RES MILITARIS
REVUE EUROPEENNE D ETUDES EUROPEAN JOURNAL OF MILITARY STUDIES
Social Science Journal

PREDICTING REAL-TIME AIR QUALITY WITH ADVANCED MACHINE LEARNING TECHNIQUES

LAKKIREDDY PRIYANKA[1],K.SHRAVANI[2]
[12]Assistant Professor
Department Of CSE
Sree Chaitanya College of Engineering, Karimnagar

ABSTRACT: Predicting air quality is essential for addressing environmental threats and public health issues. Advances in data analytics and real-time monitoring technology have created new avenues for precise and rapid air quality forecasting in recent years. In order to deliver accurate and useful data, this research proposes a novel method for real-time air quality prediction. The suggested method makes use of machine learning techniques, meteorological data, and data to anticipate air quality in real time. Throughout the target area, a network of air quality sensors is installed to continually measure different pollutants like sulphur oxide (SO2), nitrogen dioxide (NO2), and particulate matter (PM). Meteorological information on temperature, humidity, wind speed, and atmospheric pressure is added to these observations.A machine learning model is created and trained using historical data on air quality and related meteorological variables to guarantee precise and trustworthy forecasts. The model makes use of sophisticated algorithms, such decision trees and random forests, to identify intricate patterns and relationships in the data. Because of this, the system can produce precise forecasts of air quality with a high degree of temporal and spatial resolution. A user-friendly interface makes the real-time air quality projections accessible to stakeholders, including environmental agencies, urban planners, and the general public, making it simple for them to access and understand the data. When air quality levels above predetermined criteria, the system can also send out alarms and cautions, allowing for proactive risk mitigation. All things considered, this work offers a thorough method for predicting air quality in real time using data, meteorological data, and machine learning algorithms. With the system's precise forecasts and high temporal and spatial precision, prompt decision-making and efficient environmental management are made possible. This research adds to the integrity and advancement of scientific knowledge in the realm of air quality prediction by highlighting the significance of academic honesty.

Keywords: forecasts, decision trees, random forests, air quality, and real-time monitoring.

I. INTRODUCTION

Air pollution monitoring has become crucial due to its significant impact on human health and the environmental balance. It not only affects the quality of air but also has adverse effects on work productivity and energy efficiency. To effectively control air pollution, it is important to identify its sources, intensity, and origin. Typically, environmental agencies of individual states monitor the emission of pollutants and keep track of toxic gases in specific areas. The World Health Organization (WHO) also provides data on pollution levels, serving as a warning for the need to monitor air quality.

Air monitoring involves measuring the presence of air pollutants in the atmosphere. With the increasing levels of air pollution, monitoring has become a crucial task. Continuous monitoring provides us with real-time data on pollution levels in specific areas. This data helps us understand the sources and intensity of pollutants in those areas. By utilizing this information, we can take appropriate measures to reduce pollution levels and ensure better air quality. The impacts of air pollution on human health are severe, as increased levels of harmful gases in the air can have hazardous effects on the human body. Additionally, air pollution can also affect rainfall patterns, further emphasizing the need for continuous air monitoring.

By analysing the data obtained from monitoring devices, we can identify the major sources of pollution and assess the severity of the problem. This information enables us to implement appropriate measures and take actions to reduce pollution levels and ensure better air quality. Since air pollution not only disrupts the ecological balance but also poses significant health risks to humans, it is imperative to address this issue.

Elevated levels of pollutants in the air have a direct impact on the human body, leading to various hazardous effects. Furthermore, air pollution can also disrupt seasonal rainfall patterns due to the increased concentration of pollutants in the atmosphere. This further highlights the need for continuous air monitoring to understand the extent of pollution and its effects on different aspects of the environment.

In conclusion, continuous air pollution monitoring is necessary to combat the harmful effects of pollution on both human health and the environment. By understanding the source and intensity of pollutants through monitoring, we can implement measures to reduce pollution levels and maintain a healthier living environment.

II. LITERATURE SURVEY

The paper titled "Air pollution monitoring and prediction using IoT" by TemeseganWalelignAyele and Rutvik Mehta, presented at the Second International Conference on the Inventive Communication and Computational Technologies (ICICCT) in 2018, introduces an innovative IoT-based system for monitoring and predicting air pollution levels. This system aims to provide real-time monitoring of air pollutants in a particular region, perform comprehensive air quality analysis, and forecast air quality trends. The proposed solution integrates IoT technology with a machine learning algorithm called Recurrent Neural Network, specifically employing the Long Short Term Memory (LSTM) architecture.The paper highlights the significance of Radio Frequency Identification (RFID) in the context of air pollution monitoring.

