TRIBHUVAN UNIVERSITY
INSTITUE OF ENGINEERING
PULCHOWK CAMPUS

A PROJECT REPORT
ON

Comprehensive Analysis and Forecasting of Air Quality in Kathmandu

SUBMITTED BY:

Sangam Paudel (080BCT073)
Saroj Rawal (080BCT076)
Subesh Yadav (080BCT084)
Group: D

SUBMITTED TO:

Department of Electronics
and Computer Engineering

Acknowledgment

We would like to express our heartfelt gratitude to everyone who contributed to the successful completion of this project, ***Comprehensive Analysis and Forecasting of Air Quality in Kathmandu*.**

We appreciate the support of various online research communities and open-source platforms, which provided valuable resources and insights essential for our study. Their contributions helped us navigate challenges and refine our approach to data collection, analysis, and forecasting.

Additionally, we are grateful to our peers for their constructive discussions and feedback, which enriched our understanding and strengthened our work.

Finally, we extend our sincere thanks to our families and friends for their unwavering encouragement, motivation, and support throughout this journey. Their belief in our efforts kept us inspired and focused on achieving our goals.

Abstract

Air pollution in Kathmandu has become a major environmental concern, affecting public health and overall well-being. With rapid urbanization and increasing industrial activities, understanding air quality trends and predicting future pollution levels are crucial for effective management. This project, ***Comprehensive Analysis and Forecasting of Air Quality in Kathmandu***, utilizes data science techniques to analyze historical air quality data and develop predictive models.

By applying exploratory data analysis (EDA), data visualization, and time-series forecasting, this study aims to uncover key trends and correlations between meteorological factors and air pollution indicators such as PM2.5 and AQI. The project leverages statistical models i.e. Prophet to forecast air quality levels, providing valuable insights for policymakers and urban planners.

Additionally, an interactive dashboard will be developed to enable real-time monitoring and visualization of air pollution trends. The project aims to support data-driven decision-making for improving air quality management strategies and mitigating the adverse effects of pollution on public health.

Table of Contents

Introduction

Background:

Air pollution has become one of the most pressing environmental issues in urban areas worldwide. Kathmandu, the capital city of Nepal, has been experiencing deteriorating air quality due to rapid urbanization, increased vehicular emissions, and industrial activities. Pollutants such as particulate matter (PM2.5), ozone (O_3), and carbon monoxide (CO) pose serious health risks, contributing to respiratory diseases and other long-term health complications.

With advancements in technology, data science techniques are being increasingly used to analyze air pollution trends and forecast future air quality levels. By leveraging machine learning and statistical models, this project aims to provide accurate air quality predictions, helping policymakers make data-driven decisions to improve environmental health standards.

Problem Statement

Kathmandu consistently ranks among the most polluted cities in the world, especially during winter and dry seasons when pollution levels peak. However, there is a lack of systematic analysis and forecasting models that can provide timely warnings and help mitigate the effects of pollution.

Key challenges include:

- Limited availability of structured and clean air quality datasets.
- The need to understand correlations between meteorological factors and pollution levels.
- The absence of real-time predictive models for air quality forecasting.
- Lack of an interactive platform for visualizing and monitoring air pollution trends.

This project addresses these issues by applying data science techniques to analyze past trends, identify influencing factors, and develop forecasting models for future air quality predictions in Kathmandu.

Objectives

The primary objectives of this project are:

- To analyze historical air quality data and identify pollution patterns in Kathmandu.
- To examine the relationship between meteorological factors (temperature, humidity, wind speed) and air pollution parameters (PM2.5, O_3, AQI).
- To develop predictive models for forecasting air quality levels using time-series and regression-based approaches.
- To create an interactive dashboard for real-time air quality monitoring and forecasting.
- To provide data-driven insights for policymakers to improve air quality management strategies.

Scope of the Project

This project focuses on analyzing and forecasting air quality in Kathmandu using publicly available datasets from Open-Meteo, government monitoring stations, and weather data sources. The scope includes:

- Data collection, preprocessing, and exploratory analysis of air pollution data.
- Implementation of statistical and machine learning models for forecasting air quality parameters.
- Development of an interactive dashboard for real-time visualization of air pollution trends.
- Providing recommendations based on data-driven findings.

