TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Air is one of the most essential natural resources for the existence and survival of the entire life on this planet. As this is the largest growing industrial nation, as known India is producing record amount of pollutants specifically Co2, pm2.5 etc and other harmful aerial contaminants. Air pollution in cities has become a cause for fear and has been a major topic of concern. The Indian air quality standard pollutants are indexed in terms of their scale, these air quality indexes indicates the levels of major pollutants on the atmosphere. The main causes associated with air pollution are the burning of fossil fuels, agriculture, exhaust from factories and industries, residential heating, and natural disasters. We collect the data from the Indian government database and start calculating the individual index of the pollutant for every available data points and find their respective AQI for the region. By predicting the air quality index, we can backtrack the major pollution causing pollutant and the location affected seriously by the pollutant across India. By this we can extract various techniques to obtain heavily affected regions on a particular region. PM2.5 refers to tiny particles in the air that reduce visibility and cause the air to appear hazy when levels are elevated. The levels of PM10 and PM2.5 are also increased in air and reduced the air quality, results in adverse effect on living beings. This system exploits machine learning models to detect and predict the data set consisting of atmospheric conditions. Air pollution can cause long-term and short-term health effects. Air quality evaluation is an important way to monitor and control air pollution. Air Quality Index(AQI), is used to measure the quality of air. Fine particulate matter (PM2.5) is significant among the pollutant index because it is a big concern to people's health when its level in the air is relatively high. The air quality of a particular city selected by the user and groups it into different categories like good, satisfactory, moderate, poor, very poor, severe based on AQI (Air Quality Index).

To reduce the effects of harmful concentrations with the ability to predict the occurrence of peak values of concentration, various models need to be developed. One challenge in this regard for pollution data is diversity of data that exists. We subsequently identify the

accuracy of these models for predication of pollution. Monitoring has become a major job as air pollution has been increasing day by day.

Air pollution plays an important role in health alerts when air pollution levels might exceed the specified levels. Hence, continuous monitoring of the air is necessary. The primary goal is to predict air pollution level in City with the ground data set.

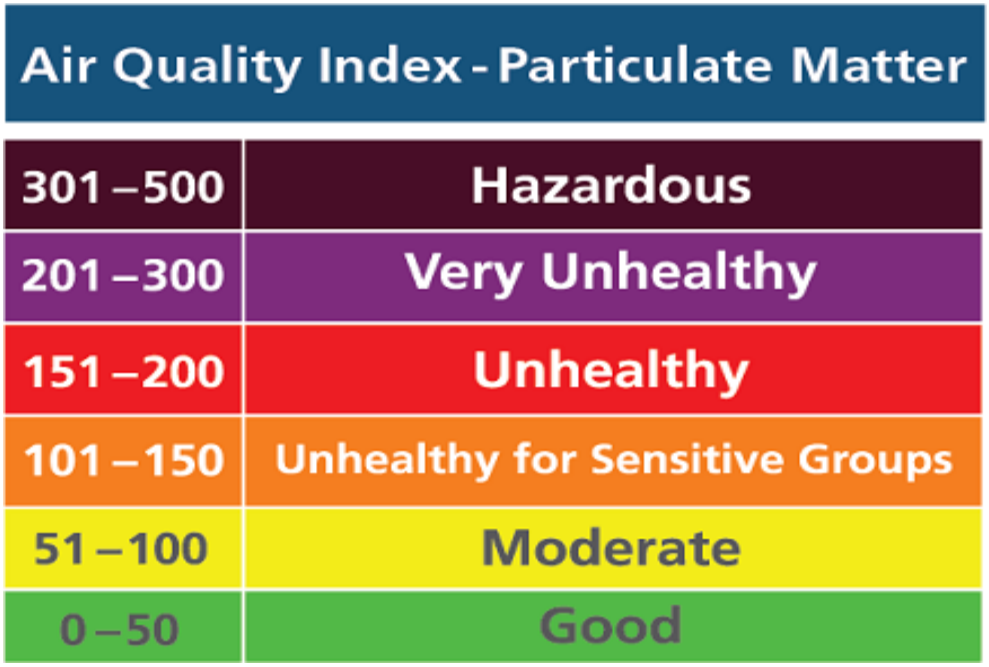

Fig 1.1: AQI

CHAPTER 2
LITERATURE REVIEW

CHAPTER 2

LITERATURE REVIEW

The difficulty of the conventional monitoring instruments is their large size, heavy weight and extraordinary costlier. These lead to inadequate deployment of the monitoring stations. In order to be effective, the locations of the monitoring stations need careful placement because the air pollution situation in urban areas is highly related to human activities (e.g. construction activities) and location-dependent (e.g., the traffic choke-points have much worse air quality than average). IOT Based Air Pollution Monitoring System monitors the Air Quality over a webserver using internet and will activate an alarm when the air quality goes down beyond a certain level, means when there is amount of harmful gases present in the air like CO2, smoke, alcohol, benzene, NH3, NOx and LPG. The system will show the air quality in PPM on the LCD and as well as on webpage so that it can be monitored very easily. Temperature and Humidity is detected and supervised in the system. An Air Pollution Monitoring System for monitoring the combination of major air pollutant gases has been designed, developed, and observed with the wireless standard. This system measures combination of gases such as CO, NO2 and SO2, and using semiconductor sensors. The hardware unit integrates a single-chip microcontroller, air pollution sensors array, a GSM-Module and a GPS-Module. The Central-Server is a high-end personal computer application server with internet connectivity. The hardware unit gathers air pollutants levels (CO, NO2, and SO2), and packs them in a frame with the GPS physical location, time, and date. The frame is finally uploaded to the GSM-Modem and transmitted to the Central-Server via wireless network. The Environmental air pollution has significant influence on the combination of constituents in the atmosphere leading to effects like global warming and acid rains. To avoid such harmful imbalances in the nature, an air pollution measuring system is utmost important. The traditional air quality monitoring system, controlled by the Pollution Control Department, is extremely costlier. Wireless Sensor Networks are a new and very challenging research field for embedded system design automation, as their design must enforce stringent constraints in terms of power and cost. This attempts to develop an effective solution for pollution measuring using wireless sensor networks (WSN). It

focuses on development of a prototype for a Wireless Sensor Network (WSN) that supervises various environmental guidelines of interest in urban areas based on ZigBee protocol.

