VISVESVARAYA TECHNOLOGICAL UNIVERSITY
“Jnana Sangama”, Belagavi– 590018, Karnataka

PROJECT REPORT

ON

“Air Quality Prediction Using Machine Learning”

Submitted in partial fulfillment of the requirements for the
Project (15CSP85)

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE & ENGINEERING

Submitted by:
YASHASHWINI R [1JT16CS061]
PALLAVI SINGH [1JT15CS034]

Under the Guidance of
Mr. Srinidhi Kulkarni [Assistant Professor]
Mr. Saravana MK [Assistant Professor]
Department of Computer Science and Engineering

Department of Computer Science and Engineering
Jyothy Institute of Technology
Tataguni, Off. Kanakapura Road,
Bengaluru – 560082
2020 – 2021

JYOTHY INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering

Tataguni, Off. Kanakapura Road,
Bengaluru - 560 082

CERTIFICATE

This is to certify that the project work entitled “**Air Quality Prediction Using Machine Learning**” is a bonafide work carried out by Ms. Yashashwini R [1JT16CS061], Ms. Pallavi Singh [1JT15CS034] in partial fulfillment for the award of **Bachelor of Engineering** in **Computer Science & Engineering** under Visvesvaraya Technological University, Belagavi during the year 2020-2021. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said degree.

Signature of Guide	Signature of Guide	Signature of HOD
....	....	....
Mr. Srinidhi Kulkarni	Mr. Saravana MK	Dr. Prabhanjan. S
Assistant Professor	Assistant Professor	Prof & head of Dept - CSE
Dept of CSE	Dept of CSE	JIT, Bengaluru
JIT, Bengaluru	JIT, Bengaluru	

External Viva

Name of the Examiners   Signature with Date

1)

2)

JYOTHY INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering

Tataguni, Off. Kanakapura Road,
Bengaluru - 560 082

VISION AND MISSION

Vision

To be a center of excellence in Computer Science and Engineering education, focus on research, innovation and entrepreneurial skill development with professional competency.

Mission

M1: To provide state of the art ICT infrastructure and innovative, research-oriented teaching-learning environment and motivation for self-learning & problem-solving abilities by recruiting committed faculty.

M2: To encourage Industry Institute Interaction & multi-disciplinary approach for problem-solving and adapt to ever-changing global IT trends.

M3: To imbibe awareness of societal responsibility and leadership qualities with professional competency and ethics.

DECLARATION

We, hereby declare that the entire work embodied in this report has been carried out by us during IV year of BE degree, at Jyothy Institute of Technology, Bangalore under the guidance of Mr. Srinidhi Kulkarni, Assistant Professor, Dept. Of CSE, JIT and Mr. Saravana MK, Assistant Professor, Dept of CSE, JIT Bangalore, affiliated to Visvesvaraya Technological University, Belagavi. The work embodied in this project work is original and it has not been submitted in part or full for any other degreein any university.

Place: Bengaluru

Date:

YASHASHWINI R [1JT16CS061]
PALLAVI SINGH [1JT15CS034]

ACKNOWLEDGEMENT

This project is a result of accumulated guidance, direction and support of several important persons. We take this opportunity to express our gratitude to all, whose contribution in this project work can never be forgotten.

It is a pleasure to acknowledge the assistance and contribution of various individuals in this project. Firstly, we express our sincere gratitude to **Dr. K. Gopalakrishna,** Principal, JIT for providing us the facilities.

We take this opportunity to express our sincere gratitude to **Dr. Prabhanjan S, Head, Dept. of CSE** for his technical counsel and assistance at every stage of the project.

We extend our profound gratitude to **Dr. Madhu BR**, Assistant Professor, Dept. of **CSE, JIT** for encouragement in order to complete this project.

We also extend our gratitude to **Mr. Srinidhi Kulkarni, Assistant Professor, Dept. of CSE, JIT** and **Mr. Saravana MK, Assistant Professor, Dept. of CSE, JIT** for her personal guidance, keen interest and encouragement which helped us in successfully complete this endeavor.

Last but not the least we would like to thanks our parents and friends without whose constant help, the completion of the project would have not been possible.

YASHASHWINI R [1JT16CS061]
PALLAVI SINGH [1JT15CS034]

ABSTRACT

In this paper, it is aimed to predict the Air Quality Index (AQI) by the use of Machine learning algorithms. To reach this, the key parameters have been selected which can affect the Air quality index are temperature, humidity, pressure, wind speed, PM10 and SO_2 respectively. Air quality of certain states in India can be used as one of the major factors determining pollution index also how well the city's industries and population is controlled. Urbanized Air quality monitoring has been a constant challenge with the advent of industrialization. Air pollution causes conspicuous damage to the environment as well as to human health resulting in acid rain, heart diseases, global warming and skin cancer to all humankind. This paper addresses the challenge of predicting the Air Quality Index (AQI), with the goal to reduce the pollution before it gets unfavorable and also suggests mankind to move places in advance, using ensemble techniques for predicting the Air Quality Index (AQI). This paper investigates how effective some available prediction models are in predicting the Air Quality Index (AQI) values provided some input data, based on the pollution and meteorological information in India. We carry out regression analysis on the dataset, and our results shows which meteorological factors impact the AQI values most and how helpful the predictive models are to help in air quality prediction.

TABLE OF CONTENTS

Sl no.	Description	Page no.
1	**INTRODUCTION**	**3**
	1.1 Introduction	
2	**LITERATURE REVIEW**	**6**
3	**SYSTEM DESING**	**8**
	3.1 system overview	
	3.2 detailed structure overview	
4	**PROPOSED SYSTEM**	**10**
5	**METHODOLOGY**	**12**
	5.1 Challenges of collecting data and cleaning it	
	5.2 Machine learning algorithms for air quality index prediction:	
6	**SYSTEM REQUIREMENTS SPECIFICATION**	**15**
	6.1 Hardware required	
	6.2 Software requirement	
	6.3 Technologies	
	6.4 Functional requirements	
	6.5 Non – functional requirements	
	6.6 Software Quality Attributes	
7	**IMPLEMENTATION**	**27**
8	**SNAP AND SCREEN SHOTS**	**31**
9	**FUTURE WORK**	**35**
10	**CONCLUSION**	**36**
11	**REFERENCES**	**37**

LIST OF FIGURES

Fig no.	Description	Page no.
1.1	AQI	4
3.1	System Overview	8
3.2	Detailed Structure Overview	9
	3.2.1 Forecast module	10
	3.2.2 2 Live module	10
3.3	Flowchart	11
8.1	Home page	19
8.2	overview	19
8.3	Pre pandemic	21
8.4	Life- in- Pandemic	21
8.5	Life- in- Pandemic	21
8.6	Life- in- Pandemic	21
8.7	Forecast	22
8.8	Live module	22

