International Journal for Modern Trends in Science and Technology, 6(12): 373-378, 2020

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST061270
Available online at: http://www.ijmtst.com/vol6issue12.html

Impact of COVID-19 Pandemic on the Air Quality Levels of Delhi

Jaash Sehgal[1] | K.C. Tripathi[2] | M.L. Sharma[3]

[1] Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India,
[2] Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India,
[3] Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India,

To Cite this Article
Jaash Sehgal, K.C. Tripathi and M.L. Sharma, "Impact of COVID-19 Pandemic on the Air Quality Levels of Delhi", *International Journal for Modern Trends in Science and Technology*, 6(12): 373-378, 2020.

Article Info
Received on 16-November-2020, Revised on 09-December-2020, Accepted on 12-December-2020, Published on 17-December-2020.

ABSTRACT

Air Pollution [2] has been an issue of major concern in many cities of India like Delhi, Lucknow, Kolkata etc. But due to the Covid-19 pandemic lockdown throughout the nation, there was seen a significant change in the air quality of these cities, as there was little to no movement of people in the cities. In this paper, machine learning [3] techniques are used to predict the air quality levels in Delhi. Also, those levels which would have been obtained if the pandemic induced lockdown had not occurred are extrapolated [4]. Then, the two results are compared with each other in order to analyze the impact of Covid-19 on air quality levels.

KEYWORDS:*Air Pollution, Covid-19, Lockdown, Delhi, PM2.5 [5], PM10 [6].*

I. INTRODUCTION

In the developing countries like India, the rapid increase in population and economic upswing in cities have led to environmental problems such as air pollution, water pollution, noise pollution and many more. Air pollution is a mix of particles and gases that can reach harmful concentrations both outside and indoors. Its effects can range from higher disease risks to rising temperatures. Soot, smoke, pollen, methane, and carbon dioxide are a just few examples of common pollutants. Air pollution has direct impact on human's health. There has been increased public awareness about the same in our country. Global warming, acid rains, increase in the number of asthma patients are some of the long-term consequences of air pollution.In the capital of India i.e. Delhi air pollution has been an issue of major concern for years now. But when the lockdown was imposed due to the Covid-19 pandemic, a shift in trend was seen in the air pollution levels. There was decrease in the air pollution and the quality of air has been improved.

In this paper, the data was gathered from Kaggle [7] and it was taken day-wise. By using this data, we built a prediction algorithm which will predict the air quality index on a day. Also, data was extrapolated from the existing data in order to get the data of the air quality which would have occurred if the pandemic hadn't hit. Further, the original data and the results obtained by providing the extrapolated data to the algorithm were compared.

The results of the two main pollutants i.e. PM2.5 and PM10 were generated. The root mean square error was calculated for each of the pollutants and also after the application of Principal Component Analysis [8] for both. The results were satisfactory and proved that Covid-19 has had a positive impact on Air Pollution in Delhi.

The remainder of this research paper is organized in the following way:

In Section 2, we highlight the work done in some related fields. Section 3 provides the all the information about the data. In Section 4, we describe methodology which was actually implemented in building this project. Section 5 talks about the results and the analysis of the project. In Section 6, the final conclusion and future scope of the work is given.

II. RELATED WORK

Air Pollution forecasting has been a major field of research in recent years. As in metropolitan cities of India like Delhi, Kolkata, Chennai etc. have been affected heavily by the deteriorating air quality. Air pollution forecasting in mainly done using machine learning and deep learning [9] models and algorithms. Deep Learning is a class of machine learning algorithms that uses multiple layers to extract higher-level features from the raw input progressively. A popular deep learning architecture is RNN [10] which is used to model sequential data. It contains cyclic connections where the outputs from previous time steps are fed as input to the current time step. In RNNs, errors are backpropagated, and weights are updated using a technique called Back Propagation Through Time (BPTT). RNN can model sequence of data so that each sample can be assumed to be dependent on previous ones. Recurrent neural networks are even used with convolutional layers to extend the effective pixel neighborhood. Another popular deep learning architecture is LSTM. Long Short-Term Memory [11] (LSTM) networks are a modified version of recurrent neural networks, which makes it easier to remember past data in memory. The vanishing gradient problem of RNN is resolved here. LSTM is well-suited to classify, process and predict time series given time lags of unknown duration. It trains the model by using back-propagation.

