Harnessing Twitter Data for Analyzing Public Reactions to Transportation Policies: Evidences from the Odd-Even Policy in Delhi, India

Rounaq BASU [a], Aparup KHATUA [b], Arnab JANA [c], Saptarshi GHOSH [d]

[a] *Department of Civil & Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; E-mail: rounaq@mit.edu*
[b] *Temasek Laboratories (TL@NTU), Nanyang Technological University, Singapore; E-mail: aparupkhatua@gmail.com*
[c] *Centre for Urban Science and Engineering, IIT Bombay, India; E-mail: arnab.jana@iitb.ac.in*
[d] *Department of Computer Science and Engineering, IIT Kharagpur, India; E-mail: saptarshi@cse.iitkgp.ernet.in*

Abstract: Pre-post analysis of public policies faces significant challenges in terms of data collection. With the advent of popular social media platforms like Facebook and Twitter, it has become easier to access relevant data which enables evaluation of the success of a policy. The paper proposes a methodology for collection and processing of Twitter data to understand public sentiment and reaction to a public policy. The methodology is demonstrated through a case study of a particular transportation policy (the Odd-Even policy) that was recently undertaken in Delhi, India. A large set of tweets (microblogs posted on Twitter) related to the policy was collected, and analyzed. It was observed that the citizens are positive overall with respect to the policy. However, dissatisfaction related to certain issues and consequences of the policy was also prominent. Our analysis, although post-policy implementation, aids in identifying such issues so that remedial measures can be taken.

Keywords: Transportation policy evaluation; Sentiment analysis; Social media; Citizen participation; Odd-Even policy.

1. INTRODUCTION

Transportation choice is strongly interrelated with public policy. Since the 1970's, urban transportation policies in the US have focused on tackling this important question – "*How can an existing urban transportation system be better managed so that increasing travel demand could be satisfied without building more capacity?*" (1). The environmental repercussions of prolonged reliance on private automobiles, especially single occupancy vehicles, promoted this train of thought and laid down the foundation of travel demand management (TDM). Implementation of rideshare programs is a classic example of TDM. The primary focus of any TDM activity is to meet present mobility needs along with energy conservation and improvement in air quality.

The pervasive use of automobiles is considered as a major source of air pollution problems at the local level and related global environmental issues (2). Urbanization and economic growth of other nations over the last two decades have promoted similar travel behavior. Various policies have been implemented worldwide to curb the growing dependence on automobiles, e.g. a well-connected multi-modal public transport network, congestion pricing, etc. Evaluation of these policies and the extent of their reception by the public are extremely critical. A study by Bengston, Fletcher and Nelson (3) found a lack of empirical evaluations of urban public policies. Hatzopoulou and Miller (4) commented that modeling tools which can be used for assessment of the impacts of policies on their proposed goals have not caught up with recent technological developments.

The need for transportation analysts to become more responsive to the challenges associated with policy evaluation was highlighted by Thomas and Schofer (5). Keeping these discussions in mind, this study aims to focus on evaluation of transportation policies through analysis of public reactions. Developing nations face a lack of robust data collection methods. Some techniques like installation of CCTV cameras at road crossings to measure the traffic flow are applicable in few developed countries, but not in India. Moreover, the huge population makes it difficult to carry out detailed field surveys. Because of these factors, utilizing data from online sources (which can be collected in huge volumes and at minimal cost) is a viable alternative to gauge the opinion of the public on Government policies. The potential of such online sources is increasing with time, as millions of people in India (and other developing countries) are becoming active on Twitter as noticed in recent years (6). This paper addresses this pressing issue (the lack of robust, cost-effective and time-effective data collection methods) by proposing the use of Big Data for the abovementioned analysis in developing countries. However, it must be noted that such techniques can prove to be useful for developed countries as well.

The objectives of the paper are two-fold: (a) To propose a methodological framework to extract and process Twitter data, and (b) To demonstrate the usefulness of Twitter data in determining the effect or reactions to public policies. The second objective is attempted through analysis of a specific case study from India: the Odd-Even policy implemented in Delhi, the capital city of India. The Government of Delhi issued a decree on December 28, 2015 that motor cars having a registration number ending with an odd digit would be allowed to ply only on odd dates of the month, and those with ending with an even digit for even dates of the month. This restriction was applicable from 8:00 AM to 8:00 PM, and was relaxed only on Sundays. An experimental run was carried out from January 1, 2016 to January 15, 2016 in order to gauge the impact on vehicular pollution levels before actual policy implementation. An interesting highlight of the policy is the exclusion of this restriction on vehicles occupied solely by women and vehicles driven / occupied by handicapped people. With an attempt to reinvigorate the use of public transit as well as remain mindful of the needs of specific population cohorts (inclusiveness), this policy is a first-of-its-kind in India. The widespread media coverage and public involvement in critiquing the policy led to its inclusion as a case study in this paper.

When a policy like the *Odd-Even policy* is introduced by the Government, it is important for the Government to know answers to questions like: (i) Does the policy face majorly support or opposition from the public? and (ii) What are the problems being faced by the public due to the policy (if any)? This study focuses on whether such questions can be answered using information crowdsourced from online social media like Twitter. The remainder of the paper is organized as follows. Section 2 briefly summarizes the importance of Twitter as a data source, and the associated advantages and limitations. Section 3 presents the methodology for data collection, cleaning (or pre-processing) and analysis. The aforementioned case study is analyzed in Section 4, while Section 5 provides final remarks about this method and its usefulness for future research.

2. LITERATURE REVIEW

Since the advent of the Internet in the 1990s, academicians have displayed particular interest in examining the behavior of human social groups that are linked by electronic networks. This is guided by the assumption that virtual behavior is a reflection of actual behavior in the real world. However, harnessing social media for gaining insights into human behavior provides steep challenges, the most critical of which is the sheer volume of available data (7). Availability of vast amounts of data on a national (often international as well) scale, colloquially known as Big Data, is productive for understanding transportation behavior. Big Data facilitates flows of information and mechanisms of better understanding of heterogeneous individuals.

2.1 Twitter as a data source

Twitter (https://twitter.com) is a widely used microblogging platform, which enables users to communicate through short text messages called microblogs or tweets (at most 140 characters long). Users can interact socially by 'following' other users - if user A is following user B, then all tweets posted by B will be made available to A in real time. Twitter has become one of the most popular social networking sites in today's Web, with more than 500 million tweets being posted every day on average (8). The growth in popularity of Twitter has coincided with a rise in the usage of smartphones, which can be used to directly post tweets by users on the move. Hence, multiple Twitter users post tweets throughout the day, describing their thoughts and actions, which are available in real time (within a few seconds).

Another reason for the importance of Twitter as a data source is the huge amount of data that can be collected at minimal cost. The Twitter platform provides ways by which a large amount of tweets can be collected through computer programs in real time (using the Twitter API – see https://dev.twitter.com/overview/documentation for details). For instance, the Twitter streaming API is a publicly available resource that allows access to the global stream of tweets in real time. Various filters can be applied to get a data stream of tweets that are of interest, e.g., tweets being posted within a certain geographical region, or tweets containing some specific keywords, or tweets being posted by a selected group of users, and so on. The data collected can then be analyzed (e.g., using data mining techniques) to get a quick understanding of the present situation. For instance, if traffic gets congested in a certain region of a city, Twitter users stuck in the congestion might post tweets to inform about the situation. Such posts can be used to immediately inform the traffic authorities of the problems in that region.

Because of the large amounts of data available on users' emotions and opinions, crowdsourced data from Twitter has frequently been used to understand public reaction to various events / objects. For instance, Twitter data has been used to predict whether a movie is likely to succeed (9), the outcome of political elections (10), and so on. In this work, we use Twitter data to understand the public reaction to a Government policy related to transportation. Prior works have already shown that Twitter can be a valuable data source for information on transportation incidents such as accidents (11), and for understanding customer satisfaction with transportation services (12). Twitter data was also proposed as an alternate novel method for estimating satisfaction levels of public transit riders through sentiment analysis (13). However, the use of Twitter data in the field of transportation has remained extremely limited till date. In this work, we show that crowdsourced Twitter data can also be very useful in evaluating transportation policies introduced by the Government.

