Analyzing Air pollution mitigation methodology: Odd Even Campaign

Apoorva Tyagi
student
School of Computer Science and Engineering
Lovely Professional University, Phagwara, Punjab, India

Tejashwa Kumar Tiwari
student
School of Computer Science and Engineering
Lovely Professional University, Phagwara, Punjab, India

Gunseerat Kaur
Asstt. Prof.
School of Computer Science and Engineering
Lovely Professional University, Phagwara, Punjab, India

Abstract: Air pollution has seen an exponential growth in recent decade, major reason being the increase in contributing factors. Daily activities such as movement of vehicles, construction sites, factory waste has increased drastically. The pollutants are not only contributing to the green house effect but also affecting the humans and their surroundings. The air index has been degraded year by year eventually coming to a situation where a sharp increase can be seen in air-borne diseases among masses in major metropolitan cities. In India, the capital of the country has not only suffered from dreaded situation like smog but every year the air quality around degrades during winter months making it difficult for residents to survive. During winter the condensation principle settles this highly polluted surrounding air closer to the surface of earth making it difficult for people to conduct daily day to day activities. As a suggested measure the government in Delhi elaborated an Odd-Even Campaign, the success of this campaign has been studied in this paper through techniques of sentiment analysis.

keywords: sentiment analysis, air pollution, odd-even campaign, twitter analysis, crisis, mitigation.

I. Introduction

Air pollution is one of the leading causes of death in the world. This air pollution, which includes both outdoor and indoor air pollution, causes around 7 million premature deaths per year. Most deaths are due to stroke, chronic obstructive pulmonary disease, acute respiratory infections, heart disease and lung cancer[1]. More than 80% of people in urban areas monitoring air pollution are exposed to air quality levels above the World Health Organization (WHO) level of 10 µg / m3[2], with low and middle income countries being the most affected.

Today, air pollution has become one of the biggest problems for any country. India has already reached an alarming situation. According to the WHO Global Air Pollution Database, 14 of the 15 most polluted cities in India are from India[3]. Prolonged contact with particles can lead to respiratory and cardiovascular diseases such as asthma, bronchitis, lung cancer and myocardial infarction. In 2010, India had about 620,000 deaths from airborne disease[4]. "The WHO-sponsored Central Pollution Control Board (CPCB) describes India as one of the world's worst polluted countries[5]. Delhi is one of the most polluted cities in the world today. The central regulatory authority has recently set stricter standards for various air pollutants and pollutants, but the revised standard for pm 10 has not been established. Air polluting contributions have been on an increase due to surging human activities lately. The increase in diseases puts up economic burden and increases the costing. Air pollution increase is not only affecting the humans but also costing the local rule regulating organizations. The expenditures to eradicate and create awareness is high, whereas the citizens suffering have to pay in high amounts for treatments as well.

II. Related work

Over the years various pollutions related models have been discussed and proposed for providing an instantaneous check over the air-quality index in detail ensuring its users can gain relative information, one such model was proposed by David Hasenfratz[6] and his team who facilitated feeding data into a low cost gas sensor that could monitor the conditions and forward results to a smartphone based application[7]. Handheld devices are popular amongst the new generations which also presents as a viable form of creating awareness among its users. Visibility monitoring, weather conditions, nearby constructions sites, traffic jams can provide with much viable information in order to detect pollution in a particular location[8]. Chen[9] and his team proposes an IoT based monitoring system that perceives, evaluates the parameters and produces current information and predicts future state of the air-index[10]. Abdullah Kadri[3] and his team has worked on deploying an effective ad-hoc monitoring system that adheres to information collected by different nodes based on a geographical region and then regulate a plot to predict conditions and forecasting air-quality in future[2].

