Air Quality Monitoring and Prediction through IOT Sensor Network data Using Machine Learning and Deep Learning Algorithms

N. Sai Sirisha, N. Maheshwari, R. Nitheesh, S. Soniya
Department Of CSE, Hyderabad Institute of Technology and Management, India

I. ABSTRACT

Air pollution, such as the quantity of PM2.5 and PM10, is one of the most significant variables affecting human health. Due to a lack of air pollution measuring stations near the user, it is challenging to obtain accurate information. In order to efficiently measure and forecast air quality using various machine learning and deep learning techniques, our research uses data gathered by both fixed and mobile IOT sensors that are mounted on vehicles patrolling the area. By utilising LSTM, SVM, random Forest Algorithm, Gradient Boosting algorithms for learning and training data, we proposed a method for the prediction of air pollution. The test's findings demonstrate the performance of the air pollution prediction system and model that proposed deviation.

II. INTRODUCTION

As our world becomes more technologically advanced, so does the number of industries and services required to keep up. As the number of industries increases, so does pollution. Air pollution, in particular, poses a far greater threat than others because we are exposed to the air every minute of our lives. It is important not only to measure air quality, but also to predict air quality so that necessary measures can be taken to minimize the impact of future air pollution.

Nine of the ten most polluted cities in the world are located in India. Pollution levels around the world are increasing at an alarming rate. Of these, air pollution is particularly dangerous as it makes breathing difficult and dangerous. Air quality is measured by counting the number of pollutants such as NO_2, CO, SO_2, PM 2.5, PM 10 etc. Air pollution can cause a variety of breathing problems and lung diseases including lung cancer. Despite efforts to monitor and store all data, they need improvement to work effectively. Our project uses data collected by both real-time and stationary IoT sensors and uses machine learning and deep learning to predict the air quality for the next 30 days.

The main goal of this project is to predict and forecast the air quality for the next 30 days and compare the different algorithms used in this project to predict air quality based on the RMSE (Root Mean Square Error) values around the selection of the best fit algorithm for this model.

III. LITERATURE SURVEY

There was an existing hybrid approach to efficiently monitor air quality using multiple static sensors as well as mobile IoT Sensors. Static sensors can provide a holistic view by providing a continuous feed of information. On the other hand, mobile sensors can provide more accurate region-specific data to reduce errors create a predictive model to use the collected data and provide quick information about air quality around people. We also developed a visualization tool to better analyse and predict air quality, providing insights for professional researchers and general users.

The main contributions of our work can be summarized as follows:

- In this approach to integrating fixed and mobile IoT sensors to measure and predict air quality data.
- We demonstrate the feasibility and effectiveness of our method by analysing the prediction results of different machine models.
- We developed a visualization tool to show the relative distribution of air pollutants, focusing on PM10 and PM2.5 which provide a visual understand of the air quality around people.

PM2.5 and PM10

Minute particles suspended in air or water is alluded to as particulate matter (PM). Tiny solid or liquid particles in air are known as aerosols. PM10 incorporates particles that are ten µm in distance across, PM2.5 those yet 2.5 µm. The perniciousness of suspended particles is a result of the particles with a measurement of however

10µm. They'll be transmitted straightforwardly very high from development exercises (industry, private, agribusiness, transport) and regular sources (backwoods fires, volcanic emissions, and so on.). Particles additionally can be designed straightforwardly inside the air by physio-compound responses between contamination as of now gift inside the air. PM2.5 and PM10 are minute particles gift

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

inside the air and openness thereto is unimaginably destructive for wellbeing. when how much those particles will increment and infiltrate profoundly in to the lungs, you'll mastery assortment of wellbeing impacts like respiratory disadvantage, burning or itching sensation inside the eyes and so forth.

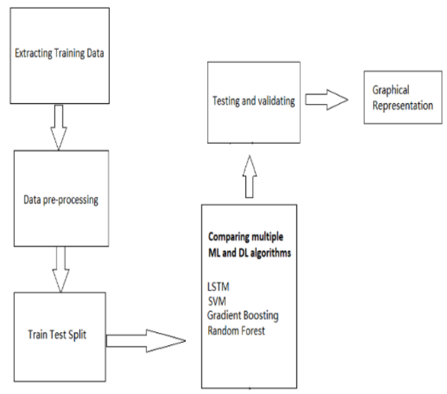
Disadvantages of Existing System The system existing so far has only focused on the hardware part. It does not take into account the accuracy of the prediction.