CHAPTER 3
SYSTEM DESIGN

3.1 System overview

System design is the phase where we address a solution to the problem statement we mentioned earlier and plan the entire process as to achieve all the requirement we specified in the requirement specification stage of our project. In other terms we are starting with a design that would help us understand how to solve all the problems specified by the requirements specification. The system design is used to understand the different modules in the system and the development of each module and a detailed description of the entire system.

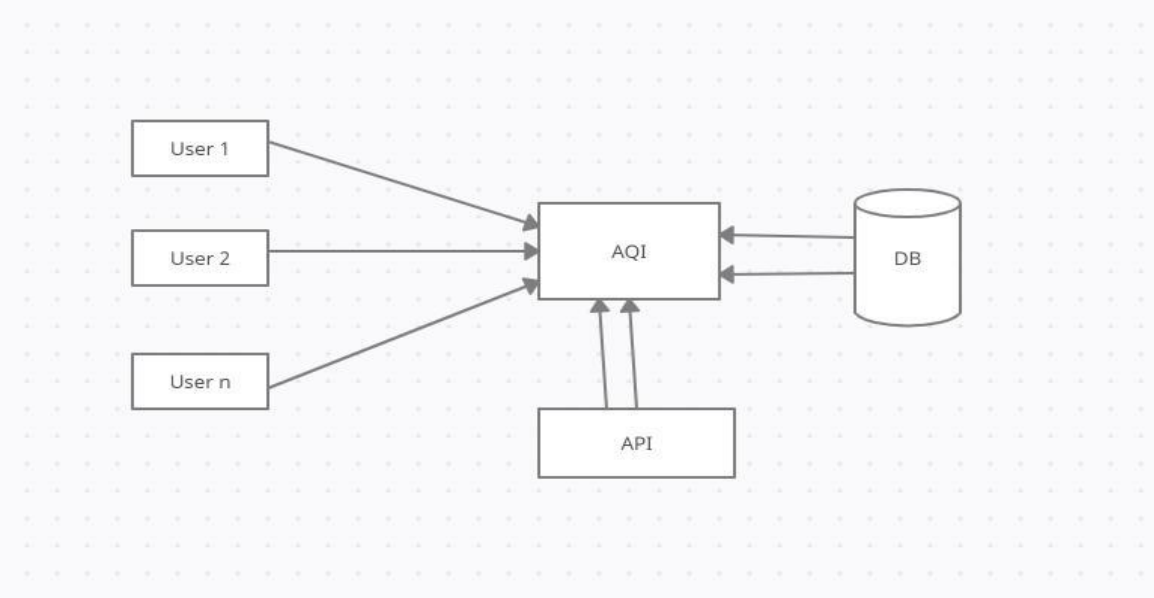

Fig 3.1: System Overview

3.2 Detailed Structure Overview

Detailed structure overview has more text to reach the necessary level of detail about the system's functioning. In the diagram shown in the Fig 4.2 feature extraction process is done by the user who logs in to the system.

The Student Proctoring System is designed to be user specific, depending upon the role of the user who enters the system, several modules will be allocated to that particular user.

