IJIRCCE

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

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 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007071 |

Machine Learning Based Air Quality Prediction

Meghana Suryanarayana, N A Nikhitha, Rashmi Gupta, K A Neha, Prof. Nethravathy V

Dept. of Computer Science and Engineering, Bangalore Institute of Technology, Bangalore, India

ABSTRACT: In recent years, people have been paying more and more attention to air quality because it directly affects people's health and daily life. Effective air quality forecasting has emerged as a popular research topic. This initiative, however, has numerous difficulties, including the unpredictability of the data sources and the volatility in pollutant concentration over time. To address this issue, we provide an enhanced ARIMA-based air quality forecast approach.

KEYWORDS: Air Quality Index System, Arima model, Pre-processing, Feature Extraction, Prediction.

I. INTRODUCTION

Information on air quality monitoring in real time can be obtained from the devices, stations, and satellite meteorological data currently in use. Nevertheless, this is utterly important to forecast the trend of air pollution in the future and is far from sufficient. Based on this, we suggest fusing the predictive data, i.e., the forecast data on weather conditions, with the available historical data on air quality and meteorological data, supported by machine learning methods, in order to explore mining data correlation and construct a successful predictive model. The ability to anticipate occurrences of poor air quality has been extensively studied, but most of these studies are constrained by the lack of longitudinal data, making it challenging to take seasonal and other factors into consideration.

On the basis of dataset gathered by Gujarat's Environmental Protection Administration, many prediction models have been created (EPA). For predicting air quality index (AQI) level predictions, machine learning techniques such as adaptive boosting (AdaBoost), artificial neural network (ANN), random forest, stacking ensemble, and support vector machine (SVM) produce promising results. Additionally, we introduce the ARIMA model to predict the air quality index in the future.

II. RELATED WORK

This section presents the existing methods and relevant approaches which are surveyed as follows.

While working on the project we have found several applications that is related to our work. Visually different methodologies followed to predict the Air quality as well as evaluation of performance of the models proposed. Many algorithms and new approaches are introduced for the analysis and prediction process.

Yu Jiao, Zhifeng Wang1, Yang Zhang et.al[1] In view of the increasing attention paid by the state to environmental governance in recent years and the continuous deterioration of air quality, this paper proposes a prediction model of environmental quality based on Long Short-Term Memory (LSTM). This paper uses data provided by the environmental protection department to predict Air Quality Index (AQI) through temperature, PM2.5, PM10, SO2, wind direction, NO2, CO and O3. In this study, it is aimed to predict the Air Quality Index (AQI) by the Extreme Learning Machines (ELM) algorithm. For this purpose, six parameters have been selected which can affect the AQI. These are temperature, humidity, pressure, wind speed, PM10 and SO2 respectively. In this work, we propose instead, the use of deep learning methods, particularly, recurrent neural networks for predicting the gas concentrations based on the outputs of these sensors. This paper presents a first study of using Gated Recurrent Unit (GRU) neural network models for gas concentration prediction. The GRU networks achieve on average, a 44.69% and a 25.17% RMSE improvement in concentration prediction on a gas dataset when compared with Support Vector Regression (SVR) and Multilayer Perceptron (MLP) models respectively.

Burhan Baran[2]Proposes a prediction model (AQI) by using extreme learning machines. ELM is a single-layer and feed-forward artificial neural network model. Three different activation functions, sine, sigmoidal and hardlimit were used by ELM for the classification of training and test data.10-fold cross validation technique is used.

Shu Wang, Yuhuang Hu, Javier Burgues, Santiago Marco and Shih-ChiiLiu[3]This paper presents a first study of using Gated Recurrent Unit (GRU) neural network models for gas concentration prediction. The GRU networks achieve on average, a 44.69% and a 25.17% RMSE improvement in concentration prediction on a gas dataset when compared with Support Vector Regression (SVR) and Multilayer Perceptron (MLP) models respectively.

