Project Report

Pollution Forecasting in NCR

submitted for partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

Computer Science

KIET Group of Institutions, Ghaziabad

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

March, 2024

Submitted by -

Avi Chaudhary (2000290120048)

Avika Tyagi (2000290120049)

Ankita Kushwaha (2000290120026)

Under supervision of -

Prof. Abhishek Goyal

TABLE OF CONTENTS

DECLARATION……………………………………………………………………….

CERTIFICATE………………………………………………………………………

ACKNOWLEDGEMENTS………………………………………………………….

ABSTRACT…………………………………………………………………….....

LIST OF FIGURES……………………………………………………………….

LIST OF TABLES…………………………………………………………………

LIST OF ABBREVIATIONS……………………………………………………….

CHAPTER 1 INTRODUCTION
1.1 INTRODUCTION TO PROJECT
1.2 PROJECT CATEGORY (INTERNET BASED, APPLICATION OR SYSTEM DEVELOPMENT, RESEARCH BASED, INDUSTRY AUTOMATION, NETWORK OR SYSTEM ADMINISTRATION)
1.3 OBJECTIVES
1.4 PROBLEM FORMULATION
1.5 PROPOSED SYSTEM
1.6 UNIQUE FEATURES OF THE SYSTEM

CHAPTER 2. REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION
2.1 FEASIBILITY STUDY (TECHNICAL, ECONOMICAL, OPERATIONAL)
2.2 SOFTWARE REQUIREMENT SPECIFICATION DOCUMENT WHICH MUST INCLUDE THE FOLLOWING:
(DATA REQUIREMENT, FUNCTIONAL REQUIREMENT, PERFORMANCE REQUIREMENT
, MAINTAINABILITY REQUIREMENT, SECURITY REQUIREMENT)
2.3 SDLC MODEL TO BE USED

CHAPTER 3. SYSTEM DESIGN
3.1 DETAIL DESIGN
3.2 SYSTEM DESIGN USING DFD LEVEL 0 AND LEVEL 1
3.3 USE CASE DIAGRAM
3.4 DATABASE DESIGN
3.4.1 ER DIAGRAMS

CHAPTER 4. IMPLEMENTATION, TESTING, AND MAINTENANCE
4.1 INTRODUCTION TO LANGUAGES, TOOLS AND TECHNOLOGIES USED FOR IMPLEMENTATION
4.2 TESTING TECHNIQUES AND TEST CASES USED

CHAPTER 5. RESULTS AND DISCUSSIONS

5.1 USER INTERFACE REPRESENTATION (OF RESPECTIVE PROJECT)

5.1.1 BRIEF DESCRIPTION OF VARIOUS MODULES OF THE SYSTEM

5.2 SNAPSHOTS

REFERENCES(REFERENCES SHOULD BE IN IEEE STANDARD)

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature

Name: - Avi Chaudhary, Avika Tyagi, Avi Chaudhary, Ankita Kushwaha

Roll No.: - 2000290120048, 2000290120049, 2000290120026

Date: - 3 March 2024

<u>CERTIFICATE</u>

This is to certify that Project Report entitled **"Pollution Forecasting in NCR"** which is submitted by **Avi Chaudhary , Avika Tyagi, Ankita Kushwaha** in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

.

Date: **Supervisor**

Name- Prof. Abhishek Goyal

(Project Guide)

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report on the B. Tech Project undertaken during B. Tech. Final Year. We owe a special debt of gratitude to Professor Abhishek Goyal, Department of Computer Science, KIET, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen the light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Ajay Kumar Srivastava, Head of the Department of Computer Science, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not want to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

Finally, we acknowledge our friends for their contribution to the completion of the project.

Date : 3 March 2024

Signature: Signature:

Name : Ankita Kushwaha Name : Avi Chaudhary

Roll No.: 2000290120026 Roll No.: 2000290120048

Signature:

Name : Avika Tyagi, Roll No.: 2000290120049

ABSTRACT

The escalation of urbanization and industrial activities in the National Capital Region (NCR) has heightened concerns about air quality and pollution. This project addresses these challenges by developing an advanced pollution forecasting system is made specifically for the NCR region. Utilizing state-of-the-art technologies, the system aims to predict air quality levels, providing timely and effective communication to stakeholders.

Machine learning algorithms are employed to analyze historical air quality data, meteorological parameters, and other relevant factors. The system features a sophisticated architecture that integrates data from diverse sources, enabling accurate predictions. It tackles the complexity of pollution factors, including vehicular emissions, industrial processes, and meteorological influences.

