DISASTER MANAGEMENT (CE8.401)

PROJECT REPORT

LIVE AQI DATA & DATA PREDICTIONS OF INDIAN CITIES

Submitted to :

Prof. Jofin George

Prof. Shubham Singhal

Submitted by :

TEAM 13

Harshit Kashyap : 2022201050

Priyank Mahour : 2022201047

Ujjwal Prakash : 2022202009

Garima Tripathi : 2021101123

ABSTRACT

Our project focuses on advancing air quality monitoring and assessment methodologies through the integration of machine learning techniques and image recognition models. Leveraging real-time data on pollutant concentrations from various Indian cities, we developed predictive models to estimate Air Quality Index (AQI) values, providing insights into air pollution levels. Additionally, we employed ImageInception V3, a pre-trained convolutional neural network, to analyze visual cues extracted from images and assess pollution severity. Through meticulous data collection, preprocessing, and model development, our project offers a comprehensive framework for evaluating air quality dynamics. Our findings contribute to the advancement of environmental monitoring initiatives and facilitate evidence-based decision-making to mitigate air pollution and safeguard public health.

TABLE OF CONTENTS

LIST OF FIGURES :

INTRODUCTION:

Air pollution is a pressing environmental issue with significant implications for public health and the ecosystem. It encompasses the presence of harmful substances in the Earth's atmosphere, often resulting from human activities such as industrial processes, vehicular emissions, and agricultural practices. The detrimental effects of air pollution on human health, including respiratory diseases, cardiovascular disorders, and even premature death, underscore the urgent need for effective monitoring and management strategies.

What is Air Pollution?

Air pollution refers to the contamination of the air with harmful substances, known as pollutants, which can have adverse effects on human health, the environment, and the climate. These pollutants can be of natural origin, such as dust and pollen, or anthropogenic, arising from human activities like burning fossil fuels and industrial emissions.

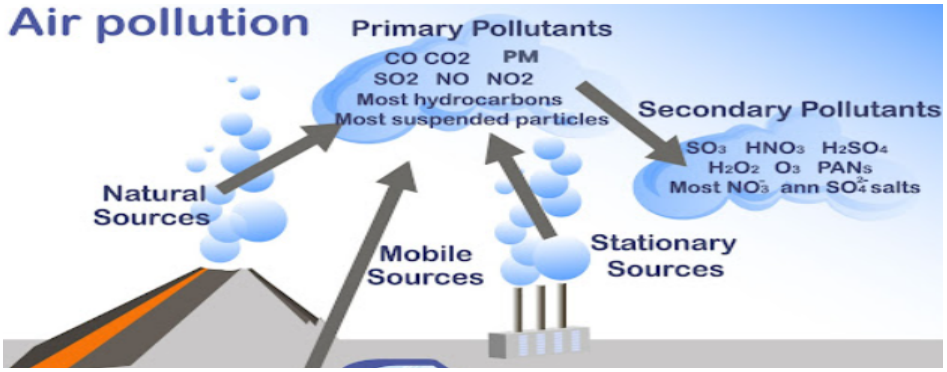
What is Air Quality Index (AQI)?

The Air Quality Index (AQI) is a numerical scale used to measure and communicate the quality of ambient air in a specific location. It provides a standardized way to assess the concentration levels of various pollutants in the air and their potential health effects on individuals. The AQI typically ranges from 0 to 500, with lower values indicating good air quality and higher values corresponding to increasingly hazardous conditions.

The Air Quality Index (AQI) was officially launched in India on April 6, 2015, by the Ministry of Environment, Forest and Climate Change (MoEFCC). This initiative was part of the government's efforts to improve air quality monitoring and public awareness about air pollution across the country.

Pollutants Determining AQI:

Pollutants of Air Pollution



- Particulate Matter (PM2.5 and PM10):

Particulate matter refers to tiny particles suspended in the air, categorized based on their size. PM2.5 particles have a diameter of 2.5 micrometers or less, while PM10 particles have a diameter of 10 micrometers or less. These particles can originate from various sources, including vehicle exhaust, industrial emissions, construction activities, and wildfires. PM2.5 and PM10 can penetrate deep into the lungs and enter the bloodstream, causing respiratory and cardiovascular problems. High levels of particulate matter are associated with reduced visibility, haze, and smog.

- Ozone (O3):

Ground-level ozone is a secondary pollutant formed by chemical reactions between nitrogen oxides (NOx) and volatile organic compounds (VOCs) in the presence of sunlight. Ozone pollution is more prevalent in urban areas with high traffic congestion and industrial activity. Exposure to elevated levels of ozone can irritate the respiratory system, trigger asthma attacks, and reduce lung function. Ozone pollution is often more pronounced during the summer months and can contribute to the formation of smog.

- Nitrogen Dioxide (NO2):

Nitrogen dioxide is a reddish-brown gas primarily emitted from vehicle exhaust, industrial processes, and combustion of fossil fuels. NO2 can irritate the respiratory system, aggravate existing respiratory conditions such as asthma, and increase the risk of respiratory infections. Long-term exposure to NO2 is associated with decreased lung function and respiratory symptoms.

- Sulfur Dioxide (SO2):

Sulfur dioxide is a colorless gas produced by the combustion of fossil fuels containing sulfur, such as coal and oil. Major sources of SO2 emissions include power plants, industrial facilities, and vehicles. SO2 can irritate the respiratory system, exacerbate asthma, and contribute to the formation of particulate matter. In addition to its health effects, SO2 can also lead to acid rain, which can harm ecosystems and corrode buildings and infrastructure.

- Volatile Organic Compounds (VOCs):

Volatile organic compounds are a diverse group of chemicals that can evaporate into the air at room temperature. Common VOCs include benzene, toluene, xylene, and formaldehyde, which are emitted from sources such as vehicle exhaust, industrial processes, and household products. VOCs can react with other pollutants in the atmosphere to form ozone and secondary particulate matter. Prolonged exposure to VOCs can cause respiratory irritation, headaches, nausea, and damage to the liver, kidneys, and central nervous system.

- Ammonia (NH3):

Ammonia is a colorless gas with a pungent odor, commonly emitted from agricultural activities such as fertilizer application and livestock farming. NH3 can react with other pollutants in the atmosphere to form fine particulate matter, contributing to haze and smog. High levels of ammonia can also have adverse effects on human health, causing respiratory irritation and exacerbating asthma symptoms.

These pollutants play a significant role in air pollution and can have detrimental effects on human health, ecosystems, and the environment. Monitoring and reducing emissions of these pollutants are essential for improving air quality and protecting public health. The Air Quality Index (AQI) provides a standardized framework for assessing the concentration levels of these pollutants and communicating air quality information to the public in a clear and understandable manner. By understanding the sources and health effects of pollutants, policymakers, industries, and individuals can take proactive measures to mitigate air pollution and promote sustainable development practices.