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

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007071 |

Aditya C R , Chandana R Deshmukh , Nayana D K, Praveen Gandhi Vidyavastu et.al[4]The meteorological and traffic factors, burning of fossil fuels, industrial society parameters such as power plant emissions play significant roles in air pollution. Among all the particulate matter that determine the quality of the air, Particulate matter (PM 2.5) needs more attention. We compare four simple machine learning algorithms, neural network, k-nearest neighbor, support vector machines and decision tree.

Veljanovska, K.; Dimoski, A.[5] In this paper performances of four simple machine learning algorithms i.e, neural network, k-nearest, support neighbour vector machines and decision tree were compared to identify the algorithm that is best suited for real-time implementation. It was found out that Neural Networks has the maximum accuracy of 92.3%.

Kostandina Veljanovska , Angel Dimoski et.al[6] The Air pollution and its prevention are constant scientific challenges during last decades. This paper is one scientific contribution towards this challenge.The air pollution database contains data from measurements stations in capital city of the Republic of Macedonia. The results are promising and it was proven that implementation of these algorithms could be very efficient in predicting air quality index

Jamal Toutouh1, Sergio Nesmachnow2, and Diego Gabriel Rossit et.al[7] They proposed using CGANs to train generative models to create synthesized pollution data, in this case, NO2 concentration, according to a given road data traffic density. The idea is to deal with the lack of data suffered by data-driven methods for modelling, predicting, and forecasting ambient air pollution. They have proposed three different CGANs according to the complexity of the ANNs architectures used for the generator and the discriminator CGAN: Three different neural networks architectures for the generator and the discriminator in the proposed CGAN: CGAN-1, three-layer perceptron (256 neurons per hidden-layer); CGAN-2, four-layer perceptron (128 neurons per hidden-layer); and CGAN-3, four-layer perceptron (256 neurons per hidden-layer)the ANNs.

Zhongjie Fu et al et.al[8]The paper aims at developing an email system that will help even a novice visually impaired person to utilize the services for communication without previous training. Radhika M.Patil, Dr. H. T. Dinde, Sonali. K. Powar et.al[10]Bayesian network model, an air quality prediction model based on machine learning, is used to predict the air quality in Hangzhou. Six air pollutants SO2, NO2, O3, CO, PM10 and PM2.5 are used as the evaluation factors of the model, the AQI value is used as the output result of the model.

Radhika M.Patil, Dr. H. T. Dinde, Sonali. K. Powar[9]AQI can be defined as it is a numerical value that the government agencies use to measure the levels of air pollution in the atmosphere and communicate it with population Formation of AQI: Sub-index calculation and Aggregated index calculation

III. SYSTEM ANALYSIS AND DESIGN

A. *Existing system and their drawbacks*

To develop an appropriate application, the existing air quality index prediction systems are reviewed. The drawbacks of existing systems are:

- The system that is currently used is not accurately suitable for future prediction.
- ANN will take more time to train. The network is reduced to a certain value of the error on the sample means that the training has been completed and this value does not give us optimum results .