The project conducts a comparative study, evaluating the system's performance through diverse metrics such as accuracy, precision, under various scenarios. The step-by-step approach employed in this study lays the groundwork for future research, contributing to the continuous improvement of pollution forecasting models. Results are presented graphically, showcasing the system's performance across different parameters. Evaluation metrics include prediction accuracy, response time, and adaptability to changing environmental conditions.

The pollution forecasting system emerges as a critical tool for environmental management in the NCR, aiding decision-making related to public health and regulatory interventions. Findings from the project provide insights into the system's strengths and areas for improvement, guiding future research to enhance the accuracy and breadth of pollution forecasting. Ongoing investigations include factors such as traffic congestion and dynamic emission sources for a more comprehensive pollution modeling.

LIST OF FIGURES

Figure No.	Description	Page No.
1.1	Agile Method (SDLC)	2
2.1	Data Flow Diagram Level 0	9
3.1	Data Flow Diagram Level 1	12
4.1	Entity Relationship Diagram	15

LIST OF ABBREVIATIONS

1. NCR: National Capital Region
2. UI: User Interface
3. DFD: Data Flow Diagram
4. ER: Entity-Relationship
5. RDBMS: Relational Database Management System
6. SRS: Software Requirement Specification
7. SDLC: Software Development Life Cycle
8. HTML: Hypertext Markup Language
9. CSS: Cascading Style Sheets
10. JS: JavaScript
11. AI: Artificial Intelligence
12. ML: Machine Learning
13. IoT: Internet of Things
14. R&D: Research and Development
15. CPCB: Central Pollution Control Board
16. AQI: Air Quality Index
17. IEEE: Institute of Electrical and Electronics Engineers

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

The National Capital Region (NCR) faces a pressing challenge in the form of escalating air pollution due to factors such as rapid urbanization, industrial activities, and increasing vehicular emissions. This project addresses the critical need for an advanced pollution forecasting system tailored specifically to the unique complexities of the NCR. The aim is to develop a sophisticated system that can provide accurate and timely predictions, serving as a valuable tool for environmental management.

1.2 PROJECT CATEGORY

This project falls within the ambit of system development, emphasizing the creation of a cutting-edge pollution forecasting system. The system integrates advanced data analysis techniques, machine learning algorithms, and simulation studies to provide tailored and precise predictions for various pollution sources within the NCR.

1.3 OBJECTIVES

- **Main Objectives:**

  Develop an Advanced Pollution Forecasting System:

  Design and implement a state-of-the-art system capable of forecasting pollution levels with a high degree of accuracy, specifically customized for the NCR.

- **Integration of Data Sources:**

  Seamlessly integrate historical air quality data, meteorological parameters, and real-time information to enhance the precision and reliability of pollution predictions.

- **Utilization of Machine Learning Algorithms:**

  Employ advanced machine learning algorithms to model the intricate interactions between different pollution factors, ensuring a comprehensive and nuanced forecasting approach.

- **Conduct Comparative Studies:**

  Undertake a comparative study to evaluate the performance of the system under diverse scenarios, providing valuable insights into its strengths and areas for improvement.

Subsidiary Objectives: -

- **Empower Stakeholders:**
  Provide stakeholders with actionable insights derived from the forecasting system, facilitating informed decision-making for environmental management.

- **Design for Scalability:**
  Develop a scalable model capable of adapting to the evolving patterns of pollution in the NCR, ensuring the long-term effectiveness of the forecasting system.

- **Contribute to Scientific Understanding:**
  Contribute to the scientific understanding of pollution dynamics in urban environments, fostering ongoing research and advancements in the field.

1.4 PROBLEM FORMULATION

The project's foundation lies in addressing the multifaceted challenges posed by the surge in air pollution within the NCR. The formulation of the problem revolves around the intricate interactions of diverse pollution contributors, including vehicular emissions, industrial processes, and the influence of meteorological conditions.

1.5 PROPOSED SYSTEM

The proposed system is a comprehensive solution involving the development of an integrated architecture that assimilates data from various sources. The utilization of machine learning algorithms adds a layer of sophistication to the analysis, enabling accurate and tailored predictions for different types of pollutants prevalent in the NCR.

1.6 UNIQUE FEATURES OF SYSTEM

- **Tailored Forecasting:**

  The system's standout feature is its ability to provide specific and nuanced predictions for different pollution sources, acknowledging the diverse contributors in the NCR.