Concentration Levels:

The concentration levels of these pollutants are measured in micrograms per cubic meter ($\mu g/m^3$) for particulate matter and parts per million (ppm) for gases like ozone, nitrogen dioxide, and sulfur dioxide. These concentration levels are then converted into a numerical value on the AQI scale, which ranges from 0 to 500, with specific breakpoints indicating different levels of health concern.

The AQI serves as a valuable tool for assessing air quality and informing the public about potential health risks associated with exposure to air pollution. By monitoring and understanding the concentration levels of key pollutants, policymakers, health professionals, and individuals can take proactive measures to mitigate the adverse effects of air pollution on public health and the environment.

Concentration levels of pollutants in the air can vary widely depending on factors such as industrial activity, vehicular emissions, geographical location, and weather conditions. Understanding what constitutes healthy and dangerous levels of pollutants is crucial for assessing air quality and its impact on human health. Here's a more detailed explanation:

Healthy Concentration Levels:

Healthy concentration levels of pollutants are those that pose minimal risk to human health and the environment. The World Health Organization (WHO) and national environmental agencies often establish air quality guidelines that define healthy concentration levels for different pollutants. These guidelines are based on scientific research and aim to protect public health by limiting exposure to harmful substances in the air.

For example, the WHO guideline for PM2.5 recommends an annual average concentration of 10 micrograms per cubic meter ($\mu g/m^3$) and a 24-hour average concentration of 25 $\mu g/m^3$. Similarly, the guideline for nitrogen dioxide (NO_2) suggests an annual average concentration of 40 micrograms per cubic meter ($\mu g/m^3$).

Dangerous Concentration Levels:

Dangerous concentration levels of pollutants exceed established air quality guidelines and can have severe health effects, especially for vulnerable populations such as children, the elderly, and individuals with pre-existing health conditions. Exposure to high levels of pollutants can lead to respiratory and cardiovascular diseases, exacerbate asthma, and increase the risk of premature death.

In Indian cities, certain regions experience dangerously high levels of air pollution, particularly during specific seasons or due to local sources of emissions. For example:

- **Delhi**:

Delhi, the capital city of India, often faces severe air pollution, especially during the winter months. PM2.5 levels in Delhi have been known to exceed 500 $\mu g/m^3$ during episodes of smog, far surpassing the WHO guidelines. High levels of vehicular emissions, industrial activity, construction dust, and crop burning in neighboring states contribute to Delhi's poor air quality.

- **Kanpur**:

Kanpur, located in the state of Uttar Pradesh, is another Indian city with consistently high levels of air pollution. PM2.5 concentrations in Kanpur frequently exceed 300 µg/m³, posing significant health risks to its residents. Industrial pollution, vehicular emissions, and biomass burning are major contributors to air pollution in Kanpur.

- **Patna**:

Patna, the capital city of Bihar, also grapples with severe air pollution, particularly during the winter months. PM2.5 levels in Patna often exceed 400 µg/m³, leading to respiratory problems and other health issues. Vehicular emissions, construction activities, and open burning of waste contribute to Patna's poor air quality.

- **Mumbai, Maharashtra:**

Mumbai, the financial capital of India, experiences moderate to poor air quality, primarily due to vehicular emissions, industrial activities, and construction dust. The city's proximity to the Arabian Sea influences its air quality, with sea breezes helping to disperse pollutants. However, during the winter months, pollution levels can spike due to atmospheric conditions and increased traffic. PM2.5 levels in Mumbai typically range from 20 to 100 µg/m³, with higher concentrations observed in densely populated areas and industrial zones.

- **Chennai, Tamil Nadu:**

Chennai, the capital city of Tamil Nadu, generally has better air quality compared to some other Indian cities. However, pollution levels can still reach unhealthy levels, particularly during the winter months when weather conditions favor the accumulation of pollutants. Vehicular emissions, industrial activity, and construction dust contribute to Chennai's air pollution. PM2.5 concentrations in Chennai typically range from 10 to 60 µg/m³, with occasional spikes during pollution episodes.

- **Bengaluru, Karnataka:**

Bengaluru, often referred to as the Silicon Valley of India, experiences moderate air pollution levels attributed to vehicular emissions, industrial activity, and construction dust. While pollution levels are lower compared to cities like Delhi and Kanpur, Bengaluru still faces challenges in maintaining air quality standards, especially in areas with heavy traffic congestion and industrial clusters. PM2.5 levels in Bengaluru typically range from 10 to 50 µg/m³, with higher concentrations observed in traffic hotspots and industrial areas.

- **Hyderabad, Telangana:**

Hyderabad, the capital city of Telangana, experiences moderate air quality, with pollution levels influenced by vehicular emissions, industrial activity, and urban development. While pollution levels in Hyderabad are lower compared to some northern cities, efforts to combat air pollution are essential to maintain healthy air quality standards. PM2.5 concentrations in Hyderabad typically range from 15 to 70 µg/m³, with variations observed across different seasons and locations within the city.

- **Kolkata, West Bengal:**

Kolkata, the capital city of West Bengal, faces challenges related to air pollution, primarily due to vehicular emissions, industrial activity, and biomass burning in nearby agricultural areas. The city's proximity to the Bay of Bengal influences its air quality, with sea breezes helping to disperse pollutants. However, pollution levels can still exceed health standards, especially during the winter months. PM2.5 concentrations in Kolkata typically range from 20 to 100 µg/m³, with higher levels observed in densely populated areas and industrial zones.

Guidelines for concentration levels of particulate matter

Guidelines

$PM_{2.5}$:	10 $\mu g/m^3$ annual mean 25 $\mu g/m^3$ 24-hour mean
PM_{10}:	20 $\mu g/m^3$ annual mean 50 $\mu g/m^3$ 24-hour mean

OBJECTIVE AND SCOPE

Objectives:

1. Real-Time Air Quality Monitoring: The primary objective of this project is to develop a web application capable of providing real-time monitoring of Air Quality Index (AQI) data in various Indian cities. Users will have access to up-to-date information on AQI levels and pollutant concentrations, allowing them to make informed decisions about outdoor activities and health precautions.

2. Predictive Modeling: Another objective is to implement predictive models that can forecast AQI levels based on historical data and current environmental factors. These models will utilize machine learning algorithms, such as deep learning and linear regression, to analyze past trends and predict future air quality conditions. This forecasting capability will enable users to anticipate changes in air quality and take preventive measures accordingly.

