ISSN: 2584-0495 Vol. 2, Issue 7, pp. 984-990

International Journal of Microsystems and IoT

ISSN: (Online) Journal homepage: https://www.ijmit.org

Global Air Pollution Prediction using Decision Tree

Ashish Kumar, Jeslin Shanthamalar J., and Nancy Noella RS

Cite as: Kumar, A., Shanthamalar J., J., & Nancy Noella, R. S. (2024). Global Air Pollution Prediction using Decision Tree. International Journal of Microsystems and IoT, 2(7), 984–990. https://doi.org/10.5281/zenodo.13293676

Published online: 17 July 2024

DOI: https://doi.org/10.5281/zenodo.13293676

International Journal of Microsystems and IoT
ISSN: 2584-0495; Vol. 2, Issue 7, pp. 984-990; DOI: https://doi.org/10.5281/zenodo.13293676

Check for updates

Global Air Pollution Prediction using Decision Tree

Ashish Kumar, Jeslin Shanthamalar J, and Nancy Noella RS

Department of Computer Science and Engineering, Sathyabama Institute of Technology, Chennai, Tamil Nadu, India

ABSTRACT
Air pollution is a worldwide environmental concern that significantly impacts the health and well-being of individuals across the globe. The impact of air pollutants on the health of human beings, ecosystems, and the climate has spurred significant concern and necessitated comprehensive analysis and understanding. This study provides an in-depth examination of global air pollution trends, sources, and their consequences, drawing from a vast and diverse dataset. The "Global Air Pollution Dataset" serves as the foundation for this analysis, offering a wealth of information that encompasses air quality measurements, emissions data, meteorological variables, and more. The dataset combines historical records with real-time updates, providing a rich resource for researchers, policymakers, and environmental enthusiasts. The proposed model achieved 98% in terms of accuracy using Decision Tree Algorithm to analyze the Pollutants as a public health concern that include particulate matter, carbon monoxide, nitrogen dioxide and ozone.

KEYWORDS
Machine Learning, Deep Learning, Decision Tree Algorithm, Random Forest

1. INTRODUCTION

Air pollution is a vast global environmental challenge with consequences for public health, ecosystems, and climate change. The presence of pollutants such as particulate matter, nitrogen oxides, carbon monoxide, and organic compounds in the atmosphere has led to a host of adverse effects, including respiratory diseases, reduced air quality, and altered weather patterns[1]-]10].

The issue of air pollution has emerged as one of the most environmental challenges confronting our planet. The degradation of air quality due to pollutants emitted from various human activities has consequences for both human health and the environment. From densely populated urban centers to remote rural areas, no corner of the globe remains untouched by the pervasive impacts. Due to this pollution nowadays in the clear dark sky at night, we can't see our Milky Way galaxy, while in our history looking back, our eyes can see the Milky Way galaxy without any equipment it didn't end here animals and insects like beetles used to travel in the night sky straight without any hesitation. They were following a galaxy-white strip-like structure. These days we can't see because of pollution and it causes some myopia disease which affects human eyes as well as animals[11]-[15].

The air pollution is multifaceted and complex, encompassing a spectrum of health effects, environmental degradation, and socioeconomic implications. Exposure to polluted air is linked or we can say that lead to health problems, ranging from respiratory diseases and lungs cancer to neurological damages.

Air pollution exacts a toll on ecosystems, disrupting ecological processes and biodiversity.

Acid rain, a consequence of air pollution caused by sulfur dioxide and nitrogen oxides emissions, poses a threat to forests, freshwater bodies, and aquatic life. Moreover, airborne pollutants can accumulate in soil and water, entering food chains and posing risks to wildlife and human populations.

The global challenge of air pollution requires concerted efforts at local, national, and international levels, encompassing policy interventions, technological innovations, and behavioral changes. Strategies to mitigate air pollution encompass a range of measures, including the adoption of cleaner energy sources, improved vehicle emissions standards, enhanced industrial pollution control, and sustainable land management practices.

2. RELATED WORK

Doreswamy et al. investigated the ML predictive models for forecasting Particulate Matter that must be pollutants in the air. But according to the analysis they only went to Taiwan for their analysis. Madan T et al. Air pollution is very dangerous and should be decreasing as fast as they could. These are the things they must keep in their mind for prediction or for analysis. Madhuri VM et al. The presence of air pollutants in air is accessed. These can also be able to predict by using probability and statistics. So, they used some of these algorithms and worked on that. C. Wu et al. To improve the air quality, He was trying to show that air pollution has gone through several stages in the evolution of single pollutants into composite pollutants which had been caused by the existence of many or several pollutants. Bhalgat P et al. Work as Monitoring and ensuring that good air quality has become a top priority for governments in urban areas. global warming, acid rain, or an increased number of patients.

