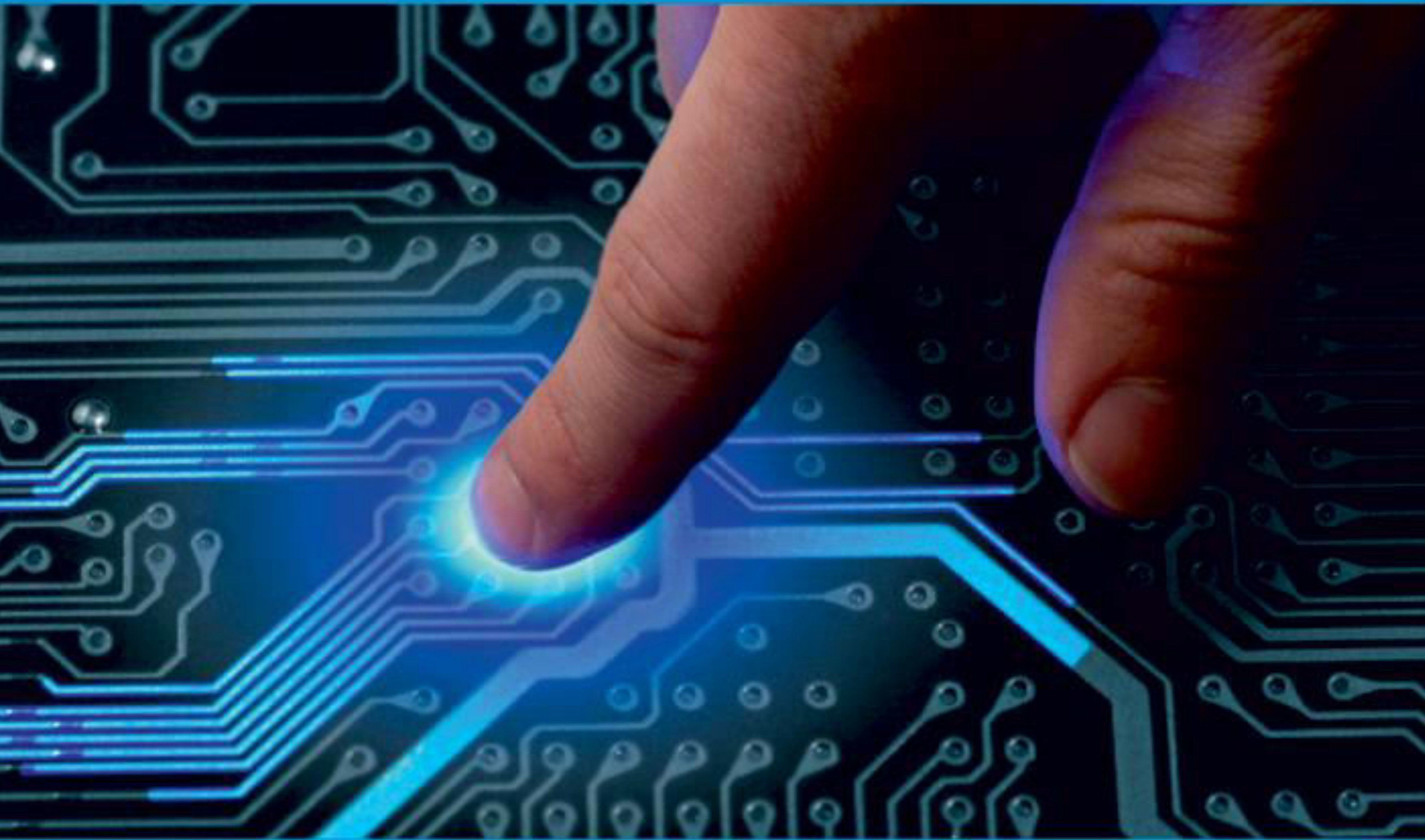
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 12, December 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Quality of Air Prediction using Machine Learning

Sowmya, Sahana

Assistant Professor, Department of CSE, Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Udupi, Karnataka, India

Assistant Professor, Department of CSE, Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Udupi, Karnataka, India

ABSTRACT: The importance of air quality monitoring has increased in many places, particularly in industrialized and metropolitan areas. Air quality is harmed by pollution, which comes from a variety of sources including the production of power, transportation, and fuel use. Inhaling toxic chemicals into the atmosphere poses serious risks to people's health in smart cities. Particulate matter is made up of little solid or liquid particles that are extremely dangerous to inhale and can cause a number of health problems. The component of these particles that are especially harmful to human health is PM2.5, or particles smaller than 2.5 micrometres. Predicting air quality has therefore become extremely important. The project's main goals are to investigate air quality and forecast air pollution using machine learning techniques. It focuses on predicting PM2.5 levels by analysing different air contaminants.