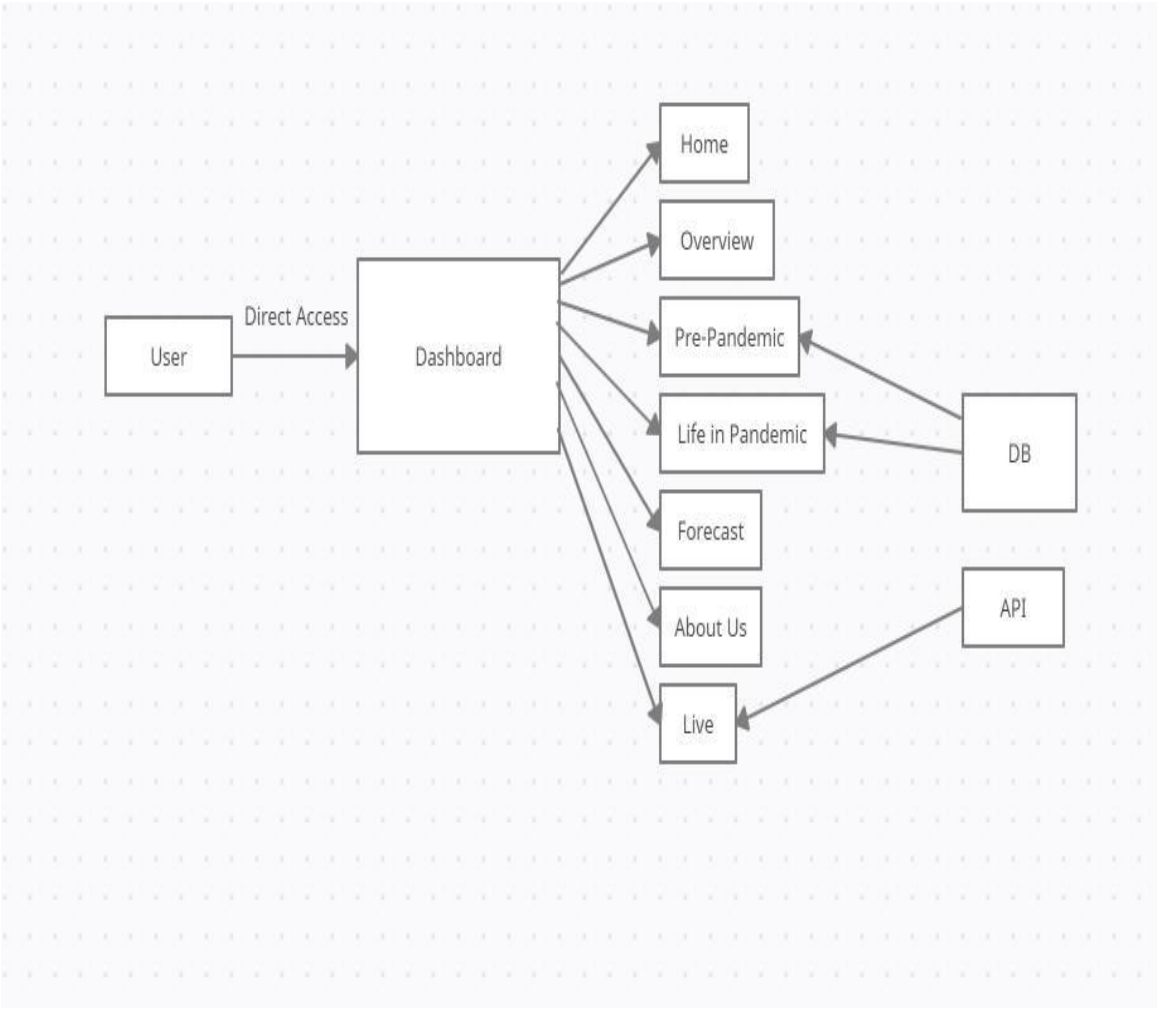

Fig 3.2: Detailed Structure Overview

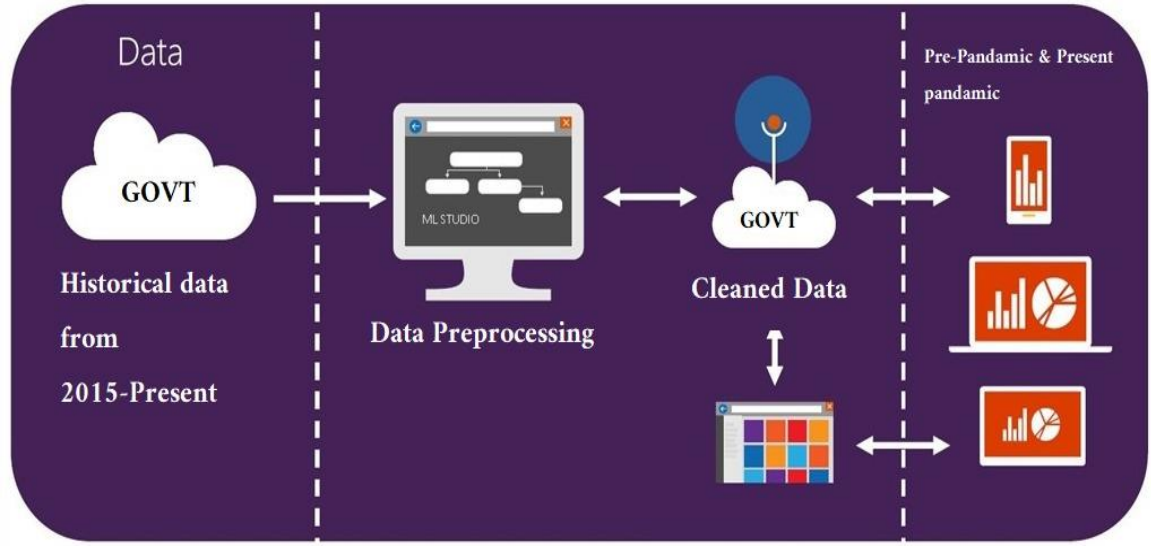

Fig 3.2.1: Forecaste module

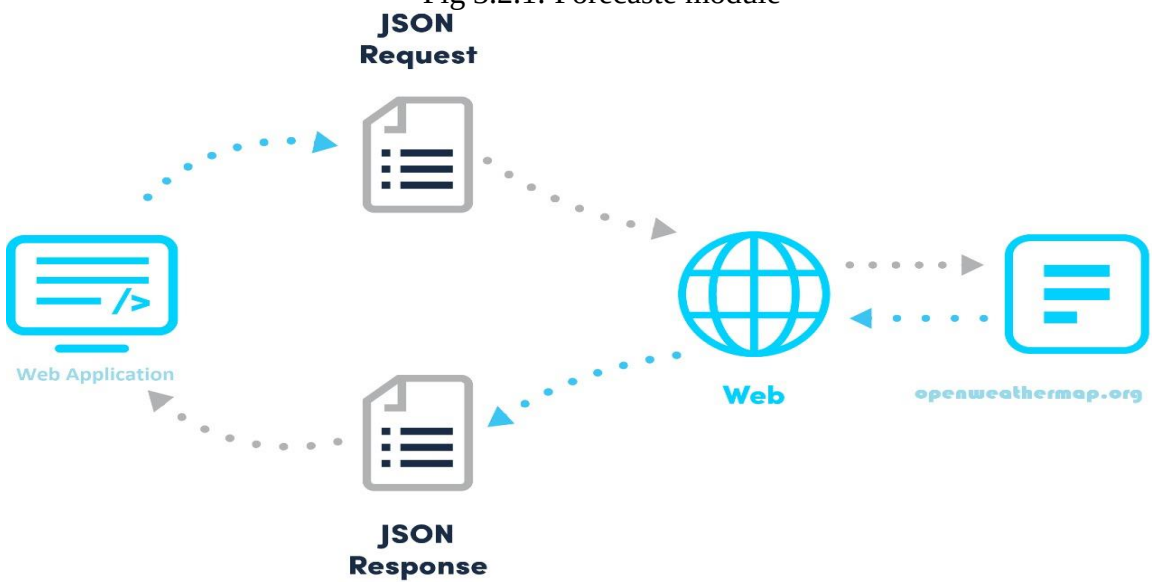

Fig 3.2.2 Live module

3.3 Flow Chart

A flowchart is a formalized graphic representation of a logic sequence, work or manufacturing process, organization chart, or similar formalized structure. The purpose of a flow chart is to provide people with a common language or reference point when dealing with a project or process. Flowcharts use simple geometric symbols and arrows to define relationship

- A flow chart can also be defined as a diagrammatic representation of an algorithm, a step-by-step approach to solve a task.
- There are certain symbols which are used in a flow chart Namely:
- Rectangle: A rectangle represents a process
- Decision: A diamond represents a decision which depends on certain factors in the program.