CHAPTER 1

INTRODUCTION

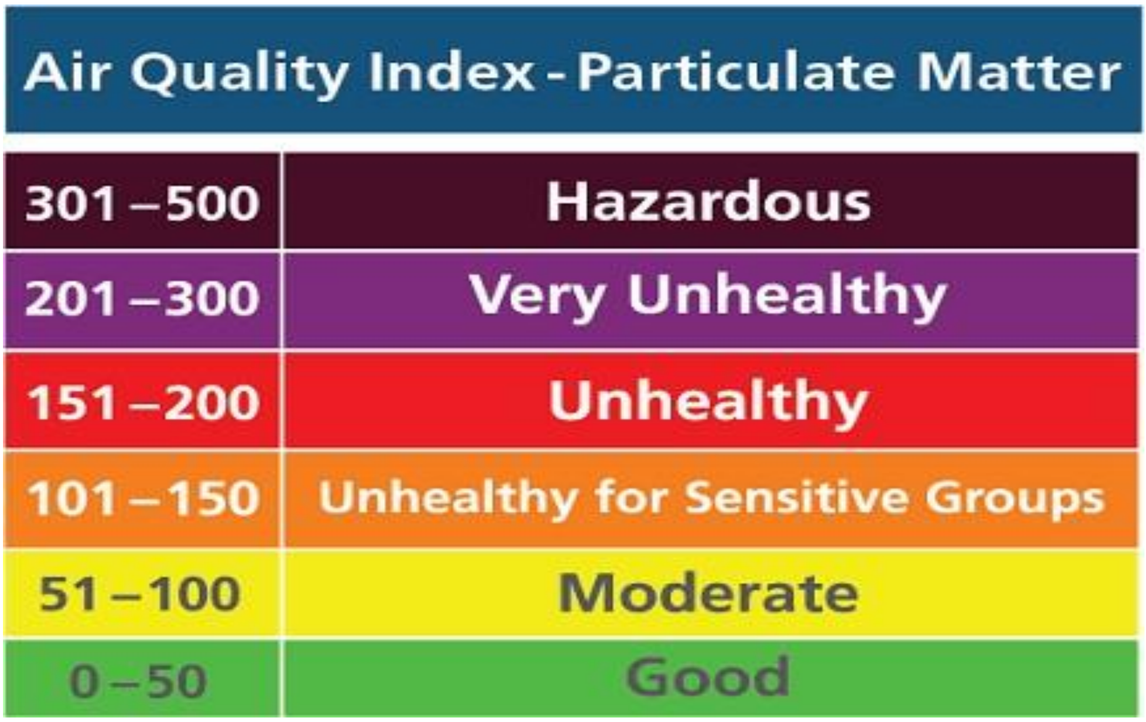
Air is one of the most essential natural resources for the existence and survival of the entire life on this planet. As this is the largest growing industrial nation, as known India is producing record amount of pollutants specifically Co2, pm2.5 etc and other harmful aerial contaminants. Air pollution 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 jobas 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.



Air Quality Index - Particulate Matter

301 – 500	Hazardous
201 – 300	Very Unhealthy
151 – 200	Unhealthy
101 – 150	Unhealthy for Sensitive Groups
51 – 100	Moderate
0 – 50	Good

Fig 1.1: AQI

The AQI is divided into six categories. Each category corresponds to a different level of health concern. Each category also has a specific color. The color makes it easy for people to quickly determine whether air quality is reaching unhealthy levels in their communities.

EPA establishes an AQI for five major air pollutants regulated by the Clean Air Act. Each of these pollutants has a national air quality standard set by EPA to protect public health:

- ground-level ozone
- particle pollution (also known as particulate matter, including PM2.5 and PM10)
- carbon monoxide
- sulphur dioxide
- nitrogen dioxide

Worldwide, air pollution is responsible for around 1.3 million deaths annually according to the World Health Organization (WHO). The depletion of air quality is just one of harmful effects due to pollutants released into the air. Other detrimental consequences, such as acid rain, global warming, aerosol formation, and photochemical smog, have also increased over the last several decades. The recent rapid spread of COVID-19 has prompted many researchers to investigate underlying pollution-related conditions contributing to COVID-19 pandemics in countries. Several shreds of evidence have shown that air pollution is linked to significantly higher COVID-19 death rates, and patterns in COVID-19 death rates mimic patterns in both high population density and high PM2.5 exposure areas.

All the above mentioned raises an urgent need to anticipate and plan for pollution fluctuations to help communities and individuals better mitigate the negative impact of air pollution. To do so, air quality evaluation plays a significant role in monitoring and controlling air pollution. The Environmental Protection Agency (EPA) tracks the commonly known criteria pollutants, i.e., ground-level ozone (O3), Sulphur dioxide (SO2), particulates matter (PM10 and PM2.5), carbon monoxide (CO), carbon dioxide (CO2), and nitrogen dioxide (NO2). These substances are in compositions of a common index, called the Air Quality Index (AQI), indicating how clean or polluted the air is currently or forecasted to become in areas.

As the AQI increases, a higher percentage of the population is exposed. Different countries have their air quality indices, corresponding to different air quality standards. In the United States, the US Environmental Protection Agency monitors six pollutants at more than 4000 sites: O3, PM10, PM2.5, NO2, SO2, and lead. Rybarczyk and Zalakeviciute reviewed a selection of the 46 most relevant journal papers and found more studies with O3, NO2, PM10 and PM2.5, and less on an overall AQI.

The work presented in this paper focuses on the development of AQI prediction models for acute air pollution events 1, 8, and 24 h in advance. The following machine learning (ML) algorithms are investigated, i.e., random forest, adaptive boosting (AdaBoost), support vector machine, artificial neural network, and stacking ensemble methods to train models. As well, this research observes how prediction performance decays over longer time frames, and the precision is measured with three commonly used scale-dependent error indexs.

CHAPTER 2

LITERATURE REVIEW

1. The difficulty of the conventional monitoring instruments is their large size, heavy weightand 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.

2. The global effects of air pollution have led to the implementation of control policies with the installation of certified air monitoring stations that would allow the enforcement of such protocols. However, a large scale deployment of fixed air monitoring stations is not feasible due to economic and/or operational limitations. Subsequently, affordable, automated and portable stations may provide an alternative solution. Nonetheless, the typical sensors embedded in the RAMP monitors (usually electro-chemical or metal-oxide transducers) are affected by the environmental variations and each sensor's limitations. In general, in-situ calibration procedures must include a mechanism to identify what exogenous parameters should be included in the pollutant estimation model. The criterion may combine meteorological information or introduce heuristics such as timestamps. As a result, we consider that taking advantage of the environmental dynamics may improve the accuracy of the estimations. For this reason, this work proposes a combination of meteorological information and heuristics, in addition to a decision rule to find an optimal combination of sensors for the estimation model.

3. 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).

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. This is observed through a small device that can be placed anywhere in a city. First, it is studied the operation of ZigBee protocol. Second, it was chosen and tested a ZigBee module and sensors from the market. Then, it was developed a module that supervises: humidity, temperature, light, carbon monoxide, carbon dioxide and oxygen. These data are measured and sent regularly to a base station connected to a computer. These data are stored and processed for presentation on the Internet in this Environment Observation and Forecasting System (EOFS) is an application for supervising and providing a forecasting about environmental circumstances.

The air pollution monitoring system which involves a context model and a flexible data acquisition policy. The context model is used for understanding the status of air pollution on the remote Place. It can provide an alarm and safety guideline depending on the condition of the context model. It also supports the flexible sampling interval change for effective the tradeoff between sampling rates and battery lifetimes. In this Pollution Map is a new automated system that monitors the air quality of urban cities and displays the information using a web service.

The system collects pollution data using mobile hardware modules, transmits the data regularly using GPRS to a back-end server, and integrates the data to generate a pollution map of the city using its geographical information system. The pollution map is available at any time from an easy-to-view website. The proposed system consists of a Mobile Data Acquisition Unit (Mobile DAQ) and a fixed Internet-Enabled Pollution Monitoring Server (Pollution-Server). The Mobile-DAQ unit combines a single-chip microcontroller, air pollution sensors array, a General Packet Radio Service Modem (GPRS-Modem), and a Global Positioning System Module (GPS-Module). The Pollution-Server is a high-end personal computer application server with Internet connectivity. The Mobile-DAQ unit gathers air pollutants levels (CO, NO2, and SO2), and packs them in a frame with the GPS physical location, time, and date.

4. Anikender Kumar, PramilaGoyal (2011) presented the study that forecasts the daily AQI value for the city Delhi, India using previous record of AQI and meteorological parameters with the help of Principal Component Regression (PCR) and Multiple Linear Regression Techniques. They perform the prediction of daily AQI of the year 2006 using previous records of the year 2000-2005 and different equations. After that this predicted value then compared with observed value of AQI of 2006 for the seasons summer, Monsoon, Post Monsoon and winter using Multiple Linear Regression Technique. Principal Component Analysis is used to find the collinearity among the independent variables. The Principal components were used in Multiple Linear Regression to eliminate collinearity among the predictor variables and also reduce the number of predictors. The Principal Component Regression gives the better performance for predicting the AQI in winter season than any other seasons. In this study only meteorological parameters were considered or used while forecasting the future AQI but they have not considered the ambient air pollutants that may cause the adverse health effects. Huixiang Liu (et al.2019) have taken two different cities Beijing and Italian city for the study purpose. They have forecasted the Air Quality Index (AQI) for the city Beijing and predicting the concentration of NOxin an Italian City depending on two different publicly available datasets. The first Dataset for the period of December 2013 to August 2018 having 1738 instances is made available from the Beijing Municipal Environmental Centre which contains the fields like hourly averaged AQI and the concentrations of PM2.5, O3, SO2, PM10, and NO2 in Beijing. The second Dataset with 9358 instances is collected from Italian city for the period of March 2004 to February 2005. This dataset contains the attributes as Hourly averaged concentration of CO, Non methane Hydrocarbons, Benzene, NOx, NO2. But they focused majorly on NOxprediction as it is one of the important predictor for Air Quality evaluation. They used Support Vector Regression (SVR) and Random Forest Regression (RFR) techniques for AQI and NOx concentration prediction. SVR shows better performance in prediction of AQI while RFR gives the better performance in predicting the NOx concentration.