3. Image Recognition for Pollution Assessment: Additionally, the project aims to integrate an image recognition model, such as ImageInception, to assess the severity of pollution in images. By analyzing visual cues related to air pollution, such as haze, smog, and particulate matter, this model will classify images into different pollution categories, providing users with a visual representation of air quality conditions.

4. User Interface and Experience: The project seeks to design an intuitive and user-friendly interface for the web application, ensuring ease of navigation and accessibility for users. Interactive features, such as maps, charts, and filters, will be incorporated to enhance the user experience and facilitate data exploration.

Scope:

1. Geographical Coverage: The project will focus on providing AQI data for various cities across India, with an initial emphasis on major metropolitan areas known for air pollution issues. The geographical coverage may expand over time to include additional cities and regions based on user demand and data availability.

2. Pollutant Parameters: The web application will monitor and report on the concentration levels of key pollutants known to affect air quality, including particulate matter (PM2.5 and PM10), ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and others. These pollutants will be measured in micrograms per cubic meter (μg/m³) or parts per million (ppm), as per standard air quality monitoring protocols.

3. Model Complexity: The predictive models implemented in the project will range in complexity, from simple linear regression models to more advanced deep learning algorithms. The choice of model will depend on factors such as data availability, computational resources, and performance requirements.

5. Data Sources: The project will utilize data from reliable sources, such as government agencies, research institutions, and environmental monitoring stations. APIs or databases providing real-time or historical AQI data for Indian cities will be accessed and integrated into the web application's database for visualization and analysis.

In summary, the project aims to develop a comprehensive web application for real-time monitoring of AQI data in Indian cities, incorporating predictive modeling for AQI forecasting, image recognition for pollution assessment, and user-friendly features for enhanced engagement and usability.

METHODOLOGY

DATA COLLECTION

1) Ambee-Data External API:
   Purpose: Utilized to fetch live data on Air Quality Index (AQI) and concentration of pollutants.
   Functionality: The API provided real-time updates on AQI and concentrations of pollutants such as PM2.5, PM10, nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO), and ozone (O3).
   Integration: Incorporated the API into the project's data pipeline to continuously fetch and update live data for analysis.
2) Kaggle Datasets:
   Purpose: Used for historical data analysis and comparison with real-time data.
   Contents: Kaggle datasets included historical records of AQI and pollutant concentrations from various locations and time periods.
   Methodology: Leveraged Kaggle datasets to identify trends, patterns, and seasonal variations in air quality over time.
3) Government Websites:
   Purpose: Accessed for official air quality data and regulatory information.
   Data Availability: Government websites provided comprehensive datasets on AQI, pollutant levels, and related environmental policies and regulations.

WEB APPLICATION FOR VISUALIZATION OF POLLUTANTS

The aim of this project is to develop a user-friendly web application for visualizing pollutant concentrations and predicting the Air Quality Index (AQI) using machine learning techniques. Monitoring and predicting AQI is crucial for assessing environmental quality and its impact on public health. To achieve this goal, we utilized Streamlit for web development and implemented two machine learning models:

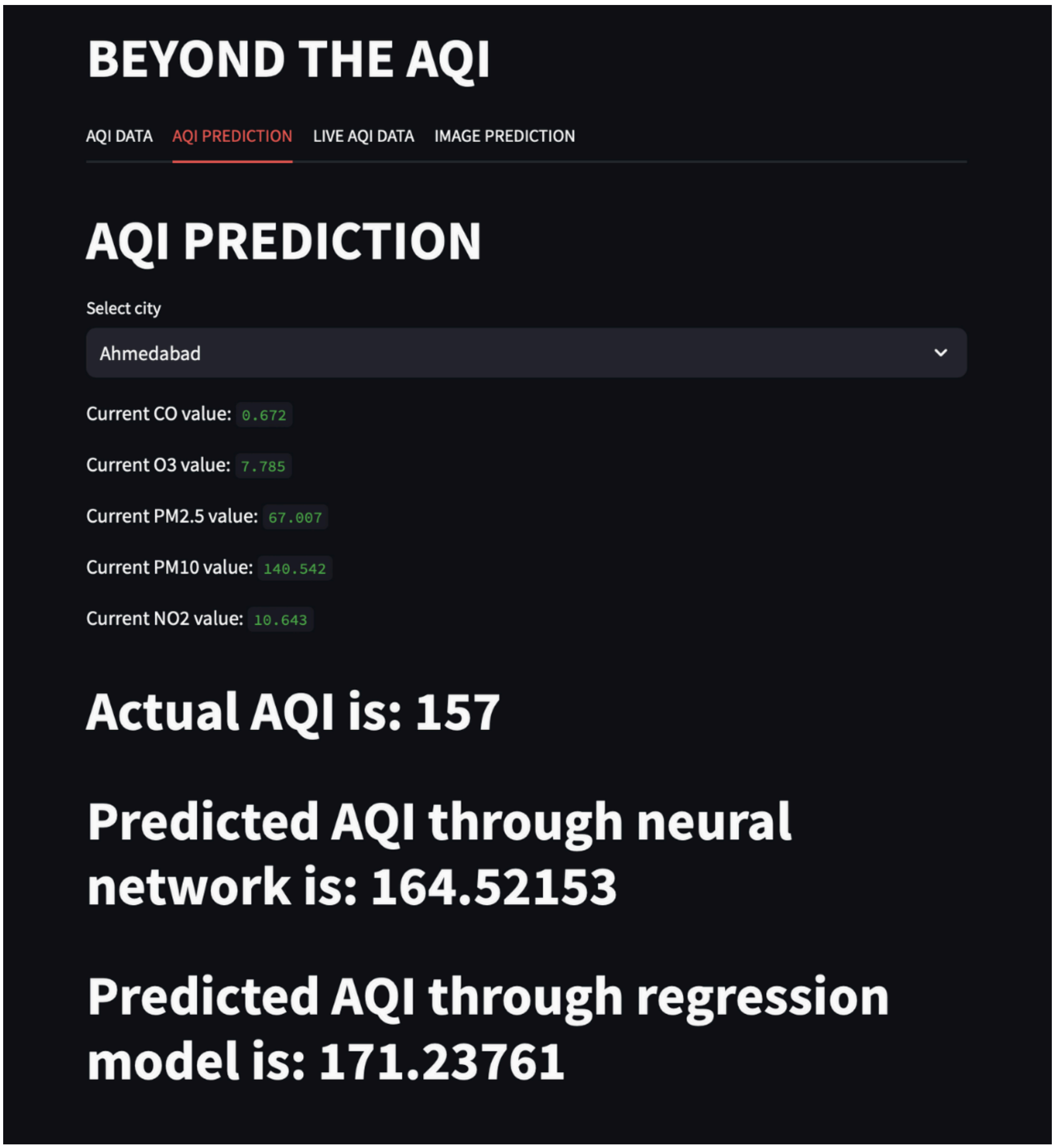
We collected data from multiple websites spanning the years 2015 to 2023, which includes pollutant concentration measurements for SO2, NO2, NOx, toluene, benzene, and xylene. The dataset underwent preprocessing steps to handle missing values and outliers, ensuring data integrity for subsequent analysis.