Connecting arrows: A line is a connector which shows the relationship various processes

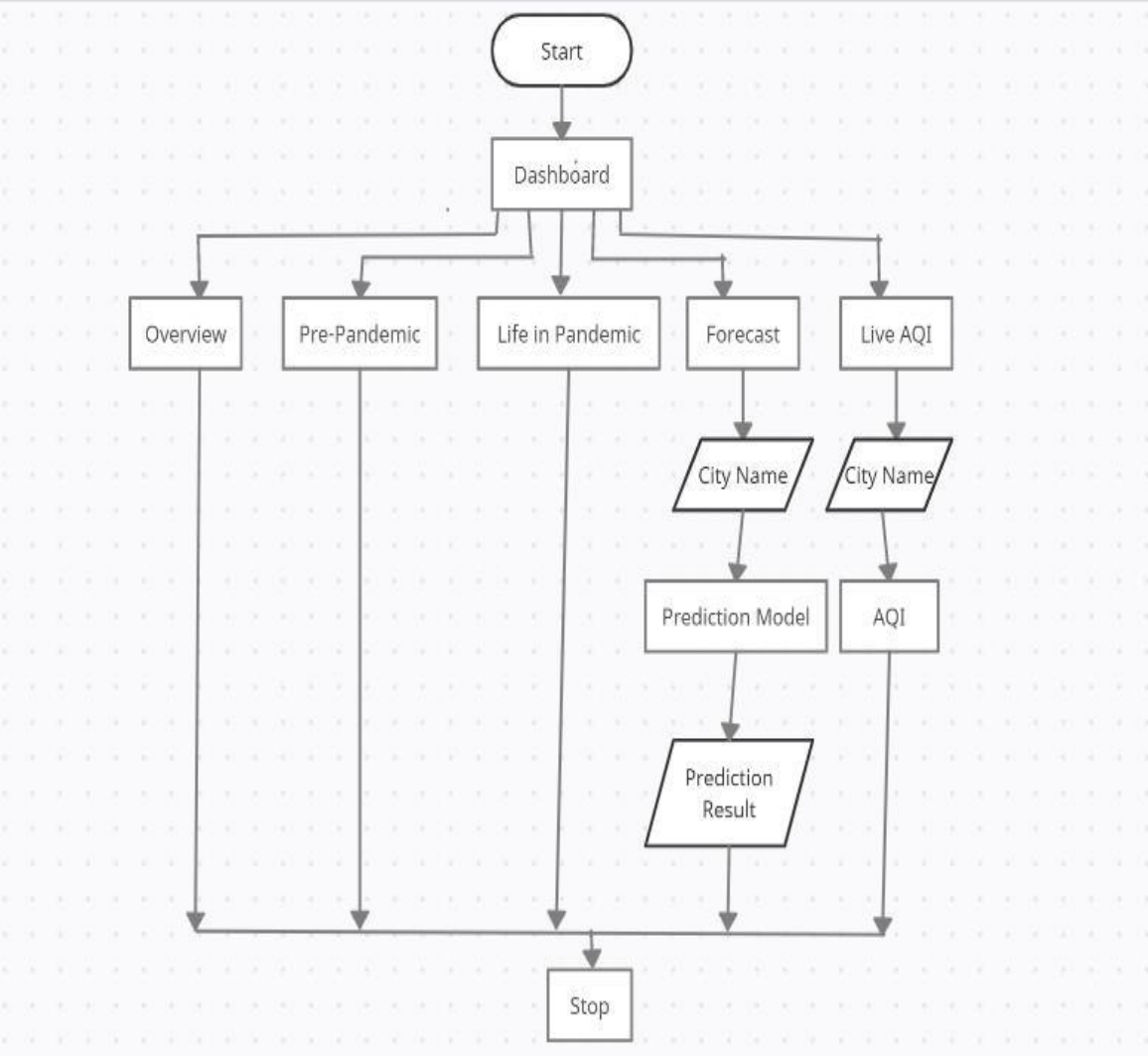

Fig 3.3: Flowchart

CHAPTER4

PROPOSED SYSTEM

Proposed system:

Drawbacks in existing system as follows:

• Inadequate Air Quality Monitoring equipment.

• Inadequate human resources.

• Paucity of funds.

• Inadequate awareness creation.

• Indiscriminate and clandestine quarrying and mining activities.

• Inadequate international co-operation between countries for appropriate technology sharing and transfer.

Proposed system resolves all the drawbacks mentioned above using machine learning and python. As the first point tells, the monitoring equipment is either inadequate or the costing of the equipment is very high which is difficult for installing devices for monitoring the AQI. The proposed system solves this issue by collecting history data and making prediction of the future AQI in the earliest.

Second issue with the existing system is about human resources. The installed device requires constant monitoring with allocated human resource. This leads to a huge dependency on the human resource. This issue is resolved by proposed system as the AQI engine is completely automated which eliminates dependency from the humans for alerting the AQI level.

Inadequate international co-operation between countries for appropriate technology sharing and transfer. This is a dependency of the outside world for sharing real time data which will be used for further alerts. This is resolved in proposed system as the project is integrated with the live rest API for retrieving mere real time AQI level and all the pollutants percentages that qualifies whether the existing is a better place for any individual.

CHAPTER 5

METHODOLOGY

5.1 Challenges of collecting data and cleaning it.

Data collection is the most important part of data science from the integrity and correct of data is important to any kind of data work. The criteria that followed for data cleaning are:

- Validity of data: The data that is obtained should be relevant to the study.
- Uniformity of data: The data obtained from different resources should ideally be of same type.
- Consistency: The data collected should have same format.

The data that was used for the study is from govt of India and some of the other 3rd party websites which provide the history of air quality data for different pollutants.

5.1.1 Data analysis process.

Data analysis is a process of collecting and organizing data to draw helpful conclusions from it. This process of data analysis uses analytical and logical reasoning to gain information from the data. The objective of data analysis is to find meaning in data so that the derived knowledge can be used to make proper decisions.

