Pollution Detection and Prediction System

Project Report

Submitted in partial fulfillment of the requirements for the

Degree of Bachelors in Engineering

by

Rohit Martires	Swizel Monteiro	Vranda Asolkar	Vaylon Fernandes
(Reg no. 1613032)	(Reg no. 1613033)	(Reg no. 1613001)	(Reg no. 1613017)

Under Guidance of

Dr. Shreyas Simu	Dr. Varsha Turkar	Mr. Vassant Salgaonkar
(Internal Guide)	(Internal Guide)	(External Guide)

12/21/19

Contents

1 INTRODUCTION **3**
1.1 Project area 3
1.2 Motivation 3
1.3 Gap Analysis 3
1.4 Objectives of the project 4
1.5 Dataset and study area 4

2 LITRATURE SURVEY **5**

3 METHODOLOGY **8**
3.1 MACHINE LEARNING ALGORITHMS 11
3.2 AIR DISPERSION MODEL 12
3.2.1 Gaussian Dispersion Model 12
3.2.2 Lagrangian Model 13
3.2.3 Puff-Plume Model 13

4 RESULTS **13**
4.1 Machine Learning module 14
4.2 Air Dispersion module 18

5 CONCLUSION **19**

References **19**

List of Figures

1 Methodology flow diagram 9
2 IOT process flow 10
3 Machine Leaning process flow 11
4 parameter tuned max depth 15
5 parameter tuned max features 15
6 parameter tuned min sample leaf 15
7 parameter tuned min sample split 15
8 parameter tuned n estimators 16
9 parameter tuned activation function 16
10 parameter tuned number of neurons 16
11 parameter tuned solver 17
12 parameter tuned kernel 17
13 parameter tuned alogorithm 18
14 parameter tuned weights 18
15 predicted and actual data using samples form the test data . . . 18
16 predicted and actual data using samples form the train data . . . 18
17 Gaussian air dispersion model 19

1 INTRODUCTION

With the rapid growth of economy, industrial activities are increasing more frequently, leading to faster rate of pollution this rate is only increasing and if not kept in check can cause harmful effects to mankind and other living organisms. This effects are amplified if no regulation is kept in effect. Environmental pollution is one of the most serious problems facing humanity and other life forms on our planet today, industrial pollution contributing a major share in it. Industrial pollution is generally referred to the undesirable outcome when factories or other industrial plants emits harmful by-products and waste into the environment such as emissions to air or water bodies. The six major types of pollutants are carbon monoxide, hydrocarbons, nitrogen oxides, particulates, Sulphur dioxide and photochemical oxidants.

1.1 Project area

The Paris agreement's central aim is to strengthen the global response to the threat of climate change by keeping a global temperature rise this century well below 2 degrees Celsius above pre-industrial levels.[1] Long term exposure to polluted air and water cause's chronic health problems making the issue of industrial pollution into a severe one .It also lowers the air quality in surrounding areas which causes many respiratory disorders affecting both lungs and heart. Not just the humans, but the marine life is greatly deteriorating and affected with the extent of increasing industrial pollution. However with effective measures, the ill effect of industrial pollution could be reduced significantly. The prevention and control of industrial pollution are highly encouraged by government worldwide. Simple things like purchasing energy-efficient equipment and products made from recycled materials for your organization. Having industrial pollution control policies in place and guiding strictly upon them.

1.2 Motivation

The Draft Environment Laws (Amendment) Bill, 2015 was published by the Ministry of Environment, Forest and Climate Change (MoEFCC) on October 7, 2015. The objectives of the Draft Bill are to provide for "effective deterrent penal provisions" and to introduce "the concept of monetary penalty for violations and contraventions".There are no effective strict rules for pollution monitoring and control in industries yet. But the government of India is in the midst of making industrial pollution control and monitoring laws more strict.

