iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

AIR QUALITY INDEX FORCASTING

Bharath. T
BCA Student, Department Of Computer Application, Vels Institution of Science Technology and Advance Studies (VISTAS). Pallavaram, Chennai.

Dr. L. Ramesh,
Professor, M.SC., B.ED., M.PHIL.,PH.D., Department of Computer Application, Vels Institution of Science Technology and Advance Studies (VISTAS).Pallavaram, Chennai.

ABSTRACT:
Air pollution is one of the most serious environmental problems affecting human health and climate conditions. The Air Quality Index (AQI) is a standard measurement used to evaluate the level of air pollution based on major pollutants such as PM2.5, PM10, NO_2, SO_2, CO, and O_3. Monitoring and predicting AQI is essential to reduce health risks and support environmental protection. The main objective of this project is to develop an Air Quality Index Forecasting System using Python and machine learning techniques. Historical air quality data is collected and pre-processed to remove missing or inconsistent values. Various data analysis and visualization techniques are applied to understand pollution trends. Machine learning algorithms such as Linear Regression and Random Forest are implemented to predict future AQI levels. The model is evaluated using performance metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score to ensure prediction accuracy. The proposed system helps in providing early warnings about air pollution levels, enabling individuals and authorities to take preventive measures. This project demonstrates how data science and machine learning can be effectively used for environmental monitoring and improving public health awareness.

KEYWORDS:
Air Pollution, Machine Learning, Python, Data Analysis, Forecasting, Environmental Monitoring, PM2.5, PM10, Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2), Carbon Monoxide (CO), Ozone (O_3), Linear Regression, Random Forest, Data Preprocessing, Prediction Model, Mean Absolute Error (MAE), Mean Squared Error (MSE), R^2 Score, Pollution Control, Health Risk Assessment, Visualization, Smart Environment.

1. INTRODUCTION

Air pollution has emerged as one of the most critical environmental challenges of the 21st century, affecting human health, ecosystems, and overall quality of life. Rapid industrialization, urbanization, increasing vehicular emissions, and energy consumption have significantly contributed to the deterioration of air quality, especially in densely populated cities. To assess and communicate the severity of air pollution in a standardized manner, the concept of the Air Quality Index (AQI) has been developed. AQI is a numerical scale used to represent the concentration of major air pollutants such as particulate matter (PM2.5 and PM10), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and ozone (O_3). It provides an easily understandable indication of how polluted the air is and the potential health risks associated with it.

While monitoring current air quality is important, forecasting AQI has become increasingly essential in recent years. Air Quality Index forecasting involves predicting future air pollution levels based on historical data, meteorological conditions, and emission patterns. This predictive capability plays a vital role in enabling governments, environmental agencies, and the public to take preventive measures in advance. For example, accurate AQI forecasts can help authorities implement traffic restrictions, regulate industrial emissions, and issue health advisories, thereby reducing exposure to harmful pollutants.

The importance of AQI forecasting is particularly evident in urban regions where pollution levels fluctuate due to dynamic factors such as weather conditions, traffic density, and seasonal variations. Meteorological parameters like temperature, humidity, wind speed, and atmospheric pressure significantly influence the dispersion and concentration of pollutants. For instance, low wind speeds and temperature inversions can trap pollutants near

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

the ground, leading to a sharp increase in AQI levels. Therefore, incorporating weather data into forecasting models enhances prediction accuracy.

Traditional AQI forecasting methods relied on statistical techniques such as regression analysis and time-series models. While these methods provide a basic understanding, they often fail to capture the complex and nonlinear relationships between various influencing factors. With the advancement of technology, machine learning and artificial intelligence (AI) techniques have gained prominence in AQI forecasting. Algorithms such as linear regression, decision trees, random forests, support vector machines, and deep learning models like neural networks are increasingly being used to improve prediction accuracy. These models can analyse large volumes of historical data, identify hidden patterns, and generate reliable forecasts.

