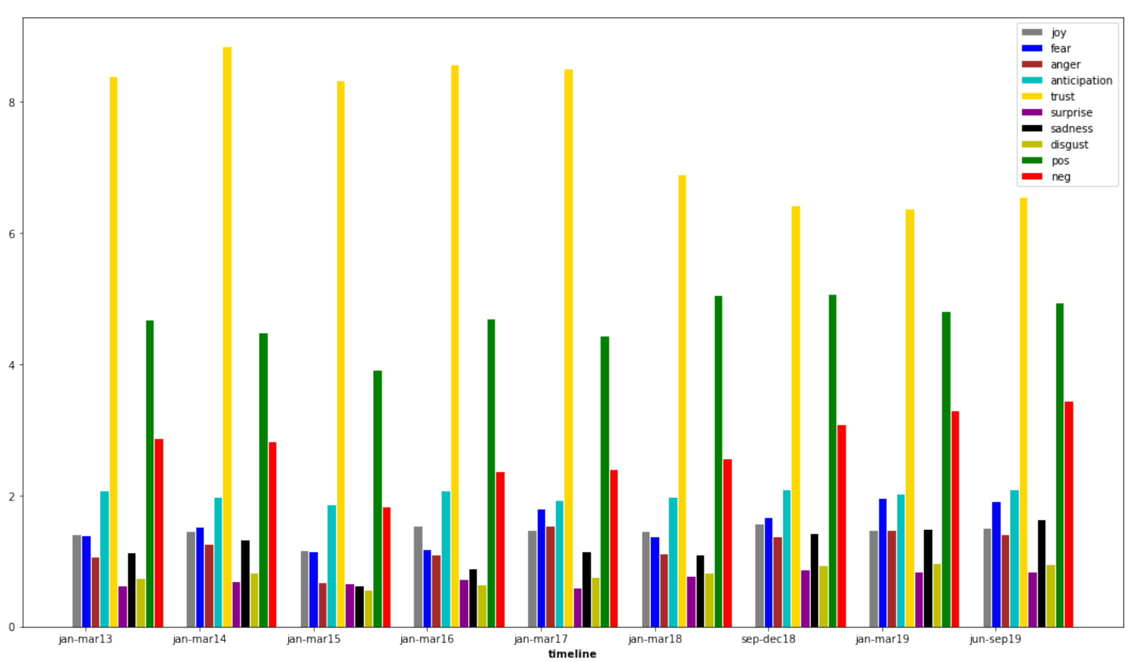

Figure 6.21 Indian economy categorized emotion wise

Chapter 7

Key Observations

7.1 Some event specific observations

Here, we list some event specific observations from the experiments done.

1. English and Hindi news media have had contrasting reactions to the events like Sabarimala.

2. Sentiment towards demonetisation has turned negative over time, as the negative effects became clearer.

3. Positivity wrt. Modi increases after Balakot. Observed in both common people tweets and news tweets, albeit a lot more in the news tweets. Did this positive reaction of media help in enhancing the popularity of Narendra Modi during and after this time?

4. Trust has decreased in the context of Indian economy over the last few years, while sadness has increased. This is in line with India's economic state the past few years including decreasing GDP year on year.

5. Fear has been a common emotion expressed in the context of lynching incidents, which is not surprising.

   Some of these observations are fairly obvious, and in a sense confirm the obvious, while some other observations raise more questions than answers and require further research.

7.2 Some general observations

Here, we list some general observations that seem true across the board.

1. Our study confirms the "exhausted majority" observation of the Hidden Tribes study. Majority of people have neutral opinion. Can be observed in all the five cases.

2. The activity at the time of the event is a lot more than at other times, observed in all cases.

3. Neutral content decreases disproportionately during and after the event. Are more of those neutrals becoming polar, or are most of those neutrals leaving the conversation which leaves only the polar opposites in the debate, similar to the exhausted majority sentiment?

