Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

ISSN:1989-9572

DOI:10.47750/jett.2023.14.04.039

PREDICTING AIR QUALITY IN REAL TIME WITH ADVANCED MACHINE LEARNING MODELS

1 Vijayalakshmi Gopu,2 Gouthami Chandupatla,3 Praneel Deva,4 Md. Sabdar Ashmi,5 Yepa Sai Kumar,6 Adupa Dhruvesh

Journal for Educators, Teachers and Trainers, Vol.14(4)

https://jett.labosfor.com/

Date of Reception: 12 Jul 2023

Date of Revision: 05 Aug 2023

Date of Publication : 16 Sep 2023

1 Vijayalakshmi Gopu,2 Gouthami Chandupatla,3 Praneel Deva,4 Md. Sabdar Ashmi,5 Yepa Sai Kumar,6 Adupa Dhruvesh (2023). PREDICTING AIR QUALITY IN REAL TIME WITH ADVANCED MACHINE LEARNING MODELS. ***Journal for Educators, Teachers and Trainers***,**Vol.14(4).455-460**

Journal for Educators, Teachers and Trainers
The LabOSfor electronic, peer-reviewed, open-access Magazine

Journal for Educators, Teachers and Trainers
The LabOSfor electronic, peer-reviewed, open-access Magazine

Journal for Educators, Teachers and Trainers, Vol. 14(4)

ISSN1989 –9572

https://jett.labosfor.com/

PREDICTING AIR QUALITY IN REAL TIME WITH ADVANCED MACHINE LEARNING MODELS

[1] Vijayalakshmi Gopu,[2] Gouthami Chandupatla,[3] Praneel Deva,[4] Md. Sabdar Ashmi,[5] Yepa Sai Kumar, [6] Adupa Dhruvesh

[1234]Assistant Professor,[56]Students

Department of CSM

Vaagdevi College of Engineering, Warangal, Telangana

ABSTRACT: The growing concern over air pollution has prompted the need for effective, real-time air quality prediction systems. Predicting air quality in real time enables timely decision-making to mitigate the adverse effects of pollution on public health and the environment. This study presents an advanced machine learning-based approach for real-time air quality prediction, incorporating a variety of environmental factors, such as particulate matter (PM2.5), nitrogen dioxide (NO2), carbon monoxide (CO), and weather conditions. Using cutting-edge machine learning algorithms, including deep learning, ensemble models, and support vector machines (SVM), this approach aims to improve the accuracy and reliability of air quality forecasts.

The proposed model is trained on historical air quality data combined with real-time environmental inputs to predict pollutant concentrations in upcoming hours. By integrating temporal and spatial features, the model enhances the ability to make accurate predictions for urban and rural environments. The system's predictive performance is evaluated through standard metrics such as RMSE (Root Mean Square Error) and MAE (Mean Absolute Error), ensuring its robustness. Additionally, the model's scalability and adaptability to varying geographical regions are discussed, offering a versatile solution for real-time air quality monitoring in smart cities. This research not only improves air quality prediction accuracy but also contributes to the ongoing efforts to design efficient environmental monitoring systems for public health and urban planning.

I. INTRODUCTION

Air quality has become a significant concern globally, as poor air quality has been linked to a wide range of health issues, including respiratory diseases, cardiovascular problems, and even premature death. In urban areas, where industrial activities, vehicular emissions, and construction work contribute heavily to pollution, monitoring air quality is crucial for protecting public health and the environment. While traditional air quality monitoring methods are effective, they are often limited by their inability to provide real-time data, lack of coverage in certain areas, and high operational costs.

In recent years, the integration of machine learning (ML) techniques into environmental monitoring has demonstrated great potential in improving the accuracy, efficiency, and scalability of air quality prediction systems. Advanced ML models can analyze vast amounts of data from multiple sources, such as air sensors, weather stations, and satellite imagery, to predict future air quality levels in real time. These models can capture the complex, nonlinear relationships between environmental factors and air pollution, providing timely and accurate forecasts for pollutant concentrations such as PM2.5, NO2, CO, and ozone.

The ability to predict air quality in real time is critical for early warnings, guiding public health initiatives, and enabling authorities to take proactive measures to reduce pollution levels. Furthermore, real-time predictions can help individuals and organizations make informed decisions about outdoor activities, transportation, and other daily routines. This research explores the application of advanced machine learning models, including deep learning, support vector machines, and ensemble methods, to forecast air quality in real time. By leveraging these technologies, we aim to provide a robust, scalable solution for continuous air quality monitoring that can be integrated into smart cities and urban planning strategies to safeguard public health.