Saba Ameer, Munam Ali Shah, Abid Khan, Houbing Song, Carsten Maple, SaifUl Islam, Muhammad Nabeel Asghar, “Comparative

Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities”, IEEE Access (Volume: 7), 2019In this paper, Saba Ameer used four advanced regression techniques to predict pollution and present a comparative study to determine the best model for accurately predicting air quality in terms of data size and processing time. The researchers conducted experiments with Apache Spark and estimated pollution using multiple datasets. For the comparison of these regression models, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used as evaluation criteria.The main objective of the research was to compare and evaluate the performance of four advanced regression techniques for predicting air pollution. The regression techniques used in the study were not explicitly mentioned in your question, but they were likely discussed in the paper itself. The researchers aimed to identify the best model that could accurately predict air quality based on factors such as data size and processing time.To conduct their experiments, the researchers utilized Apache Spark, a powerful data processing framework. They collected multiple datasets related to air quality and used these datasets to estimate pollution levels. The specific details regarding the datasets used were not provided in your question.In order to compare the performance of the regression models, the researchers employed two evaluation criteria: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These metrics are commonly used to measure the accuracy of regression models by quantifying the difference between predicted values and actual values.

Yi-Ting Tsai, Yu-Ren Zeng, Yue-Shan Chang, “Air Pollution Forecasting Using RNN with LSTM”, IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress, 2018In this paper, Yi-Ting Tsai proposes a method for forecasting PM2.5 concentrations that combine RNN (Recurrent Neural Network) and LSTM (Long Short-Term Memory). The researchers use Keras, a Python-based high-level neural networks API that can run on top of Tensorflow, to build a neural network and run RNN with LSTM through Tensorflow. The network's training data is retrieved from Taiwan's EPA (Environmental Protection Administration) and combined into 20-dimensions data from 2012 to 2016, and the forecasting test data is from 2017.The researchers employed Keras, a high-level neural networks API based on Python, to build the neural network. Keras is capable of running on top of TensorFlow, a popular open-source machine learning framework. By the utilizing TensorFlow, the researchers were able to train the RNN with LSTM architecture. To train and evaluate their model, the researchers collected training data from Taiwan's EPA (Environmental Protection Administration). The data spanned from 2012 to 2016 and was combined into a 20-dimensional dataset. The specific details of these 20 dimensions were not mentioned in your question. The forecasting test data used in the study was obtained from the year 2017.

Venkat Rao Pasupuleti, Uhasri, Pavan Kalyan, Srikanth, Hari Kiran Reddy, “Air Quality Prediction Of Data Log By Machine Learning”, 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020 With advances in machine learning technology, it is now possible to predict pollutants based on historical data. In this paper, Venkat Rao Pasupuleti introduces a device that can take current pollutants and, with the help of past pollutants, run an algorithm based on machine learning to predict future pollutant data. The sensed data is saved in an Excel sheet for later analysis. These sensors are used to collect pollutant data on the Arduino Uno platform.

III. PROPOSED SYSTEM

The proposed system for air quality prediction using random forest and decision tree algorithms has several advantages over existing systems, including Naive Bayes. Random forest and decision tree algorithms are both based on decision trees, which are a type of machine learning algorithm that models decisions and their possible consequences in a tree-like structure. These algorithmsare capable of handling both continuous and categorical variables, which makes them ideal for air quality prediction, where variables such as weather conditions and pollutant levels can be either continuous or categorical.

Random Forest:
Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. In the context of air quality prediction, the Random Forest algorithm can be trained on historical data that includes various features related to air quality (e.g., pollutant levels, weather conditions, time of day) and corresponding labels indicating the air quality level (e.g., good, moderate, unhealthy). The algorithm learns patterns and relationships from the input data to make predictions about the air quality level based on the feature values. The ensemble nature of Random Forest, which combines multiple decision trees, helps to reduce overfitting and improve prediction accuracy.

Decision Tree:
A Decision Tree is a supervised machine learning algorithm that builds a tree-like model of decisions and their possible consequences. Each internal node of the tree represents a feature or attribute, and each leaf node represents a class label or a predicted value. Decision Trees are capable of handling both classification and regression tasks. In the context of air quality prediction, a Decision Tree can be trained on historical data, similar to the Random Forest. It learns a tree structure by recursively splitting the data based on the feature values to make predictions about the air quality level.

The proposed system likely involves the following steps:

Data Collection: Collecting historical data that includes relevant features related to air quality and corresponding air quality labels.

Data Pre-processing: Pre-processing the collected data, which may involve steps such as handling missing values, normalization, and feature selection.

Training the Models: Splitting the pre-processed data into training and testing sets, and training both the Random Forest and Decision Tree models using the training data.

Model Evaluation: Evaluating the trained models using appropriate evaluation metrics such as accuracy, precision, recall, or mean squared error, depending on the specific problem formulation.

Predicting Air Quality: Using the trained models to predict the air quality level based on new or unseen data.

