An Effective Prediction of Air Pollution by Using Machine Learning Models

Dr. G. Kavitha[1], T. Somasekhar Reddy[2], S. Arun Kumar[3] & U. Venu[4]

[1] Assistant Professor, [2,3,4] B-Tech, IV ECE Student,

[1,2,3,4] Department of Electronics and Communication Engineering, Dr. M.G.R Educational and Research Institute, Tamilnadu, India.

Abstract: Air pollution is a global issue posing health risks and environmental degradation. Governments in developing countries, particularly India, are tasked with regulating air pollutant levels. IoT-based air quality monitoring systems collect data on pollutants like PM2.5, NO2, SO2, ozone, and CO. The project aims to create a hybrid machine learning algorithm that synergizes the advantages of both Random Forest and Linear Regression models to provide precise predictions of air pollution levels. A user-friendly website will be developed to disseminate real-time air quality alerts to the public. The project focuses on enhancing the machine learning model, optimizing location-based recommendations, implementing user authentication, integrating a database for data storage, refining push notifications, fostering community engagement, and ensuring robust privacy measures. The objective of this project is to improve the accuracy and timeliness of predictions related to air pollution, providing decision-makers with actionable insights to effectively address concerns about air quality.

Keywords: *Air pollutant levels, Random Forest algorithm, Linear Regression algorithm, Hybrid Machine learning algorithm, User authentication, Robust privacy.*

1. Introduction

Environmental health relies heavily on air quality, a fundamental element that directly influences the health and well-being of individuals and communities. With the rise of urbanization and industrialization, the need for effective air quality monitoring systems has become increasingly crucial. The "AN EFFECTIVE PREDICTION OF AIR POLLUTION BY USING MACHINE LEARNING MODELS" aims to address this challenge by combining cutting-edge technologies, including Internet of Things (IoT) devices and machine learning algorithms, to provide accurate predictions, real-time location-based recommendations, and a dynamic user feedback loop.

The project begins with the deployment of IoT devices equipped with various sensors to measure key environmental parameters such as particulate matter, gas concentrations, temperature, and humidity. These devices enable continuous monitoring of air quality, creating a robust dataset that forms the foundation for predictive modelling.

To predict air quality, a hybrid machine learning model is employed, merging the strengths of Random Forest and Linear Regression algorithms. This approach aims to enhance prediction accuracy, considering both the complexities of air quality dynamics and the interpretability of the model. The integration of machine learning not only provides accurate predictions but also sets the stage for continuous improvement through feedback-driven enhancements.

The existing web interface serves as a user-friendly platform where individuals can access air quality predictions, receive alerts, and contribute to the system's learning through feedback. As part of the proposed work, this interface will undergo significant enhancements to provide dynamic and personalized recommendations, community engagement features, and stronger privacy and security measures.

The ultimate goal of effective air pollution by using machine learning models is to empower individuals with timely and relevant information about their local air quality, enabling them to make well-informed decisions that have a positive impact on both their health and the environment. As air quality continues to be a pressing global concern, this project seeks to contribute to a more sustainable and healthier future by harnessing the potential of technology to address real-world challenge.

2. Literature Survey

1. “Air Quality Index Forecasting via Genetic Algorithm-Based Improved Extreme Learning Machine” by Chunhao Liu, Guangyuan Pan, Dongming Song, and Hao Wei. In this paper the authors describes a method called GA-KELM for predicting air quality. It uses a combination of genetic algorithms and machine learning to make better predictions about air quality, including factors like pollution levels. This method is more accurate and faster than other methods tested in experiments.

2. “Air Quality Prediction and Monitoring using Machine Learning Algorithm based IoT sensor- A researcher’s perspective” by G. Kalaivani Research Scholar, Dr. P. Mayilvahanan. In this paper the authors Strives to make a valuable contribution to the realm of air quality prediction and monitoring through the integration of IoT and ML technologies, with the overarching objective of safeguarding human health against the detrimental impacts of air pollution.