Ziyue Guan and Richard O. Sinnot (2018) used the various machine learning algorithms to predict the PM2.5 concentration. Data were collected from the official website of Environment Protection Agency (EPA) for the city Melbourne that contains PM2.5 air parameter and they have also collected the unofficial data from Airbeam which is the mobile device used to measure PM2.5 value. The machine Learning Algorithms Artificial Neural Network (ANN), Linear Regression (LR) and Long Short Term Memory (LSTM) recurrent neural network were used for the PM2.5 prediction but out of these algorithms LSTM gives the best performance ad predict the high PM2.5 value with reasonable Accuracy. HeidarMaleki (et al.2019) predicted the hourly concentration values for the ambient air pollutants NO2, SO2, PM10, PM2.5, CO and O3 for the stations Naderi, Havashenasi, MohiteZist and Behdasht in Ahvaz, Iran which is the most polluted city in the world. They have also calculated and predicted Air Quality Index (AQI) and Air Quality Health Index (AQHI) for the four air quality monitoring stations in Ahvaz mentioned above. They used Artificial Neural Network (ANN) machine learning algorithm for the prediction of air pollutants concentration (hourly) and two air quality indices AQI and AQHI over the August 2009 to August 2010.

Input to ANN algorithms involves the factors such as meteorological parameters, Air pollutants concentration, time and date. Aditya C R (et al.2018) employed the machine algorithms to detect and forecast the PM2.5 concentration level on the basis of dataset containing atmospheric conditions in a specific city. They also predicted the PM2.5 concentration level for a particular date . First of all they classify the air as polluted or not polluted by using Logistic Regression algorithm and then Auto Regression algorithm was used to predict the future value of PM2.5 depending upon previous records. Nidhi Sharma (et al.2018) had gone through the detailed data analysis of air pollutants from 2009-2017 and also proposed the critical observation of 2016-1017 air pollutants trend in Delhi, India. They have predicted the future trends of various pollutants as Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), Suspended Particulate Matter (PM), Ozone (O3), Carbon Monoxide (CO) and Benzene. By using data analytics Time series Regression forecasting they have predicted the future values of the pollutants mentioned earlier on the of previous records. According to this study results the AnandVihar and Shadipur monitoring stations of Delhi are under the study. The result shows that there is a drastic increase in PM10 concentration level, NO2 and PM2.5 are evidently increased showing the increased pollution in Delhi. CO is predicted to reduce by 0.169mg/m3 , there is increase in NO2 concentration level for coming years by 16.77 µg/m3 , Ozone is predicted to increase by 6.11mg/m3 , Benzene reduce by 1.33mg/m3 and SO2 is forecasted to increase by 1.24µg/m3 Volume 5, Issue 8, August – 2020 International Journal of Innovative Science and Research Technology ISSN No:-24

Mohamed Shakir and N.Rakesh (2018) have analysed the proportion of various air pollutants (NO, NO2, CO, PM10 and SO2) with respect to the time of the day and the day of the week and estimated the effect of environmental parameters as temperature, wind speed and humidity on the air pollutants mentioned above with the help of WEKA tool. The data was collected from pollution control board of Karnataka. By using ZeroR algorithm in WEKA tool the study come up with the results that shows that the concentration levels of air pollutants increase during the working days and especially during the peak hours of the day and decrease during week-ends or holidays. Using Simple K-means Clustering algorithms the study shows the relationship or dependencies between the environmental factorslike Temperature, wind speed and humidity and the air pollutants like NO, NO2, PM10, CO and SO2. KazemNaddaf (et al.2012) used the AirQ software proposed by WHO that provides the quantitative data on the impact of PM10, SO2, NO2 and O3 on the health of the people in Tehran City of Iran which is the most populated city in Iran. Health impacts under the consideration were all cause mortality, cardiovascular diseases and the respiratory diseases. “The results of the study shows that the air pollutant PM10 had the highest heath impact on the 8,700,000 inhabitants of Tehran City and also caused an excess of total mortality of 2194 deaths out of 47284 in a year”. The total number of excess cases of mortality due to SO2, NO2 and Ozone are 1458, 1050 and 819 respectively. These results shows that Tehran suffered from critical problem of air pollution and for Tehran there is a need to reduce the health burden of air pollution. YusefOmidiKhaniabadi (et al.2016) the main aim of this study is to discover the relation or association between health impacts such as mortality rate of cardiovascular diseases and the air pollutants as PM10, NO2 and O3 over the years 2014 and 2015 for Kermanshah city in Iran. They used the AQI software proposed by WHO for this purpose. The number of premature deaths for cardiovascular diseases is of 188 related to PM10, 33 related to NO2 and 83 related to O3.The results of this study indicates that if there is 10µ/m3 increase in PM10, NO2 and O3 concentration level the mortality risk will

Increase by 1.066, 1.012 and 1.020 respectively.

The prediction of air quality is becoming essential for minimizing the environmental imbalances further effectively addresses the air pollution. There are different type of numerical as well as statistical tools for the prediction and analysis of air pollution. The emergence of advanced computing/analysis techniques from traditional computing methods to recent soft computing techniques are effectively addresses the air quality prediction. The traditional approach for air quality prediction uses mathematical and statistical techniques.

In these techniques, initially a physical model was designed and then data is coded with mathematical differential equations. But such methods suffers from disadvantages like they provide limited accuracy as they were unable to predict the extreme points i.e. the pollution maximum and minimum cut-offs cannot be determined using such approach. Also, such methods were lengthy and inefficient approach for better output prediction. But with the advancement in technology and research, an alternative to traditional methods has been proposed i.e. Artificial Intelligence (AI) techniques can be used for prediction purposes. Among various types of soft computing techniques, the following are the major air pollution predictive model techniques. ¬ Artificial Neural Networks (ANN). ¬ Support Vector Machines (SVM). ¬ Fuzzy Logic (FL). ¬ Hidden Markov Model (HMM). ¬ Genetic Algorithm. ¬ Particle Swarm Intelligence. ¬ Hybrid soft computing techniques.

A. ARTIFICIAL NEURAL NETWORK With the pioneering work of McCulloch & Pitts, Artificial Neural Networks (ANN) has its roots in wide interdisciplinary history from the early 1940's .ANN raised as a mechanism to mimic the human's brain processes. ANN is an intelligent system that has the capacity to learn, memorize and create relationships among the data. ANN is made up by the simple processing units, the neurons, which are connected in a network by a large number of weighted links where the acquired knowledge is stored and over which signals or information can pass. The prediction of air quality, effectively addressed by the prediction of various air pollutants like Sulphur, Nitrogen, carbon monoxide, ozone, suspended particulate matter (SPM) by divided the data set into training , validation and verification further simulation using ANN. ANN was effectively addresses the prediction of Sulphur Dioxide distribution and the future concentration in the air by modeling the Sulphur Dioxide concentration and its distribution from the air pollution station B. SUPPORT VECTOR MACHINES(SVM) Support vector machines (SVMs, also support vector networks are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis.

The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. The SVM model provides a promising alternative and advantageous in times series data analysis for predicting the level of air pollutants. The potential of applying SVM model in ambient air pollutant prediction studied and projected as a most promising approach in prediction of PM10 pollutant.

