Air Quality Expectation utilizing AI Calculations

[1] D.Wasiha Tasneem, [2]M.Lakshmi, [3]C.Ishaq Shareef
[1,2,3]Assistant Professor
[1,2,3] Department of Computer Science & Engineering,
[1,2,3]Ashoka Women's Engineering College

Abstract: In many industrial and urban regions, the government has made air quality monitoring and protection a top priority. Weather and traffic patterns, the use of fossil fuels, and other industrial processes all contribute to air pollution. It is time to adopt models that will help us track the concentrations of air contaminants, as pollution continues to rise (so2,no2,etc). Urban areas, in particular, are suffering from the effects of this toxic gas accumulation in the atmosphere. Since environmental sensing networks and sensor data are readily accessible, a growing number of academics are turning to Big Data Analytics. So2 concentrations in the environment can be predicted using machine learning approaches in this study. – Oxygen-containing sulfate aerosols may cause irritation to the skin, eyes, nose, throat, and lungs. So2 levels in the next years or months may be predicted using models based on historical data.

1. INTRODUCTION

There has been an upsurge in environmental issues in emerging nations such as India as a result of fast population growth and an economic boom in cities. Air pollution has a direct effect on people's health. There has been an increase in public awareness of the same in our country. Global warming, acid rain, and an increase in the number of asthma patients are all long-term repercussions of air pollution. Using accurate air quality forecasts may help reduce the effect of pollution on both humans and the environment.One of the most important goals for society, therefore, is to improve air quality forecasts. Sulphur dioxide is a gas. It is a significant source of air pollution. In addition to being colorless, it also has an unpleasant, harsh odor. Chemicals like sulfuric acid, sulfurous acid, and sulfuric acid may be readily synthesized from it. Someone is health might be adversely affected by inhaling sulfur dioxide. Nasal, throat, and airway irritation may cause coughing, wheezing, shortness of breath, or tightness in the chest. A rise in sulphur dioxide levels in the atmosphere might alter a habitat's suitability for plant and animal species.[6] The suggested approach is able to anticipate future Sulphur Dioxide concentrations [6].

RELATEDWORK

Student researchers used machine learning algorithms to estimate air quality indexes (AQI) in a specific location in this study. It is common practice to use the Air Quality Index (AQI) to gauge how clean or polluted the air really is. For example, the agencies[8] keep track of the concentration of gases such as asso2 or no2, or even CO2. According to this model, the air quality index may be predicted using data from past years and a specific year in the future as Gradient Decay Boosted Multivariable Regression. They increased the model's efficiency by using cost estimation for predicting problems. They claim that their model can accurately estimate the air quality index for a whole county, a state, or a restricted area based on pollutant concentration data from the past.

Artificial Neural Networks and Kriging are utilized in this article to predict air pollution levels in numerous places in Mumbai and Navi Mumbai based on historical data from the meteorological department and the Pollution Control Board. The proposed model is tested and implemented using MATLAB for ANN and R for Kriging, and the results are given.

A multilayer perceptron (ANN) protocol and linear regression are employed in this system to forecast the following day's pollution. The system aids in the prediction of the next date for pollution details based on fundamental criteria and the analysis of pollution details and the forecasting of future pollution. Using Time Series Analysis, future data points were identified and air pollution was predicted[3].

The suggested system accomplishes two critical functions. I PM2.5 concentrations may be determined using atmospheric parameters. The PM2.5 level for a certain day may be predicted using this method. To determine whether a sample of data has been contaminated or not, logistic regression is used. Predicting future PM2.5 levels based on historical data is a common use of autoregression. This study's major objective is to forecast City's air pollution level using field data[9].

2. DATASET

Dataset/Source:Kaggle

Structured/Unstructureddata: Structured DatainCSVformat.

1)stn_code2)sampling_date
3) state
4) location
5) agency
6)type7)so28)no29)rspm
10) spm
11)location_monitoring_station12)pm2_5
13)date

Dataset Description:

Around 450000 entries are included in the collection, which covers all of India's states. Only Maharashtra's Dataset was used. As a result, we have 60383 entries in our database. The following 13 characteristics are included in this dataset.

