International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

DETECTION AND PREDECTION OF AIR POLLUTION USING ML MODELS

[1]M.Suresh, [2]Sawa Triveni, [3]Gundareddy Pravallika, [4]Tumula Gopi Krishna,[5]Kotha Srikanth Reddy, [6]Vangala Srilakshmi Naga Siromani
[1,2,3,4,5,6]Assistant Professor, Department of CSE in Narasaraopet Institute Of Technology

ABSTRACT

In the populated and developing countries, governments consider the regulation of air as a major task. The meteorological and traffic factors, burning of fossil fuels, industrial parameters such as power plant emissions play significant roles in air pollution. Among all the particulate matter that determine the quality of the air, Particulate matter (PM 2.5) needs more attention. When it's level is high in the air, it causes serious issues on people's health. Hence, controlling it by constantly keeping a check on its level in the air is important. In this paper, Logistic regression is employed to detect whether a data sample is either polluted or not polluted. Autoregression is employed to predict future values of PM2.5 based on the previous PM2.5 readings. Knowledge of level of PM2.5 in nearing years, month or week, enables us to reduce its level to lesser than the harmful range. This system attempts to predict PM2.5 level and detect air quality based on a data set consisting of daily atmospheric conditions in a specific city.

Keywords — Pollution detection, Pollution Prediction, Logistic Regression, Linear Regression, Autoregressio

1. INTRODUCTION

Air is perhaps the most fundamental characteristic assets for the endurance and presence of the each and each life on this planet. All types of life including plants and creatures rely upon air for their fundamental endurance and presence. Consequently, all the living beings require great nature of air which is liberated from destructive gases to proceed with their life. The expanding populace, its cars and businesses are contaminating all the air at a disturbing rate. Air contamination can cause different long haul and transient wellbeing impacts.

The advancement of air quality prescient models can be useful as such models can give early admonitions of contamination levels expanding to unacceptable levels. Contamination fixations in metropolitan zones are fundamentally from vehicular fumes, manufacturing plants, and limited scope ventures. Air contamination can influence our wellbeing and climate from numerous points of view. A huge number of untimely passing happen each year because of breathing in a high level of contamination focuses like PM10, PM2.5, CO, Nitrogen Oxides (NO+NO2). In the previous few years, the substantial natural stacking has prompted the weakening of air quality in metropolitan and modern zones in Ghaziabad. This welcomed us to zero in our examination on air quality in Ghaziabad. The assignment of controlling and improving air quality level has pulled in a lot of public consideration. The principle focal point of this paper is to investigate the reasonable AI methods that will help in better anticipating of the air contamination fixation.

The information is gathered from CPCB (Central Pollution Control Board) online information and sensors over the objective district. At that point the dissemination of suspended particles like PM10, PM2.5, SO2, and NO2 contaminated climate air are recognized.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

The information is analyzed utilizing information mining methods for as long as 5 months (2017-2018). In foreseeing the air contamination, we likewise discovered those meteorological factors to a great extent influences the air contamination. Along these lines, in our examination, we additionally consider some meteorological variables to anticipate the contamination of air like Temperature, Least Temperature, Maximum Temperature, Wind speed and Relative Humidity and a few different highlights. This will help in the forecast of air quality in metropolitan also, modern regions of Ghaziabad and this could fill in as a significant reference for government offices in assessing present and contriving future air contamination approaches. Our examination centers around expectation of air contamination level of a specific or explicit locale. In air contamination forecast, model exactness, productivity and flexibility are key contemplations.

2. LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things r satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

Witten, Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016. "Data Mining: Practical Machine Learning Tools and Technique" may become a key reference to any student, teacher or researcher interested in using, designing and deploying data mining techniques and applications. This book also deals with various aspects relevant to undergraduate or research programmes in machine learning, intelligent systems, bioinformatics and biomedical informatics.

Li S., and Shue L., "Data mining to aid policy making in air pollution management," Expert Systems with Applications, vol. 27, pp. 331-340, 2004.