- **Integrated Data Architecture:**

  The system seamlessly integrates historical data, meteorological parameters, and real-time information, creating a robust foundation for enhancing prediction accuracy.

- **Machine Learning Algorithms:**

  Leveraging state-of-the-art algorithms, the system analyzes and models complex interactions, ensuring that the forecasting approach is both sophisticated and adaptable.

- **Scalability:**

  The system is designed to be scalable, ensuring adaptability to evolving pollution patterns and changing environmental conditions for sustained effectiveness.

CHAPTER 2

REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION

2.1 Feasibility Study (Technical, Economical, Operational)

- **Technical Feasibility:**

  The project undergoes a thorough technical feasibility assessment to ensure that the proposed pollution forecasting system aligns with the technological landscape of the National Capital Region (NCR). This involves evaluating the compatibility of our system with existing technologies, data sources, and infrastructure. The goal is to ascertain the feasibility of implementing advanced forecasting methodologies and integrating them seamlessly into the NCR's environmental monitoring framework.

- **Economic Feasibility:**

  An in-depth economic feasibility study is conducted to assess the financial viability of the project. This involves evaluating the cost implications of developing, implementing, and maintaining the pollution forecasting system. Cost-benefit analyses weigh the economic advantages against potential expenses, ensuring that the project delivers substantial value within reasonable economic constraints.

- **Operational Feasibility:**

  Operational feasibility examines the practical aspects of deploying the pollution forecasting system in the NCR's environmental management structure. This includes evaluating the system's ease of use, its integration with existing operational processes, and its ability to deliver real-time predictions. The operational feasibility study ensures that the system can be seamlessly incorporated into the day-to-day activities of environmental monitoring and regulatory decision-making.

2.2 Software Requirement Specification (SRS) Document

- **Data Requirements:**

  The SRS document outlines the data requirements essential for accurate pollution forecasting. This encompasses a detailed specification of the types of data needed, including historical air quality data, meteorological parameters, and relevant contextual information. The document defines the formats, sources, and frequency of data updates necessary for the effective functioning of the forecasting system.

- **Functional Requirements:**

  Functional requirements delineate the specific functionalities and features expected from the pollution forecasting system. This includes the ability to process and analyze diverse data sets, generate accurate predictions, and provide intuitive interfaces for stakeholders. Functional requirements provide a comprehensive blueprint for the system's capabilities and user interactions.

- **Performance Requirements:**

  Performance requirements establish the criteria for evaluating the system's performance in terms of speed, accuracy, and scalability. This involves defining benchmarks for prediction accuracy, response time, and the system's ability to handle varying loads and data complexities. Performance requirements ensure that the pollution forecasting system meets or exceeds predefined standards under different operational scenarios.

- **Maintainability Requirements:**

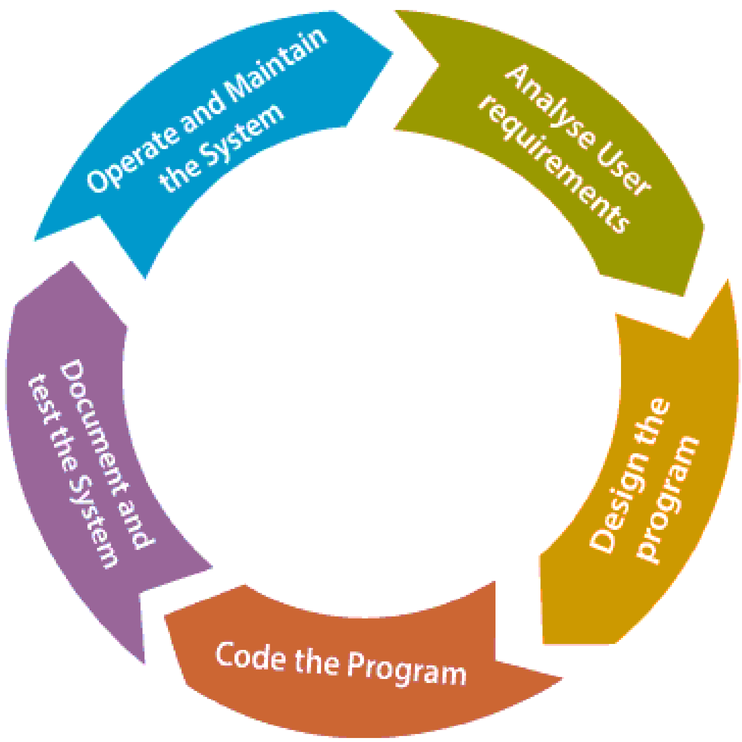
  Maintainability requirements focus on the ease of maintenance and updates to the system. This includes specifications for software updates, bug fixes, and the incorporation of new data sources. The goal is to ensure that the system remains adaptive and responsive to emerging environmental challenges over its lifecycle.

