Synopsis

On

"Evaluation of Machine Learning and Deep Learning Models for Daily Air Quality Index Prediction"

Submitted By

Shreya G. Bhende

Guided by

Prof. S. V. Raut

Prof. D. G. Ingale
Project In-charge

Dr. R. V. Durge
Head of Department

Dr.Rajendra Gode Institute of Technology & Research Amravati

2025-2025

Introduction:

Air pollution has emerged as one of the most critical environmental and public health challenges in the 21st century. Rapid urbanization, industrialization, and the increasing number of vehicles have led to a sharp rise in pollutant emissions, causing deterioration in air quality across many cities worldwide. Poor air quality has been directly linked to respiratory diseases, cardiovascular issues, and reduced life expectancy, making its monitoring and prediction a vital aspect of sustainable development and public health planning. The **Air Quality Index (AQI)** is a standardized measure used globally to indicate air pollution levels and communicate associated health risks to the general population. It aggregates data from multiple pollutants such as **particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO), and ozone (O3)** into a single numerical scale. This helps governments and agencies provide daily advisories, health warnings, and take precautionary measures to mitigate pollution-related hazards. Traditional statistical approaches for AQI forecasting often fail to capture the non-linear and complex interactions between pollutants and environmental factors. In recent years, **Machine Learning (ML)** techniques such as Linear Regression, Support Vector Machines (SVM), and Random Forests have been employed to model these relationships more effectively. While these models have shown promising results, they often struggle with time-series dependencies inherent in AQI data.

Aim & Objectives:-

1. To develop and evaluate machine learning (ML) and deep learning (DL) models for accurate prediction of daily Air Quality Index (AQI) in Delhi.
2. To analyze the performance of various ML/DL algorithms (e.g., Random Forest, XGBoost, LSTM, CNN) for AQI forecasting.
3. To utilize historical air quality, weather, and pollutant concentration data for model training.
4. To assess the predictive capabilities of models in handling seasonal variation and pollution spikes.
5. To provide actionable air quality predictions that can aid policy-making and public awareness.

To ensure real-time prediction capability for better public health response.

Scope:-

The scope of this project encompasses the development and implementation of a predictive analytics model utilizing deep learning techniques for stock market forecasting and trend analysis. The project focuses on:

1. **Data Sources**: Historical AQI data from CPCB, weather data from IMD, and other open datasets.
2. **Comparative Analysis**: Performance comparison between traditional ML models and advanced DL architectures.
3. **Urban Relevance**: Specific case study on Delhi, which faces severe air pollution issues.
4. **Forecast Horizon**: Short-term predictions (next 24–72 hours) for real-time applicability.
5. **Public Utility**: Results can be integrated into mobile/web apps for public alerts.
6. **Policy Impact**: Helping government and organizations create better pollution control strategies.

Software Requirements:-

1. **System**: Windows/Linux
2. **Programming Languages**: Python (Pandas, NumPy, Scikit-learn, TensorFlow, Keras, PyTorch)
3. **Data Handling**: SQL / CSV datasets
4. **Visualization**: Matplotlib, Seaborn, Plotly
5. **Development Tools**: Jupyter Notebook, Google Colab, PyCharm
6. **Version Control**: Git/GitHub

Hardware Requirement:-

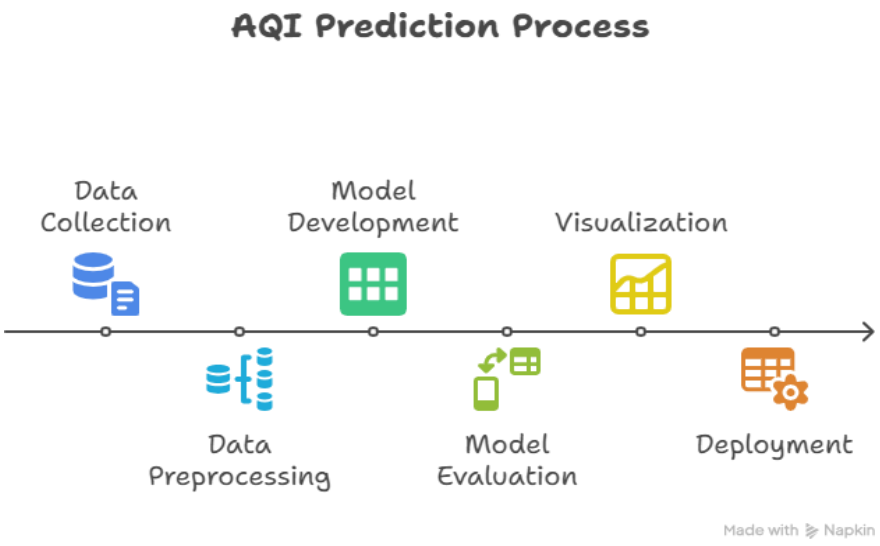
1. **Processor**: Intel i5 or above
2. **RAM**: 8 GB minimum (16 GB recommended for DL)
3. **GPU**: NVIDIA GPU with CUDA support (for DL model training)
4. **Storage**: Minimum 500 GB HDD/SSD
5. Internet connectivity for dataset access

Technology Used:

1. Machine Learning (Random Forest, Gradient Boosting, Support Vector Regression)
2. Deep Learning (LSTM, GRU, CNN)
3. Data Preprocessing & Feature Engineering
4. Time Series Forecasting Techniques
5. Statistical Performance Metrics (RMSE, MAE, R^2)
6. Data Visualization and Dashboard Development

Working

1. **Data Collection**: Gather AQI and weather data from CPCB and IMD.
2. **Data Preprocessing**: Handle missing values, normalize features, and perform feature selection.
3. **Model Development**: Train ML models like Random Forest and XGBoost. Train DL models like LSTM and CNN for sequential prediction.
4. **Model Evaluation**: Compare models using statistical accuracy metrics. Analyze seasonal and pollution spike handling capability.
5. **Visualization**: Create graphs and charts showing predicted vs. actual AQI values.
6. **Deployment**: Create a web-based dashboard for real-time AQI prediction display.
    - LSTMs are a type of Recurrent Neural Network (RNN) specialized for time-series forecasting.
    - They retain memory of past trends and can identify long-term dependencies in stock price movements.



Application:-

1. Government agencies for pollution control planning.
2. Health advisories for sensitive populations.
3. Urban planning and traffic management.
4. Academic research in environmental data science.
5. Integration into mobile applications for public awareness.

Data-driven policy-making in environmental health

Reference -:

1. Central Pollution Control Board (CPCB) – Air Quality Data.
2. IMD (Indian Meteorological Department) – Weather Data.
3. Brownlee, J. (2018). Deep Learning for Time Series Forecasting.
4. Li, X., et al. (2022). "Air quality prediction using machine learning and deep learning models."
5. WHO Reports on Global Air Pollution, (2022)