2.2 Advantages and limitations of Twitter as a data source

As stated earlier, the primary advantages of Twitter as a data source are (i) the huge amounts of data available, (ii) the ease of accessing the data, and (iii) the real time nature of the data. For instance, if public opinion regarding a Government policy is to be evaluated, it is much easier to obtain the opinion of thousands of people through Twitter using computer programs, than to conduct a survey and convince even a few hundred people to participate in the survey.

However, Twitter data has certain limitations as well. The amount of data is so large that it is infeasible for human beings to process all the data. Many tweets are merely conversational / sarcastic in nature, and do not provide any meaningful information. Further, there is lot of duplication of content, since the same tweet (or very similar tweets) is frequently posted by multiple users. Hence, the meaningful information is usually hidden within a large amount of conversational content. In this scenario, computer programs and techniques in Computer Science fields like Data Mining, Natural Language Processing, Machine Learning, etc., have to be used to extract meaningful information from the tweets.

In recent years, there has been a lot of advancement in the Computer Science fields stated above, and reliable methods are now available for many tasks like inferring the sentiment in a piece of text. However, most of the existing techniques work well for formal English text, whereas tweets are often written in an informal way using abbreviations, colloquial terms, etc. (this is primarily due to the extremely short size of tweets - at most 140 characters). Hence it is still a challenge for computer scientists to develop improved natural language processing techniques for tweets (14). As better techniques are developed, they can then be used for better evaluation of the public reaction to policies.

Another potential limitation of Twitter data is that the user population that generates the data is largely dominated by people who are comfortable with modern technology (generally reflective of the younger population cohorts), and people having access to the Internet; hence, the data might not always reflect the voice of all sections of the society. However, it can be postulated that Twitter provides a far more representative and larger sample than might be possible with traditional survey techniques.

3. DATA AND RESEARCH CONTEXT

It is known that users in Twitter annotate their tweets with certain keywords called hashtags, which refer to the event / issue on which the tweet is being posted. It was observed that tweets related to the Odd-Even policy in Delhi were being annotated with the four hashtags: *#oddEven*, *#oddEvenRule*, *#oddEvenPlan*, and *#oddEvenFormula*. Hence, the Twitter Streaming API was used to collect all tweets containing any of the hashtags mentioned above, posted during the period from December 19, 2015 to January 31, 2016. In total, 213,219 tweets posted during this period and containing the hashtags mentioned above were collected.

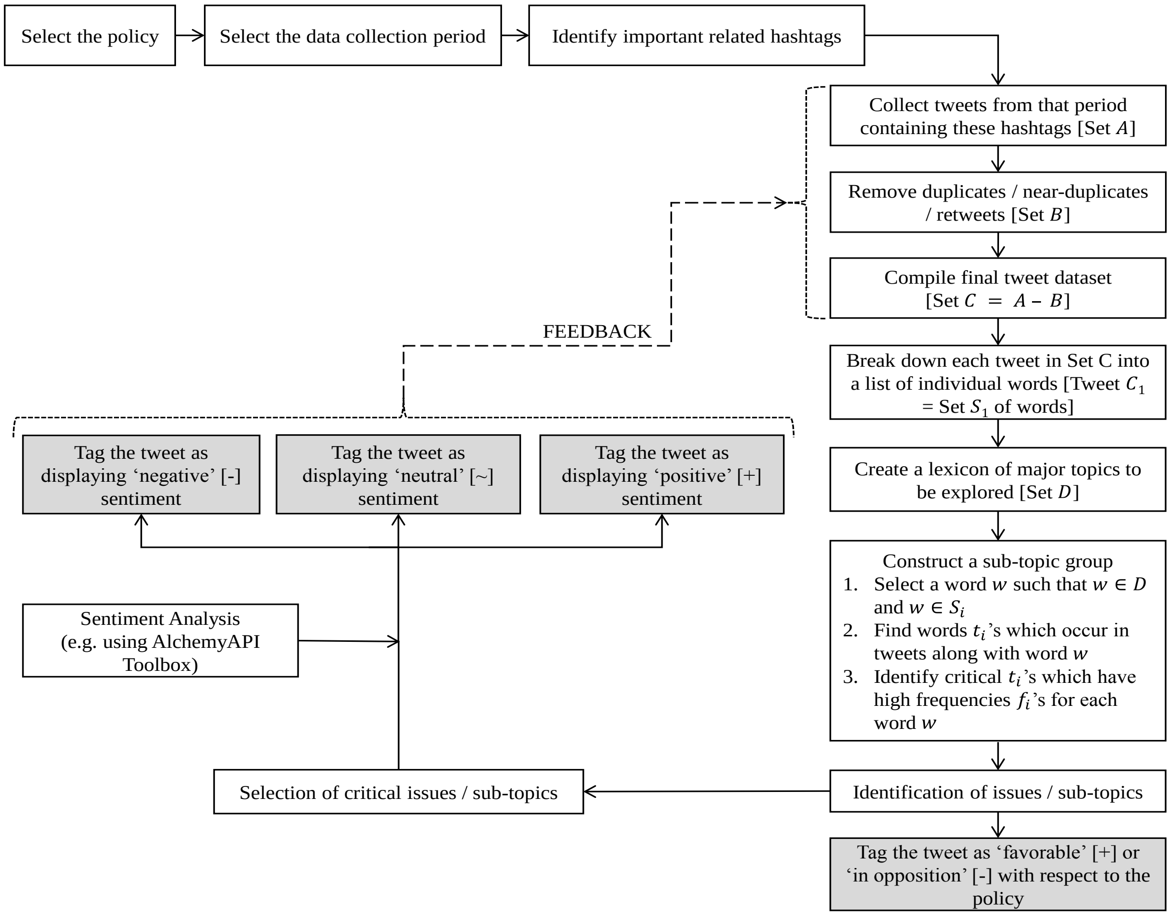


Figure 1. Flowchart of the (generic) methodology for using Twitter data to analyze public reactions to Government policies

A flowchart of the entire methodology adopted in this study is presented in Figure 1, which may prove to be useful for utilizing Twitter data to explore the aforementioned objectives with respect to any public policy. The individual steps are discussed in the rest of this paper in the context of our chosen case study.

Prior works have observed that there is significant duplication of content in the Twitter stream. Particular tweets are often retweeted / re-posted by many different users, and multiple users post very similar tweets. Since such duplicate content leads to information overload for most analyses, standard steps were taken to identify duplicate tweets and remove the duplicates (15). In brief, two tweets are considered duplicates if they have a significant fraction of words in common, and the chronologically latter tweet is ignored subsequently. After this process of removing duplicates and near-duplicates, a set of 69,534 distinct tweets were left. The fact that only 32.6% of tweets can be considered distinct lends weight to the argument of significant duplication of content in Twitter, thereby rendering this step of the methodology particularly important.

Each collected tweet contains the text of the tweet, information on the user who posted the tweet, the UTC timestamp when the tweet was posted, some attributes of the user who posted the tweet (e.g., the number of their social links), and some attributes of the tweet (e.g., how many times the tweet has been retweeted). However, for the present analysis, only the text of the tweet was considered. The text of tweets often contains keywords called hashtags (starting with the symbol #). Table 1 shows some sample tweets from our dataset, and the hashtags.

Table 1. Sample Tweets related to Key Issues

#Hashtags	Sample Tweets
#oddevenplan	Air pollution levels in Delhi dropped by 300% on day 2 which is remarkable #OddEven ***#OddEvenPlan*** #OddEvenMovement
#oddevenformula	Odd-even: No manic Monday, Delhiites tweet about jam-free roads ***#OddEvenFormula***
#DelhiForOddEven	***#DelhiForOddEven*** DTC buses estimated to have carried 4 million commuters on Day1 of #OddEven
#carpooling	How Delhi's odd-even policy has made #***carpooling*** into an idea whose time has come
#OddEvenFails	#oddeven is not a success due to empty roads on holidays, but ***#OddEvenFails*** due to idea of development fails against d idea…
#women	***#OddEven*** is putting #women to risk. Forced to travel alone or by buses, how safe r they?