KEYWORDS: Air quality, PM 2.5, Prediction

I. INTRODUCTION

Global economic and social development has led to increased ground-level air pollution in many large cities, particularly in rapidly developing countries like China and India. For a considerable portion of the population, this pollution has a major role in contributing to a variety of health difficulties, including respiratory disorders, bronchitis, lung cancer, throat and eye conditions, asthma, skin infections, and heart problems. Air pollution is a menace to our world as well as a health issue.

Over the past 20 years, climate change has become a major worldwide issue and has been a major topic of discussion in international forums. Acid rain is a result of air pollution and damages trees, wildlife, and soil. The primary cause of both urban smog and global climate change is air pollution. It is essential to develop an early warning system with precise forecasts.

Air pollution causes:
Air pollution is caused by the presence in the atmosphere of toxic substances, mainly produced by human activities, even though sometimes it can result from natural phenomena such as volcanic eruptions, dust storms and wildfires, also depleting the air quality.

Anthropogenic air pollution sources are:
•Combustion of fossil fuels, like coal and oil for electricity and road transport, producing air pollutants like nitrogen and sulfur dioxide.
•Emissions from industries and factories, releasing large amount of carbon monoxide, hydrocarbon, chemicals and organic compounds into the air.
•Agricultural activities, due to the use of pesticides, insecticides, and fertilizers that emit harmful chemicals.
•Waste production, mostly because of methane generation in landfills.

II. RELATED WORK

Air pollution is one of the most significant environmental challenges worldwide, with adverse impacts on human health, ecosystems, and climate. The growing interest in using data-driven techniques, such as the Internet of Things (iot) and machine learning (ML), has led to several advancements in air pollution monitoring and prediction systems.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The following works focus on different aspects of air quality prediction, including the use of iot, machine learning models, and big data techniques.

Ayele and Mehta (2018) explored [1] the use of IoT for air pollution monitoring and prediction. Their study highlights that IoT-based systems collect real-time data on air quality and other environmental factors, which can then be processed to predict pollution levels. The authors demonstrate the potential of integrating IoT sensors with predictive models to create more effective and real-time monitoring systems. They suggest that combining IoT with machine learning algorithms can lead to more accurate air quality predictions, which are essential for environmental monitoring and policymaking.

Castelli et al. (2020) employed[2] machine learning models to predict air quality in California, focusing on the prediction of air pollutants using various data sources. The authors demonstrated that machine learning could be effectively applied to air quality forecasting, with the models showing promise in predicting various pollutant concentrations, including particulate matter and ozone. Similarly, Jeya and Sankari (2020) highlighted the use of deep learning techniques for air pollution prediction. Their deep learning model showed promising results, particularly in forecasting air pollution in urban environments where pollution levels can be highly dynamic.

Rybarczyk and Zalakeviciute (2021) explored[3] the impact of the COVID-19 pandemic on air quality using a machine learning approach. The study found that the lockdowns and restrictions imposed during the pandemic resulted in a significant reduction in air pollution in many urban areas. Using machine learning models, the authors were able to assess the changes in air quality patterns and highlight the potential for using these models to understand the effects of external factors, such as global crises, on pollution levels.

Sanjeev (2021) reviewed[4] various machine learning algorithms for the analysis and prediction of air quality. The study covers several models, including decision trees, support vector machines (SVM), and artificial neural networks (ANN). The author concluded that no single algorithm is universally superior for air quality prediction; rather, the choice of algorithm depends on the specific context, such as the type of pollutants being measured, data availability, and computational resources.

Problem Statement and Motivation:
Monitoring and protecting the quality of the air is now essential in many industrial and urban locations. Air pollution is a serious issue that is caused by fuel use, energy generation, and transportation.

III. OBJECTIVE

Developing a machine learning model with accurate air quality metric prediction is the aim. The goal of this model is to provide accurate and timely air quality forecasts, which will aid in environmental and public health monitoring.

