



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Innovative Machine Learning for Air Pollution Prediction

Dr.Akula Suneetha¹, M. Kiran Kumar², P. Karthik³, M. Raghu Narayana⁴, G. Prakash Babu⁵

Associate Professor, Department of CSE, KKR & KSR Institute of Technology, Guntur – Andhra Pradesh, India¹

B. Tech Student, Department of CSE, KKR & KSR Institute of Technology, Guntur – Andhra Pradesh, India^{2,3,4,5}

ABSTRACT: The promise of air pollution prediction is in its ability to completely change how we address and reduce air pollution. Urban planning, environmental sustainability, and public health all depend on timely and accurate air quality forecasts. Authorities can take proactive steps to reduce exposure risks and efficiently manage pollution sources by supplying accurate forecasts. The model aims to forecast pollution levels with high accuracy and temporal granularity by analysing historical air quality data, meteorological parameters, and emission sources using sophisticated algorithms. Regression and time-series analysis are two machine learning techniques that can be integrated to create a dynamic and adaptive prediction model that can capture intricate relationships between a variety of variables.

KEYWORDS: Air Quality Index (AQI), Web Scrapping, Support Vector Machine, Logistic Regression, Decision Tree Classification, Pycert.

I. INTRODUCTION

The promise of air pollution prediction is in its ability to completely change how we address and reduce air pollution. Urban planning, environmental sustainability, and public health all depend on timely and accurate air quality forecasts. Authorities can take proactive steps to reduce exposure risks and efficiently manage pollution sources by supplying accurate forecasts. The project's main goal is to divide air quality into three categories using classification algorithms: poor, medium, and high. By giving the public and policymakers a more intuitive understanding of pollution levels, these classifications help them make more informed decisions. The predictions are made more precise by including PM2.5 and PM10 levels, which provide information about the fine particulate matter that has a major influence on respiratory health.

PyCaret is a potent Python machine learning automation library, and the project incorporates it to improve the prediction model's efficiency and accuracy. PyCaret accelerates the process of testing various algorithms and fine-tuning hyperparameters by streamlining the entire machine learning workflow. This guarantees that the prediction model stays responsive and adaptive to the changing nature of air quality data while also speeding up the model's development. With the integration of these developments, the project makes a valuable contribution to environmental science and public health by providing an advanced air quality prediction tool. Real-time air quality classification can reduce exposure risks and promote a healthier, more sustainable living environment by enabling individuals, local government agencies, and industry to act promptly.

With PyCaret automation integrated, this project presents a strong air quality level classification system with improved precision. Through the integration of PM2.5 and PM10 parameters, the model offers refined perspectives on the effects on respiratory health. The optimized process guarantees flexibility in response to changing air quality information, enabling people and organizations to take prompt action.

II. LITERATURE SURVEY

[1] Rishanti Murugan, Naveen Palanichamy, "Smart City Air Quality Prediction using Machine Learning", Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021). Air pollution is a leading challenge to deal with globally, especially in smart cities. Air pollution explicates as any material from any atmospheric source that exists in any form, either liquid, solid, or gas phase, causing destruction or the capability to alter the usual characteristics of the atmosphere and increasing the health risk to living things or causing the environment and ecosystem to be unbalanced

[2] Shreyas Simu, Varsha Turkar, Rohit Martires, Vranda Asolkar, Swizel Monteiro*, Vaylon Fernandes, "Air Pollution Prediction using Machine Learning", IEEE 2020. The primary objective of this research is to monitor the air pollutants emitted from an industry/factory and predict the future spread of these pollutants also to output these results in the form of reports for industries/factories. To accomplish this work, two objectives were identified that are: To develop a module that uses Machine Learning (ML) models with pollution data to predict future emission parameter values, to create a module that simulates the movement of fluid particles (pollutants) in the air using air dispersion models with meteorological data.

[3] Tanisha Madan, Shraddha Sagar, Deepali Virmani, "Air Quality Prediction using Machine Learning Algorithms –A Review" 2020 2nd International Conference on Advances in Computing, Communication Control and Networking. Air Pollution can cause severe disease like lungs cancer, brain disease and even lead to death. Machine learning algorithms helps in determining the air quality index. Various research is being done in this field but still results are still not accurate. Dataset are available from Kaggle, air quality monitoring sites and divided into two Training and Testing. Machine Learning algorithms employed for this are Linear Regression, Decision Tree, Random Forest, Artificial Neural Network Support Vector Machine

[4] Sriram Krishna Yarragunta, Mohammed Abdul Nabi, "Prediction of Air Pollutants Using Supervised Machine Learning ", IEEE 2021. Air Quality Index is a device used for determining the air quality and the causes of air pollution. Air pollution is widely regarded as a concern in India. The highest possible values that are considered fit for public medical services are relatively higher than the daily air quality quantified attributes. The situation is particularly bad in larger cities like Delhi, where the AQI has reached an all-time high of 999.