C. FUZZY LOGIC the term "fuzzy logic" was introduced with the proposal of fuzzy set theory is a form of many-valued logic, deals with reasoning that is approximate rather than fixed and exact. Fuzzy Logic deals with reasoning and provides a better overview in the form of rules that defines all the conditions that are required for predicting the air pollution prediction. In sugarcane processing industry, fuzzy logic can be used to classify and quantify levels of pollution as poor, ordinary, very good and excellent. The Mamdani fuzzy inference system provides the results for prediction of the air quality in and around the sugarcane industry.

D. HIDDEN-MARKOV MODEL (HMM) A Hidden Markov model (HMM) is a classical approach for time series analysis and prediction. A HMM is based on the relationships between the attributes of particular data items and a data set. Hidden Markov Model (HMM), a probabilistic function of a Markov Chain, enables the prediction of PM2.5, using the meteorological measurements and its observation levels.

E. GENETIC ALGORITHM Genetic Algorithm is based on Darwin’s Theory.

It begins with arbitrary created individual population and then fitness is evaluated and parents are selected from the individuals. Genetic Algorithms effectively addresses the change of the accumulation of the surrounding atmosphere and prediction of the thickness of the air pollutants. Genetic Algorithms are effectively applied to extract the optimal feature subset of a large database containing pollutant concentration measurements, and feeds to a nearest neighbor algorithm in order to predict the daily maximum concentration for pollutants.

F. PARTICLE SWARM OPTIMIZATION (PSO) Particle Swarm Intelligence is a populated search method that resembles a school of flying birds. Particles are candidate solutions to problem in hand. Each particle adjusts its flying according to its own flying experience and its companion’s flying experience. A PSO adopted to train multi-layer perceptron to predict air quality parameters more effectively. Particle swam optimization algorithm devised with the characters of pellucid principle and physical explication, to evaluate the grade of atmospheric pollution and multi pollutants.

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.

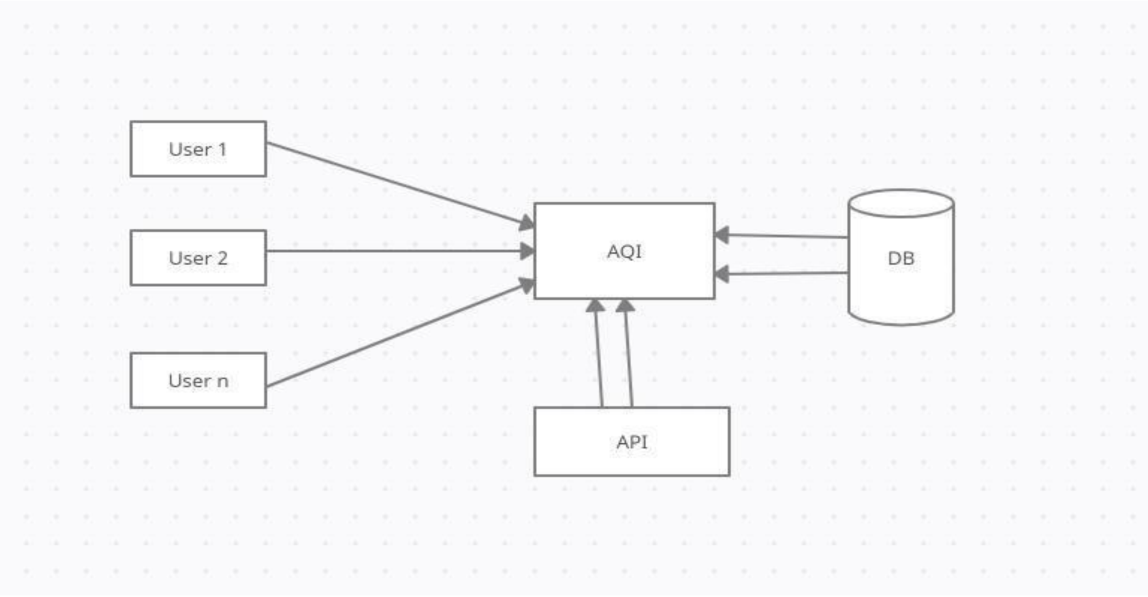


Fig 3.1: System Overview

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.

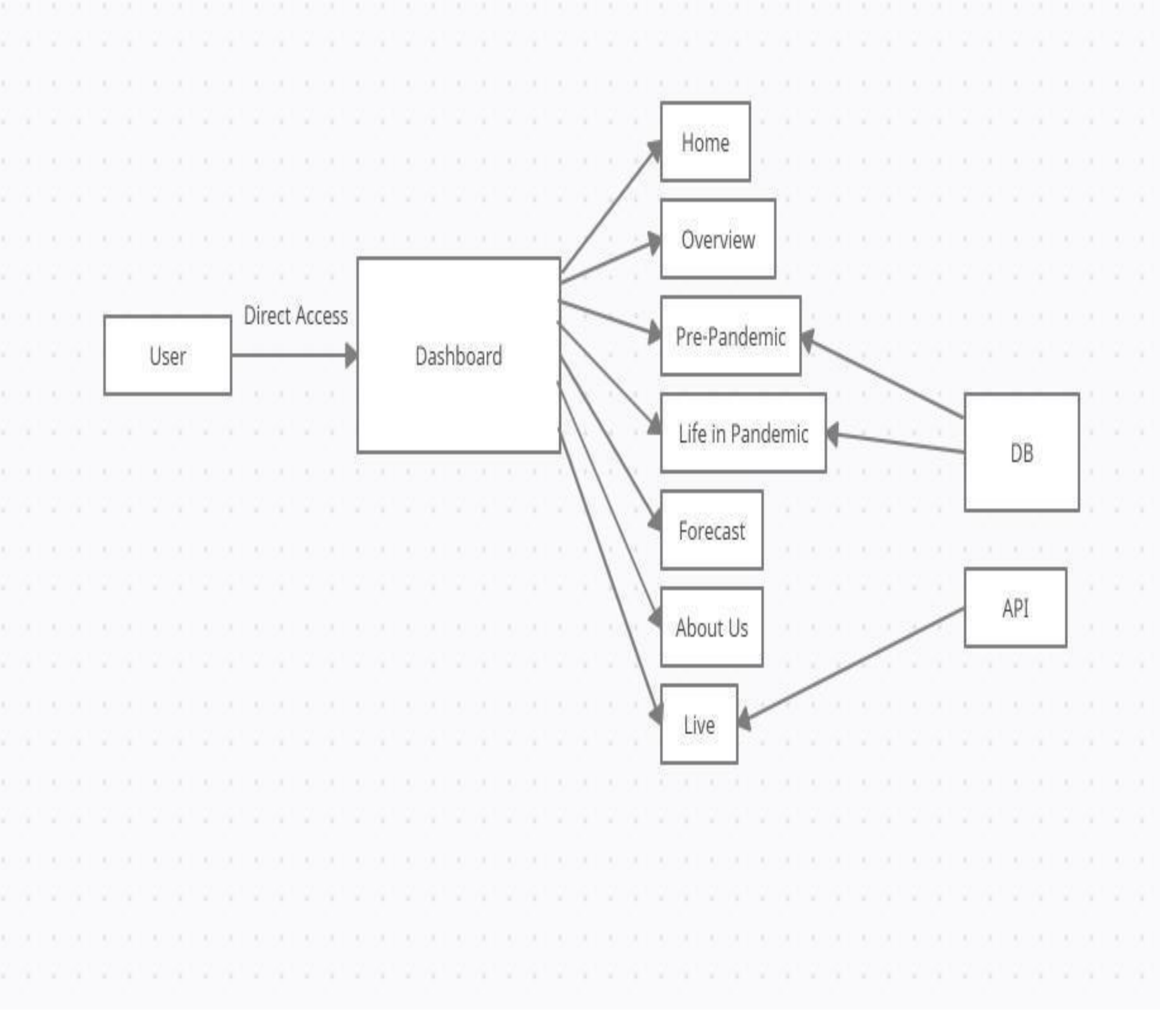


Fig 3.2: Detailed Structure Overview

The complete flow diagram of the web page where in their multi section in it namely.

Home – complete overview of air quality its pollutants which is responsible for air quality index.

Overview – it is a summary of data analysis for the past history data which describes the AQI.

