AIR POLLUTION PREDICTION USING RANDOM FOREST AND DECISION TREE ALGORITHMS

Dr.PRASADA RAO M[1] ,VINNELA.A[2] ,SAI GANESH.B [3] ,

PRIYANKA.B[4] ,SIVATEJA.P[5]

Assoc Prof,Mtech ,Ph.D,Dept of CSE, Wellfare Institute of Science Technology & Management,Visakhapatnam
[1](319129510001), CSE Student, Wellfare Institue of Science Technology & Management
[1](319129510008), CSE Student, Wellfare Institue of Science Technology & Management
[1](319129510009), CSE Student, Wellfare Institue of Science Technology & Management
[1](319129510032), CSE Student, Wellfare Institue of Science Technology & Management

ABSTRACT

Due to human activities, industrialization and urbanization air is getting polluted. The major air pollutants involved are Carbonmonoxide(co), Nitric oxide(NO),Nitrogen dioxide(N02) ,Ozone(O3).The concentration of air pollutants in ambient air is governed by the meteorological parameters such as atmospheric wind speed, wind direction, relative humidity, and temperature. In Earlier techniques such as Probability, Statistics etc. were used to predict the pollutants involved in air, but those methods are very complex to predict .So we are and going to use Data minin g techniques for the better approach to predict the airpollution and quality of air which are need to predict air, relative humidity by considering various parameters. such as CO,NO2, NO, Temperature etc. So,to predict the pollution of air whether its increasin g year by year or decreasing.We are going to use Decision tree and Random forest algorthims in Data mining techniques.

Keywords: Air pollution prediction, Data mining, Smart city, Time series, Random Forest Algorithm, Complexity, Effectiveness, Practicable.

INTRODUCTION

Air is a basic requirement for the survivaland development of all lives on Earth. It affects health and influences the development of the economy. Today, due to the development of industrialization, the increase in the number of private cars, and the burning of fossil fuels, air quality is decreasing, with increasingly serious air pollution. There are many pollutants in the atmosphere, such as SO2, NO2, CO2, NO, CO.internationally, a large number of scholars have conducted research on air pollution and air quality forecasts, concentrating on the forecasting of contaminants. Air pollution affects the life of a society, and even endangers the survival of mankind. During the Industrial Revolution, there was a dramatic increase in coal use by factories and households, and the smog caused significant morbidity and mortality, particularly when combined with stagnant atmospheric conditions. Humans are very sensitive to humidity, as the skin relies on the air to get rid of moisture. The process of sweating is your body's attempt to keep cool andmaintain its current temperature.
relies on the air to get rid of moisture.
The process ofsweating is your body's

attempt to keep cool and maintain its current temperature.

If the air is at 100-percent relative humidity, sweat will not evaporate into the air. As a result, we feel much hotter than the actual temperature when the relative humidity is high. If the relative humidity is low, we can feel much cooler than the actual temperature because our sweat evaporates easily, cooling us off.

Existing System

In Earlier techniques such as Probability, Statistics etc. were used to predict the pollutants involved in air, but those methods are very complex to predict .sometimes the data is manuplated easily and data is not shown accurately .So we are going to use Data mining techniques and Decision tree , Random forest algorthims for the better approach to predict the air pollution .

Proposed System:-

The air pollution prediction using Data mining techniques considers two methods learning algorithms such as decision tree and Random forest algorithms. The Data mining techniques for the better approach to predict the air pollution and quality of air which the need to predict air, relative humidity . considering various parameters. such as CO,NO2, NO, Temperature etc. So,to predict the pollution of air whether its increasing year by year or decreasing.The air pollution prediction using Data mining techniques considers two methods learning algorithms such as decision tree and Random forest algorithms. The Data mining techniques for the better approach to predict the airpollution and quality of air which the need to predict air, relative humidity .considering various parameters. such as CO,NO2, NO, Temperature etc. So,to predict the pollution of air whether its increasing year by year or decreasing.

METHODOLOGY

- •DECESSION TREE ALGORITHRANDOM

RANDOM FOREST ALGORITHM

DECISION TREE:

Decision Tree is one of the supervised learning algorithms which it is used to represent the decision that is made based on the condition. It is used for both classification and regression. The Decision tree is always constructed from top to bottom. The first node from the top is called as root node. The last nodes is called as a leaf node. Internal nodes are present in between the root node and leaf nodes. Based on some condition the internal nodes are split and finally, the decisions are made. In the real time as the number of variables increases tree grows larger and algorithm becomes complex. In Decision tree we have two types, they are classification and regression trees. Classification tree is used to classify the dataset. so that it is easy to analyze the data. But using this algorithm we cannot make a prediction. The Regression tree is a tree mainly used to predict continuous values.Growth of tree depends on factors like the attribute which is chosen to make a prediction,condition used for the split the tree; deciding when to stop or terminate the growth of the tree.
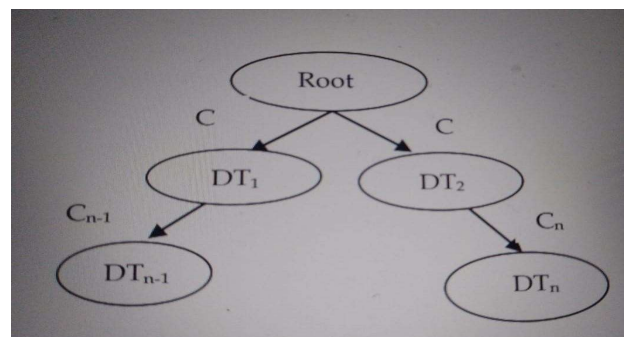