4. Polarity before the event (pre-event state) is negligible, increases drastically during the event, and remains even after a significant amount of time after the event. Comparing the pre-event and post-event states, the change in polarity is seen to be drastic.

A controversial event leaves us in a polarized world, even after the event is not a part of the headlines and does not affect us in a direct way.

Chapter 8

Limitations

- Limitations of translation and transliteration.
  - For code-mixed data, we transliterated Hindi words to English. Transliterating Hindi to English transliterates only the words and leaves the sentence structure same which might affect the sentiment analysis.
  - For Hindi tweets we did not have a good sentiment analyser so we translated them to English first. Translation changes the words and the negativity of the words may not be captured in the translated words, potentially affecting the sentiment analysis.
- Limitations of sentiment analyser.
  - The sentiment analyser uses naive bayes and might be biased towards toxicity of specific words, rather than getting the sentiment of the entire sentence.
  - Upon manual analysis it was found that sarcasm went undetected by the sentiment analyser. For example, ”*Incredible journey Jungle to Shakha to London to Ayodhya to Godhra to Chaiwala to Delhi to Demonetisation to Pakora to Horse trading*” - this tweet was given a polarity score of +0.9 by the sentiment analyser.
  - Upon manual analysis it was found that the sentiment analyser was able to give an accuracy of 65% in the best case scenario. A better analysis tool with better performance could help us more.
- Limitations of emotional category finding tool.
  - Each word is associated with one or more of the eight emotions. The analysis counts the total number of words with each emotion over the whole corpus.
  - Ideally we would want an emotion to be associated with each tweet.
  - Hence even though the overall emotion of the whole corpus is captured by word-by-word emotion categorisation, tweet-by-tweet emotion categorisation might help us gauge the emotions expressed better.

Chapter 9

Conclusion

We tried to look for solutions for the following questions in this thesis, as mentioned in the Introduction itself:

- Is society inherently polarised?
- How does a divisive event affect the polarity?
- How does news media react to controvertial events and play a role in the conversation on social media platforms such as Twitter?

While we don't have definitive answers, what we observed is that:

- No, the society seems fairly neutral when major events have not happened.
- A divisive event leaves the society polarized, atleast as observed (and maybe enforced) on social media.
- It would be wrong to generalize media, but from the observed incidents, we can conclude that media is susceptible to intentional/unintentional bias in a variety of cases, including when serving to audiences of different language or culture.

Related Publications

- Ashutosh Ranjan, Dipti Misra Sharma, Radhika Krishnan : *Polarization and its Life on Social Media: A Case Study on Sabarimala and Demonetisation*. Proceedings of ICON 2020: 17th International Conference on Natural Language Processing, IIT Patna, India, 18th-20th December, 2020.

Bibliography

[1] S. Ahuja and G. Dubey. Clustering and sentiment analysis on twitter data. In *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*, pages 1–5. IEEE, 2017.

[2] J. Aldred, A. Astell, R. Behr, L. Cochrane, J. Hind, A. Pickard, L. Potter, A. Wignall, and E. Wiseman. *The world's 50 most powerful blogs*, 2008. URL https://www.theguardian.com/technology/2008/mar/09/blogs.

[3] E. W. Anderson. Customer satisfaction and word of mouth. *Journal of service research*, 1(1): 5–17, 1998.

[4] A. Balahur, M. Turchi, R. Steinberger, J. M. P. Ortega, G. Jacquet, D. Küçük, V. Zavarella, and A. El Ghali. Resource creation and evaluation for multilingual sentiment analysis in social media texts. In *LREC*, pages 4265–4269. Citeseer, 2014.

[5] C. Banea, R. Mihalcea, and J. Wiebe. Porting multilingual subjectivity resources across languages. *IEEE Transactions on Affective Computing*, 4(2):211–225, 2013.

[6] P. Barberá. Social media, echo chambers, and political polarization. *Social Media and Democracy: The State of the Field, Prospects for Reform*, page 34, 2020.

[7] M. Bautin, L. Vijayarenu, and S. Skiena. International sentiment analysis for news and blogs. In *ICWSM*, 2008.