III. Problem Statement and Objectives

In this regard, carefully examining the dataset, it should be clear to identify patterns (increase in air pollution) and correlation factors for the major levels of air pollution in India and some key locations in India. We will also look at the implications of a government-sponsored initiative such as the Phase II Odd Even pilot project. The odd-even experiment was created with the initial idea of reducing vehicles on the roads to prevent excessive inflow and out flow of traffic in the city. The factor which greatly influenced the ideology's success and downfall was the compliance of the citizens. Another implication to this ruling could be an active monitoring system which would take into account the nearby activities leading to the campaign being implemented only in some regions[7]. The major concern of the residents was that the campaign although lead to more shared pools in the region for a specific day but eventually the concern was inconvenience caused to some.

The major cause for pollution in the National capital region is still uncertain, in parts of the city it turns out to be vehicular activity, otherwise certain regions have seen massive construction sites, industrial pollution has also been the major contributing factor[11]. With this paper an aim is to provide an insight to the pollution controlling agencies to look forward and deploy more stringent strategies to cope with the control of air pollution. The objectives of the study have been highlighted:

- Examin the air pollution in separate geographical regions of the country
- Identifying the metereological, weather and activity patterns to ensure the monitoring parameters are precise
- To construct a predictive model that can fetch the pre-determined contexts and examine the contents of major pollutants such as PM 2.5, NO_2, etc

IV. Methodology

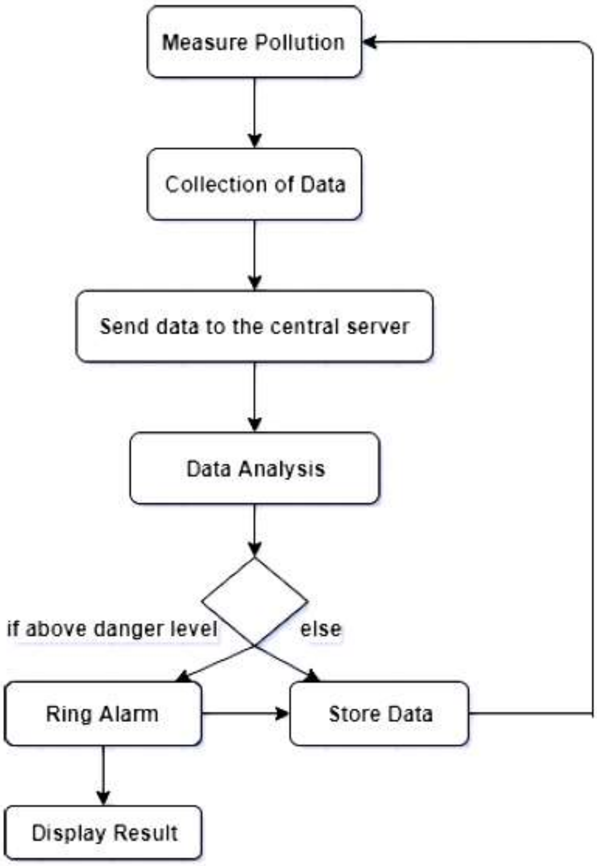


Figure 1 Data flow diagram for the proposed model

a) *Measuring Pollution:* Air quality index is the measure of conducting study on air quality it runs from 0 to 500 and it degrades towards larger values. Table 2 defined the levels of pollutions and their degree of severity[12].

Table 1 Air quality index measures

Air Quality Index (AQI) Values	Levels of Health Concern	Colors
0-50	**Good**	**Green**
51-100	**Moderate**	**Yellow**
101-150	**Unhealthy for Sensitive Groups**	**Orange**
151 to 200	**Unhealthy**	**Red**
201 to 300	**Very Unhealthy**	**Purple**
301 to 500	**Hazardous**	**Maroon**

b) *Collection of Data:* and sending it to the central server: After measuring the air quality index, the process of collecting the data related to air pollution is done so that the data can help in analysis of these effects. After data collection, the data is sent to the central server so that the server can do different analysis on different factors.

c) *Data Analysis:* When data is sent to the central server, data analysis process is started by the server and it tries to analyze every aspects and information it can retrieve from the data using different tools and techniques on different factors and attributes contributing in the process.

d) *Data storage and Results:* After getting the data analyzed, several conditions are checked w1hich are usually set by the data analyst. If the analysis received from pollution data has value more than the condition set, then it triggers an alarm and the data result is shown saying that it is in critical phase and it might harm people's health. And if it does not trigger the alarm, then the data is simply stored in the database and gets back to the first process.