3. “Machine Learning algorithms for air pollutants forecasting” by Marius Dobrea, Andreea Badicu, Marina Barbu, Oana Şubea, Mihaela Balanescu, Geroge Suciu, Andrei Birdici, Oana Orza, Ciprian Dobre. In this paper the authors objective is to address the issue of air pollution and its negative effects worldwide. It aims to propose and present numerous machine learning algorithms are available for forecasting upcoming air quality.

4. “Air Quality Prediction of Data Log by Machine Learning” by Venkat Rao Pasupuleti, Uhasri, Pavan Kalyan, Srikanth, Hari Kiran Reddy. In this paper the authors aim to introduce a device that can monitor and predict air pollutants based on past data using machine learning algorithms. The device is designed to sense and collect pollutant data using sensors on the Arduino Uno platform.

5. “Prediction of daily PM10 concentration using machine learning” by Oumaima Bouakline, Khadija Arjdal, Kenza Khomsi, Noureddine Semane, Noureddine Semane, Salem Nafiri, Houda ajmi. In this paper the authors present Humidity and Temperature monitoring in the smart garden. In this system information from the sensors and feedback from the actuators are transmitted to the server and can be observed using a smartphone. To analyze collected measurements of air quality over ten years from the Zrektouni air quality station. in Casablanca (Morocco) and predict the daily average of particulate matter with a diameter less than 10 µm (PM10) using machine learning models.

3. Existing Work

An effective prediction of Air Pollution by Using Machine Learning Models" involves the use of Internet of Things (IoT) devices that are equipped with sensors to monitor air quality parameters comprehensively. These sensors measure things, like particulate matter, gas concentrations, temperature and humidity. The information collected by these sensors is used in a machine learning model that employs algorithms such, as Random Forest or Linear Regression to predict air pollution levels. Additionally, the project includes a web interface where users can access the predictions and receive alerts.

4. Proposed Work

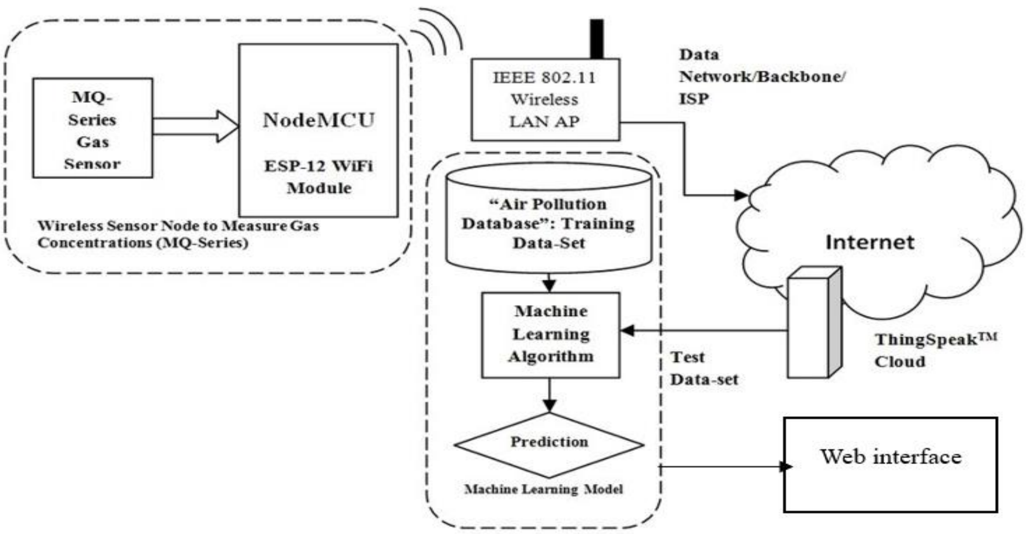
The proposed work introduces a comprehensive strategy to elevate the existing system. By developing hybrid machine learning model, the project strives to significantly improve he precision and dependability of air pollution predictions. The integration of dynamic data sources, including real-time feeds and historical data, promises to enhance the system's adaptability to changing environmental conditions. The introduction of user authentication and personalization features is designed to offer individuals a more tailored experience, allowing for customized preferences, personalized alerts, and specific recommendations. Community engagement features, such as forums and social sharing functionalities, seek to foster a collaborative environment, encouraging users to share information and insights. The integration of a database aims to streamline data management, supporting data-driven decision-making and continuous model improvement. The optimization of the alert system, guided by user feedback, ensures more timely and informative notifications, enhancing the user experience. Strengthened privacy measures and compliance with data protection regulations contribute to a more secure and trustworthy system. Finally, the implementation of continuous monitoring ensures that users receive the latest information even during periods of inactivity, positioning the project as a cutting-edge solution in the realm of effective air quality prediction systems.

