

IoT Based Air Quality Prediction using SVM and Random Forest

Devata R. Anekar¹, Ashish M. Hase², Dinesh Bharat Nikam³, Dinesh Balkawade⁴, Arti R. Masare⁵

Assistant Professor, Department of Information Technology

UG Scholar, Department of Information^{2,3,4,5}

Sinhgad Academy of Engineering, Pune, Maharashtra, India

Savitribai Phule Pune University, Pune, India

Abstract: *Internet of Things (IoT) may be a worldwide System of “smart devices” which will sense and connect with their surroundings and interact with users and other systems. Global air pollution is one of the major concerns of our era. The level of pollution has increased with time by a lot of things like the increase in population, increased vehicle use, industrialization, and urbanization which ends up in harmful effects on human wellbeing by directly affecting the health of the population exposed to it. Air quality goes down when enough amount of harmful gases are present in the air like carbon dioxide, smoke, alcohol, benzene, NH₃, and NO₂. To analyses, we are developing an IoT Based pollution Monitoring System which we'll monitor the Air Quality over an internet server. Existing monitoring systems have inferior precision, low sensitivity, and need laboratory analysis. Therefore, improved monitoring systems are needed.*

Keywords: Air Quality Monitoring; Machine Learning; Air Quality Index

I. INTRODUCTION

The Environment is nothing but everything that encircles us. The environment is getting polluted due to human activities and natural disasters, very severe among them is air pollution. The concentration of air pollutants in ambient air is governed by meteorological parameters such as atmospheric wind speed, wind direction, relative humidity, and temperature. If the humidity is more, we feel much hotter because sweat will not evaporate into the atmosphere. Urbanization is one of the main reasons for air pollution because the increase in the transportation facilities emits more pollutants into the atmosphere and another main reason for air pollution is Industrialization. The major pollutants are Nitrogen Oxide (NO), Carbon Monoxide (CO), Particulate matter (PM), SO₂, etc. Carbon Monoxide is produced due to the deficient Oxidization of propellants such as petroleum, gas, etc. Nitrogen Oxide is produced due to the ignition of thermal fuel; Carbon monoxide causes headaches, vomiting; Benzene is produced due to smoking, it causes respiratory problems; Nitrogen oxides cause dizziness, nausea; Particulate matter with a diameter of 2.5 micrometer or less than that affects more to human health. Measures must be taken to minimize air pollution in the environment. Air Quality Index(AQI), is used to measure the quality of air. Earlier classical methods such as probability, statistics were used to predict the quality of air, but those methods are very complex to predict the quality of air. Due to the advancement of technology, now it is very easy to fetch data about the pollutants of air using sensors. Assessment of raw data to detect the pollutants needs vigorous analysis. Convolution Neural networks, Recursive Neural networks, Deep Learning, Machine learning algorithms assure accomplishing the prediction of future AQI so that measures can be taken appropriately. Machine learning which comes under artificial intelligence has three kinds of learning algorithms, they are Supervised Learning, Unsupervised Learning, Reinforcement Learning. In the proposed work we have used a supervised learning approach. There are many algorithms under supervised learning algorithms such as Linear Regression, Nearest Neighbor, SVM, kernel SVM, Naive Bayes, and Random Forest. Compared to all other algorithms SVM and Random forest give better results, so our approach selects them to predict accurate air pollution.

II. LITERATURE SURVEY

Chakradhar ReddyK and Nagarjuna Reddy K[1], The supervised machine learning technique (SMLT) was used to gather several pieces of information from the dataset, including variable recognition, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatment and analysis, data cleaning/preparation, and data representation. Their findings



provide a valuable guide to sensitivity analysis of model parameters in terms of success in air quality pollution prediction through accuracy measurement.

Yun-Chia Liang and Yona Maimury[2], including adaptive boosting (AdaBoost), artificial neural network (ANN), random forest, stacking ensemble, and support vector machine (SVM), produce promising results for air quality index (AQI) level predictions. A series of experiments, using datasets for three different regions to obtain the best prediction performance from the stacking ensemble, AdaBoost, and random forest, found the stacking ensemble delivers consistently superior performance for R² and RMSE, while AdaBoost provides best results for MAE.

Madhuri VM, Samyama Gunjal GH[3], the proposed work was based on a supervised learning approach using different algorithms such as LR, SVM, DT, and RF. The result shows that AQI predictions obtained through RF are promising which are analyzed with the result. Akshatha S and Jayaram M N[4], addressed the Objective of the framework was to utilize different sensors and servers to plan a proficient air quality checking framework without affecting the common habitat and give live updates to keep away from clashes. Mauro Castelli and Fabiana Martins Clemente [5], proposed a famous AI strategy, support vector relapse (SVR), to figure poison and particulate levels and to foresee the air quality list (AQI). Among the different tried other options, outspread premise work (RBF) was the sort of portion that permitted SVR to get the most reliable expectations. Utilizing the entire arrangement of accessible factors uncovered a more effective technique than choosing highlights utilizing head part investigation. &e introduced results show that SVR with RBF bit permitted to precisely foresee hourly poison focuses, similar to carbon monoxide, sulfur dioxide, nitrogen dioxide, ground-level ozone, and particulate matter 2.5, just as the hourly AQI for the territory of California.