[8] BBC. *Balakot air strike: Pakistan shows off disputed site on eve of India election*, 2019. URL https://www.bbc.com/news/world-asia-47882354.

[9] D. Bhaskar. *Why millions throng Sabarimala shrine*, 2016. URL https://daily.bhaskar.com/news/NAT-TOP-why-millions-throng-sabarimala-shrine-1755207.html.

[10] D. Bhaskar. *800 saal purani parampara tooti*, 2019. URL https://epaper.bhaskar.com/new-delhi/194/03012019/cph/1.

[11] I. A. Bhat, V. Mujadia, A. Tammewar, R. A. Bhat, and M. Shrivastava. Iiit-h system submission for fire2014 shared task on transliterated search. In *Proceedings of the Forum for Information*

Retrieval Evaluation, FIRE '14, pages 48–53, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3755-7. doi: 10.1145/2824864.2824872. URL http://doi.acm.org/10.1145/2824864.2824872.

[12] V. Bijukumar. When religious faith mutilates gender equality: Women entry in sabarimala temple in kerala. *ANTYAJAA: Indian Journal of Women and Social Change*, 4(2):238–244, 2019. doi: 10.1177/2455632719880857. URL https://doi.org/10.1177/2455632719880857.

[13] F. Bravo-Marquez, M. Mendoza, and B. Poblete. Meta-level sentiment models for big social data analysis. *Knowledge-based systems*, 69:86–99, 2014.

[14] BrightLocal. *Local Consumer Review Survey*, 2015. URL https://www.brightlocal.com/research/local-consumer-review-survey-2015/.

[15] P. R. Centre. *Social media outpaces print newspapers in the U.S. as a news source*, 2018. URL https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/.

[16] M. I. Common. *The Hidden Tribes of USA*, 2018. URL https://hiddentribes.us/.

[17] T. Economist. *The dire consequences of India's demonetisation initiative*, 2016. URL https://www.economist.com/finance-and-economics/2016/12/03/the-dire-consequences-of-indias-demonetisation-initiative.

[18] R. Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56 (4):82–89, 2013.

[19] M. A. R. A. Garibay, I. Polarization in social media assists influencers to become more influential: analysis and two inoculation strategies, 2020.

[20] H. Gene Zucker. The variable nature of news media influence. *Annals of the International Communication Association*, 2(1):225–240, 1978.

[21] M. Gerber. Predicting crime using twitter and kernel density estimation. *Decision Support Systems*, 61, 05 2014. doi: 10.1016/j.dss.2014.02.003.

[22] Glean. *Media monitoring, media measurement, analytics and market intelligence for companies, not-for-profit organizations and government agencies*, 2017. URL https://glean.info/.

[23] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3):211–223, 2001.

[24] O. Habimana, Y. Li, R. Li, X. Gu, and G. Yu. Sentiment analysis using deep learning approaches: an overview. *Science China Information Sciences*, 63(1):1–36, 2020.

[25] S. Han. *Python freeware translate*, 2008. URL https://pypi.org/project/py-translate/.

[26] Y. He and D. Zhou. Self-training from labeled features for sentiment analysis. *Information Processing & Management*, 47(4):606–616, 2011.

[27] J. Henrique. *Get Old Tweets Python API*, 2018. URL https://github.com/Jefferson-Henrique/GetOldTweets-python.

[28] A. Hernandez-Suarez, G. Sanchez-Perez, K. Toscano-Medina, H. Perez-Meana, J. Portillo-Portillo, V. Sanchez, and L. García Villalba. Using twitter data to monitor natural disaster social dynamics: A recurrent neural network approach with word embeddings and kernel density estimation. *Sensors*, 19(7):1746, Apr 2019. ISSN 1424-8220. doi: 10.3390/s19071746. URL http://dx.doi.org/10.3390/s19071746.