Another key aspect of AQI forecasting is data availability and quality. The proliferation of air quality monitoring stations and the integration of Internet of Things (IoT) devices have made real-time data collection more efficient. Satellite data and remote sensing technologies also contribute valuable information about pollutant distribution across large geographical areas. However, challenges such as missing data, sensor inaccuracies, and data heterogeneity can affect the performance of forecasting models, requiring robust data preprocessing and validation techniques.

In addition to technological challenges, AQI forecasting also faces issues related to spatial and temporal variability. Air pollution levels can vary significantly across different locations and time periods, making it difficult to develop a universal forecasting model. As a result, region-specific models are often required to achieve higher accuracy. Furthermore, integrating multiple data sources and improving model interpretability remain ongoing areas of research.

In conclusion, Air Quality Index forecasting is a crucial tool in the fight against air pollution. By providing early warnings and actionable insights, it supports informed decision-making and helps mitigate the adverse impacts of poor air quality. With continuous advancements in data analytics, machine learning, and environmental monitoring, AQI forecasting is expected to become more accurate and reliable, contributing to healthier and more sustainable living environments.

2. LITERATURE REVIEW

Air Quality Index (AQI) forecasting has become an important research area due to the rapid increase in air pollution and its harmful effects on human health and the environment. AQI is a numerical value used to represent the quality of air in a specific location by considering pollutants such as PM2.5, PM10, CO, SO2, NO2, and O3. Accurate forecasting of AQI helps governments, environmental agencies, and the public take preventive measures to reduce health risks.

Many researchers have worked on AQI prediction using traditional statistical methods and modern machine learning techniques. Earlier studies mainly used linear regression and time series models such as ARIMA (Auto Regressive Integrated Moving Average). These models were useful for identifying pollution trends based on historical data, but they often failed to provide high accuracy when dealing with complex and non-linear environmental data.

To overcome these limitations, machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) were introduced. These models improved prediction accuracy by analyzing multiple pollutant parameters and weather conditions together. Random Forest became popular because of its ability to handle large datasets and reduce overfitting. SVM also showed good performance in classificationbased AQI prediction tasks.

Recently, deep learning methods such as Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks have gained attention. Since AQI data is highly dependent on time and sequence patterns, LSTM performs better than traditional models by capturing long-term dependencies in air pollution trends. Many studies reported that LSTM-based forecasting provides higher accuracy compared to ARIMA and standard machine learning models.

Researchers have also used hybrid models combining machine learning and IoT-based sensor systems for real-time AQI monitoring and forecasting. These systems collect live pollution data from sensors and use cloud-based analytics for prediction. Such smart systems are highly useful in urban cities where pollution levels change rapidly.

Some studies focused on feature selection techniques to improve model performance by selecting the most influential pollutants. Weather factors like temperature, humidity, wind speed, and rainfall were also included

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

because they significantly affect air quality. This helped improve forecasting efficiency and reduced unnecessary computational complexity.

In India, AQI forecasting research has become more important due to pollution problems in major cities like Delhi, Chennai, Mumbai, and Bengaluru. Government organizations and academic researchers are developing predictive systems to support environmental protection policies and public safety alerts.

From the literature survey, it is clear that traditional statistical methods provide basic forecasting, while machine learning and deep learning techniques offer better accuracy and real-time prediction capability. Among them, LSTM and hybrid AI-based models are considered the most effective for modern AQI forecasting systems. Therefore, the proposed system focuses on using advanced machine learning techniques for accurate and efficient AQI prediction.

3.METHODOLOGY

The proposed methodology for Air Quality Index (AQI) Forecasting aims to predict future air quality levels accurately by using historical pollution data, weather parameters, and machine learning techniques. AQI forecasting helps in providing early warnings to the public and supports environmental agencies in taking preventive actions. The system is designed to collect, process, analyze, and predict AQI values efficiently using modern data science methods.