In this paper we are applied mining process to extract knowledge from weather dataset .Data collected from DAV BDL public school, Bhanur, Medak weather station. Data set includes four years period [2011-2015] of daily weather observations. We are trying to apply data mining techniques clustering, Association and classification. We collected Weather data parameters temperature, pressure, humidity and dew point, wind speed rain fall and wind direction .These parameters one is related other parameter. For outlier detection, data analysis and experimental

results we are used weka data mining tool and graphs from Excel tool provide a very useful and accurate knowledge in a form of tables and graphs. This knowledge can be used to obtain decision making for different areas like Agriculture, Air pollution; Disaster Management and also for prediction .Our future work includes building an automatic, efficient and accurate system to predict weather

3. System analysis

3.1 Existing system:

The characteristics of air pollution to changes the everyday pose emerging challenges for these existing research problems in new problem settings to predicting the weather. On the one hand, users often write tweets in a very casual manner. Acronyms, misspellings, and special tokens make tweets noisy, and

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

techniques developed for formal documents are error-prone on tweets.

Disadvantages:

• Most cost effective

• Performance is less

3.2 Proposed system:

The prediction models proposed based on different city air pollution data record collecting can also be adapted to other whether report , while might require some changes. But before considering model adaptations, we need to be clear on whether the three geo location problems on cities, i.e., prediction of home location on pollution rate, city location and mentioned location, are applicable to the target platform or not. For example, based on city mentioned location prediction on some image and video sharing platforms and other whether reporting platforms may not be applicable to based on prediction the weather reports.

Advantages:

• Less cost effective

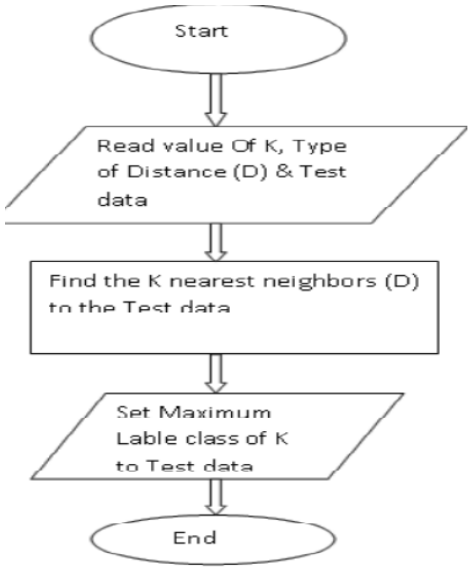
• Performance is high

4. ALGORITHMS

4.1 K-Nearest Neighbors Algorithm:

KNN is a nonparametric supervised learning technique that uses training sets to segment data points into given categories. In simple classifications, the word collects information from all educational cases and similarities based on the new case. Look at the training for the most similar (neighbor) K cases and predict the new instance (x) by summarizing the output variables for these K cases. Classification is the class value mode (or most commonly). A flow diagram of the KNN algorithm is shown in Figure 2

FLOWCHART

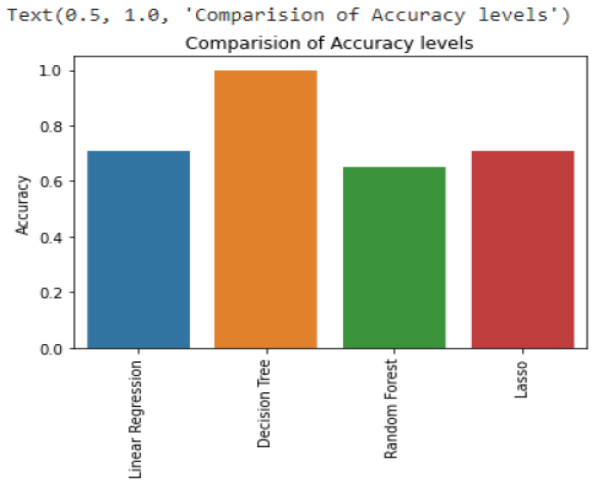


KNN Algorithm Flowchart

5. Results

	StationId	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	Xylene	AQI	AQI_Bucket
0	AP001	2017-11-24	71.36	116.76	1.75	20.55	12.40	12.19	0.10	10.76	109.26	0.17	5.92	0.10	NaN	NaN
1	AP001	2017-11-25	81.40	124.50	1.44	26.50	12.08	10.72	0.12	15.24	127.09	0.20	6.50	0.06	184.0	Moderate
2	AP001	2017-11-26	78.52	129.06	1.26	26.00	14.85	10.28	0.14	26.96	117.44	0.22	7.95	0.08	197.0	Moderate
3	AP001	2017-11-27	88.76	135.32	6.60	30.85	21.77	12.91	0.11	33.93	111.61	0.29	7.63	0.12	198.0	Moderate
4	AP001	2017-11-28	64.18	104.06	2.56	28.07	17.01	11.42	0.09	19.00	138.18	0.17	5.02	0.07	188.0	Moderate

