Prediction of Air Quality Influential Factors with AtmosphericAir Present Pollutants Using Multiple Regression Methodology

Chebrolu Ramadevi[1], S Ravindran[2]

Assistant professor, Computer science Engineering, NRI Institute of Technology, AndhraPradesh, India.
SQL Developer, ICT Infracon LLP, 11-Sumangalam Society, B/H Asia School, Drive In Road, Ahmedabad-380054, India.

ABSTRACT

Globally the temperature and environmental behavior factors has been changing and raising day by day. Change in pollution, air quality causes serious problem to survival of humans, and other living creatures as well. So, an air quality index is used by government agencies to communicate to the publichow much the air currently polluted. AQI information is obtained by averaging readings from an air quality sensors and other sources as well. Our objective is to identify the most important parameters that causes increasing of temperature more and to predict the value of temperature based on past dataset. In this paper we implemented regression methodology by considering the Air Quality data of Italian city. The Ground Truth hourly averaged concentrations for CO, Hydrocarbons, Benzene, Nitrogen Oxides and Nitrogen Dioxide pollutants were used. After doing the experimentation coefficients of thevariables were estimated to know how correlate each other. R-squared and adjusted r- squared metricsused to evaluate the model. The Residual scores are described after implementation of the multiple regression methodology.

Keywords: *AQI, regression, atmospheric pollutants, Particulate matter, dependent and independent variable*

1. Introduction

AQI is the most important index which can be used to assess the current air quality as well as pollutionlevel in the particular region. Public health risks increase as the AQI rises, especially affecting children,the elderly, and individuals with respiratory or cardiovascular issues. During these times, governmental bodies generally encourage people to reduce physical activity outdoors, or even avoid going out altogether. Air quality index is used by government agencies to communicate to the public how much the air currently polluted [1]. Different countries have their own air quality indices, corresponding to different national air quality standards. Some of the standards used are Air Quality Health Index by Canada, Air Pollution Index by Malaysia, and the Pollutant Standards Index by Singapore, Air Quality index by India etc. AQI information is obtained by averaging readings from an air quality sensors andother sources as well [2].

There are several factors causing to increase the more air pollution level pollutants such as ozone, nitrogen dioxide, Sulphur dioxide, among others. The reasons are such as due to vehicle traffic, forest fires, industries, human utilities or anything that can increase air pollution [3].

AQI level is based on the level of six atmospheric pollutants, namely

- Sulfur dioxide (SO2),
- Nitrogen dioxide (NO2),
- Suspended particulates smaller than 10μm or PM10,
- Suspended particulates smaller than 2.5μm or PM2.5,
- Carbon monoxide (CO),
- Ozone (O3) measured at the monitoring stations throughout each city

1.1. AQI level indicator

To indicate the level of pollution will use generally AQI scale based on the obtained or calculated valuewill be compared with the scale measurement indicator. Based on the indication range the status or severity of pollution will determine.

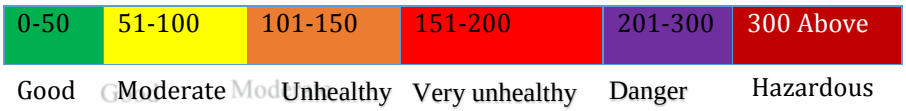

Figure 1: AQI Scale for severity prediction

Seasonal and daily AQI calculation revealed that air quality status in the study region under various classes ranging from good, moderate, satisfactory and unacceptable class for different AQI calculation.

1.2. Side effects of low AQ levels

Effects of air pollution have impaired human health for a long time [3], [4]. Ever since the industrial revolution, humans have been pumping out large volumes of carbon dioxide and other greenhouse gases. Global warming and greenhouse effect is one of the most notable effects of air pollution at the global level [5]–[7]. On an individual level, air pollution causes the following effects on human health.

- Heart, breathing related disease
- Lung Diseases
- Cancer
- Harming healthy life
- Damage Central Nervous System
- Global Climate Changes
- Effects on humans, animals, materials, plants as well
- Leads to acid rains as well

These are the various reasons that cause more damage to our environment, human lives as wellso we started working on air quality data set to estimate how the variables will impact each other as well as we implemented prediction modelling.

