Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 2):
pp. 3128-3140
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.02.330

Exploratory Evaluation Of Air Pollution In Dehradun A Machine Learning Approach

Neelam Singh[1], Harneet Kaur[2], Smitha Kothari Badola[3]

[1]Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, neelamjain.jain@gmail.com

[2]Department of Computer Application, Graphic Era Deemed to be University, Dehradun, kaurharneet122@gmail.com

[3]Research Scholar and Assistant Professor, School of Management, Graphic Era Hill University.

Abstract

Air Pollution is the presence of any gaseous substance, dust particles or liquid vapors which are in atmospheric concentration that lead to harmful effects to the environment. This pollution leads to many harmful diseases having acute and chronic effects on human beings and it can be seen that automobile emissions contribute to more than 60%of air pollution. Also, burning of soil fuel indoors is also one of the factor.

This paper estimates the damage and effects of air pollution on living beings. Air pollution in interurban areas is a major and serious problem. Regular increase in air pollution is a major concern to the health system of human beings as well as plants and animals. Emissions from vehicles are of prime concern since these are base level sources and accordingly maximum affect everyone. This paper has made an attempt to find out the influence of air pollution on Human Health through machine learning approaches. This study focuses on AQI (Air Quality Index) to evaluate the possible implications on human health with the help of linear regression model.

Keywords: Air Pollution, Machine Learning, Data mining, AQI, Linear Regression.

1. Introduction

Natural physiognomies of the atmosphere can be transformed by any physical, chemical or biological agent leading to contamination of indoor and outdoor environment and is termed as 'Air Pollution' as stated by Wark et al (1998) and

Steinfeld (1998). These agents can be generated through various human activities including industrial facilities, household ignition devices, and vehicles or could be forest fire as well. Pollutants like Carbon Monooxide (CO), particulate material (like PM2.5), Sulphur Dioxide (SO2) and Nitrogen Dioxide (NO2), impose majorthreat on human health. According to Choubin et al.(2020) major respiratory diseases like asthma,various allergies aggravates due to increase in pollution affecting human life. How does air pollution affect our health?

- Respiratory Disease
- Cardiovascular Disease
- Cancer

Who gets affected the most due to Air Pollution?

Air pollution affects everyone's wellbeing, but certain groups can be adversely affected by this which includes people in old age, children and people with ailments or allergies. Nearly 9 out of 10 person residing in metropolitan areas are affected by air pollution.

Environmental pollution is one of the major problematic concern in all over the world including developed, underdeveloped and developing countries .The effect of air pollution is very much severe on human including as well on plants and animals and on our surrounding as discussed by Soni et al. (2019). Air pollution causes harmful diseases and cause various environmental problem at regional and global. Air pollution causes acute and chronic effects on human being. It is valued that 60 % of air pollution in India mainly caused by automobile emissions.

Pollutants such as SOx, NOx, SPM, and RSPM, as well as inorganic, organic, and metal pollution, are mostly generated and monitored by automotive emissions. According to the World Health Organization's 2014 Global Health Report on Air Pollution, India has risen to become the world's ninth largest industrial nation while also being the worlds majorly polluted as mentioned by Pooja et al. (2014). The district of Dehradun's ambient air quality is rapidly decreasing. In the Dehradun district, the most significant sources of air pollution is through vehicles. According to a study conducted by , the number of registered vehicles increased by 11% over the last 12 years, with cars increasing by 14% and two-wheelers increasing by 75%. The current transportation system, it was estimated, was primarily operated by private operators using temp and buses.Singh et al (2013) conducted a study which shows more than 1300 tempos were operating in Dehradun's various routes, this number has increased to more than triple along with local buses and private taxis. Vehicle emissions are difficult to control because they are emitted directly into the lower troposphere, and because Dehradun is a valley, dilution and dispersion of pollution takes a long time, especially during the winter season, causing a variety of health effects such as headaches, nausea, eye irritation, bronchial problems, and visibility.

