**Transp. in Dev. Econ. manuscript No.**
(will be inserted by the editor)

# Public Opinion Analysis of Transportation Policy using Social Media Data: Case Study on the Delhi Odd-Even Policy

**Pranamesh Chakraborty · Anuj Sharma**

the date of receipt and acceptance should be inserted later

**Abstract** Twitter, a microblogging service, has become a popular platform for people to express their views and opinions on different issues. Sentiment analysis of tweets can help in understanding public opinion on different government decisions. This paper used Twitter data to extract sentiments of people during the Phase 1 and Phase 2 of the odd-even policy implemented by the Delhi government to curb the air pollution and improve traffic flow. In this study, we used four different lexicon based approaches: Bing, Afinn, National Research Council (NRC) emotion lexicon, and Deep Recursive Neural Network based Natural Language Processing software (CoreNLP) to extract sentiments from tweets and thereby assess overall public opinions. The daily trend obtained for each phase was normalized with the number of tweets and then compared using Granger causality test. The causality test results showed that the trends obtained during the two phases were significantly different from each other. In particular, public sentiments were found to mostly turn negative during the later stage of the Phase 2 which indicates fading away of the public enthusiasm and positiveness towards the policy during the later stages of the policy implementation.

**Keywords** twitter · sentiment analysis · odd-even · delhi

P. Chakraborty (corresponding author)
PhD student, Department of Civil, Construction and Environmental Engineering
Iowa State University
Tel.: +1-515-357-9076
E-mail: pranames@iastate.edu

A. Sharma
Associate Professor, Department of Civil, Construction and Environmental Engineering
Iowa State University
Tel.: +1-765-430-0023
E-mail: anujs@iastate.edu

## 1 Introduction

Delhi and its National Capital Region, with a population of 16.2 million constitutes 1.35% of India's urban population [1]. A combination of 200 km of metro rail, buses, and different para transit options (auto rickshaws) serve the public transportation of Delhi. Although the share of cars in community trips in Delhi is relatively low compared to other major cities of the world, recent studies have ranked Delhi as the "worst" polluted city in terms of environment performance index [2]. To tackle the high pollution level, the Government of the National Capital Territory of Delhi, India, implemented the odd-even policy, where only odd numbered cars were allowed to operate on odd numbered dates and cars with even number plates on even dates between 08:00-20:00 hours. This policy was implemented in two phases. Phase 1 was implemented in winter season, between January 1 and January 15, 2016 and the Phase 2 during summer from April 15 to April 30, 2016. Twenty different categories were exempted from this rule which included motorized two-wheelers, electric and hybrid cars, cars driven by women, cars of very/very important persons such as parliamentarians, emergency vehicles such as ambulance, fire brigade, etc.

Similar such odd-even policy has been implemented in the past in other parts of the world too. This includes Buenos Aires (Argentina), Bogota (Columbia), Mexico City (Mexico), Manila (Philippines), Lagos (Nigeria), and Beijing (China) during the Olympic Games [3]. However, past studies suggest that although such driving restriction policies may reduce pollution and congestion in short run, but in long run people learn to cope with the restrictions by shifting to two-wheelers, buying second older cars, etc. [4]. Hence, recent studies have been done to find out the efficacy of the policy implementation in Delhi.

Tiwari et al. [5] and Garg et al. [6] conducted separate studies to find out the impact of the Phase 1 of the odd-even policy on the atmospheric pollution level. Both the studies found that the policy didn't had any significant impact on reducing pollution level. Zanouda et al. [7] also performed a study to understand public opinion on the odd-even policy based on the political and spatio-temporal factors. Their study used Twitter data of before, during, and after-period of Phase 1. The study showed that public perception of the experiment outcome was strongly influenced by their political affiliation for the people residing outside Delhi and India. However, these studies didn't cover the Phase 2 of the policy implementation and hence couldn't compare if similar observations prevailed during the $2^{nd}$ phase. Other studies have looked into both phases of the policy implementation and compared the similarities/dissimilarities between the two phases.

Chelani [8] analyzed the concentration of fine particulate matter, $PM_{2.5}$ (particulate matter 2.5 mm or less in diameter) during the odd-even policy dates. Similarity and causality analysis were performed to find out the local and regional influence. The study found that $PM_{2.5}$ was higher during Phase 1, which was in winter compared to Phase 2 which was in summer. However, it's worth mentioning that the atmospheric conditions during winter do not allow dispersion of pollutants compared to other seasons which might result in higher pollutant concentration during Phase 1. Kumar et al. [9] also conducted similar study to find out the influence of the policy on fine and coarse particles. Their study suggested that even though certain hours of the trial days generated cleaner air, but the overnight emissions from heavy goods vehicles made them overall ineffective. $PM_{2.5}$ and $PM_{10}$ (particulate matter 10 mm or less in diameter) concentration were found to increase during Phase 1 compared to the previous year (2015) emissions while the concentrations decreased for Phase 2 compared to the previous year. However, it was also observed that wind speed and ambient temperature could be possible explanations for such observations rather than odd-even policy implementation. Mohan et al. [10] evaluated car, two-wheeler, bus, and auto rickshaw flow rates and also car occupancy rates during the traffic restriction experiment period. Their study found that the car flow rates decreased by 20% while other modes of vehicles increased during the analyses period. Although, no significant difference in proportion of car users were observed between Phase 1 and Phase 2, but car-occupancy rates were found to decrease in Phase 2 suggesting that less people opted for car-sharing in Phase 2. However, they also state that these differences might be due to small sample errors and not due to the policy. Thus, it can be seen that significant amount of research has been performed to evaluate the effects of the odd-even policy on the environmental and traffic aspect. Studies have been also performed to evaluate their differences in the effects between the two phases. However external factors (different atmospheric conditions during the two phases) or low sample sizes (for car occupancy rates) didn't allowed them to find out the exact differences between the impacts of the two phases. The main objective of this study is to understand the public opinion towards the policy and its change over the two phases of the implementation. Social media platforms (e.g. Twitter and Facebook) can be used to extract such information and mine them to get useful information.

