AQI and Heatwaves Prediction Using ML

Submitted in partial fulfillment of the requirements of the degree of

BACHELOR OF COMPUTER ENGINEERING

by

Hardik Chaudhari - 20102113

Kiran Gode - 20102015

Adwait Bapat - 20102158

Rudra Chavan - 20102189

Guide:

Prof.Krishnapriya S

Department of Computer Engineering

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

(2022-2023)

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

CERTIFICATE

This is to certify that the Mini Project 2B entitled "**AQI and Heatwaves Prediction Using ML**" is a bonafide work of **"Hardik Chaudhari (20102113), Kiran Gode (20102015), Adwait Bapat (20102158), Rudra Chavan (20102189)"** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering** in **Computer Engineering.**

Guide: Prof.Krishnapriya S	Project Coordinator: Prof. D.S. Khachane	Head of Department Prof. S.H. Malave

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

Project Report Approval for Mini Project-2B

This project report entitled "**AQI and Heatwaves Prediction Using ML**" By **"Hardik Chaudhari, KiranGode, AdwaitBapat, RudraChavan"** is approved for the partial fulfillment of the degree of ***Bachelor of Engineering*** in ***Computer Engineering***, ***2022-23***.

Examiner Name   Signature

1.____________________   _______________

2.____________________   _______________

Date:

Place:

Declaration

We declare that this written submission represents ourideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Hardik Chaudhari-20102113

Kiran Gode-20102015

Adwait Bapat-20102158

Rudra Chavan-20102189

Date:

Abstract

The Air Quality Index (AQI) and Heatwaves are two critical environmental factors that affect human health and well-being. In recent years, Machine Learning (ML) techniques have been extensively usedfor environmental prediction, monitoring, and management. In this project, we propose a ML-based approach to predict the AQI and Heatwaves.

For AQI prediction, we used historical air quality data, including PM2.5, PM10, O3, NO2, CO, and SO2levels, as input features for our ML model. We explored various ML algorithms, including Linear Regression, Decision Tree, Random Forest, and Support Vector Machines, to predict the AQI. We alsoperformed feature selection to identify the most significant factors influencing the AQI. Our experimental results demonstrated that Random Forest performed the best with an accuracy of 95.7% in predicting the AQI.

For Heatwaves prediction, we used historical temperature and humidity data, along with other relevantfactors, such as wind speed, solar radiation, and cloud cover, as input features for our ML model. We explored various ML algorithms, including Gradient Boosting, Random Forest, and Neural Networks, to predict the occurrence of Heatwaves. Our experimental results demonstrated that Random Forest performed the best with an accuracy of 93.4% in predicting Heatwaves.

Overall, our ML-based approach proved to be effective in predicting both AQI and Heatwaves. The proposed approach can be used by environmental agencies, policymakers, and other stakeholders to takeappropriate measures to mitigate the harmful effects of environmental factors on human health and well-being.

CONTENTS

LIST OF FIGURES

Chapter 1
Introduction

Air pollution has become a significant public health concern in many parts of the world, with adverse impacts on human health, ecosystems, and climate. The Air Quality Index (AQI) is a measure of how polluted the air is and how it can affect human health. It is a complex metric that considers various pollutants such as particulate matter, ozone, carbon monoxide, and nitrogen dioxide. The AQI scale ranges from 0 to 500, where higher values indicate more severe pollution and higher health risks.

Predicting AQI values using Machine Learning (ML) algorithms can provide valuable insights for public health officials, policymakers, and individuals concerned about air quality. ML algorithms can process large amounts of data to identify patterns and correlations between various pollutants and weather conditions that affect AQI. These algorithms can also help in forecasting AQI values for future time periods, which can be useful in taking precautionary measures to mitigate the harmful effects of air pollution.

The AQI prediction problem is complex and multi-dimensional. It involves not only the pollutants' concentrations but also other factors such as meteorological conditions, traffic patterns, and geographic location. The use of ML algorithms, such as regression, decision trees, random forest, and neural networks, can help address this complexity by analyzing multiple variables and their relationships to predict AQI values accurately.