In few odd cases the pollutant level for data is missing anf hence lower variables have been considered in those surmount cases, the missing variables were majorly for PM 2.5, NO2, SO2, CO. In order to reduce inconsistency of such missing values, average values of past days have been filled instead. In other few cases no data was reported on some days and hence that data has been merged as well.

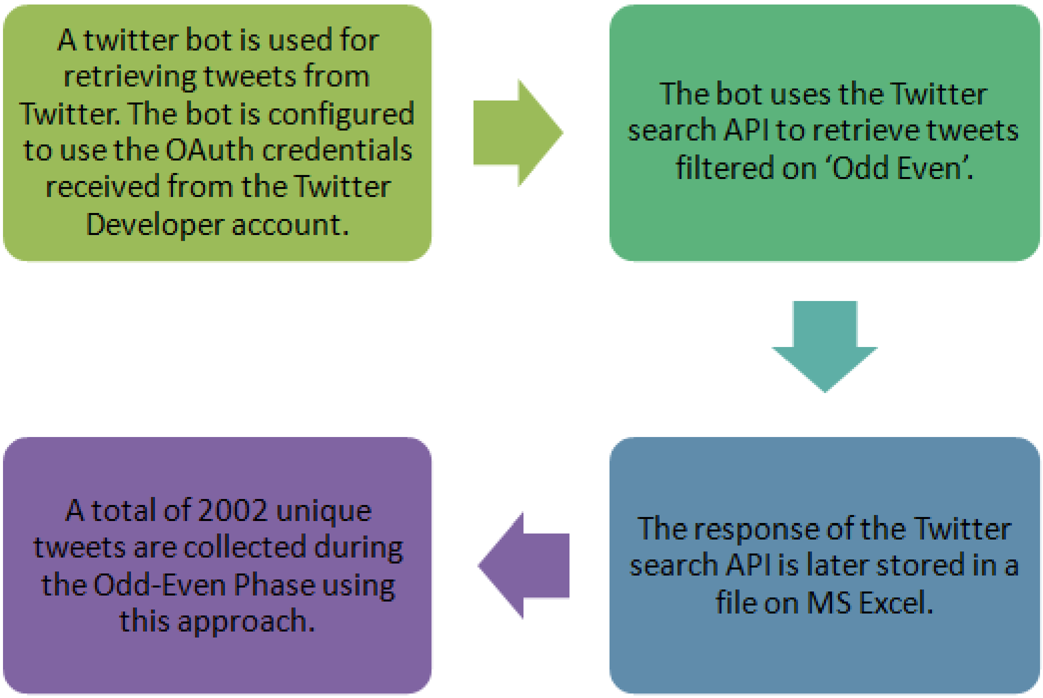
- Variables Transformation
  - Deploying a multi-linear regression model with all variables and their categories.
  - Neural network without data transformations was taken into consideration.
- Missing values and Outliers
  - Any missing values in the dataset were re-featured with average values and not null
  - In case of no data being available for certain days the records were omitted accordingly to not cause any deviations
  - The days which had some considerable values have been graded to present a valid value and not absurd results.
  - Days when Outliers were present, we replaced those values with their average.
- Combining all the data together
  - After cleansing the dataset of any anomalies, Microsoft SQL server was opted for model creation.
- Dimensionality Reduction
  - Removed all the columns and rows which are not needed for the creation of the model.