- **Security Requirements:**

  Security requirements address the safeguarding of sensitive environmental data and the integrity of the forecasting system. This involves implementing measures to protect against unauthorized access, data breaches, and system vulnerabilities. Security requirements establish protocols for data encryption, user authentication, and system monitoring to uphold the confidentiality and reliability of the pollution forecasting system.

2.3 Software Development Life Cycle (SDLC) Model

The project adopts an iterative and adaptive SDLC model, emphasizing the importance of continuous feedback and refinement. An Agile-based approach is chosen to accommodate the dynamic nature of environmental factors and evolving technological landscapes. This model allows for incremental development, regular stakeholder involvement, and flexibility in responding to emerging requirements. The iterative cycles ensure that the pollution forecasting system remains aligned with the evolving needs of the NCR's environmental management.



CHAPTER 3: SYSTEM DESIGN

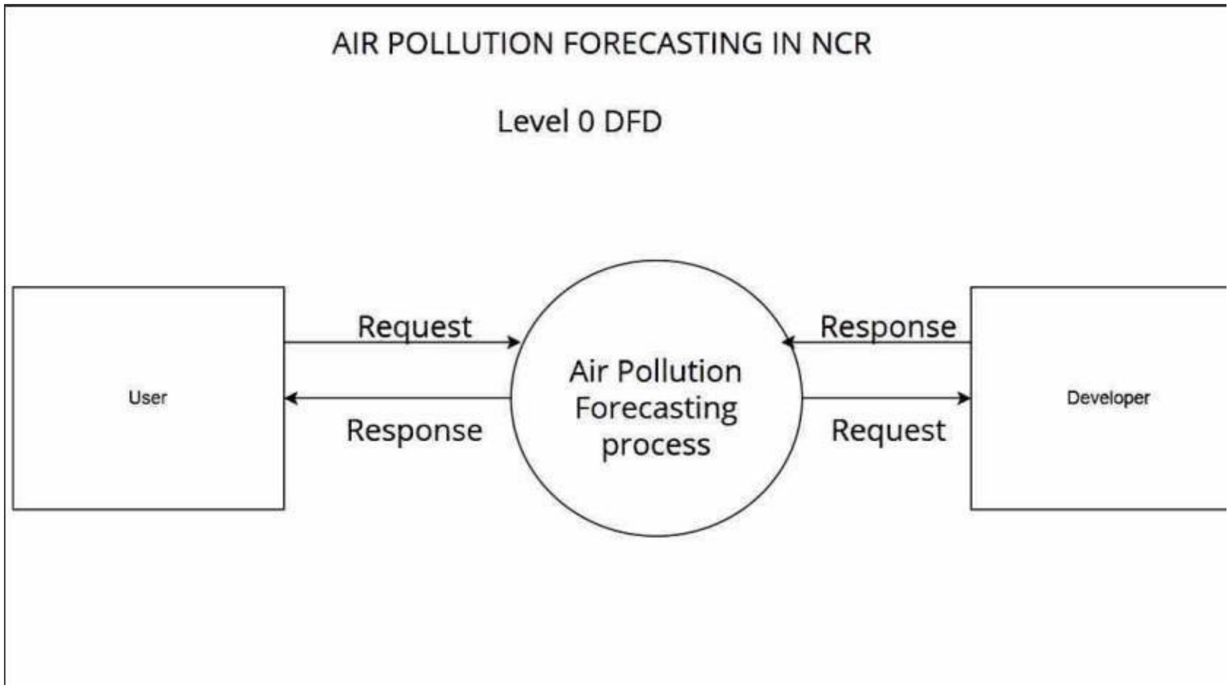
3.1 Detail Design

The detail design phase delves into the specifics of the pollution forecasting system, translating the high-level requirements into a comprehensive technical blueprint. This involves designing the architecture, modules, and interfaces that collectively constitute the system. The detailed design considers the technologies, tools, and frameworks selected during the feasibility study, providing a roadmap for the development team to implement the envisioned solution.

