OPEN ACCESS

Volume: 11

Special Issue: 1

Month: July

Year: 2023

E-ISSN: 2582-0397

P-ISSN: 2321-788X

Impact Factor: 3.025

Received: 10.05.2023

Accepted: 20.06.2023

Published: 01.07.2023

Citation:
Kumbhar, Shreedhar Maruti, and R. Meghana. "Machine Learning for Air Quality Prediction." *Shanlax International Journal of Arts, Science and Humanities*, vol. 11, no. S1, 2023, pp. 129–34.

DOI:
https://doi.org/10.34293/sijash.v11iS1-July.6327

Machine Learning for Air Quality Prediction

Shreedhar Maruti Kumbhar
Department of Masters of Computer Applications
Rajarajeswari College of Engineering

Meghana R
Department of Masters of Computer Applications
Rajarajeswari College of Engineering

Abstract

Environmental protection measures cannot be properly guaranteed in the present as a result of rapid industrialization. The main issue hurting the the standard of living in country is the escalating severity of environmental issues. Therefore, in in order to comprehend the potential air pollution process beforehand, we need to put together a reasonably good air quality forecasting model. Setting up and implementing appropriate control measures to lessen Air pollution is extremely important, in accordance with model's forecast results.Techniques for mining data, such as neural networks, mutual information theory, and intelligent optimisation algorithms are all extensively used in this article. As a training set and test set, we employ the fundamental information from open monitoring locations that predicts long-term air quality.Today, determining the the state of the air has emerged as among the most crucial significant activities for the residents of many industrial and metropolitan areas. Many kinds of pollution are caused by the usage of fuel, electricity, transportation, etc. have a detrimental effect on the air's quality. The grade of living in smart cities is significantly being impacted by the buildup of hazardous gases. In accordance with address the escalating levels of air pollution, we must put into place efficient air quality testing and prediction models that collect information on pollutant concentrations and offer evaluations of regional air pollution. Inhaling these minuscule solid or liquid droplets, which make up particulate matter, can have a serious negative impact on one's health.

Keyword: Air Quality, Machine Learning, Electrictyprediction, Quillbot

Introduction

Without explicit programming, a system of computer algorithms known as "machine learning" is accomplished of learning from experience and advancing itself. Artificial intelligence is exemplified via machine learning. predicts a result that can be applied to produce actionable insights using statistical techniques and data.

Automating processes such as fraud detection, proactive maintenance, and portfolio optimization, and so forth is another application of machine learning.

Programming as usual is considerably different from machine learning. In traditional programming, each rule would be written after consultation with the assistance of an expert in the industry for which the product was being developed. Every rule is supported

logically, and the computer will act in accordance with conclusion that follows the logical claim. As the system grows in complexity, more rules must be written. It can quickly become tough to maintain.

The machine is able to forecast the outcome when we provide a comparable scenario. However, if the machine is provided a fresh example, it fights to forecast, much like a person.

The core of (ML)Machine learning is learning and inference. the initialway the machine learns is by identifying patterns. The data allowed for this finding to be made. One among the skills of the data scientist is to choose the facts to deliver the computer with care. heir most important skills.

Literature Survey

Fuzzy logic and autoregressive techniques for measuring air quality and forecasting [1] Techniques for AI have been created and implemented in recent years to solve environmental issues. Two prototypes for the calculation and forecasting of state of air are presented in this paper. In accordance to identify chemicals that can be harmful harm sensitive persons in metropolitan settings and interfere with their usual activities, we first create a novel computational paradigm for airqualityassessment. In this model, we suggest employing the Sigma operator to statistically evaluate air quality parameters utilising their historical data information and identifying their detrimental effects on Airquality based on toxicity limits, frequency averages, and deviations of toxicological tests.