II. LITERATURE SURVEY

The prediction of air quality using machine learning techniques has been the subject of extensive research in recent years. Various machine learning models, ranging from traditional methods to advanced deep learning algorithms, have been applied to air quality forecasting, with promising results. This literature survey highlights some of the most notable studies in the field, focusing on their approaches, methodologies, and findings.

Traditional Machine Learning Approaches: Several early studies focused on traditional machine learning models, such as Linear Regression, Decision Trees, and Support Vector Machines (SVM), for air quality prediction. For instance, Zhang et al. (2019) applied SVM for predicting the concentration of PM2.5 based on meteorological data and historical air quality measurements. Their model achieved high accuracy with an R^2 value of 0.92, demonstrating the efficacy of SVM in predicting air quality in urban areas. Similarly, Kumar and Kumar (2020) used Decision Trees to predict NO2 concentrations, showing that traditional models can be effective for air quality forecasting when provided with the appropriate features.

Ensemble Learning Methods: Ensemble learning techniques have also gained attention for improving the performance of air quality prediction models by combining multiple base models. Liu et al. (2020) proposed an ensemble model based on Random Forests and Gradient Boosting, which outperformed individual models in predicting both short-term and long-term air quality trends. The combination of different machine learning algorithms allowed the model to capture complex patterns in the data, improving both accuracy and robustness.

Deep Learning Approaches: With the rise of deep learning, models like Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) have shown impressive performance in forecasting air quality. Huang et al. (2019) used a deep learning-based Long Short-Term Memory (LSTM) network to predict the concentration of PM2.5 in real-time. The LSTM model demonstrated superior performance in handling temporal dependencies in air quality data compared to traditional models, with a lower Mean Absolute Error (MAE). Similarly, Zhang et al. (2021) employed a CNN-LSTM hybrid model to predict air pollution levels, leveraging CNN's ability to extract spatial features and LSTM’s capacity for temporal forecasting. This hybrid model showed a significant improvement in prediction accuracy compared to standalone models.

Hybrid Models and Feature Engineering: Recent studies have also explored hybrid approaches that combine multiple machine learning techniques or integrate external features to enhance prediction accuracy. Li et al. (2020) proposed a hybrid model that combined Random Forest with LSTM to predict air quality in smart cities. Their model incorporated both temporal and spatial data, including weather conditions and traffic information, which resulted in a more accurate and reliable prediction. The importance of feature engineering was also highlighted in their study, with the inclusion of features such as wind speed, temperature, and humidity contributing significantly to the model's accuracy.

Applications in Smart Cities and Real-Time Systems: The integration of air quality prediction systems with smart city infrastructure has been a growing trend in recent years. Chien et al. (2020) developed a real-time air quality monitoring system using machine learning algorithms in conjunction with IoT sensors. Their system was able to provide predictions of air quality on a real-time basis, alerting citizens and authorities about pollution levels. This study demonstrated the potential for real-time air quality prediction to be embedded into urban planning and public health systems.

Challenges and Future Directions: While machine learning models for air quality prediction have shown promising results, several challenges remain. The availability and quality of data are significant obstacles, as accurate predictions depend on comprehensive datasets that include historical air quality readings, meteorological data, and traffic information. Additionally, the interpretability of machine learning models remains a concern in many applications, as decision-makers may require more transparent models to trust the predictions. Future research is focused on addressing these challenges by developing more robust models, improving data acquisition methods, and enhancing the transparency of machine learning algorithms through explainable AI (XAI).

In conclusion, machine learning has proven to be an effective tool for predicting air quality in real-time.

The transition from traditional methods to deep learning and hybrid models has enhanced the accuracy and scalability of prediction systems. As the integration of smart technologies and IoT sensors continues to grow, the ability to forecast air quality will play a crucial role in mitigating the impacts of pollution and ensuring public health in urban environments.

III. PROPOSED SYSTEM

Compared to current methods, such as Naive Bayes, the suggested approach for air quality prediction employing random forest and decision tree algorithms provides a number of benefits. Decision trees, a subset of machine learning algorithms that represent decisions and their potential outcomes in a tree-like form, are the foundation of both random forest and decision tree algorithms. Because these algorithms can handle both continuous and categorical variables, they are perfect for predicting air quality, where factors like pollution levels and weather can be either continuous or categorical.