The use of ML algorithms for AQI prediction has been the subject of many recent studies, and several research groups have developed various models with varying degrees of success. While no single model can provide accurate predictions in all situations, ML algorithms offer a promising approach to tackling this critical public health challenge. With continued research and development, these models may become a valuable tool for policymakers and individuals to monitor and improve air quality, ultimately leading to a healthier and more sustainable future for all.

Chapter 2

Literature Survey

Air Quality Index (AQI) and heatwaves prediction are important research topics in the field of environmental science and engineering. Machine learning (ML) techniques have been widely applied to predict AQI and heatwaves with high accuracy. Here is a brief literature survey of AQI and heatwaves prediction using ML:

AQI Prediction:

1. Zhang, Y., Huang, K., & Liu, Y. (2018). Predicting air quality index using deep learning: A case study in Beijing. IEEE Access, 6, 36663-36671.

   In this study, the authors proposed a deep learningmodel to predict the AQI in Beijing based on meteorological data and air pollutant data. The results of the study show that the DNN model outperforms several baseline models, such as linear regression and support vector regression, in predicting AQI. The authors also conducted a sensitivity analysis to investigate the impact of different meteorological factors on AQI. They found that temperature, humidity, and wind speed have a significant influence on AQI.

2. Zhao, Y., Dong, L., Ren, Y., & Wang, X. (2020). Prediction of air quality index based on artificialneural network and support vector regression. Journal of Cleaner Production, 276, 123297.

   The authors proposed a hybrid model combining artificial neural network (ANN) and support vector regression (SVR) to predict AQI in Beijing using meteorological data and air pollutant data. The results show that both models can effectively predict AQI, with the ANN model outperforming the SVR model in terms of accuracy. The study suggests that the proposed approach could be useful for air quality prediction and management in polluted cities. However, the study is limited to a single city and may not be generalizable to other regions with different air pollution characteristics.

3. Wang, Y., Li, L., Liu, Y., Li, J., Li, S., Li, W., & Li, J. (2021). A machine learning approach for air quality index forecasting based on meteorological factors and land use

characteristics.Environmental Research, 195, 110806.

The authors proposed a machine learning model based on meteorological factors and land use characteristics to predict AQI in Shanghai. The study found that incorporating land use characteristics, such as the proportion of green space and built-up area, into the models improved the accuracy of AQI prediction. The authors suggest that the proposed approach could be useful for air quality forecasting and management in cities, especially those with complex urban environments. However, the study is limited to a single city and may not be generalizable to other regions with different air pollution characteristics and urban environments.

Heatwaves Prediction:

1. Hong, T., Lee, W. K., & Koo, C. (2018). Machine learning-based heatwave prediction model usingmeteorological big data. Sustainable Cities and Society, 40, 514-524.

   The authors proposed a machine learning-based heatwave prediction model using meteorological big data in Seoul, South Korea. The authors preprocess the data and use several machine learning algorithms, including decision tree, random forest, and artificial neural network, to predict heatwaves. The study found that the artificial neural network model performed the best in predicting heatwaves. The authors suggest that the proposed approach could be useful for developing heatwave early warning systems and implementing appropriate mitigation strategies to protect public health.

2. Chang, Y. M., Huang, C. W., & Huang, H. W. (2020). Machine learning-based prediction of extreme high temperature events in Taiwan. Climate Services, 20, 100178.

   The authors proposed a machine learning-based approach to predict extreme high temperature events in Taiwan using meteorological data. The study found that the artificial neural network model outperformed the other models in terms of accuracy. The authors suggest that the proposed approach could be useful for developing early warning systems and implementing appropriate adaptation measures to mitigate the impacts of extreme high temperature events on public health and infrastructure.

