IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12319

Machine Learning-Based Prediction of Air Quality Index

Pratik Avhad [1], Dhananjay Kale[2], Saurabh Zanje[3], Yash Thapa[4]

Student, Department of Computer Engineering, RGCOE, Ahmednagar, India[1-4]

Abstract: In day today's World pollution is major issue in the country and also whole world. This paper presents a machine learning (ML)- based approach for forecasting air quality. Governments primarily employ air quality monitoring systems to regulate the release of harmful substances into the atmosphere. In addition to ensuring the population's general welfare and quality of life, this also helps to support the agricultural and industrial sectors. Based on the amount of PM2.5 in thegathered dataset, air quality Predicted.ML techniques such as Logistic Regression (LR), Support Vector Machine (SVM), RandomForest (RF), and Neural Network (NN) are examined. A model is built using training data, and its performance is assessed using test data. Precision, recall, F1 score, and support are utilized as performance evaluation metrics. In places with a high population density, such major cities, air pollution is a serious issue. A range of emissions caused by human activities have animpact on the quality of the air, including driving, using electricity, and burning fuel. All businesses, from early-stage startups to major platform providers, have mademachine learning and its components one of their main areas of Concentration. In the realm of machine learning, an artificial intelligence device gathers sensor data and learns how to behave.

Keywords: Air Quality Index, Machine Learning, Regression, Random Forest, Particulate Matter.

I. INTRODUCTION

Most important factor that effect of Human health is air pollution. There are some reasons are Industrialization, Urbanization,Globalalisation, Agriculturization that causes Air pollution rapidly increasing day by day. And because of air pollution human suffer from different diseases like includelung cancer, bronchitis, heart problems,throatand eye issues, asthma, and respiratory system diseases.

By predicting the air quality index, the worstconsequences on the environment and human health can be avoided.

II. LITERATURE REVIEW

Samriddhi Banara Teena Singh, examined that the three current methods for predicting air pollution in order to come up with a solution. The methods employed were linear regression, random forest regression, and convolutional neural networks. There have been calculated error rates. Also, it means that the value increases as the RMSEdecreases.

With an MSE of 936, we discovered that theRandom Forest algorithm delivered the best results for city day data. After carefully examining each of the three methods, theConvolutional Neural Network algorithm delivered the best results for city hour data with an MSE of 1834. The recommended model is suitable for visualizing air quality, according to preliminary findings [1].

RM Fernando, WMKS Ilmini. Author explained the value that qualitatively represents the state of air quality is known asthe air quality index, or AQI for short. The threat to human health and the environment increases with the air quality index. Human activity is the main cause of air pollution [2].

N. Srinivasa Gupta et. el. Author usesdifferent Machine learning methods to forecast the AQI with accuracy. The performance of the three top data miningmodels (SVR, RFR, and CR) for accurately forecasting AQI data in some of India's mostpopulous and polluted cities was tested in thecurrent study. The class data was equalised using the synthetic minority oversamplingtechnique to produce better and more dependable findings. [3].

Vidit Kumar, Sparsh Singh (AQI) an instrument used to gauge air quality is the Air Quality Index. The proposed statistical modelcan also be utilised to help deterministic air quality models, which are common and can offer helpful spatial scenarios of air quality conditions, as part of the data assimilation process. Air, but are not as good as statistical models in terms of future predictions. . This inclusion shouldimprove the predictive power of these air quality models [4].