G. K. Kang et al. Observing and maintaining air quality has become important in many places. The quality of air has been affected because of pollution. Y. Chen et al. Inhaling polluted air for a severe long time can lead to many physical and mental problems. K. Kujaroentavon et al. The goal of doing work is to keep rules based on the quality of air. The results of this study were correctly classified into training. W. Bin In The Air Quality Index of capitals, their cities to find the distribution across all cities. They used to know about the data by using some of the graphs and analyzing them. Veljanovska and A. Dimoski They found that the pollution Affecting human's respiratory systems is a vast cause for increase in death rate and shortened lifespan and increased the risk of spreading diseases.

G. Parthasarathy and B. N. Chatterji It involves determining the density of observed data points with k as the number of the neighbors. K. H. Brodersen et al. It is used in the performance on information retrieval or checks the overall performance. On the basis of accuracy on two limitations. F. A. Gers as they have been trained in their neural network to predict a clean speech if there is some noise available in their background, so they used a noise feature. The first method is Phase error-based filtering which performs based on time difference. The second method is Correlation shaping. M. Zhao and X. Li as this is based on the air quality but they have distributed it in a way of a smart city. By making a weather list system it will improve overall understanding of the environment.

V. M. Niharika and P. S. Rao Present in the atmosphere in a way that can harm humans, plants or animals, and the environment will be damaged by those substances, and it will show the effect on the weather or on the sudden climatic change. D. Mishra as used the Neuro-Fuzzy model at that time so for the air quality forecasting they chose to deal with the Nitrogen dioxide (NO2) this is the pollutant that had been chosen for the analysis. T. Lan et al. as they decided to work on the disease which is called as a spread, they used to store the data in the format of ionograms so they decided to train and process. Xiaosong Zhao et al. as they used deep learning to go for the prediction and then for classification, they decided to use a Support machine, Random forest, and the RNN model used in three different industrial places[16]-[20].

Venkat-Rao Pasupuleti et al. as it used to analyze the air quality, so they used the air quality Index for only India to and used a useful approach for resolving the quality analysis of air. Soubhik Mahanta et al. decided to do this prediction because they don't want any of the people to face danger or suffer from the pollution. Zhili Zhao et al. They decided to check for every hour, which means they will show the hourly prediction which means they have a huge impact on weather or pollution. Kostandina Veljanovska and Angel Dimovski as I understand in his paperwork he is using AI tools so that it can make decisions on its own not by a programmer that will feed all the data and say it to show the values. Kumar et al. The findings of the research offer practical insights for enhancing air quality monitoring and management strategies, showcasing their relevance in real-world scenarios.

3. DESCRIPTION

The Title of the dataset that is used in the proposed work is Global Air Pollution. The proposed work uses a dataset that contains 23463 instances with 12 attributes. These Columns are Country, City, AQI (Air Quality Index) Value, AQI Category, Carbon Monoxide gas, Ozone, Nitrogen Dioxide gas, and Particulate Matter. These are so harmful if you make contact between them.

Describing the proposed work of the program as histogram, Bar Graph, Regression Plot, Box plot, and Heat Map is used. By talking about Categories it depend on AQI conditions AQI is good condition, Moderate condition, Unhealthy for the sensitive groups of situation, Very Unhealthy, or Unhealthy, or Hazardous disease or situation. By using a heat map, we can verify the level of air pollution is present in that place or the country. It will help in predicting air pollution according to the data provided. Also by using Hierarchical Clustering to show the air quality means air pollution level according to country based.

By using median and mode values of AQI taken for prediction here. Also, by using a Heat map here we separated which gasses are polluting which of the specific places we like to take first from the proposed system by taking Ozone to see which places are polluted in what ways like good, moderate, Unhealthy or Hazardous.