Forecast for Delhi's daily indicator of air quality [2] air pollution's effects concerning human health via ambient air have received significantrecent years' emphasis, environmental science in regard to air pollution, research has focused on forecasting indicators of Air Quality. Governmental organisations can use the Air Quality Index (AQI), which mightbe calculated using a formula based on an extensive examination of the air concentration pollutants, to describe the state of the Air Quality at a specific location. The drive of present work is to builda model for predicting daily AQI prediction that can serve as the foundation for decision-making procedures. First, The Environmental Protection Agency of the United States (USEPA) approach a number of factors were used to compute the AirQuality Index contaminants, including RSPM (Respirable Suspended Particulate Matter), (SO2) Sulphur Dioxide, (NO2) Nitrogen Dioxide and Particulate Matter in Suspended Form (SPM). The formula's breakpoint and sub-index concentrations, however, were calculated by utilising the (NAAQ)National Ambient Air Quality in India Standard.

Techniques for nonlinear and linear models predicting the quality of the airin cities [3] In this investigation, modelling techniques including the application of linear and nonlinear predict Lucknow, India's urban AirQuality. The (RSPM)respirable suspended particulate matter, SO2, and NO2 in the environment were predicted using the partial least squares method regression (PLSR), multivariate polynomial regression (MPR), and (ANN)artificial neural network approach-based models using meteorological (air temperature, relative humidity, wind speed), and air quality nursing data (SPM, NO2, SO2) of five years (2005-2009). Multilayer Perceptron Network (MLPN), Radial-Basis Function Network (RBFN), and Generalised Regression Neural Network (GRNN) are threenumerous (ANN) artificial neural networks models that were created.

Modelling for AirQuality is industrial complex and assessing model performance [4] Due to a concentration of industrial activity, Jamshedpur, an Indian steel metropolis in the east, is experiencing rising air contamination levels. The Gausian dispersion model was used to analyse the Industrial Source Complex in the short term. predict the effect of NOX emissions arising from different air pollution sources, including factories, cars, and household sources. The proportion of NOx from industrial, vehicular, and household sources contributed 53, 40, and 7%, respectively. In order to assess the model's performance, Additional statistical investigation was conducted. out.

by comparing the concentrations of measured and projected NOx. With a 68% accuracy rate, the model's performance was deemed to be satisfactory.

Evaluation of PM10 and PM2.5 concentration efficacy models for predicting urban traffic intersections throughout the winter [5]To assess There are several models available. may be used to calculate the air quality around roadways. To identify which model works the best, it is desirable to compare the operational performance of numerous models that are pertinent to local conditions. Three air quality replicas have been developed to assess the air quality at the town of Guwahati's major crossroads. been developed. identified: the "modified General Finite Line Source Model" (M-GFLSM) of particulates, the "California Line Source" (CALINE3) model, and the "California Line Source for Queuing & Hot Spot Calculations" (CAL3QHC) model.

Existing Work

Many research on the use of have been undertaken to measure air quality, including satellite wireless sensor networks, remote sensing, and dispersion models. The air is the mathematical models like the Box and Gaussian models. How does air pollution move across the atmosphere? The traditional dispersion models serve the primary purposes of meteorology in terms of traffic volumes. These models don't take into consideration specific situations and settings, such as human mobility and concentrations, and rely only on parameters that replicate the dispersion of pollution. Meanwhile, a correct model of dispersion depends on information about, for example, wind speed and traffic pollutants. In rare circumstances, accuracy cannot be assured due to considerations like these.The SOM Neural Network model is used To forecast the state of the air in the current system. Self-Organizing Map (SOM) is the name of the facts visualisation technique TeuvoKohonen suggested. High dimensional data can be analysed by reducing its dimensions to a map. more easily understood. It also exemplifies clustering by combining information that is related. After the introduction of a SOM variation called Growing Self Organising Ma, this clustering capability was warmly accepted.The SOM's fundamental flaw is that it demands neuron weights be sufficient and essential to cluster inputs. The groupings found in the map may not be totally accurate or instructive if a SOM is given too little information or too much superfluous information in the weights.