Social media has gradually evolved to be a popular platform for people to express their opinions on current trending topics. Twitter is such a platform where users can post brief text updates (maximum 140 characters) or multimedia such as images or audio clips. Researchers are using the tweets to find out general public perceptions on a variety of topics [11]. These sources have been used for monitoring political sentiments, predicting election results [12], detecting tension in online communities [13], understanding sentiments on new product launch [14], etc. Besides these, Twitter has also been used for tracking complex real-time events like natural disasters [15] [16], road hazards detection [17], and disease propagation [18]. Collins et al. [19] used Twitter data for monitoring sentiments of the riders of the Chicago Transit Authority public transit system. Luong et al. [20] also conducted a similar study to find out public opinion of the light rail service in Los Angeles. Sasaki et al. [21] performed a feasibility study using Twitter as a sensor for detecting transportation information. Recently, Sharma et al. [22] used Deep Belief Network (DBN) to classify the sentiments of tweets posted by users during the odd-even policy in Delhi. They used six different models based on the DBN classifier to find out the performance of the proposed methods. This paper on the other hand used four different lexicon-based approaches to perform the sentiment analysis of the tweets collected during the Phase 1 and Phase 2 of the odd-even policy implementation. The accuracy of the methods were evaluated based on 500 randomly sampled tweets from the given dataset. Then, the overall daily trends of the sentiments during the two phases obtained from the different methods were compared based on Granger causality test to check the similarity/dissimilarity between the two phases. The next section provides the details of the data used in this study. Section 3 gives the details of the methodology used in this study followed by the results in Section 4. The final section provides the conclusion along with the limitations of the study and the scope of future work.

## 2 Data Description

The twitter data for the Phase 1 of the policy implementation were bought using the Full Archive Search API (application program interface) provided by GNIP [23]. For the Phase 2 of the policy implementation, Twitter Streaming API [24] was used for collecting tweets in real-time related to the odd-even policy. The popular hashtags (e.g., #OddEven, #oddevenplan, #OddEvenRule, #odd- evenformula, #OddEvenPolicy, #oddevendobara) were used for extracting the relevant tweets for the two phases of the policy implementation. Relevant tweets were downloaded from 9 days before the policy implementation to the entire 15 days of the policy implementation. However, the number of tweets were found to be significantly lower (less than 100) during the last 4 days of the Phase 2. This may be probably due to technical issues in the downloading process or maybe due to sampling used by Twitter Streaming API. So, the last 4 days of both the phases were removed from the analysis. A total of 650,000 tweets were obtained during the Phase 1 from 23rd December, 2015 to 11th January, 2016 while 180,000 tweets were obtained during the Phase 2 from 6th April, 2016 to 25th April, 2016. Using the Full Archive Search API (Enterprise version) for downloading the tweets during the Phase 1 of the policy resulted in larger number of tweets compared to the tweets of the Phase 2 obtained using the free Twitter Streaming API. Figure 1 shows the number of tweets collected during each day of the analysis period. It can be seen that the volume of tweets increased during the initial period of the policy implementation (around 1st January and 15th April) for both the phases. Retweets were also included in the data source with the assumption that if a user retweets, it indicates that the user is also having a similar opinion or sentiment.

### 2.1 Pre-processing

The twitter data often contains noise such as RT for retweets, external website links or URLs, @usernames, etc. which needs to be removed before the sentiment analysis task. Our preprocessing step involved the standard preprocessing steps used in previous literature [25], [26] which includes removal of (a) URLs, (b) hashtags, (c) stop words such as ‘a’, ‘an’, ‘the’, etc., (d) usernames, (e) unnecessary spaces, (f) punctuation marks, (g) numbers, and (h) special characters such as emotions or non-English alphabets such as Hindi, etc. Finally, the stemming process is applied to convert all inflected words to its root form called stem. For example, automatic, automation, and automate are converted to its stem form automate. Snowball stemmer [27], the popular stemming package is used for this purpose.