[29] T. Hindu. *Will Balakot air strikes influence how India votes?*, 2019. URL https://www.thehindu.com/elections/lok-sabha-2019/will-balakot-air-strikes-influence-how-india-votes/article26555839.ece.

[30] T. Hindu. *Two women under 50 enter Sabarimala temple*, 2019. URL https://www.thehindu.com/news/national/kerala/two-women-under-50-enter-sabarimala-temple/article25887406.ece.

[31] T. Hindu. *Two women make it to Sabarimala*, 2019. URL https://www.thehindu.com/todays-paper/two-women-make-it-to-sabarimala/article25894662.ece.

[32] T. Hindu. *Shashi Tharoor slams BJP for surge in mob lynchings*, 2019. URL https://www.thehindu.com/news/national/shashi-tharoor-slams-bjp-for-surge-in-mob-lynchings/article29481103.ece.

[33] Y. Hu and W. Li. Document sentiment classification by exploring description model of topical terms. *Computer Speech & Language*, 25(2):386–403, 2011.

[34] M. Hulden, M. Silfverberg, and J. Francom. Kernel density estimation for text-based geolocation. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, page 145–150. AAAI Press, 2015. ISBN 0262511290.

[35] D. India. *One more arrested in Alwar cow vigilantism incident*, 2017. URL https://www.dnaindia.com/india/report-one-more-arrested-in-alwar-cow-vigilantism-incident-2388187.

[36] A. International. *Hate Crime Reports On An Alarming Rise – Reveals Amnesty International India's 'Halt The Hate'*, 2019. URL https://amnesty.org.in/news-update/hate-crime-reports-on-an-alarming-rise-reveals-amnesty-international-indias-h

[37] A. Jazeera. *India: Demonetisation takes its toll on the poor*, 2016. URL https://www.aljazeera.com/economy/2016/11/16/india-demonetisation-takes-its-toll-on-the-poor.

[38] A. Joshi, P. Bhattacharyya, and B. R. A fall-back strategy for sentiment analysis in hindi: a case study. 12 2010.

[39] T. Kincl, M. Novák, and J. Přibil. Improving sentiment analysis performance on morphologically rich languages: Language and domain independent approach. *Computer Speech Language*, 56: 36 – 51, 2019. ISSN 0885-2308. doi: https://doi.org/10.1016/j.csl.2019.01.001. URL http://www.sciencedirect.com/science/article/pii/S0885230816300109.

[40] L.-W. Ku, Y.-T. Liang, H.-H. Chen, et al. Opinion extraction, summarization and tracking in news and blog corpora. In *AAAI spring symposium: Computational approaches to analyzing weblogs*, volume 100107, pages 1–167, 2006.

[41] R. A. Laksono, K. R. Sungkono, R. Sarno, and C. S. Wahyuni. Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes. In *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, pages 49–54. IEEE, 2019.

[42] B. Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.

[43] Y. Liu, L. Han, and X. Cai. A small sample text classification algorithm based on kernel density estimation. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 8:235–244, 2015.

[44] S. Loria. textblob documentation. *Release 0.15*, 2, 2018.

[45] S. Mahapatra and J. Plagemann. The social media strategies of indian political parties. *German Institute of Global and Area Studies (GIGA)*, 2019.

[46] P. Mancini. Media fragmentation, party system, and democracy. *The International Journal of Press/Politics*, 18(1):43–60, 2013.

[47] B. Marar. *Kerala High Court PIL*, 1991. URL https://indiankanoon.org/doc/1915943/.

[48] W. Medhat, A. Hassan, and H. Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113, 2014.

[49] A. Micu, A. E. Micu, M. Geru, and R. C. Lixandroiu. Analyzing user sentiment in social media: Implications for online marketing strategy. *Psychology & Marketing*, 34(12):1094–1100, 2017.

[50] Mint. *SC refers individual cases against demonetisation to constitution bench*, 2016. URL https://www.livemint.com/Politics/i2HFNZCdECO3Hn1qqXpyLI/SC-refers-individual-cases-against-demonetisation-to-constit.html.