Methods of data analysis:

1.Collaborate your needs

Begin to analyze data or drill down into any analysis techniques, it’s crucial to sit down collaboratively with all key, decide on your primary campaign or strategic goals, and gain a true understanding of the types of insights that will best benefit progress or provide with the heights of vision needed to evolve.

2.Establishing questions.

Once core objectives are outlined, consider which questions will need answering to help achieve the mission. Data analysis is the most important data analytics techniques that will shape the very foundations of success.

3.Setting up KPI’s

As the data is ready, started to gather the raw data considering to offer potential value, and established clear-cut questions for insights to answer, a host of key performance

indicators (KPIs) that will help track, measure, and shape progress in a number of key areas. KPIs are important to both analysis methods in qualitative and quantitative research.

KPI is one of the primary methods of analyzing data.

5.2 Machine learning algorithms for air quality index prediction:

5.2.1 Linear regression:

Linear Regression is an algorithm based on the machine learning are depends on supervised learning which performs a regression task. Depending on independent variables linear regression gives a target prediction value which is most likely used for finding the relationship among variables and forecasting. Depending on the connection among the established and the independent variables, different regression models differ, they are being considered and list of independent variables used.

$y = mx+c$

In the above expression y indicates labels to data and x indicates the input training data (input parameter). Value of x is used to predict the value of y which gives best fit line for finding the best m and c values during training the model.

c = intercept

m = slope of line

When the best m and c esteems, the best fit line. So when long last utilizing model for expectation, it will foresee the estimation of y for the information estimation of x.

5.2.2 Random Forest Regressor:

Random forest consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. The concept behind random forest is a simple but powerful one.

Need for random forest to perform well are:

1.There must be some actual signal in our features so that models built using those features do better than random guessing, discretized levels. The process of converting regression tasks to classification tasks is problematic, as it ignores the magnitude of the numeric data and consequently is inaccurate. Other researchers have worked on predicting concentrations of pollutants. It focuses on learning multiple tasks that have

commonalities that can improve the efficiency and accuracy of the models. A variety of regularizations can be utilized to enhance the commonalities of the related tasks, including the nuclear norm, spectral norm, Frobenius norm, and so on. However, most of the former machine learning works on air pollutant prediction did not consider the similarities between the models and only focused on improving the model performance for a single task. Therefore, we decided to use meteorological and pollutant data to perform predictions of hourly concentrations on the basis of data models.

CHAPTER 6

SYSTEM REQUIREMENTS SPECIFICATION

4.1 Hardware required

- Processor: Pentium IV/III
- Hard disk: minimum 80GB
- RAM: minimum 2GB

4.2 Software requirement

- Operating System: Windows
- Application: Anaconda
- Tool: Juypter Notebook/Spyder
- Python version: 3.8
- Framework: flask
- Supporting tools: notepad++
- Libraries: numpy, pandas, sklearn, joblib, matplotlib, seaborn, plotly

4.3 Technologies

- Machine Learning
- Data Science
- Web Technologies

4.4 Functional requirements

In the functional requirements we focus on documenting the operations and activities that our project is supposed to perform and they include the following. Functional requirements involve calculations, technical details, data manipulation and processing, and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describe all the cases where the system uses the functional requirements.

4.5 Non – functional requirements

In terms of non-functional requirements, we should be mainly focusing on the performance requirements to be intact.

Performance: - The performance of the Application can be determined by it responsive time, time to complete the given task. For example, when Application is made to start up it shouldn't take more than 3 second to load initial screen. Also, it should be made sure that app will not hindrance to the user Input.

Scalability:- App should able to adopt itself to increased usage or able to handle more data as time progress. For example, when the user data (caches, stored data etc.) increases app should be capable of handling them without delay by optimizing the way storage is done and accessed.

Responsiveness: - Application should be responsive to the user Input or to any external interrupt which is of highest priority and return back to same state. For example, when app gets interrupted by call, then app should able to save state and return to same state/ page which was there before it got interrupted.

Use-ability:- User should be able to understand the flow of App easily i.e. users should able to use App without any guideline or help from experts/manuals. If user experience needs to be explained then its not good UX.

Reliability:- The application should be reliable to perform the business , I.e. when user perform some important action it should be acknowledged with confirmation.

Security :- All the app data should be secured and be encrypted with minimum needs so that it's protected from outside environment also from internal attack.

Availability :- There should be a common plane where the user can access your application to install and look for regular updates give feedback.

Screen Adaption:- Now a days lot of mobile devices comes with different screen sizes and layout , So your application should to able to render it's layout to different screen sizes. Along with automatic adjustment of Font size and image rendering.

Network Coverage:- As we all know all Apps work well with Wi-Fi but also care should be taken care to handle slow connection while experience Wi-Fi black spots or when connected to mobile Network. App should be able to look out for Wi-Fi if not available then automatically switch to mobile network.

Accessibility:- It is a feature which makes physically challenged people make use of your Application.

Performance :- When user opens the app the app should able to load menu within 5 seconds with all thumbnail images , you doesn't want to make customer wait for app to respond for long time.

Reliability :- When user is done with selecting the menu and proceeding to check out there should be a way for user to see summary.

Security:- Users info like personal contact , payment methods should be protected and should not be accessible to unauthorized personals and also there should not be a way for user to manipulate the application for their gain or bypass necessary means.

4.6Software Quality Attributes

Software quality attributes are used to measure the products performance and we need to make sure the software being developed is up to the industry standards and to ensure that is to ensure our system meets all the below mentioned quality attributes and we have explored all these attributes to ensure our system meets these qualities:

i.Reliability

Reliability is an important quality in any product for that matter. Reliability can be defined as the probability that a system performs user required functionality correctly at a specified environment in a given period of time. Since the product is being used for a specific reason if the product isn't reliable then there is no point in using it. People can go for other similar products.

ii.Maintainability

The system being developed should be easier to perform maintenance on. Maintainability refers to the easiness of maintaining a software system. There are two types of software maintenance operations: corrective maintenance, adaptive maintenance. Any issues that may occur shouldn't cause any large-scale damage and any repairs to be done should be easy to perform.

iii.Portability

Portability is one of the biggest advantages with any system. Portability is the property of a software system that can be easily transported from one hardware/software platform to another. If the system can be taken to any place without having to go through a lot of trouble, then that system has the biggest advantage.