V. SENTIMENT ANALYSIS OF ODD-EVEN CAMPAIGN

The ideology to come up with this analysis is to form a model which can present a predictive model to analyze how a particular campaign is affecting the people. Delhi was chosen for a trial version of this campaign, predicting the success is although uncertain, sentiment analysis although can be implemented to generate a basic understanding of how effective a particular system is. The Odd-Even Program, a pollution control measure, was launched in the state capital on 4 November. Under the scheme, vehicles should drive alternately on even and odd data according to their registration number between 8:00 and 20:00. The scheme attracts a fine of Rs 4,000. The regulation is not effective on Sundays. Twitter comes out as an emerging platform where opinions are posted enthusiastically by the users. Their views and opinions can be monitored and utilized to generate some efficient results. These can also be effective and opinionated views on variety of issues or proving the effectiveness of the particular system[13]. The main motive is to devise a model for the classification of twwets into three basic most categories, positive, negative and neutral and make use of freely available data resources to cite information and use it accordingly.

- **Text mining of tweets for odd even phase : Data extraction**

The tweets have been retrieved using a feature of hashtags, which are a prevalent method of defining a topic of the tweet. A corpus of tweets on the same topic can be created using hashtags. Twitter offers several Application Programming Interfaces (APIs) that can automatically extract data for each event using the specified parameters. We collected tweets from November 2019 with the hashtags #OddEven #Delhi #oddeven. The resulting record consists of 2002 tweets. The nine samples of the corpus which have retweets count of more than 100 are given in the following Table 1



Firgure 2 Process methodology for data extraction

Table 2 Displaying some of the tweets used for text mining

Sr. No	Text	Retweeted
1	"Odd numbered car, even date—so took out the cycle for commute to work. Enjoyed it more, in the sunny	1028
99	"On d 1st day of #oddeven, PM 10 levels went down from an alarming high at 6 pm As Dy CM @msisodia said in his pressure, while #OddEven is not d only factor it is still one of d factors"	316
704	"Audi should launch a new car for Delhi called for the Audi-1. The only car you can run all days during #OddEven"	178
1002	"Arvind kejriwal ji, we have 2 simple questions for you! #OddEven"	102
1152	"Did #OddEvenphase 2 in #Delhi deliver results?"	239
1498	"Massive #TrafficJams on	156

	#Delhi Roads... Want #OddEven days again.”	
1499	“#OddEven Impact Petrol, diesel sales down 25% in #Delhi since odd-even scheme kicked”	142
1765	“#Delhi ‘s #OddEven car rule to curb smog comes into effect on Friday”	97
1830	“Delhi #OddEven campaign is a boon”	101

VI. Results

The results have been tabulated keeping in view how the classes have been outlined, in the figure 2 given below the graph shows the variation between negative and positive sentiments varying from values -5 being highly negative and +3 being highly positive. 0 stands as a neutral sentiment. Combining the factorizations the positive holds more ground than negative but majority of the tweets were neutral with the keywords. At a particular time frame the majority remained for the neutral responses only throughout.