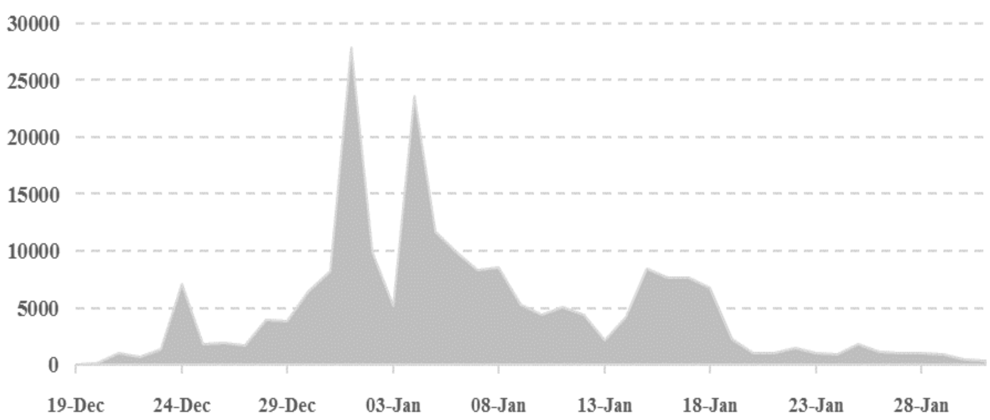


Figure 2. Number of tweets per day over the data collection period

As stated earlier, the tweets were collected over the period from December 19, 2015 to January 31, 2016. Figure 2 shows the number of tweets posted per day during this period. News of the proposed policy was released by the Government on December 24, 2015 and it is reflected from the high tweet levels on that day as compared to its neighbors (see Figure 2). The policy document was approved and released on December 28, which led to further outbreak of emotion and opinion over the next few days. Since the policy period was from January 1, 2016 to January 15, 2016, a sharp spike in Twitter activity was noticed on the first day of implementation (January 1) expectedly. Tweets dwindled over the next two days which constituted a weekend (January 3 being a Sunday) because of a drop in the number of trips, especially work trips. January 4, a Monday, again witnessed a sharp rise in tweet levels with a slow decline over the remaining days of the policy. The final rise was seen on the last day of the policy, i.e. January 15. The policy remained a part of public discussion even after its conclusion. Interestingly, comparatively high tweet levels were observed on both January 16 and January 17 (Saturday and Sunday respectively). This was a contrast to the tweet behavior on the previous weekends while the policy was in effect, and was a result of people actively discussing their overall opinion about the policy.

Table 2 reports the spatial distribution, on the basis of the latitude and longitude, of our tweets in tabular format. It is worth to note that around 3700 of our tweets, out of 0.02 million total tweets, are geo-tagged tweets. However, if we assume this as a representative sample then more than 50% of our sample is from Delhi (latitude of Delhi: 28.6; longitude of Delhi: 77.2) and surrounding regions. Table 2 reports a few important locations, mostly from and around Delhi, of our dataset. Figure 3, the heat map of our twitter data, is the graphical depiction of Table 2. As mentioned, this graphical depiction reconfirms that a significant number of tweets are from Delhi and surrounding region. However, it is also noteworthy that social media users have also participated in this Odd-Even policy discussion from other parts of India.

Table 2. Spatial distribution of geo-tagged tweets

Longitude	Latitude	# of tweets
77.09	28.40	1426
72.87	18.84	487
77.56	12.73	284
76.97	28.20	224
77.56	12.92	149
77.50	28.44	145

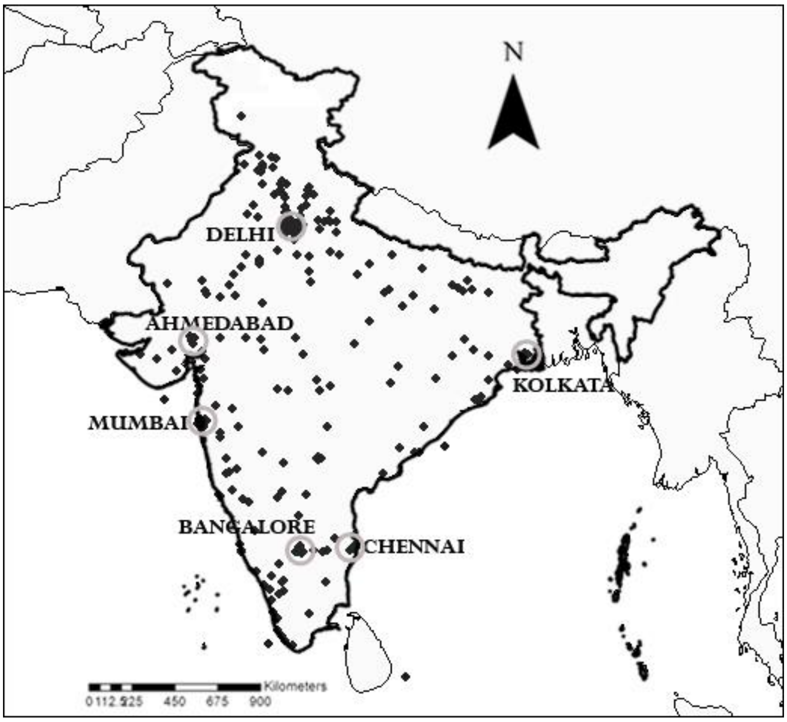


Figure 3. Spatial distribution of geo-tagged tweets.

4. DATA ANALYSIS AND FINDINGS

We propose an integrated framework with three steps of data analysis that would lead to identification of public reaction to the policy. As shown in Section 4.1, the first step is to identify a specific set of popular high-volume themes that emerge out of the Twitter conversations regarding the policy. This would be followed by selection of keywords from some of these important themes and exploration of their temporal distribution, as outlined in Section 4.2. We would expect some keywords to be associated with anticipation of the new policy, which would eventually mellow down with passage of time. The final step, as shown in Section 4.3, is to gauge sentiments and emotions from these keywords so that one can associate overall public reactions to the policy. In combination, these three analyses present a comprehensive integrated framework to evaluate public reaction to policies.

We now discuss in detail the analysis of the collected data. For this section, each tweet was considered as a set of words, after ignoring a standard set of English stop words (like 'is', 'am', 'you') and the URLs contained in the tweets (if any). Few context-specific words like ‘odd’, ‘even’, ‘oddeven’, ‘Delhi’, etc. were contained in more than 90% of the tweets, hence these words were also ignored. Further, all words were converted to lower-case to unify lower and upper case occurrences of the same word. For instance, the tweet "*Delhi CM Arvind Kejriwal seeks PM Modi's help on the odd-even vehicle plan to curb air pollution https://t.co/7acapjs9xs*" would be considered as the set of words {delhi, cm, arvind, kejriwal, seeks, pm, modi, help, odd-even, vehicle, plan, curb, air, pollution}.

4.1 Topic Modeling

Our first aim was to identify the different topics (themes) that were being discussed by the Twitter population. Since our dataset has close to 70,000 tweets, it is infeasible to manually identify the topic discussed in each tweet. Hence, we use topic models, which are statistical models for discovering topics in a text corpus. We use the widely used topic modeling algorithm Latent Dirichlet Allocation (LDA) (16). LDA identifies topics as mixtures of terms (words), and treat documents as mixtures of topics. Each topic is represented by a probability distribution over the vocabulary of all terms, where the probability values convey the affinity for a given term to a particular topic. In simple terms, each topic (as identified by LDA) is a set of frequently co-occurring terms, and each document is assigned probabilities of belonging to the different topics. Further details of the LDA algorithm are omitted for brevity.

The LDA algorithm takes three input parameters: the number of topics **N**, a hyper parameter $\boldsymbol{\alpha}$ for the Dirichlet prior topic distribution, and a hyperparameter $\boldsymbol{\beta}$ for the Dirichlet prior word distribution (details omitted for brevity). Choosing optimal values for these parameters is a challenging problem and is not the focus of this work. Using the harmonic mean of log likelihood, we identified that the optimal number of topics would be 15 (see Figure 4). However, we also noticed that the keyword groups for five topics were quite difficult to translate into easily comprehensible themes. On the other hand, we were able to assign intelligible and relevant themes to the remaining ten topics. Hence, we use **N** = 10 topics, and $\boldsymbol{\alpha} = 0.5$, $\boldsymbol{\beta} = 0.1$; similar values for the parameters have been used in prior works which have attempted topic modeling over tweets (17).