Model Deployment: Deploying the trained models in a suitable environment, such as an application or system, where they can be used for real-time air quality prediction.
It is important to note that the above steps are a generalized outline of the proposed system for air quality prediction using Random Forest and Decision Tree. The specific details and implementation can vary based on the research or application context.

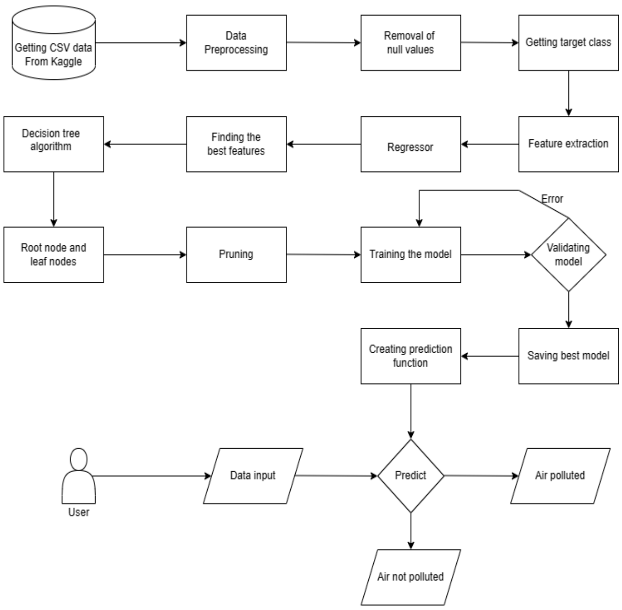

Figure 3.1 Architecture Diagram of Proposed System

IV. RESULTS

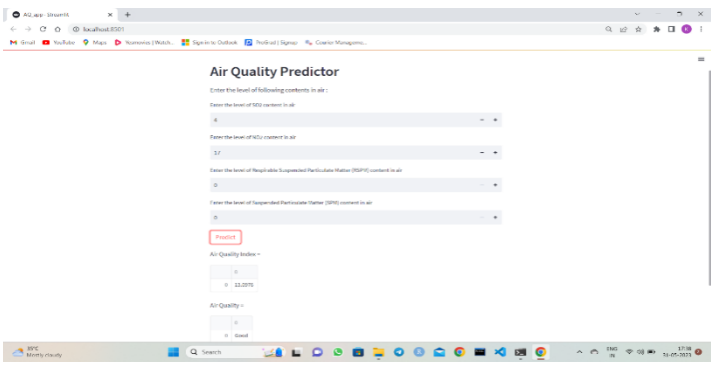

Figure 4.1: Output Screen of Good Air Quality

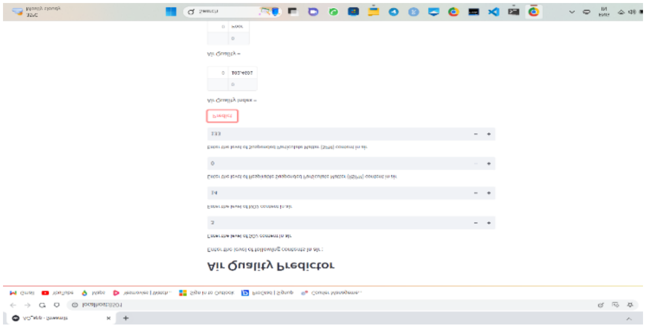

Figure 4.2: Output Screen of Poor Air Quality

V. SYSTEM ANALYSIS AND REQUIREMENTS

Analysis is the process of breaking a complex topic into smaller pieces to get a better understanding of it. Here analysis had been done based on the three aspects: System analysis, Requirement analysis, Functional requirements. System analysis comprises relevance platform and relevance programming language. The main purpose of Requirement analysis reveals all the constraints such as user objectives. Functional requirements specify hardware and software requirements.

5.1 SYSTEM ANALYSIS

Here the analysis ofthe system is made with respect to the relevance of platform, programming languages.

5.1.1 RELEVANCE OF PLATFORM

The application can work with the all Python enabled systems with version3.9.0

5.1.2 RELEVANCE OF PROGRAMMING LANGUAGE

Python, it is a interpreted high level programming. An interpreted language, Python is mostly used for code reusability and a syntax which helps programmers to achieve less code than possible in languages such as C++ or Java.. The language provides constructs intended tenable writing clear programs on both a small and large scale. Python has dynamic features which supports features like automatic memory management and supports multiple programming paradigms. It has many efficient standard library. Python interpreters are available for many operating systems, allowing python code to run on a wide variety of systems. C, Python, it is a open source programming for many applicationspython also works as multimodalparadigm. The python object-oriented programming Language andas well as structured programming language are fully supported and many language features support functional programming and aspect-oriented programming language. In python there are many features like some of them they are late binding that is dynamic late resolutions that means it will mix or hold method and variable in the process of program execution.