Rajeev Tiwari, Shuchi Upadhyay, Parv Singhal [6], The Aim was to foster a fake neural organization for air quality expectation that can perform with obliged dataset with exceptionally vigorous element to deal with the information including commotion and mistakes. Dataset utilized arrangements with contamination in the U.S. including four significant toxins (Nitrogen Dioxide, Sulfur Dioxide, Carbon Monoxide and Ozone) on everyday schedule for the time-frame of year 2008 to 2017.

Shivam Sharma and Nishu Soni [7], The point of the framework was to execute an air quality checking gadget utilizing the Internet of Thing which is otherwise called IoT for controlling air contamination and improve the nature of air. This framework estimates the continuous air quality list, temperature, dampness which is shown on a site with the assistance of web.

Mrs. A. Gnana Soundari Mrs. J. Gnana Jeslin M.E, Akshaya A [8], they proposed model skilled for effectively foreseeing the air quality list of a complete region or any state or any limited locale furnished with the authentic information of toxin focus. In our model by executing the proposed boundary diminishing details, we accomplished preferred execution over the standard relapse models. our model has 96% precision on foreseeing the currently accessible dataset on anticipating the air quality list of entire India, likewise we use AHP MCDM method to find of request of inclination by comparability to ideal arrangement.

Vineeta , Ajit Bhat , Asha S Manek , Pranay Mishra[9] , they represented the data has been collected from the two sources in the Bengaluru region: government website and static sensors built using Arduino. The level of CO is measured using three machine algorithms namely Random Forest Regression (RFR), Decision Tree Regression (DTR) and Linear Regression (LR). The results show that RFR gives least error of the three and hence more accuracy. Command line interface has also been created to see the CO level prediction. Fabiana Martins Clemente, Aleš Popovič, Sara Silva, and Leonardo Vanneschi[10], they employ a popular machine learning method, support vector regression (SVR), to forecast pollutant and particulate levels and to predict the air quality index (AQI). Among the various tested alternatives, radial basis function (RBF) was the type of kernel that allowed SVR to obtain the most accurate predictions. Using the whole set of available variables revealed a more successful strategy than selecting features using principal component analysis. The presented results demonstrate that SVR with RBF kernel allows us to accurately predict hourly pollutant concentrations, like carbon monoxide, sulfur dioxide, nitrogen dioxide, ground-level ozone, and particulate matter 2.5, as well as the hourly AQI for the state of California. Classification into six AQI categories defined by the US Environmental Protection Agency was performed with an accuracy of 94.1% on unseen validation data. S. L. Bangare et al. [9-15] have worked in the brain tumor detection. N. Shelke et al [16] given LRA-DNN method. Suneet Gupta et al [17] worked for end user system. Gururaj Awate et al. [18] worked on Alzheimers Disease. P. S. Bangare et al [19] worked on the object detection. Kalpana Thakare et al

[20-25] have worked on various machine learning algorithms. M. L. Bangare et al. [26-28] worked on the cloud platform. Rajesaheb R. Kadam et al [29] and Sachindra K. Chavan et al. [30] have discussed security issues with cloud.

III. OBJECTIVES OF SYSTEM

- The goal is to develop machine learning-based air quality forecasting techniques that are as accurate as possible
- To Train the dataset using Support Vector Machine Algorithm and Random Forest Techniques
- To use pygmeps library. To plotting area wise on Map after receiving data from sensors
- To mark alert on map according to location / area. Red if huge, Orange if minor and Green if clear.

IV. IMPLEMENTATION DETAILS OF MODULE

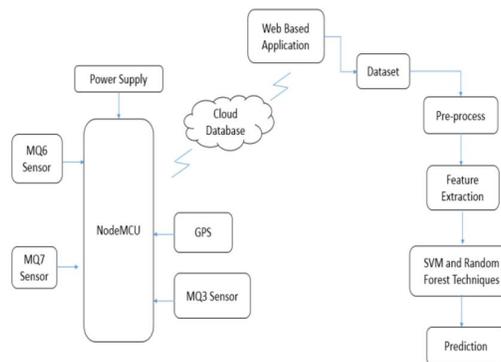


Figure: System Architecture

The system is designed with a combination of hardware and software. The system includes node MCU which is been used as the main controller and other gas sensors are been interfaced to node MCU and GPS is used to get the location of the specific area. The data collected from the sensor is been trained using SVM Techniques like Plotting on Map after receiving data from sensors. For Plotting, we are using pygmeps library. The alert is indicated on the map according to location/area. Red if huge, Orange if minor and Green if clear.