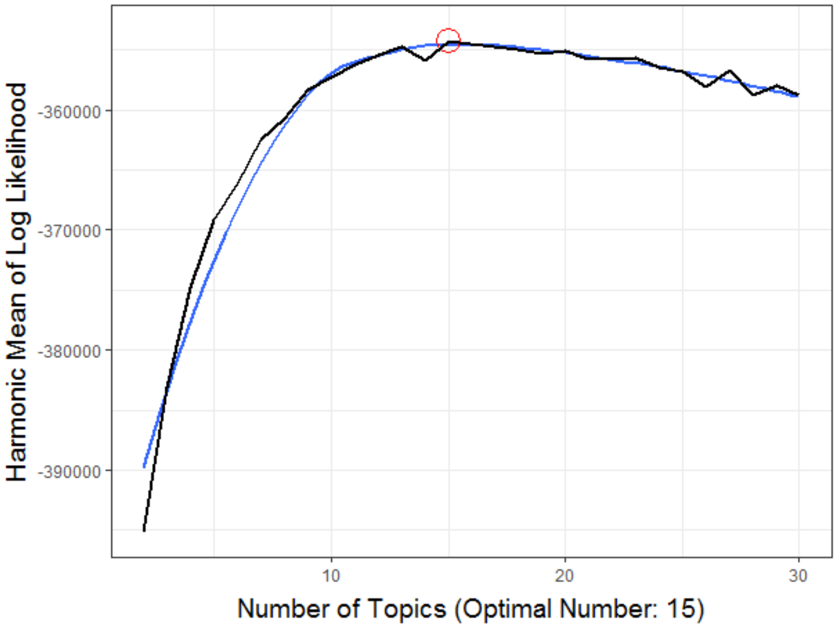


Figure 4. Optimal Topic Selection using Harmonic Mean of Log Likelihood

In summary, we were able to identify 10 topics, where each topic contains words that frequently co-occurred in the tweets. Table 3 shows the topics identified by LDA, along with some words from each topic. Manually observing the words in each topic, we also attempted to assign a 'theme' to each topic.

Table 3. Topics identified by topic modeling over the data

Topic ID	Theme	Few representative words
1	Exemption for women, CNG vehicles including buses, etc.	cars, women, cng, buses, exempted
2	Role of political & media influences	kejriwal, arvind, chief, minister, ndtv
3	Role of electronic media, police, civic volunteers, etc.	Plan, TimesOfIndia, police, volunteers
4	Impact on traffic	traffic, roads, working, Monday
5	Impact on air pollution	pollution, air, quality, levels, reduce
6	Alternative transport mechanisms	metro, public, transport, bus
7	Role of government and judiciary	govt, government, trial, court, asks
8	Public **support** towards the policy	oddEvenSuccess, arvindkejriwal, aamadmiparty
9	Public **opposition** towards the policy	Modi, fail, failure, oddEvenFails
10	General conversation	car, work, follow, happy, new, year

It is evident from Table 3 that a variety of themes were being discussed in relation to the policy, including the role of political parties, electronic media, police, and so on. Some public litigations were filed in relation to the policy, hence the role of the judiciary was also discussed. There were posts both supporting the policy (these posts mentioned Aam Aadmi Party, the ruling party in Delhi, which deployed the policy) as well as opposing the policy (these posts mentioned terms like Modi (the leader of the opposition party), fail, failure, etc.). The impact of the policy on traffic and pollution was also discussed. People also posted about alternative means of transport that came into increased usage.

The determination of discrete themes is extremely useful for the next step of the analysis, since emotions and usage frequency associated with keywords pertaining to a given theme will be useful in judging the importance of that theme. A relevant example would be that of alternative public transport. Did the other available options such as the Delhi Metro prove to be equal to the task of catering to increased footfall or did chaos ensue due to further crowding? Such questions can be answered only if we identify sentiments and emotions associated with the keywords for Theme No. 6, i.e. alternative transport mechanisms. Real-time policy evaluations would aid in reactionary policy interventions, such as increasing the frequency of metro service if the current system is unable to cope with the rise in passengers.

4.2 Temporal Analysis of Selected Issues

We next analyze the temporal variation of the number of tweets containing certain keywords (hashtags). This analysis gives us an idea of which issues were frequently discussed at different points in the duration of the study.

Figure 5a shows the temporal variation of tweets on some *negative* keywords (which seem to criticize the policy), like *#oddEvenFails*, *#oddEvenCorruption*. Most of the tweets containing these negative keywords were posted in the early stages (around January 1). However, these keywords were not posted as frequently in the later stages.

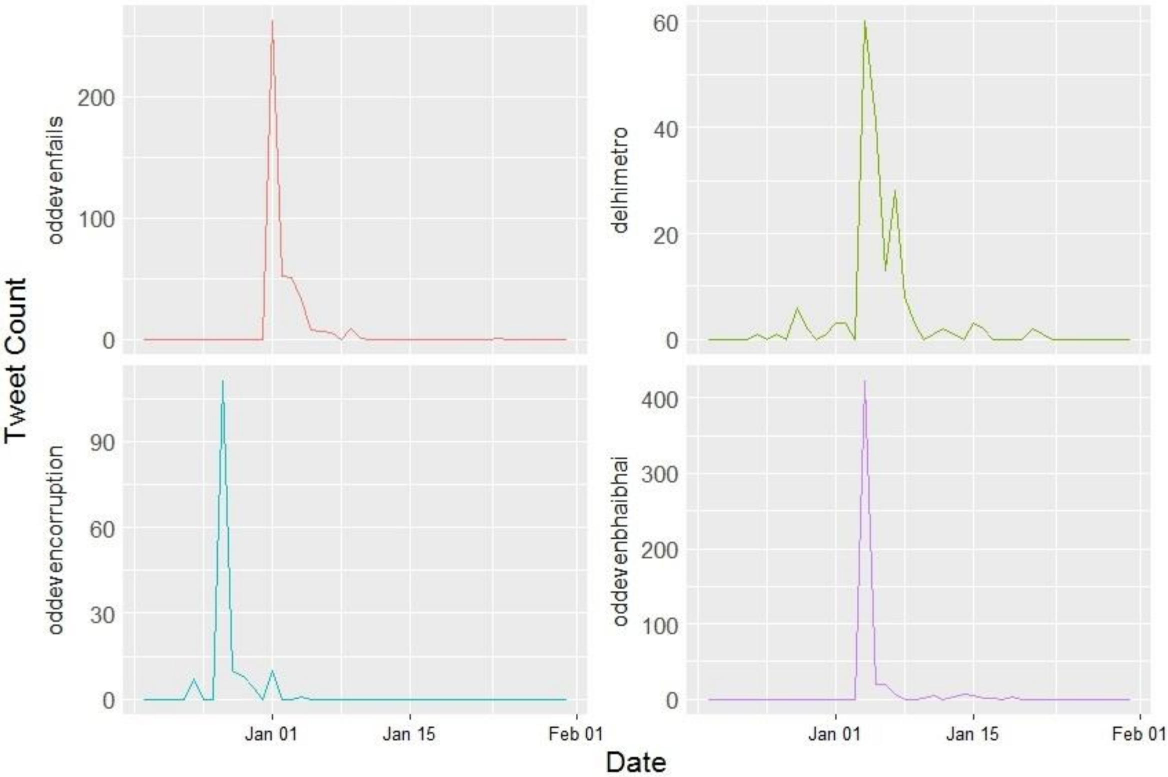


Figure 5a. Temporal analysis of negative keywords

Figure 5b shows the temporal variation of tweets containing some *positive* keywords (which supports the policy), like *#oddEvenSuccess*, *#kudosDelhi*, *#iPledgeForOddEven*. Some of these keywords were posted in the early stages (like *#kudosDelhi*, *#iPledgeForOddEven*) while some others were posted in the later stages as well (like

#congratsDelhi, *#delhiLovesKejriwal* were majorly posted immediately after January 15, the last day of the policy).

Overall, the observations from Figure 5a and 5b show that most of the posts containing negative keywords were posted in the early stages, while the posts containing positive keywords were posted both in the early and later stages. In general, these observations indicate that the pubic sentiment regarding the policy improved with time, thereby pointing towards a positive experience during the implementation period.