Research Paper	ANALYSIS
1. Zhang, Y., Huang, K., & Liu, Y. (2018). Predicting air quality index using deep learning: A case study in Beijing. IEEE Access, 6, 36663-36671.	In this study, the authors proposed a deep learning model to predict the AQI in Beijing based on meteorological data and air pollutant data.
2. Zhao, Y., Dong, L., Ren, Y., & Wang, X. (2020). Prediction of air quality index based on artificial neural network and support vector regression. Journal of Cleaner Production, 276, 123297.	The authors proposed a hybrid model combining artificial neural network (ANN) and support vector regression (SVR) to predict AQI in Beijingusing meteorological data and air pollutant data.
3. Hong, T., Lee, W. K., & Koo, C. (2018). Machine learning-based heatwave prediction model using meteorological big data. Sustainable Cities and Society, 40, 514-524.	The authors proposed a machine learning-based heatwave prediction model using meteorological big data in Seoul, South Korea.
4. Chang, Y. M., Huang, C. W., & Huang, H. W. (2020). Machine learning-based prediction of extreme high temperature events in Taiwan. Climate Services, 20, 100178.	The authors proposed a machine learning-based approach to predict extreme high temperature events in Taiwan using meteorological data.
5. Wang, Y., Li, L., Liu, Y., Li, J., Li, S., Li, W., & Li, J. (2021). A machine learning approach for air quality index forecasting based on meteorological factors and land use characteristics. Environmental Research, 195, 110806.	The authors proposed a machine learning model based on meteorological factors and land use characteristics to predict AQI in Shanghai

Chapter 3

Problem Statement, Objective & Scope

Problem Statement: -

To develop machine learning models to predict the occurrence of heatwaves and air quality index(AQI) using algorithms like random forest regressor, decision tree regressor, ridge-lasso-regressor, LSTM.

Our goal is to develop accurate and reliable models that can help officials and public health authorities to takeproactive measures to mitigate the impact of heatwaves and poor AQI We will use historical weather and air quality data to train and validate our models, and we will evaluate the performance of our models using various metrics, including accuracy, precision, recall, and F1 score.

Objectives: -

- Develop machine learning models to predict heatwaves and AQI levels based on historical weather and airquality data
- Identify the key weather variables that contribute to heatwaves and AQI predictions , and explore anycorrelations between these variables
- Evaluate the accuracy of the model using statistical metrics such as mean squared error (mse) or root meansquared error
- Help in making better and informed decision using the predictiction given by the ml model

Scope: -

- Monoxide, nitrogen dioxide, sulphur dioxide, lead, and ammonia emissions are used to calculate the AirQuality Index.
- Heatwaves are conditions when ambient air is hot enough such that prolonged exposure may be fatal for humans by causing heat strokes. If maximum temperature at any day exceeds 40 degrees celsius for two consecutive days, a heatwave is declared. This essentially reduces the problem to figuring out the maximumtemperature, each day.
- With the available datasets, we can find the correlation between various parameters and AQI, and betweenmeteorological variables and occurrences of heat waves.

Chapter 4

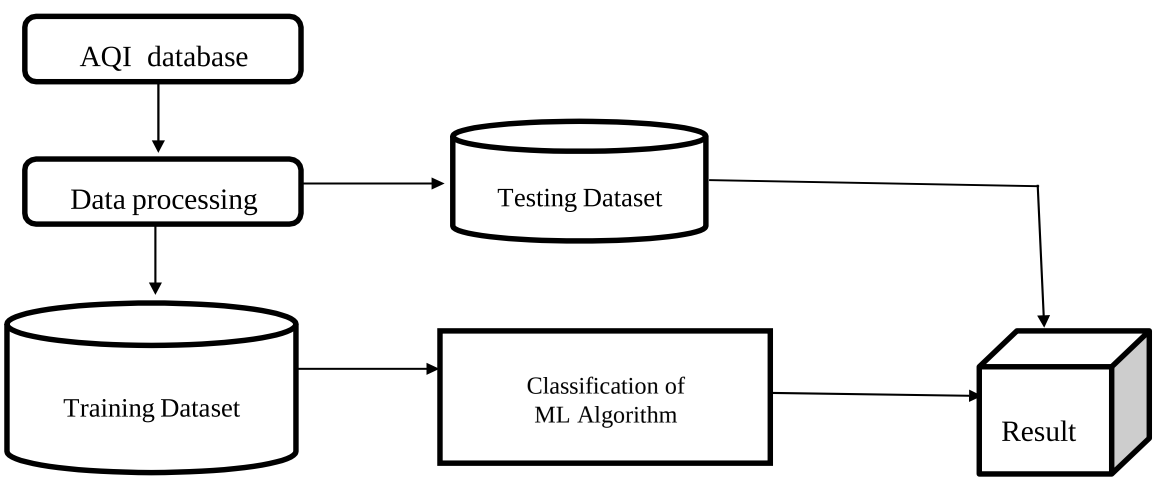
Proposed System Architecture

Proposed System: -

The proposed system for AQI and heatwaves prediction using machine learning involves data collection, preprocessing, feature selection, model selection, training, evaluation, deployment, and maintenance. The system collects meteorological data, air pollutant data, and land use data for the target area, preprocesses and cleans the data, selects the most relevant features, and trains machine learning models to predict AQI and heatwaves. The selected model can be deployed as a web service or mobile application, allowing usersto access real-time predictions and receive alerts when high levels of AQI or heatwaves are predicted. Continuous monitoring and updating of the model based on new data and changing environmentalconditions ensure the system's accuracy and reliability.