The Streamlit web application allows users to interactively explore historical pollutant concentration data and receive real-time updates on AQI. Features include dynamic visualizations, customizable time ranges, and informative tooltips to enhance user experience. The user interface design prioritized clarity and accessibility, enabling users of varying technical backgrounds to engage with the application effortlessly.

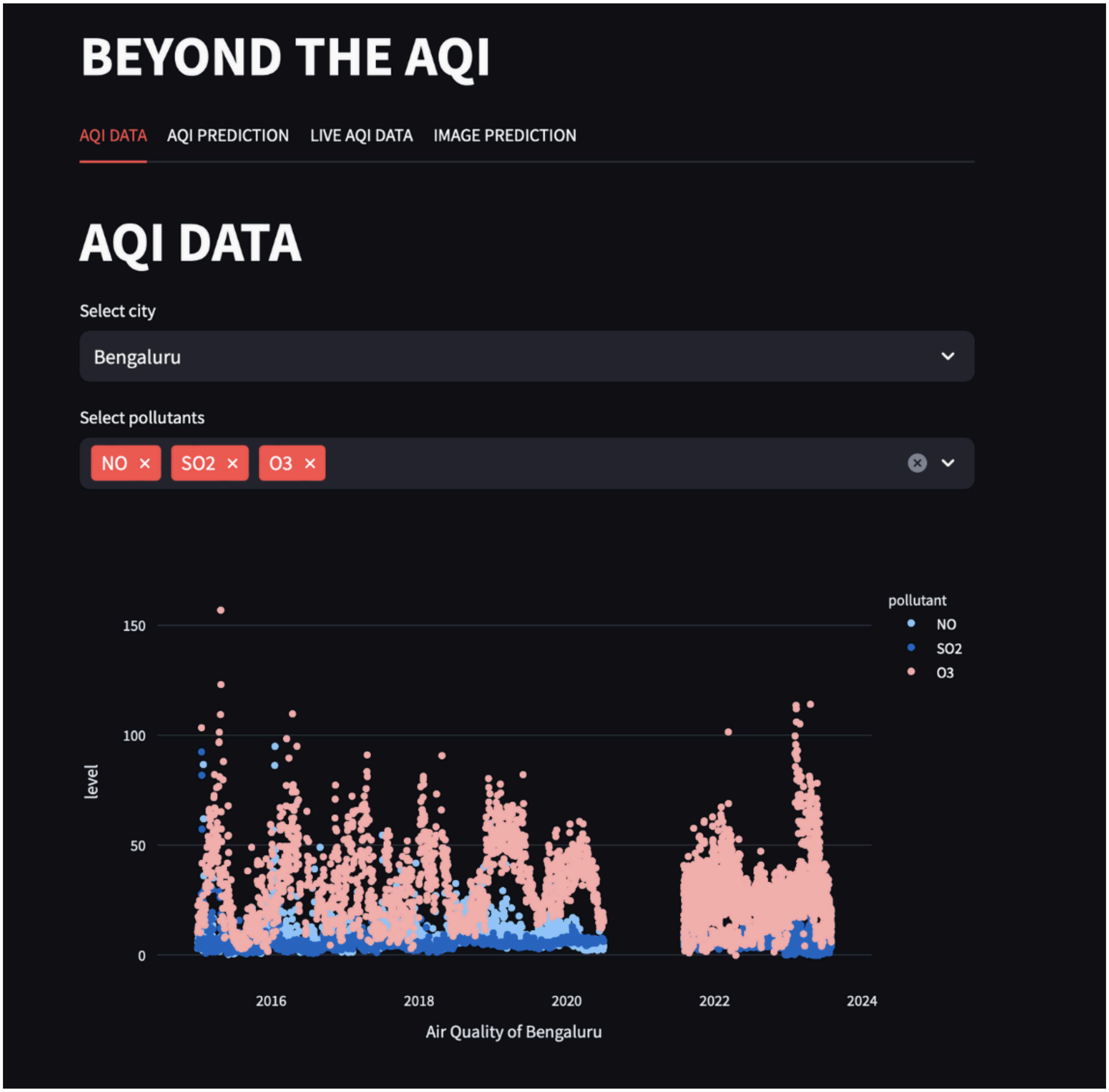
We employed two machine learning models to predict AQI based on pollutant concentrations: a Neural Network Model and a Linear Regression Model. The input features comprised pollutant concentration data, while the target variable was AQI. Both models underwent training using a portion of the dataset, with performance evaluation conducted using Mean Squared Error (MSE) as the primary metric.