Covid-19 is another field in which entire world is involved right now. Be it effect of pandemic on people or its effect on stock prices of companies, on crops etc., everyone in the world is focused on the research of impact of Covid-19 on the world.

III. DATA USED

Data Acquisition

Delhi covers an area of 1484 km2 out of which 783 km2 is under the rural area, and 700 km2 is under the urban area. It is bordered by Haryana state on three sides and by Uttar Pradesh to the east. The current population of Delhi is around 19 million. According to the United Nations' World Urbanization Prospects, Delhi would become the most populous city in the world by 2028.

The data of the air pollutants was gathered from Kaggle. The data provided was of 26 cities and each city had data of its various stations. The data was provided by day and by hour. I took the data of the stations of Delhi, which were 37. These stations had a lot of varying data which ranged from starting of 2015 to July, 2020. I took up those stations having consistent data i.e. from January 2018 to July 2020. This reduced the stations down to 10 stations.

The dataset has a total of 16 columns i.e. Station ID, Date, PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, Xylene, AQI, AQIBucket. Table 1 shows the list of pollutants in the data along with their unit of measurement.

Table 1. List of Pollutants

S. No.	Parameters	Unit
1.	PM2.5	ug / m3
2.	PM10	ug / m3
3.	NO	ug / m3
4.	NO2	ug / m3
5.	NOx	ug / m3
6.	NH3	ug / m3
7.	CO	ug / m3
8.	SO2	ug / m3
9.	O3	ug / m3
10.	Benzene	ug / m3
11.	Toluene	ug / m3
12.	Xylene	ug / m3

Pre-processing

The collected data contained several missing values and extreme values that deviate from other observations on data which may indicate variability in measurement, experimental errors or a novelty. This was handled by setting the outliers to null values. The missing data present in the dataset acts as noise which affects the performance forecasting model. So, the missing data was populated using the technique called interpolation. It is a method of constructing new data points within a range of a discrete set of known data

points. It can be linear, bilinear, piecewise, polynomial, spline, cubic, bicubic, etc. The linear interpolation technique was used to fill the missing values. So, the final dataset considered has 35088 samples for each pollutant and the weather parameters. The descriptive statistics of the data are shown in Table 2.

	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene
mean	110.23	218.0	34.06	45.4	55.16	39.35	1.57	14.3	43.90	3.59	24.83
std	88.12	137.4	43.52	32.5	55.10	24.14	2.40	9.40	53.10	4.83	34.95
min	0.48	0.08	0.01	0.02	0.00	0.07	0.00	0.04	0.02	0.02	0.00
25%	48.88	112.6	7.22	23.2	19.69	24.89	0.71	7.68	20.65	0.65	2.62
50%	83.12	188.7	17.27	37.3	37.71	34.50	1.12	12.2	32.54	2.18	12.88
75%	143.8	291.4	41.07	59.7	70.17	47.97	1.75	18.6	50.18	4.90	32.57
max	1000	976	436.8	397	453.6	418.9	48.3	130	963	185.4	396.8

Table 2.Summary of Statistics of Pollutants

IV. Proposed Methodology

Principal Component Analysis

The data was first standardized using Standard Scalar in the Scikit Learn library in python. After this the process of Principal Component Analysis began manually. First the covariance matrix was obtained, from which eigen values and eigen vectors were obtained. The eigen values were used in order to reduce the features from 11 to 6. This helped in making the entire process faster and simpler going ahead.

Prediction of Data

We took the data of station DL003 from January 1st, 2018 to July 7th, 2020 and split it into two parts, first was before lockdown which was from January 1st, 2018 to March 22nd, 2020 and second was during lockdown which was from March 23rd, 2020 to July 7th, 2020. Then a model was created using Scikit Learn library in python. Various models used experimented with, but we landed with Random Forest Classifier. The before Lockdown data was used as the training data and the model was trained using this data itself. The during lockdown data was taken as testing data and predictions were made using this data. Then, the results of these predictions were compared with the testing data and accuracy was obtained as 73.73%.