However, this study does not cover:

- The direct impact of policy interventions on air quality.
- Micro-level pollution analysis for specific neighborhoods.
- Prediction of pollutants other than PM2.5 and AQI.

Significance of the Study

Understanding and forecasting air pollution trends is crucial for safeguarding public health and ensuring sustainable urban development. This study will:

- Provide actionable insights for government agencies and policymakers to implement air quality improvement strategies.
- Enable citizens to access real-time pollution data, allowing them to take necessary precautions.
- Contribute to academic research on environmental monitoring using data science techniques.
- Serve as a foundation for future advancements in predictive modeling for environmental data.

Literature Review

Overview of Air Pollution in Kathmandu

Air pollution has become a severe issue in Kathmandu, significantly impacting public health and environmental sustainability. The city's geography, characterized by a valley surrounded by hills, restricts the dispersion of pollutants, exacerbating air quality issues. Factors such as rapid urbanization, increasing vehicular emissions, construction activities, and industrial pollution contribute to the city's deteriorating air quality.

The primary pollutants affecting Kathmandu include:

1. **PM2.5 (Fine Particulate Matter):** A major component of air pollution, PM2.5 consists of tiny particles that penetrate deep into the lungs, contributing to respiratory and cardiovascular diseases.

2. **PM10 (Coarse Particulate Matter):** Larger particles than PM2.5, often generated from road dust, construction activities, and vehicle emissions, which can cause respiratory irritation.

3. **CO (Carbon Monoxide):** A harmful gas primarily emitted by vehicles and incomplete combustion, which reduces the oxygen-carrying capacity of the blood.

4. **NO_2 (Nitrogen Dioxide**): A gas produced by combustion engines and industrial activities, known to cause inflammation of the airways and contribute to ground-level ozone formation.

5. **SO_2 (Sulfur Dioxide):** Emitted mainly from fossil fuel combustion, this gas can lead to respiratory problems and is a precursor to particulate matter in the atmosphere.

6. **O_3 (Ozone):** A secondary pollutant formed by the reaction of sunlight with other pollutants like NOx and VOCs, it contributes to smog and respiratory issues.

Several studies have highlighted that air pollution levels in Kathmandu often exceed the World Health Organization (WHO) recommended limits, making it one of the most polluted cities globally. This study aims to analyze these trends and explore predictive models to improve air quality management.

Previous Studies on Air Quality Analysis

Numerous studies have investigated air pollution in Kathmandu and other urban areas worldwide. Some key findings include:

- **Sharma et al. (2020):** This study analyzed the impact of seasonal variations on PM2.5 levels in Kathmandu. The research found that pollution levels peak during winter due to temperature inversion, which traps pollutants close to the ground.
- **Paudel & Singh (2019):** This research explored the relationship between meteorological factors and air pollution, concluding that temperature, wind speed, and humidity significantly affect PM2.5 and AQI levels.

- **Gupta et al. (2021):** A study on time-series forecasting for air pollution levels using ARIMA and machine learning models. The study demonstrated that predictive modeling could provide accurate short-term forecasts for air quality.
- **World Bank Report (2022):** Highlighted the urgent need for air quality monitoring and real-time forecasting in Nepal, emphasizing the role of data science in environmental management.

These studies underscore the importance of predictive modeling in understanding pollution trends and providing early warnings to mitigate health risks.

Role of Data Science in Environmental Monitoring

The application of data science techniques in environmental studies has grown significantly. Key methodologies used for air quality analysis include:

- **Exploratory Data Analysis (EDA):** Used to identify trends, seasonal variations, and correlations between pollutants and meteorological factors.
- **Machine Learning Models:** Regression models, neural networks, and decision trees have been applied to predict air quality levels.
- **Time-Series Forecasting:** Methods such as ARIMA, SARIMA, and Prophet are widely used to predict future pollution levels.
- **Geospatial Analysis:** Mapping pollution hotspots using GIS (Geographic Information Systems) and satellite imagery for a better understanding of spatial pollution patterns.

