IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005121|

A Machine Learning and Cloud Based Air Quality Assessment System

Mankar Poonam, Pinjari Sana, Mansuri Saniya, Prof. K.D. Tamhane

UG Student, Dept. of Computer Engineering, PREC, Savitribai Phule Pune University, Maharashtra, India

UG Student, Dept. of Computer Engineering, PREC, Savitribai Phule Pune University, Maharashtra, India

UG Student, Dept. of Computer Engineering, PREC, Savitribai Phule Pune University, Maharashtra, India

Asst. Prof, Dept of Computer Engineering, PREC, Savitribai Phule Pune University, Maharashtra, India

ABSTRACT: Air plays a vital role in the life of a human being as they are dependent on air for living and without it, they will not survive. The quality of air a human inhale also reflects in the quality of life he lives. Due to rapid industrialization urban population has increased manifolds as humans flock to the cities for jobs. This industrialization has also a created a problem of clean air as hazardous gases are sent in the air by these industries. Due to bad air quality, it has given rise to lot of respiratory diseases which is the cause of many deaths each year. The urbanization and industrialization have become an integral part of our life and cannot be stopped but there is a need of a mechanism by using which a human can understand the air quality around him and try to avoid it. Thus, this problem of air pollution can be handled by using latest technologies such as machine learning, cloud computing and mobile computing together to design a system which will alert a user of the quality of the air around him. So, to achieve the target of air pollution prediction system we will design an Android application using which a user will send an air quality assessment request and get and view a response of air quality around him. The admin server which will be a standalone application will first fetch the air quality assessment requests. Then the system will fetch the air quality parameters for a specific address sent to it by the user. Then a training dataset will be created from these parameters for machine learning algorithm LightGBM to train. Then the system will apply LightGBM regression on the training dataset which will return the next values of the air quality parameters according to date. Thus, the air quality assessment results will be sent to the user who can view them on his mobile phone. All this interaction with the admin and user will be handled by using Google cloud platform as cloud provider. Thus, by using our system a user can get correct assessment of the air quality of a certain place and try to avoid it for good.

KEYWORDS: Air Quality, Air Pollution, API, Cloud Computing, Mobile Computing, Machine Learning, Prediction, LightGBM.

I.INTRODUCTION

Today India is growing at a very fast pace so is its population. Due to growing work force and the prosperity many are able to buy new vehicles. India has become an automobile market hub where companies from all over the world are attracted to grab a share. Due to increase in vehicles, there is exponential increase in air pollution. This air pollution becomes very bad in many months in year at specific time of the day. The air pollution has given rise to many respiratory diseases. So, a human should avoid the places with high air pollution and try to take alternate route for work. There should be a mechanism where a citizen can enquire or get forecasted air quality readings.

As the computer has become a daily and necessary part of our life so are the applications and resources it provides to handle many daily problems. Today machine learning has become so powerful that it can analyze any kind of historical data from many years which the humans will difficult to do. There are many machine learning algorithms that can be used to analyze the historical data and predict the future predictions. One of the many machine learning algorithms is LightGBM algorithm which can work on any historical data to give perfect predictions. So, the main motivation of our project is to.

- To develop a better and easy to use air quality assessment system.
- To design a new air quality assessment system using mobile computing, cloud computing and machine learning together.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005121|

- To help a commuter avoid air pollution areas.
- To make use of historical data from as many years as possible.
- To make the air quality assessment system available 24x7 remotely.

Thus, the rest of the paper is structured as follows:

- Section 2. explains literature survey which help us in designing an air quality prediction system using studies published by other researchers.
- Section 3. explains problem definition with goals of the new air quality prediction framework, methodology i.e., mathematical model and algorithms to be used by the system and explains proposed system with block diagram or system architecture and working of the system.
- Section 4. shows the discussion with charts of how much time will be taken to collect and parse historical air quality data and get results from it using LightGBM algorithm.