Pre Pandemic -it describe the pollution rate of different cities in India before covid-19.

Life –in pandemic - it describe the pollution rate of different cities in India post covid-19.

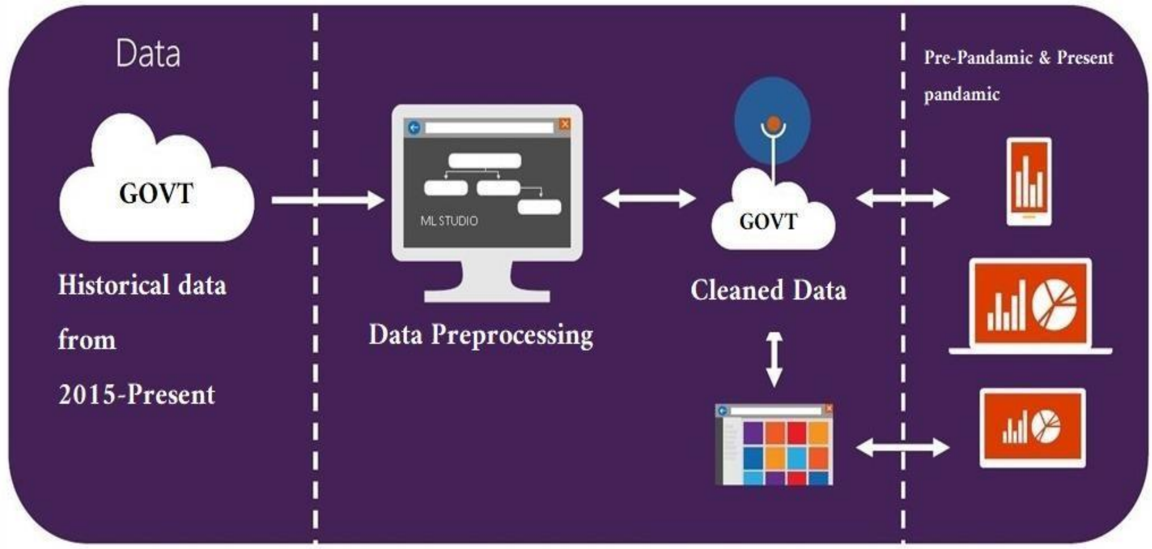

Data
GOVT
Historical data from 2015-Present
ML STUDIO
Data Preprocessing
GOVT
Cleaned Data
Pre-Pandamic & Present pandamic

Fig 3.2.1: Forecaste module

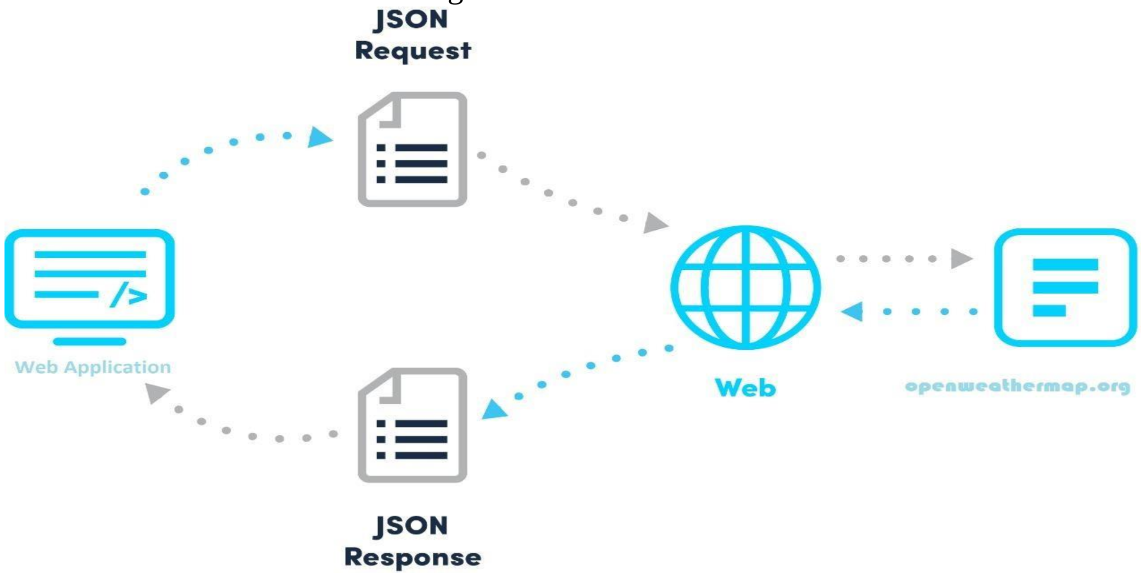


Fig 3.2.2 Live module

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

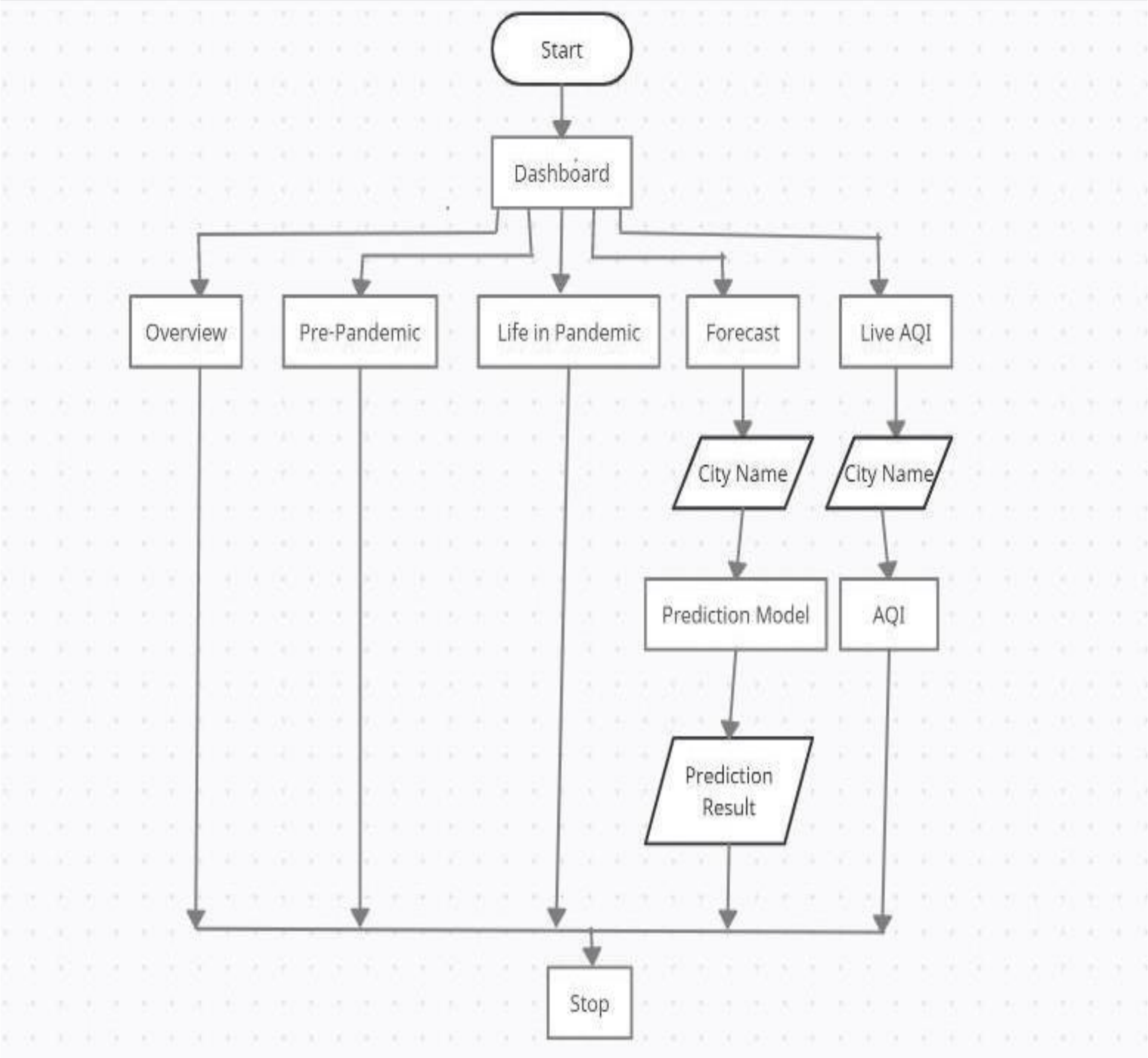


Fig 3.3: Flowchart

CHAPTER 4

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, Fresenius 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 performancefor 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 it’s 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 too 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.