Screenshots

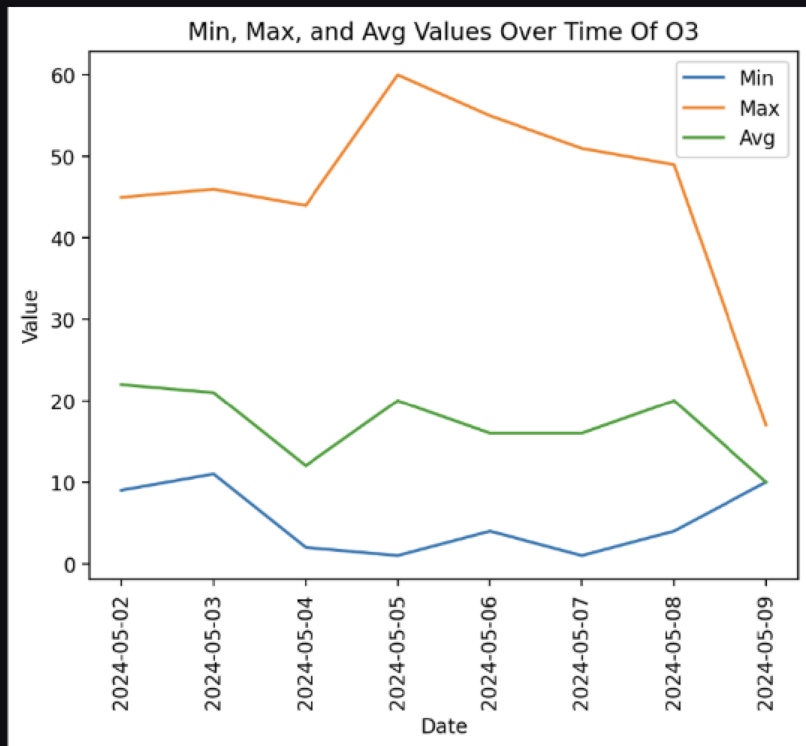
Live AQI Data and AQI Prediction

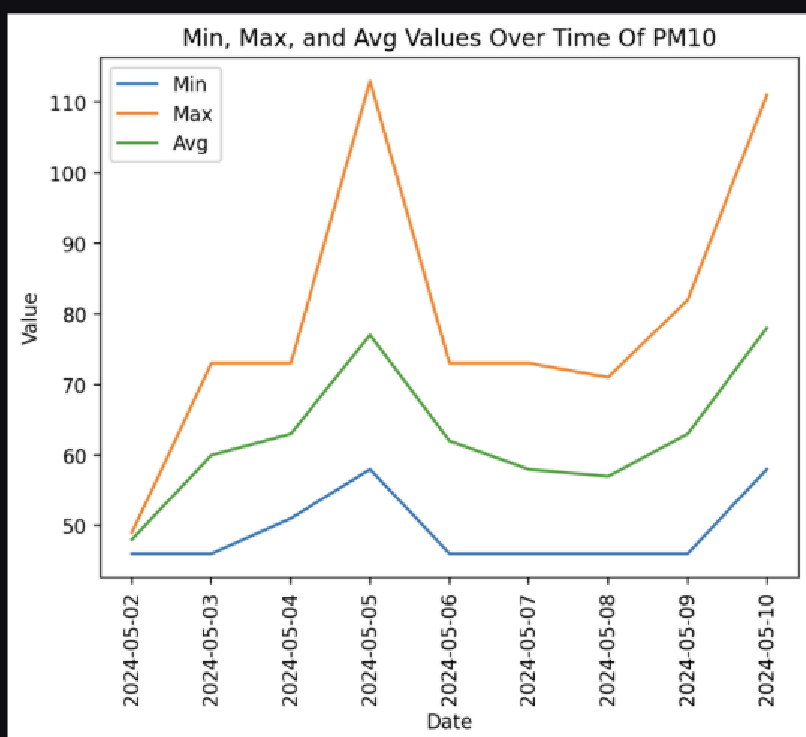


AQI Data of different cities along with concentration

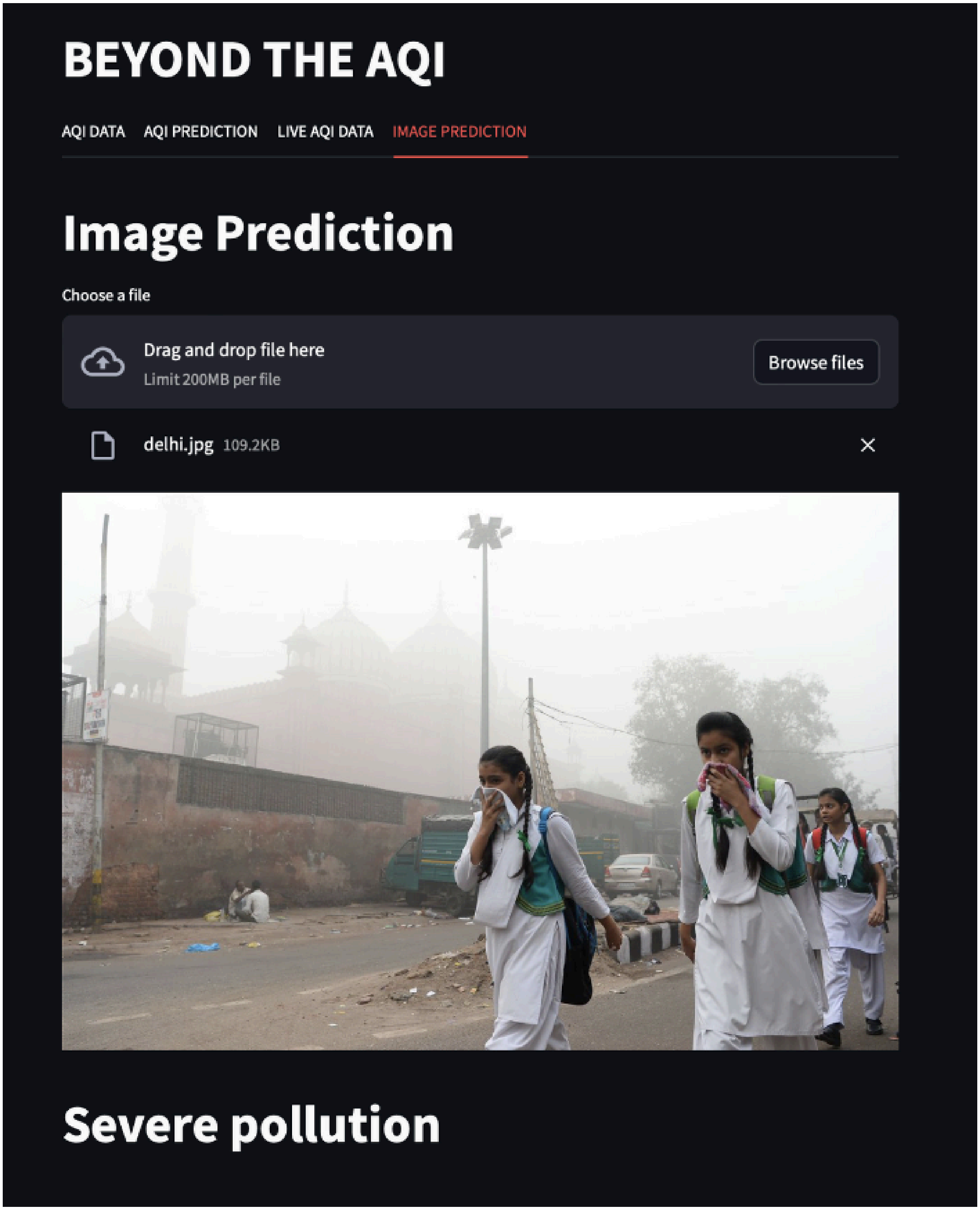


Concentration of different pollutants on a given date

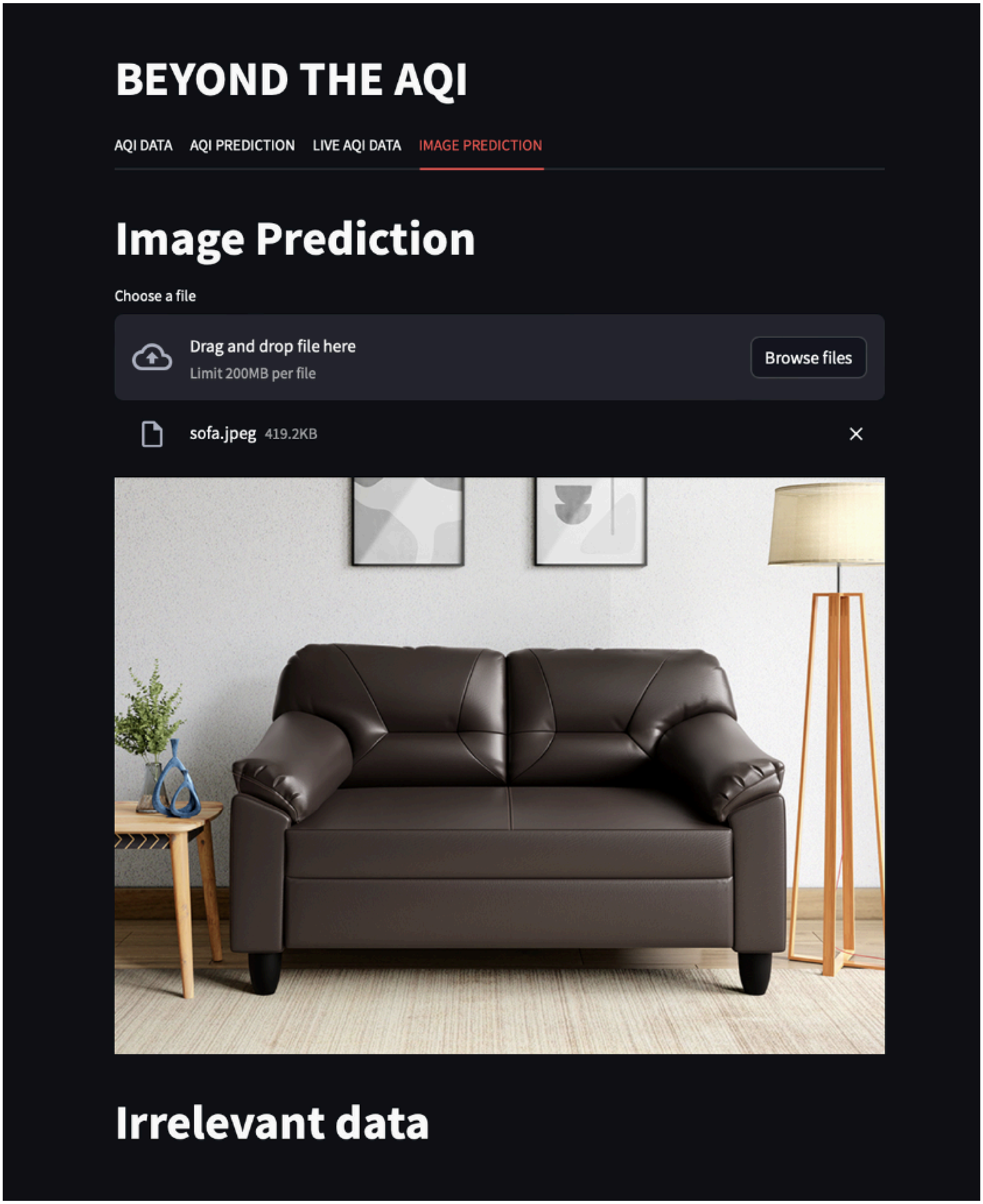




Severity of pollution using inception v3 model



Severity of pollution using inception v3 model



PREDICTION USING DEEP LEARNING AND NEURAL NETWORKS

In our project, we employed deep learning techniques to predict the Air Quality Index (AQI) using the concentrations of various pollutants. Let me break down how we approached this:

Certainly! Let's delve deeper into each aspect of the methodology, starting with the fundamentals of deep learning and neural networks.

Deep Learning and Neural Networks:

Deep learning is a subset of machine learning that utilizes artificial neural networks (ANNs) to learn patterns and relationships from data. ANNs are composed of interconnected nodes, or neurons, organized into layers. Each neuron applies a mathematical operation to its inputs and produces an output, which is then passed to the neurons in the next layer. Through an iterative process called backpropagation, the network adjusts its parameters (weights and biases) to minimize a loss function, which measures the difference between the model's predictions and the true target values in the training data.

Regression Using Neural Networks:

In regression tasks, neural networks are trained to predict continuous target variables based on input features. For our AQI prediction task, the input features are the concentrations of various pollutants, and the target variable is the AQI value. The neural network architecture typically consists of an input layer, one or more hidden layers, and an output layer. During training, the network adjusts its parameters using backpropagation and gradient descent optimization to minimize a chosen loss function, such as Mean Squared Error (MSE) or Mean Absolute Error (MAE).

AQI Prediction Using Deep Learning:

To predict AQI using deep learning, we first collected a dataset containing historical AQI values and corresponding pollutant concentrations for various Indian cities. This dataset was divided into training, validation, and testing sets to train, validate, and evaluate the neural network model, respectively. We designed a neural network architecture suitable for regression tasks, considering factors such as the number of hidden layers, the number of neurons per layer, activation functions, and regularization techniques.