5. Methodology

The methodology for the "An effective Prediction of Air Pollution by using Machine Learning Models" project unfolds through a systematic process. Commencing with the deployment of IoT devices for live data collection regarding parameters of air quality, including particulates, gas concentrations, the project focuses on preprocessing to clean and normalize the collected data. The heart of the methodology lies in the development of a hybrid machine learning model, seamlessly integrating the strengths of Random Forest and Linear Regression algorithms. Feature engineering and hyperparameter tuning enhance the model's understanding and optimize its performance. Ensemble techniques, such as bagging and boosting, are explored to further elevate predictive accuracy and robustness.

To enhance interpretability, the project incorporates advanced feature engineering techniques and model-agnostic interpretability tools. Real-time adaptability mechanisms ensure the model dynamically adjusts to changing environmental conditions. Continuous model monitoring, coupled with rigorous validation and testing protocols, guarantees ongoing evaluation and refinement. Utilizing separate datasets for training, validation, and testing, the methodology ensures unbiased model evaluation. This comprehensive approach aims not only for accurate air pollution predictions but also for a model that adapts to real-time changes, fosters interpretability, and undergoes continuous improvement, positioning the project at the forefront of innovative solutions in environmental monitoring.

6. System Architecture



7. System Specifications

7.1. Hardware Components

➢ Arduino UNO: The Arduino Uno, a widely-used microcontroller board crafted for hobbyists, students, and professionals to easily create electronic projects. It is built on the ATmega328P microcontroller and comes with a variety of built-in components, including digital and analog input/output pins, USB interface, power jack, and a reset button.

➢ DHT11 Sensor: The DHT11 serves as an economical, basic digital sensor for measuring temperature and humidity. Utilizing a capacitive humidity sensor and a thermistor, it gauges the ambient air and delivers a digital signal on the data pin, eliminating the need for analog input pins. While it is relatively straightforward to use, obtaining data requires precise timing. The main drawback is that new data can only be acquired from the sensor once every 2 seconds.

➢ MQ2 Sensor: MQ2 Gas Sensor is a Gas Sensor of the Metal Oxide Semiconductor (MOS) type primarily employed for detecting gases such as Methane, Butane, LPG, Smoke, and more. Often referred to as Chemiresistors, this sensor detects gases by monitoring the alteration in resistance of the sensing material upon contact with the gas.

➢ MQ135 sensor: The MQ-135 is a sensor module that can detect different gases in the air, including things like ammonia, carbon dioxide, and methane. It works by changing its voltage output depending on the concentration of the gas it detects. People use it in projects to check air quality or identify specific gases in the environment.

➢ BMP180 Sensor: The BMP180 is a digital barometric pressure sensor designed to measure atmospheric pressure and temperature. It is manufactured by Bosch Sensortec and is commonly used in various applications, including weather stations, altimeters, and drones.

➢ Node MCU: The NodeMCU is an open-source electronics platform built upon the ESP8266 WiFi module. It is commonly used for building Internet of Things (IoT) projects and applications.