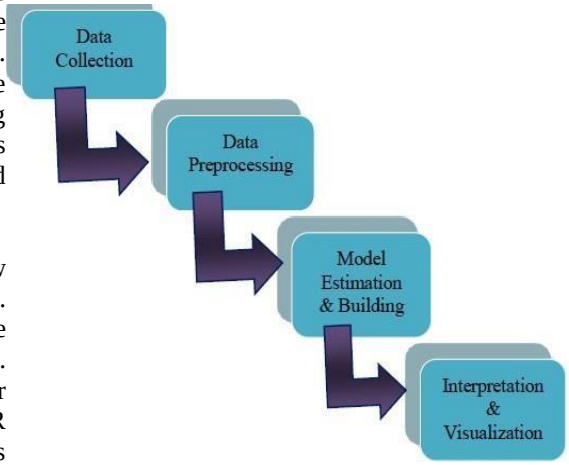
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12319

Madhuri VM, Samyama Gunjal author Discuss how climatic factors such atmospheric wind speed; wind direction, relative humidity, andtemperature affect the concentration of air pollution. To gauge air quality, one can utilise the Air Quality Index. The proposed work isa supervised learning method using differentalgorithms such as LR, SVM, DT and RF.The results show the AQI predictions obtain ed RF promising and analyzed with results [5].



Avan Chowdary Gogineni author performed a literature review and identified some machine learning algorithms to predict AQI. We identified linear regression, LASSO regression, and ridge Regression and SVR algorithms from our review of the literature. We preprocessed the data and successfully trained the linear regression, LASSO regression, ridge regression, and SVR algorithms. The same dataset was used to build each model. In this study, we considered MAE, RMSE, and r- squared error to assess model performance. Wecan conclude that the Ridge regression and LASSO regression models show better performance at lower MAE mean squared error and higher r-squared [6].

Miss Ruchita Nehete, Prof. D. D. Patil author disscuss in this project, different models Capable of predicting and classifying AQI levels in different pollution domains were tested and their performance successfully appraised. Interesting relationships between meteorological and pollutant data werediscovered using exploratory data analysis and feature engineering techniques used for thepredictive model. We gain many notableresults from our predictive models that areworth highlighting. Various approaches to managing null values produced varying resultsfor each model, but it appeared that just removing entries with null values was the best course of action. Use the classifier to predict AQI bands dynamically using values between AQI predicting PM2.5. 5 bands; the classifier appears to perform better.

Different models Capable of predicting andclassifying AQI levels in different pollution domains were tested and their performancesuccessfully appraised. Interestingrelationships between meteorological and pollutant data were discovered usingexploratory data analysis and feature engineering techniques used for the predictive model. We gain many notable results from ourpredictive models that are worth highlighting. Various approaches to managing null values produced varying results for each model, but itappeared that just removing entries with null values was the best course of action. Use the classifier to predict AQI bands dynamically using values between AQI predicting PM2.5. 5bands; the classifier appears to perform better.

Regression models helps in data analysis applications, but the conclusion is that classifier models perform better in air quality prediction [7].

III. SYSTEM ARCHITECTURE

Fig. 1 system Architecture.

1. Data collection

You are aware that machines initially pick up knowledge from the data you provide them. Your machine learning models mustbe given accurate data in order to uncoverthe proper patterns. The correctness of your model will depend on the caliber of the data you supply the computer. You willobtain inaccurate outcomes or irrelevant forecasts if your data is incomplete or out-of-date.

2. Preparing the Data

To withdraw, clean the data by removing anyunnecessary information, such as duplicatevalues, missing values, rows, and columns. You could even need to alter the row and column or row and column indexes in your dataset. Visualization of data to better understand its structure and to understand relationships between various variables and categories.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

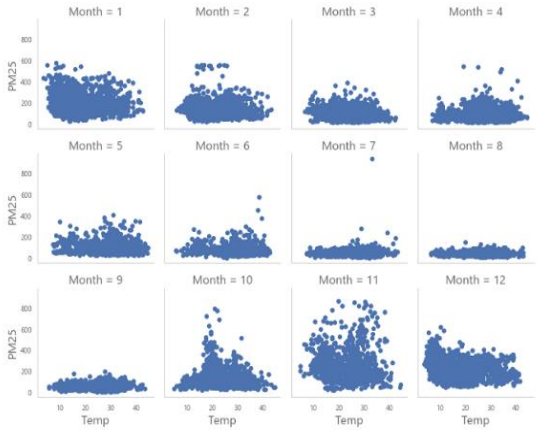
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12319

3. Model Estimation & Building:-

The results you obtain from applying a machine learning algorithm to the data you gather are determined by a machine learning model. Selecting a model appropriatefor the job at hand is crucial. The most crucial phase of machine learning is rain. To detect patterns and make predictions, you input prepared data to a machine learning modelduring training. In order to complete the set oftasks, the model must learn from the data.

4. Interpretation and Visualization:-



After training the model, check its performance. This is done by testing the performance of the model on unpublished data. The invisible data used is the test set you split our data into earlier.
How Do Machine Learning Steps Work in Python?
Begin by importing any necessary modules, as shown.

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import accuracy_score
```

Cleaning of Data:-

```
df2 = df.drop(labels=['PM10', 'NO', 'NO2', 'NOx', 'NH3', 'CO', 'SO2', 'O3','AQI','AQI_Category','Da
```

	Date	PM25	Month	Hour	Temp	Rel_Humidity	Wind_Speed
0	2020-04-30 23:00:00	59.68	4	23	28.0	70.0	2.0
1	2020-04-30 20:00:00	50.73	4	20	30.2	65.0	1.0
2	2020-04-30 17:00:00	42.53	4	17	36.6	46.0	3.0
3	2020-04-30 14:00:00	47.40	4	14	37.8	40.0	3.0
4	2020-04-30 11:00:00	63.73	4	11	34.6	50.0	3.0

1) Handle Missing Values:

```
df.isnull().values.any()
```

We can see this when we visualize. For example, in the following visualization of PM 2.5 levels as a function of temperature for each month:
import datetime as dt
df['Date'] = pd.to_datetime(df['Date'])

ML Models

1. Linear Regression
Linear regression is the most fundamental type of regression analysis. It is assumed that the dependent variable and the predictor have a linear relationship (s). We try to calculate the best fit line in regression to describe the relationship between predictors and predictive/dependent variables.

2. Decision trees
Decision trees are a type of predictive modeling that aids in mapping the various decisions or solutions to a given outcome. Different nodes make up a decision tree. The root node is the beginning of the decision tree, which is typically the entire dataset in machine learning. The endpoint of a branch, or the final result of a series of decisions, is represented by a leaf node.

3. Support Vector Machine
SVM is a simple yet powerful Supervised Machine Learning algorithm that can be used to build both regression and classification models. The SVM algorithm works well with both linearly and non-linearly separable datasets.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023
DOI: 10.17148/IJARCCE.2023.12319

IV. CONCLUSION

In this study, we discuss three key issues in the field of urban air computing: interpolation, estimation, and analysis of air qualityparameters. The answer to these questions offers a solution that forecasts air quality and the negative effects of air pollution on human health. These three issues are addressed in different models by the bulk of the currentsystem.

REFERENCES

[1]. Samriddhi Banara Teena Singh "Air Pollution Forecasting using Machine Learning and Deep Learning Techniques" Volume 18 Issues 5
[2]. RM Fernando#, WMKS Ilmini, and DU Vidanagama "Air Quality Prediction Using Machine Learning" #35-CS-18- 0001@kdu.ac.lk ID 379.
[3] N. Srinivasa Gupta, Yashvi Mohta, Khyati Heda, "Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis" ID 4916267,doi.org/10.1155/2023/4916267
[4] Vidit Kumar, Sparsh Singh, Zaid Ahmed "Air Pollution Prediction using Machine Learning Algorithms: A Systematic Review" ISSN: 2278- 0181Vol. 11 Issues 12, December 2022
[5].Madhuri VM, Samyama Gunjal GH, "Air quality prediction using Machine learning supervised Learning approach" ISSN 2277-8616VOLUME 9, ISSUE 04, APRIL 2020
[6] Avan Chowdary Gogineni"Prediction of Air Quality Index Using Supervised Machine Learning" May 2022, 371 79
[7] Miss Ruchita Nehete, Prof. D. D. Patil "Air Quality Prediction Using Machine Learning" ISSN: 2320-2882 Volume 9, Issue 6 June 2021.