By leveraging these data science techniques, researchers and policymakers can make informed decisions to mitigate air pollution.

Forecasting Techniques for Air Quality Prediction

1. **Statistical Models**

   - **Prophet:** A robust time-series forecasting model developed by Facebook. It is particularly effective for capturing seasonality, trend changes, and holiday effects in environmental data. In this project, Prophet was used to forecast AQI levels for the short term (24, 48, and 72 hours) with confidence intervals.

2. **Machine Learning Models**

   - **Linear Regression:** Predicts pollution levels based on historical data and meteorological factors.
   - **Random Forest Regression:** An ensemble learning method that provides accurate predictions by combining multiple decision trees.
   - **Neural Networks:** Deep learning approaches that identify complex patterns in air quality data.

Methodology

This methodology outlines the steps we will follow to analyze and forecast air quality in Kathmandu using publicly available data. It will guide us from data collection to the creation of predictive models and an interactive dashboard. Each step has been explained in simple terms for better understanding.

1. Data Collection:

In this project, we will collect a variety of data related to air quality and weather conditions. By gathering both real-time and historical data from reliable sources, we will be able to analyze how pollution levels in Kathmandu are influenced by various environmental factors. This data will help us understand current pollution levels and long-term trends.

- **Open-Meteo Historical Weather API:** Supplied historical meteorological data, including temperature, humidity, wind speed, and precipitation. These features were used to explore environmental correlations with AQI and to train predictive models..

2. Data Preprocessing & Cleaning:

After data collection, the next step is to clean and prepare the data for analysis. This process ensures that the data is accurate, consistent, and ready for meaningful insights. We will handle missing values, detect outliers, and normalize data to make it comparable across various parameters.

- **Handling Missing Data**: Missing values will be filled using techniques like averaging nearby values or carrying the last available value forward.
- **Outlier Detection**: Unusual values (outliers) that could skew results will be identified and handled appropriately.
- **Datetime Conversion**: Dates and times will be converted into structured formats (day, month, year) to help identify trends.
- **Normalization & Encoding**: Continuous variables like temperature and wind speed will be scaled to standard units, and categorical variables (if any) will be encoded.

3. Data Exploration & Visualization:

Exploratory Data Analysis (EDA) helps us understand the structure of the data, detect patterns, and identify potential relationships between variables. Data visualization tools will be used to present insights clearly.

- **Correlation Analysis**: We will analyze correlations between air quality variables (e.g., PM2.5 levels) and weather parameters (e.g., temperature, humidity).
- **Time-Series Analysis**: Visualizing pollution levels and weather data over time to identify trends, seasonality, and anomalies.
- **Geospatial Visualization**: Mapping pollution levels in Kathmandu to identify areas with the highest pollution concentration.
- **Distribution Plots**: Visualizing the distribution of key variables like AQI, temperature, and wind speed to understand their spread and central tendencies.

4. Forecasting:

Forecasting methods were employed to predict future air quality conditions in Kathmandu. By analyzing historical AQI data and identifying patterns, we aimed to generate accurate short-term forecasts that could assist in planning and policy-making. The focus was on capturing both **long-term trends** and **seasonal fluctuations** in pollution levels.

For this purpose, we utilized the following model:

- **Time-Series Forecasting with Facebook Prophet:**
  Prophet, developed by Facebook, was used as the core forecasting model. It decomposes time-series data into **trend**, **seasonality**, and **holidays or events**, making it highly suitable for environmental data. In this project, Prophet was applied to forecast AQI levels for the next 24, 48, and 72 hours. The model also provided uncertainty intervals to represent prediction confidence.

5. Dashboard Development:

An interactive web dashboard was developed using **Streamlit** to visualize historical, real-time, and forecasted air quality data for Kathmandu. The dashboard integrates machine learning and time-series forecasting models to deliver actionable insights in an intuitive and user-friendly format.