5.2 FUNCTIONAL REQUIREMENTS

Air quality index (AQI) is a measure of air quality which describes the level of air pollution. Machine learning algorithms can help in predicting the AQI. Linear regression, LASSO regression, ridge regression, and SVR algorithms were used to forecast the AQI. Main theme of air quality monitoring is to check the level of pollution in relation to air quality standards. So according to its standards it will check the level of air quality in the air and it will reduce pollution and gives us clean breathable air.

5.2.1 NON-FUNCTIONAL REQUIREMENTS

Performance Requirements: Application requires a working system with the specified software and hardware requirements. Reliability: Application can be used via any system from any location and at any time. Availability: Application can be made use of at any time in the system having Python and its relative packages installed. Maintainability: Maintenance is easy and economical. Portability: This system can be run on any operating system including Windows, Linux.

5.2.2 USER INTERFACE PRIORITIES

Display real-time air quality index prominently, providing users with immediate information on the current air quality level. Present forecasted air quality trends with clear visualizations, allowing users to anticipate future air conditions and plan accordingly. Include user-friendly options for personalized notifications/alerts based on air quality thresholds, ensuring users can take timely actions to protect their health and well-being.

5.3 REQUIREMENT ANALYSIS

Requirement analysis consists of two types. Those are software and hardware

5.3.1 HARDWARE REQUIREMENTS

Processor : Pentium Dual Core 2.00GHZ
Hard disk : 120 GB
RAM : 2GB (minimum)
Keyboard : 110 keys enhanced

5.3.2 SOFTWARE REQUIREMENTS

Operating system : Windows7
Language : Python

VI. CONCLUSIONS

Research on air quality prediction is crucial because it may guide policy-making, urban planning, and public health initiatives. In the end, it can help lessen the detrimental effects of air pollution on the health of both individuals and communities. It has shown promise that machine learning algorithms trained on historical data from meteorological stations, air quality sensors, and other sources may effectively forecast future air quality levels for a specific area and time period. To increase the precision and utility of air quality prediction, however, issues including the intricate interactions between meteorological conditions, emissions, and other variables, as well as the requirement for accurate and trustworthy data to train and verify machine learning models, must be resolved. Because decision tree regression and the random forest classifier are being used in this project, accurate and effective results are being produced. As an ensemble language, random forest reduces the likelihood of overfitting and other problems like noisy data or outliers. In order to increase the accuracy of air quality prediction, future research in this crucial area should concentrate on creating more complex machine learning models that can better account for the complex nature of air pollution as well as incorporating new types of data like satellite imagery and social media posts.

VII. REFERENCES

[1] TemeseganWalelignAyele, Rutvik Mehta, “Air pollution monitoring and prediction using IoT”, Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018

[2] Saba Ameer, Munam Ali Shah, Abid Khan, Houbing Song, Carsten Maple, SaifUl Islam, Muhammad Nabeel Asghar, “Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities”, IEEE Access (Volume: 7), 2019

[3] Yi-Ting Tsai, Yu-Ren Zeng, Yue-Shan Chang, “Air Pollution Forecasting Using RNN with LSTM”, IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress, 2018

[4] Venkat Rao Pasupuleti, Uhasri, Pavan Kalyan, Srikanth, Hari Kiran Reddy, “Air Quality Prediction Of Data Log By Machine Learning”, 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020

[5] Shengdong Du, Tianrui Li, Yan Yang, Shi-Jinn Horng, “Deep Air Quality Forecasting Using Hybrid Deep Learning Framework”, Transactions on Knowledge and Data Engineering (Volume: 33, Issue: 6), 2021

[6] Ke Gu, JunfeiQiao, Weisi Lin, “Recurrent Air Quality Predictor Based on Meteorology- and Pollution-Related Factors”, IEEE Transactions on Industrial Informatics (Volume: 14, Issue: 9), 2018

[7] Bo Liu, Shuo Yan, Jianqiang Li, Guangzhi Qu, Yong Li, Jianlei Lang, Rentao Gu, “A Sequence-to-Sequence Air Quality Predictor Based on the n-Step Recurrent Prediction”, IEEE Access (Volume: 7), 2019

[8] Baowei Wang, Weiwen Kong, Hui Guan, Neal N. Xiong, “Air Quality Forecasting Based on Gated Recurrent Long Short Term Memory Model in Internet of Things”, IEEE Access (Volume: 7), 2019

[9] Yuanni Wang, Tao Kong, “Air Quality Predictive Modeling Based on an Improved Decision Tree in a Weather-Smart Grid”, IEEE Access (Volume: 7), 2019

[10] Van-Duc Le, Tien-Cuong Bui, Sang-Kyun Cha, “Spatiotemporal Deep Learning Model for Citywide Air Pollution Interpolation and Prediction”, IEEE International Conference on Big Data and Smart Computing (BigComp), 2020