4.6 Software 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 t

CHAPTER 7

IMPLEMENTATION

7.1 Software Implementation

The student Proctoring system was implemented using Python (Django), HTML, CSS, BOOTSTRAP, the process of adopting and integrating a software application into a workflow Prior to implementation, the software should is selected by assessing needs, budget, potential benefits, obstacles, and so forth.

Student Proctoring System (SPS) is a solution tool that is designed to track, maintain and manage all the data generated by a institute, including the grades of a student, their attendance, their activities records, and they can also request for applying leave, etc.,

7.1 UI DESIGN

User interface design is the design of user interfaces for machines and software, such as computers, home appliances, mobile devices, and other electronic devices, with the focus on maximizing usability and the user experience. The goal of user interface design is to make the user's interaction as simple and efficient as possible, in terms of accomplishing user goals (user-centered design). Good user interface design facilitates finishing the task at hand without drawing unnecessary attention to itself. Graphic design and typography are utilized to support its usability, influencing how the user performs certain interactions and improving the aesthetic

Appeal of the design; design aesthetics may enhance or detract from the ability of users to use the functions of the interface. The design process must balance technical functionality and visual elements (e.g., mental model) to create a system that is not only operational but also usable and adaptable to changing user needs.

Interface design is involved in a wide range of projects from computer systems, to cars, to commercial planes; all of these projects involve much of the same basic human interactions yet also require some unique skills and knowledge. As a result, designers tend to specialize in certain types of projects and have skills centered on their expertise, whether that be software design, user research, web design, or industrial design.

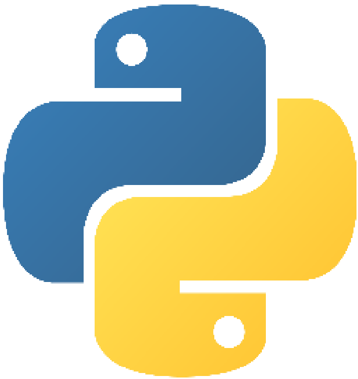
There are several phases and processes in the user interface design, some of which are more demanded Upon than others, depending on the project.

- Functionality requirements gathering
- User and task analysis
- Information architecture
- Usability testing
- Software maintenance

7.1.1 **Programming Languages Python:**

Python is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming.



`

Python 3.8.5

Python is often described as a "batteries included" language due to its comprehensive standard library. Python is often used as a support language for software developers, for build control and management, testing, and in many other ways. SCons for build control.

7.2.2 Integrated development environment (IDE)

Visual Studio Code:

Visual Studio Code is a source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.

Vs Code 3.8.6

Visual Studio Code combines the simplicity of a source code editor with powerful developer tooling, like IntelliSense code completion and debugging.

First and foremost, it is an editor that gets out of your way. The delightfully frictionless edit-build-debug cycle means less time fiddling with your environment, and more time executing on your ideas.

7.2.3 *Database*

SQLite 3.8.7

SQLite stores the entire database (definitions, tables, indices, and the data itself) as a single cross-platform file on a host machine. It implements this simple design by locking the entire database file during writing. SQLite read operations can be multitasked, though writes can only be performed sequentially.

7.2.4 HTML, CSS, JS

3.8.8 Front end:

HTML provides the basic structure of sites, which is enhanced and modified by other technologies like **CSS** and **JavaScript**. **CSS** is used to control presentation, formatting, and layout. **JavaScript** is used to control the behavior of different elements.

Features of HTML:

It is a very easy and simple language. It can be easily understood and modified.

It is very easy to make an effective presentation with HTML because it has a lot of formatting tags.

It is a markup language, so it provides a flexible way to design web pages along with the text.

It facilitates programmers to add a link on the web pages (by html anchor tag), so it enhances the interest of browsing of the user.

CSS Features:

Using **CSS**, you can control the color of the text, the style of fonts, the spacing between paragraphs, how columns are sized and laid out, what background images or colors are used, layout designs, and variations in display for different devices and screen sizes as well as a variety of other effects.

CSS is compatible with all the devices.

With the help of CSS, website maintenance is easy and faster.

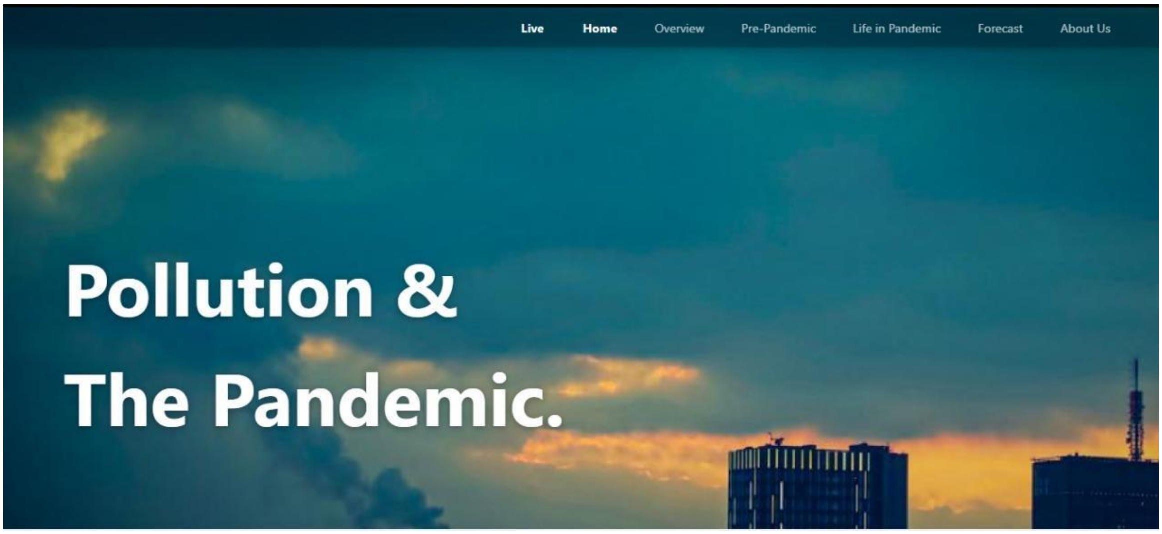
JavaScript Features:

JavaScript was created in the first place for DOM manipulation. Earlier websites were mostly static after JavaScript it was created dynamic Web sites.

Functions in JS are objects. They may have properties and methods just like another objects.