4. IMPLEMENTATION

This paper presents a technique for global air pollution. Structural changes in the architecture of global air pollution prediction. As part of data collection this data is collected based on countries and cities and some of the pollutant gasses and their AQI Values with their Category. The name of the selected gasses which proposed work taken is:

i. Carbon Monoxide gas (CO)
ii. Ozone layer (O3)
iii. Nitrogen Dioxide gas (NO2)
iv. Particulate Matter (PM2.5)

Before Starting every city has its own AQI values and their Categories but here to say that the proposed work to see how it will affect the pollution level in the country. AQI Values of every gasses plays a crucial role in the prediction.

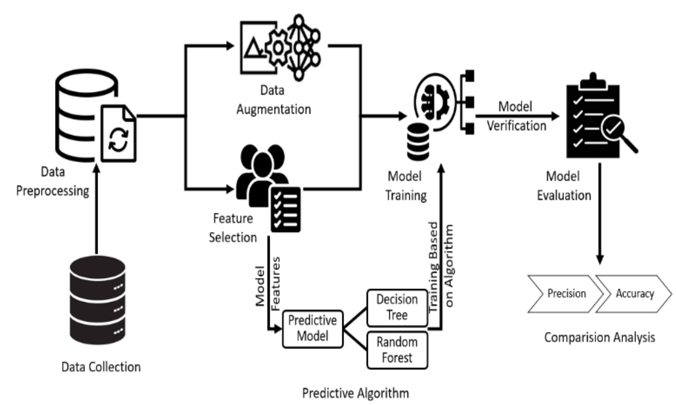

Fig. 1 Architecture Diagram

Carbon Monoxide [CO]

Carbon monoxide (CO) is a gas that lacks color and odor. It is emitted by sources like kerosene and gas space heaters, along with gas stoves, which can degrade indoor air quality. exposure to CO can lead to symptoms such as dizziness, loss of consciousness, and even death[22].

Ozone [O3]

Ground-level ozone, produced through chemical reactions involving nitrogen oxides and specific organic compounds, presents health hazards and serves detrimental purposes. It can diminish lung function and exacerbate conditions like asthma. It damages vegetation during the growing season.

Nitrogen Dioxide [NO2]

Nitrogen Dioxide (NO2) is classified as one of the nitrogen oxides. While it can naturally enter the air from the stratosphere, at ground level, NO2 primarily originates from emissions from cars, trucks, buses, and power plants.

Particulate Matter [PM2.5]

Atmospheric Particulate Matter consists of intricate combinations of small solid particles as well as liquid particles that are in the air. Inhalation of these particles can lead to severe heart and lung issues.

4.1 Data Processing

First we will divide all the numeric values and non-numeric values. In numeric values these aqi_value, co_aqi_value as a numeric value, ozone_aqi_value, no2_aqi_value there wherever is value those belong to numerical categories and pm2_5_aqi_values. For non-numeric values are aqi_category, co_aqi_category these are the one who doesn't contains numeric, ozone_aqi_category, no2_aqi_category these only contains object and pm2_5_aqi_category now here the density of values in each numeric column is here. So, these are the densities in each of the numeric columns.

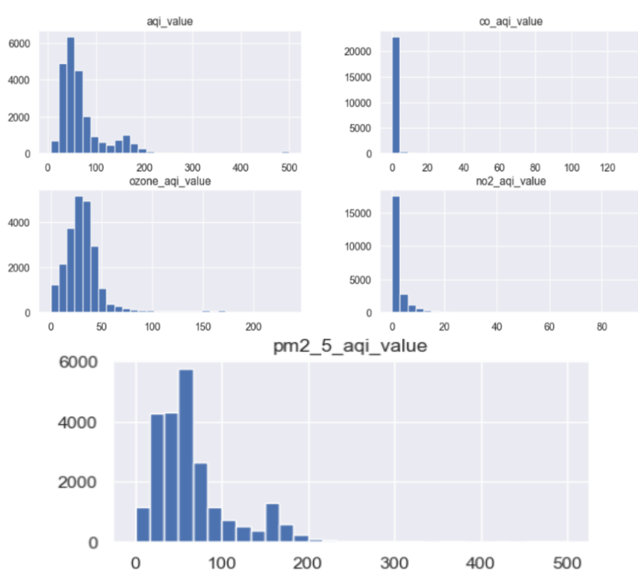

Fig. 2 Density of Polluted Air

4.2 Feature Selection and Data Augmentation

Count	176.000000
Mean	133.312500
std	355.508745
min	1.000000
25%	10.750000
50%	28.500000
75%	81.000000
max	2872.000000

Fig. 3 Country Details

For the feature selection proposed work have taken the country column as in Fig.3, so now by describe the country we get the data about the country.