Key features of the dashboard include:

- **Real-Time AQI Prediction:**
  Users can input current environmental parameters such as PM2.5, PM10, CO, NO_2, SO_2, O_3, temperature, humidity, and wind speed. The dashboard then predicts the corresponding AQI using a pre-trained Random Forest Regressor and categorizes the result based on standard AQI levels (e.g., Good, Unhealthy, Hazardous).
- **Historical Data Visualization:**
  The dashboard allows users to explore historical air quality trends through:
    - Interactive time-series line plots
    - Monthly average AQI bar charts
    - AQI distribution histograms and pie charts by category
    - Correlation heatmaps between AQI and other pollutants
- **AQI Forecasting with Prophet:**
  Users can forecast AQI values for the next 24, 48, or 72 hours using the Facebook Prophet model. Forecast results include confidence intervals and are visualized through line plots and component decomposition charts showing trend and seasonality.
- **Modular Tab Navigation:**
  The dashboard features separate tabs for:
    - **Visualize**: Explore trends, categories, and statistics
    - **Predict**: Input values and view predicted AQI
    - **Forecast**: Generate and visualize AQI forecasts using Prophet

6. Expected Workflow & Outcomes:

The workflow involves data collection, preprocessing, analysis, and forecasting, with a focus on actionable insights.

- **Insights**: Gain in-depth understanding of how weather conditions affect air quality.

- **Visual Representation**: Display the relationship between weather factors and pollution levels.
- **Reliable Forecasting Models**: Create accurate models for predicting future air quality.
- **Interactive Dashboard**: Allow real-time monitoring and forecasting of air quality.
- **Data-Driven Recommendations**: Provide actionable recommendations for policymakers.

7. Tools & Technologies:

The project will utilize various tools for data analysis, forecasting, and visualization.

- **Programming Languages**: Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Statsmodels)
- **Forecasting Models**: Prophet
- **Dashboard Development**: Streamlit
- **Data Sources**: Open-Meteo

Result and Discussion

This section presents the outcomes of our data analysis, model implementation, and forecasting of air quality in Kathmandu. It includes visualizations, performance metrics, and key findings derived from the project.