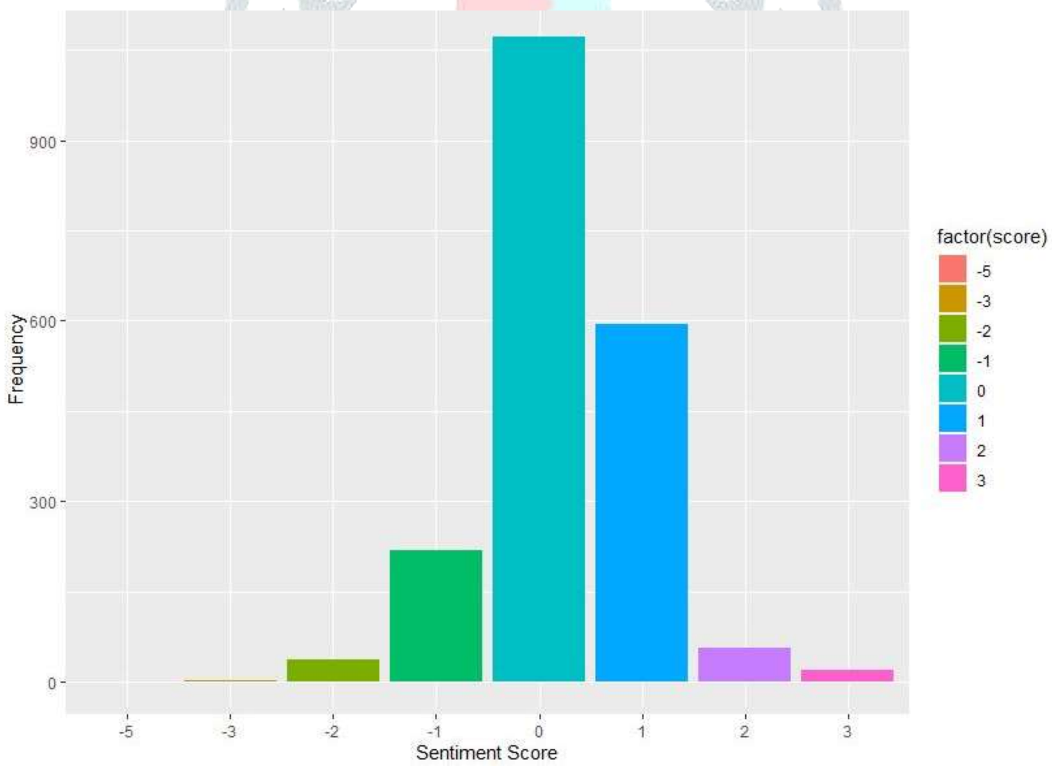


Figure 2 customer sentiment scores

Framing a better response further figure 3 shows clearly how the polarity of the tweets went from being positive to negative.