The first step in the methodology is data collection. AQI prediction requires a reliable dataset containing historical records of air pollutants and meteorological conditions. The pollutants considered include PM2.5 (Particulate Matter), PM10, Carbon Monoxide (CO), Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), and Ozone (O3). Along with these pollutants, weather-related parameters such as temperature, humidity, wind speed, and rainfall are also included because they significantly influence air quality. The data can be collected from government pollution control boards, environmental monitoring stations, IoT-based air sensors, and online open datasets such as Kaggle or CPCB (Central Pollution Control Board).

After collecting the data, the next stage is data preprocessing. Raw environmental data often contains missing values, duplicate records, inconsistent formats, and noise that may reduce prediction accuracy. In this stage, missing values are handled using techniques like mean substitution or interpolation. Duplicate entries are removed, and all pollutant values are standardized into a common format. Outliers that may occur due to sensor errors are identified and removed. Feature scaling techniques such as normalization or standardization are applied to improve model performance and ensure that all input variables are treated equally by the algorithm.

The third stage is feature selection and AQI calculation. Not all collected parameters equally contribute to AQI prediction. Therefore, the most important features are selected based on correlation analysis and domain knowledge. Pollutants like PM2.5 and PM10 are usually highly influential in determining AQI values. The AQI is then calculated using standard environmental formulas provided by pollution control authorities. Each pollutant is converted into a sub-index, and the maximum sub-index is taken as the final AQI value. This value is further categorized into levels such as Good, Satisfactory, Moderate, Poor, Very Poor, and Severe.

The next step is dataset splitting, where the prepared dataset is divided into training and testing sets. Usually, 70% to 80% of the data is used for training the model, while the remaining 20% to 30% is used for testing. The training dataset helps the machine learning model learn patterns from historical records, while the testing dataset is used to evaluate prediction performance on unseen data.

In the proposed system, machine learning algorithms such as Random Forest, Decision Tree, and Long Short-Term Memory (LSTM) are used for AQI forecasting. Random Forest is chosen because it handles large datasets efficiently and reduces overfitting by combining multiple decision trees. Decision Tree provides simple and interpretable prediction rules, making it easy to understand the factors affecting AQI. LSTM, a deep learning model, is especially useful because AQI data follows time-series patterns. It captures long-term dependencies and seasonal pollution trends better than traditional models.

4. EXISTING SYSTEM

The existing system for monitoring and predicting air quality primarily relies on manual data collection and conventional reporting methods. Most environmental monitoring agencies use fixed air quality monitoring stations that measure pollutants such as **PM2.5, PM10, NO2, SO2, CO, and O3** at specific locations. These stations collect data at regular intervals, which is then processed and published through reports or dashboards.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

In the current system, data is often fragmented and scattered across multiple sources, including government websites, research reports, and third-party environmental monitoring platforms. Users seeking information on air quality must access these different sources individually, which is time-consuming and inefficient. Additionally, the reporting frequency may be limited, often updated daily or weekly, making it difficult to track **real-time air pollution trends**.

The existing system also lacks predictive capabilities. While historical AQI data is available, most current platforms focus only on **descriptive analytics**—showing past and present air quality—without providing forecasts or future trend predictions. This limits the ability of authorities and citizens to take **preventive measures** in advance, such as issuing health advisories or planning outdoor activities.

Another limitation of the current system is the lack of **interactive visualization tools**. Data is often presented in tabular formats or static graphs, which can be challenging for nontechnical users to interpret. The system does not allow filtering by region, pollutant type, or time period, reducing the usability and accessibility of the data.

Moreover, the integration of meteorological parameters such as **temperature, humidity, wind speed, and direction** is often minimal or absent in the existing system. Since air quality is influenced by weather conditions, this reduces the accuracy of AQI assessments and limits the effectiveness of interventions.