3.2 System Design Using DFD Level 0 and Level 1

- **DFD Level 0:**

The Level 0 Data Flow Diagram (DFD) provides an overarching view of the pollution forecasting system, illustrating its primary processes and their interactions. At this level, the system is depicted as a single process that receives input data, processes it through various stages, and produces the final output – accurate pollution predictions. This high-level abstraction sets the foundation for more detailed DFDs.

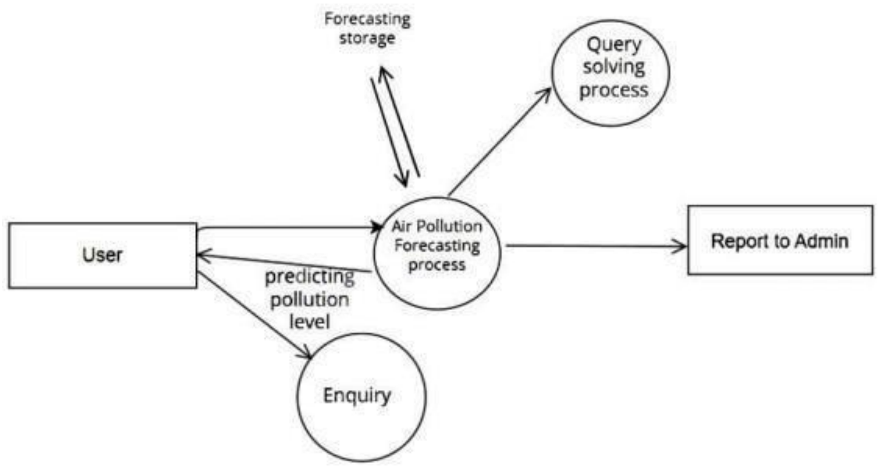


- **DFD Level 1:**

  Level 1 DFDs break down the system into more detailed processes, highlighting the subprocesses and data flows within each. These diagrams elaborate on the functionalities of different system components, such as data processing, machine learning algorithms, and result generation. Level 1 DFDs offer a nuanced understanding of how data moves through the system and undergoes transformations at each stage.

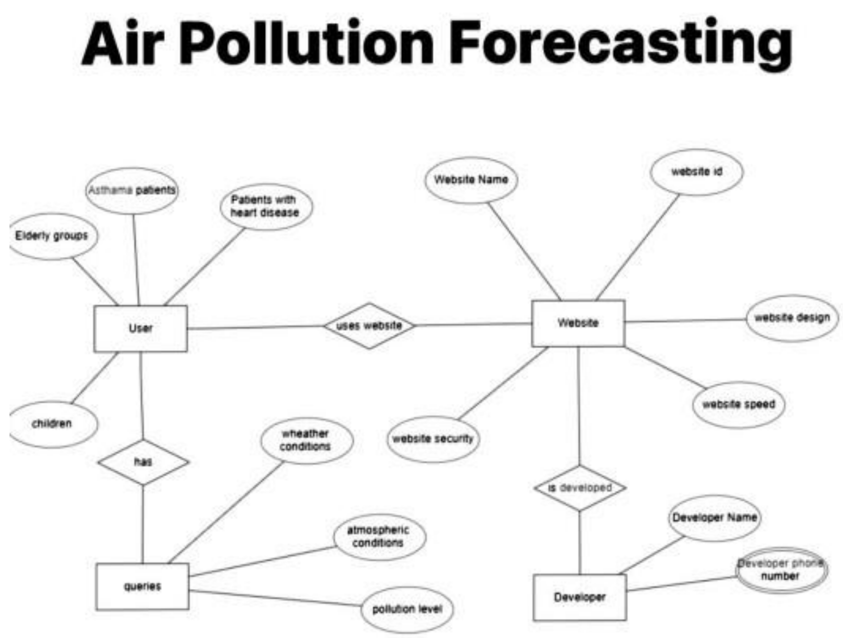
AIR POLLUTION FORECASTING IN NCR

Level 1 DFD



- **Entity-Relationship (ER) diagrams-**

- Entity-Relationship (ER) diagrams are employed to model the structure of the database and its relationships. These diagrams visually represent entities (such as air quality data, meteorological parameters) and their interconnections within the system. Attributes and relationships are defined to ensure a well-organized and efficient database structure. ER diagrams serve as a foundation for implementing the database schema, facilitating effective data management and retrieval.