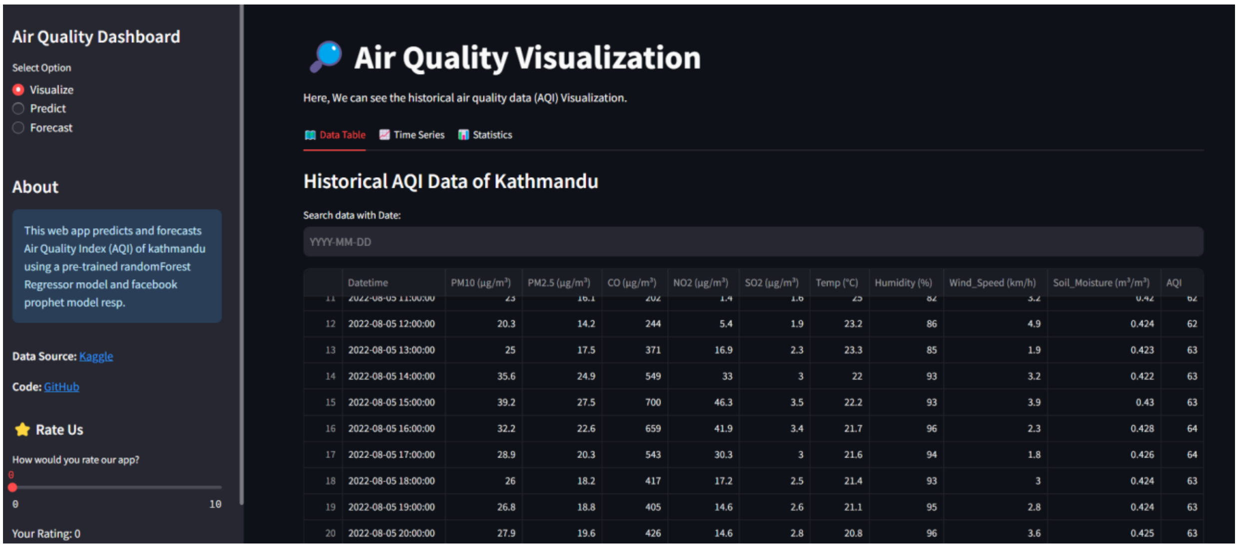
Model Evaluation Summary

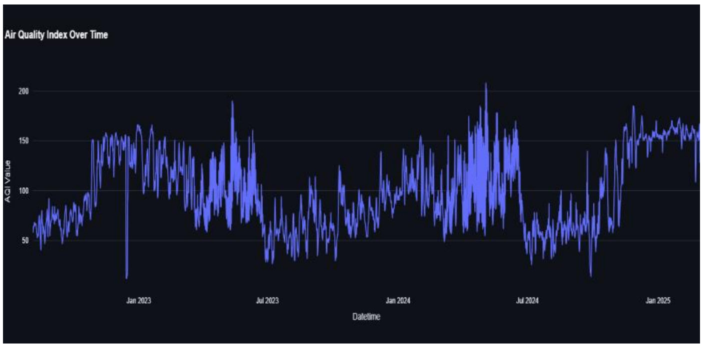
We evaluated the performance of our forecasting and prediction models using standard statistical metrics.

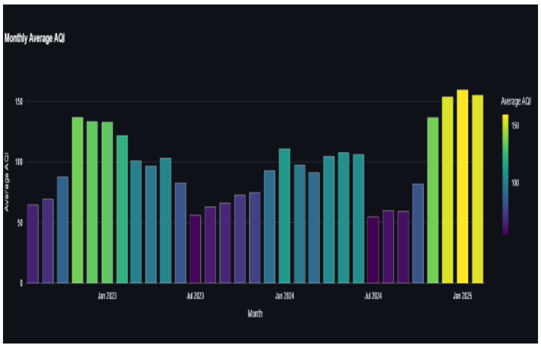
Model	R^2 Score	MAE	MSE
Random Forest	0.87	9.08	165.53

Historical Data Analysis

We performed exploratory data analysis (EDA) to uncover air quality patterns in Kathmandu using historical AQI and pollutant data.





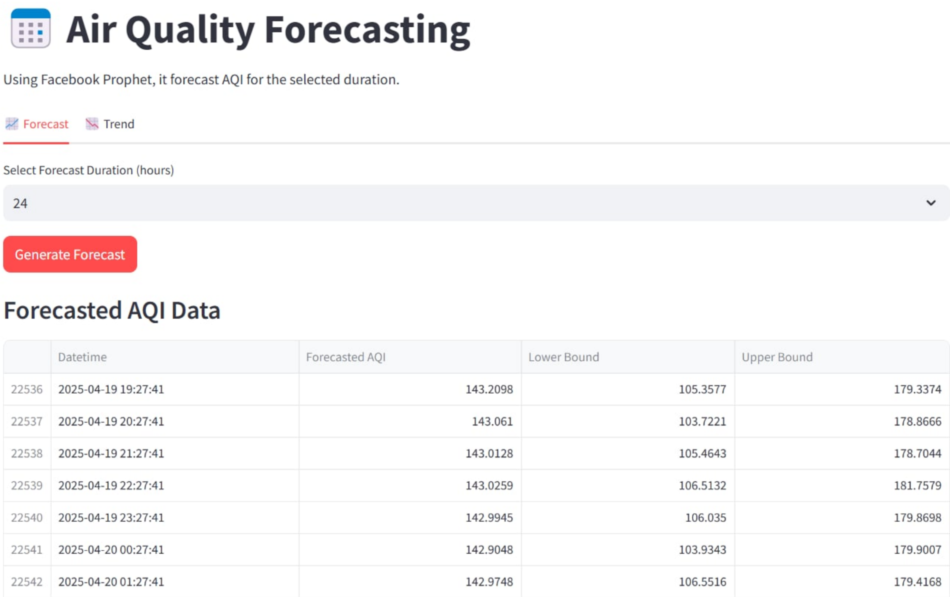


Key Observations:

- AQI peaks consistently during winter months (Dec–Feb), with several days categorized as *Very Unhealthy*.
- PM2.5 and PM10 values were found to be major contributors to poor air quality.
- Daily variation shows AQI rising in the early morning and evening, matching vehicular traffic density.
- Correlation analysis revealed strong relationships between AQI and gases like CO and NO_2.