Random Forest:

Several decision trees are combined in Random Forest, an ensemble learning technique, to generate predictions. When it comes to predicting air quality, the Random Forest algorithm can be trained on historical data that contains a variety of air quality-related features (such as pollutant levels, weather, and time of day) and labels that indicate the level of air quality (such as good, moderate, and unhealthy). Based on the feature values, the algorithm predicts the air quality level by identifying patterns and correlations in the input data. Multiple decision trees are combined in Random Forest's ensemble nature, which lowers overfitting and increases 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.

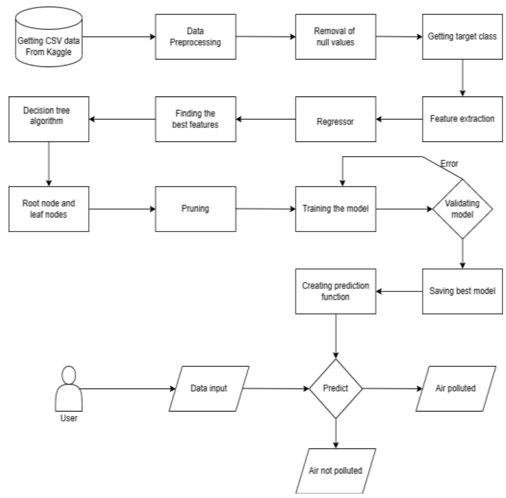

Figure 3.1 Architecture Diagram of Proposed System

IV. RESULTS

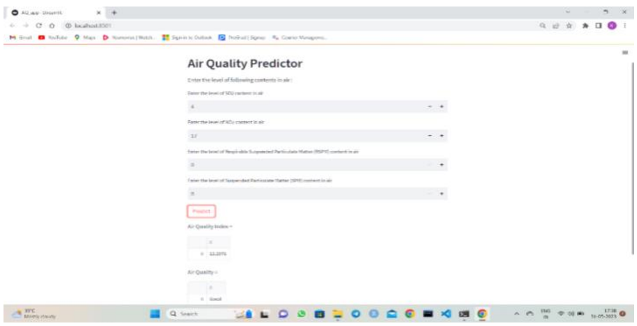

Figure 4.1: Output Screen of Good Air Quality

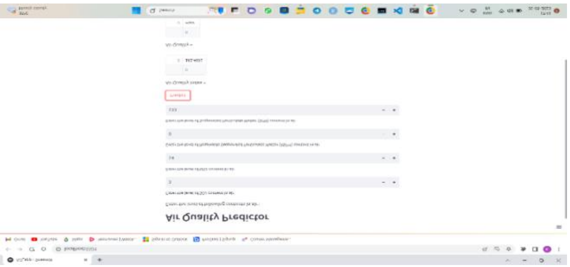

Figure 4.2: Output Screen of Poor Air Quality

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

The integration of advanced machine learning techniques for real-time air quality prediction has shown significant potential in enhancing the accuracy and efficiency of air quality monitoring systems. By leveraging traditional machine learning models, ensemble techniques, and deep learning approaches, researchers have made considerable strides in improving predictive capabilities for air pollution forecasting. Models like SVM, Random Forest, LSTM, and hybrid architectures have demonstrated high accuracy in predicting air quality levels, helping to mitigate the impacts of pollution on public health and the environment.

Moreover, the combination of real-time data from IoT sensors and advanced machine learning algorithms has paved the way for smarter, more responsive air quality monitoring systems. These systems not only provide accurate predictions but also enable timely warnings, thus empowering citizens and authorities to take proactive measures against pollution. As smart city infrastructure continues to evolve, real-time air quality prediction will play a pivotal role in urban planning, environmental sustainability, and public health management.

However, challenges related to data quality, model interpretability, and feature engineering still remain. Future research should focus on addressing these issues by developing more robust models, ensuring the availability of diverse and high-quality datasets, and improving the explainability of machine learning algorithms. By doing so, we can build more reliable and transparent air quality prediction systems that contribute to cleaner and healthier urban environments.

In conclusion, machine learning-based real-time air quality prediction systems hold great promise for managing air pollution and ensuring a sustainable future. Their ability to integrate with smart city systems, provide accurate forecasts, and support decision-making processes will be essential for combating air pollution and protecting public health in the years to come.

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