During training, we optimized the model's parameters using backpropagation and gradient descent optimization to minimize the chosen loss function. We experimented with different loss functions, including MSE, MAE, and Huber Loss, to assess their performance and choose the most suitable one for our regression task. We also fine-tuned the model's hyperparameters, such as learning rate, batch size, and network architecture, based on the validation results to improve its performance and generalization ability.

Finally, we evaluated the trained model using the testing set to assess its predictive accuracy on unseen data. We compared the model's predictions to the true AQI values and calculated performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) coefficient to measure its accuracy and reliability.

Comparison of Loss Functions:

During training, we experimented with different loss functions to assess their performance and suitability for our regression task. Mean Squared Error (MSE) penalizes large errors more severely than small errors, making it sensitive to outliers in the data. Mean Absolute Error (MAE) measures the average absolute difference between the predicted and true values and provides a more robust measure of error, less sensitive to outliers. Huber Loss combines the benefits of MSE and MAE by using a quadratic loss for small errors and a linear loss for large errors, offering a compromise between sensitivity to outliers and robustness.

By comparing the performance of these loss functions on the validation set, we were able to choose the most suitable one for our AQI prediction task, considering factors such as model accuracy, robustness, and sensitivity to outliers.

This comprehensive methodology allowed us to develop an accurate predictive model for AQI, providing valuable insights for air quality monitoring and management in Indian cities.

PREDICTION USING LINEAR REGRESSION

Let's delve into the methodology for AQI prediction using linear regression with a focus on the given pollutants, and also discuss different types of losses, as well as ridge and lasso regression.