## 3 Methodology

Sentiment analysis of tweets involves determination of the polarity of the tweets, whether it is expressing positive, negative or neutral sentiment towards the topic/subject. Hence sentiment classification can be also termed as polarity determination. Four different classes of twitter sentiment analysis approaches have been identified in the literature [21]

– Machine learning
– Lexicon based
– Hybrid (Machine learning & Lexicon based)
– Graph based

A majority of machine-learning methods involves building a classifier from machine-learning domain trained on different features to detect sentiment of tweets. The common classifiers used are Support Vector Machines, Naïve Bayes, Logistic Regression, Random Forest, and Conditional Random Field. Details of such application can be found in [26], [28]. Lexicon based methods, on other hand, use a pre- determined list of positive and negative terms to determine the polarity of the tweets. Hybrid approach methods combine lexicon-based and machine-learning methods while graph-based methods use social network properties to achieve better performance [29], [30]. This study used four different lexicon-based methods to determine overall public opinion on the odd-even policy implemented by the Delhi government. Details of the methods used in this study are discussed next.

Different lexicon-based methods exist which utilize different sets of pre-determined list of opinion words to determine the overall polarity of the tweet or text. In this paper, we used four such lexicon-based methods namely (a) Afinn (b) NRC (c) Bing and (d) Stanford CoreNLP to determine the sentiments of the extracted tweets. The Afinn lexicon [31] includes 2,477 English words with positive words scored from 1 to 5 and negative words from -1 to -5. The word list is focused on language commonly used in microblogging platforms like Twitter and hence contains acronyms, web jargons and slang words too. The NRC lexicon [32] is also a similar word-emotion association lexicon containing more than 14,000 distinct words created by using the crowdsourcing Amazon Mechanical Turk. The Bing [33] lexicon is also a similar lexicon dictionary containing around 6800 positive and negative English opinion words. All these words assign points for each positive and negative word in a tweet and then sum up these points to find out the overall sentiment of the tweet. The Stanford CoreNLP method [34], on other hand, not only uses the positive and negative words, but also utilizes the order of the words to build the overall polarity of the sentiment. The model is based on Recursive Deep Neural Network that builds on top of grammatical structures.

### 3.1 Evaluation of sentiment classification

Each of these four sentiment classification methods are applied on the extracted tweets to find out the sentiments. Five

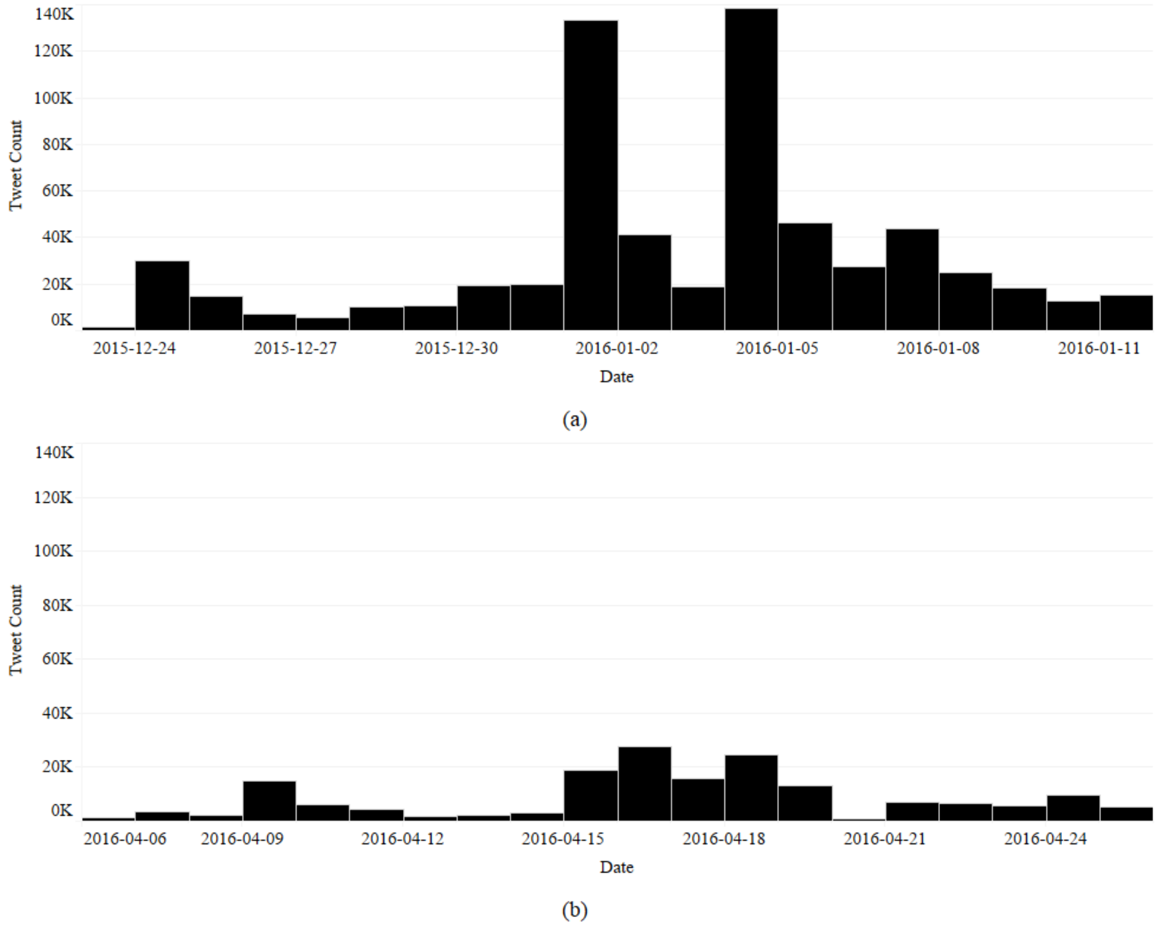



Fig. 1: Daily number of tweets collected during (a) Phase 1, and (b) Phase 2 of the odd-even policy