2. Physiography of Dehradun

The Himalayas are to the North, in South are the Shivalik range, to the East is Ganga, and the Yamuna to the west of the Doon Valley. The city of Dehradun is enclosed by the Song River in the east, by the Tons River in the west, by the Himalayan peaks on the north, and by the Sal woods on the south. Dehradun is situated in the middle of 2 rivers i.e. Ganga and Yamuna which are the main rivers of India, both of which have a splendid location. Dehradun is surrounded onall sides by lush forest, and a series of streams and canals cut through the city in a north-south orientation. The city is set in an interesting topographical setting, with great hills to the east and north and Shivalik range to the south.

3. LITERATURE REVIEW

According to Venegas et al.(2014) Air Pollution prediction is done with the help of two common methodologies of deterministic and stochastic methods. To model and monitor Air Pollution, Diffusion model, a deterministic method is widely considered by Ranzato et al. (2012). In contrast to deterministic methods, Air Pollution prediction using statistical approaches is widely acceptable as factors like temperature, rainfall, air-pressure, wind and humidity plays a major role in disseminating pollutants as discussed by Wang et al. (2015). Cai et al. (2009) implemented neural networks to predict Air Pollution along with some associated parameters like hours and day of week, traffic, past 3 years air pollutant concentration, wind-speed and direction. In Rey Station at Tehran using Support vector machine (SVM) and partial least square (PLS) method prediction of CO concentration was done by Delavar et al. (2019).

3.1 Air Quality Index (AQI):

AQI is a mechanism that alerts the public when pollution levels are unsafe. Air Quality Index (AQI) assess ozone (smog) and particle pollution (small particles from power plants, ash, and factories, vehicle emission and exhaust, dust, pollen, and other pollutants), as studied by Zeinalnezhad et al. (2020). AQI levels are conveyed through newspapers, radio, television, and internet on a regular interval. Current air quality data can be checked to assist in taking steps to protect oneself, children, and others from harmful levels of pollution.

This paper estimates the changes in AQI for Dehradun region based on the implementation of Linear Regression techniques.

3.2 Regression Analysis:

Regression analysis is a statistical method which focuses on modelling the relationship amid dependent (target) and independent (predictor, one or more)

variables as employed by Zeinalnezhad et al. (2020). This relationship defines how the value of a dependent variable is fluctuating with respect to an independent variable when other independent variables are said tobe fixed. It is mainly used for forecasting, prognosis, time series modelling, and determining the causal-conclude relationship between variables. Different types of Regression models exist like

Linear Regression, Support Vector Regression, Polynomial Regression, Logistic Regression,Decision tree and random forest regression. In this study we used the Linear Regression method to show the Air Quality Index Graph using datasets of the year 2018, 2019 and 2020.

4. Methodology

The study was done in three phases. In the first phase, data is collected from Kaggle for various regions of Dehradun for the year 2018, 2019 and 2020 as provided by S. Singh (2020). The data collected is the monitoring data of ambient quality of air of Dehradun. Second phase involves the analysis of data, which is done through linear regression model. In the last phase the result are presented to show how the AQI has changed over these three years. Figure 1 shows the Air Quality graph of Dehradun for the year 2021, December.