IV. PROPOSED SYSTEM

Rather than focusing on the hardware and data collection, the proposed project is interested in using sensor-collected datasets to measure air quality through PM2.5 and PM10 levels using various Machine Learning and Deep Learning algorithms for efficient determination, monitoring and prediction.

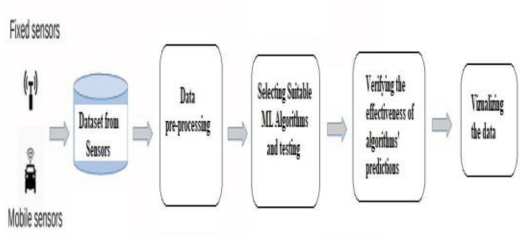
V. DESIGN

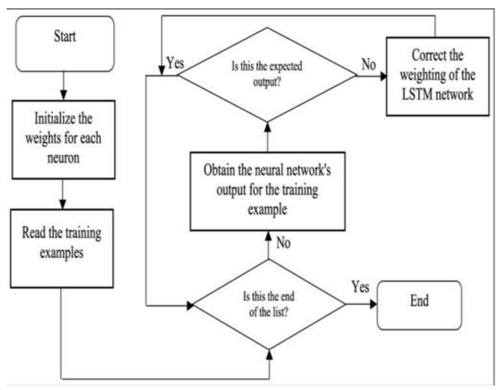
- The design phase is the time when the necessary preparations and moves are made taken. The fundamental criteria have all been planned. Configuration addresses identifying programming components, pointing the links between the components, and keep track of plan selections. plan that included itemising and constructing the scheme. The project is made to meet all of user's requirements. And the test scenarios are met. For the purpose of identifying vulnerabilities in the code. Taking into account all of these, we work to develop software that is, more effective and components of hardware. In the proposed system, multiple ML techniques are used to predict air quality.
- Dl algorithms, too. When the software is executed, a popup appears asking to put a dataset online. After that, the user can prepare the dataset to obtain the prediction results. In order to better understand how well each algorithm functions and which algorithm is the best among them, the algorithms employed are also compared.
- In the below process model, the training data is extracted and pre-processed first. All the NA and invalid values are cleaned up. After that, the data set is split. The training data is fitted to LSTM, SVM, Random Forest and Gradient Boosting algorithms. Testing and validation follow. The last step is graphical representation.



Architecture of Project

The system is a software product, first, the users upload the records into the window. The users just need to run the model to see if their hardware supports our product. To implement our product, users can upload any valid dataset they want to view and run the model. The architecture consists of data collected from fixed and mobile IoT sensors. Pre-process the collected data to eliminate unwanted values and tuples. Do feature extraction to find useful features and eliminate everything else. This allows for faster data processing and more accurate results. The processed data is now fed into 4 models for training and testing. We used LSTM, gradient boosting, SVM and random forest in this project. After the data is trained and tested, we get the result. To check the accuracy of the algorithms, we further validated the model. Once the results are satisfied, visualise the data in the form of charts to better understand the project.





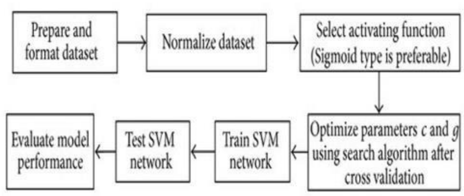
Algorithms and Flowcharts

1. ***LSTM Algorithm***

Long Short-Term Memory (LSTM) is a deep learning system on recurrent neural networks. In the sequence prediction challenge, recurrent neural networks can learn sequential dependencies. They store knowledge about previous inputs for an indefinite period of time and use it to anticipate the next output. It is used for processing, forecasting and classification of time series data.

