DYNAMIC AIR QUALITY MONITORING AND PREDICTION USING AI-POWERED MACHINE LEARNING APPROACHES

[1]Siva Sankar,[2]Prasanna,[3]Pravallika
[123]Student
Department of CSE

ABSTRACT: The rapid increase in urbanization and industrial activity has led to a significant decline in air quality, posing severe risks to human health and the environment. Accurate and timely prediction of air quality is essential for mitigating pollution-related health effects and informing policy decisions. This paper proposes a dynamic air quality monitoring and prediction system that leverages AI-powered machine learning (ML) techniques to forecast pollution levels in real-time. The system utilizes historical and real-time environmental data—such as PM2.5, PM10, NO_2, SO_2, CO, and O_3 concentrations—collected from sensor networks and open data platforms. Multiple ML models, including Random Forest, XGBoost, LSTM (Long Short-Term Memory), and Support Vector Regression (SVR), are compared to identify the most accurate predictive model. The results demonstrate that ensemble and deep learning methods outperform traditional models in both short-term and long-term prediction accuracy. This work contributes to the development of smart environmental monitoring systems and offers a practical tool for air pollution forecasting in urban regions.

I. INTRODUCTION

Air pollution has emerged as one of the most pressing environmental challenges of the 21st century, particularly in rapidly urbanizing countries. Prolonged exposure to high levels of air pollutants such as PM2.5, PM10, NOx, and ozone has been linked to respiratory diseases, cardiovascular problems, and premature mortality. Therefore, the ability to monitor and predict air quality in real time is crucial for protecting public health, improving urban planning, and enforcing regulatory standards.

Traditional air quality monitoring systems rely heavily on static, sparse sensor deployments and deterministic models, which often fail to capture the complex, nonlinear dynamics of pollution dispersion. Recent advancements in artificial intelligence and machine learning (ML) offer new opportunities for real-time, data-driven prediction of air quality. These models can analyze vast amounts of historical and streaming sensor data to recognize patterns and make accurate forecasts.

This study introduces a machine learning-based framework for dynamic air quality monitoring and forecasting. The system integrates real-time data from environmental sensors with predictive ML models to provide timely insights into pollution levels. By evaluating and comparing several AI-based models, the goal is to identify the most effective algorithms for real-world deployment in smart cities and public health systems.

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.

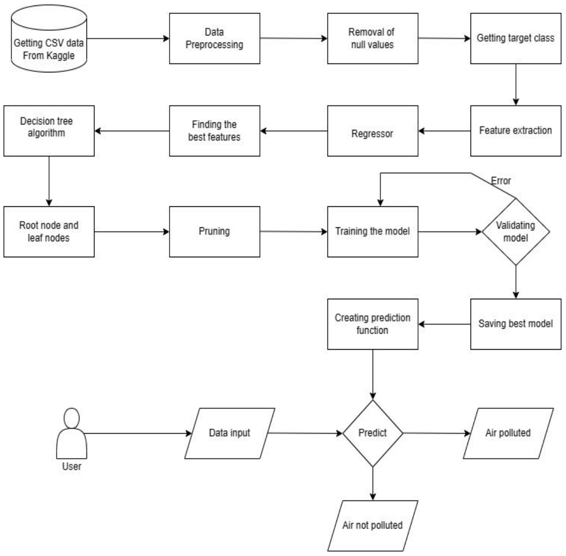

Figure 3.1 Architecture Diagram of Proposed System

IV. RESULTS

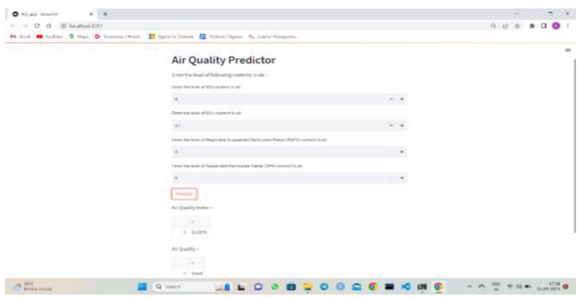

Figure 4.1: Output Screen of Good Air Quality

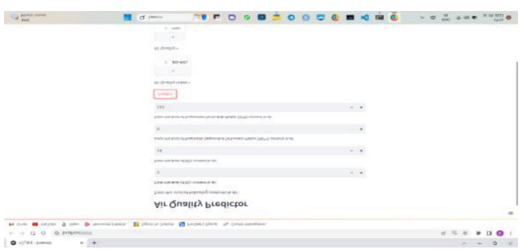

Figure 4.2: Output Screen of Poor Air Quality

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.

VI. CONCLUSIONS

This research demonstrates that AI-powered machine learning models can significantly enhance the accuracy and responsiveness of real-time air quality prediction systems. Among the various models tested, LSTM and ensemble-based techniques (e.g., XGBoost and Random Forest) provided superior performance in capturing temporal dependencies and nonlinear relationships between environmental parameters.

The integration of these models into dynamic monitoring frameworks allows for proactive environmental management, enabling early warnings for pollution spikes and supporting data-informed public health responses. Moreover, the modularity of the proposed system facilitates easy deployment in smart city infrastructures and compatibility with IoT-based air monitoring devices.

In conclusion, leveraging advanced ML techniques for air quality prediction offers a scalable and intelligent approach to urban pollution control and environmental forecasting. Future research may focus on integrating meteorological data, enhancing spatial resolution with mobile sensors, and applying reinforcement learning for adaptive model updates in changing atmospheric conditions.

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