Extrapolation of Data

This is the main component of the experiment. For extrapolation we used the SciPy library in python. The data in this step represents that data which would have occurred if the lockdown had not happened. This represents the hypothetical situation if the pandemic hadn't hit at all. The way this was done was simple. A one step ahead time series was plotted for the features and using this time series a curve was fitted on it. Now using this fitted curve data was extrapolated for the lockdown period. This process was done for the first two features i.e. PM2.5 and PM10 and also for the first two principal components.

Comparison of results

The results of the extrapolated data were compared to the original data. This was done using various statistical methods like Root Mean Square Error (RMSE). This was done for PM2.5 and PM10 and also for the first two principal components.

Applying on rest of the stations

Now, this entire above-mentioned process was applied on other 9 stations and the results were calculated.

V. Results and Analysis

The results from the comparison of were as expected. The values of PM2.5 and PM10, the two

main pollutants of my study, were worse(more) than the actual values of the two pollutants. The results are shown in the figures shown.

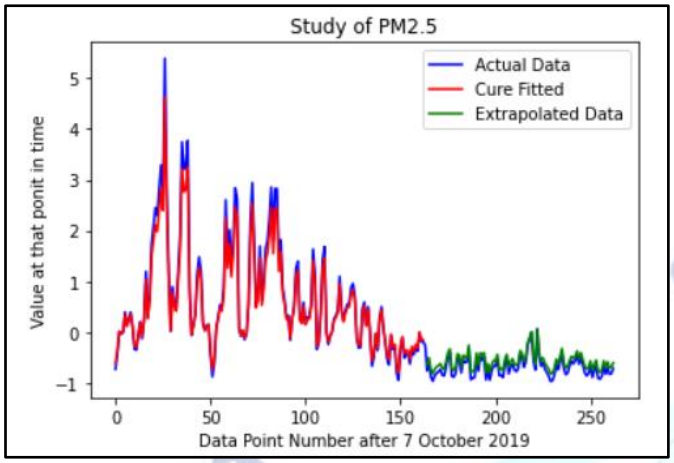

Fig 1. Time Series of PM2.5 after 7th October 2019

In figure 1, the data is shown after 7th October 2019. The blue graph shows the actual data of PM2.5 which was seen in real time, the red graph gives curve which is fitted on the actual data of PM2.5 till 22nd March 2020. The green graph is the extrapolated data of PM2.5 which is after 22nd March 2020. The green graph shows the data of PM2.5 which would have occurred if the lockdown had not taken place. The root mean square error (after 22nd March 2020) for figure 1. was calculated and was obtained to be 0.11.

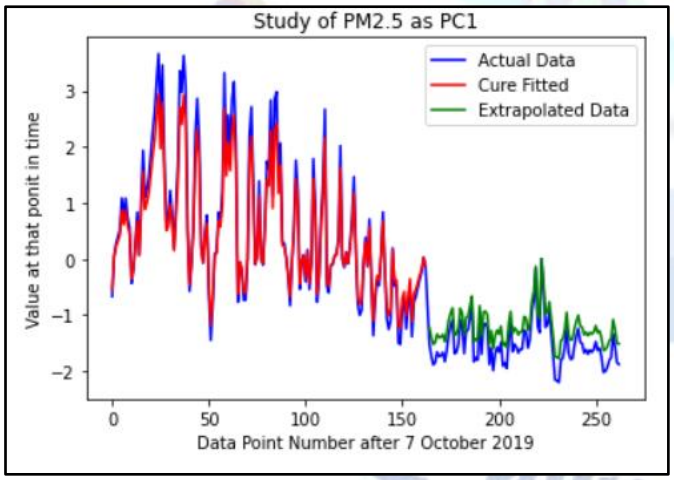

Fig 2. Time Series of PM2.5 after application of PCA

In figure 2, the data is shown after 7th October 2019. All of the data is of PM2.5 after the application of Principal Component Analysis (PCA). PM2.5 becomes the first principal component. The blue graph shows the actual data which was seen in real time, the red graph gives curve which is fitted on the actual data till 22nd March 2020. The green graph is the extrapolated data which is after 22nd March 2020. The green graph shows the data which would have occurred if the lockdown had not taken place. The root mean square error (after 22nd March 2020) for figure 2. was calculated and was obtained to be 0.30.