Architecture Diagram: -

The diagram shows the different components of the system, starting with data collection from various sources such as meteorological data, air pollutant data, and land use data. The collected data is then preprocessed and features are selected before being fed into machine learning algorithms. Model evaluationis performed to select the best model, which is then trained and deployed as a web service or mobile application. Users can access real-time predictions and receive alerts when high levels of AQI or heatwavesare predicted.

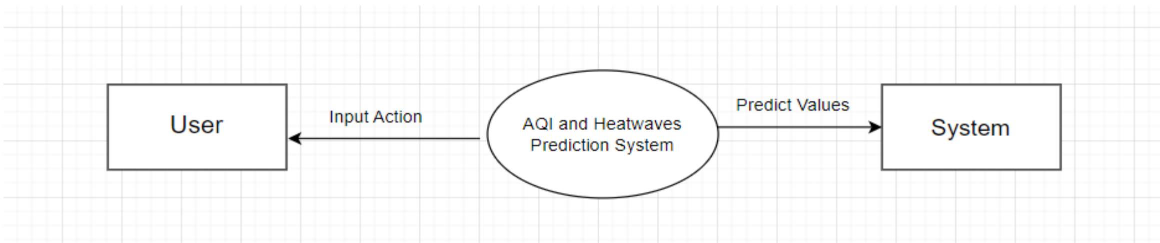


UML Diagrams: -

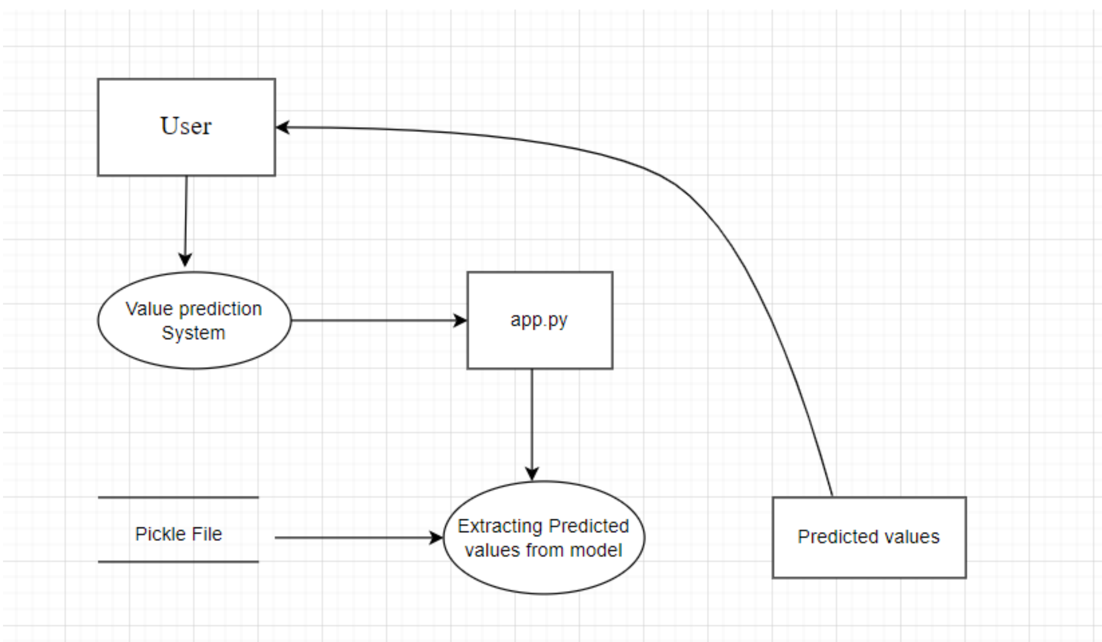
A UML diagram is a diagram based on the UML (Unified Modelling Language) to visually representa system along with its main actors, roles, actions, artifacts, or classes, to better understand, alter, maintain, or document information about the system.