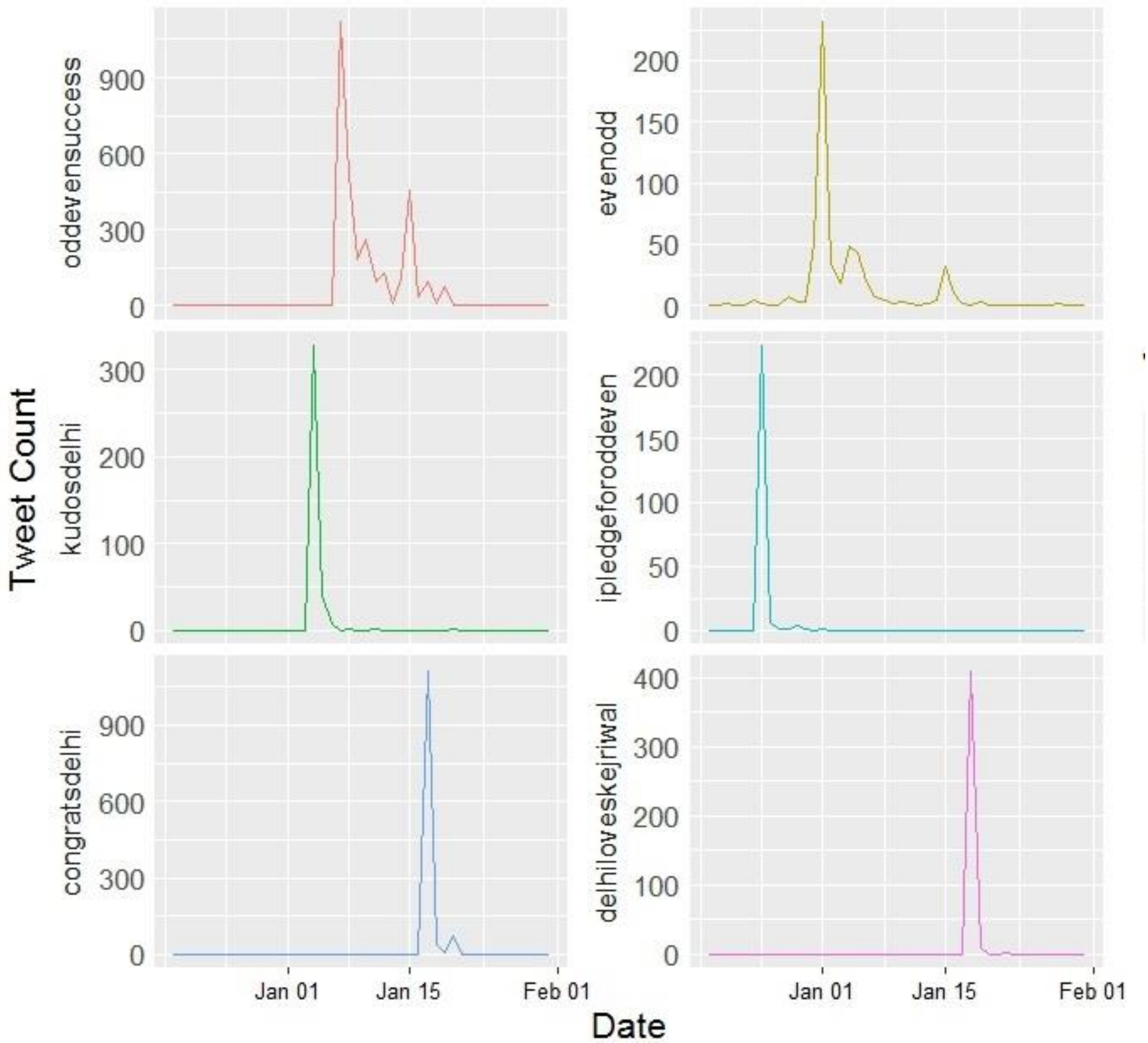


Figure 5b. Temporal analysis of positive keywords

It is interesting to observe that the keywords shown in both Figure 5a and 5b are purely transient in nature, i.e., they were discussed only during 2-3 days. On the other hand, some keywords were discussed during the entire period; Figure 5c shows the temporal variation in tweets for some such keywords, like '*carpooling*', '*public transport*', '*women*', etc. These patterns also hint about the relative importance of a particular keyword.

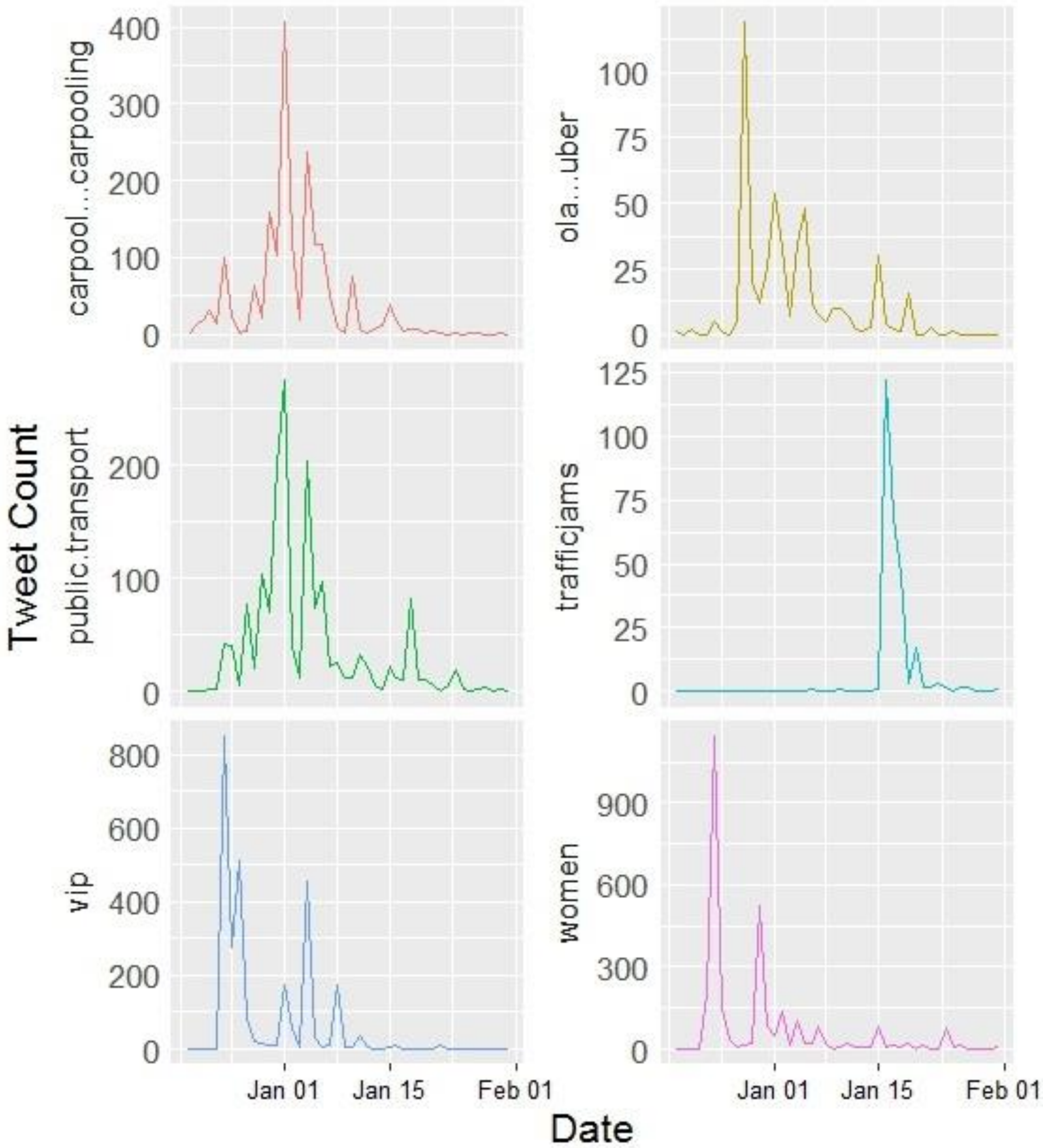


Figure 5c. Temporal analysis of some keywords that were posted throughout

4.3 Sentiment and Emotion Analysis

When a new policy like the Odd-Even policy is adopted, it is of utmost importance to gauge the public sentiment / opinion around the policy. In this section, we focus our attention on understanding the public sentiment around the policy, as evident from the tweets posted by them.