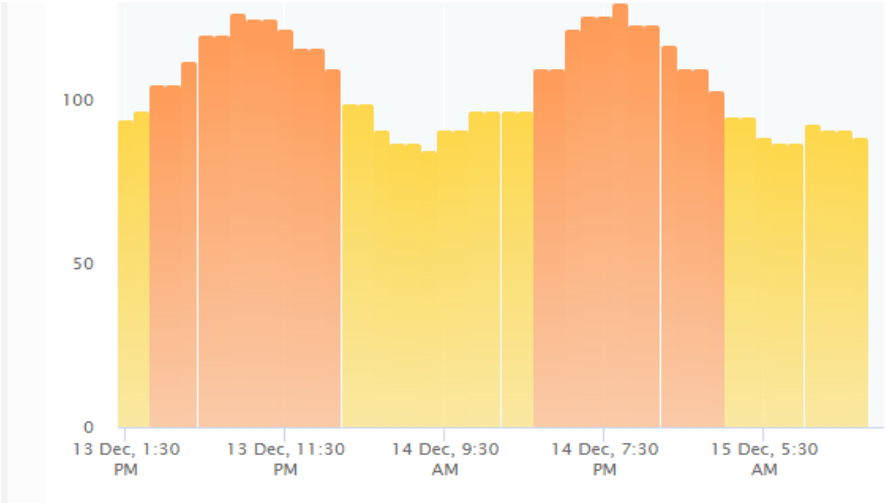

Figure 1. Dehradun Air Quality Graph Dec, 2021[17]

A variety of air pollutants can be generated through various sources, affecting the biotic environment and changing the composition of the air. The ability of the atmosphere to absorb or disperse air pollutants is dependent not only on the quantities that are blown out of the sources of air pollution but also on how the atmosphere is absorbing and dispersing them. Because of variations in climatic and topographical conditions, pollution concentrations vary spatially and temporally, causing the air pollution pattern to shift with different destinations. Vehicles, industrial, residential, and natural sources all contribute to air pollution. The

existence of air pollutants in the ambient air has an adverse influence on the inhabitant's health. Automobile emissions, industrial emissions, and the burning of fossil fuels are all contributing to severe air pollution in Uttarakhand's Dehradun region. These emissions degrade the quality of the ambient air, resulting in a range of health issues. An assessment of air pollution has been undertaken in this research study in order to determine the ambient air quality of Dehradun. SPM, RSPM, SOx, and NOx, as well as metrological limits such as temperature, humidity, and rainfall, were all regulated. Pollution levels were high in numerous sampling locations, primarily due to vehicle pollution in Dehradun and rising urbanization and industry over time.

4.1 Implementation:

Implementation is performed on the collection of data set from the sources of Air Quality Index of Uttarakhand state of different districts of year 2018, 2019 and 2020. Linear Regression model is used to check the AQI changes during these years.

4.1.1 Linear Regression model

It is a supervised learning model which predicts a dependent variable value (d) based on independent variable (i). The hypothesis function for Linear Regression can be defined as:

$$d = \theta_1 + i.\theta_2 \quad (1)$$

Where

i : input training data, d : labels to data

θ1 : intercept

θ2 : coefficient of x

Best fit line is generated once the best value of θ1 and θ2 is generated.

Linear Regression is applied on monitoring data of ambient quality of air as shown in Table1, 2and 3 respectively.

Table 1. Year 2018 Monitoring Data Of Ambient Quality Of Air (S. Singh(2020))

State / UT	City / town / village	Location	No. of monitoring days	AQI		
				Minimum (24-	Maximum (24-hourly	Annual

				hourly average)	average)	Average
Uttarakhand	Dehradun	Raipur Road, Near parag Diary	10	21	27	23
		Clock Tower, PWD Guest House	15	23	26	25
		Himalaya Drug Co. Near ISBT	11	25	27	26
	Haldwani	Govt. Women Hospital	101	7	32	11
	Haridwar	SIDCUL, Haridwar	6	13	22	19
	Kashipur	BSNL Office, Kashipur	48	10	24	14
	Rishikesh	Nagar PalikaParishad	15	16	23	21
	Rudrapur	SIDCUL Office	57	10	24	13

Equation 2, 3 and 4 are used to plot the regression line for the year 2018 data from table 1, in figure 2, for the year 2019 data from table 2, in figure 3 and for the year 2020 data from table 3,in figure 4 respectively.