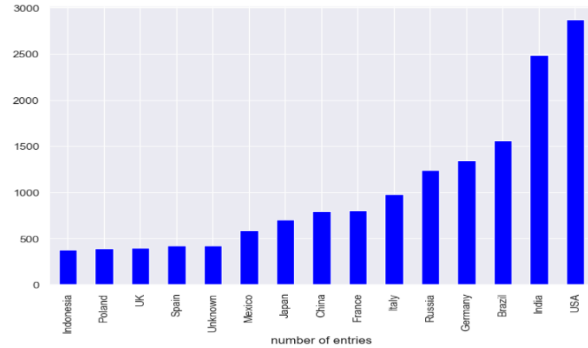

Fig. 4 Well Represented Country

Now, for the proposed work going to check the values present in the or we can say the name of the country in Fig.4 which is greater or larger we have to reduce the size of the name of the country as in short form. Here the proposed work going to show the top or most well represented country that are present in the data is

Proposed system had 23462 locations in 175 countries. There is too much information to fit onto the screen without dividing it into smaller pieces, focusing on something more particular or doing something else about it. To do something about it, first, work took a more detailed look at the most represented countries within the dataset. It shows the well represented details of the country regarding different air pollution measurements. In order to do so, the proposed work used heat maps to show the air's quality index of countries.

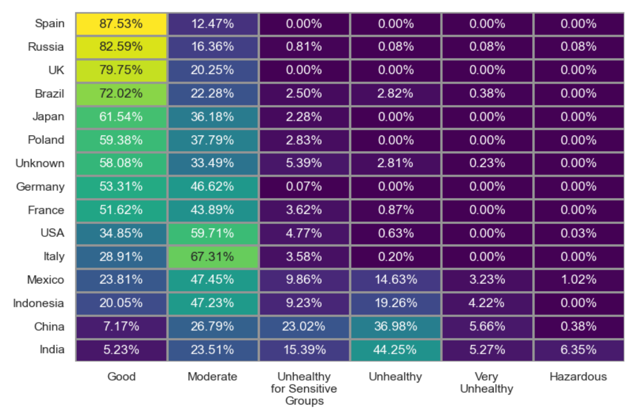

Fig. 5 Air Quality Index of most Risky Countries

Here in Fig.5 in this table everyone can see that there is Quality of air where and in which country how much pollutants are there. This one here only shows the Air Quality index of the country without based on the different gasses. Like this there are more heat maps proposed work have created and those all are based on the different gasses and countries. Each specific gas has their own heat maps based on the pollution level. We can clearly see that the condition is good, Moderate, Unhealthy, Hazardous all the data have been distributed among the others.

Here we can easily evaluate the data which is present in the dataset like here in the proposed work used boxplot to show the air quality index by using AQI Value is

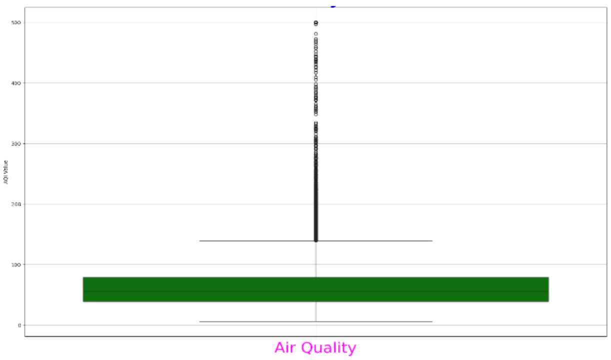

Fig. 6 Box Plot for Air Quality Index with Outliers

Here we saw that there are many outliers in the box plot so here we can't be able to get the accurate measurement or clear distribution of the Quality of air. So, by using the Interquartile range proposed work fixed the issue. Here proposed work used the formula for removing the outliers which is as follows:

First Quartile Q1 (25th percentile) = 39.0
Median Q2 (50th percentile) = 55.0
Third Quartile Q3 (75th percentile) = 79.0

IQR = Q3 - Q1 = qn(0.75) - qn(0.25)
79.0 - 39.0 = 40.0

The upper quartile (Q3) spans a distance equal to 1.5 times the interquartile range (IQR) from the third quartile. This measurement extends to the highest recorded data point within this interval from the dataset. so,

< Q1 - 1.5 * IQR
39.0 - 1.5 * 40.0 = -21.0

A distance equal to 1.5 times the interquartile range (IQR) is subtracted from the lower quartile (Q1). This distance is then extended downward to the lowest observed data point within this range from the dataset.