[5] Chavi Srivastava, Shyamli Singh "Estimation of Air Pollution in Delhi Using Machine Learning Techniques", 2018 International Conference on Computing, Power and Communication Technologies (GUCON) In this paper, we implemented different classification and regression techniques like Linear Regression, SDG Regression, Random Forest Regression, Decision Tree Regression, Support Vector Regression, Artificial Neural Networks, Gradient Boosting Regression and Adaptive Boosting Regression to forecast the Air Quality Index of major pollutants like PM_{2.5}, PM₁₀, CO, NO₂, SO₂ and O₃. The techniques are then evaluated using Mean square error, Mean absolute error and R², which show that Support Vector Regression and Artificial Neural Networks are best suited for predicting the air quality in New Delhi

[6] Jayant Kumar Singh, Amit Kumar Goel, "Prediction of Air Pollution by using Machine Learning Algorithm", 2021 7th International Conference on Advanced Computing & Communication Systems (ICACCS) The Air Pollution Forecasting System: Air Quality Index (AQI) is a record that gives the public the degree of contamination related with its wellbeing impacts. The AQI centers around the different wellbeing impacts that individuals may encounter dependent fair and square and long stretches of introduction to the poison concentration. The AQI values are not quite the same as nation to nation dependent on the air quality norm of the country. The higher the AQI level more noteworthy is the danger of wellbeing related problems. The by and large point of this venture is to make a student calculation that will have the option to foresee the hourly contamination focus. Additionally, an Android application will be built up that will provide the clients about the constant contamination convergence of PM_{2.5} alongside the hourly forecasted value of the toxin fixation from the student calculation. The Android application will also recommend data of the less dirtied.

[7] Rohit Adke, Suyog Bachhav, "Air Pollution Prediction using Machine Learning", Volume: 06 Issue: 10 | Oct 2019, AS per World Health Organization [WHO] air pollution is infectivity of the indoor or outdoor environment by any chemical and biological agent which changes characteristics of the environment. Industrial Emission: As the new technologies are evolving day by day so new factories are being installed. This has caused air pollution by emitting harmful smoke, gases etc.

[8] Pakki Pavan Sai, Raghupathi RajaSekhar, "Prediction of Air Quality Index by Using Machine Learning", Volume 7 ~ Issue 11 (2022) pp: 01-07 It is a contamination of air due to presence of substance in atmosphere that are harmful to health of humans and other living beings and cause damage to climate materials. Not just humans are harmed by air pollution; many animals, forests, crops are as well. Therefore, a project was created to analyse the air quality using various machine learning approaches. There are different methods in machine learning technique. It can be used for appropriate data analysis, air quality prediction and air quality determination. "AQI" Air quality index is a number that indicates how polluted the air is at given time and given place. There are many pollutants that cause air pollution

[9] Shailesh Munge Sagar Kharche, "Air Quality Prediction based on Machine Learning", Volume 9, Issue 7 July 2021 | ISSN: 2320-2882 One major basic right is clean air that is integral to the concept of citizenship and it's while not a doubt, the responsibility of every subject to try to do his/her half to stay the air clean. Air quality prognostication has been looked into because the key answer of early warning and management of pollution. The Air pollutants data is retrieved from the sensors which square measure processed during a unified schema and hold on as a dataset.

[10] N. Srinivasa Gupta, Yashvi Mohta, "Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis", Volume 2023, Article ID 4916267, 26 pages

Humans can only survive because of air. Its quality must be monitored and understood for our wellbeing. An index for reporting air quality is called the air quality index (AQI). It measures the impact of air pollution on a person's health over a short period of time. The purpose of the AQI is to educate the public on the negative health effects of local air pollution.

[11] Miss. S. Jayasri, Mrs. R. Vijayalakshmi, "Analysis of Real-Time Fine Particulate Matter (PM2.5) Concentrations for Air Pollution Using Machine Learning" Vol 4, no 6, pp 3393-3400 June 2023 By predicting the air quality index for a certain place, machine learning is used to forecast India's air quality. The air quality index of India is a commonly used indicator of pollution levels (so₂, no₂, rspm, spm, etc.) through time. The project may be driven by the growing concern over air pollution and its detrimental effects on public health and the environment. Initiating this project reflects a commitment to addressing and mitigating air pollution issues.