1.3 Gap Analysis

The previous work in this field included setting up monitoring stations to measure the amount of pollutants in the atmosphere .This was done using network of sensors topologically arranged as a grid.[2] These sensors are usually placed in a large area for e.g. around a city or a large industry. The placement of sensors around a large area can reduce the accuracy of the model. This happens because of external factors like weather, noise, etc. affect the accuracy of the sensors to a large extent. Also, these models can't accurately reflect the extent of pollution a particular industry is causing due to the large area. Some previous models also predicted pollutant concentration for the future based on machine learning algorithms. Also there were some models that predicted dispersion of pollutants from a point source i.e. the area to which the pollutants will spread to.[3] No previous models have linked the emission rate of the stack and the dispersion from the source(stack) together i.e. they have not been predicting the emission rate and calculating the spread in the same model.

The proposed model will however predict emission rate at the stack and use it as a source to calculate dispersion using air dispersion models. This will be done by placing sensors at the source (stack). Placing the sensors at the source will help the sensors take more accurate readings as the sensors will not be affected by external factors to a large extent as compared to the previous models. This will allow us to take more accurate readings and hence make more accurate predictions of future emission rates.

1.4 Objectives of the project

The main objective of the project is to monitor the pollutants emitted from an industry/factory and predict the future dispersion of these pollutants also to then output these results in the form of reports for industries/factories. To accomplish the main objective can be broken down into 3 sub-objectives that are

- To create a series of gadgets which can measure the emission parameters and the meteorological parameters and also that can withstand the environment conditions present near the stack/chimney and transmit this information to a server.
- To develop a module that uses machine learning models with data acquired from the cloud server to predict future emission parameter values.
- To create a module that simulates the movement of fluid particles (pollutant) in air using air dispersion models with meteorological data.

1.5 Dataset and study area

The objective of the prediction module is to predict the emission rate at the stack/chimney. To build and test the prediction model dummy data was generated using Gaussian distribution to randomly generate values for the feature set. The feature set is chosen on the bases on what factors will correlate to the emission rate and the type of pollutants emitted from the stack.

The feature set consists of independent variables: day, month, type of industry, size of industry, and output efficiency of industry and dependent variables: emission rate.

Definitions:

- Type of industry: what the industry/factory produces which will correlate to what gases are emitted out of the stack/chimney, this will be in the form of labeled classes.
- Size of industry: how big is the industry/factory which will correlate how much the maximum is outputted, this will be represented in the form of a scale from 1 to 10.
- Output efficiency: the amount of output it produces each day divided by the total amount of output it can ideally produce.
- Emission rate: this is defined as the amount of pollutants released from the stack per unit time.

The prediction module is also used on the collected metrological data to predict the air velocity and direction, this is then applied to calculate the dispersion of pollutants in air.

The feature set of this prediction model consists of independent variables: day, month, ambient temprature, ambient pressure, moisture content and dependent variables: wind velocity and wind direction.

2 LITRATURE SURVEY

- A Lagrangian Dispersion Model For Calculating Concentration Distribution Within A Built-Up Domain April 1996 H. Kaplan And N. Dinar 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.
- A novel approach to atmospheric dispersion modelling: The Puff-Particle Model 1998 By Peter De Haan' And Mathias W. Rotach 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.

- A 3D Lagrangian Particle Model For The Atmospheric Dispersion Of Toxic Pollutants 2002 S. Raza, R. Avilaand J. Cervantes The Lagrangian Monte Carlo particle dispersion models works very efficiently for the atmospheric dispersion of effluents. In order to incorporate the effect of vertical wind shear the modifed dispersion coefficient should be used with the Gaussian plume model. As an alternative, a much reliable 3D numerical model as presented here, may be used at a slightly higher computational cost. Such a numerical model may also be used in a complex topography region with mountains, where the conventional Gaussian models are not suitable.

- A Simplified Non-Linear Model Of NOx Emissions In A Power Station Boiler 1996 N. Li and S. Thompson 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.

- Prediction Of Boilers Emission Using Polynomial Networks May 2006 Moustafa.Elshafei, Mohamed A.Habib, Mansour Al-Dajani 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. The prediction ofNOx and 02 are essential for efficient operation of the boiler while maintaining the NOx pollutant within a tolerable limit. The proposed softsensor has a simple modular structure for low cost implementation. The softsensor can be integrated with the boiler control system for optimization of boilers operation.