$$(intercept) = \frac{\sum y \sum x2 - \sum x \sum xy}{(\sum x2) - (\sum x)2} \quad (2)$$

$$y = a + bx \quad (3)$$

$$b(\text{slope}) = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x2 - (\sum x)2} \quad (4)$$

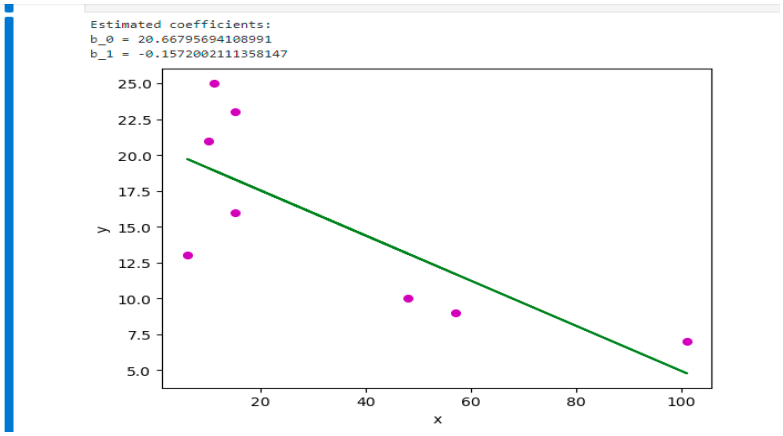

Figure 2. Best Fit Line for Year 2018

Table 2. Year 2019 Monitoring Data Of Ambient Quality Of Air (S. Singh(2020)

State / UT	City/ Town / Village	Location	No. of monitoring days	AQI		
				Minimum (24-hourly average)	Maximum (24-hourly average)	Annual Average
Uttarakhand	Dehradun	Raipur Road, Near parag Diary	68	21	26	23
		Clock Tower, PWD Guest House	70	23	27	25
		Himalaya Drug Co. Near ISBT	68	23	28	26
	Haldwani	Govt. Women Hospital	103	5	32	8
	Haridwar	SIDCUL, Haridwar	66	16	23	20
	Kashipur	BSNL Office, Kashipur	97	13	24	14
	Rishikesh	Nagar PalikaParishad	78	19	25	22
	Rudrapur	SIDCUL Office	87	12	23	14

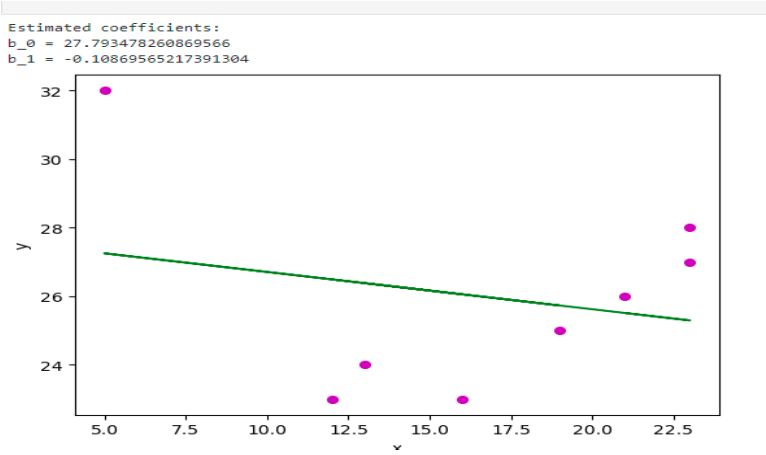

Figure 3. Best Fit Line for Year 2019

Table 3. Year 2020 Monitoring Data Of Ambient Quality Of Air (S. Singh (2020))

State / UT	City/Town/Village	Location	AQI		
			Minimum (24-hourly average)	Maximum (24-hourly average)	Annual Average
Uttarakhand	Dehradun	Raipur Road, Near parag Diary	6	26	21
		Clock Tower, PWD Guest House	7	27	22
		Himalaya Drug Co. Near ISBT	8	29	23
	Haldwani	Govt. Women Hospital	5	28	8
	Haridwar	SIDCUL, Haridwar	3	17	10
	Kashipur	BSNL Office,	8	29	13