CHAPTER 8

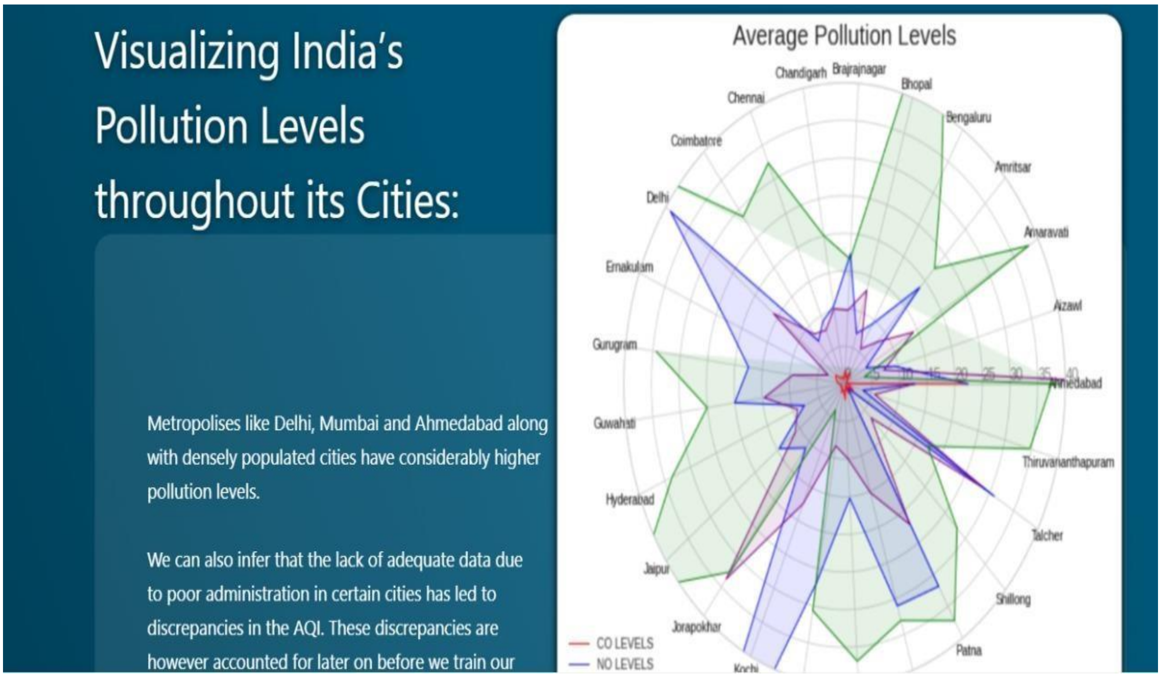
SNAP AND SCREENSHOOTS



Pollution &

The Pandemic.

Fig 8.1: Home Page

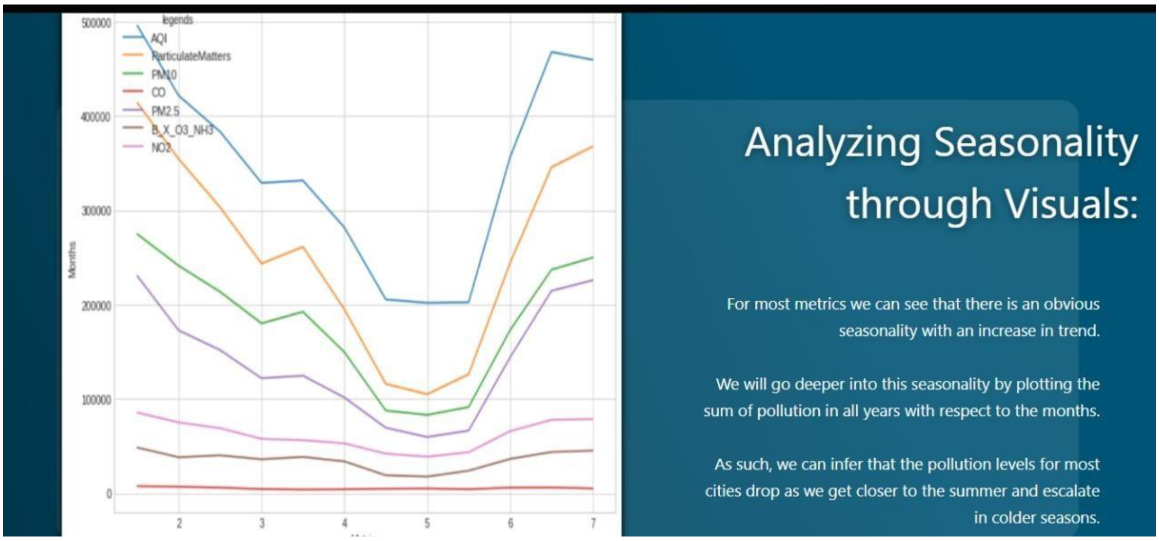


Visualizing India's Pollution Levels throughout its Cities:

Metropolises like Delhi, Mumbai and Ahmedabad along with densely populated cities have considerably higher pollution levels.

We can also infer that the lack of adequate data due to poor administration in certain cities has led to discrepancies in the AQI. These discrepancies are however accounted for later on before we train our

Fig 8.2: Overview



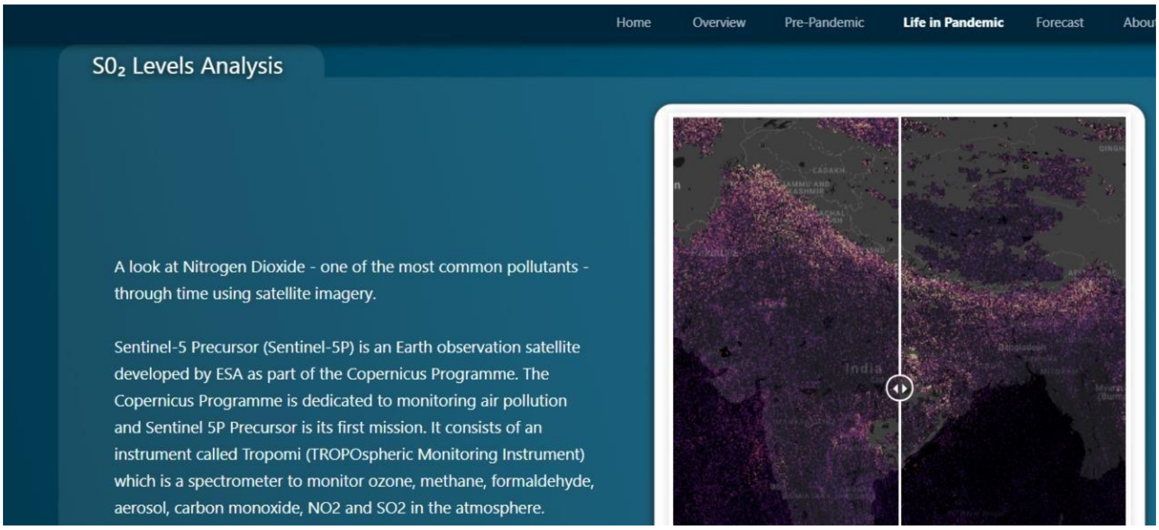
Analyzing Seasonality through Visuals:

For most metrics we can see that there is an obvious seasonality with an increase in trend.

We will go deeper into this seasonality by plotting the sum of pollution in all years with respect to the months.

As such, we can infer that the pollution levels for most cities drop as we get closer to the summer and escalate in colder seasons.

Fig 8.3: Pre-pandemic

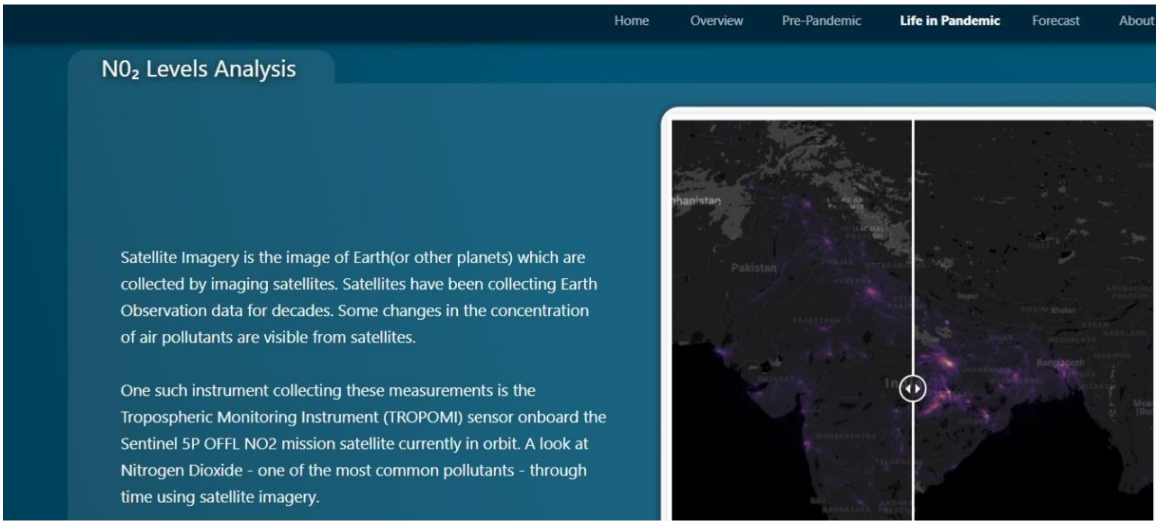


S0$_2$ Levels Analysis

A look at Nitrogen Dioxide - one of the most common pollutants - through time using satellite imagery.