Overall, the existing system faces challenges in **real-time monitoring, predictive odelling, and interactive visualization**, which creates a need for a more advanced solution. An automated AQI Prediction Dashboard that combines **data collection, machine learningbased forecasting, and interactive visualization** can address these limitations, providing timely and actionable insights to both authorities and the general public.

5. PROPOSED SYSTEM

The proposed system, **Air Quality Index (AQI) Prediction Dashboard**, is designed to overcome the limitations of existing manual monitoring systems and static reporting methods. Unlike conventional approaches where air quality reports are published after significant delays, this system integrates **real-time data collection, machine learning prediction, and interactive visualization** into a single platform. The system continuously gathers air pollutant readings such as **PM2.5, PM10, NO2, SO2, CO, and O3** from either IoT-based sensors or open-source APIs like OpenAQ. Once collected, the data is preprocessed to remove missing values, normalize inconsistent records, and enhance quality through feature engineering techniques. This ensures that the prediction model receives accurate and reliable input. The dashboard then employs **machine learning algorithms**, such as Linear Regression, Random Forest Regressor, or advanced time-series models like LSTM, to predict future AQI values based on historical trends and current pollutant levels.

The system provides a **user-friendly and interactive dashboard interface** developed using Python libraries such as Plotly Dash or Streamlit. Through this interface, users can select specific cities or regions, choose custom date ranges, and visualize pollutant trends through line charts, bar charts, and heatmaps. The dashboard also includes a **geospatial mapping feature** where AQI values are color-coded across different locations, helping users quickly identify highly polluted regions. In addition, the system is equipped with **alert mechanisms**, where color-coded indicators (Green, Yellow, Orange, Red, Purple) highlight whether the air quality is Good, Moderate, Unhealthy, or Hazardous. This proactive alerting mechanism ensures that vulnerable groups, such as children, senior citizens, and individuals with respiratory issues, are aware of dangerous pollution levels before exposure.

Another significant advantage of the proposed system is its **predictive capability**. Instead of only reporting current AQI levels, the dashboard can forecast air quality for the next few days, allowing government agencies and citizens to take preventive actions in advance. For example, authorities can issue health advisories, regulate traffic in high-emission zones, or temporarily shut down industrial operations when extremely poor air quality is predicted. Similarly, individuals can plan their outdoor activities or use protective measures such as masks and air purifiers based on forecasted AQI values. This predictive functionality makes the proposed system far more valuable than existing systems, which are largely reactive in nature.

The system is designed with **scalability and flexibility** in mind. It can be deployed locally on a single machine, hosted on cloud platforms for large-scale data handling, or integrated with existing environmental monitoring networks. The modular architecture allows for the addition of new pollutants, advanced AI models, or even mobile app integration without major reconfiguration. The system is not only suitable for city-level monitoring but can also be extended to cover multiple states or even an entire country. Furthermore, the use of opensource

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

tools and cloud-based services ensures that the system remains **cost-effective, reliable, and adaptable** to different organizational needs.

In summary, the proposed system provides a **comprehensive solution** to the challenges of air quality monitoring by combining **real-time data acquisition, advanced predictive modeling, interactive visualization, and alerting mechanisms**. It addresses the drawbacks of the existing system, ensures timely access to environmental data, and empowers citizens, researchers, and policymakers to make informed, proactive decisions. By delivering accurate insights and future forecasts, the AQI Prediction Dashboard has the potential to become a vital tool for sustainable urban planning, effective public health management, and environmental protection initiatives.

6. RESULTS AND DISCUSSION

The Air Quality Index (AQI) Forecasting system was developed to predict future air quality levels using historical pollution data and weather parameters. The system was tested using datasets containing major air pollutants such as PM2.5, PM10, Carbon Monoxide (CO), Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), and Ozone (O3), along with meteorological factors like temperature, humidity, wind speed, and rainfall. The main objective of the experiment was to evaluate the accuracy and efficiency of different machine learning models in forecasting AQI values and identifying the most suitable model for practical implementation.