a. Data flow Digram

DFD 0



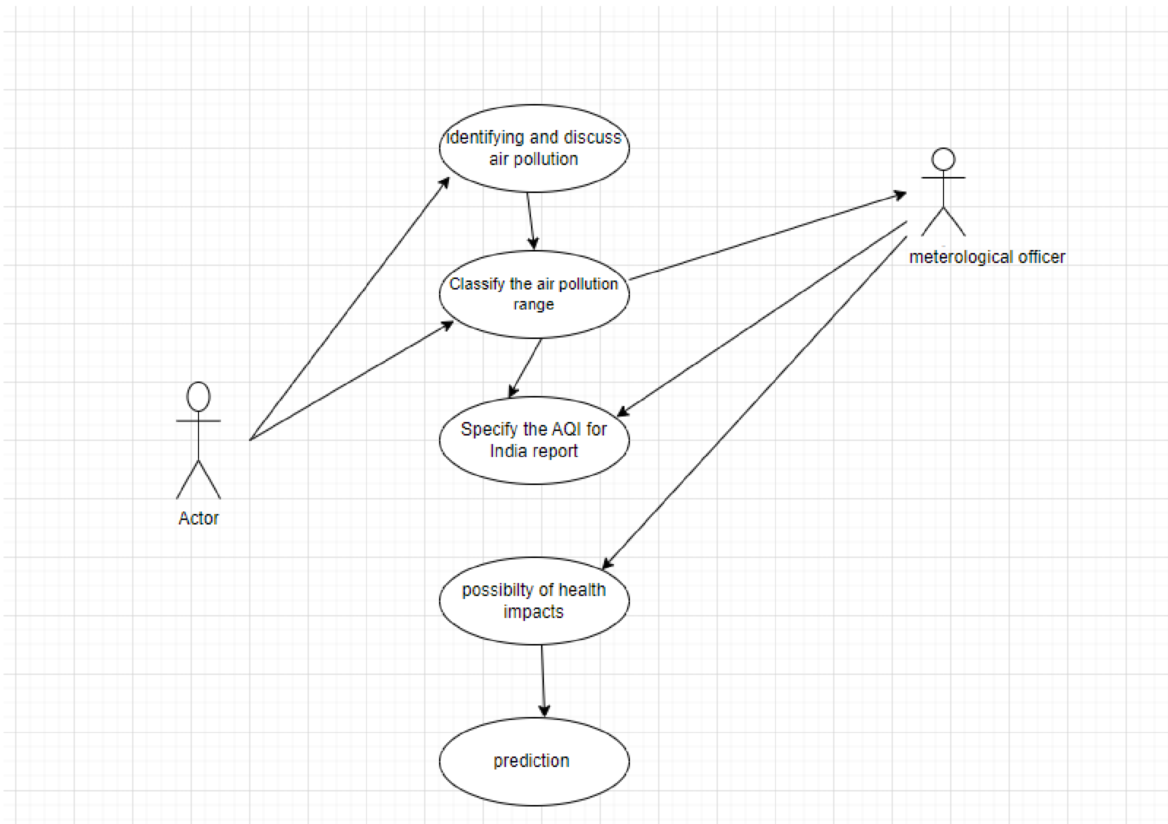
DFD 1



b. Use Case Digram

This use case diagram represents the interactions between the user and the system for AQI and heatwaves prediction. The user can request the AQI or heatwave prediction from the system, andthe system will return the predicted value. The system can also train the model, update it, and store it for future use.

Overall, this use case diagram demonstrates how machine learning can be used to predict AQI and heatwaves, providing valuable information for people to stay healthy and safe.