RANDOM FOREST ALGORITHM:-

It is defined as a set of decision trees to do regression and classification. Classification is

used to find out the majority voting. Regression is used to calculate the mean value. This algorithm is more accurate, robust, and can handle a variety of data such as binary data, categorical data, and continuous data. Random Forest is nothing but multiple decision trees. 75% of the dataset is considered for the training. The training data is subjected to sampling and based on attribute sampling different decision trees are constructed by applying the Random Forest.
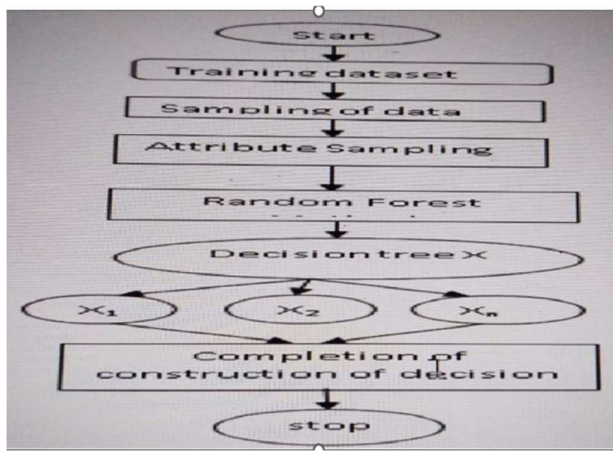
OUTPUT SCREENS:-
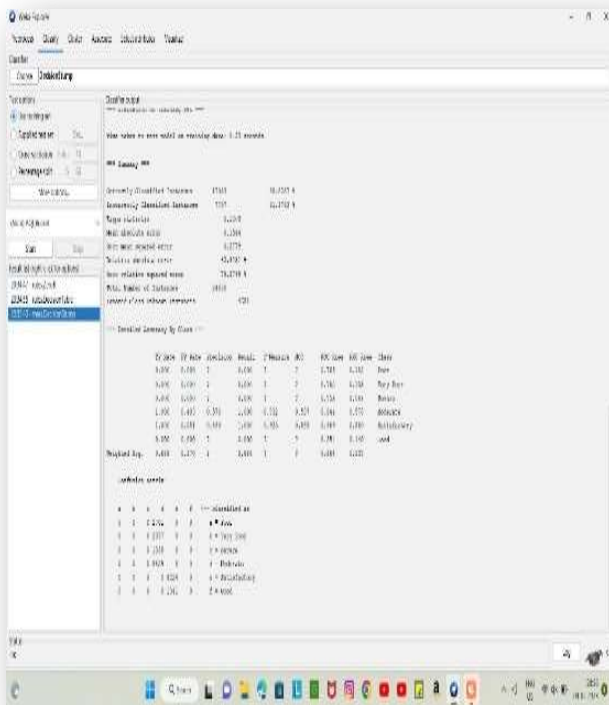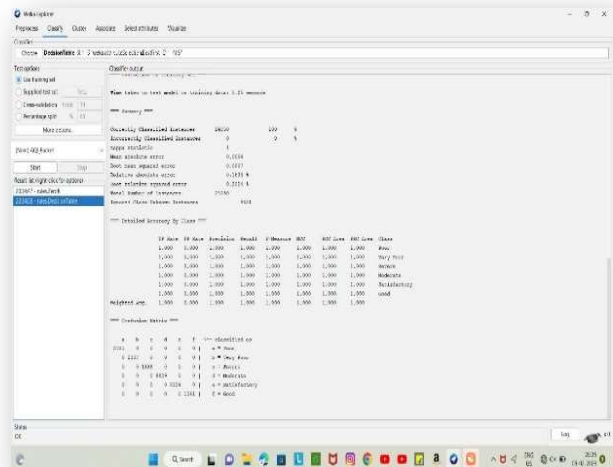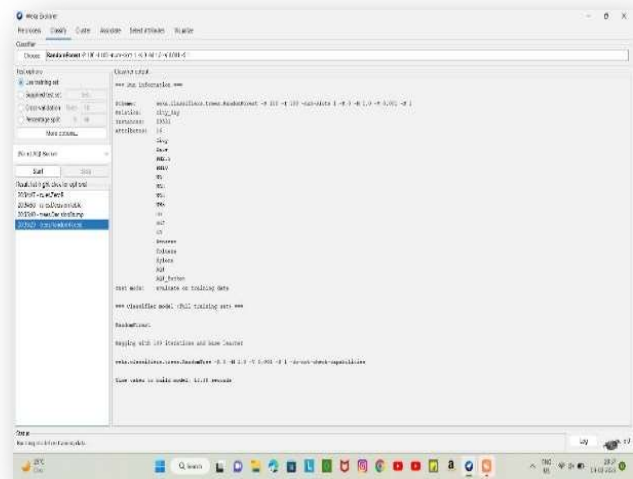
LITETURE SURVEY