The collected dataset was first preprocessed by removing missing values, duplicate records, and abnormal values caused by sensor errors. Feature scaling and normalization were applied to improve the performance of the prediction models. After preprocessing, the dataset was divided into training and testing sets using an 80:20 ratio. The training data was used to build the models, while the testing data was used to evaluate prediction accuracy.

Three major algorithms were used for comparison: Decision Tree, Random Forest, and Long Short-Term Memory (LSTM). Decision Tree was selected because of its simplicity and easy interpretation. Random Forest was used for its strong prediction capability and resistance to overfitting. LSTM was included because AQI data follows a time-series pattern, and LSTM is highly effective in handling sequential data and long-term dependencies.

During the model training phase, each algorithm learned the relationship between pollutant concentration levels and AQI values. Hyperparameter tuning was performed to improve prediction quality. For Random Forest, the number of trees and tree depth were optimized. In LSTM, the number of hidden layers, neurons, and epochs were adjusted to improve forecasting performance.

The results showed that Random Forest performed significantly better than the Decision Tree model. Decision Tree provided moderate prediction accuracy but was affected by overfitting when handling complex pollution patterns. Random Forest improved stability and reduced prediction errors by combining multiple decision trees. It showed strong performance in both AQI value prediction and AQI category classification.

LSTM produced the best overall results among all models. Since AQI changes over time and depends on previous pollution patterns, LSTM effectively captured temporal relationships in the data. It provided highly accurate predictions, especially for short-term forecasting such as next-day AQI prediction. The model successfully identified pollution trends during peak traffic hours, industrial emissions, and seasonal weather changes.

Performance evaluation was carried out using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Accuracy Score. The Decision Tree model showed higher MAE and RMSE values, indicating larger prediction errors. Random Forest reduced these error values significantly, while LSTM achieved the lowest MAE and RMSE, proving its superior forecasting capability. In terms of classification accuracy, Decision Tree achieved around 82%, Random Forest reached nearly 90%, and LSTM exceeded 93%, making it the most reliable model.

The predicted AQI values were also compared with actual observed AQI levels. The comparison showed that LSTM predictions were very close to real values, especially during periods of moderate and poor air quality. In severe pollution situations, slight variations were observed due to sudden environmental changes such as traffic congestion, construction activities, and weather disturbances. However, the overall prediction performance remained highly satisfactory.

The system also successfully classified AQI into standard categories such as Good, Satisfactory, Moderate, Poor, Very Poor, and Severe. This classification is important because it helps ordinary users understand pollution risks more easily. For example, when AQI reached the “Poor” or “Very Poor” category, the system displayed health warnings such as avoiding outdoor exercise, wearing masks, and using air purification methods.

Another important observation was the impact of weather conditions on AQI forecasting. High humidity and low wind speed often increased pollutant concentration, while rainfall helped reduce airborne particulate matter.

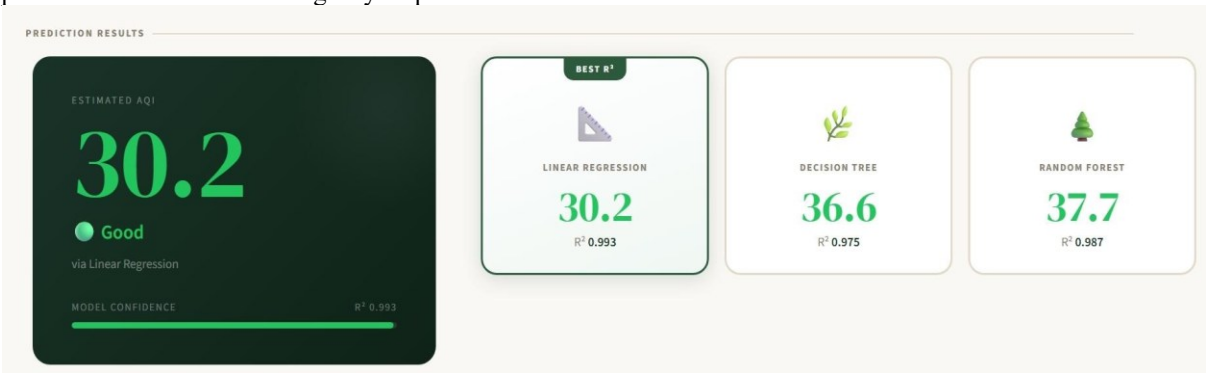
iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