CHAPTER 7

SNAP AND SCREENSHOOTS

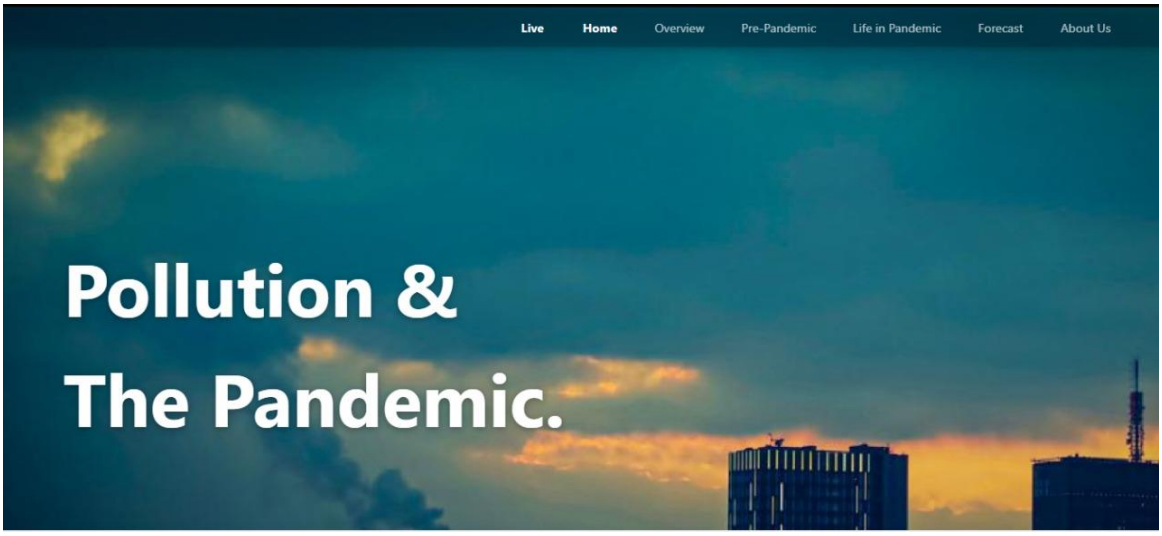

Fig 7.1: Home page

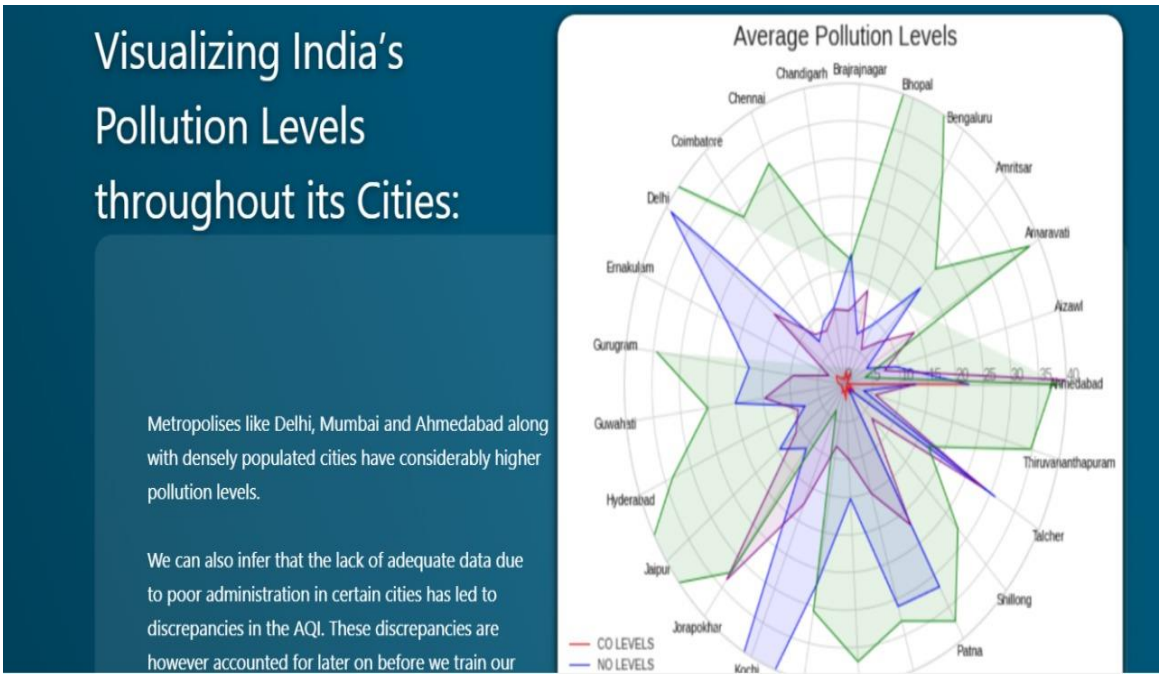

Fig 7.2: overview

Fig 7.3: Prepandamic

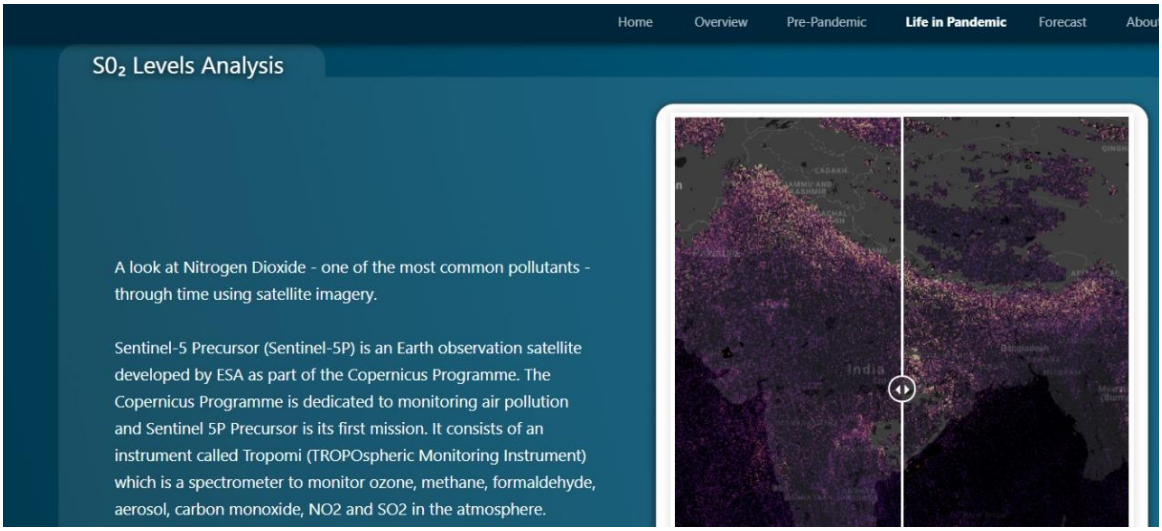

Fig 7.4: Life- in- Pandemic

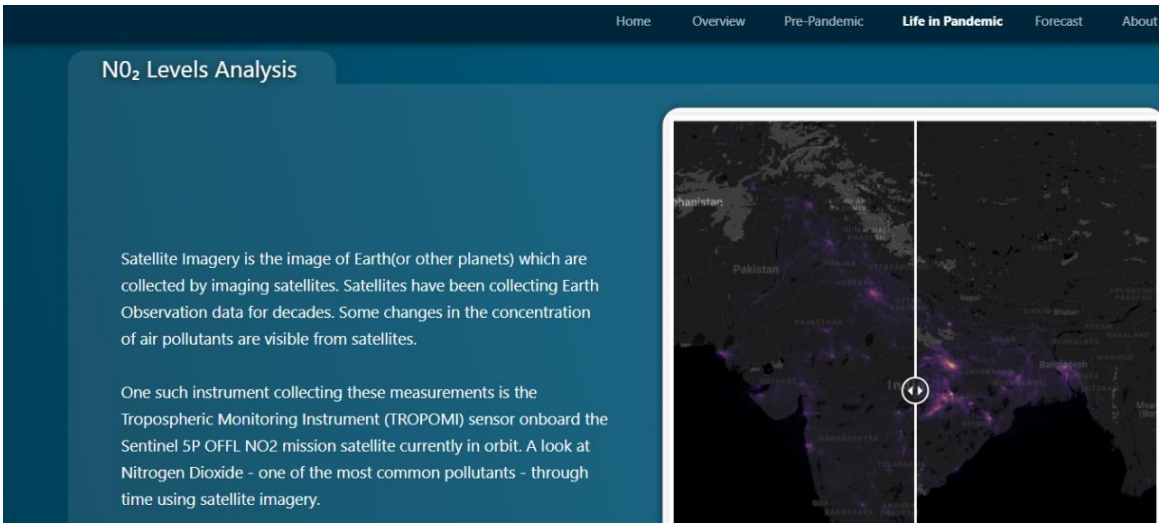

Fig 7.5 : Life- in- Pandemic

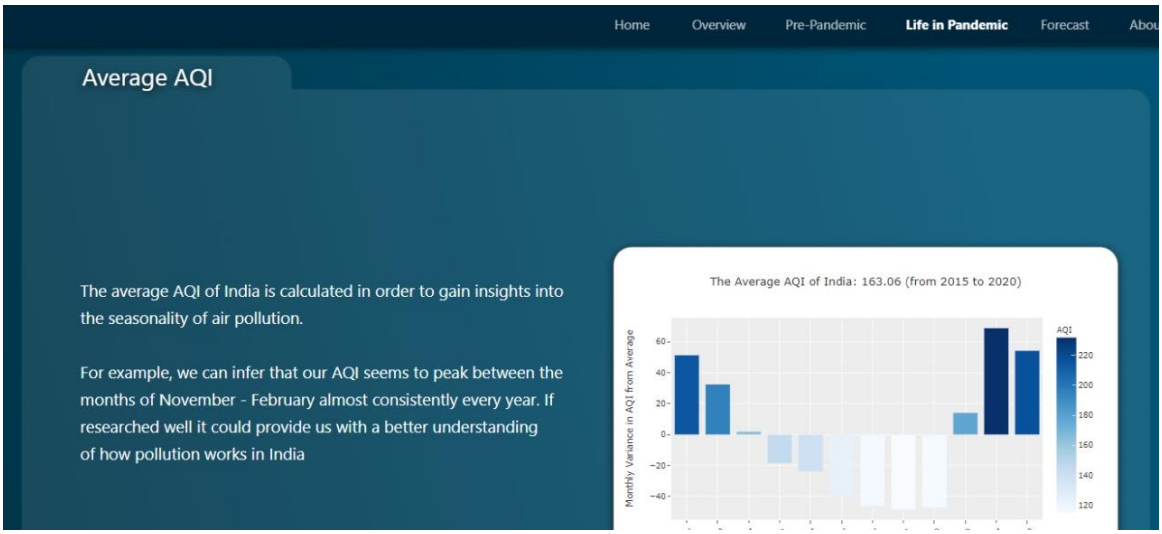

Fig 7.6: Life- in- Pandemic

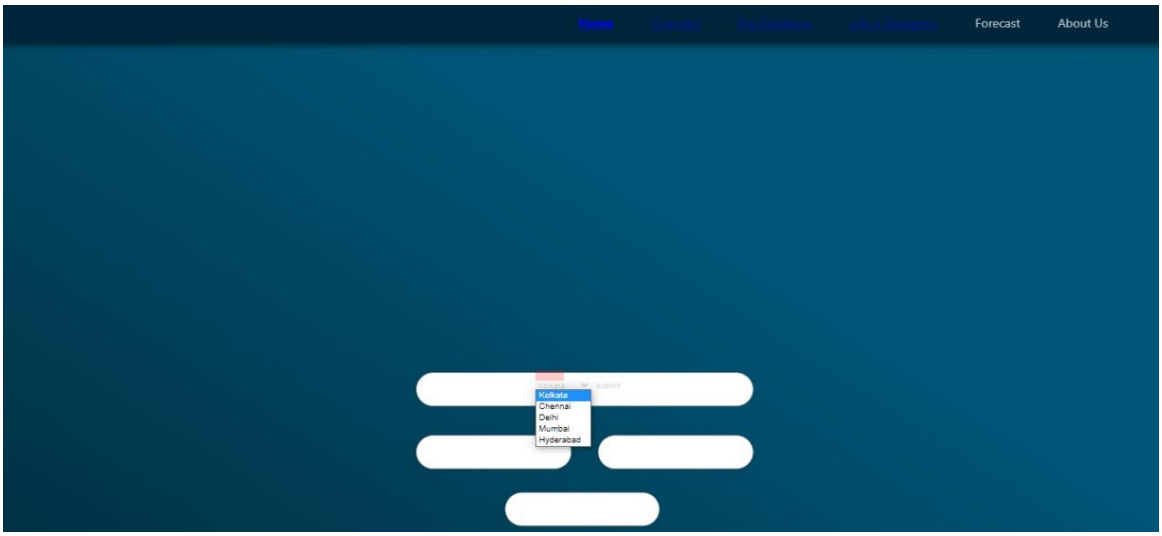

Fig 7.7: Forecast

	co	no	no2	o3	so2	pm2_5	pm10	nh3	city
0	283.72	0.02	7.54	26.82	9.54	10.43	12.39	1.54	Mumbai
1	881.20	0.00	25.36	39.70	28.37	47.22	66.78	10.51	Delhi
2	480.65	0.05	18.34	16.45	18.84	16.28	31.14	5.13	Kolkata
3	534.06	0.05	23.65	16.09	14.54	16.23	23.13	8.23	Chennai
4	273.71	0.01	7.03	16.27	0.68	3.33	3.99	1.22	Bengalūru
5	300.41	0.07	17.82	16.99	18.12	17.29	20.48	1.06	Hyderabad
6	707.63	0.00	21.42	21.46	9.42	36.25	42.99	3.36	Ahmadābād
7	287.06	0.12	10.63	11.80	8.11	5.60	7.01	2.09	Pune
8	467.30	0.44	19.54	6.71	13.47	26.83	29.36	1.54	Sūrat
9	687.60	0.00	6.51	91.55	8.94	64.95	70.92	6.02	Lucknow
10	794.41	0.00	19.02	39.34	12.52	71.23	87.76	5.38	Patna
11	467.30	0.00	4.84	59.37	4.77	35.60	37.26	3.90	Bhopal
12	220.30	0.00	1.65	22.35	1.18	1.70	2.40	0.60	Coimbatore
13	460.63	0.29	27.76	16.63	35.29	22.59	29.38	0.82	Vishākhapatnam
14	360.49	0.00	11.48	22.89	5.07	10.99	14.37	1.43	Kochi