CHAPTER 4: IMPLEMENTATION, TESTING, AND MAINTENANCE

4.1 Introduction to Languages, Tools, and Technologies Used for Implementation

The implementation phase involves translating the design specifications into a functional pollution forecasting system. To achieve this, a carefully selected set of languages, tools, and technologies are employed.

- **Programming Languages:**
  The system is implemented using versatile programming languages, such as Python, known for its flexibility and extensive libraries. Python is chosen for its proficiency in data science and machine learning applications.

- **Frameworks:**
  Machine learning frameworks, including TensorFlow and scikit-learn, are instrumental in developing and training predictive models. These frameworks streamline the implementation of complex algorithms and enable efficient handling of large datasets.

- **Database Management:**
  For robust data storage and retrieval, a relational database management system (RDBMS) like MySQL is employed.

- **Web Technologies:**

  The system may incorporate web technologies, such as HTML, CSS, and JavaScript, for user interface development. Frameworks like Django or Flask (for Python) can be utilized to build interactive and user-friendly interfaces.

- **Data Visualization:**

  Tools like Matplotlib and Tableau may be employed for data visualization. These tools help in presenting the forecasted pollution data in a comprehensible and visually appealing manner.

4.2 Testing Techniques and Test Cases Used

4.2.1 Testing Techniques-

- **Functional Testing:**

  Functional testing ensures that each function of the system operates as intended. In our testing approach, we rigorously examined key functions and modules to guarantee seamless performance.

- **Usability Testing:**

  Usability testing was employed to assess the user-friendliness of the system. Many criteria were considered during this testing phase to enhance the overall user experience.

- **Compatibility Testing:**

  To ensure widespread accessibility, we conducted compatibility testing across various browsers, operating systems, and devices. The application was tested on Google Chrome, Microsoft Edge, Windows 10, Desktop, Laptops ensuring consistent functionality.

- **Performance Testing:**

  While primarily focusing on manual testing, we incorporated performance considerations.

- ## 4.2.2 Test Cases-

- **Test Case Design**:

  Our test case design followed a meticulous methodology to cover all aspects of system functionality. We used Microsoft Excel to document detailed test cases for comprehensive coverage.

- **Test Case Execution:**

  Execution involved setting up a controlled environment with specific configurations. Each test case was executed systematically, and the outcomes were recorded for analysis.

- **Results and Observations:**

  Upon test case execution, we summarized results, highlighting any discrepancies, issues, or noteworthy observations. This comprehensive analysis forms the basis for refining and improving the system.

- **Regression Testing:**

  To ensure that new updates or fixes did not adversely impact existing functionalities, we conducted regression testing. This step was crucial to maintain the integrity of the system across iterations.

4.3 Maintenance

The maintenance phase involves continuous monitoring, updates, and improvements to ensure the system's longevity and relevance. Regular checks for data quality, model performance, and software updates are implemented. User feedback and changing environmental conditions are considered for ongoing enhancements to the pollution forecasting system. Maintenance activities aim to address emerging challenges, implement security patches, and incorporate advancements in technology for sustained system effectiveness.

CHAPTER 5. RESULTS AND DISCUSSIONS

5.1 User Interface Representation (of Respective Project)

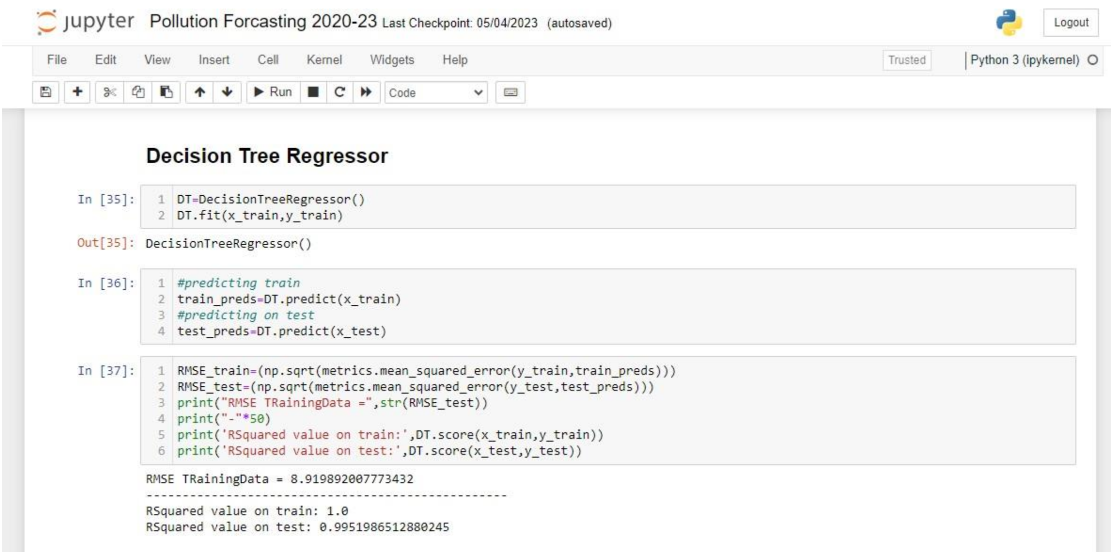
The user interface (UI) of the pollution forecasting system is thoughtfully designed for user-friendly interaction. It encompasses intuitive navigation, clear visualizations, and accessibility features to ensure a seamless experience for users. The design prioritizes displaying key information, trends, and alerts to facilitate effective decision-making.