Sentiment analysis from text is a well-researched field in the Computer Science community, and several techniques to infer the sentiment in a piece of text are available. For sentiment analysis of a given text, we look for words / phrases that carry a positive or negative connotation, and attempt to infer the sentiment reflected in the text based on presence of such sentiment-words. For a given piece of text (e.g., a tweet), we get a sentiment where a negative score implies that the text reflects negative sentiment, a score of 0.0 implies that the text is neutral, and a positive score reflects positive sentiment. The key step is to use comprehensive lexicons for identifying the sentiment-words. In this study, we have combined two lexicons: (i) Hu & Liu's opinion lexicon, and (ii) AFINN-111 (18). Combination of two different lexicons increases the power of sentiment detection.

Once we have a sentiment-score for each tweet, we can also derive the sentiment-score of a keyword (e.g., a hashtag). For this, we collect all tweets which contain that keyword, and then observe the sentiment-scores of all these tweets. The cumulative sentiment score is the sum total of the scores for each tweet; a positive cumulative sentiment-score indicates that the overall sentiment around this keyword was positive. The average sentiment score is the cumulative sentiment score divided by the tweet volume (number of tweets containing the keyword). We also report the fraction of tweets (containing the keyword) that have positive scores and negative scores.

Table 4 reports the summary of our findings. Panel 1, of Table 4, reports the top keywords in terms of tweet volume. In addition to tweet volume, we also report the cumulative sentiment score, average sentiment score, and the fraction of positive and negative tweets (related to a specific keyword). Panel 2 and 3 reports the most positive and negative keywords (in terms of average sentiment score). We did not report hashtags/keywords with less than 100 tweets in our corpus.

Table 4. Sentiment analysis of keywords

Keywords	Number of Tweets	Overall Sent. Score	Avg. Senti Score	% of +ve Tweets	% of -ve Tweets
Top Trending Topics					
oddeven	78,406	13,707	0.17	37	30
pollution	22,531	-10,728	-0.48	21	52
oddevenformula	5,653	-2,949	-0.52	15	58
oddevensuccess	3,105	-1,637	-0.53	10	48
oddevenplan	2,896	-387	-0.13	26	33
Positive Trending Topics					
kudosdelhi	379	1,132	2.99	100	0
evenodd	307	477	1.55	56	5
congratsdelhi	1,223	940	0.77	56	19
delhiloveskejriwal	420	310	0.74	47	9
evenoddplan	221	157	0.71	43	16
Negative Trending Topics					
delhimetro	179	-189	-1.06	12	53
umanchain4pollutionfreedelhi	467	-465	-1.00	2	66
oddevencorruption	151	-134	-0.89	2	91
oddevenfails	428	-265	-0.62	5	32
oddevenbhaibhai	497	-281	-0.57	15	61

OddEven was the most popular keyword (volume-wise) and it was associated with overall positive sentiment. Next, *pollution* appeared in the second position; note that pollution control was the main objective of the Odd-Even policy. The overall sentiment-score around the keyword pollution was negative (-0.48), which shows that the public was not fully satisfied with the impact of the policy on pollution.

From Table 4, we can note that total number of positive tweets were higher than that of negative tweets. Figure 6 further confirms that the number of positive sentiment tweets was much larger than that of negative-sentiment tweets. Hence, in general, people supported the policy. However, there was large negative sentiment around certain keywords. For example, *OddEvenFails* displayed strong negative sentiment. Supporters of other political parties used these hashtags to criticize this policy. The public was also dissatisfied with some specific issues like the performance of the Delhi metro rail system (see the keyword *delhimetro* in Table 4).

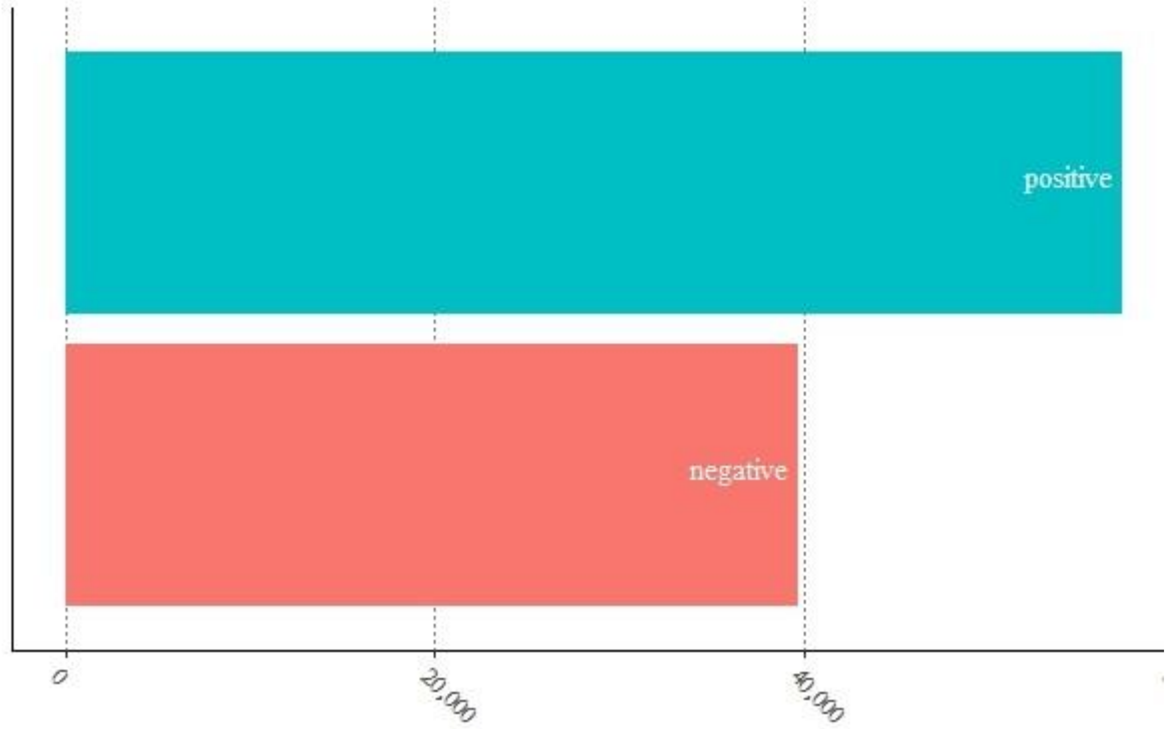
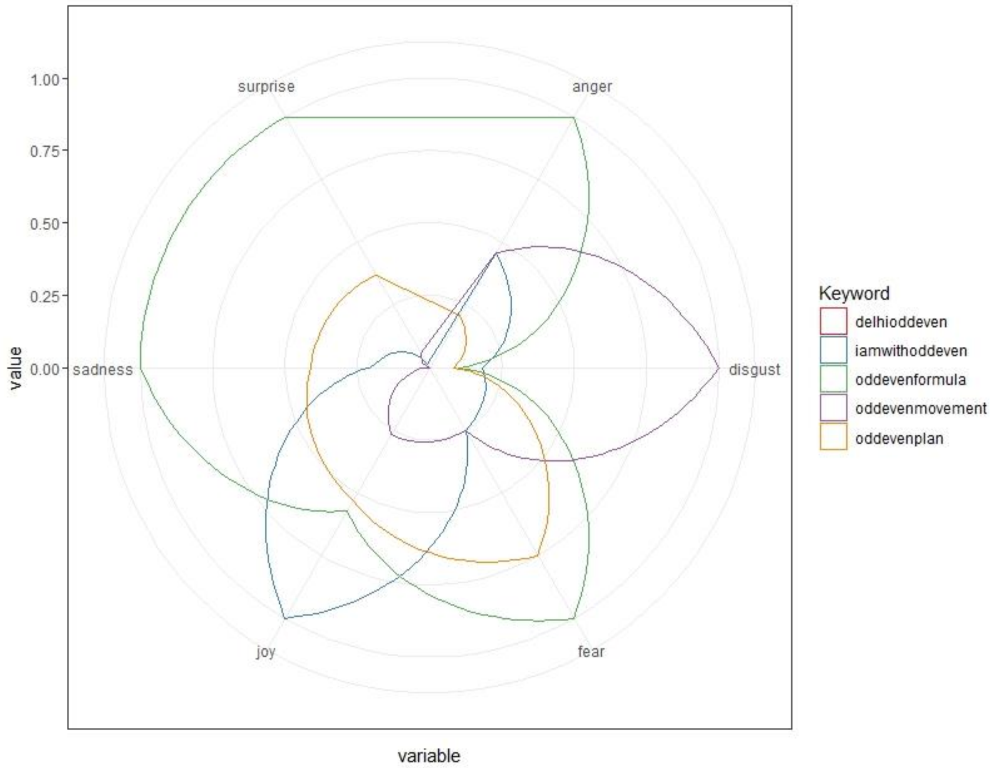


Figure 6. Cumulative Sentiment Analysis of our data

The analysis till now has used sentiment analysis techniques, which infer the polarity of a text as either positive or negative. However, such binary classification cannot capture the emotion / mood of the user who posted the content. On the other hand, emotion analysis can indicate the *joy* or *surprise* in a positive tweet. Again, it is important to note that *anger* and *disgust* are both negative, but they have some differences in terms of emotions. We now adopt emotion analysis techniques, which can identify the mood of the user posted the content.