hundred randomly sampled tweets are selected and hand-annotated into three classes: positive, negative and neutral. Precision and recall (Equations 1 and 2 respectively) are computed for each class and then the average precision and recall is determined. Finally, the F-measure (Equation 3) is also computed. The metrics are same as the ones used in past studies [13, 35].

$$precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{1}$$

$$recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{2}$$

$$F = 2 \times \frac{TruePositive}{TruePositive + FalsePositive} \tag{3}$$

### 3.2 Daily trend comparison

The sentiment value assigned to each tweet by the given lexicon-based methods are determined and aggregated to get the overall sentiment of each day, denoted as $SS_d^m$, similar to the study by Collins et al. [19]. Here, $d$ denotes the date and $m$ denotes the method. Since the scale of sentiment scores given by each method is different, $SS_d^m$ is normalized based on the range of sentiment score ($RS^m$) given by method $m$. $RS^m$ is obtained by finding out the absolute difference of maximum and minimum sentiment score assigned to the analyzed tweets by the method $m$. Also, to account for the varying number of tweets obtained for each day, $SS_d^m$ is also normalized by the daily number of tweets ($n_d$). The daily normalized sentiment strength, denoted by $NSS_d^m$, is given by Equation 4. The variation of $NSS_d^m$ depicts the daily trend of public opinion on the odd-even policy.

$$Normalized\ Sentiment\ Strength\ (NSS_d^m) = \frac{SS_d^m}{n_d \times RS^m} \tag{4}$$

The daily trend of the sentiment scores obtained by each method are then compared to find out if the two phases of policy implementation had similar reactions among public or not. While studies have been done comparing the effect of the two phases of the odd-even policy on pollution and traffic level, our objective here is to find out if the public sentiments were similar during the two phases. To facilitate this, Granger causality test [36] is performed.

Granger causality test checks if predictions of a variable $Y$ can be improved by using its own past values and also past values of another variable $X$ rather than using its own past values only. Let $Y$ follows a univariate linear autoregressive models given by Equation 5. Then, the autoregression is augmented including the lagged values of the variable $X$, given by Equation 6. Granger test can be performed by an $F$-test which reports the Wald's statistics for the joint hypothesis given in Equation 7. Rejection of the null hypothesis can be inferred as "$X$ Granger-causes $Y$". It is worth mentioning here that Granger causality isn't in true sense a test for "causality". It is simply a $F$-test which checks if the predictions of $Y$ can be improved using past values of $X$ too along with that of $Y$, and it is denoted as "$X$ Granger-causes $Y$" instead of true causality.

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + ... + \alpha_m y_{t-m} + \varepsilon_t \tag{5}$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + ... + \alpha_m y_{t-m} + \beta_1 x_{t-1} + ... + \beta_l x_{t-l} + \varepsilon_t \tag{6}$$

$$\beta_1 = \beta_2 = ... = \beta_l \tag{7}$$

## 4 Results

Table 1 shows the sentiment scores assigned by each sentiment classification method to five sample tweets. Except CoreNLP, the remaining three methods do not take into the structure of the sentence. Instead, they look into words in isolation, giving positive points for positive words and negative for negative words. Hence they fail to classify correctly some complex structured tweets (e.g., Tweet # 5 in Table 1). Also as discussed in Section 3, each sentiment analysis method has a different scale of scoring. Hence the sentiment scores of the sample tweets in Table 1 can be found to differ for different methods. To recall, this issue is handled by normalizing the daily sentiment scores of each method based on the corresponding range of the corresponding method (see Equation 4).

In order to find out the overall accuracy of these methods, 500 randomly sampled tweets are hand-annotated and compared with the output obtained from each sentiment analysis method. Table 2 shows the precision, recall, and $F$-score (Equations 1, 2 and 3) of each method. As shown in Table 2, the precision, recall, and $F$-score of CoreNLP are substantially lower compared to the ones obtained for the remaining three methods. One possible reason can be the CoreNLP method is presently designed for analysing sentiments of proper English sentences rather than tweets which generally consist of a significant amount of web jargon. Hence CoreNLP is found to perform poorly in this case. Due to poor performance of CoreNLP in the test dataset, it has been excluded from further analysis and the remaining of the analyses have been performed using the remaining three algorithms (Bing, Afinn and NRC).