Anotherissue with SOMs is that it could be challenging to create an ideal mapping inwhich categories differ from one another. Instead, anomalies in the map frequently arise when two comparable groupings show up in various locations on the same map. In order to create many areas of similar neurons, clusters frequently split into smaller clusters. This can be avoided by properly initialising the map, however that is not an option if the ultimate map's state is obscure.

Proposed Methodology

These days, air pollution is a serious issue. Air pollution results from an undesired substance, contaminant, or agent being existing in the air. The air quality index, or AQI, is used to label the air around us. In this study, we will use the RF method to expect the presence of air contaminants in the atmosphere. The goal of this research is to investigate if we can predict the quantity of concentration in the air given other environmental parameters by using a publicly available dataset and (ML)machine learning techniques.The Indian government (CPCB) states that the AQI ranges from 0 to 500, with 0 denoting good AirQuality and 500 denoting severe air quality. Eight major pollutants must be taken into account in order to calculate the AirQuality index (AQI): lead (Pb), carbon monoxide (CO), ozone (O3), nitrogen dioxide (NO2), sulphur dioxide (SO2), and particulate matter (PM 10 and PM 2.5). It is essential to have data for at least three pollutants, one of which must be either PM10 or PM2, to be able tocalculate AQI.5. There are various concentrations and

each pollutant corresponding to health consequences in the AQI range of 0-500.
Using a Random Forest Classifier, the suggested method is designed to first forecast the class of the AirQuality in the area. The system was developed with Python. The dataset is used as an input. The dataset is examined using the RF Classifier, and generates results.

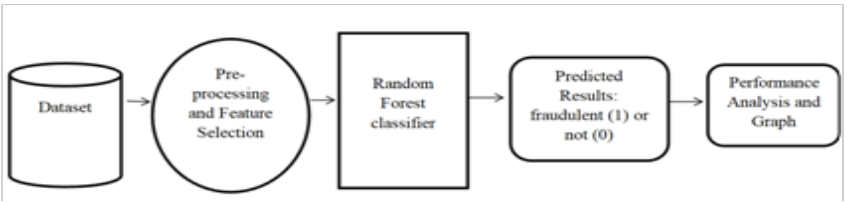

Figure 1 Proposed Architecture

Implementation

Data Collection

The actual procedure of creating a (ML)Machine learning model and accumulating data starts now. This stage is critical since the amount and quality of data we can gather will determine how effectively the model performs.The methods for acquiring the data are numerous and involve Web scraping and manual interventions. The dataset is placed in the model folder. The collection is taken from the well-known kaggle dataset repository. The dataset link is provided below.

Data Preparation

Amass data and prepare it for training. Remove duplicates, correct errors, handle missing numbers, normalise, convert data types, as well as anything else that may be required for cleaning up.By randomizing the data, we eliminate the effects of the exact sequence in which we gathered and/or otherwise prepared our data.Conduct additional exploratory research, such as data visualisation, to identify key correlations between variables. or class imbalances (bias alert!).There are many options for instruction and assessment.

Model Selection

We activethe algorithm for computer learning A random forest is the Random Forest Classifier. classifier. We used this approach since we had a train accuracy of 99%.
Algorithm for Random Forests:Let's explain the algorithm using everyday language. Consider how far you wish to travel., somewhere you will enjoy.So how can you locate a location that you will enjoy? You can perform an online search, browse travel reviews blogs and websites, or inquire your friends for recommendations.

Saving the Trained Model

When you are comfortable utilising your trained and tested prototype in a production-ready environment, the head step is to store it in a.h5 or.pkl file using a library like pickle.Make sure Pickle is configured properly in your location.The prototype will now be dumped as an a.pkl file and imported into the module.

Dataset

There are 29532 different informational items in the gathering. Each of the 16 columns in the dataset is described below.City: Types of cities

The date is dd/m/yyy.
Particulate Matter (PM2.5) 2.5 microns in ug/m3
Particulate Matter (PM10) 10 microns in ug/m3
Nitric Oxide in ug/m3 is NO.
Nitric oxide, measured in ug/m3, is NO2.
NOx: Any nitrous oxide present in ppb
NH3: Ammonia in ug per m3.
CO: Carbon Monoxide, measured in mg/m3.
Sulphur dioxide measured in ug/m3: SO2
O3: Ozone in a m3 of air.