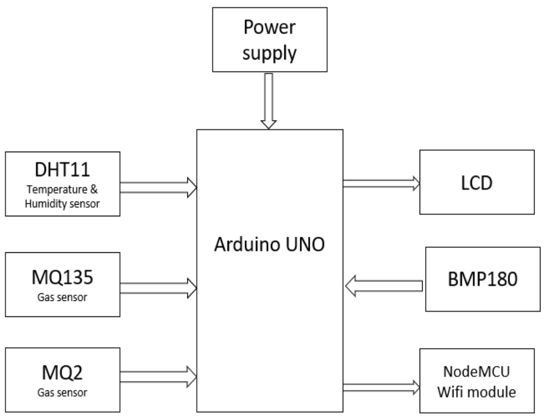
7.2. Software Components

➢ Arduino Software: Arduino The IDE, an open-source software, is primarily utilized for coding and compiling the code for the Arduino Module. As the official Arduino software, it simplifies the code compilation process to the extent that even individuals without prior technical knowledge can easily delve into the learning process.

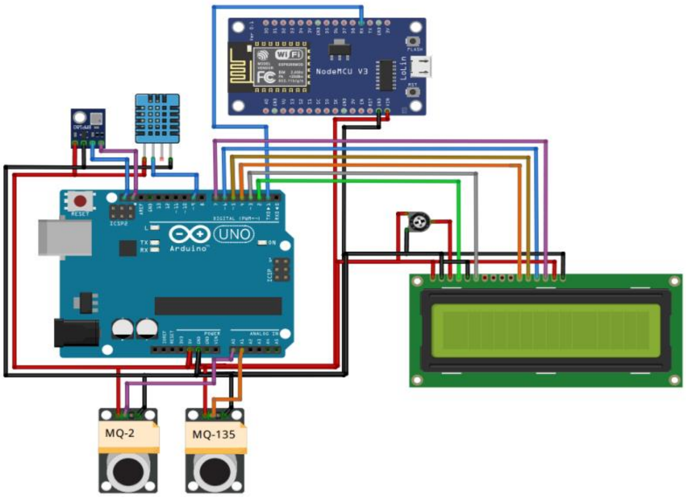
➢ Python: Python is a versatile, dynamic, high-level, and interpreted programming language with a general purpose. Supporting an Object-Oriented programming approach for application development, it is known for its simplicity and ease of learning, offering a plethora of high-level data structures.

8. DESIGN

8.1. Block Diagram



8.2. Circuit Diagram



9. Working

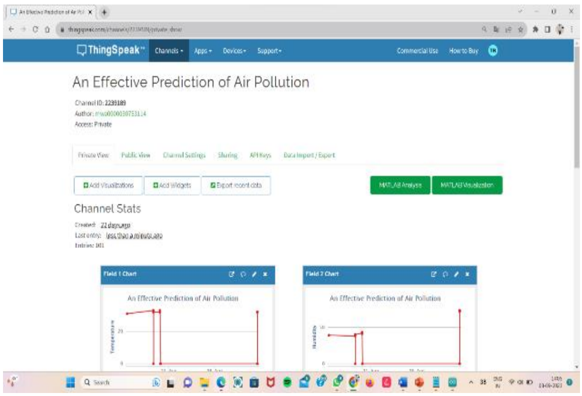
The working of, "An effective Prediction of Air Pollution by using a Machine Learning Models" operates through a systematic and collaborative workflow. It begins with the continuous monitoring of air quality parameters by IoT devices, which collect real-time data from the environment. The collected data undergoes thorough cleaning and preprocessing in the Data Processing Module, ensuring its quality and readiness for model training. The core of the system lies in the Hybrid Machine Learning Model, where both Random Forest and Linear Regression algorithms collaboratively contribute to predicting air pollution levels with enhanced accuracy and robustness.