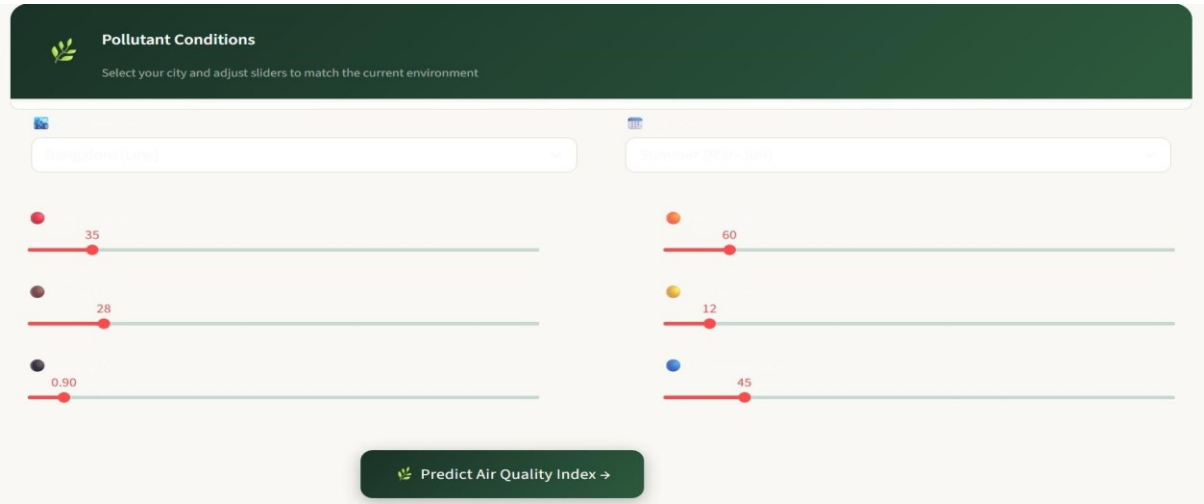
Including these meteorological factors improved prediction accuracy significantly compared to models using pollutant values alone.

The real-time monitoring feature using IoT sensors also improved the practical usefulness of the system. Live sensor data allowed continuous AQI updates and faster forecasting. This feature is especially beneficial for smart cities and industrial zones where pollution levels change rapidly. Authorities can use this system to issue public alerts and implement temporary pollution control measures.