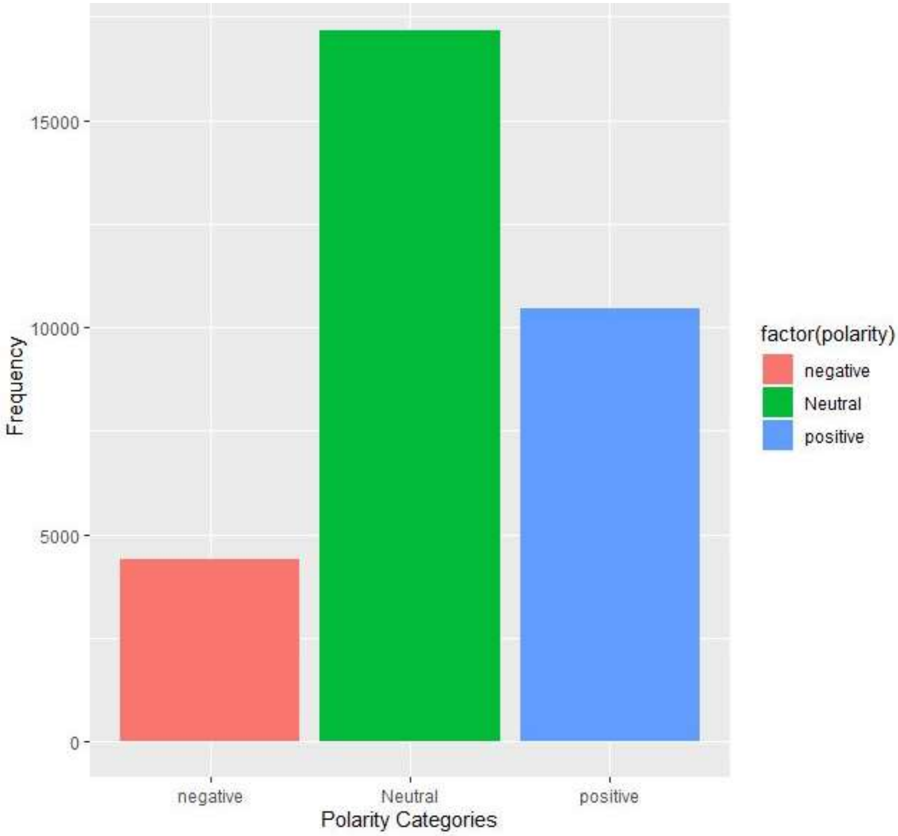


Figure 3 Public reaction score- overall

VII. Conclusion

From the Sentiment analysis of the tweets collected for 'Odd-Even' theme, it's clear that Twitterati for the foremost part holds neutral feedback towards this rule. Twitterati largely holds positive sentiment regarding Odd Even scheme with a mixture of anticipation and negative results. Close levels between anticipation and negative sentiments reflects that majority of people are unsure whether or not it'll work at this point. Similarly, some people don't seem to be happy with the previous Odd-Even campaign that occurred in 2017.PM 2.5 and PM 10 showed a gradual decrease within the level of pollutants considering the previous day of implementation i.e. Nov 3, 2019 and also the actual day when campaign began, i.e. Nov 4, 2019. There is a robust chance that any gains from Odd-Even policy in terms of air quality levels were entirely eclipsed by "other sources of pollution".

References

[1] X. Du, O. Emebo, A. Varde, N. Tandon, S. N. Chowdhury, and G. Weikum, "Air quality assessment from social media and structured data: Pollutants and health impacts in urban planning," *2016 IEEE 32nd Int. Conf. Data Eng. Work. ICDEW 2016*, pp. 54–59, 2016.

[2] K. B. Shaban, A. Kadri, and E. Rezk, "Urban air pollution monitoring system with forecasting models," *IEEE Sens. J.*, vol. 16, no. 8, pp. 2598–2606, 2016.

[3] A. Kadri, E. Yaacoub, M. Mushtaha, and A. Abu-Dayya, "Wireless sensor network for real-time air pollution monitoring," *2013 1st Int. Conf. Commun. Signal Process. Their Appl. ICCSPA 2013*, pp. 1–5, 2013.

[4] E. G. Snyder *et al.*, "The changing paradigm of air pollution monitoring," *Environ. Sci. Technol.*, vol. 47, no. 20, pp. 11369–11377, 2013.

[5] W. Jiang, Y. Wang, M. H. Tsou, and X. Fu, "Using social media to detect outdoor air pollution and monitor air quality index (AQI): A geo-targeted spatiotemporal analysis framework with sina weibo (Chinese twitter)," *PLoS One*, vol. 10, no. 10, Oct. 2015.

[6] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, "IPSN'12 - Proceedings of the 11th International Conference on Information Processing in Sensor Networks," *IPSN'12 - Proc. 11th Int. Conf. Inf. Process. Sens. Networks*, pp. 1–5, 2012.

[7] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent Twitter sentiment classification," *Proc. 49th Annu. Meet. Assoc. Comput. Linguist.*, vol. 49, no. 12, pp. 151–160, 2011.

[8] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surv. Tutorials*, vol. 16, no. 4, pp. 1801–1819, 2014.

[9] C. Xiaojun, L. Xianpeng, and X. Peng, "IOT-based air pollution monitoring and forecasting system," *2015 Int. Conf. Comput. Comput. Sci. ICCCS 2015*, pp. 257–260, 2015.

[10] L. Zhou, Y. Zhu, and K. K. R. Choo, "Efficiently and securely harnessing cloud to solve linear regression and other matrix operations," *Futur. Gener. Comput. Syst.*, vol. 81, pp. 404–413, Apr. 2018.

[11] Z. Xiaomei, Y. Jing, Z. Jianpei, and H. Hongyu, "Microblog sentiment analysis with weak dependency connections," *Knowledge-Based Syst.*, vol. 142, pp. 170–180, Feb. 2018.

[12] T. V. Tran, N. T. Dang, and W. Y. Chung, "Battery-free smart-sensor system for real-time indoor air quality monitoring," *Sensors Actuators, B Chem.*, vol. 248, pp. 930–939, 2017.

[13] A. F. Y. Chao and H. L. Yang, "Using Chinese radical parts for sentiment analysis and domain-dependent seed set extraction," *Comput. Speech Lang.*, vol. 47, pp. 194–213, Jan. 2018.