The rest of the paper is structured as follows, section 2 briefed about methodology and data set description. In section 3 we put up obtained results and graphs. Conclusion and future work are definedin section 4.

2. Methodology Data Set Information:

The dataset formulated with an Air Quality Chemical Multi-Sensor Device embedded with 5 metal oxide chemical sensors which records hourly averaged responses. The device was located on the field in a significantly polluted area, at road level, within an Italian city. The dataset has 9358 instances whichwere recorded from March 2004 to February 2005 [8].

Attribute Information:

- ✓ True hourly averaged concentration CO in mg/m^3
- ✓ PT08.S1 (tin oxide) hourly averaged sensor response
- ✓ Non Metanic Hydrocarbons concentration in microg/m^3
- ✓ Benzene concentration in microg/m^3
- ✓ PT08.S2 (titania)

- ✓ NOx concentration in ppb
- ✓ PT08.S3 (tungsten oxide)
- ✓ NO2 concentration in microg/m^3
- ✓ PT08.S4 (tungsten oxide)
- ✓ PT08.S5 (indium oxide)
- ✓ Temperature
- ✓ Relative Humidity
- ✓ AH Absolute Humidity

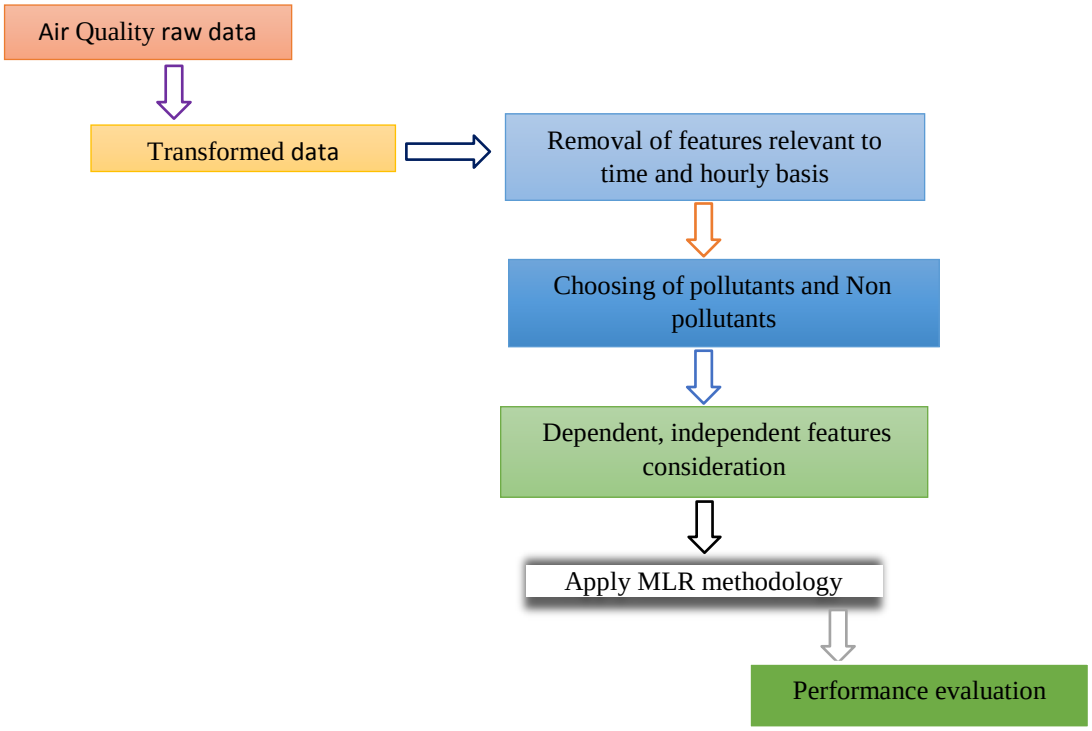

Figure 2: Proposed approach for Temperature prediction from Air Quality data

Initially the data collected we collected from data repository. We filtered the data by removing irrelevant features. We considered the features of pollutants which are important and needed for analysis. By considering the data we defined dependent and independent variables corresponding to implement regression methodology [9].