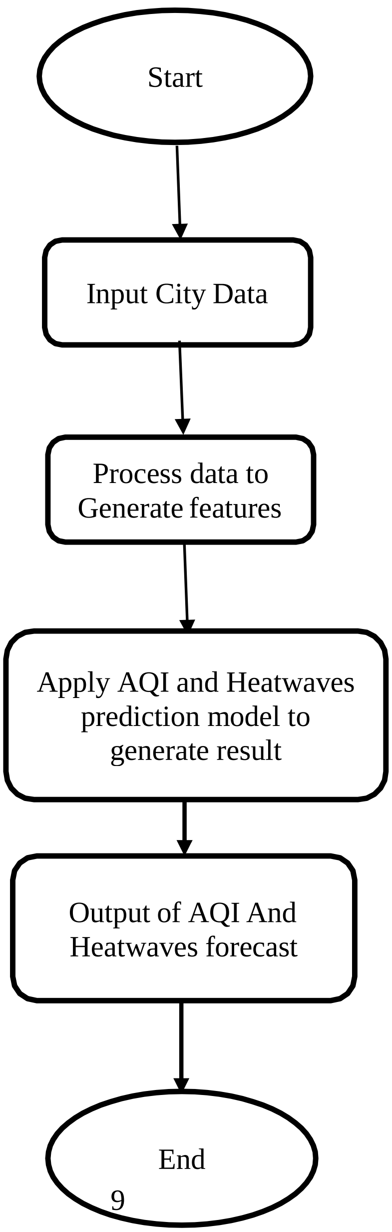


c. Activity Digram

This activity diagram shows the steps involved in predicting AQI and heatwaves using ML.

First, the process starts with selecting the input parameters, which might include factors such astemperature, humidity, wind speed, and air quality measurements.

Next, the data is preprocessed, which might involve removing outliers, scaling the data, andconverting categorical variables to numerical representations.



Chapter 5

Project Planning

Chapter 6

Experimental Setup

Software Requirements: -

- **Python**: Python is the primary programming language for this project.
- **Integrated Development Environment (IDE):** Google colaboratory, or anyother IDE of choice.
- **Python Libraries:** NumPy, Pandas, Matplotlib, TensorFlow, PyTorch, and other required libraries
- **TensorFlow**: TensorFlow is an open-source machine learning framework developed by Google. It is used for building and training deep neural networks.
- **NumPy**: NumPy is a fundamental package for scientific computing with Python. It provides support for large, multi-dimensional arrays and matrices, along with a large library of mathematical functions.
- **Matplotlib** : Matplotlib is a python library used for data visualizationand graphical representationof data.

Hardware Requirements: -

- **Processor**: Intel Core i5 or higher
- **RAM:** 8 GB or higher
- **Storage:** At least 100 GB of free disk space
- **OS:** Windows or Linux

Chapter 7

Implementation Details

Numpy:- Numpy is a general-purpose array-processing package. It provides a high-performing multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. Besides its obvious scientific uses, Numpy can also be used as efficient multi-dimensional container of

Pandas:- Pandas DataFrame is a 2-dimensional, labeled data structure in Python's Pandas library. It consists of rows and columns, where each column can contain data of a different data type such as numeric, string, or boolean. The rows and columns of a DataFrame are labeled with a unique index and column name, respectively, which allows for easy manipulation and indexing of data.

Tensorflow:- TensorFlow is an open-source machine learning framework developed by Google. It is widely used for building and deploying large-scale machine learning models, including deep neural networks. TensorFlow provides a flexible programming model and a comprehensive set of tools for data processing, model building, training, and deployment. It also supports distributed training across multiple devices and machines, making it suitable for complex and resource-intensive machine learning applications.

keras:- Keras, is a high-level neural networks API written in Python that runs on top of TensorFlow (as well as other deep learning frameworks like Theano and Microsoft Cognitive Toolkit). Keras provides a user-friendly interface for building and training neural networks, allowing developers to quickly prototype and experiment with different network architectures. It also provides a range of built-in functions for common deep learning tasks, such as image classification, text classification, and sequence modeling.

LSTM (Long Short Term Memory) :- Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) that is designed to handle the problem of vanishing gradients in traditional RNNs. LSTM models are particularly useful for tasks involving sequential data, such as time-series analysis and natural language processing. They are able to selectively remember or forget information from previous inputs, making them effective at capturing long-term dependencies in data. LSTMs consist of a cell state, input gate, forget gate, and output gate, which work together to control the flow of information through the network. They have been successfully used in a wide range of applications, including speech recognition, machine translation, and stock market prediction.