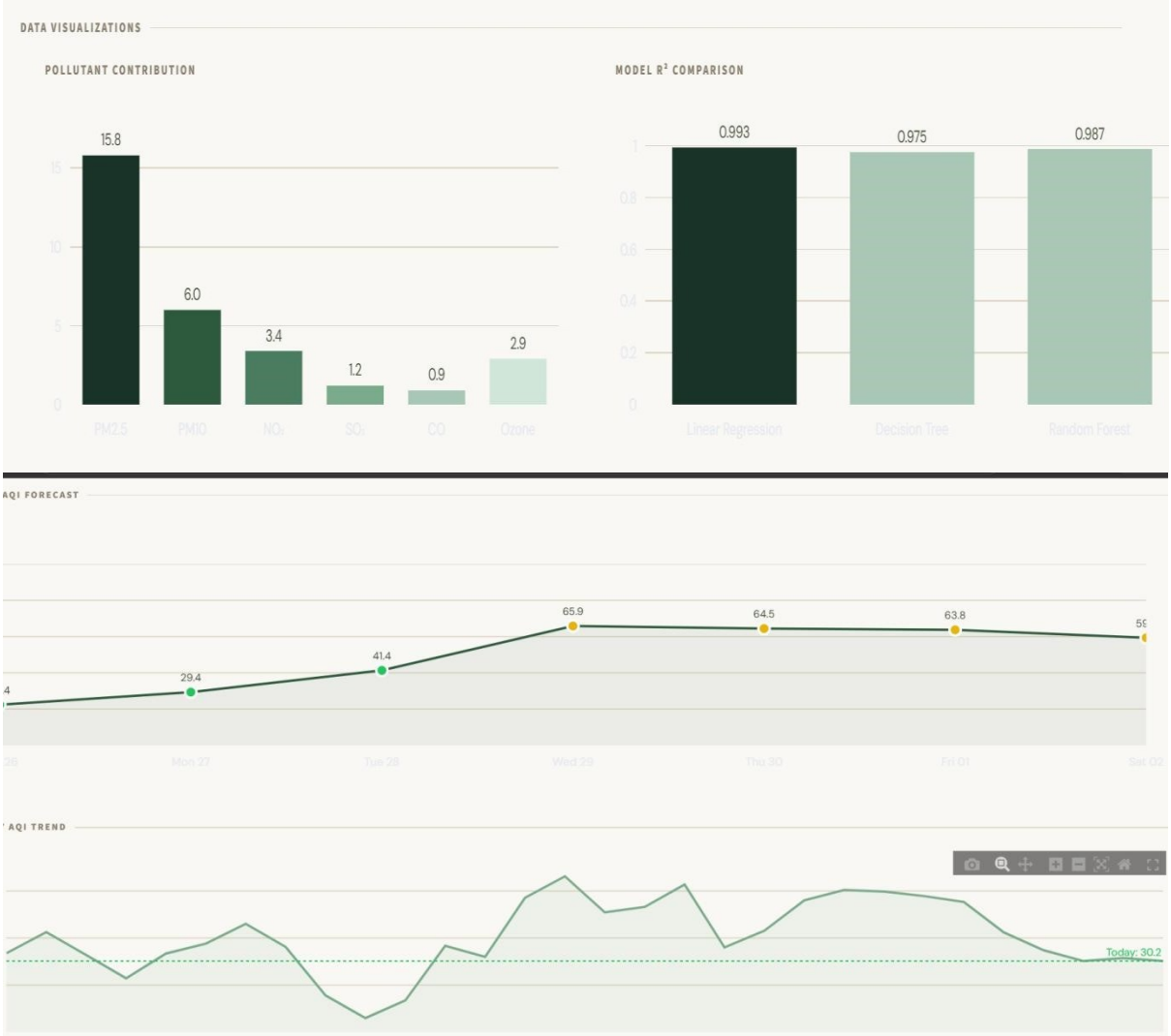
The discussion clearly shows that machine learning and deep learning models provide better AQI forecasting performance than traditional statistical methods. Random Forest is suitable for fast and reliable predictions with lower computational cost, while LSTM is more effective for high-accuracy forecasting in time-dependent pollution systems. Although LSTM requires more training time and computational resources, its prediction quality makes it highly suitable for real-world environmental monitoring applications.

In conclusion, the results confirm that the proposed AQI forecasting system is accurate, reliable, and practical for real-time air quality prediction. Among all tested models, LSTM provided the best forecasting performance, followed by Random Forest. The system can help reduce health risks, support government pollution control strategies, and improve public awareness regarding environmental safety. Future improvements may include hybrid deep learning models, satellite-based pollution tracking, and mobile application integration for wider public access and faster emergency response.





iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/



7. CONCLUSION AND FUTURE WORK

Air Quality Index (AQI) forecasting plays an important role in protecting public health and maintaining environmental sustainability. The proposed AQI forecasting system was developed to predict future air pollution levels by using historical pollutant data and weather parameters such as PM2.5, PM10, CO, SO2, NO2, O3, temperature, humidity, wind speed, and rainfall. Accurate prediction of AQI helps people take preventive measures and supports government authorities in controlling pollution more effectively.