B. *System Design*

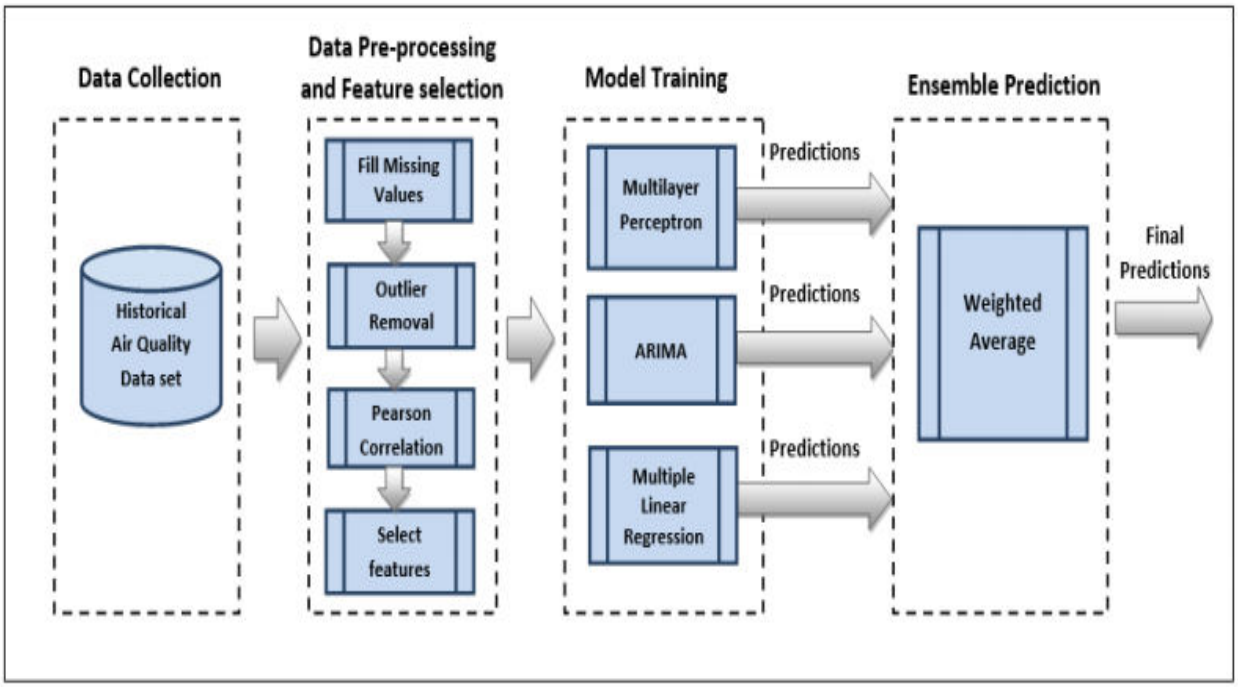
In this section, we will discuss the Architectural System design of our proposed system and Description of the various modules involved. Module and their description are listed below:

1. Air Quality Records: It is the step where the records of the air quality is obtained from various different trusted sources. Preferably from the Government sources. Here we are using Gujarat air quality dataset to perform analysis.
2. Data Pre-processing: The data collected will be cleaned as per the requirements like removal of the null values and corrupted data.

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

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007071 |



3. *Feature Extraction:*The required features and attributes among the data is separated and is extracted

4. *Building and Training the Module:*Learning Algorithm: The model is fed with the learning algorithm and is trained with the training data.

5. *Train Model*: The learning algorithm trains the model and is ready to accept the test data.

6. *Validate Model*: Test data is fed to the model and the same is validated using the data available.

7. *Model performance report*: data is fed into the model the performance report is taken.

IV. ALGORITHMS

Algorithm design refers to a method or a mathematical process for problem-solving and engineering algorithms. The design of algorithms is part of many solution theories of operation research, such as dynamic programming and divide-and-conquer.

Purpose: To predict the gas concentrations and air quality index over the period of time.

*Input:*Gas concentrations along with date, time and place where they are recorded.
(Co2,No2,PM2.5,PM10,O3,So2,Date,Place,Time)

*Output:*Prediction of Air quality Index and Gas Concentrations over a period of time.

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

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007071 |

Algorithm:

Step 1: Acquire and read the dataset.

Step 2: For each value in the dataset,

- Perform pre processing such as filling the missing values, outlier removal.
- After performing the pre processing extract the features that are mainly required to predict the air quality index.

Step 3:Divide the dataset into 2 sets i.e., train and test.

Step 4: Here We are using ARIMA forecasting algorithm to implement this project. An ARIMA model for time series analysis and prediction is a class of statistical models. ARIMA is a further extension of the Autoregressive Move Average and includes the concept of integration. Implement the model on the trained data.

AR: Auto regression. One observation is connected to several trailing observations through a dependent link in this paradigm.
Integrated: The use of raw data distinguishing for a stable time series.
Average moving: Using the moving average model's residual error dependence between an observation and the time delay.
Auto Regression (AR) Model: Auto regression is a time series model that uses previous stored data to predict the future value. It is a fairly basic concept that may lead to accurate predictions of a succession.

Step 5: Calculate accuracy and plot the graphs for all the respective gas predictions and also the air quality index

V. RESULTS

Given below are the graphs depicting the predictions of different gases and air quality index.