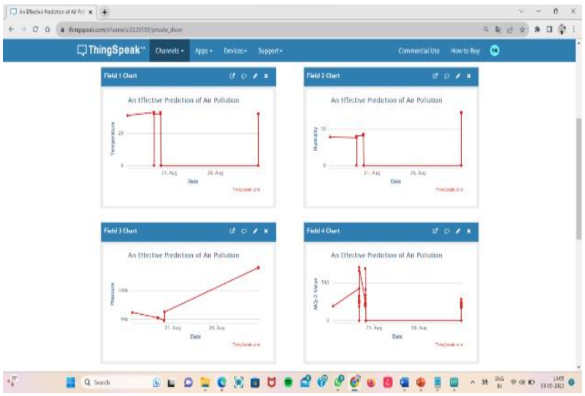
Advanced feature engineering, hyperparameter tuning, and the integration of ensemble techniques further refine the model's predictive capabilities. The Interpretability and Explain ability Module ensures that the model's predictions are not only accurate but also comprehensible, fostering trust and understanding. The Real-time Adaptability Module allows the model to dynamically adjust to changing environmental conditions, ensuring its effectiveness in dynamic scenarios. Continuous monitoring and rigorous validation and testing protocols contribute to the ongoing evaluation and improvement of the model's performance over time.

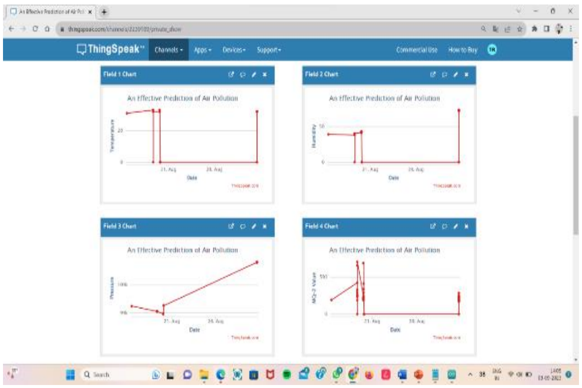
The User Interface serves as the gateway for users to access air pollution predictions, receive alerts, and interact with the system. Simultaneously, the Database acts as a central repository for user feedback, historical data, and model training data, facilitating data-driven decision-making, trend analysis, and continuous improvement of the machine learning model. This collaborative and iterative approach ensures that the project not only predicts air pollution effectively but also engages users and adapts to evolving environmental conditions, positioning it as an innovative solution in the realm of environmental monitoring.

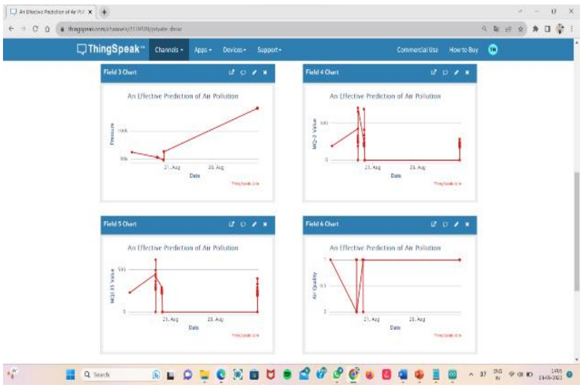
10. Result

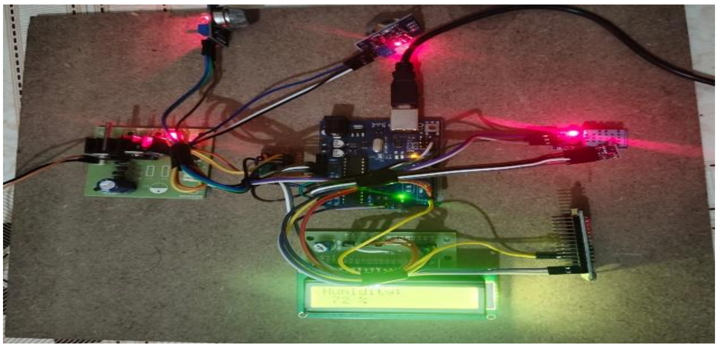
The IoT devices equipped with MQ2, MQ135, DHT11, and BMP180 sensors, the primary outcome revolves around the real-time acquisition and visualization of environmental data. The sensors, each serving a specific purpose in monitoring air quality, temperature, humidity, and barometric pressure, contribute to a comprehensive dataset. The values obtained from these sensors are transmitted and stored on the ThingSpeak server, where they are graphically represented, allowing for a visual interpretation of fluctuations and patterns over time.