V. CONCLUSION

As a result, our project is to check the quality of the exposed level in the air pollution. Our project was designed to help a person to detect and predict the air quality in a particular area. Air Pollution is the major affecting factor to our environment. Not only affecting the environment but also affecting human health. The web-based application is developed to predict air quality. The gas sensors were used for identifying the gases and GPS for getting the location of the area and the graph is plotted accordingly.

REFERENCES

- [1]. Chakradhar Reddy K. and Nagarjuna Reddy K , “The Prediction of Quality of the Air Using Supervised Learning”, IEEE 2021
- [2]. Yun-Chia Liang and YonaMaimury , “Machine Learning-Based Prediction of Air Quality ” , Applied Science , 2020
- [3]. Madhuri V M, SamyamaGunjal GH, SavithaKamalapurkar , “Air Pollution Prediction Using Machine Learning Supervised Learning Approach ”, International Journal Of Scientific & Technology Research Volume 9, Issue 04, April 2020
- [4]. Mauro Castelli and Fabiana Martins Clemente, “A Machine Learning Approach to Predict Air Quality in California” , Research Article , 2020
- [5]. Rajeev Tiwari, ShuchiUpadhyay, ParvSinghal , “Air Pollution Level Prediction System”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) , 2019

- [6]. Vineeta , Ajit Bhat , Asha S Manek , Pranay Mishra, "Machine Learning based Prediction System for Detecting Air Pollution" IJERT,2019
- [7]. Mrs. A. Gnana Soundari Mrs. J. Gnana Jeslin M.E, Akshaya A."Indian Air Quality Prediction And Analysis Using Machine Learning,"IJERT 2020
- [8]. Fabiana Martins Clemente,Aleš Popovič,Sara Silva, and Leonardo Vanneschi"A Machine Learning Approach to Predict Air Quality in California", 2020.
- [9]. S. L. Bangare, G. Pradeepini, S. T. Patil, "Implementation for brain tumor detection and three dimensional visualization model development for reconstruction", ARPN Journal of Engineering and Applied Sciences (ARPN JEAS), Vol.13, Issue.2, ISSN 1819-6608, pp.467-473. 20/1/2018 http://www.arpnjournals.org/jeas/research_papers/rp_2018/jeas_0118_6691.pdf
- [10]. S. L. Bangare, S. T. Patil et al, "Reviewing Otsu's Method for Image Thresholding." International Journal of Applied Engineering Research, ISSN 0973-4562, Volume 10, Number 9 (2015) pp. 21777-21783, © Research India Publications <https://dx.doi.org/10.37622/IJAER/10.9.2015.21777-21783>
- [11]. S. L. Bangare, G. Pradeepini, S. T. Patil, "Regenerative pixel mode and tumor locus algorithm development for brain tumor analysis: a new computational technique for precise medical imaging", International Journal of Biomedical Engineering and Technology, Inderscience, 2018, Vol.27 No.1/2. <https://www.inderscienceonline.com/doi/pdf/10.1504/IJBET.2018.093087>
- [12]. S. L. Bangare, A. R. Khare, P. S. Bangare, "Quality measurement of modularized object oriented software using metrics", ICWET '11: Proceedings of the International Conference & Workshop on Emerging Trends in Technology, February 2011, pp. 771–774. <https://doi.org/10.1145/1980022.1980190.1>.
- [13]. S. L. Bangare, G. Pradeepini and S. T. Patil, "Brain tumor classification using mixed method approach," 2017 International Conference on Information Communication and Embedded Systems (ICICES), 2017, pp. 1-4, doi: 10.1109/ICICES.2017.8070748.
- [14]. S. L. Bangare, S. Prakash, K. Gulati, B. Veeru, G. Dhiman and S. Jaiswal, "The Architecture, Classification, and Unsolved Research Issues of Big Data extraction as well as decomposing the Internet of Vehicles (IoV)," 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), 2021, pp. 566-571, doi: 10.1109/ISPCC53510.2021.9609451.
- [15]. S. L. Bangare, G. Pradeepini, S. T. Patil et al, "Neuroendoscopy Adapter Module Development for Better Brain Tumor Image Visualization", International Journal of Electrical and Computer Engineering (IJECE) Vol. 7, No. 6, December 2017, pp. 3643~3654. <http://ijece.iaescore.com/index.php/IJECE/article/view/8733/7392>
- [16]. N. Shelke, S. Chaudhury, S. Chakrabarti, S. L. Bangare et al. "An efficient way of text-based emotion analysis from social media using LRA-DNN", Neuroscience Informatics, Volume 2, Issue 3, September 2022, 100048, ISSN 2772-5286, <https://doi.org/10.1016/j.neuri.2022.100048>
- [17]. Suneet Gupta, Sumit Kumar, Sunil L. Bangare, Shibili Nuhmani, Arnold C. Alguno, Issah Abubakari Samori, "Homogeneous Decision Community Extraction Based on End-User Mental Behavior on Social Media", Computational Intelligence and Neuroscience, vol. 2022, Article ID 3490860, 9 pages, 2022. <https://doi.org/10.1155/2022/3490860>.
- [18]. Gururaj Awate, S. L. Bangare, G. Pradeepini and S. T. Patil, "Detection of Alzheimers Disease from MRI using Convolutional Neural Network with Tensorflow",arXiv, <https://doi.org/10.48550/arXiv.1806.10170>
- [19]. P. S. Bangare, S. L. Bangare, R. U. Yawle and S. T. Patil, "Detection of human feature in abandoned object with modern security alert system using Android Application," 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI), 2017, pp. 139-144, doi: 10.1109/ETIICT.2017.7977025
- [20]. Kalpana S. Thakare, Viraj Varale, "Prediction of Heart Disease using Machine Learning Algorithm", Bioscience Biotechnology Research Communications (Special issue) Volume 13, Issue 12, 2020 (Dec 2020 issue).
- [21]. Kalpana S. Thakare, A. M. Rajurkar, "Shot Boundary Detection of MPEG Video using Biorthogonal Wavelet Transform", International Journal of Pure and Applied Mathematics, Volume 118, No. 7, pp. 405-413, ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version), url: <http://www.ijpam.eu>