Sentinel-5 Precursor (Sentinel-5P) is an Earth observation satellite developed by ESA as part of the Copernicus Programme. The Copernicus Programme is dedicated to monitoring air pollution and Sentinel 5P Precursor is its first mission. It consists of an instrument called Tropomi (TROPOspheric Monitoring Instrument) which is a spectrometer to monitor ozone, methane, formaldehyde, aerosol, carbon monoxide, NO2 and SO2 in the atmosphere.

Fig 8.4: Life- in- Pandemic

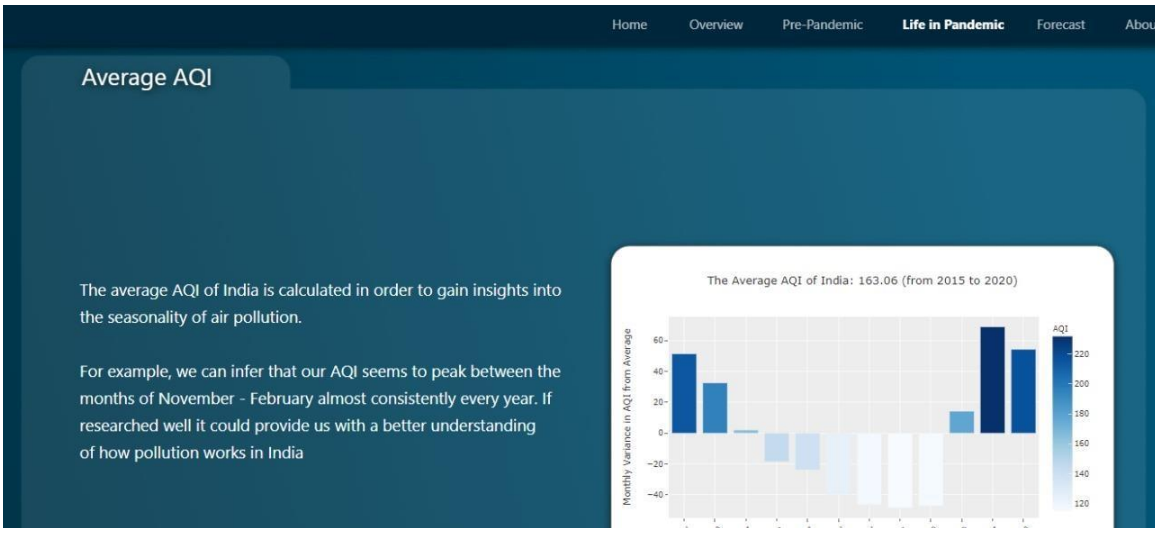


NO₂ Levels Analysis

Satellite Imagery is the image of Earth(or other planets) which are collected by imaging satellites. Satellites have been collecting Earth Observation data for decades. Some changes in the concentration of air pollutants are visible from satellites.

One such instrument collecting these measurements is the Tropospheric Monitoring Instrument (TROPOMI) sensor onboard the Sentinel 5P OFFL NO2 mission satellite currently in orbit. A look at Nitrogen Dioxide - one of the most common pollutants - through time using satellite imagery.

Fig 8.5: Life- in- Pandemic


Home
Overview
Pre-Pandemic
Life in Pandemic
Forecast
Abou
The Average AQI of India: 163.06 (from 2015 to 2020)
Monthly Variance in AQI from Average
60
40
20
0
−20
−40
AQI
220
200
180
160
140
120

Average AQI

The average AQI of India is calculated in order to gain insights into the seasonality of air pollution.

For example, we can infer that our AQI seems to peak between the months of November - February almost consistently every year. If researched well it could provide us with a better understanding of how pollution works in India

Fig 8.6: Life- in- Pandemic

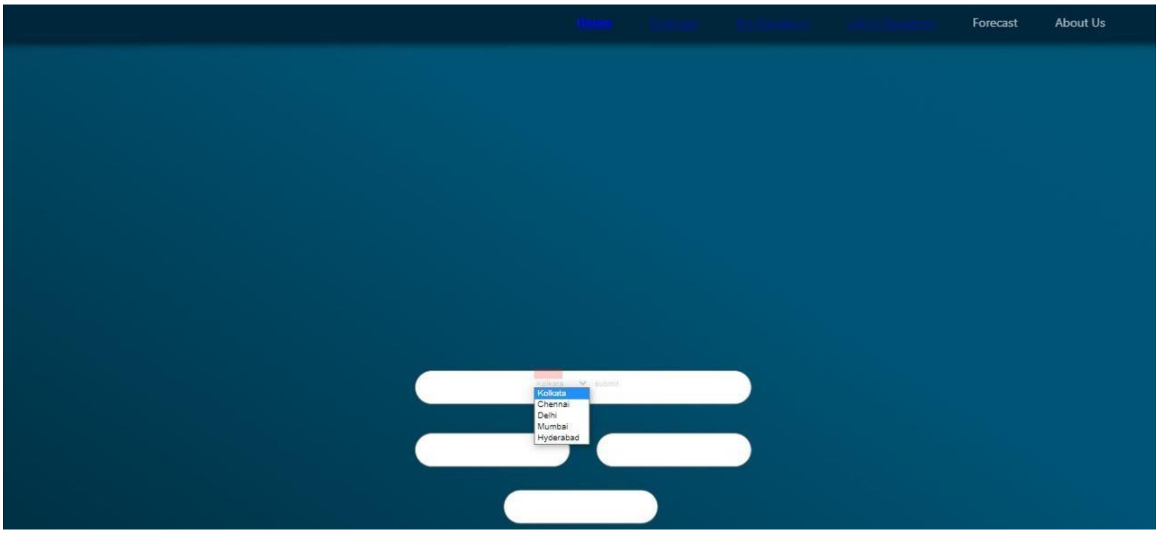
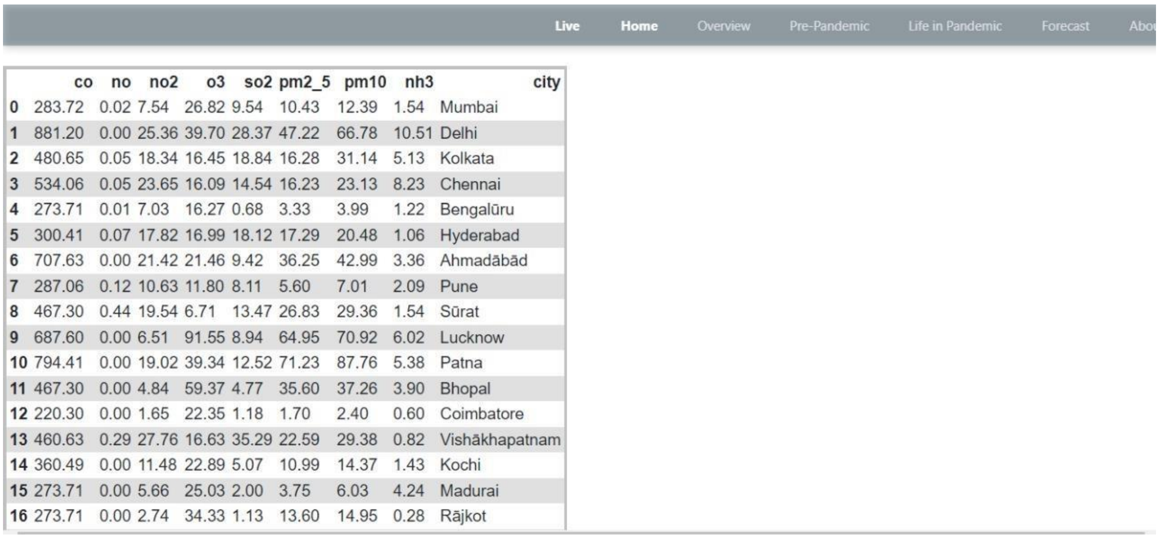


Fig 8.7: Forecast



	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

ig 8.8: Live module

CHAPTER 9

FUTURE WORK

The current proposed system works on static data, where in the data obtained 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 problem.

CHAPTER 10

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.