Literature survey

|  |  |  |  |
| --- | --- | --- | --- |
| Sr no. | Name of Author | Title of paper | Description |
| 1 | Gaganjot Kaur Kang, Jerry Zeyu Gao, Sen Chiao, Shengqiang Lu, and Gang Xie | Air Quality Prediction: Big Data and Machine Learning Approaches | * This paper reports recent literature study, reviews and compares current research work on air quality evaluation based on big data analytics, machine learning models |
| 2 | Chen Xiaojun1, Liu Xianpeng2, Xu Peng3 | IOT- Based Air Pollution Monitoring and Forecasting System | * It was also observed that adding more meteorological factors, the prediction performance is greatly improved. * They used past 5 years data for training the model. Which shows large sample data can improve performance and train the model well. * The overall experiment used IOT and neural network for air pollution monitoring and forecasting. |
| 3 | Temesegan Walelign Ayele , Rutvik Mehta | Air pollution monitoring and prediction using IoT | * The proposed work on an air pollution monitoring and prediction system is enables us to monitor air quality with the help IoT devices * For predicting the LSTM is implemented. It has a quick convergence and reduces the training cycles with a good accuracy |
| 4 | Zhongshan Yang, Jian Wang | A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction | * In this paper, fuzzy comprehensive evaluation was used to determine the main air pollutants and evaluate the level of air pollution. * The experimental results showed that the proposed model has the best accuracy and stability compared to the other five individual models and five combined models. |
| 5 | Dixian Zhu 1,\*, Changjie Cai 2, Tianbao Yang 1 and Xun Zhou 3 | A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization | * They have developed efficient machine learning methods for air pollutant prediction. * They have formulated the problem as regularized MTL and employed advanced optimization algorithms for solving different formulations. |
| 6 | Mahmoud Reza Delavar 1,\* , Amin Gholami 2, Gholam Reza Shiran 3, Yousef Rashidi 4 , Gholam Reza Nakhaeizadeh 5, Kurt Fedra 6 and Smaeil Hatefi Afshar 2 | A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran | * A comparative study of machine learning methods including NARX, ANN and SVR has been employed for air pollution prediction and the NARX finally selected as the optimum one. |
| 7 | Xiuwen Yi1,2, Junbo Zhang2,1,+, Zhaoyuan Wang1,2, Tianrui Li1, Yu Zheng2,1,3 | Deep Distributed Fusion Network for Air Quality Prediction | * They proposed a DNN-based approach to predict air quality. * This approach achieves a higher accuracy in both general cases and sudden changes |
| 8 | QU Hongquan , PANG Liping | A Dynamic Method to Estimate Source Emission  Rate and Predict Contaminant Concentrations  A | * This paper develops a dynamic method to estimate source emission rate and predict contaminant concentrations in an enclosed space. * Based on a variable structure model ofconcentration, this method uses EKF algorithm in combination with least squares method to realize state prediction and parameter estimation at the same time. * This method could realize to track and real-time predict contaminant concentration, and identify source emission rate accurately and efficiently. |
| 9 | N. Li and S. Thompson | A Simplified Non-Linear Model Of NOx Emissions In A Power Station Boiler | * This paper has presented a model of NO emissions for a power plant boiler. It is modelled from the extended Zeldovich mechanism and require only a few physical parameters obtained from experiments. * A set of new test data is used to compare the simulated values with real measurements. It is shown that good results are   obtained from the model with  real plant input variables.   * The model can also be used in other applications such as for optimising boiler operation and combustion control system design. |
| 10 | S.H. Yu, Y.S. Koo, E.Y. Ha and H.Y. Kwon | PM-10 Forecasting using Neural Networks Model | * In this paper, we study on the factors to affect the PM-10 pollution and develop a PM-10 prediction model using MLP neural network model. * Especially, neural model has an advantage that there doesn’t need to analyze the input data before the data are used, like regression model. * To improve the performance of the model, it needs to shorten the learning period from year to quarter month and to learn and predict PM-10 with multiple networks according to the PM-10 levels. |
| 11 | Moustafa.Elshafei, Mohamed A.Habib, Mansour Al-Dajani | Prediction Of Boilers Emission Using  Polynomial Networks | * The paper provided an efficient polynomial network solution to the problem of on-line monitoring of NOx emission from industrial boilers. * The effect of six variables were studied using 3D CFD simulation model and used by polynomial networks for prediction of NOx and 02 in the exhaust flue. |
| 12 | S. Raza, R. Avilaand J. Cervantes | A 3D Lagrangian Particle Model For The Atmospheric  Dispersion Of Toxic Pollutants | * The Lagrangian Monte Carlo particle dispersion models work very efficiently for the atmospheric   dispersion of effluents.   * In order to incorporate the effect of vertical wind shear the modiffed dispersion coeffcient should be used with the Gaussian plume model. |
| 13 | H. Kaplan And N. Dinar | A Lagrangian Dispersion Model For Calculating  Concentration Distribution Within A Built-Up  Domain | * In this paper we present a diagnostic model for calculating concentration distribution of a passive scalar in a built-up area. * The model requires measurements of the wind velocity and direction at a reference height above the obstacles. * The model is able to predict 3-d concentration distributions and to identify concentration accumulation at specific points. the model succeeds in predicting concentration distribution quantitatively and qualitatively and can be used to study many air pollution phenomena. |
| 14 | By Peter De Haan’ And Mathias W. Rotach | **A** novel approach to atmospheric dispersion modelling: The Puff-Particle Model | * In the present paper, an approach to model dispersion is presented which aims at combining the advantages of puff models and particle models. * The resulting model type is called Puff-Particle Model (PPM). In the PPM, a few hundred puffs are simulated in three-dimensional space, as compared to many thousand particles usually required in pure particle models. * The concept of the PPM is very simple: while puff growth is described by the concept of relative dispersion (thus accounting for eddies smaller than the puff), the effect of meandering (i.e. the variation between the trajectories of different puffs) due to larger eddies (larger than the actual puff size) is simulated by introducing puff-centre   trajectories derived from particle trajectories from a particle model. |