Table 2: Precision Recall values for each method

| Method | Precision | Recall | F-score |
|---|---|---|---|
| Bing | 0.688 | 0.683 | 0.685 |
| Afinn | 0.703 | 0.673 | 0.688 |
| NRC | 0.632 | 0.76 | 0.69 |
| CoreNLP | 0.483 | 0.388 | 0.43 |

### 4.1 Daily trend estimation and comparison

The sentiments obtained from the tweets are then aggregated to obtain the daily trend of the normalized sentiment strength, *NSS* (Equation 4). To recall, normalization is performed to take into account the daily variation of the number of tweets and the scale of sentiment scores assigned by each method. To compare the trends during the two phases of the policy implementation, relative dates are used with Day 1 being the start date of the policy implementation (i.e., 1$^{st}$ January and 15$^{th}$ April respectively). Figure 2 shows the trend of *NSS* obtained from each sentiment analysis method during the analyses periods. It can be seen from Figure 2 that even though people were enthusiastic during the initial period of the Phase 2 (April 15$^{th}$), however the sentiment scores kept decreasing during the later phase of the policy implementation period. On the other hand, sentiment scores were steady and mostly positive even during the later stage of the Phase 1. To quantitatively compare the similarity/dissimilarity of the trends obtained during the two phases, Granger causality test is performed. Akaike's Information Criterion is used to find out the optimal lag length ($l$ in Equations 6 and 7). The optimal lag length ($l$) was found to be equal to 2 for all methods. We used default standard errors in this study. In future studies, cluster robust variance estimators (i.e., clustered standard errors) can be used to get more robust estimates. Table 3 provides the regression coefficients ($\beta_1, \beta_2$ of Equations 6 and 7) along with corresponding $p$-values, the overall $F$-statistics, and overall $p$-value for each sentiment analysis method. To recall, the null hypothesis states that there is no Granger causality between the two time-series. Hence, higher $p$-values ($> 0.05$) given in Table 3 show that we fail to reject the null hypothesis that there is no Granger causality between the trend of sentiments obtained during the two phases. None of the regression coefficients were found to be significant at 95% confidence level. In other words, there is significant difference between the trends obtained from the two phases. This is also evident from Figure 2 which shows the sentiments scores dropped during the later stage of the Phase 2 of the policy implementation.

Table 1: Sentiment score assigned by the different methods

| # | Tweet | Sentiment Score assigned by method Bing | Afinn | NRC | CoreNLP |
|---|---|---|---|---|---|
| 1 | Another vile attempt to make #OddEven a failure, this time with a dangerous excuse. | -4 | -6 | -2 | -5 |
| 2 | #OddEven is a gimmick apart from reducing vehicles on roads. Kejri is burning tax money on publicity and people keep choking on diesel fumes | -3 | -2 | -0.9 | -1 |
| 3 | Delhi #oddeven: @Olacabs, @Uber surge pricing puts commuters in a jam, reports @mallicajoshi | -1 | 0 | 1 | 0 |
| 4 | Smooth ride today #OddEven #OddEvenDobara less of cars only odd ones showing up... Good show #delhi | 1 | 1 | 1 | -1 |
| 5 | Even in terms of traffic, #OddEven doesn't seem to be as successful as last time. | 1 | 3 | 1 | -3 |
| 6 | "RT @IndiaToday: #OddEvenPlan was supported not just by Delhi people but even, judges car-pooled and walked to work: #Kejriwal | 2 | 2 | 1 | -1 |