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362



Bar Graph of Air pollution

6. Conclusion

The proposed system will definitely help in improving the prediction of air pollution in our smart city. Multivariate Multistep Time Series Prediction Using Random Forest technique improve the performance and reduce the

complexity of the air pollution prediction model. Also, here we are using feature selection technique which makes our prediction even better.

With the advancement of Machine Learning Techniques Real-time air quality prediction and evaluation is desirable for future smart cities. Our study focuses on prediction of air pollution level of a particular or specific region. In air pollution prediction, model accuracy, efficiency and adaptability are key considerations. This project reports our recent study in predicting air pollution level in Ghaziabad region, and compares result of different algorithms.

This system has the feature for the people to view the predicted amount of pollution on their mobile phones through websites. Letting civilians also involved in the process adds an extra value to it. As civilians are equally aware and curious about the environment, this concept of air pollution prediction system is beneficial for the welfare for the society. And it is implemented using the latest technology. The paper also highlights some observations on challenges and needs.

Reference:

1. Witten, Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

2. Li S., and Shue L., "Data mining to aid policy making in air pollution management,"

Expert Systems with Applications, vol. 27, pp. 331-340, 2004.

3. Gu, Ke, Junfei Qiao, and Weisi Lin. "Recurrent Air Quality Predictor Based on Meteorology and Pollution Related Factors." IEEE Transactions on Industrial Informatics (2018).

4. García Nieto, P.J., Sánchez Lasheras, F., GarcíaGonzalo, E. et al. “Estimation of PM10 concentration from air quality data in the vicinity of the major steel works site in the metropolitan area using machine learning techniques”Stoch Environ Res Risk Assess (2018), https://link.springer.com/article/10.1007/s0047 7- 018-1565-6.

5. Hu, Ke, Ashfaqur Rahman, Hari Bhrugubanda, and Vijay Sivaraman. "Hazeest: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors." IEEE Sensors Journal 17, no. 11 (2017): 3517-3525.

6. K. B. Shaban, A. Kadri, and E. Rezk, “Urban Air Pollution Monitoring System With Forecasting Models” IEEE Sensors Journal, vol. 16, no. 8, pp. 2598–2606, Apr. 2016.

7. Xiao Feng, Qi Li, Yajie Zhu, “Artificial Neural Network Forecasting of PM2.5 Pollution using Air Mass Trajectory based Geographic Model and Wavelet

Transformation" Atmospheric Environment Journal, www.elsevier.com/locate/atmosenv, 2015.

8. M. S. Baawain and A. S. Al-Serihi, "Systematic approach for the prediction of ground-level air pollution (around an industrial port) using an artificial neural network," Aerosol and Air Quality Research, vol. 14, pp. 124–134, 2014.

9. W.-Z. Lu and D. Wang, "Learning machines: Rationale and application in groundlevel ozone prediction," Applied Soft Computing`, vol. 24, pp. 135–141, Nov. 2014.

10. A. Sotomayor-Olmedo, M. A. Aceves-Fernández, E. Gorrostieta-Hurtado, C. Pedraza-Ortega, J. M. RamosArreguín, and J. E. Vargas-Soto, "Forecast Urban Air Pollution in Mexico City by Using Support Vector Machines: A Kernel Performance Approach," International Journal of Intelligence Science, vol. 3, no. 3, pp. 126–135, Jul. 2013.

11. Shweta Taneja, Dr. Nidhi Sharma, Kettun Oberoi, Yash Navoria ,"Predicting Trends in Air Pollution in Delhi using Data Mining", IEEE(2016)