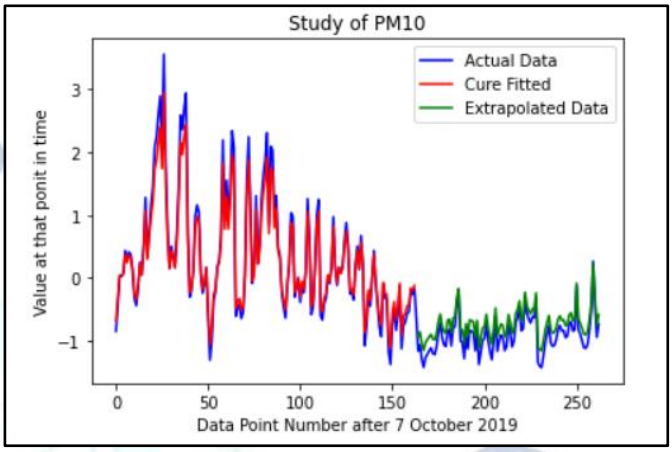

Fig 3. Time Series of PM10 after 7th October 2019

In figure 3, the data is shown after 7th October 2019. The blue graph shows the actual data of PM10 which was seen in real time, the red graph gives curve which is fitted on the actual data of PM10 till 22nd March 2020. The green graph is the extrapolated data of PM10 which is after 22nd March 2020. The green graph shows the data of PM10 which would have occurred if the lockdown had not taken place. The root mean square error (after 22nd March 2020) for figure 3. was calculated and was obtained to be 0.18.

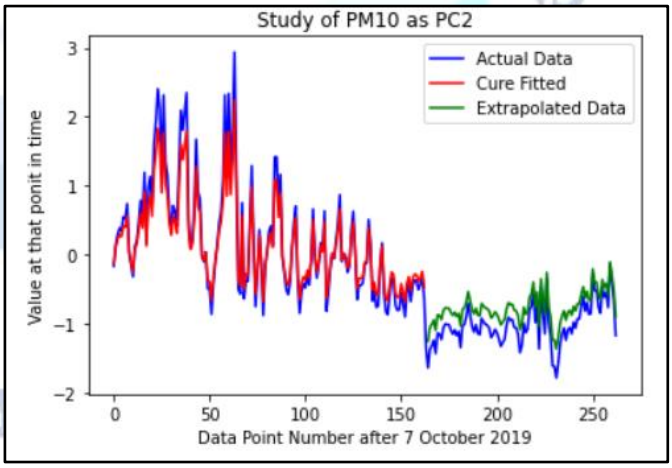

Fig 4. Time Series of PM10 after application of PCA

In figure 4, the data is shown after 7th October 2019. All of the data is of PM10 after the application of Principal Component Analysis (PCA). PM10 becomes the second principal component. The blue graph shows the actual data which was

seen in real time, the red graph gives curve which is fitted on the actual data till 22nd March 2020. The green graph is the extrapolated data which is after 22nd March 2020. The green graph shows the data which would have occurred if the lockdown had not taken place. The root mean square error (after 22nd March 2020) for figure 4. was calculated and was obtained to be 0.24.

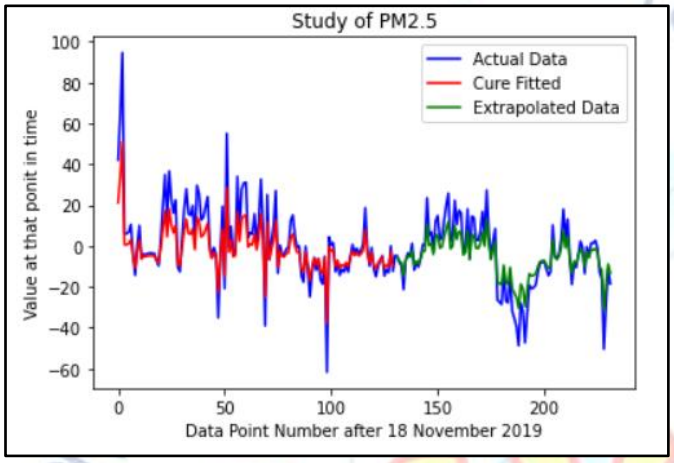

Fig 5. Time Series of PM2.5 of all 25 stations

In figure 5, the data is shown after 18th November 2019. All of the data is of PM2.5 of all the 25 stations after the application of Principal Component Analysis (PCA). The blue graph shows the actual data which was seen in real time, the red graph gives curve which is fitted on the actual data till 22nd March 2020. The green graph is the extrapolated data which is after 22nd March 2020. The green graph shows the data which would have occurred if the lockdown had not taken place. The root mean square error (after 22nd March 2020) for figure 5 was calculated and was obtained to be 7.08.