In this the author Pasupuleti et al. (2020) the decision tree, linear regression, random forest. Major air pollutants are taken and meterological condition are taken using Arduino Platform. Random forest gives better accurate result due to overfitting that reduces errors But drawback is Random forest uses more memory and high cost.

Haotian Jing & Yingchun Wang(2020) had predicted the air quality index using XG Boost.It uses the weak classifier and shortcoming of the previous weak classifier to form a strong classifier thus reducing the error between predicted and actual values .It uses the K- cross validation .The mean absolute error and coefficient of determination is determined to predict the difference between actual and predicted value.The drawback faced is that it takes the previous value and is affected by outlier unwanted pollutant in the air.

Desislava Ivanova and Angel Elenkov (2019) had used Rasberry Pi platform with Multilayer perceptron algorithm of machine learning to predict the air pollutant accurately.Multilayer perceptron overcome problem of classification which is used for discrete values and regression which is used for continuous value. In this author uses discrete values and had used multilayer perceptron using backpropagation and therefore input did not pass to the activation function resulting in 0 or 1indicating how big the difference between the predicted and actual value. The coefficient of determination($R2$) obtained is better but more can be improved by increment feeding. In Elseviere 2018,The tittle is Forecasting air pollution load in Delhi using data analysis tools.The methods used are Time series regression.With limitations are Here we are using time series with regression.so again regression is limited to the linear relationship.It is easily affected by outliers.

Soubhik et al. (2018) had compared various algorithms like Linear regression, Neural network regression, Lasso regression, ElasticNet regression, Decision Forest, Extra trees, Boosted decision tree, XGBoost, KNN, and Ridge regression to predict air pollutant level.Better accuracy is obtained by extra trees because features are arranged in decreasing order of importance to predict the upcoming value

CONCLUSION:-

The concentration of air pollutants in ambient air is governed by the meteorological parameters such as atmospheric wind speed, wind direction,relative humidity,and temperature. Air pollution prediction, is used to measure the predict the quality of the air. So,it predict the polluted air whether its increasing year by year or decreasing.We are going to use Decision tree and Random forest algorthims in Data mining techniques.To give the acuaracy in data and amount of pollution increasing year by year or decreasing year by year with the analysis made from years.

REFERENCES:-

[1]. Kostandina Veljanovska1 & Angel Dimoski2, Air Quality Index Prediction Using Simple Machine Learning Algorithms,2018, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS).

[2]. Xiaosong Zhao , Rui Zhang, Jheng-Long Wu, Pei-Chann Chang and Yuan Ze University, A Deep Recurrent Neural Network for Air Quality Classification, 2018, Journal of Information Hiding and Multimedia Signal Processing .

[3]. Savita Vivek Mohurle, Dr. Richa Purohit and Manisha Patil, A study of fuzzy clustering concept for measuring air pollution index,2018, International Journal of Advanced Science and Research

[4]. Aditya C R , Chandana R Deshmukh , Nayana D K and Praveen Gandhi Vidyavastu , Detection and Prediction of Air Pollution using Machine Learning Models,2018, International Journal of Engineering Trends and Technology (IJETT) .

[5]. Shan Zhang , Xiaoli Li &Yang Li , Jianxiang Mei, Prediction of Urban PM2.5 Concentration Based on Wavelet Neural Network,2018,IEEE.

[6].Timothy M. Amado & Jennifer C. Dela Cruz, Development of Machine Learning-based Predictive Models for Air Quality Monitoring and Characterization, 2018, IEEE .

[7]. Arwa Shawabkeh , Feda Al-Beqain , Ali Rodan, Maher Salem, Benzene Air Pollution Monitoring Model using ANN and SVM,2018,IEEE .

[8]. Gaganjot Kaur Kang, Jerry Zeyu Gao, Sen Chiao, Shengqiang Lu, and Gang Xie, Air Quality Prediction: Big Data and Machine Learning Approaches,2018, International Journal of Environmental Science and Development .

[9]. Nicolás Mejía Martínez, Laura Melissa Montes, Ivan Mura and Juan Felipe Franco, Machine Learning Techniques for PM10 Levels Forecast in Bogotá,2018,IEEE.

[10]. Lidia Contreras Ochando, Cristina I. Font Julian, Francisco Contreras Ochando, Cesar Ferri,Airvlc: An application for real-time forecasting urban air pollution,2015, Proceedings of the 2 nd International Workshop on Mining Urban Data, Lille, France.