III. METHODOLOGIES

Support Vector Machine (SVM) is a powerful classification algorithm that can also be used for regression or outlier detection. The classifier version is referred to as Support Vector Classifier (SVC). Effective in high-dimensional spaces, robust to overfitting, and can work well in situations where the number of features is greater than the number of samples.

Logistic regression is a linear model for binary classification. It estimates the probability that a given instance belongs to a particular category. Simple, efficient, and works well when the relationship between features and the log-odds of the outcome is approximately linear.

K-Nearest Neighbors (KNN) is a non-parametric and lazy learning algorithm. It classifies a data point based on the majority class of its k nearest neighbors. Simple and effective, works well for small datasets.

A decision tree is a flowchart-like tree structure where each node represents a decision or test on a feature, each branch represents the outcome of the test, and each leaf node represents a class label. Easy to understand and interpret, requires little data preprocessing.

PyCaret is designed to help data scientists and machine learning practitioners with various tasks, such as data preprocessing, feature engineering, model selection, hyperparameter tuning, and model evaluation. PyCaret aims to automate many of the repetitive tasks in the machine learning process

IV. SYSTEM ARCHITECTURE

The system starts by gathering real-time data from multiple sources, including environmental monitoring stations, weather stations, and online databases, on air quality parameters like AQI, PM_{2.5}, PM₁₀, temperature, and humidity. To clean and get it ready for analysis, the gathered data is pre-processed. This include addressing missing values, data normalization, and feature extraction that is pertinent. With the pre-processed data, the system trains classification algorithms like Random Forest, Decision Trees, Support Vector Machines (SVM), and Neural Networks using PyCaret. PyCaret improves efficiency by automating the selection of algorithms, hyperparameter tuning, and model evaluation. The efficacy and accuracy of the trained models in forecasting air quality levels are evaluated by applying suitable performance metrics.

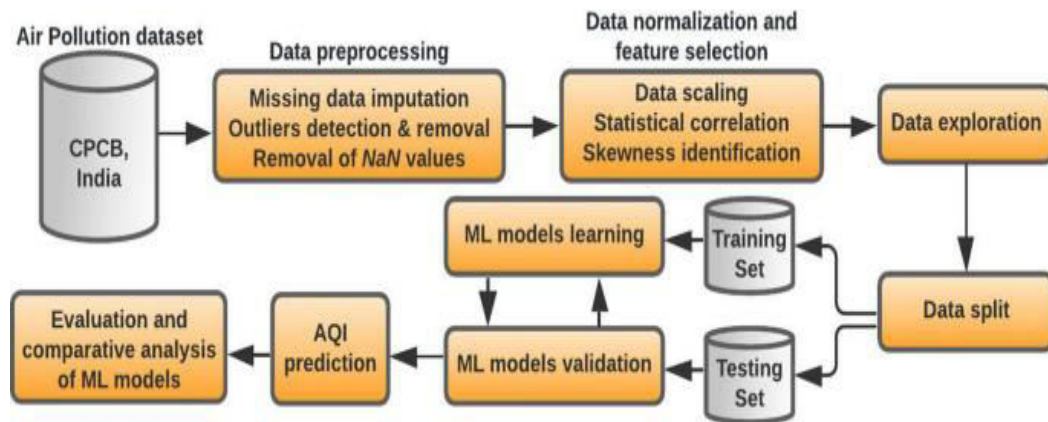


Fig - 1: System Architecture

The trained models are assessed for accuracy and efficacy in air quality prediction using appropriate performance metrics. The models are prepared to make predictions on fresh data inputs after they have been trained and assessed. The input parameters are used by the system to predict the three levels of air quality: poor, medium, and high.

The system architecture describes about the trained model of our project implementation. The various datasets are collected from the free environment. This dataset is pre-processed and the different NaN values are calculated in the data preprocessing. These calculated values are normalised and feature selection is done.