11. Discussion

In this hardware simulation, we've undertaken a critical step to assess the reliability of our data collection setup by simulating sensor data and observing the output on ThingSpeak. Our primary focus has been on ensuring the accuracy and consistency of the simulated data, which is essential for reliable air quality monitoring. We've initiated the validation and calibration processes, which will be pivotal in guaranteeing the accuracy of our data in real-world scenarios. The successful integration of simulated data with ThingSpeak is a significant milestone, indicating the capability of our chosen IoT platform. These insights and achievements from the simulation phase will serve as a solid foundation for refining our system before progressing to the deployment of physical sensors and real-time data monitoring.

12. Conclusion

In conclusion, our project successfully addresses the critical issue of air pollution by combining IoT-based monitoring, hybrid machine learning algorithms, and a user-friendly website. We have demonstrated the effectiveness of our hybrid model, which merges Random Forest and Linear Regression algorithms, in predicting air quality accurately. The website we developed provides a valuable platform enabling users to access real-time air quality data, visualize historical trends, and receive timely alerts when air quality deteriorates.

This project contributes to public health by empowering individuals with the knowledge they need to enable informed decision-making regarding outdoor activities, health precautions, and environmental awareness. While the initial implementation is promising, we acknowledge that further enhancements and refinements are possible. Future work may include expanding the range of supported IoT devices, adding air quality forecasting capabilities, and incorporating user feedback for continuous improvement. Overall, our project serves as a valuable tool in the ongoing efforts to combat air pollution and its negative impacts on both human health and the environment.

References

[1] Vivek Singh "Indian air quality prediction and analysis using Machine learning" Journal publication of International Research for Engineering and Management- Volume: 10 Issue: 05, oct-2021.

[2] Kaushal Kumar, Dr.B.P. Pande "Air pollution prediction with machine learning: a case study of Indian cities" International journal of Environmental Science and Technology" DOI: 10.1007/s13762-022-04241-5, May 2022.

[3] N. Srinivasa Gupta , Yashvi Mohta, Khyati Heda, Raahil Armaan, B. Valarmathi, and G. Arulkumaran "Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis" Journal of Environmental and Public Health Volume 2023, Article ID 4916267, January 2023.

[4] Madhuri VM, Samyama Gunjal GH, Savitha Kamalapurkar "Air Pollution Prediction Using Machine Learning Supervised Learning Approach" International journal of Scientific & Technology research-Volume 9, Issue 04, April 2020.

[5] Vidit Kumar, Sparsh Singh, Zaid Ahmed, Ms. Nikita Verma "Air Pollution Prediction using Machine Learning Algorithms: A Systematic Review" International Journal of Engineering Research & Technology-Vol. 11 Issue 12, December 2022.

[6] Venkat Rao Pasupuleti, Uhasri, Pavan Kalyan, Srikanth, Hari Kiran Reddy "Air Quality Prediction of Data Log by Machine Learning" International Conference on Advanced Computing & Communication Systems- July 2020.

[7] Oumaima BOUAKLINE, Oumaima BOUAKLINE, Kenza KHOMSI, Noureddine SEMANE, Abdelhak ELIDRISSI, Salem NAFIRI, Houda NAJMI "Prediction of daily PM10 concentration using machine learning" International Conference on Electronics, Control, Optimization and Computer Science-DOI: 10.1109/ICECOCS50124.2020.9314380.

[8] G. Kalaivani, Dr. P. Mayilvahanan "Air Quality Prediction and Monitoring using Machine Learning Algorithm based IoT sensor- A researcher's perspective" International Conference on Communication and Electronics System-DOI: 10.1109/ICCES51350.2021.948915, January 2021.

[9] Tanisha Madan, Shraddha Sagar, Deepali Virmani "Air Quality Prediction using Machine Learning Algorithms –A Review" International Conference on Advances in Computing, Communication Control and Networking-DOI: 10.1109/ICACCCN51052.2020.9362912, April 2020.

[10] Marius Dobrea, Andreea Badicu, Marina Barbu, Oana, Ciprian Dobre "Machine Learning algorithms for air pollutants forecasting" International Symposium for Design and Technology in Electronic Packaging-May 2020.