II.RELATED WORK

This section describes the fundamentals of various techniques that can be used in designing an air quality forecasting using machine learning based LightGBM regression. It helps in understanding various ideas put forward by other researchers and how the drawback in their system can be overcome to design a better and reliable air quality prediction framework. Some of the ideas with technique and drawbacks are mentioned below:

In 2019 ZHANG et al. [1] put forward a paper which mainly emphasises on "LightGBM algorithm-based model to predict air quality" technique to predict air quality. This research is efficient to cover all the things needed for a to develop a successful air quality prediction framework but it lacks to demonstrate how a user can access the system using a simple mobile interface remotely as mobile and cloud computing.

In 2019 Li et al. [2] put forward a paper which mainly emphasises on "air quality prediction using logistic regression and random forest classification" technique to predict air quality. This research is efficient to cover all the things needed for a to develop a successful air quality prediction framework but it lacks to demonstrate how a user can access the system using a simple mobile interface remotely as mobile and cloud computing.

In 2019 Nandini K et al. [3] put forward a paper which mainly emphasises on "air quality prediction using logistic regression and decision tree" technique to predict air quality. This research is efficient to cover all the things needed for a to develop a successful air quality prediction framework but it lacks to demonstrate how a user can access the system using a simple mobile interface remotely as mobile and cloud computing.

In 2020 ZHANG et al. [4] put forward a paper which mainly emphasises on "cancer detection using LightGBM" technique to detect cancer. This research is efficient to cover all the things needed for a to develop a successful air quality prediction framework but it lacks to demonstrate how a user can access the system using a simple mobile interface remotely as mobile and cloud computing.

In 2020 Singh et al. [5] put forward a paper which mainly emphasises on "air quality prediction using machine learning techniques" technique to predict air quality. This research is efficient to cover all the things needed for a to develop a successful air quality prediction framework but it lacks to demonstrate how a user can access the system using a simple mobile interface remotely as mobile and cloud computing.

.

III.PROPOSED ALGORITHM

This section describes all the issues which we can face while designing an air quality prediction system using LightGBM algorithm and how to resolve it using a methodology to solve the problem and design a successful system.

Problem Statement

As the industrialization increases so there is increase in the pollution levels by the gases released from the industries. This gas makes the air quality bad for the human beings to inhale causing respiratory diseases. Industry cannot be stopped and so does urbanization but a alert mechanism can be designed which will predict the next pollution levels of a specific place. By using this kind of system, a commuter can avoid the polluted areas and keep his respiratory health properly. So, a new air quality prediction system can be designed using latest technologies such as cloud computing, machine learning and mobile computing together which will show the next pollution levels to a user.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005121|

This system has to be available 24x7 and can be used on the mobile phone as a mobile app remotely. Thus, by using this system the problem of inhaling bad quality can be avoided. So, our major goals and objectives to solve this problem is stated as follows:

- To predict air quality of a specific location using latitude and longitude.
- To make the system accessible remotely using mobile app.
- To make use of latest technologies such as machine learning, cloud computing and mobile computing together to create a new air quality assessment system.
- To accumulate data needed for machine learning.
- To implement and get results from LightGBM algorithm for prediction.

Methodology

This section will study the mathematical conditions and algorithms to be used for designing a air quality prediction system. These are explained as follows:

Mathematical Model

Our air quality prediction system framework can be explained in two sets with probability, success and failure conditions.

- User Module:
  Set (U)= {U0, U1, U2, U3, U4}

U0∈U = Register and authenticate mobile app.

U1∈U = Enter an address.

U2∈U = Send air quality prediction request.

U3∈U = Get air quality prediction response.

U4∈U = View air quality prediction response.

- Admin Module:
  Set (A)= {A0, A1, A2, A3, A4, U4}

A0∈A = Fetch air quality prediction requests.

A1∈A = Fetch air quality historical data for request.

A2∈A = Create training dataset from historical data.

A3∈A = Train and apply LightGBM.

A4∈A = Get and send air quality prediction response.

U4∈A = View air quality prediction response.