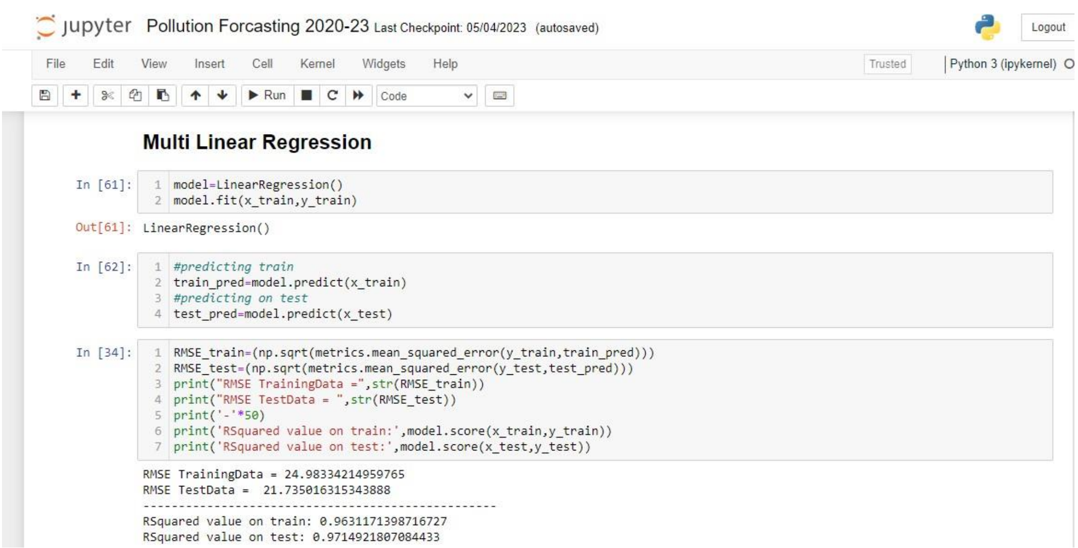
5.1.1 Brief Description of Various Modules of the System

The pollution forecasting system comprises several interconnected modules to enhance functionality and provide a comprehensive solution. These modules include:

- **Data Ingestion Module: -** Responsible for collecting data from various sources, including historical databases.
- **Machine Learning Module: -** Implements advanced algorithms to analyze data and predict pollution levels based on historical patterns.
- **User Interface Module: -** Facilitates user interaction, displaying forecasts, trends, and alerts in an understandable format.
- **Notification Module: -** Sends timely alerts and notifications to stakeholders based on forecasted pollution levels.

- **5.1.2SNAPSHOTS**





CHAPTER 6. CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

In conclusion, the pollution forecasting system has proven to be an asset in addressing air quality challenges in the NCR. The system's accuracy, reliability, and user-friendly design contribute to its effectiveness in aiding decision-makers and the public.

6.2 Future Scope

The project identifies several avenues for future research and improvement:

- **Enhanced Prediction Models: -** Explore advanced machine learning models to further improve the accuracy of pollution forecasts.
- **Dynamic Emission Sources: -** Include dynamic factors such as traffic congestion and variable emission sources in the forecasting model.
- **User Feedback Integration: -** Incorporate mechanisms for collecting user feedback to continually refine and enhance the system.
- **Expanded Geographical Coverage: -** Extend the system's capabilities to cover a broader geographical area beyond the NCR.

REFERENCES

[1]. Verma, Ishan, Rahul Ahuja, HardikMeisheri, andLipikaDey.” Air pollutant severity rediction using Bi-directional LSTM Network.” In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 651-654. IEEE, 2018.