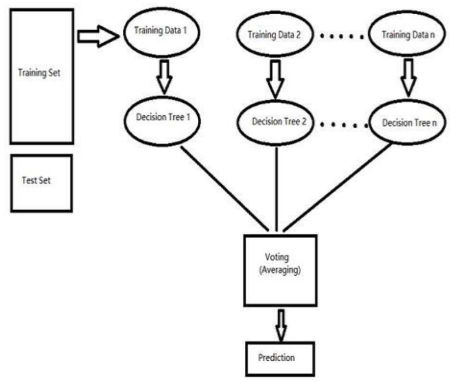
2. ***SVM Algorithm***

A popular Supervised Learning technique for tackling classification and regression issues is the Support vector Machine (SVM). The majority of the time, it is utilised in machine learning to address categorization issues. The SVM method's objective is to identify the ideal decision boundary or line for classifying n-dimensional space into groups so that following data points can be quickly assigned to the appropriate category. The best potential choice boundary is a hyperplane.



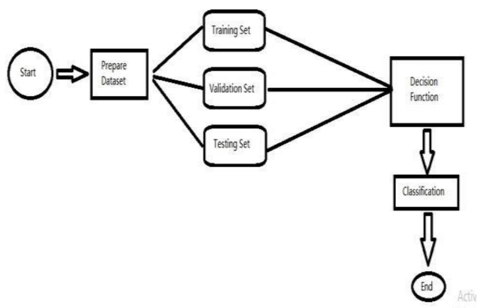
3. Random Forest-Algorithm

The supervised learning method is used by Random Forest, a well-known machine learning algorithm. In machine learning, it can be applied to both classification and regression issues. It is based on ensemble learning, which is a technique for combining various classifiers to handle a challenging problem and enhance the performance of the model. Random Forest is a classifier that increases the predicted accuracy of a dataset by averaging the outcomes of multiple decision trees applied to various subsets of the dataset. "The random forest collects predictions from each decision tree and predicts the final result depending on the majority of votes, as opposed to relying just on one decision tree.



4. Gradient Boosting Algorithm

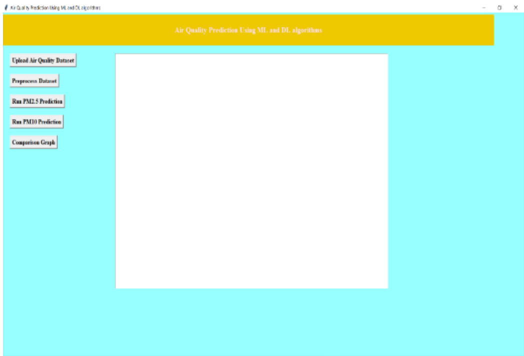
One method for lowering a model's bias inaccuracy is gradient boosting Every prediction fixes the mistakes made by the one before it. It makes prediction using a collection of flimsy prediction models, such decision trees.



VI. RESULTS

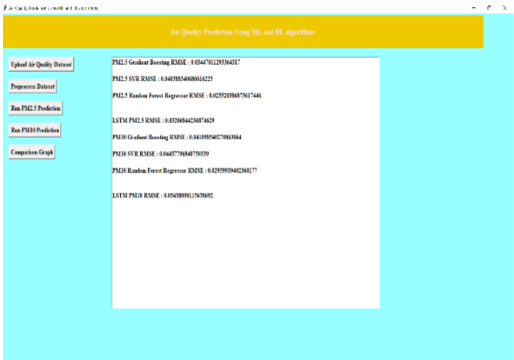
The window figure was made using the graphical user interface programme Tkinter. The window has five buttons, including Upload Air quality dataset, which the administrator or user can utilise to upload the necessary dataset. Next is the "pre-process dataset" button. It's On Click function cleans the dataset and performs attributes selections by checking the data for empty spaces, missing data, and noisy data. The "Run PM2.5 predictions" button displays the PM 2.5 values predicted by the SVR (Support vector Regression), random Forest, Gradient Boosting, and LSTM algorithms for the following 30days. Expects values are shown in the output along with forecasted values. Additionally, RMSE (root Mean Square Error).