> Q3 + 1.5 * IQR
79.0 + 1.5 * 40.0 = 139

So, after using that formula for the boxplot we get the clarity for the Air Quality index value.

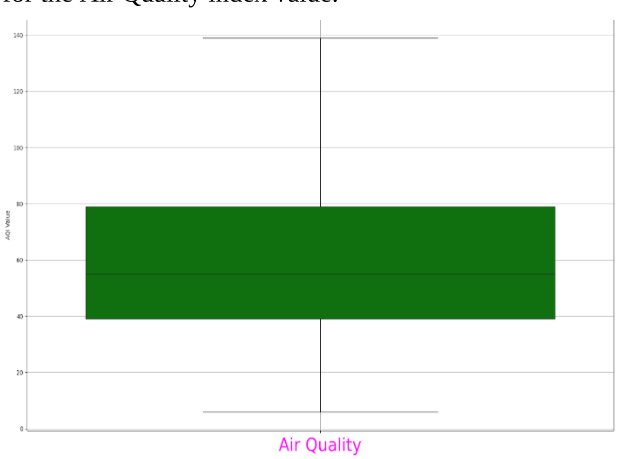

Fig. 7 Box Plot for Air Quality Index without Outliers

4.3 Model Training

Here in this Global air pollution proposed work there used the dataset to train models of air pollution. By using sklearn.model_selection proposed work used the dataset to train and test the model by splitting it into several parts. At first it took 20% percent of the data to train and then test like we can say that here in the proposed work used algorithms to evaluate the data like Decision tree algorithm.

Prediction / Actual	Good	Hazardous	Moderate	Unhealthy	Unhealthy for Sensitive Groups	Very Unhealthy
Good	8166	0	0	0	0	0
Hazardous	0	138	0	0	0	0
Moderate	0	0	7260	0	0	0
Unhealthy	0	0	0	1703	0	0
Unhealthy for Sensitive Groups	0	0	0	0	1299	0
Very Unhealthy	0	0	0	0	0	204

Fig. 8 Training Data Result

Prediction / Actual	Good	Hazardous	Moderate	Unhealthy	Unhealthy for Sensitive Groups	Very Unhealthy
Good	2042	0	0	0	0	0
Hazardous	0	34	0	0	0	0
Moderate	0	0	1815	0	0	0
Unhealthy	0	0	0	426	0	0
Unhealthy for Sensitive Groups	0	0	0	0	325	0
Very Unhealthy	0	0	0	0	0	51

Fig. 9 Testing Data Result

Decision Rule: if Xi ≤T, go to the left child; otherwise, go to right child

Entropy: Entropy(S) = −n∑i=1 P(xi)log2, P(xi)

Gain(S, Attribute) = Entropy(S) - ∑ |Sv|/|S| (Entropy(Sv))
where,
v = {attribute values)

On the Gain value we can get the arrangement of a decision tree algorithm.

By splitting I got the values as
{'min_samples_split': 27, 'min_samples_leaf': 46, 'max_leaf_nodes': 40, 'max_depth': 10, 'criterion': 'entropy'}

By using these splitting over decision tree proposed work got the value as for testing data of these 20% training the data proposed work got,

Prediction / Actual	Good	Hazardous	Moderate	Unhealthy	Unhealthy for Sensitive Groups	Very Unhealthy
Good	2039	0	3	0	0	0
Hazardous	0	19	0	15	0	0
Moderate	15	0	1793	0	7	0
Unhealthy	0	0	0	421	5	0
Unhealthy for Sensitive Groups	0	0	0	9	315	1
Very Unhealthy	0	2	0	9	0	40

Fig. 10 Testing Data Result

The accuracy of the Decision tree algorithm is approximately 98%. After using Randomized algorithm and Decision Tree Algorithm proposed work found that Decision Tree Algorithm is the best algorithm for the dataset[23]-[35].

4.4 Model Training

In the proposed work that explored all the gasses and on the basis of the good, moderate, unhealthy, very unhealthy, hazardous these are the conditions of the gasses all over the places globally.