We use a Naïve Bayesian classification method to explore our data for emotion identification. Naive Bayes is a combination of probabilistic algorithms for predicting the category (here, emotion) of a sample like a tweet. Bayesian classifiers assign the most likely category (emotion) of a tweet on the basis of its feature vector. We consider a feature as a piece of information that the algorithm can use to label the category for a particular tweet. In our text categorization task, this feature is equivalent to the word frequencies of a particular text. If features are independent of a given class then this can be represented as follows: $P(X|C) = \prod_{i=1}^{n} P(X_i \mid C)$, where $\mathbf{X} = (X_1, \ldots, X_n)$ is a feature vector and C is a category (emotion). This method returns the best fit for a particular text (tweet) in one of the six categories as follows - anger, disgust, fear, joy, sadness, and surprise. For example, the following tweet - "*successful in handling expected huge rush on all the lines on the first working monday of the year delhi metro statement oddeven"* will be classified into the category "joy". The efficiency of Naïve Bayes classification depends on the feature vector lexicon. In this study, we have employed the emotion lexicons developed by Strapparava et al. (19).

Figure 7 reports the emotion of tweets containing some specific keywords like *carpool*, *cab service* (like *ola*, *uber*) or *women*. *Women*-related tweets show anger because many people felt that women would be forced to take carpools due to the policy, and they would not be safe in carpools. Topics like *iAmWithOddEven* are classified as "joy". Figure 8 reports the emotion analysis of all tweets in the entire corpus. The results again show that people generally supported the government on this policy, because presence of sadness or anger is much lower compared to presence of positive emotions like joy and trust.



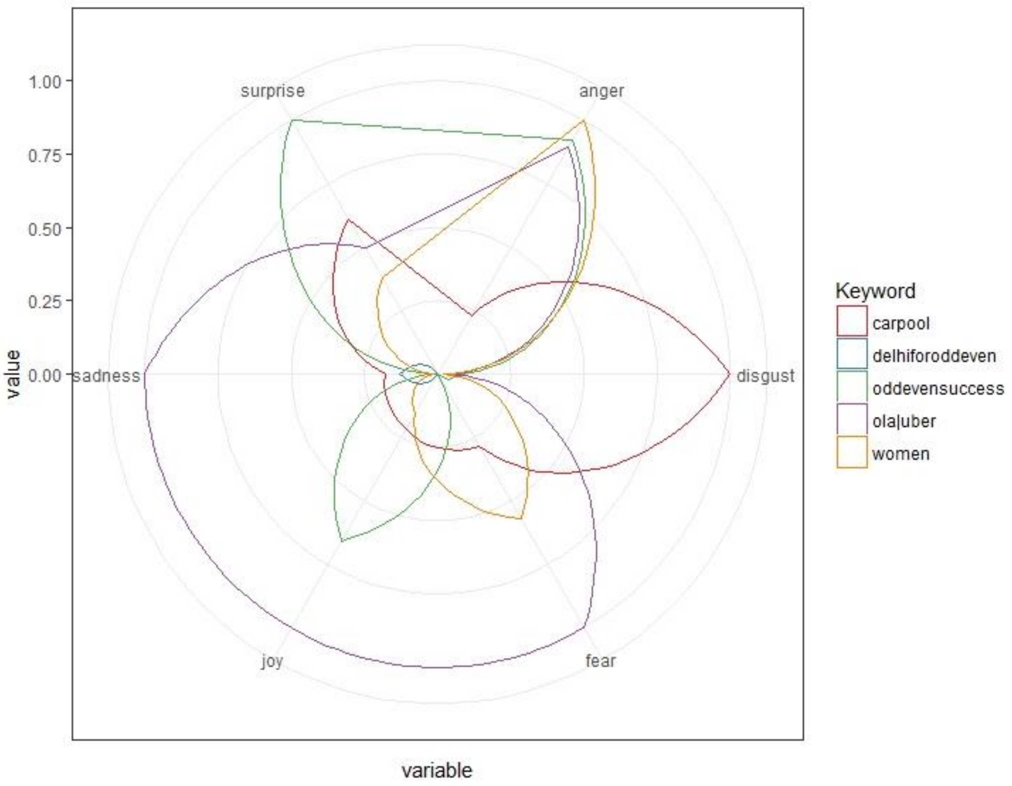


Figure 7. Emotion Analysis of tweets containing selected keywords

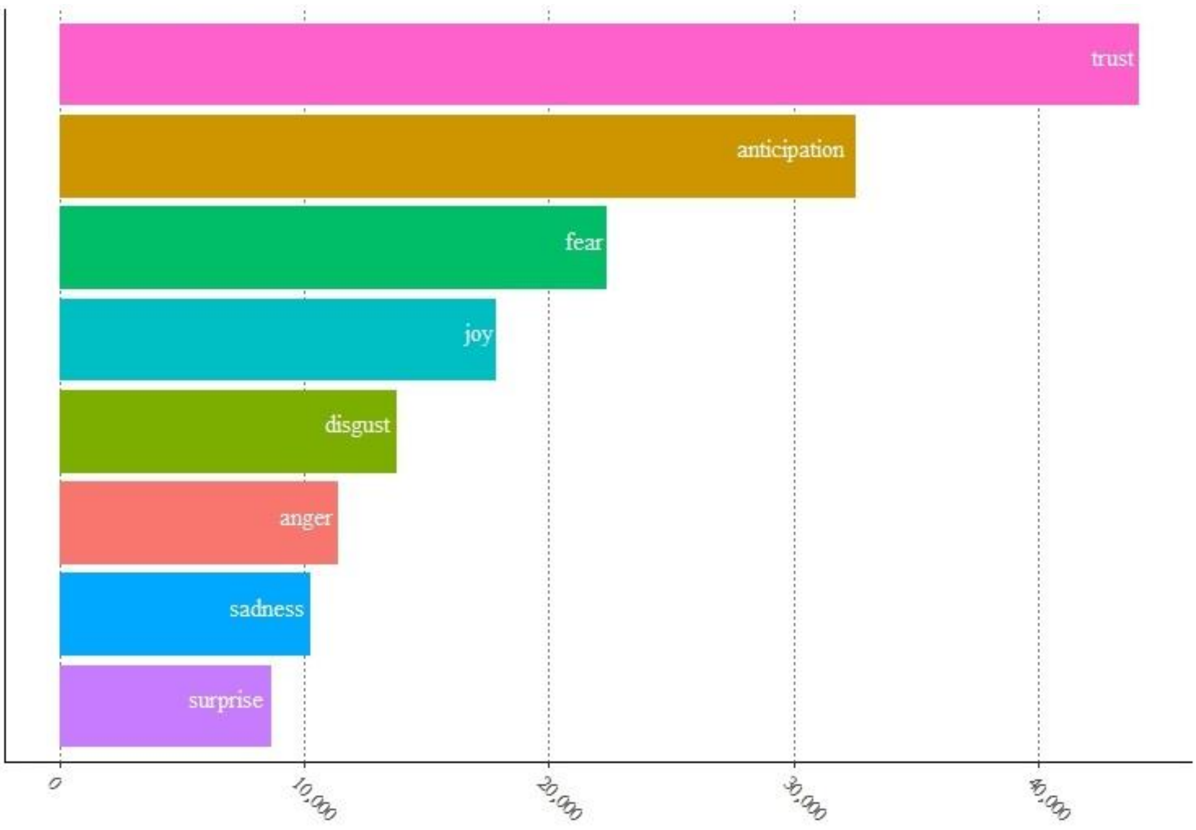


Figure 8. Emotion Analysis of all tweets in the entire dataset

5. DISCUSSIONS AND CONCLUSION

The emergence of information and communication technologies over the past decade have influenced the functioning of local, state and national governments. Use of these technologies in the decision-making process enables inclusion of citizens apart from providing a sense of ownership and community involvement to the individual. Prattipati (20) noted that there is a wide variance of the use of such technologies in government initiatives and functioning across the world. Developing nations are lagging behind in this aspect, which may be attributed to non-inclusion of expert manpower in the government ranks. A case study of India was presented by Gupta, Dasgupta and Gupta (21) while exploring this issue, and it was found that there is much left to achieve on this front. To that effect, this study proposes a methodology for enabling the use of Twitter data in order to analyze public sentiments and consequences of Government policies. The methodology is demonstrated through a case study of the Odd-Even policy in Delhi, India.