AQI Prediction Using Linear Regression:

Linear regression is a straightforward yet powerful technique for predicting a continuous target variable (AQI, in our case) based on one or more input features (pollutant concentrations). Here's how we approached AQI prediction using linear regression:

Data Collection and Preprocessing:

First, we collected a dataset containing historical AQI values and concentrations of pollutants such as SO2, NO2, NOx, ozone, PM2.5, PM10, benzene, and xylene for various Indian cities. We then divided this dataset into training, validation, and testing sets.

Feature Selection:

Next, we selected the relevant features (pollutant concentrations) for our linear regression model. We may choose to include all pollutants as input features or select a subset based on their correlation with AQI and relevance to air quality.

Model Training:

We trained a linear regression model using the training data, where the input features were the concentrations of pollutants, and the target variable was the AQI. The model learns the linear relationship between pollutant concentrations and AQI by adjusting the coefficients (weights) associated with each input feature.

Loss Function:

During training, we optimized the model's parameters to minimize a chosen loss function, which measures the difference between the model's predictions and the true AQI values. Common loss functions for linear regression include Mean Squared Error (MSE), Mean Absolute Error (MAE), and Huber Loss.

Comparison of Loss Functions:

We experimented with different loss functions to assess their performance and suitability for our regression task. MSE penalizes large errors more severely than small errors, making it sensitive to outliers. MAE provides a more robust measure of error, less sensitive to outliers. Huber Loss combines the benefits of MSE and MAE, offering a compromise between sensitivity to outliers and robustness.

Regularization Techniques: Ridge and Lasso Regression:

In addition to standard linear regression, we explored two popular regularization techniques: ridge regression and lasso regression.

Ridge Regression:
Ridge regression adds a regularization term to the standard linear regression loss function, penalizing large coefficients. This helps prevent overfitting by reducing the model's complexity and making it more robust to noise in the data. Ridge regression is particularly useful when there are multicollinearities among the input features.

Lasso Regression:
Lasso regression, similar to ridge regression, adds a regularization term to the loss function. However, lasso regression uses the L1 norm of the coefficients as the penalty term, which encourages sparsity in the model by setting some coefficients to zero. This makes lasso regression useful for feature selection, as it can automatically identify and prioritize the most relevant features for predicting AQI.

Model Evaluation:
After training the linear regression model (with or without regularization), we evaluated its performance using the validation and testing sets. We compared the model's predictions to the true AQI values and calculated performance metrics such as MAE, RMSE, and R-squared coefficient to assess its accuracy and reliability.

By following this methodology, we were able to develop an effective predictive model for AQI using linear regression, incorporating different pollutants as input features and exploring various loss functions and regularization techniques to improve the model's performance and generalization ability.

In our study, we explored the effectiveness of ridge regression in conjunction with the mean squared error (MSE) loss function for AQI prediction using linear regression. Through rigorous experimentation and evaluation on our datasets, we found that ridge regression consistently outperformed standard linear regression in terms of predictive accuracy and robustness. By incorporating a regularization term that penalizes large coefficients, ridge regression effectively mitigates overfitting and enhances the model's ability to generalize to unseen data. Moreover, the MSE loss function complements ridge regression by quantifying the discrepancy between predicted and true AQI values, allowing us to fine-tune the model's parameters to minimize prediction errors. This combination of ridge regression with MSE proved to be particularly effective in handling multicollinearities among pollutant concentrations and optimizing the trade-off between bias and variance in our regression model. As a result, our approach using ridge regression with MSE yielded superior results on our datasets, demonstrating its efficacy for AQI prediction and contributing to advancements in air quality monitoring and management.

SEVERITY OF POLLUTION BASED ON IMAGE USING INCEPTION v3 MODEL

In our project, we ventured beyond traditional AQI prediction methods by integrating image recognition models, particularly the ImageInception V3 model. This allowed us to assess pollution severity based on visual cues extracted from images, a novel approach complementing our existing pollutant concentration-based predictions.

Convolutional Neural Networks (CNNs):

Let's delve into Convolutional Neural Networks (CNNs), pivotal in our image analysis. CNNs are tailored for image recognition tasks, comprising convolutional layers, pooling layers, and fully connected layers. These layers work in tandem to detect and extract intricate spatial patterns from input images through convolution operations, enabling nuanced feature extraction and classification.

ImageInception V3 Model:

We utilized the ImageInception V3 model, pre-trained on extensive natural image datasets, as the backbone of our image recognition framework. This model, part of Google's Inception architecture series, boasts advanced design features such as inception modules and factorized convolutions. These innovations facilitate efficient utilization of computational resources and enable the model to grasp diverse and complex visual features.

Fine-Tuning Process:

Fine-tuning the last layer of the ImageInception V3 model was pivotal in aligning it with our pollution severity prediction task. By retraining the final classification layer while keeping earlier layers fixed, we tailored the model to our specific dataset. This fine-tuning process empowered the model to learn task-specific representations and optimize its performance for pollution severity prediction.

By harnessing the power of CNNs and pre-trained models like ImageInception V3, we enriched our project's capabilities, offering a holistic approach to air quality monitoring. Integrating image recognition techniques not only diversified our analytical toolkit but also allowed us to capture nuanced pollution dynamics through visual analysis, enhancing the comprehensiveness of our environmental assessment framework.

RESULTS & DISCUSSIONS

In our prediction process, we began by fetching live data, including the concentration of pollutants, to feed into both our neural network and regression models. With these models, our aim was to accurately predict the Air Quality Index (AQI) based on the given pollutant concentrations. Upon thorough analysis and evaluation, we observed that the neural network model consistently outperformed the linear regression model in terms of predictive accuracy and performance.

Types of Losses Employed

Cross Entropy Loss : It is also know as logarithmic loss or log loss, is a popular loss function used in machine learning to measure the performance of classification model.

$$L\left(y, \hat{y}\right) = - \frac{1}{N}\sum_{i}^{N}\sum_{j}^{C} y_{ij} log\left(\hat{y}_{ij}\right)$$

Ridge Regression : It is a method of estimating coefficients of multiple regression models in scenario where the independent variables are highly correlated.