METHODOLOGY

Our overall setup consists of network of sensors that will be mounted in a specific industry, the data collected from these sensors will be stored on the server.

These sensors measure the air parameters in terms of ambient air as well as stack emission.

On this data we apply various machine learning algorithms for prediction of emission rate.

The air dispersion models are then applied on the predicted emission rate to calculate the dispersion of pollutants from the source that is at the stack level.

The entire system is basically divided into two broad categories

1. IOT:

The IOT process flow begins with measurement of meteorological air parameters using sensors, which is divided into ambient air parameters and stack emission parameters.

We considered these parameters to predict the value of V (velocity of wind) and Q (emission rate).

This data from the sensors is uploaded and stored on the cloud setup.

1. MACHINE LEARNING:

The machine learning process flow begins with the generation of dummy data in bulk using python. The dummy data was not truly random, it was correlated with various meteorological air parameters so that the machine could be trained well. It was observed that by adding more metrological factors, the prediction performance is greatly improved and large sample data can also improve performance and train the model well.

Next, machine learning algorithms were implemented on the created dummy data to predict the value of Q-emission rate and V-velocity of wind.

For this purpose, the data was divided into training set (80%) and test set (20%). If the error reduces for training as well as test data, the process is continued. Else if the error reduces only for training data, but increases for test data then the process is terminated. This could greatly increase chances of generalizability of the algorithm.

The performance check was then conducted on the predicted emission rate .The mean square error was measured in each case to check for accuracy.

Optimization of the various algorithms was done in such a way so as to reduce the error as minimum as possible thereby increasing the accuracy of prediction. On this basis, the best algorithm was selected.

To calculate the extent of pollution spread we used Gaussian dispersion model since it was the most optimal model in terms of computing power.

In this way the entire process of prediction of pollution and calculation of its spread is done.