Rishikesh	Kashipur Nagar PalikaParishad	6	24	20
Rudrapur	SIDCUL Office	8	22	13

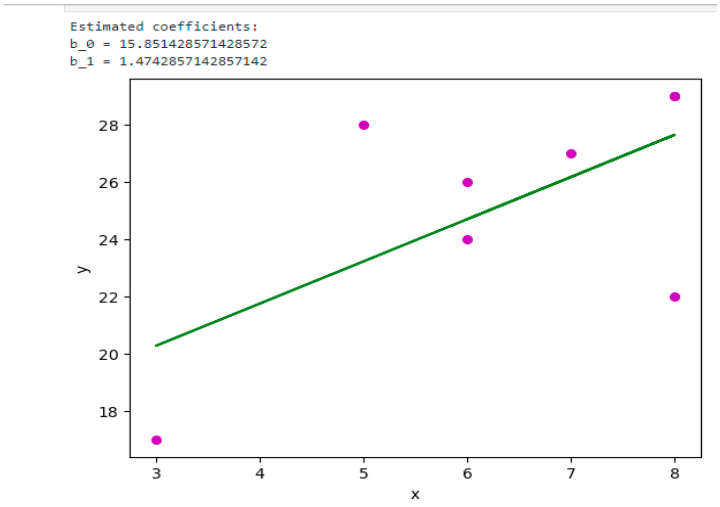

Figure 4. Best Fit Line for Year 2020

5. Result Analysis and Discussion

From the Regression results calculated for the year 2018, 2019 and 2020 as shown in table 4, figure 5, we came to the result that more the negative value of coefficient, the more will be the airquality index which leads to the more poor quality of air.

Table 4. Estimated Coefficient for the Year 2018, 2019 and 2020

Year	**b_0**	**b_1**
2018	20.667	-0.157
2019	27.793	-0.1086
2020	15.85	1.4742

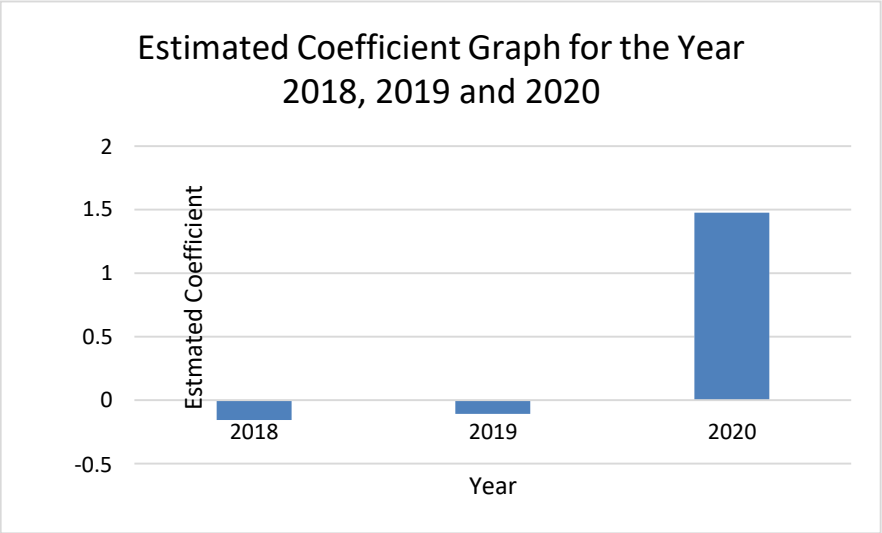

Figure 5. Coefficient Graph for the Year 2018, 2019 and 2020

As in year 2019 the Air quality index is more than in the year 2020 which results in the more poorair quality in year 2019 but in 2020 there is a improvement in the quality of air because of Covid-
19 lockdown and lesser use of transportation or vehicular facilities and also less usage of pollution sources.