The sample date is the date on which the data was recorded, whereas the station code identifies which station collected the data. The name of the agency that recorded the data may be found in the state and location fields. "Type" denotes where data was collected, such as "industrial," "residential," "commercial," etc." Measured concentrations of sulfide, nitrogen, respirable suspended particulate matter, and suspended particulate matter are represented by the abbreviations so2, no2, rspm, and spm. sampling date has been replaced with date. Airborne particulate matter (PM) with a diameter smaller than or equal to 2.5 micrometers (approximately 3% of the diameter of a human hair) is referred to as PM2.5[4]. However, the vast majority of the values in this column are empty.

SplittingforTesting:80% of the data was used for training, while 20% was used to test.

PreprocessingandFeatureSelection:

Only Maharashtra State's data was

researched and put to algorithms. Due to these reductions (to 60,383) the state column is no longer relevant[2].

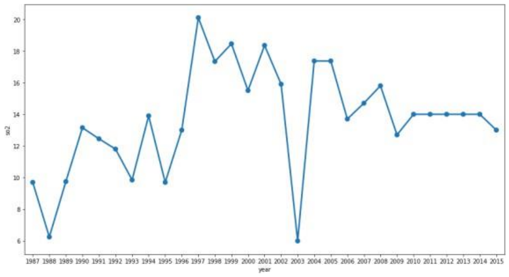
We removed pm2 5 since it contained only null values.

To put it another way, the agency's name has nothing to do with how bad the state is in terms of pollution. stn code, in a similar vein, is of little use.

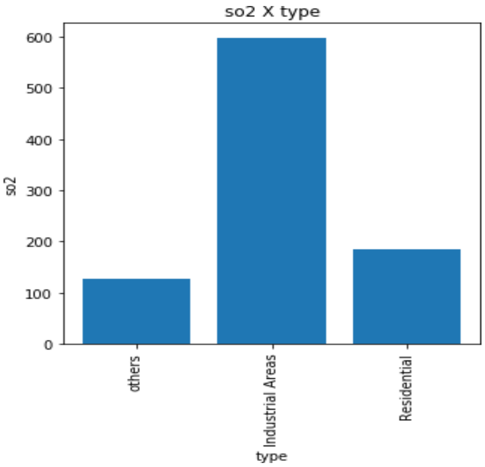
Because the date is a better representation of sampling date, the latter will be removed to save space. We do not require the location of the monitoring station in our study, thus the location monitoring station property is also superfluous.

3. EXPLORATORYDATAANALYSIS:

Below is a chart depicting changes in SO2 concentration over time. 1997 and 2001 were the greatest and lowest years respectively. However, over the last several years, it has been steady.



- This graph demonstrates that industrial locations have the greatest concentrations of so2.



- In this graph, we can deduce that Nagpur is most polluted, with Akole, Amravati, and Jalna coming in a close second and third, respectively.[11].

4. RESULTANDDISCUSSION:

We areabletoidentify the future datapoints using TimeSeriesAnalysis.

Modelsused forthesameare:

1) ARmodel:(autoregressivemodel)

TestMSE:166.358

Regression equations are used to estimate the future time step based on data from prior time steps in an autoregressive model.

You can get accurate predictions on a wide variety of time series issues using a simple principle.

yhat=b0 +b1*X1

An input value X is substituted for the training data to arrive at b0 and b1 coefficients, which are then used to calculate yhat.

When applied to a time series, this approach may be used to get input variables from observations made at prior time steps (referred to as lag variables).

Based on observations from the previous two time steps, for example, we may estimate the value of the following time step (t+1) (t-1 and t-2). This might look like this in a regression model:

X(t+1)=b0+b1*X(t-1) +b2*X(t-2)

Because the regression model uses data from the same inputvariableatprevioustimesteps,itisreferredtoasanautoregression (regressionofself)[10].

2) ARIMAMODEL:

The ARIMA family of statistical models may make time series data analysis and prediction simpler.

ARIMA is a more complex version of the AutoRegressive Moving Average (ARMA), which incorporates the integration principle.

In statistics, this is known as autoregression (AR). A model that takes into account the link between an observation and a number of lagged observations.

I am integrated. By subtracting an observation from an observation from a prior time step and applying differencing to the resulting data, we may stabilize our time series.

MA stands for the Moving Average. As the name suggests, this model makes advantage of the relationship between an observation and a moving average residual error

Each of these elements is stated clearly in the model as a scalar parameter.