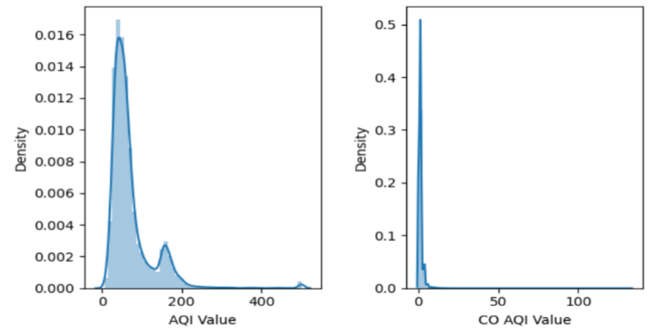

Fig. 11 Values based on polluted gasses

In the first view we can see that AQI Value is given based on a country and city. The order to analyze the graph is 'Moderate', 'Good', 'Very Unhealthy', 'Unhealthy for Sensitive Groups', 'Unhealthy', 'Hazardous'.

In the second one we can clearly analyse that this AQI value is dependent on the Carbon Monoxide gas in the country and its order to see the gasses is 'Good', 'Unhealthy for Sensitive Groups', 'Moderate'.

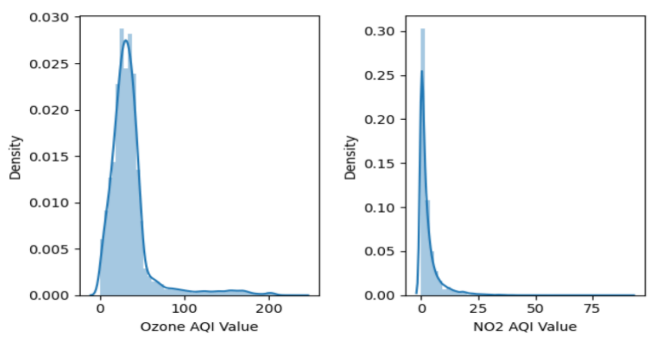

Fig. 12 Values based on the Ozone and Nitrogen Dioxide gasses

In Fig.12 we can clearly analyse that this AQI value is dependent on the Ozone gas in the country and its order to see the gasses is 'Good', 'Moderate', these are normal 'Unhealthy for Sensitive Groups', 'Unhealthy', these can go to severe 'Very Unhealthy', if unhealthy is there then big or serious issues [27].

In Fig.12 We can clearly analyse that this AQI value is dependent on the Nitrogen Dioxide gas in the country and its order to see the gasses is 'Good', 'Moderate'. There is no large amount of Pollution caused by NO2.

In Fig.13 we can clearly analyse that this AQI value is depend on the Particulate Matter 2.5 gas in the country and its order to see the gasses is 'Moderate', 'Good', 'Unhealthy for Sensitive Groups', 'Very Unhealthy', 'Unhealthy', 'Hazardous'.

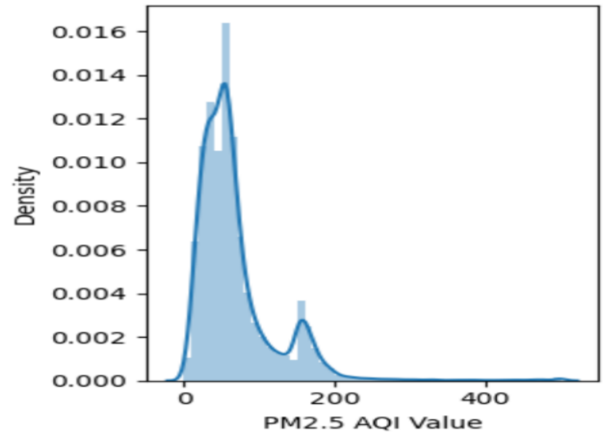

Fig. 13 Values based on Particulate Matter

Table. 1 Cross platform comparative performance

Algorithms	Accuracy
Random Forest	82.35%
Decision Tree	98.59%

So, accuracy of the Decision Tree Algorithm is more than the Random Forest Algorithm so here the proposed work preferred to use the Decision Tree Algorithm for getting the best suitable Accuracy[26].

5. COMPARATIVE ANALYSIS

Here the performance measures of existing works are given below:

Table. 1 Cross platform comparative performance

SI	Existing Method	Algorithm	Accuracy
1	K. Hu (2017)	SVR	64.2%
2	Madhuri VM (2020)	SVM	75%
3	Venkat Rao Pasupuleti (2020)	Random Forest	79%
4	Soubhik Mahanta (2019)	Extra Trees	85.3%
5	Y. Wang (2019)	BP Neural Network	91%
6	Kumar. (2023)	KNN	89%
7	Proposed Work	Decision Tree	98%

The proposed system got 98% accuracy by using Decision tree Algorithm which is quite higher than existing work.