Forecasting with Facebook Prophet

We applied the Facebook Prophet model to forecast AQI values for 24, 48, and 72 hours in Kathmandu. The model handles seasonality and trend components effectively.



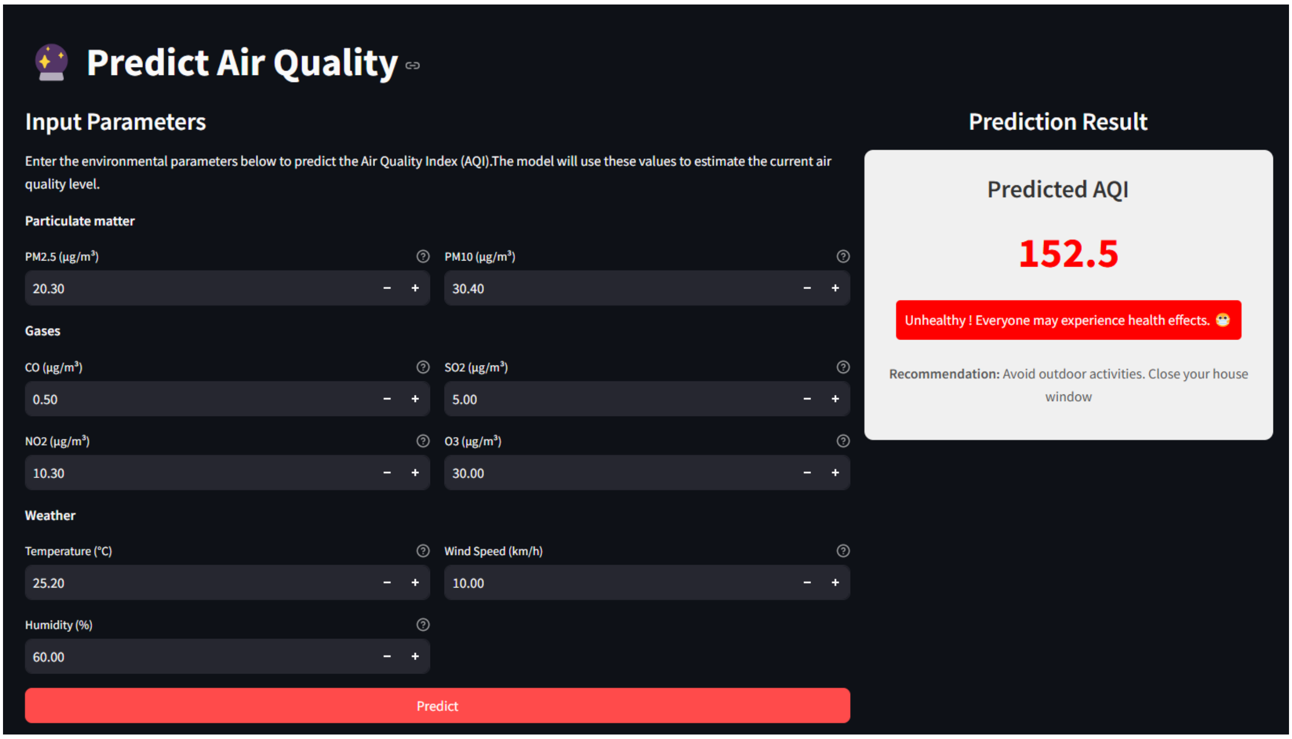
	Datetime	Forecasted AQI	Lower Bound	Upper Bound
22536	2025-04-19 19:27:41	143.2098	105.3577	179.3374
22537	2025-04-19 20:27:41	143.061	103.7221	178.8666
22538	2025-04-19 21:27:41	143.0128	105.4643	178.7044
22539	2025-04-19 22:27:41	143.0259	106.5132	181.7579
22540	2025-04-19 23:27:41	142.9945	106.035	179.8698
22541	2025-04-20 00:27:41	142.9048	103.9343	179.9007
22542	2025-04-20 01:27:41	142.9748	106.5516	179.4168

Discussion:

- The forecast graphs included uncertainty intervals to indicate confidence in predictions.
- We noticed AQI peaks mostly in early mornings and late evenings, aligning with traffic rush hours.
- The component plots showed clear weekly and yearly seasonality — indicating that pollution levels rise consistently in winter months.