V. RESULTS

At this point, our system is completely operational and prepared to take user input. Users can get real-time predictions of air quality levels by entering parameters like temperature, humidity, PM2.5, and PM10, as well as the Air Quality Index (AQI). Three discrete levels of output—High, Medium, and Poor—will enable users to rapidly evaluate the current condition of the local air quality. People, local government, and environmental agencies can all make educated decisions and take the required steps to protect public health and lessen the effects of air pollution thanks to this user-friendly interface. When we entered the data for the Andhra Pradesh cities of Bapatla and Anakapalle into our air quality forecasting system, different results were produced. With reference to pollutants like PM2.5 and PM10, the system forecasted a moderate level of air quality for Bapatla (Fig:1), which is a reasonably acceptable state. The forecast for Anakapalle (Fig:2), on the other hand, indicated a poor rating for air quality, indicating a higher concentration of pollutants and possibly harmful health effects for locals.

Select State

Select Village Name

AQI

PM2.5

PM10

Temperature

Humidity



Air Pollution status in your area

: MODERATE

Fig - 2: Test Result-1

The above figure is describing about Forecasting the Bapatla region's Air Quality. The quality of air is tested. The details for testing are displayed as the state of location, quality of the air pollution and the other different data like temperature is calculated and the final output is displayed.

Air Pollution Prediction

Select State
Andhra Pradesh

Select Village Name
Anakapalle

AQI
151

PM2.5
56

PM10
114

Temperature
30

Humidity
53

Predict

Air Pollution status in your area

: POOR

Fig - 3: Test Result-2

The above figure is describing about Forecasting the Anakapalle region's Air Quality. The quality of air is tested by collecting all the free air from the location environment. This data is processed and the values are displayed as the final result. The quality of the air status is displayed as the numerical value.

VI. CONCLUSION

In conclusion, by supplying an effective instrument for evaluating and controlling air pollution, our project offers major advantages to society. Our system helps people, communities, and policymakers make informed decisions to protect the environment and public health by precisely forecasting the air quality levels in particular areas.

First off, by making information about air quality conditions easily accessible, our project raises public awareness. People can use this information to schedule outdoor activities, safeguard populations that are more susceptible, and take preventative measures to avoid being exposed to pollutants.

Additionally, by promoting the adoption of greener practices and technologies, our project advances environmental sustainability. By increasing public awareness of the negative effects of air pollution and offering practical advice, we support initiatives to lower emissions and enhance air quality for both current and future generations.

All things considered, our project is an invaluable weapon in the fight against air pollution, enabling people and communities to take preventative action to safeguard the environment and public health. Our goal is to make significant contributions to the development of a cleaner, healthier, and more sustainable society through ongoing improvement and application.

REFERENCES

- [1] Rishanti Murugan, Naveen Palanichamy, "Smart City Air Quality Prediction using Machine Learning", Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021) .
- [2] Shreyas Simu, Varsha Turkar, Rohit Martires, Vranda Asolkar, Swizel Monteiro*, Vaylon Fernandes, "Air Pollution Prediction using Machine Learning", IEEE 2020.
- [3] Tanisha Madan, Shraddha Sagar, Deepali Virmani, "Air Quality Prediction using Machine Learning Algorithms –A Review"2020 2nd International Conference on Advances in Computing, Communication Control and Networking
- [4] Sriram Krishna Yarragunta, Mohammed Abdul Nabi, "Prediction of Air Pollutants Using Supervised Machine Learning ", IEEE 2021
- [5] Chavi Srivastava, Shyamli Singh "Estimation of Air Pollution in Delhi Using Machine Learning Techniques",2018 International Conference on Computing, Power and Communication Technologies (GUCON)
- [6] Jayant Kumar Singh, Amit Kumar Goel, "Prediction of Air Pollution by using Machine Learning Algorithm",2021 7th International Conference on Advanced Computing & Communication Systems (ICACCS)
- [7] Rohit Adke, Suyog Bachhav, "Air Pollution Prediction using Machine Learning", Volume: 06 Issue: 10 | Oct 2019
- [8] Pakki Pavan Sai, Raghupathi RajaSekhar, "Prediction of Air Quality Index by Using Machine Learning", Volume 7 ~ Issue 11 (2022) pp: 01-07
- [9] Shailesh Munge, Sagar Kharche, "Air Quality Prediction based on Machine Learning", Volume 9, Issue 7 July 2021 | ISSN: 2320-2882
- [10] N. Srinivasa Gupta, Yashvi Mohta," Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis", Volume 2023, Article ID 4916267, 26 pages
- [11] Miss. S. Jayasri, Mrs. R. Vijayalakshmi, "Analysis of Real-Time Fine Particulate Matter (PM2.5) Concentrations for Air Pollution Using Machine Learning" Vol 4, no 6, pp 3393-3400 June 2023.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details