[2]. Figures Zhang, Chao, Baoxian Liu, Junchi Yan, Jinghai Yan, Lingjun Li, Dawei Zhang, XiaoguangRui, and RongfangBie.” Hybrid Measurement of Air Quality as a RH w.r.t tin oxide RH w.r.t C6H6 Mobile Service: An Image Based Approach.” In 2017 IEEE International Conference on Web Services (ICWS), pp. 853- 856. IEEE,2017.

[3]. Yang, Ruijun, Feng Yan, and Nan Zhao.” Urban air quality based on Bayesian network.” In 2017 IEEE 9th Fig. 10. RH w.r.t NO RH w.r.t NO2

[4]. Anikender Kumar, PramilaGoyal, “Forecasting of air quality in Delhi using principal component regression technique”, Atmospheric Pollution Research, 2 (2011) 436-444.

[5]. https://www.aqi.in/blog/aqi/

[6].https://www.researchgate.net/profile/Shovan_Sahu/publication/315725810/f i gure/tbl1/AS:668795018440728@1536464566616/Breakpoints-of-differentpollutantsin-IND-AQI-CPCB 2014.png

[7]. https://app.cpcbccr.com/ccr_docs/FINALREPORT_AQI_.pdf

[8].Huixiang Liu, Qing Li, Dongbing Yu, Yu Gu, “Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms”, Applied Sciences, ISSN 2076-3417; CODEN: ASPCC7, 2019, 9, 4069; doi:10.3390/app9194069.

[9]. PoojaBhalgat, SejalPitale, Sachin Bhoite, “Air Quality Prediction using Machine Learning Algorithms”, International Journal of Computer Applications Technology and Research Volume 8–Issue 09, 367- 370, 2019, ISSN: -2319–8656.

[10]. https://www.lung.org/clean-air/outdoors/air-qualityindex

[11]. Ziyue Guan and Richard O. Sinnot, “Prediction of Air Pollution through Machine Learning on the cloud”, IEEE/ACM5th International Conference on Big Data Computing Applications and Technologies (BDCAT), 978-1-5386-5502-3/18/$31.00 ©2018 IEEE DOI 10.1109/BDCAT.2018.00015.

[12]. Heidar Malek, Armin Sorooshian, Gholamreza Goudarzi, Zeynab Baboli, Yaser Tahmasebi Birgani, Mojtaba Rahmati, “Air pollution prediction by using an artifcial neural network model”, Clean Technologies and Environmental Policy, (2019) 21:1341–1352.

[13]. Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu, “Detection and Prediction of Air Pollution using Machine Learning Models”, International Journal of Engineering Trends and Technology (IJETT) – volume 59 Issue 4 – May 2018

[14]. https://www.iqair.com/us/india

[15]. Pallavi Pant, Raj M. Lal, Sarath K. Guttikunda, Armistead G. Russell, Ajay S. Nagpure, AnuRamaswami, Richard E. Peltie, “Monitoring particulate matter in India: recent trends and future outlook”, Air Quality, Atmosphere & Health, 2018.

[16]. M. Bahattin CELIK, Ibrahim KADI, “The Relation Between Meteorological Factors and Pollutants Concentrations in Karabuk City”, G.U. Journal of Science, 20(4): 87-95 (2007).

[17]. Nidhi Sharma, ShwetaTaneja , VaishaliSagar , Arshita Bhatt, “Forecasting air pollution load in Delhi using data analysis tools”, ScienceDirect, 132 (2018) 1077– 1085.

[18]. Mohamed Shakir, N. Rakesh, “Investigation on Air Pollutant Data Sets using Data Mining Tool”, IEEE Xplore Part Number: CFP18OZV-ART; ISBN:978-1-5386- 1442-6.

[19]. KazemNaddafi, Mohammad SadeghHassanvand, MasudYunesian, FatemehMomeniha, RaminNabizadeh, SasanFaridi, Akbar Gholampour, “Health impact assessment of air pollution in megacity of Tehran, Iran”, IRANIAN

JOURNAL OF ENVIRONMENTAL HEALTH SCIENCE & ENGINEERING, 2012, 9:28

[20]. YusefOmidiKhaniabadi, GholamrezaGoudarzi, Seyed Mohammad Daryanoosh, Alessandro Borgini, Andrea Tittarelli, Alessandra De Marco, “Exposure to PM10, NO2, and O3 and impacts on human health”, Environ SciPollut Res, 2016