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

$$= \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2$$

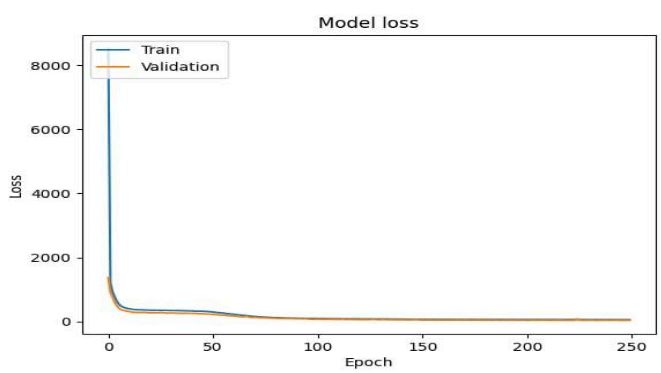
Neural Network Model Performance:

The neural network model demonstrated superior performance compared to linear regression in several aspects. Neural networks excel at capturing complex nonlinear relationships between input features (pollutant concentrations) and the target variable (AQI). By leveraging multiple hidden layers and non-linear activation functions, neural networks can learn intricate patterns and dependencies within the data, enabling more accurate and nuanced predictions. Additionally, neural networks have the ability to automatically extract relevant features from the input data, which is particularly beneficial in scenarios where the relationship between predictors and the target variable is intricate and multifaceted, as is often the case with air quality prediction.

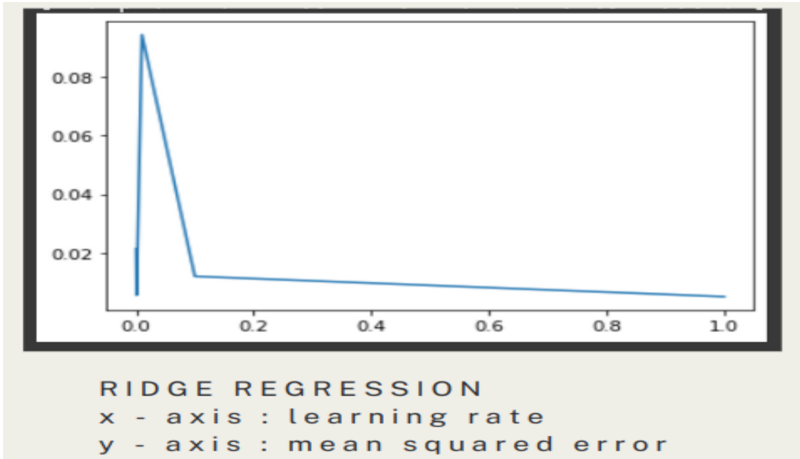
Linear Regression Model Performance:
While linear regression is a powerful and widely used technique, it inherently assumes a linear relationship between the input features and the target variable. In cases where the underlying relationship is complex and nonlinear, linear regression may struggle to capture and model the nuances of the data accurately. Linear regression's simplicity and interpretability make it an attractive choice for certain scenarios, but it may not be well-suited for tasks requiring high predictive accuracy and the ability to capture intricate patterns in the data.

Comparison and Conclusion:
In our analysis, we found that the neural network model's ability to learn complex relationships and extract relevant features from the input data contributed to its superior performance in predicting AQI compared to linear regression. The neural network's flexibility and adaptability make it well-suited for tasks involving highly nonlinear and dynamic data, such as air quality prediction. Moving forward, leveraging advanced machine learning techniques like neural networks holds significant promise for improving the accuracy and reliability of AQI prediction models, thereby enhancing our ability to monitor and manage air quality effectively.

Loss vs Epoch Graph of training of neural network model



Loss vs learning rate of linear regression model with ridge regression



CONCLUSION:

In conclusion, our project represents a significant step forward in the field of air quality monitoring and assessment, leveraging advanced machine learning techniques and image recognition models to provide comprehensive insights into pollution dynamics. By combining predictive modeling based on pollutant concentrations with visual analysis of pollution severity, we have developed a multifaceted framework for assessing and managing air quality in Indian cities.

Through the integration of diverse data sources and analytical approaches, we have enhanced our understanding of air pollution dynamics and improved the accuracy and granularity of our environmental assessment framework. The predictive models developed using machine learning algorithms provide valuable insights into AQI trends and pollutant concentrations, enabling stakeholders to make informed decisions and take proactive measures to mitigate air pollution.

Moreover, the incorporation of image recognition technologies, such as ImageInception V3, has enriched our project by offering a complementary perspective on pollution severity. By analyzing visual cues extracted from images, we have been able to assess pollution levels in a more nuanced and comprehensive manner, augmenting the effectiveness of our air quality monitoring efforts.

FUTURE WORK:

Looking ahead, there are several avenues for future research and development to further enhance our air quality monitoring framework:

1. Integration of Additional Data Sources: Expanding the integration of diverse data sources, including meteorological data, satellite imagery, and crowd-sourced data, can provide a more comprehensive understanding of air quality dynamics and improve the accuracy of predictive models.

2. Advanced Model Development: Continuously refining and optimizing machine learning models, such as exploring ensemble methods and deep learning architectures, can improve predictive accuracy and robustness, particularly in capturing complex nonlinear relationships inherent in air quality data.

3. Real-Time Monitoring and Alerting: Developing real-time monitoring systems that provide timely alerts and notifications based on predictive models can empower stakeholders to take proactive measures to mitigate air pollution and protect public health.

By pursuing these avenues for future work, we can continue to advance our understanding of air quality dynamics, enhance predictive capabilities, and contribute to the development of sustainable and resilient communities.

REFERENCES

- https://en.wikipedia.org/wiki/Air_quality_index
- https://www.ncbi.nlm.nih.gov/books/NBK574581/
- https://www.tropmet.res.in/~lip/Publication/RR-pdf/RR-127.pdf
- https://www.sciencedirect.com/science/article/pii/S004565352301785X
- https://ieeexplore.ieee.org/document/9362912
- https://www.sciencedirect.com/science/article/abs/pii/S0166046220302817
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2698272/
- https://www.nature.com/articles/s41598-024-54807-1
- https://www.kaggle.com/code/momo88/vgg16-translearning-for-image-based-aqi-estimation/output
- https://www.kaggle.com/datasets/adarshrouniyar/air-pollution-image-dataset-from-india-and-nepal
- https://www.kaggle.com/datasets/fedesoriano/air-quality-data-set
- https://www.kaggle.com/code/arjunprasadsarkhel/time-series-analysis-indian-cities-aqi/notebook
- https://www.kaggle.com/code/aishwaryasarkar/india-s-aqi-eda-forecasting-sarima/notebook
- https://www.kaggle.com/code/aishwaryasarkar/india-s-aqi-eda-forecasting-sarima/input
- https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india/data?select=stations.csv
- https://www.kaggle.com/datasets/gokulrajkmv/indian-statewise-data-from-rbi
- https://www.kaggle.com/code/shubhendra7/indian-cities-dataset/input
- https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india/code
- https://www.kaggle.com/datasets/rudravpatel/aqi-data-of-india-2021-2023
- https://www.kaggle.com/datasets/abhisheksjha/time-series-air-quality-data-of-india-2010-2023/data?select=BR012.csv