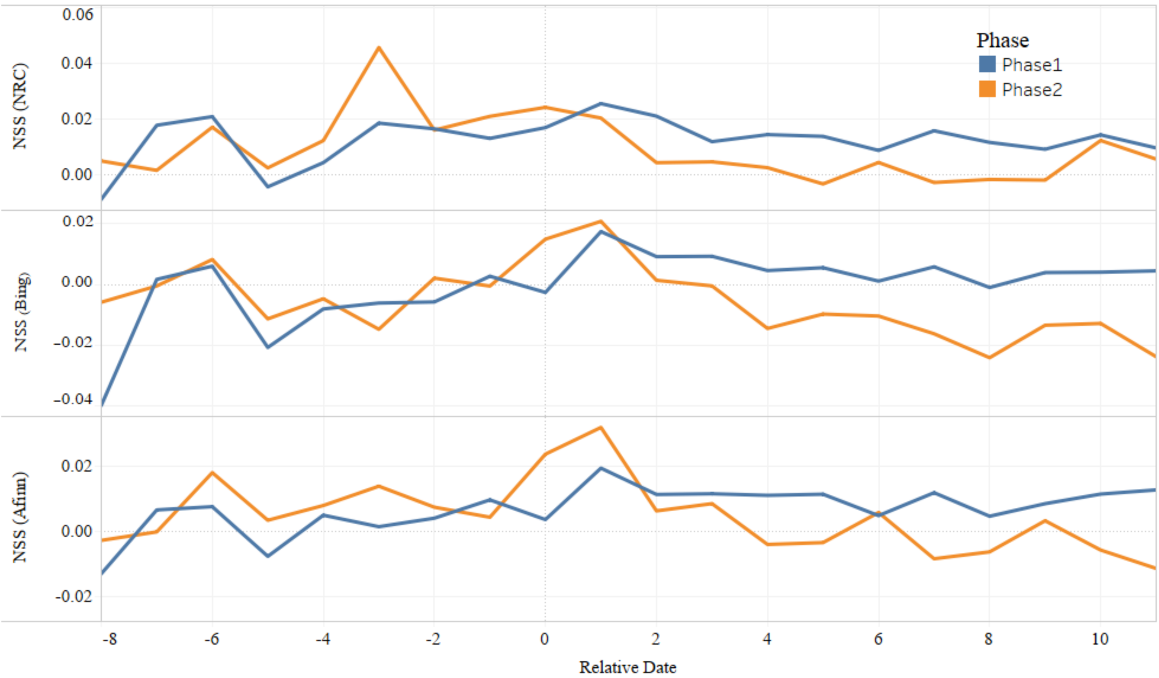



Fig. 2: Daily trend of sentiments obtained from each method

Table 3: Granger causality test results of daily trend comparison of Phases 1 and 2

| Method | $\beta_1$ ($p$-val) | $\beta_2$ ($p$-val) | $F$-value | overall $p$-val |
|---|---|---|---|---|
| NRC | 0.042 (0.62) | -0.363 (0.7) | 0.1714 | 0.43 |
| Bing | 0.146 (0.36) | -0.239 (0.36) | 0.3747 | 0.17 |
| Afinn | 0.727 (0.41) | -0.328 (0.45) | 2.8198 | 0.45 |

To further illustrate the difference in trends during the two phases, wordclouds are plotted to find out the frequently occurring words during the two phases. As shown in Figure 2, the public sentiments changed during the later stages of the Phase 2. So, the periods before the policy implementation are now treated separately and classified as: Pre-Phase 1 ($23^{rd}$ December, 2015 to $31^{st}$ December, 2015), and Pre-Phase 2 ($6^{th}$ April, 2016 to $14^{th}$ April, 2016). Phase 1 and Phase 2 are now defined from $1^{st}$ January, 2016 to $11^{th}$ January, 2016 and $15^{th}$ April, 2016 to $25^{th}$ April, 2016. The wordclouds during the Pre-Phase 1, Phase 1, Pre-Phase 2, and Phase 2 are shown in Figure 3. From Figure 3, it can be seen that the positive comments e.g., 'Iamwithoddeven', 'oddevensuccess' which were frequent during Phase 1 were not present during Phase 2. This suggests the fading out of the positive sentiments among public during the Phase 2 implementation, similar to the trend shown in Figure 2. Such observations are also in-line with the past studies done on such policy implementations. For example, Gallego et al. [4] found that such policy implementations work well in short-term only, and in long-run people cope with such restrictions by shifting to two-wheelers, buying second old cars, etc. Their study was based on the traffic policies enforced in Mexico City, Mexico and Santiago, Chile. Even for the odd-even policy in Delhi, Mohan et al. [10] observed that car-occupancy rates increased during the Phase 2 suggesting that less people opted for car-sharing during Phase 2. Although the above study didn't conduct any statistical tests to find out the differences in car-occupancy rates between the two phases and they concluded that their observation might be due to small sample errors, our study confirms statistically and qualitatively that the public sentiments changed between the two phases.

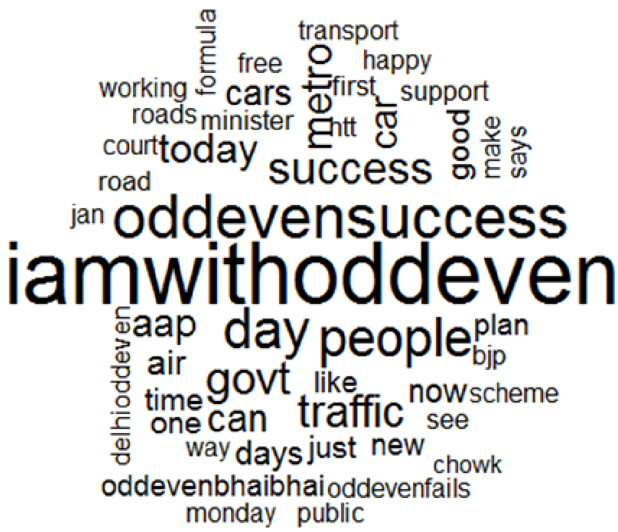
## 5 Conclusions

Social media has gradually evolved to be a popular platform for people to express their views on different topics. These resources can be used for monitoring public sentiments during different trending issues, new product launch, etc. This study used Twitter data to find out public sentiments during the odd-even policy implemented in Delhi, the national capital of India. Four different lexicon-based approaches are used for sentiment analysis purpose. Accuracy of these methods have been determined on hand-annotated test set. And finally, the daily trend of the sentiment during the two phases of the policy implementation are determined. The main objective in this study is to find out the differences in sentiments, if any, among the public regarding the two phases of the policy implementation. To fulfill this objective, causality tests are performed to check the similarity/dissimilarity between the trends obtained during the two phases. The causality results show that there is significant difference between the public sentiments trend between the two phases. The analyses also show that although people were enthusiastic during Phase 1 and the initial period of the Phase 2, however more people started giving negative views on the policy with the progress of the Phase 2. This is in-line with previous studies too which showed that in long-run people become less-enthusiastic towards such policies and learn to cope with them by buying second cars or two-wheelers, thereby reducing car-occupancy rates. Such analyses help the government to find out the growing satisfaction or dissent among the public regarding the policies. In future, this study can be extended by including data from other social media platforms. In this context, it is worth mentioning that Twitter is ranked $7^{\text{th}}$ in terms of social media usage in India. Thus, including data from other social media sources will help to provide more representative samples from the population and add more generalizability to the results. Detailed study can also be done in future to find out the exact reasons for growing public grievances and what measures can be adopted to tackle such issues. Also, more advanced machine learning algorithms (e.g., support vector machines, etc.) can be used in future to obtain better results on the sentiment analysis of the tweets.