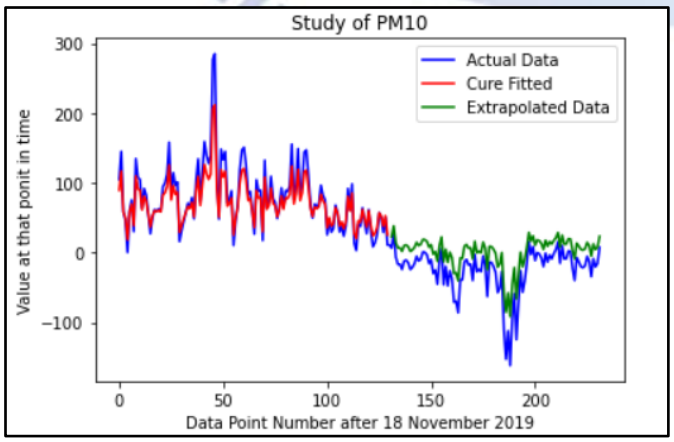

Fig 6. Time Series of PM10 of all 25 stations

In figure 6, the data is shown after 18th November 2019. All of the data is of PM10 of all the 25 stations after the application of Principal Component Analysis (PCA). The blue graph shows the actual data which was seen in real time, the red graph gives curve which is fitted on the actual data till 22nd March 2020. The green graph is the extrapolated data which is after 22nd March 2020. The green graph shows the data which would have occurred if the lockdown had not taken place. The root mean square error (after 22nd March 2020) for figure 6 was calculated and was obtained to be 24.40.

VI. CONCLUSION AND FUTURE SCOPE

Since, in many metropolitan cities of the world air pollution is a major area of concern, a significant drop in air pollution has benefitted the environment all over the world. Due to the pandemic in lockdown, there was seen a significant drop in pollutants. As, it can be concluded from the results that due to the covid-19 induced lockdown that there was a drop in the levels of major air pollutants such as PM2.5 and PM10, which in turn increased the quality of air in Delhi by reducing the air pollution. This shows that Covid-19 had a silver lining in terms of the improving the air quality, therefore the environment in Delhi.

This research paper deals only with the impact on the air quality of Delhi. This research can be extended for other metropolitan cities of India as well as the world. As most of the metro cities in the world are densely populated and high pollution causing activities take place there, so pollution levels in these regions are quite high. This study can be extended to cities like Kolkata, Chennai, New York, Beijing etc. This will provide us with a brief idea on how Covid-19 has affected these regions also.

REFERENCES

[1] World Health Organization, Emergencies, Diseases, Novel Coronavirus 2019, [online] Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[2] P. Gupta, R. Kumar, S. P. Singh and A. Jangid, "A study on monitoring of air quality and modelling of pollution control," 2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Agra, 2016, pp. 1-4, doi: 10.1109/R10-HTC.2016.7906800.

[3] EthemAlpaydin. 2010. Introduction to Machine Learning (2nd. ed.). The MIT Press.

[4] Norbert Wiener. 1964. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. The MIT Press.

[5] Cao, Q., Rui, G. & Liang, Y. Study on PM2.5 pollution and the mortality due to lung cancer in China based on

geographic weighted regression model. BMC Public Health 18, 925 (2018). https://doi.org/10.1186/s12889-018-5844-4
[6] Filonchyk, Mikalai& Yan, Haowen& Yang, Shuwen&Hurynovich, Volha. (2016). A study of PM2.5 and PM10 concentrations in the atmosphere of large cities in Gansu Province, China, in summer period. Journal of Earth System Science. 125. 10.1007/s12040-016-0722-x.
[7] Kaggle, [online] Available: https://www.kaggle.com/
[8] Jolliffe I. (2011) Principal Component Analysis. In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_455
[9] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning.
[10] L. C. Jain and L. R. Medsker. 1999. Recurrent Neural Networks: Design and Applications (1st. ed.). CRC Press, Inc., USA.
[11] Hochreiter, Sepp &Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.