IV. PROPOSED SYSTEM

For parameter evaluation, photo-based methods are essential, but handling the massive amounts of data they produce can be difficult. A graphical user interface (GUI) program will be used to build a machine learning technique in order to address this. This is combining several datasets from different sources into a large one so that machine learning algorithms may find patterns and produce extremely accurate results.

Architectural Design:
In system design, the term "architectural design" refers to the process of establishing a system's overall structure and organization, including its components, their interactions, and the design concepts that underpin them. High- level decisions about the structure and operation of the system must be made, including the selection of the proper hardware and software, the creation of the user interface, and the definition of the data model. Because it establishes the framework for the system's functionality, performance, and maintainability, architectural design is a crucial stage in system.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

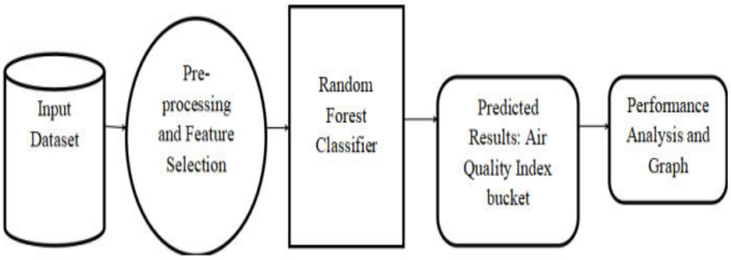


Fig 1: System design of proposed model.

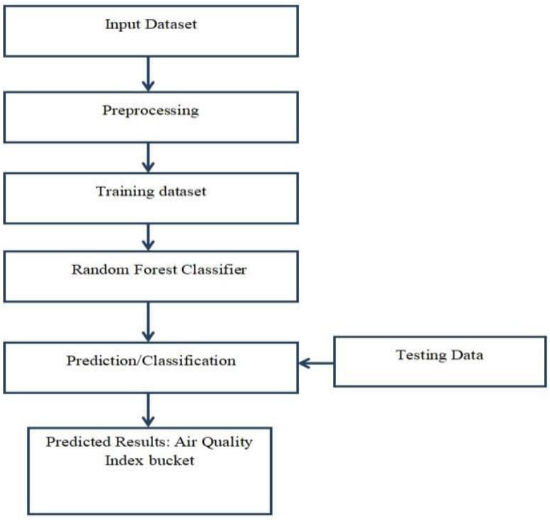


Fig 2: Process of dataflow diagram

Implementation:

A. Data collection: There are several techniques to collect the data, like web scraping, manual interventions. The dataset is located in the model folder. The dataset is referred from the popular dataset repository called Kaggle.
B. Data set: The dataset consists of 29532 individual data. There are 16 columns in the dataset.
C. Algorithm: **The Random Forest Algorithm**

How does the algorithm work?
It works in four steps:
Select random samples from a given dataset.
Construct a decision tree for each sample and get a prediction result from each decision tree. Perform a vote for each predicted result.
Select the prediction result with the most votes as the final prediction.

© 2024 IJIRCCE | Volume 12, Issue 12, December 2024| DOI: 10.15680/IJIRCCE.2024.1212096

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

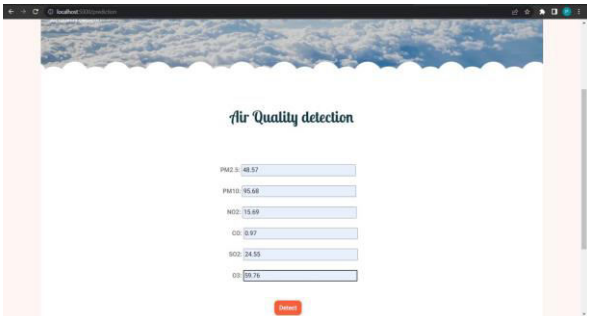


Fig. 3. Entering Pollutant values

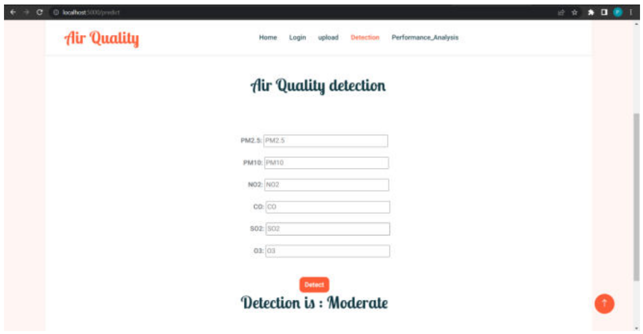


Fig 4. Detecting AQ

According to the Indian Government (CPCB), Indian AQI range is from 0-500, from 0 being good and 500 being severe. There are eight major pollutants to be taken into account for AQI calculation. Fig 3 showing entering particular city pollutant value and fig 4 shows detecting entered values air quality.