ARIMA(p,d,q) is a standard notation for rapidly identifying which ARIMA model is being used, with the parameters replaced by integer values.

This model's parameters are outlined in the following order: p: a measure of the model's lag order; how many lag observations are included in the model?

d:The degree of differencing refers to how many times the raw observations have been differed.

q:There are two terms used to describe how big a moving average window is: "order" and "size."[7]

5. CONCLUSION

We might infer from the produced bar graphs that some cities are particularly dirty and in need of immediate action. We may start preparing now for places like Pune and Mumbai, where the concentration of SO2 is rising. To make predictions for so2, we employed AR and ARIMA models. As a result, features such as location monitoring station or station code were of no value. The following are the So2 safe levels: Over the course of an hour, the average concentration was at 0.20 ppm. 0.08 ppm during the course of 24 hours. Over the course of a year, 0.02 ppm was averaged. PM2 5 is also an essential factor in predicting air quality. Future measurements are needed since these particles are linked to a variety of health issues, including cardiac arrhythmias and heart attacks, as well as respiratory issues such as asthma flareups and bronchial infections. Because the data in the date column is out of order, this model is unable to provide the predicted results. Cities face the same dilemma. It will not be beneficial if we forecast for the whole state. The AQI will now be calculated and categorization models used. This model also makes us aware of future issues and research requirements, such as PM2.5,AQI, etc., thanks to its simplicity and comprehensiveness.

REFERENCES

1. Amelia, N.R and Akhbar;

Arianingsih, Ida, 2015, "Pembuatan PetaPenutupLahanMenggunakanFotoUdarayangDibuatdenganParamotordiTamanNasionalLoreLindu(TNLL)"(TheDevelopment Of Land Cover Maps Using Aerial Photo of LorenLindu National Park /NPLL)", Warta Rimba Vol.3, Numb.3, pp.65-72,December.2015.

2. Harsono. N, Subhan.A , Sukaridhoto, S. Sudarso, A, 2006, "TeknikPemetaan Wilayah Secara Cepat dan AKurat Menggunakan GPSyang Dikoordinasikan Melalui Jaringan 3G atau yang Setara" (FastandAccurateEngineeringofAerialMappingUsingGPSCoordinated Through 3G Network and Equivalent", Proceedings oftheNationalConferenceonInformation&CommunicationTechnologyforIndonesia, 3-4May2006,Bandung.
3. Rafialy, Leonardo and Sediyono, Eko; Setiawan, Andi, "Pemanfaatan Cloud Computing Dalam Google Maps untuk Pemetaan,Informasi,AlihFungsi,LahandiKabupaten,MinahasaTenggara", Seminar NasionalTeknologiInformasi, 2013, pp.52-58.
4. Sendow, T.K and Longdong, Jefferson. "The Study Mapping (city case study Manado)". Scientific Journal Media Engineering Vol.2, Numb.1, pp.35-46, March. 2012.
5. D. Bort, "Android Is Now Available as Open Source,"Android Open Source Project.
6. C. R. Rani, A. P. Kumar, D. Adarsh, K. K. Mohan and K.V. Kiran, "Location Based Services in Android," International Journal of advances in Engineering & Technology, Vol. 3, No. 1, 2012, pp. 209-220.
7. [7]W. Kowtanapanich, Y. Tanaboriboon and W. Chadbunchachal, "An Integration of Hand-Held Computers, GPS Devices and GIS to Improve the Efficiency of EMS Data System," Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, 2005, pp. 3551-3561.
8. [8] Badanpertanahannasional. Accessedon Agustus 08, 2017.. [9]BadanPertanahanNasional.http://www.bpn.go.id/BERITA/Berita-
9. Pertanahan/tuntasdalam-10-tahun-

66621, Accessed on October 23, 2017.

10. Sahoo, B. P. S and Rath, Satyajit," Integrating GPS,GSM and Cellular Phone for Location Tracking and Monitoring," Proceedings of the International Conference on Geospatial Technologies and Applications, IIT Bombay, Mumbai, India, 2012, February 26-29.
11. Sugiyono, Metodepenelitiankuantitatif, kualitatifdan R & D. Bandung: Alfabeta., 2010.