3. Experimental Results

The analysis of regression methodology [10] done over the air quality data set and we evaluated using different evaluation metrics. The metrics we used are R-square and adjusted R-square [11]. The below table 1 represents various regression statistics and along with actual, predicted values and residual differences aswell.

Regression Statistics		Observations	Actual values	Predicted	Residuals
Multiple R	0.997436	1	13.6	15.89311573	-2.593115734
R Square	0.994880	2	13.3	13.86648819	-1.96648819
Adjusted R Square	0.994873	3	11.9	12.70700967	-1.707009668
Standard Error	3.093455	:	::::::	::::::	::::::
Observations	9356	9	10.7	10.19345773	0.106542268
		10	10.7	7.447265924	2.652734076
		11	10.3	8.241241915	2.758758085
		:		::::::	::::::
		4739	21.1	18.21656718	2.183432821
		4740	20.4	20.76494901	0.035050993
		4741	20.8	23.55561139	-2.255611387
		4742	21.3	24.77841324	-2.878413241
		:	::::::	::::::	::::::
		9347	11.8	9.228354807	1.171645193
		9348	10.4	10.40758875	-0.907588755
		9349	9.5	14.41885762	-4.71885762
		:	::::::	::::::	::::::

Table 1: Representation of regression statistics and showing actual, predicted values with residuals

Based on predicted, actual, residual scores we plotted the graph which is represented in figure 3. The curves of predicted and actual 1 represents various regression statistics and along with actual, predicted values and residual differences as well.

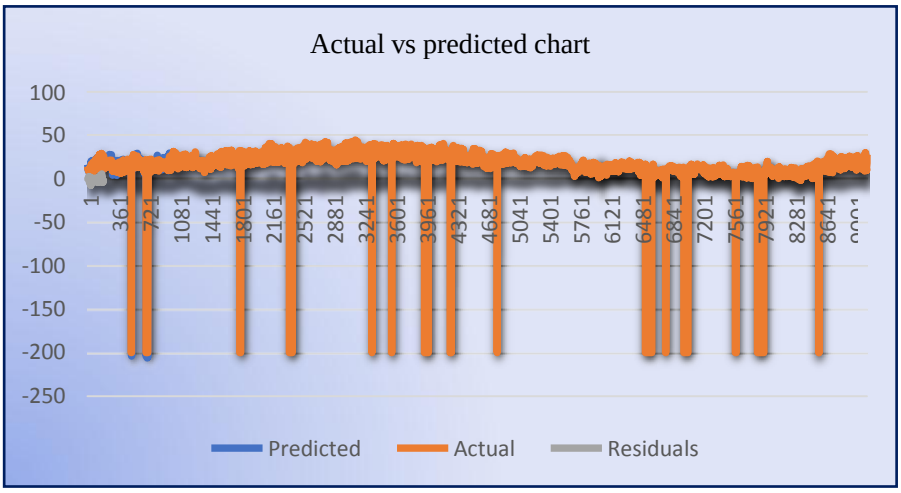

Figure 3: Graph representing Actual vs predicted from Air Quality data

	Coefficients	StandardError	t Stat	Lower95%	Upper 95%
Intercept	12.0934958	0.845781	14.29861	10.43558	13.75141
2.6	0.001400334	0.000567	2.467717	0.000288	0.002513
1360	-0.00184685	0.000451	-4.09433	-0.00273	-0.00096
150	-0.008192661	0.000275	-29.8438	-0.00873	-0.00765
11.9	-0.734283107	0.027051	-27.1445	-0.78731	-0.68126
1046	0.001137903	0.00095	1.19837	-0.00072	0.002999
166	0.00646517	0.000329	19.65011	0.00582	0.00711
1056	-0.006610725	0.000277	-23.8778	-0.00715	-0.00607
113	-0.011212918	0.000575	-19.5137	-0.01234	-0.01009
1692	0.024733971	0.000191	129.7509	0.02436	0.025108
1268	-0.004077349	0.000234	-17.4133	-0.00454	-0.00362
48.9	-0.316230013	0.002449	-129.143	-0.32103	-0.31143
0.7578	2.108685511	0.022997	91.69399	2.063606	2.153765