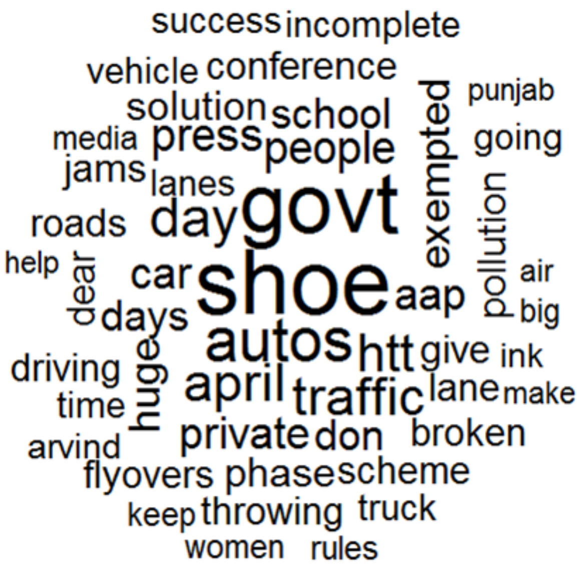
## References

1. Delhi population census data 2011. URL `https://www.census2011.co.in/census/state/delhi.html`. Accessed: 2017-12-30.
2. Angel Hsu and Alisa Zomer. An interactive air-pollution map. URL `https://www.theatlantic.com/health/archive/2014/06/the-air-we-breathe/372411/`. Accessed: 2017-03-12.
3. Lanlan Wang, Jintao Xu, and Ping Qin. Will a driving restriction policy reduce car trips? The case study of Beijing, China. *Transportation Research Part A: Policy and Practice*, 67:279–290, 2014.
4. Francisco Gallego, Juan-Pablo Montero, and Christian Salas. The effect of transport policies on car use: Evidence from Latin American cities. *Journal of Public Economics*, 107:47–62, 2013.
5. Suresh Tiwari, Abin Thomas, Prakasa S. P. Rao, Dilip Chate, Vijay Kumar Soni, Surender Singh, Sachin Ghude, Devendra Singh, and Philip K. Hopke. Pollution concentrations in Delhi India during winter 2015–16: A case study of an odd-even vehicle strategy. *Atmospheric Pollution Research*, 2018. doi: https://doi.org/10.1016/j.apr.2018.04.008.
6. Naveen Garg, Ashok K Sinha, Monika Dahiya, and P Kumar. Effect of odd-even vehicular restrictions on ambient noise levels in Delhi city. In *Advances in Mechanical, Industrial, Automation and Management Systems (AMIAMS), 2017 International Conference on*, pages 252–256. IEEE, 2017.
7. Tahar Zanouda, Sofiane Abbar, Laure Berti-Equille, Kushal Shah, Abdelkader Baggag, Sanjay Chawla, and Jaideep Srivastava. On the role of political affiliation in human perception the case of Delhi odd–even experiment. In *Social Informatics - 9th International Conference, SocInfo 2017, Proceedings*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pages 74–88. Springer- Verlag, 2017. doi: 10.1007/978-3-319-67256-4_8.
8. Asha B Chelani. Study of local and regional influence on $PM_{2.5}$ concentration during odd-even rule in Delhi using causal analysis. *Aerosol and Air Quality Research*, 17(5):1190–1203, 2017.

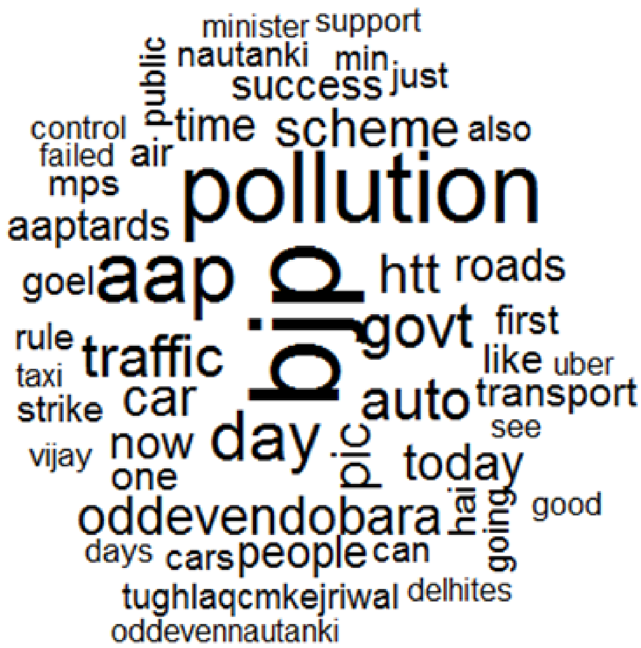

(a)