So, by studying the sets we come to notice that many elements are common in both modules and used in coordination in both sets so they be placed as

$$x \in U \cap A \text{ if } x \in U \text{ and } x \in A. \quad (1)$$

Thus, the probability of intersection of elements in both modules can be given as

$$P(U \cap A) = P(U) + P(A). \quad (2)$$

So, intersection of common elements can be shown as

$$U \cap A = \{U4\}. \quad (3)$$

The conditional probability of both modules using the same element can be shown as

$$P(U \mid A) = \frac{P(U \cap A)}{P(A)}. \quad (4)$$

Thus, we conclude that our project “A Machine Learning and Cloud Based Air Quality Assessment System” success and failure will depend upon the internet as our air quality historical data will be fetched using internet, i.e., if

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005121|

the internet connection is not good or not present the historical data will not be fetched and the project won't work, thus this is a case of failure, so our framework supports NP-Hard and not NP-Complete.

Algorithms Used

The project will use LightGBM regression algorithm for successful implementation of the framework.

- LightGBM:

This algorithm will be used to forecast the air quality of a specific location using historical data. LightGBM is also called as Light Gradient Boosting Machine. It uses a gradient boosting framework originally developed Microsoft. It is basically dependent on decision tree algorithms and can be used for machine learning tasks such as ranking and classification. Gradient based one side sampling and exclusive feature bundling are the two techniques used by this algorithm. This algorithm will be implemented in python.

Proposed System

This section is mainly divided in four main modules with other sub modules in them. The modules and submodules are explained using a block diagram or system architecture as shown in Figure.1. to illustrate them. The working of the framework or its blocks can be explained as:

- Admin and User Communication Medium:

This block in system architecture shows how an Android mobile will be used to send air quality prediction request and receive it on mobile. It also shows a standalone admin application which will analyze the user request using LightGBM and send it to the user as response.

- User Application:

This block in system architecture shows how the user will first register and authenticate. Then the user will enter a location address and send the air quality prediction request. Then the user can view the air quality prediction response on his mobile app.

- Admin application:

This block in system architecture shows how a admin will first fetch the air quality prediction request. Then he will select a specific request and fetch air quality historical data for the request address. Then it will create a training dataset which will be used to train LightGBM algorithm. Then LightGBM algorithm will be applied which will return air quality future predictions. The prediction results will be sent to the user who will view it on his mobile.

- Google Cloud Platform:

This block in system architecture shows how the user and admin communication for air quality prediction will be handled using Google cloud platform as the cloud provider.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005121|

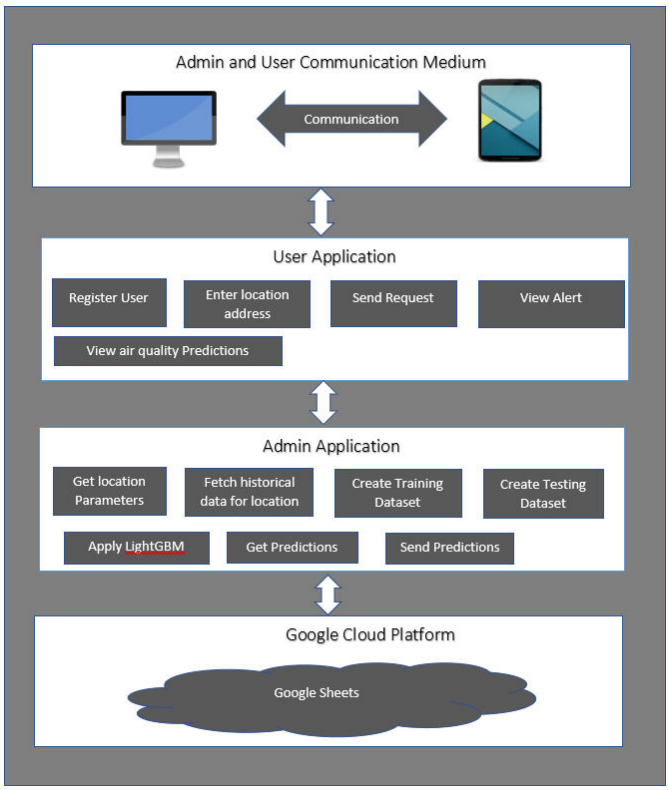

Figure 1: Block Diagram/ System Architecture

IV. SIMULATION RESULTS

In this section we are presenting two charts which mainly focuses on the time taken to train and get results from LR and SVM regression for district wise and state wise covid-19 future prediction.