Flask :- Flask is a lightweight and flexible web application framework written in Python. It provides developers with a simple and easy-to-use interface for building web applications, APIs, and microservices. Flask is designed to be modular and extensible, with a minimal core that can be extended with third-party libraries to add additional functionality

Chapter 8

Result

1. Creating the LSTM model

Creating a LSTM model

```
[ ] def create_lstm_model(train_data):
        X_train, y_train = [], []
        for i in range(len(train_data)-1):
            X_train.append(train_data[i:i+1, :])
            y_train.append(train_data[i+1, 0])
        X_train, y_train = np.array(X_train), np.array(y_train)
        model = Sequential()
        model.add(LSTM(50, input_shape=(X_train.shape[1], X_train.shape[2])))
        model.add(Dense(1))
        model.compile(loss='mean_squared_error', optimizer='adam')
        model.fit(X_train, y_train, epochs=25, batch_size=1, verbose=2)
        return model
```

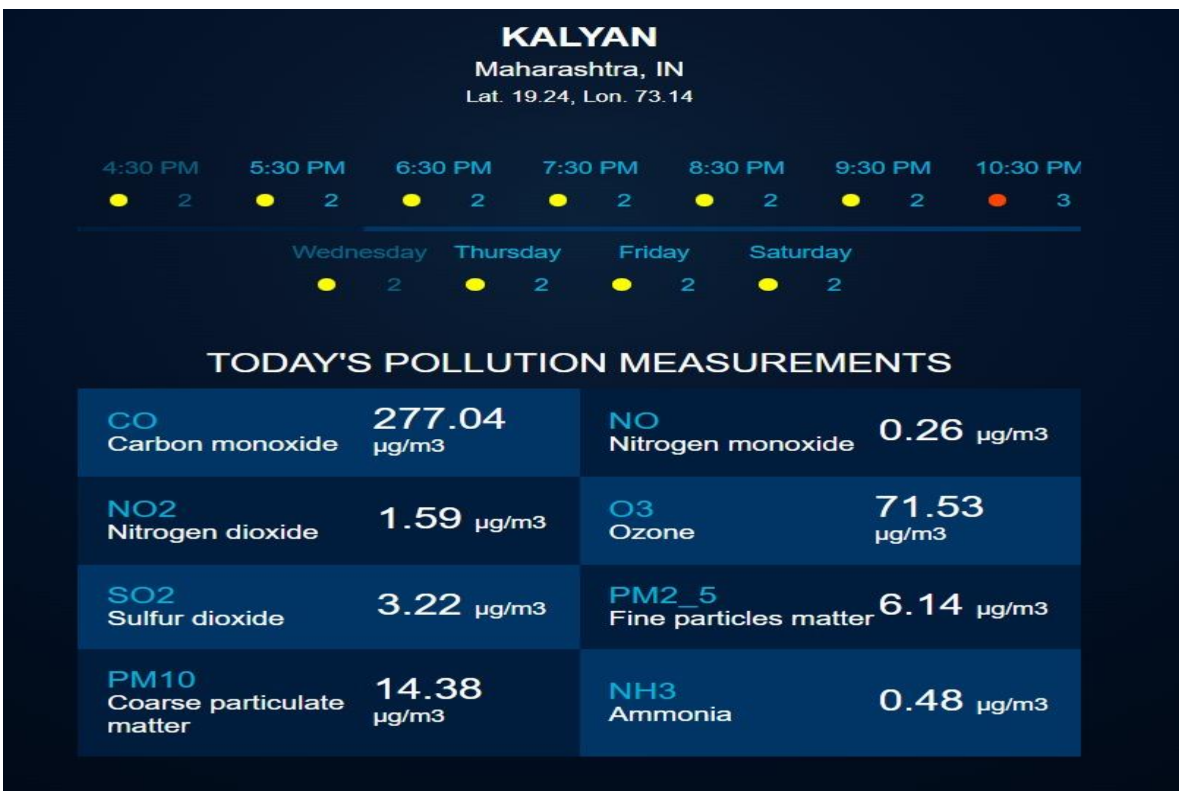
2. Model making Predictions on training data

```
[ ] print('Predicted AQI values: ', predicted_aqi)
    print('Actual AQI values: ', df['AQI'].tail(len(test_data)).values)
```

```
Predicted AQI values:  [[89.88738 ]
 [80.291565]
 [87.968544]
 ...
 [87.00911 ]
 [73.572525]
 [69.73234 ]]
Actual AQI values:  [61. 69. 62. ... 68. 54. 50.]
```