The present study only used the textual content of the tweets. However, as mentioned in the section on data collection, a lot of other information is also available from Twitter, such as the identity of the users who posted the tweets, which can be utilized as well. For instance, people who posted tweets complaining about specific issues can be contacted over Twitter, to know more about the problems they are encountering. Again, online surveys / questionnaires can be prepared and distributed to selected users via channels like Twitter / Facebook, which can potentially give lot of valuable feedback to the Government. A particular drawback of Twitter data in India is that most users wish not to disclose their location by turning off their geo-location marker on smartphones. This creates a shortcoming in analyzing the spatial distribution of users reacting to the policy.

The proposed methodology is particularly useful in identifying discrete themes that are important as per the discussions people are involved in on social media platforms. This might help guide targeted policy interventions. It was noticed in the course of our analysis that there is an overall positive sentiment with regard to the policy, but specific concerns were identified when it came to women's safety, pollution and public transport. It is extremely useful that this particular policy was simply a test run to assess its feasibility, an ideology that might be followed for policies in the future. In case the policy is implemented on a long-term basis in the near future, our analysis can help guide improvements in its implementation. For example, the Delhi Metro should be operated with a higher frequency since the current frequency is not enough to cater to the increased demand. Citizens are also concerned that this policy might induce more trips by on-demand mobility services like Uber, leading to increased pollution and congestion. Therefore, it might be conducive to implement an incentive scheme for carpooling and ridesharing, perhaps even introducing high-occupancy-vehicle (HOV) lanes as a long-term measure. Finally, in light of maintaining a safe transportation system for women, on-road police patrols can be instituted (especially after dark) and increased security cameras (CCTVs) may be installed at bus stops and metro stations.

While these policy interventions are extremely useful, we must note that this study is a post-analysis of the tweets posted during the period. The benefits of a real-time analysis (when the policy is in place) will prove to be significantly higher, so that the problems faced by the people are known immediately and can be solved. For instance, if the Government can know in real-time that people are dissatisfied due to carpooling, they can deal with this problem. There is sufficient scope to do such analysis in real time, which will increase the applicability of data collected from Twitter. Therefore, our recommendation would be to implementation test runs of a policy for a fortnight and employ the proposed methodology to identify major public concerns. Once those are incorporated in the actual policy implementation, real-time monitoring would be useful to judge the usefulness of the additional measures as well as mitigate further problems faced by the public.

Although Twitter data might prove to a viable source for development of such techniques, issues of data reliability persist. However, online social media sources like Twitter or Facebook seem to be better alternatives when comparing metrics like cost, time and public responsiveness. This is an emerging field of enquiry and subsequent discoveries or innovations will lead to better methods in data processing and analysis. We simply point out the effectiveness of such an approach and imply the usefulness of its application in shaping future policies along with enablement of citizen participation. The proposed methodology may prove to be useful in both developed and developing countries; however, we stress upon its use in developing nations like India due to lack of robust data-intensive policy evaluation techniques on a national or local scale. Due to the visible limited application of such data in the field of transportation, we envision that an increased utilization of social media data would deeply enrich intelligent transportation system (ITS) applications. As a parting note, the authors suggest exploring real-time traffic sentiment analysis as a future research direction.

REFERENCES

1. Meyer, M. D. Demand management as an element of transportation policy: using carrots and sticks to influence travel behavior. *Transportation Research Part A: Policy and Practice*, Vol. 33, No. 7–8, 1999, pp. 575-599.
2. Just, R. E., D. L. Hueth, and A. Schmitz. *The welfare economics of public policy: a practical approach to project and policy evaluation*. Edward Elgar Publishing, 2005.
3. Bengston, D. N., J. O. Fletcher, and K. C. Nelson. Public policies for managing urban growth and protecting open space: policy instruments and lessons learned in the United States. *Landscape and Urban Planning*, Vol. 69, No. 2–3, 2004, pp. 271-286.
4. Hatzopoulou, M., and E. J. Miller. Transport policy evaluation in metropolitan areas: The role of modelling in decision-making. *Transportation Research Part A: Policy and Practice*, Vol. 43, No. 4, 2009, pp. 323-338.
5. Thomas, E. N., and J. L. Schofer. Strategies for the evaluation of alternative transportation plans. *NCHRP Report*, No. 96, 1970.
6. Sharma, S. India Has 22.2 Million Twitter Users: Report. In *The Huffington Post*, New Delhi, 2015.
7. Leskovec, J., A. Rajaraman, and J. D. Ullman. *Mining of massive datasets*. Cambridge University Press, 2014.
8. ILS. *Internet Live Stats - Twitter Usage Statistics*. http://www.internetlivestats.com/twitter-statistics/. Accessed July 29, 2016.
9. Jain, V. Prediction of movie success using sentiment analysis of tweets. The International *Journal of Soft Computing and Software Engineering*, Vol. 3, No. 3, 2013, pp. 308-313.
10. Tumasjan, A., T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM*, Vol. 10, 2010, pp. 178-185.
11. Mai, E., and R. Hranac. Twitter interactions as a data source for transportation incidents. In Proc. *Transportation Research Board 92nd Ann. Meeting*, 2013.
12. Schweitzer, L. How are we doing? Opinion mining customer sentiment in US transit agencies and airlines via Twitter. *In Transportation Research Board 91st Annual Meeting*, 2012.
13. Collins, C., S. Hasan, and S. V. Ukkusuri. A novel transit rider satisfaction metric: Rider sentiments measured from online social media data. *Journal of Public Transportation*, Vol. 16, No. 2, 2013, p. 2.
14. Agarwal, A., B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. Sentiment analysis of twitter data.In Proceedings of the workshop on languages in social media, *Association for Computational Linguistics*, 2011. pp. 30-38.
15. Tao, K., F. Abel, C. Hauff, G.-J. Houben, and U. Gadiraju. Groundhog day: near-duplicate detection on Twitter. Presented at *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro, Brazil, 2013.
16. Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent dirichlet allocation. Journal of machine Learning research, 3 (Jan), pp. 993-1022.
17. Zafar, M. B., Bhattacharya, P., Ganguly, N., Gummadi, K. P., and Ghosh S., 2015, Sampling content from online social networks: Comparing random vs. expert sampling of the twitter stream, ACM Transactions on the Web 9(3), pp. 12:1 – 12:33, ACM.
18. Khatua, A., Khatua, A., Ghosh, K. and Chaki, N., 2015, January. Can# Twitter_Trends Predict Election Results? Evidence from 2014 Indian General Election. In 2015 48th Hawaii International Conference on System Sciences (HICSS), (pp. 1676-1685). IEEE.

19. Strapparava, C. and Mihalcea, R., 2008, Learning to Identify Emotions in Text, Proceedings of the 2008 ACM Symposium on Applied Computing (SAC), pp. 1556-1560.
20. Prattipati, S. N. Adoption of e-governance: differences between countries in the use of online government services. Journal of American Academy of Business, Vol. 3, No. 1/2, 2003, pp. 386-391.
21. Gupta, B., S. Dasgupta, and A. Gupta. Adoption of ICT in a government organization in a developing country: An empirical study. The Journal of Strategic Information Systems, Vol. 17, No. 2, 2008, pp. 140-154.