(b)




(c)




(d)

Fig. 3: Wordcloud of tweets during (a) Pre-Phase 1, (b) Phase 1, (c) Pre-Phase 2, and (d) Phase 2 of odd-even policy

9. Prashant Kumar, Sunil Gulia, Roy M Harrison, and Mukesh Khare. The influence of odd-even car trial on fine and coarse particles in Delhi. *Environmental Pollution*, 225:20–30, 2017.
10. Dinesh Mohan, Geetam Tiwari, Rahul Goel, and Paranjyoti Lahkar. Evaluation of odd–even day traffic restriction experiments in Delhi, India. *Transportation Research Record: Journal of the Transportation Research Board*, 2627:9–16, 2017.
11. Michael Zimmer and Nicholas John Proferes. A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66 (3):250–261, 2014.
12. Andranik Tumasjan, Timm Oliver Sprenger, Philipp G. Sandner, and Isabell M. Welpe. Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Fourth International AAAI Conference on Weblogs and Social Media*, 2010.

13. Pete Burnap, Omer F Rana, Nick Avis, Matthew Williams, William Housley, Adam Edwards, Jeffrey Morgan, and Luke Sloan. Detecting tension in online communities with computational Twitter analysis. *Technological Forecasting and Social Change*, 95:96–108, 2015.
14. Bernard J Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the Association for Information Science and Technology*, 60(11):2169–2188, 2009.
15. Hrishikesh Bakshi. *Framework for crawling and local event detection using twitter data*. Rutgers The State University of New Jersey-New Brunswick, 2011. doi: https://doi.org/doi:10.7282/T3R78DKZ.
16. Axel Bruns, Jean E. Burgess, Kate Crawford, and Frances Shaw. #qldfloods and @QPSMedia: Crisis communication on Twitter in the 2011 South East Queensland floods, 2012. URL `https://eprints.qut.edu.au/48241/`.
17. Avinash Kumar, Miao Jiang, and Yi Fang. Where not to go? Detecting road hazards using Twitter. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 1223–1226. ACM, 2014. doi: http://dx.doi.org/10.1145/2600428.2609550.
18. Vasileios Lampos and Nello Cristianini. Tracking the flu pandemic by monitoring the social web. In *Cognitive Information Processing (CIP), 2010 2nd International Workshop on*, pages 411–416. IEEE, 2010.
19. Craig Collins, Samiul Hasan, and Satish V Ukkusuri. A novel transit rider satisfaction metric: Rider sentiments measured from online social media data. *Journal of Public Transportation*, 16(2):2, 2013.
20. Thuy TB Luong and Douglas Houston. Public opinions of light rail service in Los Angeles, an analysis using Twitter data. *iConference 2015 Proceedings*, 2015.
21. Kenta Sasaki, Shinichi Nagano, Koji Ueno, and Kenta Cho. Feasibility study on detection of transportation information exploiting Twitter as a sensor. In *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.
22. Sudhir Kumar Sharma, Ximi Hoque, and Pravin Chandra. Sentiment predictions using deep belief networks model for odd-even policy in Delhi. *International Journal of Synthetic Emotions (IJSE)*, 7(2):1–22, 2016.
23. GNIP full archive search API. URL `http://support.gnip.com/apis/search_full_archive_api/`. Accessed: 2017-12-30.
24. Twitter streaming API. URL `https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data.html`. Accessed: 2017-12-30.
25. Si Si Mar Win and Than Nwe Aung. Target oriented tweets monitoring system during natural disasters. In *Computer and Information Science (ICIS), 2017 IEEE/ACIS 16th International Conference on*, pages 143–148. IEEE, 2017.
26. Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer, 2012.
27. Snowball. URL `http://snowball.tartarus.org/`. Accessed: 2017-12-30.
28. Mohammad Sadegh, Roliana Ibrahim, and Zulaiha Ali Othman. Opinion mining and sentiment analysis: A survey. *International Journal of Computers & Technology*, 2(3):171–178, 2012.
29. Vinh Ngoc Khuc, Chaitanya Shivade, Rajiv Ramnath, and Jay Ramanathan. Towards building large-scale distributed systems for Twitter sentiment analysis. In *Proceedings of the 27th annual ACM symposium on applied computing*, pages 459–464. ACM, 2012.
30. Michael Speriosu, Nikita Sudan, Sid Upadhyay, and Jason Baldridge. Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the First workshop on Unsupervised Learning in NLP*, pages 53–63. Association for Computational Linguistics, 2011.
31. Finn Årup Nielsen. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*, 2011.
32. Saif M Mohammad and Peter D Turney. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465, 2013.
33. Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.
34. Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.
35. Felipe Bravo-Marquez, Marcelo Mendoza, and Barbara Poblete. Combining strengths, emotions and polarities for boosting Twitter sentiment analysis. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, page 2. ACM, 2013.
36. Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.