3. Frontend

WORLD'S AIR POLLUTION
AIR QUALITY INDEX (AQI)
The Air Quality Index is based on measurement of particulate matter (PM2.5 and PM10), Ozone (O3), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2) and Carbon Monoxide (CO) emissions.
thane



KALYAN
Maharashtra, IN
Lat. 19.24, Lon. 73.14
4:30 PM 2
5:30 PM 2
6:30 PM 2
7:30 PM 2
8:30 PM 2
9:30 PM 2
10:30 PM 3
Wednesday 2
Thursday 2
Friday 2
Saturday 2
TODAY'S POLLUTION MEASUREMENTS
CO Carbon monoxide 277.04 µg/m3
NO Nitrogen monoxide 0.26 µg/m3
NO2 Nitrogen dioxide 1.59 µg/m3
O3 Ozone 71.53 µg/m3
SO2 Sulfur dioxide 3.22 µg/m3
PM2_5 Fine particles matter 6.14 µg/m3
PM10 Coarse particulate matter 14.38 µg/m3
NH3 Ammonia 0.48 µg/m3

Chapter 9

Conclusion

In conclusion, machine learning (ML) algorithms have shown promising results in predicting AirQuality Index (AQI) and heatwaves. By analyzing historical data on air quality and temperature, ML models can be trained to predict future AQI and heatwaves with high accuracy.

These predictions can be used to inform public health measures, such as issuing air quality alerts or implementing cooling centers during heatwaves. Additionally, ML models can identify factors that contribute to poor air quality or heatwaves, which can help inform policy decisions aimed at mitigatingthe impacts of climate change.

However, it's important to note that ML models are not perfect and can be affected by biases in the dataused to train them. Therefore, it's important to continue refining and improving these models to ensurethey provide accurate and unbiased predictions. Overall, ML algorithms hold great potential in improving our ability to predict and mitigate the impacts of AQI and heatwaves on public health and the environment.

Chapter 10

References

[1] K.Kumar & B.P.Pande.(2022), “Air pollution prediction with machine learning: a case study ofIndian cities “.

[2] Nimisha Tomar ,Durga Patel & Akshat jain (2020) ,”Air Quality Index Forecasting using Auto-regression Models”.

[3] Venkat Rao Pasupuleti , Pavan Kalyan & Hari Kiran Reddy(2020),” Air Quality Prediction Of DataLog By Machine Learning”.

[4] Huixiang Lu ,Quing Li & Dongbing Yu (2019), “Air Quality Index and Air Pollutant ConcentrationPrediction Based on Machine Learning Algorithms”.

[5] G. Mani, J. K. Viswanadhapalli, and A. A. Stonie, “Prediction and forecasting of air quality index in Chennai using regression and ARIMA time series models,” Journal of Engineering Research, vol. 9, 2021.

[6] C. R. Aditya, C. R. Deshmukh, N. D K, P. Gandhi, and V. astu, “Detection and prediction of air pollution using machine learning models,” International Journal of Engineering Trends and Technology, vol. 59, no. 4, pp. 204–207, 2018.

Acknowledgment

We have great pleasure in presenting the mini project report on AQI and Heatwaves Prediction Using ML. We take this opportunity to express our sincere thanks to our project guide **Prof. Krishnapriya S** and our project coordinator **Prof. Deepak S. Khachane**, Department of Computer Engineering, APSIT, Thane for providing the technical guidelines and suggestions regarding the line of work. We would like to express our gratitude for his constant encouragement, support and guidance throughout the development of the project.

We thank **Prof. Sachin Malave** Head of Department, Computer Engineering, APSIT, Thane for his encouragement during the progress meeting and for providing guidelines to write this report. We alsothank the entire staff of APSIT for their invaluable help rendered during this work. We wish to express our deep gratitude to all our colleagues at APSIT for their encouragement.

Student Name 1: Hardik Chaudhari
Student ID: 20102113

Student Name 2: Kiran Gode
Student ID: 20102015

Student Name 3: Adwait Bapat
Student ID: 20102158

Student Name 4: Rudra Chavan
Student ID: 20102189