Pollution Estimation from Fixed and Mobile Sensors K. Hu et al. The effectiveness of the system may be constrained to Sydney or comparable urban settings, potentially diminishing its performance when applied to diverse geographical areas. Madhuri VM, Samyama GGH complex algorithms like Random Forest may be prone to overfitting, meaning they could excessively fit the training data and perform poorly Venkat Rao Pasupuleti et al. Utilizing sensors on the Arduino Uno platform might restrict the monitoring system's coverage area, potentially resulting in incomplete data collection over a broader geographical expanse. This limitation could compromise the system's ability to capture a comprehensive picture of air quality across a large area. Soubhik Mahanta et al. Overfitting arises when a model becomes too closely attuned to the intricacies of the training data, resulting in diminished performance when applied to new or unseen data.

Y. Wang, T. Kong. The comparison of the improved algorithm to other classification prediction methods might lack comprehensiveness.

CONCLUSIONS

In conclusion, predicting global air pollution is a crucial endeavour that requires ongoing efforts and advancements in technology, data analysis, and international collaboration. As our world continues to industrialize and urbanize, the challenges associated with air quality are likely to take proactive measures to manage the impacts on human health, ecosystems, and climate. The proposed method predicted the air pollution with 98% accuracy using a decision tree algorithm Public awareness and engagement are essential components of any successful air pollution prediction and control strategy. Educating communities as to the health risks associated with the poor air quality. In conclusion, the ongoing commitment to research, technological innovation, global collaboration, and public involvement is vital for developing robust predictive models and effective strategies to decrease the impacts of air pollution on a global scale.

The mainstream here is that the proposed work is on collecting all the data of countries from different place and working on that collected data to analyse and based on that details we can tell the level of pollution, and it will have to be scalable to take the new data regularly.