[51] S. M. Mohammad. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier, 2016.

[52] S. M. Mohammad and S. Kiritchenko. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 31(2):301–326, 2015.

[53] S. M. Mohammad and P. D. Turney. Nrc emotion lexicon. *National Research Council, Canada*, 2, 2013.

[54] R. Moraes, J. F. Valiati, and W. P. G. Neto. Document-level sentiment classification: An empirical comparison between svm and ann. *Expert Systems with Applications*, 40(2):621–633, 2013.

[55] P. Munjal, M. Narula, S. Kumar, and H. Banati. Twitter sentiments based suggestive framework to predict trends. *Journal of Statistics and Management Systems*, 21(4):685–693, 2018.

[56] NDTV. *India Strikes After Pulwama Terror Attack, Hits Biggest Jaish-e-Mohammed Camp In Balakot*, 2019. URL https://www.ndtv.com/india-news/india-struck-biggest-training-camp-of-jaish-in-balakot-large-number-of-terror

[57] T. Neyazi. Social media and political polarization in india. *Seminar*, pages 31–35, 11 2017.

[58] T. R. B. of India. *Withdrawal of Legal Tender Status for 500 and 1000 Notes: RBI Notice*, 2016. URL https://www.rbi.org.in/Scripts/BS_PressReleaseDisplay.aspx?prid=38520.

[59] T. T. of India. *SC smashes gender barrier, opens Sabarimala for all women*, 2018. URL https://timesofindia.indiatimes.com/india/sc-smashes-gender-barrier-opens-sabarimala-for-all-women/articleshow/66001039.cms.

[60] T. T. of India. *Women scale one more summit; taken to Sabarimala in commando-raid style*, 2019. URL https://timesofindia.indiatimes.com/city/kochi/women-scale-one-more-summit-taken-to-sabarimala-in-commando-raid-style/articleshow/67358068.cms.

[61] T. T. of India. *Jaish terrorists attack CRPF convoy in Kashmir, kill atleast 40*, 2019. URL https://timesofindia.indiatimes.com/india/37-crpf-jawans-martyred-in-ied-blast-in-jks-pulwama/articleshow/67992189.cms.

[62] M. Ogneva. *How Companies Can Use Sentiment Analysis to Improve Their Business*, 2010. URL https://mashable.com/2010/04/19/sentiment-analysis/.

[63] O. Oyebode and R. Orji. Social media and sentiment analysis: The nigeria presidential election 2019. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 0140–0146. IEEE, 2019.

[64] A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10, pages 1320–1326, 2010.

[65] S. Poria, E. Cambria, D. Hazarika, and P. Vij. A deeper look into sarcastic tweets using deep convolutional neural networks. *arXiv preprint arXiv:1610.08815*, 2016.

[66] T. H. Post. *Demonetisation Death Toll Rises To 25 And It's Only Been 6 Days*, 2016. URL https://web.archive.org/web/20161116020908/http://www.huffingtonpost.in/2016/11/15/demonetisation-death-toll-rises-to-25-and-its-only-been-6-days/.

[67] T. W. Post. *After Pulwama, the Indian media proves it is the BJP's propaganda machine*, 2019. URL https://www.washingtonpost.com/opinions/2019/03/04/after-pulwama-indian-media-proves-it-is-bjps-propaganda-machine/.

[68] R. Prabowo and M. Thelwall. Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2):143–157, 2009.

[69] B. Quint. *Will Pulwama Balakot Make Voters Re-Elect Modi?*, 2019. URL https://www.bloombergquint.com/politics/pulwama-attack-balakot-air-strike-will-voters-re-elect-pm-narendra-modi.

[70] P. M. Roget. *Roget's Thesaurus*, 1911. URL http://www.gutenberg.org/ebooks/10681.

[71] A. K. Singh, D. K. Gupta, and R. M. Singh. Sentiment analysis of twitter user data on punjab legislative assembly election, 2017. *International Journal of Modern Education and Computer Science*, 9(9):60, 2017.