Results

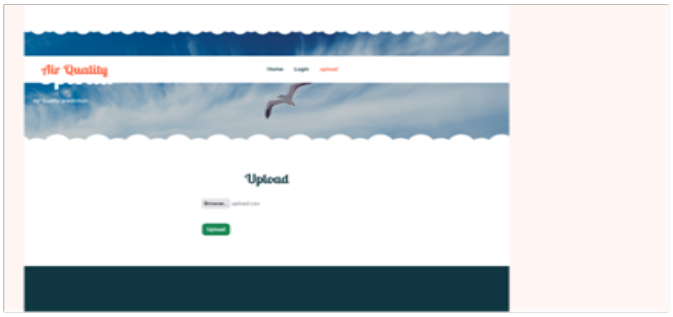

Figure 2 Upload Page

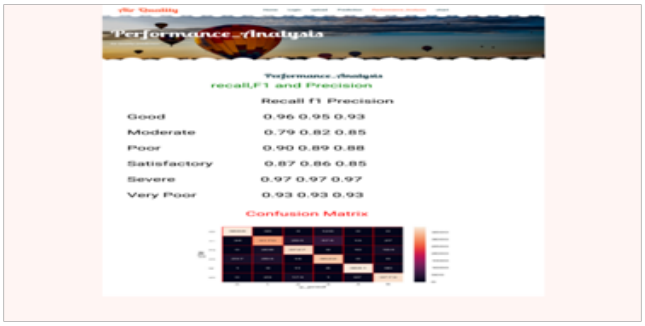

Figure 3 Performance Analysis Page

Conclusion

The model's key benefit is its ability to forecast data under various limitations based on available data, and also its ability to identify potential air pollution processes early on and implement the necessary controls to raise ambient AirQuality. The AirQuality prediction model presented and used in this work has the potential to significantly increase prediction accuracy and serve as a benchmark for subsequent air quality estimation studies. With precise and dependable results, the As of late, RF Classifier has become a viable option technique for predicting air quality.

References

1. Carbajal-Hernández, JouséJuaan "Assessment and prediction of air quality using fuzzy logic and autoregressive models." 37–50 in Atmospheric Environment 60 (2012).

2. Kumar, Anikender, and P.Goyal, "Forcasting of daily air quality index in Delhi", Science of the Total Environment 409, no. 24(2011): 5517-5523.
3. "Linear and nonlinear modeling techniques for predicting urban air quality," Science of the Total Environment 426(2012):244–255.
4. Environmental Pollution 111.3 (2001): 471-477; Sivacoumar R, et al., "Modeling air pollution in an industrial complex and evaluating model performance."
5. NamitaRaokhande and GokhaleSharad, "Evaluation of the Performance of Air Quality Models for Predicting PM10 and PM2.5 Concentrations at Urban Traffic Intersections During the Winter Season ", Science of the Total Environment 394.1(2008): 9-24.
6. Bhanarkar A.D., et al. "Evaluation of the contribution of SO2 and NO2 from various sources in the Jamshedpur region of India," Atmospheric Environment 39.40 (2005): 7745-India.
7. Bhanarkar, A. D., et al. "Evaluation of the contribution of SO2 and NO2 from various sources in the Jamshedpur region, India," Atmospheric Environment 39.40 (2005): 7745-India.
8. Milan Tripathi, "Analysis of Convolutional Neural Network-based Image Classification Techniques." JIIP 3, no. 02 (2021): 100–117. Journal of Innovative Image Processing.
9. "Deep Convolution Neural Network Model for Credit -Card Fraud Detection and Alert." Chen, Joy Iong-Zong, and Kong-Long Lai.Journal of Artificial Intelligence, Volume 3, Number 2, Pages 101–112