The chart in Figure 2. explains how much time is taken to collect the air quality historical data, parse it and create a training dataset for LightGBM algorithm to perform. It shows that fetch air quality data takes more time than other processes.

Figure 2: Data Collection Time Chart

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005121|

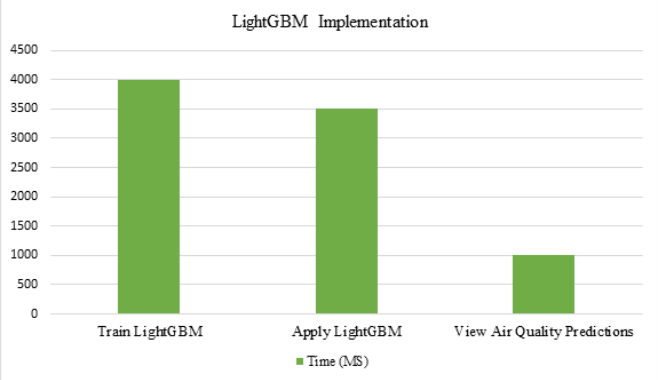

Figure 3: LightGBM Implementation Time Chart

The chart in Figure 3. explains how much time is taken to train, apply and get air quality prediction results from LightGBM algorithm. It shows that training LightGBM algorithm will take more time than other processes.

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented how we are going to design a air quality prediction system by combining various technologies such as historical air quality data, mobile computing, cloud computing and machine learning together to predict air quality of a specific place. The basic idea of this study is to help a user or commuter avoid a polluted location and keep his respiratory health in place. We studied and tried to study drawbacks of various researches [1][2][3][4][5] and decide how to improve our new air quality prediction framework. We are fetching air quality historical data of a specific address using api available on the internet. Using an api, the json data will be parsed to create a training dataset. We are using machine algorithms lightgbm regression algorithm to predict the air quality of a specific address. We are using android application as medium from where a user can send a prediction request and get response. We are using google cloud platform as our cloud provider. Thus, we conclude that by using our application a user or commuter can avoid polluted places with ease.

REFERENCES

[1] YING ZHANG, YANHAO WANG, MINGHE GAO, QUNFEI MA, JING ZHAO, RONGRONG ZHANG, QINGQING WANG and LINYAN HUANG, ‘‘A Predictive Data Feature Exploration-Based Air Quality Prediction Approach,’’ in IEEE-2019.

[2] Liying Li, Zhi Li, Lara G. Reichmann and Diane Myung-kyung Woodbridge, ‘‘A Scalable and Reliable Model for Real-time Air Quality Prediction,’’ in IEEE-2019.

[3] Nandini K and G Fathima, ‘Urban Air Quality Analysis and Prediction Using Machine Learning,’’ in IEEE-2019.

[4] Xinyu Zhang, ‘‘ION Channel Prediction Using Lightgbm Model,’’ in IEEE-2020.

[5] Anish Singh, Raja Kumar and Nitasha Hasteer, ‘‘Comparative Analysis of Classification Models for Predicting Quality of Air,’’ in IEEE-2020.

[6] J. Huang et al., ``A crowdsource-based sensing system for monitoring _ne-grained air quality in urban environments," IEEE Internet Things J., to be published.

[7] X. Li, L. Peng, Y. Hu, J. Shao, and T. Chi, ``Deep learning architecture for air quality predictions," Environ. Sci. Pollut. Res., vol. 23, no. 22, pp. 22408_22417, 2016.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details