Performance Evaluation:

Precision: The proportion of positive predictions that are actually correct. (When the model predicts default: how often is correct?)

Precision = TP / (TP + FP)

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labeled as survived, how many actually survived? High precision relates to the low false positive rate. We have got 0.788 precision which is pretty good.

Recall: The proportion of positive observed values correctly predicted. (The proportion of actual defaulters that the model will correctly predict)

Recall = TP / (TP + FN)

Recall(Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

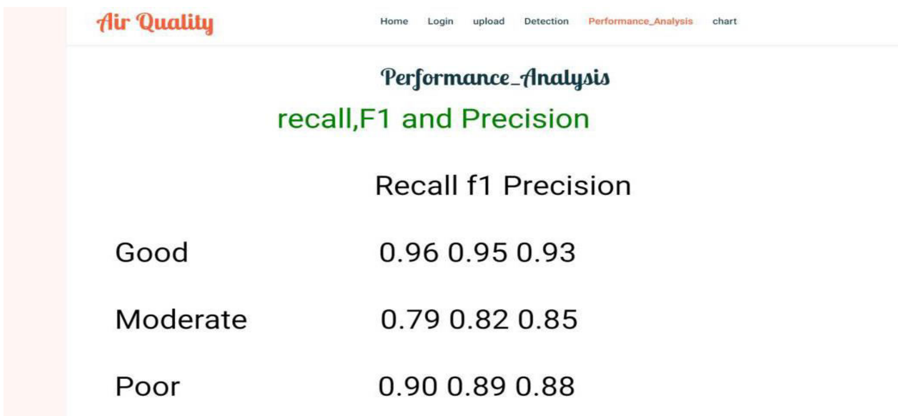


Fig 5 : Performance Analysis

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION AND FUTURE WORK

The main advantage of the model is that it can predict the data under different constraints according to the existing data, and can know the possible air pollution process in advance and take corresponding control measures, to improve the ambient air quality. The air quality prediction model proposed and implemented in this paper can greatly improve the prediction accuracy and provide a reference for future research on air quality prediction. The Random Forest Classifier has proven to be a promising method for air quality prediction, producing accurate and reliable results. The algorithm's ability to handle a large number of input variables and its robustness to overfitting make it a suitable choice. For this task, additionally, its interpretability and versatility make it suitable for implementation in a wide range of real-world applications. However, it is important to note that the performance of the Random Forest Classifier can be affected by factors such as the quality and quantity of training data, the choice of hyperparameters, and the specific problem being addressed. Further research and optimization may be needed to fully realize the potential of this method for air quality prediction.

As a future enhancement to the system, the system can be automated to detect the air quality from real time data and the results can be displayed on web application. To optimize this model it can be implemented in Artificial Intelligence environment.

REFERENCES

1. Chenghao Shi, Yu Wang, Ying Wan, Shuguang Wu, “Air Quality Prediction Based on Machine Learning”, 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE), IEEE Conference, 2022.
2. Kumar, Anikender and P. Goyal, “ Forcasting of daily air quality index in Delhi”, Science of th Total Environment 409, no. 24(2011): 5517-5523
3. Singh Kunwar P., et al. “Linear and nonlinear modelling approaches for urban air quality prediction, “ Science of the Total Environment 426(2012):244-255
4. Sivacoumar R, et al, “ Air pollution modelling for an industrial complex and model performance evaluation “, Environmental Pollution 111.3 (2001) : 471-477
5. Gokhale sharad and Namita Raokhande, “Performance evaluation of air quality models for predicting PM10 and PM2.5 concentrations at urban traffic intersection during winter period”, Science of the total environment 394.1(2008): 9-24.
6. Bhanarkar, A. D., et al, “Assessment of contribution of SO2 and NO2 from different sources in Jamshedpur region, India, “Atmospheric Environment 39.40(2005):7745-India." Atmospheric Environment 39.40 (2005): 7745-7760

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING