FORECASTING OF AIR POLLUTION USING RANDOM FOREST ALGORITHM

S.Sandhiya[1],E.Thenmozhi[2], Harsha Vardini G[3], Kirthana P[4], Niranjana SR[5]

[1]Assistant Professor, Department of Information Technology,

Panimalar Engineering College

[2]Professor, Department of Information Technology,

Panimalar Institute of Technology ,Chennai ,India.

[3,4,5]Student, Department of Information Technology

Panimalar Institute of Technology, Chennai, India.

ABSTRACT

Due to mortal conditioning, industrialization, and urbanization, air pollution has come a life- hanging factor in numerous countries around the world. Among air adulterants, Particulate Matter causes colorful ails similar as respiratory tract and cardiovascular conditions. Hence, it's necessary to directly prognosticate the air pollution attention to help the citizens from the dangerous impact of air pollution beforehand. The variation of air pollution depends on a variety of factors, similar as meteorology and the attention of other adulterants in civic areas. The end is to probe machine literacy- grounded ways for the vaticination of air pollution results with the stylish delicacy. The analysis of the dataset by supervised machine learning fashion(SMLT) captures several information like variable identification and analysis ways like univariate analysis,bi-variate, and multivariate analysis. Compare and bandy the performance of colorful machine learning algorithms from the given pollution dataset with evaluation ways.

Keywords: SMLT, Machine Learning Algorithms, Illnesses, Urban areas, Dangerous impacts.

1. INTRODUCTION

Air pollution refers to any physical, chemical or biological changes in the air. It is the contamination of air by harmful gases, dust and smoke which affects plants, animals and humans drastically. It is considered as the largest environmental threat, leading to 7 million deaths per year and other respiratory problems such as asthma and lung cancer that the report stated. Bangladesh was the most polluted country.While in India,New Delhi is the world's most polluted capital city. The countries with the worst air pollution were India, Pakistan and Bangladesh, all exceeding WHO guidelines by 10 times. Denmark, Norway, Sweden, Australia, Canada, Japan and the United Kingdom ranked among the best countries for air quality, only exceeding the guidelines by 1 to 2 times. Accurate forecasting helps people plan ahead, To aware of variations in the quality of the air they breathe.

2.OVERVIEW OF THE PROJECT

2.1 DATA PREPROCESSING:

Data preprocessing is a process of preparing the raw data and making it suitable for a machine literacy model. It's the first and pivotal step while creating machine literacy model. When creating a machine literacy design, it isn't always a case that we come across the clean and formatted data. And while doing any operation with data, it's obligatory to clean it and put in a formatted way. So for this, we use data preprocessing task.

2.2 MACHINE LEARNING:

Machine Learning provides introductory and advanced generalities of machine literacy. Our machine learning is designed for scholars and working professionals.

Machine literacy is a growing technology which enables computers to learn automatically from once data. Machine literacy uses colorful algorithms for erecting fine models and making prognostications using literal data or information.

A Machine literacy system learns from literal data, builds the vatication models, and whenever it receives new data, predicts the affair for it. The delicacy of prognosticated affair depends upon the quantum of data, as the huge quantum of data helps to make a better model which predicts the affair more directly.

Suppose we've a complex problem, where we need to perform some prognostications, so rather of writing a law for it, we just need to feed the data to general algorithms, and with the help of these algorithms, machine builds the sense as per the data and prognosticate the affair. Machine literacy has changed our way of allowing about the problem.

3.TECHNICAL REQUIREMENT

3.1 Hardware Requirements:
   a) Processor : Intel i3 or later
   b) Hard disk:Minimum 10 GB
   c) RAM : Minimum 4 GB
3.2 Software Requirements:
   a) Operating System : Windows 10 or later
   b) Tool : Anaconda with Jupyter Notebook

4. EXISTING SYSTEM

In existing system, The application was done with a novel outlier detection mechanism named Volterra graph-based outlier detection (VGOD) based on graph signal processing. The application of air prediction in existing is not user friendly for real time inputs. The accuracy of air pollution is not predicted ,so it is not mentioned. Minimum number of algorithms are used to find the outlier detection. In existing ,Data visualization is not done.

5.PROPOSED SYSTEM

The data analysis is done on the dataset, proper variable identification is done tha is both the dependent variables and independent variables are found. In Data preprocessing, errors like None and null value, Duplicate values ,Missing values, Unwanted values and Unbalanced data are rectified with the keywords DROPNOW, CLEAR, REMOVE, DISCARD, DELETE, etc. To understand visually ,Data visualization is done with graph. Histogram, piechart and plot. The proper machine learning algorithms are applied to the dataset where the pattern of data is learned. After applying different algorithms a better algorithm is used for the prediction of the outcome. The 11 types of inputs are given to forecast the output accurately like N,NO2, NOX, CO,Benzene,etc. Real time input and output are deployed in a web browser.
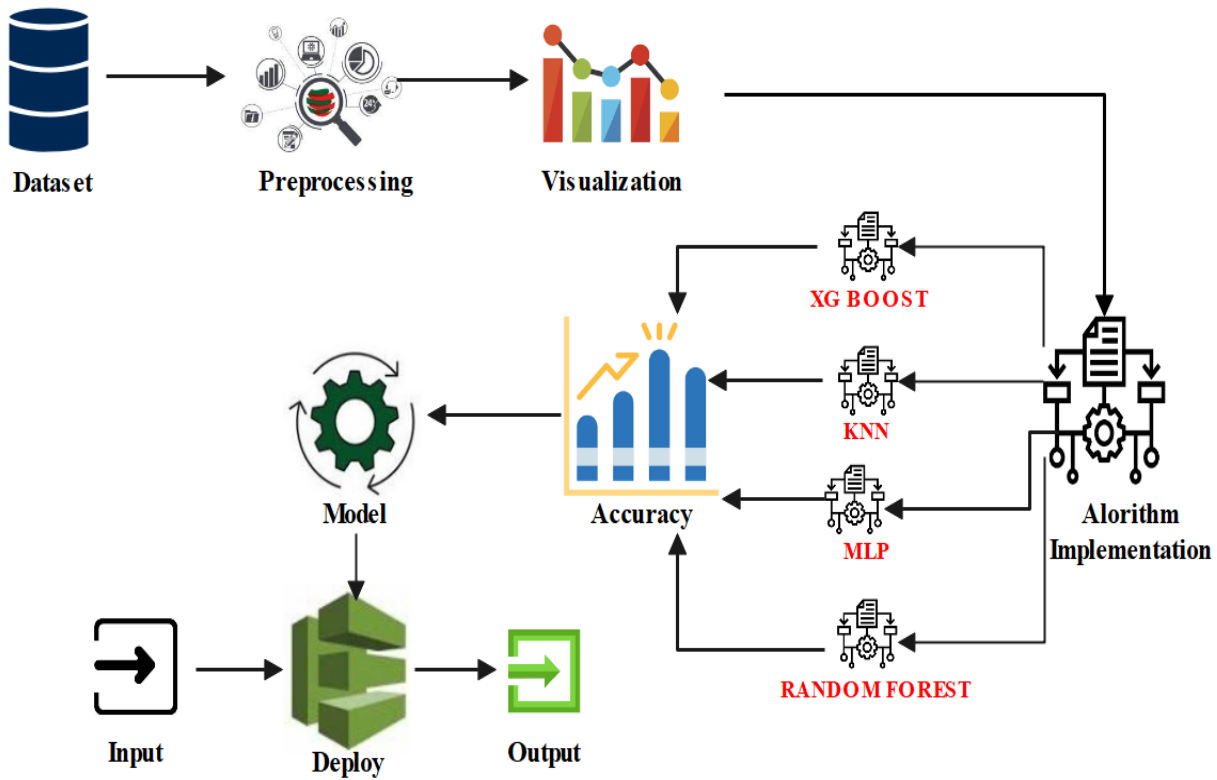
6.ARCHITECTURE DIAGRAM

Figure 1.Architecture Diagram of Air Pollution

Algorithms Used for Implementation

XG BOOST-Extreme Gradient Boosting

KNN-K-Nearest Neighbors Algorithm

MLP-Multi-Layer Perceptron

Random Forest Algorithm

7. PROGRAM OUTCOME

7.1 MODULE 1 -DATASET:

A machine learning dataset is a collection of data that is used to train the model.Dataset act as an example to teach the machine learning algorithms how to make predictions.

7.2 MODULE 2 -PREPROCESSING:

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model.A real world data generally contains null values and missing values which can't be directly used for machine learning models.It also increases the accuracy and efficiency of the machine learning model.

7.3 MODULE 3- ALGORITHM IMPLEMENTATION:

Implementing a machine learning algorithm will give a deep and practical appreciation for how the algorithm works.There are numerous micro decisions required when implementing an algorithm.

8.WORKING OF THE PROJECT:

In our application , we are giving a facility of forecasting air pollution with accuracy and we have live input fields so that any user from anywhere with specified software can access and predict the results.
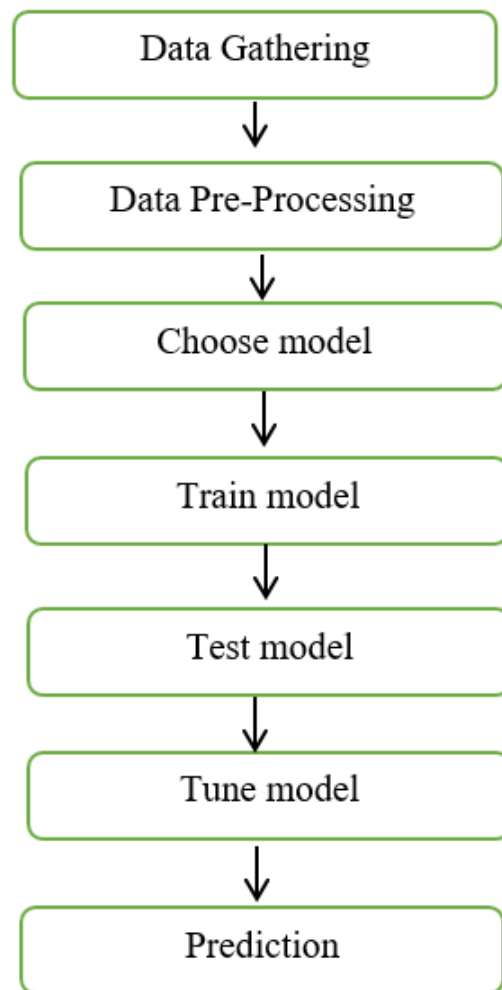
9. FLOW CHART:

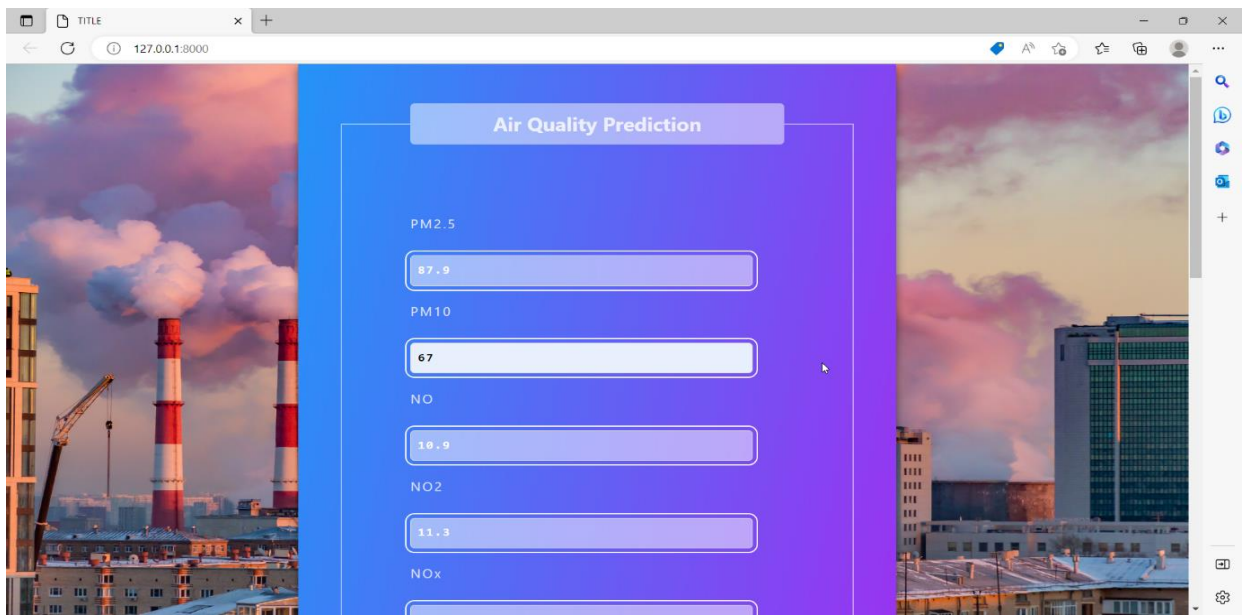
Figure 2:Flowchart

10. OUTPUT:

Figure 3

Figure4

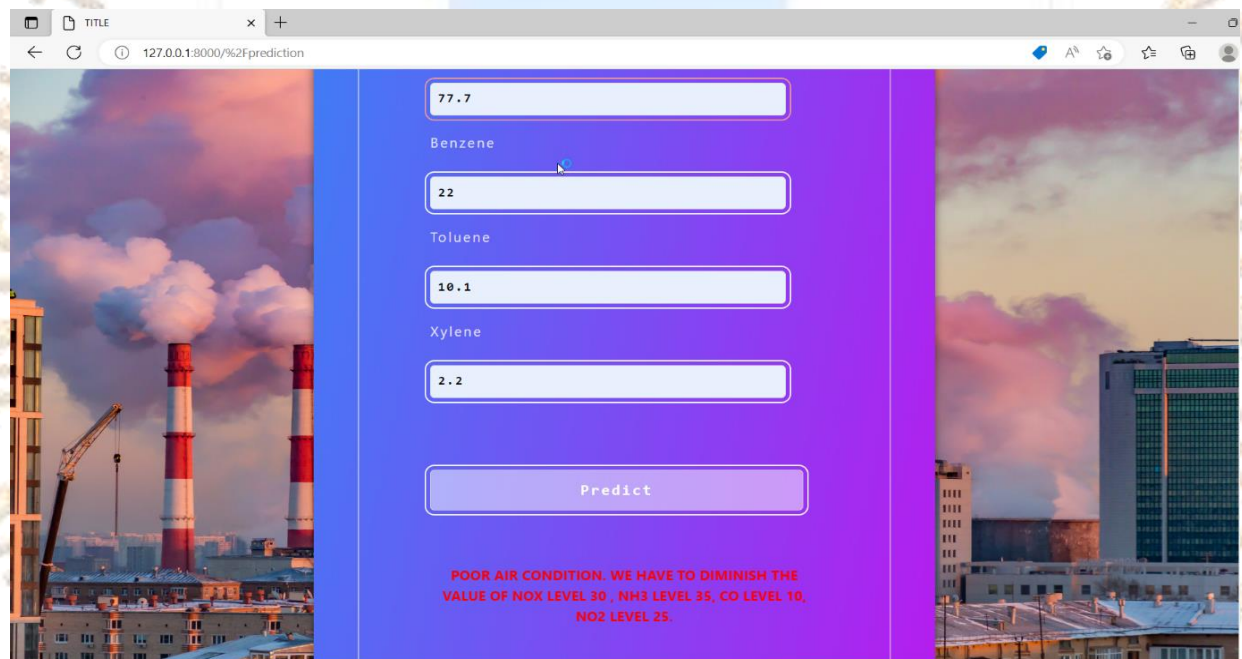## 11. CONCLUSION

Our conclusion is that the analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set of higher accuracy score algorithm will be find out. The founded one is used in the application which can help to find the Air Pollution .

12. REFERENCES

1. Aditya, C. R., Chandana R. Deshmukh, D. K. Nayana, and Praveen Gandhi Vidyavastu. "Detection and prediction of air pollution using machine learning models." *International Journal of Engineering Trends and Technology (IJETT)* 59, no. 4 (2018): 204-207.

2. Xayasouk, Thanongsak, and Hwamin Lee. "Air pollution prediction system using deep learning." *WIT Trans. Ecol. Environ* 230 (2018): 71-79.

3. Liang, Yun-Chia, Yona Maimury, Angela Hsiang-Ling Chen, and Josue Rodolfo Cuevas Juarez. "Machine learning-based prediction of air quality." *applied sciences* 10, no. 24 (2020): 9151.

4. Harishkumar, K. S., K. M. Yogesh, and Ibrahim Gad. "Forecasting air pollution particulate matter (PM2. 5) using machine learning regression models." *Procedia Computer Science* 171 (2020): 2057-2066.

5. Bekkar, Abdellatif, Badr Hssina, Samira Douzi, and Khadija Douzi. "Air-pollution prediction in smart city, deep learning approach." *Journal of big Data* 8, no. 1 (2021): 1-21.

6. Environmental pollution: An enormous and invisible burden on health systems in low- and middle-income countries. World Hosp Health Serv 2014; 50 (4): 35-40.

7. Ambient particulate pollution and the world-wide prevalence of asthma, rhinoconjunctivitis and eczema in children: Phase One of the International Study of Asthma and Allergies in Childhood (ISAAC). Occup Environ Med 2010; 67 (5): 293-300.

8. Air quality influences the prevalence of hay fever and sinusitis. Laryngoscope. 2009; 119 (3): 429-433.

9. Air pollution and development of asthma, allergy and infections in a birth cohort. The European Respiratory Journal 2007; 29: 879-888.

10. Air pollution and doctors' house calls for respiratory diseases in the Greater Paris area (2000-3). Occupational and Environmental Medicine 2007; 64: 320-324.