International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 08 | Aug 2022 www.irjet.net p-ISSN: 2395-0072

A Smart air pollution detector using SVM Classification

M.Meghana[1], Dr.R.Maruthamuthu[2]

[1]student, Department of Computer Applications, Madanapalle Institute of Technology and science, India
[2]Asst.Professor, Department of Computer Applications, Madanapalle Institute of Technology and science, India

---

Abstract - One of the top priorities for the governments of developing nations, especially India, is the control of the fast rising levels of air pollution. People can take action to reduce pollution by becoming more aware of the degree of pollution in their immediate surroundings. Fossil fuel combustion, travel habits, and industrial elements like power plant emissions all have a big impact on air pollution. The total amount of particulate matter (PM) that affects air quality. When it is concentrated heavily in the aerial medium, it poses serious health risks to people. It must therefore be controlled by regularly checking its atmospheric concentration.

Key Words: Particulate matter, SVM classifier, Regression, and Quality

1.INTRODUCTION

There can be both naturally occurring and artificial particles. Examples include dust, ash, and sea spray. Burning of solid and liquid fuels, such as when creating energy, heating a home, or driving a car, releases particulate matter (including soot). The size of the particles varies (i.e. the diameter or width of the particle). The term "PM2.5" refers to the quantity of airborne particles per cubic meter of air that have an average diameter of less than 2.5 micrometers Another name for it is fine particulate matter, or PM2.5. When airborne levels of tiny particulate matter (PM2.5) are quite high, it poses a substantial risk to people's health and is a significant portion of the pollutant index. PM2.5, or particulate matter 2.5, lowers visibility and causes the air to appear hazy when concentrations are high. The identification of air pollution and forecasting of PM2.5 levels have been accomplished using a variety of machine learning models based on a data set made up of daily atmospheric conditions. Dan Wei forecasted Beijing's air quality using the Naive Bayes classification and support vector machine algorithms to get the lowest possible error. José Juan Carbajal developed the fuzzy inference technique, which he then applied to categorize parameters using logic and include them in an air quality score.

1.1 Naïve Bayes Classification

A group of classification methods built on the Bayes Theorem is known as naive Bayes classification. Every pair of features being categorised independently from one another is not a common principle shared by all of the algorithms. It is a supervised learning algorithm that uses the Bayes theorem to solve classification issues. It is mostly employed in text classification tasks with high-dimensional training data.

LITERATURE SURVEY

[1] A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization. Dixian Zhu, Changjie Cai, Tianbao Yang, and Xun Zhou

In this study, we address the problem of air quality forecasting by predicting the hourly concentration of air pollutants, such as ozone, particle matter (PM 2.5), and sulfur dioxide. One of the most used techniques, machine learning, can effectively train a model on massive amounts of data by employing powerful optimization algorithms. Although some studies have used machine learning to predict air quality, most of the earlier research has only used data from a few years and has only trained basic regression models (either linear or nonlinear) to predict the hourly air pollution concentrationBy defining the prediction across 24 hours as a multi-task learning (MTL) issue, we offer improved models in this study to forecast the hourly air pollution concentration based on meteorological data from previous days. This makes it possible for us to choose a suitable model using various regularization methods. We suggest a practical regularization by mandating that the prediction models for consecutive hours be near one another and contrast it with other common regularizations for MTL, such as ordinary Frobenius norm regularization, nuclear norm regularization, and l 2, 1 -norm regularization. Our tests demonstrated that the suggested parameter-reducing formulations and consecutive-hour-related regularizations outperform existing standard regression models and existing regularizations in terms of performance

[2]. Sachit Mahajan, Ling-Jyh Chen, and Tzu-Chieh Tsai are the authors of "An Empirical Study of PM2.5 Forecasting Using Neural Network".

In most industrialized and developing nations, significant efforts have been undertaken in recent years to restrict air pollution levels. Many efforts are being undertaken to control the levels of fine particulate matter (PM2.5), which is thought to be one of the main causes of declining public health. Forecasting PM2.5 levels accurately is a difficult undertaking that has relied heavily on model-based approaches. In this study, we investigate fresh approaches to PM2.5 hourly forecasting. In order to increase prediction

accuracy, selecting the appropriate forecasting model becomes crucial. For the prediction job, we employed the Neural Network Autoregression (NNAR) approach. Additionally, the research compares the predictive abilities of the additive Holt-Winters approach, the autoregressive integrated moving average (ARIMA) model, and the NNAR model. Utilizing actual measurement data from the Airbox Project for experimentation and evaluation, it can be seen that our suggested method makes predictions accurately and with a relatively little amount of error.