[72] S. Sohangir, N. Petty, and D. Wang. Financial sentiment lexicon analysis. In *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, pages 286–289. IEEE, 2018.

[73] D. Tang, F. Wei, B. Qin, T. Liu, and M. Zhou. Coooolll: A deep learning system for twitter sentiment classification. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pages 208–212, 2014.

[74] G. Tarabout. Chapter 3: Religious Uncertainty, Astrology and the Courts in South India. In D. Berti, A. Good, and G. Tarabout, editors, *Of Doubt and Proof. Legal and Ritual Practices of Judgment*, pages 59–75. Ashgate, Apr. 2015. URL https://hal.archives-ouvertes.fr/hal-01246885.

[75] F. Times. *India demonetisation fails to purge black money*, 2018. URL https://www.ft.com/content/7dbe0e14-8d8a-11e7-a352-e46f43c5825d.

[76] N. Times. *sabarimala mandir: tooti varshon purani parampara, isliye mandir me mahi-laon ke pravesh pe thi rok*, 2018. URL https://navbharattimes.indiatimes.com/astro/dharam-karam/religion-and-spiritualism/know-difference-between-brahmachari-and-naishtika-brahmachari-sabrimala-templ

[77] N. Times. *Sabarimala: Ayappa ke darshan kar do mahilaon ne yu racha itihaas, kuch logonko hi thi jaankari*, 2019. URL https://navbharattimes.indiatimes.com/state/other-states/other-cities/a-darshan-in-secrecy-how-cops-planned-pre-dawn-sabari-trek/articleshow/67359187.cms.

[78] I. Today. *Here is what PM Modi said about the new Rs 500, Rs 2000 notes and black money*, 2016. URL indiatoday.in/india/story/live-pm-narendra-modi-addresses-nation-350943-2016-11-08.

[79] I. Today. *Thanks to Demonetization and GST, India's GDP growth hit a 4-year low at 6.5 per cent*, 2018. URL https://www.indiatoday.in/education-today/gk-current-affairs/story/demonetization-gst-slowed-down-gdp-growth-of-india-1123939-2018-01-06.

[80] I. Today. *Kitne aadmi the? 3 questions that loomed after Balakot airstrike*, 2019. URL https://www.indiatoday.in/india/story/iaf-airstrike-in-balakot-pakistan-questions-that-loomed-1472646-2019-03-07.

[81] A. Tyagi, A. Field, P. Lathwal, Y. Tsvetkov, and K. M. Carley. A computational analysis of polarization on indian and pakistani social media, 2020.

[82] P. Vasterman. *From Media Hype to Twitter Storm. News Explosions and Their Impact on Issues, Crises, and Public Opinion. Edited by Peter Vasterman. INTRODUCTION*. 02 2018. ISBN 978 94 6298 217 8. doi: 10.5117/9789462982178.

[83] C. Vision. *Sabarimala Opening Dates*, 2019. URL https://chennaivision.com/sabarimala-opening-dates/.

[84] S. Vosoughi, D. Roy, and S. Aral. The spread of true and false news online. *Science*, 359(6380): 1146–1151, 2018.

[85] Wikipedia. *Kernel Density Estimation*. URL https://en.wikipedia.org/wiki/Kernel_density_estimation#Example.

[86] T. Wire. *Balakot and After, or How to Build a Strategic Doctrine for Political Gain*, 2019. URL https://thewire.in/politics/balakot-narendra-modi-2019-elections.

[87] F. Xianghua, L. Guo, G. Yanyan, and W. Zhiqiang. Multi-aspect sentiment analysis for chinese online social reviews based on topic modeling and hownet lexicon. *Knowledge-Based Systems*, 37:186–195, 2013.

[88] N. N. Yusof, A. Mohamed, and S. Abdul-Rahman. Reviewing classification approaches in sentiment analysis. In *International conference on soft computing in data science*, pages 43–53. Springer, 2015.

[89] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 83–92, 2014.