Impact on Air quality index on and before lockdown

As it can be seen from the above graph in figure 5 the air quality has improved during lockdown as there was decrease in the index of air quality, due to less automobile emissions, less dust in the lockdown period, lesser use of

6. Conclusion

It was discovered that ambient air quality parameters such as SPM, RSPM, and SO2 were foundin high concentrations in the year 2019, and that these three parameters, namely SPM, RSPM,and SO2, exceeded the maximum permissible of NAAQ (2009) in almost all sampling locations, whereas the parameter NOx was found to be within limits in all sampling locations in Doon. The city of Dehradun had a high concentration of pollution. This high degree of pollution was discovered mostly as a result of automotive emissions. It was discovered that the number of two- wheelers (two strokes) was higher, which is extremely hazardous to human health. It is rapidly spreading around the world, including India, states, and cities.

References

Cai, M., Yin, Y., & Xie, M. (2009). Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. Transportation Research Part D: Transport and Environment, 14(1), 32-41.

Choubin, B., Abdolshahnejad, M., Moradi, E., Querol, X., Mosavi, A., Shamshirband, S., & Ghamisi, P. (2020). Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. Science of The Total Environment, 701, 134474.

Delavar, M. R., Gholami, A., Shiran, G. R., Rashidi, Y., Nakhaeizadeh, G. R., Fedra, K., &Hatefi Afshar, S. (2019). A novel method for improving air pollution prediction based on machine learning approaches: a case study applied to the capital city of Tehran. ISPRS International Journal of Geo-Information, 8(2), 99.

Pooja, S., Chani, P. S., & Parida, M. (2014). Assessing the performance of public transport operation in Dehradun, Uttrakhand, India. J. Environ. Res. Develop, 8(4), 1021-1025.

Ranzato, L., Barausse, A., Mantovani, A., Pittarello, A., Benzo, M., & Palmeri, L. (2012). A comparison of methods for the assessment of odor impacts on air quality: Field inspection (VDI 3940) and the air dispersion model CALPUFF. Atmospheric Environment, 61, 570-579.

Singh, O., Arya, P., & Chaudhary, B. S. (2013). On rising temperature trends at Dehradun in Doon valley of Uttarakhand, India. Journal of Earth System Science, 122(3), 613-622.

Singh, S. S. (2020). Air quality data of Indian states 2020. Kaggle. Retrieved December 27, 2020, from https://www.kaggle.com/shikharsinghpatel/air-quality-data-of-indian-states-yearly

Soni, A., Decesari, S., Shridhar, V., Prabhu, V., Panwar, P., & Marinoni, A. (2019). Investigation of potential source regions of atmospheric Black Carbon in the data deficit region of the western Himalayas and its foothills. Atmospheric Pollution Research, 10(6), 1832-1842.

Steinfeld, J. I. (1998). Atmospheric chemistry and physics: from air pollution to climate change. Environment: Science and Policy for Sustainable Development, 40(7), 26-26.

Venegas, L. E., Mazzeo, N. A., & Dezzutti, M. C. (2014). A simple model for calculating air pollution within street canyons. Atmospheric Environment, 87, 77-86.

Wark K, Warner CF, Davis WtF. Air pollution: its origin and control. 3rd ed.Addison-Wesley; 1998.

Wang, P., Liu, Y., Qin, Z., & Zhang, G. (2015). A novel hybrid forecasting model for PM10

and SO2 daily concentrations. Science of the Total Environment, 505, 1202-1212.

Zeinalnezhad, M., Chofreh, A. G., Goni, F. A., & Klemeš, J. J. (2020). Air pollution prediction using semi-experimental regression model and Adaptive Neuro-Fuzzy Inference System. Journal of Cleaner Production, 261, 121218.