[3]. Dan Wei: Predicting the degree of air pollution in a certain city

One of the most crucial jobs for the governments of developing countries, especially China, is the management of air pollutant levels. Fine particulate matter (PM2.5) is an important component of the pollution index since its excessive levels in the air pose a serious threat to people's health. When levels are high, PM2.5, or particulate matter 2.5, reduces visibility and gives the air a hazy appearance. However, it is unclear how the concentration of these particles and traffic and weather conditions interact. Some of these cutting-edge methods have been applied to air quality research to further clarify these links. These studies used a few methodologies, primarily meteorological and occasionally traffic data, to estimate ambient air pollution levels using Support Vector Machine (SVM) and neural networks. In this experiment, machine learning techniques were applied to a dataset of daily meteorological and traffic factors in Beijing, China, in an effort to predict PM2.5 levels. Due to the uncertainties around the precise number PM2.5 level, I simplified the issue by categorizing the PM2.5 level as either "High" (> 115 ug/m3) or "Low" (= 115 ug/m3). The amount was determined using the Chinese Air Quality Level Standard, which defines mild pollution as 115 ug/m3.

[4]. Machine learning method for predicting sub-micron air pollution indicators, by Pandey, Gaurav, Bin Zhang, and Le Jian.

For the governments of emerging nations, especially China, controlling air pollution levels is quickly becoming one of their top priorities. The relationship between the concentration of submicron particles and meteorological and traffic factors is poorly understood, but submicron particles, such as ultrafine particles (UFP, aerodynamic diameter 100 nm) and particulate matter 1.0 micrometers (PM1.0), are an unregulated emerging health threat to people. e used a variety of machine learning algorithms to forecast UFP and PM1.0 levels based on observations of meteorological and traffic factors recorded at a busy roadside in Hangzhou, China, in order to throw some light on these links. We find that it is possible to predict PM1.0 and UFP levels relatively accurately and that tree-based classification models (Alternating Decision Tree and Random Forests) perform the best for both of these particles based on a detailed analysis of the more than 25 classifiers employed for this purpose. Additionally, weather factors cannot be disregarded when projecting submicron particle levels because they have a larger correlation with PM1.0 and UFP levels. The overall application value of methodically gathering and analyzing datasets using machine learning approaches for the prediction of submicron sized ambient air contaminants has been shown in this study.

[5]. Carbajal-Hernandez, Juan Luis P. and José Sánchez-Fernándeza JesúsA.Carrasco-Ochoab Fuzzy logic and autoregressive models for assessing and forecasting air quality, by JoséFco.Martinez-Trinidad

Artificial intelligence techniques have been applied in recent years to solve environmental issues. Two models for the evaluation and forecasting of air quality are presented in this paper. In order to identify harmful substances that can hurt sensitive persons in metropolitan settings and interfere with their usual activities, we first create a novel computational model for air quality assessment. In this model, we suggest employing the Sigma operator to statistically evaluate air quality parameters utilizing their historical data information and identifying their detrimental effects on air quality based on toxicity limits, frequency averages, and deviations of toxicological tests. Additionally, we present a fuzzy inference system to classify parameters through a process of reasoning and integrate them into an air quality index that categorizes pollution levels into five stages: excellent, good, regular, bad, and danger. The second model put out in this work uses an autoregressive model to forecast air quality concentrations and provides a predicted air quality index based on the previously created fuzzy inference system. We compare the air quality indices created for environmental agencies and related models using information from the Mexico City Atmospheric Monitoring System. Our findings demonstrate that our models are a useful tool for evaluating site pollution and for offering recommendations to enhance contingency actions in urban environments.

2.EXISTING SYSTEM

The current systems identify the user-selected city's air quality and categorise it according to AQI into several categories like good, satisfactory, moderate, poor, extremely bad, and severe (Air Quality Index). On a monthly, weekly, or daily basis, the data is shown. Additionally, once the values are predicted, they remain unchanged in the event that atmospheric conditions suddenly alter or traffic unexpectedly increases.

Disadvantages

- ✓ Have a limited degree of accuracy because they can't foresee when pollution will be at its lowest and highest concentrations.
- ✓ Substantial mathematical answers
- ✓ They are an insufficient strategy for more accurate production estimates

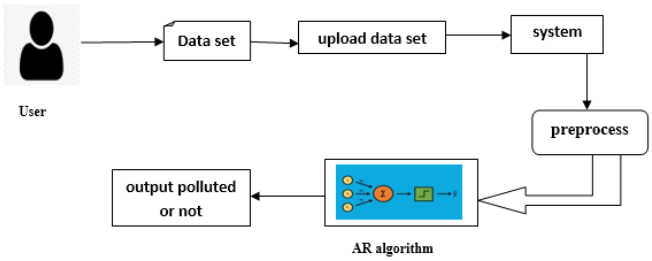
3.PROPOSED SYSTEM

The suggested system performs two duties. I Based on specified atmospheric variables, it determines the PM2.5 concentrations. (ii) Forecasts the PM2.5 concentration for a specific date. To determine if a data sample is contaminated or not, logistic regression is used. The main objective is to use ground data to anticipate the city's air pollution level. The suggested system will make it easier for regular people and meteorologists to identify and forecast pollution levels and take the appropriate measures accordingly

Advantages

- Pollution levels are simple to identify and forecast.
- A practical strategy for improved output prediction

Block chart



SVM Classification

To handle classification and regression issues, the Support Vector Machine (SVM), one of the most well-liked supervised learning techniques, is used. However, classification issues are mostly addressed by it in machine learning. The SVM method's objective is to produce the ideal decision boundary or line that can categorize n-dimensional space, allowing incoming data points to be quickly assigned to the appropriate category.