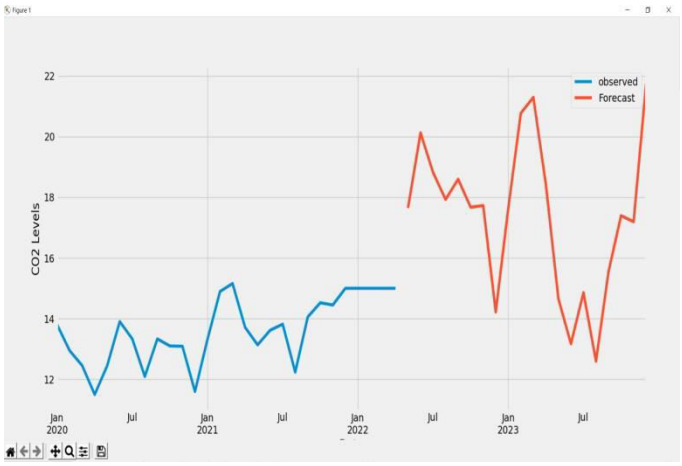


Figure 1: Snapshot of CO2 prediction levels

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

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007071 |

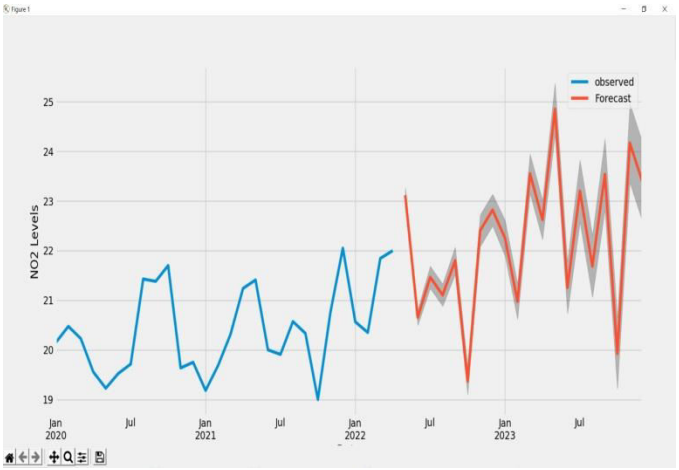


Figure 2: Snapshot of NO2 prediction levels

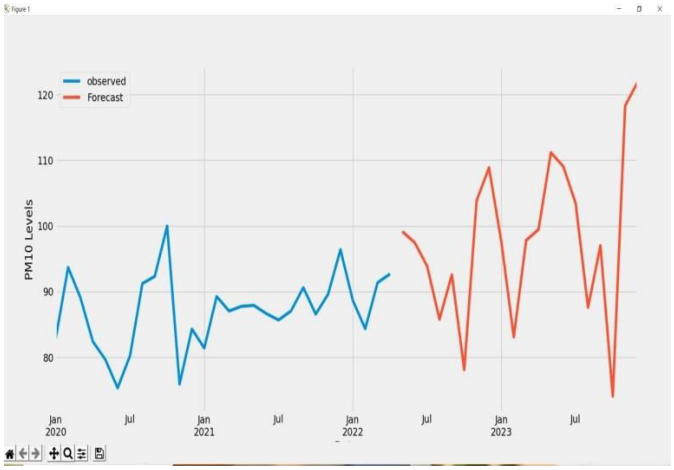


Figure 3:Snapshot of PM10 Prediction levels

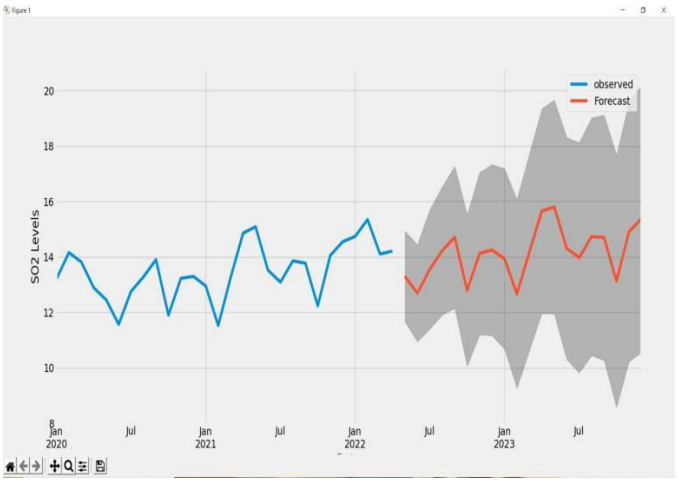


Figure 4: Snapshot Of SO2 Prediction levels

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

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007071 |

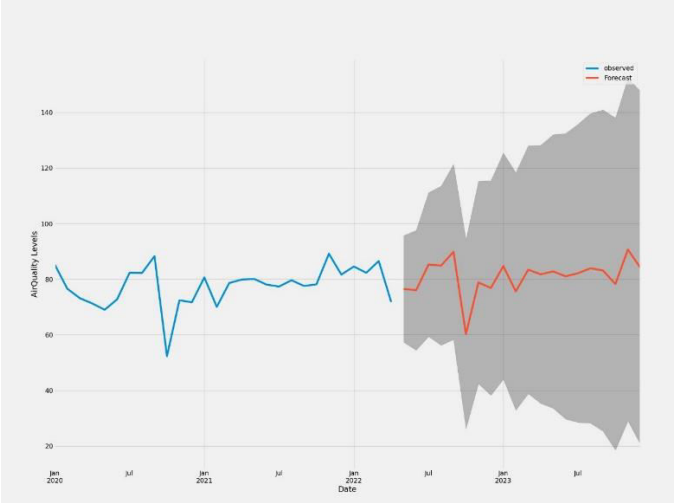


Figure 5: Snapshot Of Air Quality Index Prediction levels

VI. APPLICATIONS

- Calculate the Air Quality Index and gas concentrations of an Area.
- To Predict the Air Quality Index over a few months so that preventive measures can be taken in case of increasing pollution..

VII. CONCLUSION AND FUTURE WORK

This project is a desktop application for accurately predicting the air quality index. It offers a prediction-based tool that allows users to anticipate the air quality of the future using a set of previously gathered measurements. To stop the air quality from declining, any precautions or safety measures can be taken well in advance.
The inclusion of new parameters like temperature, pressure, and humidity can improve this job. For ease of usage, this may be made into a mobile application.

REFERENCES