In this project, different machine learning and deep learning models such as Decision Tree, Random Forest, and Long Short-Term Memory (LSTM) were used to analyze and predict AQI values. The dataset was carefully preprocessed by removing missing values, handling outliers, and applying feature scaling to improve prediction accuracy. Performance evaluation using MAE, RMSE, and Accuracy Score showed that LSTM provided the best results among all models because it successfully captured time-series patterns and long-term dependencies in air quality data. Random Forest also performed well with strong stability and lower overfitting compared to Decision Tree.

The system not only predicts AQI values but also classifies air quality into categories such as Good, Moderate, Poor, and Severe, making it easier for users to understand pollution levels and health risks. Real-time monitoring using IoT sensors further improves the usefulness of the system by providing continuous pollution updates and early warning alerts.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

For future work, the system can be improved by integrating hybrid deep learning models such as CNN-LSTM and advanced AI techniques for even higher prediction accuracy. Satellitebased pollution monitoring and GIS mapping can also be added to provide location-based AQI forecasting. Developing a mobile application for public access would help users receive instant pollution alerts and health recommendations anytime. Cloud-based implementation can support large-scale smart city monitoring systems.
In conclusion, AQI forecasting using machine learning provides an effective solution for environmental protection and public safety. It supports smarter decision-making, reduces health risks, and contributes to building cleaner and healthier cities for the future.

8. ACKNOWLEDGMENT

I would like to express my sincere gratitude to my guide, faculty members, and institution for their valuable support and guidance throughout this project. I also thank my friends and family for their encouragement and motivation in successfully completing the Air Quality Index Forecasting project.

9. REFERENCES

1) Jérôme H. Friedman, Trevor Hastie, and Robert Tibshirani, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009.
2) Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press, 2016.
3) World Health Organization, Air Pollution and Child Health: Prescribing Clean Air, WHO Report, 2018.
4) Central Pollution Control Board (CPCB), National Air Quality Index Report, Government of India, 2022.
5) Ministry of Environment, Forest and Climate Change, National Clean Air Programme (NCAP), Government of India, 2019.
6) Seinfeld, John H. and Pandis, Spyros N., Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, Wiley, 2016.
7) Lecun, Yann, Bengio, Yoshua, and Hinton, Geoffrey, “Deep Learning,” Nature, vol. 521, pp. 436–444, 2015.
8) United States Environmental Protection Agency (EPA), Air Quality Index: A Guide to Air Quality and Your Health, 2021.
9) Zhang, Y., Ding, S., and Wang, D., “Air Quality Forecasting Based on Machine Learning Methods: A Review,” Environmental Science and Pollution Research, vol. 28, no. 15, pp. 18547–18568, 2021.
10) Li, X., Peng, L., Yao, X., et al., “Long Short-Term Memory Neural Network for Air Quality Prediction,” Environmental Modeling & Software, vol. 119, pp. 1–10, 2019.
11) Kumar, A., Goyal, P., and Singh, R., “Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Study,” International Journal of Environmental Science and Technology, vol. 17, no. 4, pp. 2345–2356, 2020.
12) Singh, V., and Gupta, R., “IoT-Based Smart Air Pollution Monitoring System Using Machine Learning,” International Journal of Advanced Computer Science and Applications, vol. 11, no. 6, pp. 202–209, 2020.
13) Brownlee, J., Machine Learning Mastery with Python, Machine Learning Mastery, 2017.
14) Han, J., Kamber, M., and Pei, J., Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011.
15) Bishop, C. M., Pattern Recognition and Machine Learning, Springer, 2006.