- [22]. Kalpana S. Thakare, A. M. Rajurkar, R. R. Manthalkar, “Video Partitioning and Secured Key frame Extraction of MPEG Video”, *Procedia Computer Science Journal*, Volume 78, pp 790-798, Elsevier, 2016. Scopus DOI: <http://10.1016/j.procs.2016.02.058>, www.sciencedirect.com/science/article/pii/S1877050916000600
- [23]. Kalpana S. Thakare, A. M. Rajurkar and R. R. Manthalkar, “Content based Video Retrieval using Latent Semantic Indexing and Color, Motion and Edge Features”, *International Journal of Computer Applications* 54(12):42-48, September 2012, Published by Foundation of Computer Science, New York, USA. DOI: 10.5120/8621-2486
- [24]. Kalpana S. Thakare, Archana M. Rajurkar, R. R. Manthalkar, “A Comprehensive System Based on Spatiotemporal Features Such as motion, Quantized Color and Edge Features”, *International Journal of Wireless and Microwave Technologies (IJWMT)* ISSN 1449 (Print), ISSN: 2076-9539 (Online), Vol.1, No.3, June. 2011, DOI: 10.5815 /ijwmt
- [25]. Kalpana S. Thakare, Archana M. Rajurkar, Dr. R. R. Manthalkar, “An effective CBVR system based on Motion, Quantized color and edge density features”, *International Journal of Computer Science & Information Technology (IJCSIT)*, ISSN 0975 – 3826, Vol 3, No 2, April 2011 DOI: 10.5121/ijcsit.2011.3206 78.
- [26]. M. L. Bangare, “Attribute Based Encryption And Data Integrity For Attack on Cloud Storage”, *Journal of Analysis and Computation (JAC)*, (An International Peer Reviewed Journal), www.ijaonline.com, ISSN 0973-2861, ICASETMP-2019, pp.1-4. <http://www.ijaonline.com/wp-content/uploads/2019/07/ICASETMP67.pdf>
- [27]. M. L. Bangare, Sarang A. Joshi, “Kernel interpolation-based technique for privacy protection of pluggable data in cloud computing”, *International Journal of Cloud Computing*, Volume 9, Issue 2-3, pp.355-374, Publisher Inderscience Publishers (IEL).
- [28]. M. L. Bangare, P. M. Bangare, R. S. Apare, S. L. Bangare, “FOG COMPUTING BASED SECURITY OF IOT APPLICATION”, *Design Engineering* ISSN: 0011-9342, Year 2021 Issue 7, pp.7542-7549.
- [29]. Rajesaheb R. Kadam and Manoj L. Bangare, “A survey on security issues and solutions in live virtual machine migration”, *International Journal of Advance Foundation and Research in Computer (IJAFRC)*, (December, 2012). ISSN (2014), pp.2348-4853.
- [30]. Sachindra K. Chavan, Manoj L. Bangare, “Secure Data Storage in Cloud Service using RC5 Algorithm”, *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN: 2277-3878, Volume-2, Issue-5 November 2013, pp.139-144.