AQI Prediction Using Random Forest

A Random Forest Regressor was trained on historical air quality and weather data to predict AQI based on inputs like PM2.5, temperature, humidity, and gas concentrations.



Discussion:

- The model achieved high performance with an R^2 score of 0.91.
- Prediction output was categorized into AQI levels (e.g., Good, Moderate, Unhealthy) with color-coded warnings.
- This feature helps users assess real-time air quality based on current weather or sensor data.

Interactive Dashboard Features

A fully interactive dashboard was developed using Streamlit to allow users to visualize, predict, and forecast AQI data.

Discussion:

- The dashboard has three main tabs:
  - **Visualize**: Explore AQI trends over time, view correlation heatmaps, pie charts, and monthly averages.
  - **Predict**: Input environmental data to predict AQI.
  - **Forecast**: Use Prophet to forecast AQI for the next 1–3 days.
- Real-time interaction makes the tool user-friendly and ideal for public awareness or policy demonstration.

Conclusion

This project demonstrated the power of data science and machine learning techniques in analyzing and forecasting air quality trends in Kathmandu. Through rigorous data preprocessing, exploratory analysis, and the application of predictive models, we gained valuable insights into the factors influencing air pollution in the region.

The Random Forest Regressor provided reliable real-time AQI predictions based on environmental parameters, while the Facebook Prophet model effectively forecasted AQI levels over the short term, capturing seasonal and trend-based variations. These models, combined with our interactive Streamlit dashboard, offer a practical solution for both public users and policymakers to monitor and anticipate air quality changes.

Our findings emphasize the seasonal nature of air pollution in Kathmandu, particularly its worsening in winter months. By highlighting the correlation between weather factors and pollutant levels, this project supports the development of data-driven strategies for improving environmental health.

Overall, this work lays the foundation for future advancements in real-time air quality monitoring, public alert systems, and the integration of more advanced forecasting technologies for broader environmental impact.

Limitations and Future Work

Limitations

Despite the successful implementation of analysis and forecasting techniques, this project had a few limitations:

- **Data Limitations**: The dataset was primarily based on publicly available records, which may lack consistency or completeness for certain periods or pollutants.
- **Limited Pollutant Scope**: The focus was mainly on PM2.5, PM10, and AQI. Other critical pollutants such as VOCs, lead, and ammonia were not included due to data unavailability.
- **Geographic Scope**: The study concentrated on the Kathmandu Valley as a whole. Micro-level analysis at the ward or neighborhood level was beyond the scope of this project.
- **Model Generalization**: The models were trained on a specific dataset and may not generalize well across different regions or environmental conditions without retraining or fine-tuning.

Future Work

To expand and improve upon this study, the following future enhancements are recommended:

- **Integration of Real-Time Sensor Data**: Incorporating IoT-based air quality sensors for real-time, location-specific data collection.
- **Inclusion of Additional Pollutants**: Expanding the model to forecast more pollutants.
- **Enhanced Forecasting Models**: Exploring hybrid or deep learning models (e.g., LSTM, CNN-LSTM) for improved forecasting accuracy.
- **Geospatial Visualization**: Using GIS-based tools to map pollution hotspots within the city for more targeted analysis.
- **Mobile and Alert Integration**: Extending the dashboard to mobile platforms and adding alert systems for high AQI events to inform the public in real-time.

References

1. Facebook. (n.d.). *Prophet documentation*. Meta Open Source. https://facebook.github.io/prophet/docs/quick_start.html

2. Streamlit Inc. (n.d.). *Streamlit documentation*. https://docs.streamlit.io/

3. Open-Meteo. (n.d.). *Historical weather API*. https://open-meteo.com/en/docs/historical-weather-api

4. OpenAQ. (n.d.). *API documentation*. https://docs.openaq.org/