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

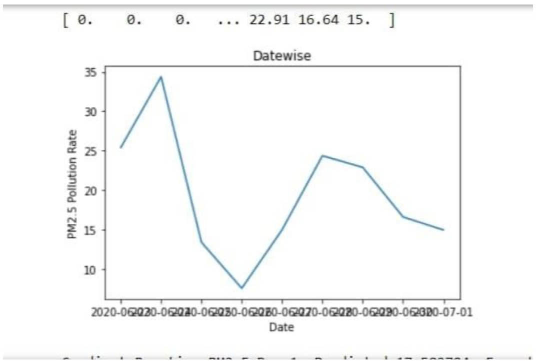


The shows the output that is obtained on clicking the button 'Upload Air quality Dataset'. The immediate action is that the user is redirected to the systems file explorer from where the dataset is uploaded and on successfully uploading the dataset it is displayed on the output screen that is on the output window.

The output screen from pressing thePm2.5 and PM10 Prediction n buttons is shown in the below figure. The results of pressing this button are the RMSE values for each of the four algorithms used to forecast values, which are shown on the window screen. The RMSE value of the Random Forest method is the lowest, and the PMSE value of the LSTM algorithm is the highest among the four, according to this result.



The graph of date vs pollution for PM2.5 values is as shown below. This graph gives us an idea about how the PM2.5 values have been so far every single day and helps us in deciphering when the values are at highest and when they are at lowest, like seasons or particular months. They also allow us to see how the patterns has been so far and predict what the future will look like if the trend continues.



VII. FUTURE SCOPE

I've began this study article with some observations and statistical regarding how I handled the dataset I got from IoT Sensors. Prediction of air quality has been increased, and the error even been reduced. I would even make an effort to develop a comparison review of air quality forecasting across several different nations. We can experiment with big datasets and improve the scalability of air quality prediction.

VIII. CONCLUSION

We learn the significance of the prediction of air quality and how it may be further developed to suit many other sectors of prediction by doing a brief study on air quality detection and prediction and its essential implementation. Suitable Machine Learning and Deep learning Algorithms are briefly examined, along with data preparation for use and training data preparation. All four methods are used to train and test the model, based on the results, the random Forest model exhibits the lowest RMSE in forecasting air quality. Thus, PM2.5 and PM10 values in the air are predicted using our improved to provide forecasts for a longer time frame.

IX. References

[1] William L Chameides, PrasadS Kasibhatla, J Yienger, and H Levy. Growth of continental-scale metro-agro-plexes, regional ozone pollution, and world food production. Science, 264(5155):74–77, 1994.

[2]David Mage, Guntis Ozolins, Peter Peterson, Anthony Webster, Rudi Orthofer, Veerle Vandeweerd, and Michael Gwynne. Urban air pollution in megacities of the world. Atmospheric environment, 30(5):681–686, 1996.48

[3] Mathilde Pascal, Gr egoire Falq, V er'ene Wagner, Edouard Chatignoux, Ma-gali Corso, Myriam Blanchard, Sabine Host, Laurence Pascal,and SophieLarrieu. Short-term impacts of particulate matter (pm10, pm10–2.5, pm2.5) on mortality in nine french cities. Atmospheric Environment, 95:175– 184, 2014.

[4] AH Legge and SV Krupa. Effects of sulphur dioxide. Air pollution and plant life, 2:135–162, 2002. 30

[5] Srinivas Devarakonda, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, and Badri Nath. Real-time air quality monitoring through mobile sensing in metropolitan areas. In Proceedings of the 2nd ACM SIGKDD international workshop on urban computing, pages 1–8, 2013. 30

[6] Mar TF, Koenig JQ. Relationship between visits to emergency departments for asthma and ozone exposure in greater Seattle, Washington. Ann Allergy Asthma Immunol. 2009; 103: 474-479 Strickland MJ, Darrow LA, et al. Short-term associations between ambient air pollutants and pediatric asthma emergency department visits. A J Respir Critical Care Med,2010, 182:307-316

[7] Lin S, Liu X, Le LH, and Hwang S-A. Chronic exposure to ambient ozone and asthma hospital admissions among children. Environ Health Perspect. 2008; 116:17251730.

[8] Islam T, McConnell R, Gauderman WJ, Avol E, Peters JM, and Gilliland F. Ozone, oxidant defense genes, and risk of asthma during adolescence. Am J Respir Crit Care Med.2009; 177(4):388-395