MAHINE LEARNING ALGORITHMS

The various machine learning algorithms that were applied are as follows:

Kth Nearest neighbors (KNN)-This is a supervised machine learning algorithm which assumes that similar things exist in close proximity (similar things are near to each other). The value of K indicates the nearest neighbors that can be taken for consideration. Its purpose is to use a database which is separated into several classes to predict classification on a new sample point. This point is classified by a majority vote of its neighbors, it is assigned to the class most common to its K nearest neighbors It can also be used for regression wherein the output of the point is the average or median value of its nearest neighbors. This is a non-parametric technique of classification since it does not make any assumptions based on the underlying data distribution. Also, in KNN there is lack of generalization which means that the training data required is very less.

Support vector Regression (SVR)-SVR is a supervised machine learning algorithm which can be used for both classification and regression problems. In this algorithm each data point is plotted in n dimensional space after which classification is performed by finding the hyperplane that will differentiate the classes very well. The hyperplane that is considered can be a linear separator of any dimension (line, plane, hyperplane). The training points are used in the decision function and are called support vectors.

Random Forest-Random forest is a method that operates by constructing multiple decision trees during training phase. The decision of the majority of the trees is chosen by the random forest as the final decision .The decision tree is a decision support tool. It uses a tree like graph to show possible consequences. When a training data set is considered with targets and features, some set of rules are formulated, these rules are used to perform predictions. It identifies the most important features out of all the available features in the training dataset. The lager the no. of trees the more accurate the results will be. Random forest classifier handles missing values and overfitting problem doesn’t exist.

Multi linear regression (MLR)-It is a technique that uses several explanatory variables to predict the outcome of response variable. A linear relationship is modelled between these independent and response variable (dependent).Prediction about one variable is done based on the information about the other variable. In this case the independent variables should not be too highly correlated with each other.

Neural Network-it is an algorithm that mimics the way the human brain operate. A neuron in a neural network is a mathematical function that collects and classifies information according to some specific architecture. A neural network contains layers of interconnected nodes .Each connection is associated with a weight that is multiplied with the input value .Each neuron has an activation function that defines the output of the neuron which is used to introduce non linearity in the network model.

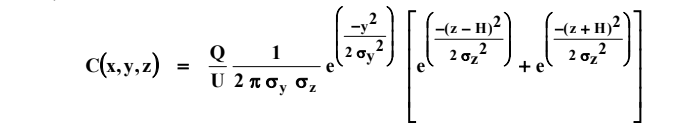
These Machine learning algorithms were implemented using python and the mean square error of each of these was measured to check for accuracy.

GAUSSIAN DISPERSION MODEL

In Gaussian dispersion model, the concentration of pollution downwind from a source is treated as spreading outward from the stack.

A point source was considered somewhere in the air where a pollutant is released at a constant rate Q (kg/s). The wind is blowing continuously in a direction x (measured in meters from the source) with a speed U (m/s). The plume spreads as it moves in the x direction such that the local concentrations C(x,y,z) (kg/m3) at any point in space form distributions which have shapes that are “Gaussian” or “normal” in planes normal to the x direction.

The total concentration at point P is given as



The parameters σy and σz (m) are the standard deviations of these Gaussian distributions, which indicate the spread of the plume in the y and z directions, respectively. They increase with the distance x from the source.

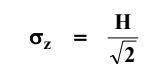
H = effective height of plume center-line (m)

hs = height of source above ground (m)

∆h = initial plume rise (m)

z = coordinate measured vertically from the ground to a point in the plume (m)

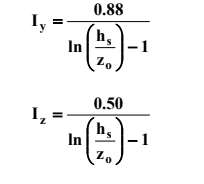
Maximum concentration occurs when-



There are many formulae and semi-empirical expressions available for determining σy and σz under different conditions of atmospheric stability. A reasonable approximation in regions near to the source when the source is elevated above the ground (such as at the top of a chimney) is

σy = Iy . x and σz = Iz . x

Where Iy and Iz are the turbulent wind speed fluctuations (turbulence intensities) in the y and z directions,



zo is the aerodynamic roughness representing different topographic ground conditions

Hence, the concentration is equal to the rate of emission from the source divided by the wind speed and then multiplied by the shaping function.