[1] Yu Jiao, Zhifeng Wang1, Yang Zhang, "Prediction of Air Quality Index Based on LSTM", 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference,2019.
[2] Burhan Baran, "Prediction of Air Quality Index by Extreme Learning Machines", Directorate of Environment and Urbanization Malatya, Turkey ,2020.
[3] Shu Wang, Yuhuang Hu, Javier Burgues, Santiago Marco and Shih-Chii Liu "Prediction of Gas Concentration Using Gated Recurrent Neural Networks", 2020 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS) ,2020.
[4] Aditya C R , Chandana R Deshmukh , Nayana D K, Praveen Gandhi Vidyavastu ," Detection and Prediction of Air Pollution using Machine Learning Models", International Journal of Engineering Trends and Technology (IJETT),2018.
[5] Veljanovska, K.; Dimoski, A. , "Air Quality Index Prediction Using Simple Machine Learning Algorithms" , International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) ,2018.
[6] KostandinaVeljanovska , Angel Dimoski, "Urban Air Quality Prediction Using Regression Analysis" ,IEEE, 2019.
[7] Jamal Toutouh1, Sergio Nesmachnow2, and Diego Gabriel Rossit, "Generative adversarial networks to model air pollution under uncertainty", 2019.
[8] Zhongjie Fu et al, "Research on air quality prediction method in Hangzhou based on machine learning", Journal of Physics: Conference Series,2021.
[9] Radhika M.Patil, Dr. H. T. Dinde, Sonali. K. Powar, "Prediction of air quality index and forecasting ambient air pollutants using machine learning algorithms", Applied Sciences, 2019.

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