Table 2: Representation of estimated regression coefficients, lower, upper boundaries

4. Conclusion & Future work

Globally the temperature and environmental behavior factors has been changing and raising day by day. Change in pollution, air quality causes serious problem to survival of humans, and other living creatures as well. AQI is the most important index which can be used to assess the current air quality as well as pollution level in the particular region. In this paper we implemented regression methodologyby considering the Air Quality data of Italian city. The Ground Truth hourly averaged concentrations for CO, Hydrocarbons, Benzene, Nitrogen Oxides and Nitrogen Dioxide pollutants were used. After doing the experimentation coefficients of the variables were estimated to know how correlate each other. To evaluate the model. The R-squared and adjusted r- squared metrics and residual scores are obtained after the implementation of the multiple regression methodology. The results obtained are Described in the above in the table 1. We did not consider the real-time data set relevant for air quality prediction analysis. In the future using the concepts of deep learning may help more to carry out further processeseffectively.

References

[1] "Air Quality & Fire Weather - NOAA Weather Program Office." https://wpo.noaa.gov/air- quality-fire-weather/ (accessed Nov. 23, 2022).

[2] "World Air Quality Index (AQI) Ranking | IQAir." https://www.iqair.com/world-air-quality- ranking (accessed Nov. 23, 2022).

[3] "Conference Title: Keywords - Global Warming, Air Pollution, Vehicle Emission, Smart System, Machine Learning | PDF | Regression Analysis | Dependent And Independent Variables." https://www.scribd.com/document/415723414/Elsevier-Paper (accessed Nov. 23, 2022).

[4] R. Rastogi *et al.*, "Examining the AQI with Effect of Agnihotra in NCR Region: Extracting Knowledge for Sustainable Society and Holistic Development with Healthcare 5.0," pp. 271– 280, 2023, doi: 10.1007/978-981-19-6068-0_27.

[5] M. W. Murad and J. J. Pereira, "Malaysia: Environmental health issues," *Encyclopedia of Environmental Health*, pp. 194–210, Jan. 2019, doi: 10.1016/B978-0-12-409548-9.11529-5.

[6] M. W. Murad and J. J. Pereira, "Malaysia: Environmental Health Issues," *Encyclopedia of Environmental Health*, pp. 577–594, Jan. 2011, doi: 10.1016/B978-0-444-52272-6.00539-0.

[7] P. Salvador, "Ozone, SOx and NOx, particulate matter, and urban air," *Encyclopedia of the Anthropocene*, vol. 1–5, pp. 7–21, Jan. 2017, doi: 10.1016/B978-0-12-809665-9.09975-4.

[8] S. de Vito, E. Massera, M. Piga, L. Martinotto, and G. di Francia, "On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario," *Sens Actuators B Chem*, vol. 129, no. 2, pp. 750–757, Feb. 2008, doi: 10.1016/J.SNB.2007.09.060.

[9] S. S. Ganesh, S. H. Modali, S. R. Palreddy, and P. Arulmozhivarman, "Forecasting air quality index using regression models: A case study on Delhi and Houston," *Proceedings - International Conference on Trends in Electronics and Informatics, ICEI 2017*, vol. 2018-January, pp. 248– 254, Feb. 2018, doi: 10.1109/ICOEI.2017.8300926.

[10] G. K. Uyanık and N. Güler, "A Study on Multiple Linear Regression Analysis," *Procedia Soc Behav Sci*, vol. 106, pp. 234–240, Dec. 2013, doi: 10.1016/J.SBSPRO.2013.12.027.

[11] J. Miles, "R Squared, Adjusted R Squared", doi: 10.1002/9781118445112.STAT06627.