15	273.71	0.00	5.66	25.03	2.00	3.75	6.03	4.24	Madurai
16	273.71	0.00	2.74	34.33	1.13	13.60	14.95	0.28	Rājkot

fig 7.8: Live module

CHAPTER 8

FUTURE WORK

The current proposed system works on static data, where in the dataobtained is from the year 2005 to 2015. The future work would be on streaming data that can actually predict the outcomes of Air Quality Index in real time which can in turn be used to alert people about the air quality in advance so as to prevent from causing health problems.

CHAPTER 9

CONCLUSION

- We have used different algorithm to find the solution and get the required output.
- The different modules has been created to get the required data from the database
- The Live module also helps the people about the current condition to the current location to stay safe or need some precautions.

REFERENCES

[1] QING TAO1, FANG LIU 1, (Member, IEEE), YONG LI, AND DENIS SIDOROV…1 School of Automation, Central South University, Changsha 410083, China" **Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional** GRU"IEEE 2009.

[2] Harsh Gupta, Dhananjay Bhardwaj, Himanshu Agrawal, Vinay Anand Tikkiwal, Arun Kumar" **An IoT Based Air Pollution Monitoring System for Smart Cities**" IEEE 2009.

[3] GALO D. ASTUDILLO 1, LUIS E. GARZA-CASTAÑON 1, AND LUIS I. MINCHALA AVILA 2, (Senior Member, IEEE**)" Design and Evaluation of a Reliable Low-Cost Atmospheric Pollution Station in Urban Environment**" IEEE 2020.

[4] EdoardoArnaudo1,*, Alessandro Farasin 1,2 and Claudio Rossi 1" **A Comparative Analysis for Air Quality Estimation from Traffic and Meteorological Data**" MDPI 2020.

[5] Xiankun Sun1, 2a, Chengfan Li1,3b*, Lan Liu2c ,Jingyuan Yin1,4d, Yongmei Lei1e ,Junjuan Zhao1f,"**Dynamic Monitoring of Haze Pollution Using Satellite Remote Sensing**" IEEE 2019.

[6] Praveen V, Delhi Narendran T, Pavithran R, ChandrasegarThirumalai, IEEE Member "**Data analysis using Box plot and Control Chart for Air Quality**" IEEE 2017.

[7] BO LIU1, (Senior Member, IEEE), SHUO YAN 1, JIANQIANG LI 1, (Senior Member, IEEE), GUANGZHI QU2, (Senior Member, IEEE), YONG LI1, JIANLEI LANG3, AND RENTAO GU 4 "**A Sequence-to-Sequence Air Quality Predictor Based on the n-Step Recurrent Prediction**" IEEE 2019 .

[8] Pearl Pullan, ChitraGautam, VandanaNiranjan," **Air Quality Management System"**IEEE 2020**.**

[9] P. Vijayakumar, AbhinavKhokhar, Archit Pal and MohikaDhawan" **Air Quality Index Monitoring and Mapping Using UAV**" IEEE 2020.

[10] Pattar Sunil Mahesh. 2 PatilBhushanRajendra 3 BodkeAkshayDnyaneshwar 4 Mr. Ulhās. V. Patil**" A SURVEY PAPER ON AIR POLLUTION MONITORING USING IOT**" IJARIIE-ISSN(O)- 2395-4396 (Vol-4) Issue-6 2018.