The two forms of SVM

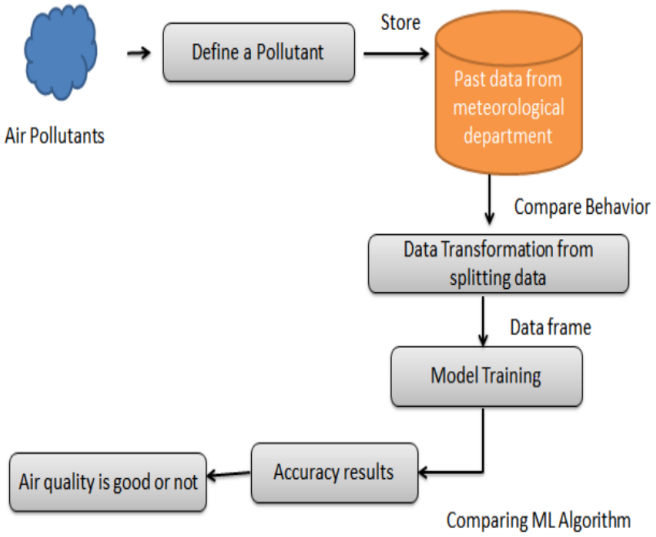
Linear SVM : Data that can be separated into two groups using just one straight line are referred to as linearly separable data, and linearly separable data is used in linear SVM. Linear SVM classifiers are used to categorize such data.

Non-Linear SVM : When a dataset cannot be classified using a straight line, it is said to have been non-linearly separated; in this case, the classifier used is known as a non-linear SVM classifier.

Regression: A dependent variable's type and the strength of its association with a number of independent variables are to be determined using the statistical technique known as regression. Regression is utilized in the fields of finance and investment. Regression problem solving is one of the most often used applications of machine learning models, particularly in supervised learning. understanding the relationship between independent factors and a product or dependent variable

Structure of the System

When the dependent variable is dichotomous, you should use logistic regression as your regression model (binary or has two classes). In this case, the data set is divided into two groups for demonstration purposes: contaminated and unpolluted. The logistic regression is a predictive analysis, as are other regression studies. The link between a single binary dependent variable and one or more independent variables can be explained using logistic regression.



4.RESULT ANALYSIS

Information was provided to participants based on PM10 concentrations discovered using a machine learning system throughout the research. The system is based on a dynamic, interactive, and always-updated smart pollutant. The machine learning model chosen has the maximum efficiency and stability after extensive testing across 6 modules with varying settings.

When using sensor inputs, the system provides an updated and calibrated method for data processing, makes PPM calculations accurate, and prepares them for presentation in compliance with authorized air quality index values. Along with the previously collected PM10 data, these values are fed into the trained model to predict the production of smog.

5.CONCLUSIONS

Air pollution regulation is increasingly becoming one of the most important responsibilities. By becoming aware of the level of pollution in their local surroundings, people can take action to lessen pollution. The results show that machine

learning models (auto regression and logistic regression) may be used to predict future air pollution levels and evaluate air quality with high accuracy. The suggested technology will make it easier for members of the general public and meteorological department staff to recognize and predict pollution levels and take the proper action in response. Additionally, this will help people in their quest for information by developing a data source for small towns, which are frequently ignored in favor of major metropolis.

REFERENCES

[1] "A machine learning technique to forecasting sub-micron air pollution indicators," by Le Jian, Bin Zhang, and Gaurav Pandey. Processes and Impacts in Environmental Science 15.5 (2013): 996–1005

[2] Predicting the level of air pollution in a specific city by Dan Wei [2014]

[3] A Model Regularization and Optimization Approach for Machine Learning in Air Quality Prediction. Tianbao Yang, Dixian Zhu, Changjie Cai, and Xun Zhou. Big data and cognitive computing [2018]. Carbajal-Hernandez, José Juan

[4] Luis P. Sánchez-Fernándeza JessA.Carrasco-Ochoa and JoséF.Co. Martnez-Trinidad: National Polytechnic Institute, Center of Computer Research, Av. Juan de Dios Batiz S/N, Gustavo A. Madero, Col. Nueva, Industrial Vallejo, 07738 Mexico, D.F. Fuzzy logic and autoregressive models for the assessment and forecasting of air quality (2012) Doi:https://doi.org/10.1016/j.atmosenv.2012.06.004

[5] Using a neural network, Sachit Mahajan, Ling-Jyh Chen, and Tzu-Chieh Tsai's paper An Empirical Study of PM2.5 Forecasting appeared in IEEE.