REFERENCES

1. Doreswamy HKS, Yogesh KM, Gad I (2020). Forecasting Air pollution particulate matter (PM2.5) using machine learning which has been based on regression models. Procedia Comput Sci 171:2057–2066.
2. Madan T, Sagar S, Virmani D. (2020). Air quality prediction using machine learning algorithms. In: 2nd international conference on advances in computing, communication control and networking (ICACCCT),140–145.
3. K. Hu, Rahman, A., Bhrugubanda, H., and Sivaraman, V. (2017). Haze Est: Machine Learning Based Metropolitan Air Pollution Estimation from Fixed and Mobile Sensors. IEEE Sensors Journal. 17(11), 3517-3525. doi: 10.1109/JSEN.2017.2690975
4. Madhuri VM, Samyama GGH, Kamalapurkar S (2020). Air pollution prediction using machine learning and supervised learning approach. Int J Sci Technol Res 9(4),118–123.
5. Madhuri VM, Samyama GGH, Kamalapurkar S (2020). Air pollution prediction using machine learning and supervised learning approach. Int J Sci Technol Res 9(4),118–123.
6. Wu, C., Hu, W., Zhou, M., Li, S., and Jia, Y.(2019). Data-driven regionalization for analyzing the spatiotemporal characteristics of air quality in China. Atmos. Environ. 203(172–182).
7. Bhalgat P, Bhoite S, Pitare S (2019). Air Quality Prediction using Machine Learning Algorithms. Int J Comput Appl Technol Res 8(9):367–370.
8. Kang, G. K., Gao, J. Z., Chiao, S., Lu,V., and Xie, G.(2018). Air quality prediction: Big data and machine learning approaches. Int. J. Environ. Sci. Develop. 9(1), 8–16.
9. Chen, Y., Wang, L., Li, F., Du, B., Choo, K.-K. R., Hassan, H., and Qin, W. (2016). Air quality data clustering using EPLS method. Inf. Fusion, 36(225-232).
10. Kujaroentavon, K., Kiattisin, S., Leelasantitham, A., and Thammaboosadee, S.,(2014). Air quality classification in Thailand based on decision tree. Proc. 7th Biomed. Eng. Int. Conf., Fukuoka, Japan. 26–28.
11. Bin, W. (2008). Analyze the spatial-temporal characteristics of air pollution in China by using air pollution index (API). M.S. thesis, Ocean Univ. of China, Qingdao, China.
12. Ramani, R.G., Shanthamalar, J. J.(2022) Automated image quality appraisal through partial least squares discriminant analysis. Int J Comput Assist Radiol Surg. 17(7):1367-1377.
13. Veljanovska, K., and Dimoski, A. (2018). Air quality index prediction using simple machine learning algorithms. Int. J. Emerg. Trends Technol. Comput. Sci. 7(1), 25–30.
14. Parthasarathy, G., and Chatterji, B. N.(1990). A class of new KNN methods for low sample problems. IEEE Trans. Syst., Man, Cybern., 20(3), 715–718. doi: 10.1109/21.57285.
15. Brodersen, K. H., Ong, C. S., Stephan, K. E., and Buhmann, J. M.(2010). The binormal assumption on precision-recall curves. Proc. Int. Conf. Pattern Recognit., Istanbul, Turkey. 4263–4266.
16. Gers, F A.(2001). Long short-term memory in recurrent neural networks,''M.S. thesis, Dept. Comput. Sci., Swiss Federal Inst. Technol., Lausanne,EPFL, Switzerland.
17. Zhao, M. and Li, X.(2011). An application of spatial decision tree for classification of air pollution index. Proc. 19th Int. Conf. Geoinform., Shanghai,China. 1–6.
18. Niharika, V. M., and Rao, P. S. (2014). A survey on air quality forecasting techniques. Int. J. Comput. Sci. Inf. Technol. 5(1), 103–107.
19. Mishra, D. (2016). Neuro-fuzzy approach to forecast NO2 pollutants addressed to air quality dispersion model over Delhi, India. Aerosol Air Qual.16(1), 166–174.
20. Lan, T., Zhang, Y., Jiang, C., Yang, G., and Zhao, Z. (2018). Automatic identification of spread F using decision trees. J. Atmos. Sol.-Terr. Phys.,179(1), 389–395.
21. Zhao, X., Zhang, R. Wu, J.L., Chang,P.C., and Ze, Y. (2018). A Deep Recurrent Neural Network for Air Quality Classification.
22. Pasupuleti, V.R., Kalyan, P., and Reddy, H.K. (2020). Air Quality Prediction of Data Log By Machine Learning. IEEE.
23. Soubhik Mahanta, T. Ramakrishnudu, Rajat Raj Jha and Niraj Tailor, Urban Air Quality Prediction Using Regression Analysis,2019, IEEE
24. Zhao, Z., Qin, J., He, Z., Li,H., Yang,Y., and Zhang,R. (2020). Combining forward with recurrent neural networks for hourly air quality prediction in Northwest of China, Environmental Science and Pollution Research.
25. Veljanovska, K., Dimoski, A.(2023). Air Quality Index Prediction Using Simple Machine Learning Algorithms,2018, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS).
26. Kumar, K., Pande, B.P.(2023). Air pollution prediction with machine learning: a case study of Indian cities. Int. J. Environ. Sci. Technol. 20, 5333–5348.
27. Wang, Y., and Kong, T. (2019). Air Quality Predictive Modeling Based on an Improved Decision Tree in a Weather-Smart Grid. IEEE Access, 7, 172892-172901, doi: 10.1109/ACCESS.2019.2956599.

AUTHORS:

Ashish Kumar is currently pursuing his B.E. Degree in Computer Science and Engineering from Sathyabama Institute of Science and Technology, Chennai. His areas of interests are Distributed Systems, Cloud Computing, System Security and Machine Learning

Corresponding Author E-mail: ashishkumar17012002@gmail.com

Dr. J. Jeslin Shanthamalar is currently working as Assistant Professor in the Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai. She completed her Ph.D. programme in the Department of Information Science and Technology, Anna University, Chennai, India, as a full-time scholar. She has more than 10 years of experience in Teaching, Research and Industry. Her areas of specialization include Data Mining, Image Processing, Machine Learning, Deep Learning and Medical Image

Email: jeslin.cse@sathyabama.ac.in

Dr. Nancy Noella R S, is currently working as an Assistant Professor in the Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai. She completed her Ph.D. program in the School of Computer Science and Engineering, VIT Chennai, India as a full scholar. She has more than 8 years of experience in Teaching and Research. Her areas of interest are Artificial Intelligence, Image Processing, Machine Learning, and Deep Learning. She has published more than 10 papers in reputed journals, patents, and conferences

Email: nancynoella.cse@sathyabama.ac.in