- PM-10 Forecasting using Neural Networks Model 2008 S.H. Yu, Y.S. Koo, E.Y. Ha and H.Y. Kwon In this paper,the study on the factors to affect the PM-10 pollution and develop a PM-10 prediction model using MLP neural network model was done. 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.

- IOT- Based Air Pollution Monitoring and Forecasting System 2015 Chen Xiaojun1, Liu Xianpeng2, Xu Peng3 Air pollution monitoring and forecasting system designed in this paper proposed a good solution to the complexity of air pollution. The use of a large number of sensors ensures monitoring accuracy, reduces monitoring cost and makes monitoring data in monitoring area more systematic and perfect. 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.

- A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction 2017 Zhongshan Yang, Jian Wang 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. More accurate forecasting could reduce the impact of air pollution on people's health and guide people's work and life. However, the chaotic nature and complexity of the air pollution data itself make prediction very difficult. In this paper, a novel hybrid model was proposed to forecast the concentration of six major air pollutants. CEEMD was used to capture the actual trend of the data and reduce the impact of noise on the prediction results, and the new optimization model combined the CS and DE models to optimize the initial weights and thresholds. To further verify the stability and accuracy of the model, a bias variance framework and DM test were used in an error evaluation system. 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.

- Air Quality Prediction: Big Data and Machine Learning Approaches January 2018 Gaganjot Kaur Kang, Jerry Zeyu Gao, Sen Chiao, Shengqiang Lu, and Gang Xie This paper reports recent literature study, reviews compares current research work on air quality evaluation based on big data analytics, machine learning models.With the advancement of IoT infrastructures, big data technologies, and machine learning techniques, real-time air quality monitor and evaluation is desirable for future smart cities. It also highlights some observations on future research issues, challenges, and needs.

- Air pollution monitoring and prediction using IoT 2018 Temesegan Walelign Ayele , Rutvik Mehta The proposed work on an air pollution monitoring and prediction system is enables us to monitor air quality with the help IoT devices. The system utilizes air sensors to detect and transmit this data to microcontroller. Then the microcontroller stores the data into the web server. For predicting the LSTM is implemented. It has a quick convergence and reduces the training cycles with a good accuracy.

- A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization February 2018 Dixian Zhu 1,*, Changjie Cai 2, Tianbao Yang 1 and Xun Zhou 3 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.

- A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran February 2019 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 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. The effective parameters for air pollution prediction have been determined in this research. This research used daily data of the pollutants. So, the quality of the proposed model can be significantly improved if hourly data is implemented .Considering the importance of air pollution problem, it is recommended that the number of air pollution measuring stations increases so as to allow for a better fit on the air pollution prediction.

- Deep Distributed Fusion Network for Air Quality Prediction 2019 Xiuwen Yi1,2, Junbo Zhang2,1,+, Zhaoyuan Wang1,2, Tianrui Li1, Yu Zheng2,1,3 They proposed a DNN-based approach to predict air quality. For this purpose they adopted a novel distributed fusion architecture to fuse heterogeneous urban data, which can simultaneously capture the individual and holistic effects from all influential factors affecting air quality. This approach achieves a higher accuracy in both general cases and sudden changes.

- A Dynamic Method to Estimate Source Emission Rate and Predict Contaminant Concentrations QU Hongquan , PANG Liping 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 of concentration, 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.

3 METHODOLOGY

In Figure 1: Methodology Flow Diagram, The 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:

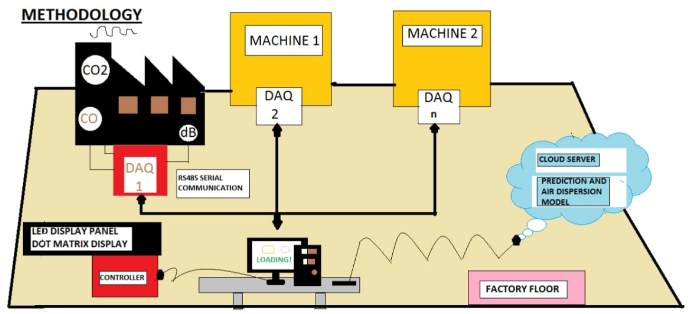

Figure 1: Methodology flow diagram

- IOT: The Iot Process Flow Begins With
  - Air Meteorological Parameters, Which Are To Be Measured Using Various Sensors. On The Basis Of Guidelines Laid By Central Pollution Control Board (Cpcb) ,It Is Divided Mainly Into Ambient Air Parameters And Stack Emission Parameters.
  - Ambient Air Parameters, Refers To Concentrations Of Pollutants In The Air, Typically Outdoor Air. The Criteria For This Is Based Upon Protection Of Human Health, Crops, Ecosystems As Well As For Planning And Other Useful Purposes.
  - Stack Emission Parameters, Parameters Required To Be Monitored In The Stack Emissions Using Continuous Emission Monitoring Systems, Are Industry Specific.
  - Consideration Of Parameters To Predict Q And V, We Considered All The Air Meteorological Parameters Inorder To Predict The Value Of V(Velocity Of Wind) And Q(Emission Rate).
  - Data Collection Using Sensors, With The Help Of Various Sensors Employed In A Specific Industry, Huge Amount Of Data Was Collected.
  - Building Up Of Cloud Architecture, The Data Collected From The Sensors Is Uploaded And Stored On The Cloud Setup. The Sensor Network Is Connected To The Cloud Server

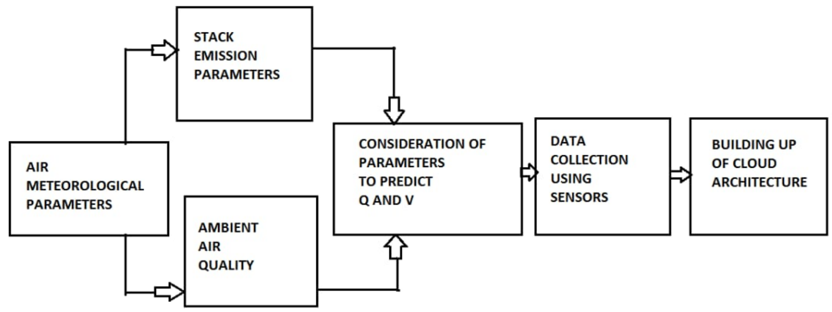

Figure 2: IOT process flow

- Machine Learning: The Machine Learning Process Flow Begins With
  - Generation Of Dummy Data In Python, Huge Amount Of Data, In Bulk Is Generated 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 Meteorological Factors, The Prediction Performance Is Greatly Improved And Large Sample Data Can Improve Performance And Train The Model Well.
  - Implementation Of Ml Algorithms, Machine Learning Algorithms Were Then 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.
  - 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 Algorithms, Various Algorithms Were Optimized 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.
  - Using Gaussian Model To Calculate Dispersion, The Extent Of Pollution Spread .We Chose Gaussian Dispersion Model Since It Was The Most Optimal Model In Terms Of Computing Power.

– Prediction Of Air Pollution, In This Manner The Entire Process Of Prediction Of Pollution And Calculation Of It's Spread Is Done.

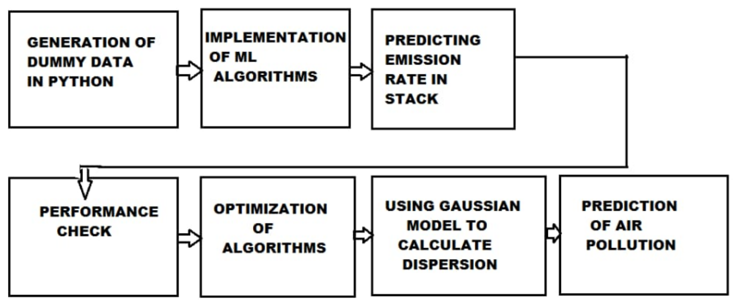

Figure 3: Machine Leaning process flow

3.1 MACHINE LEARNING ALGORITHMS

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

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

3.2 AIR DISPERSION MODEL

3.2.1 Gaussian Dispersion Model

In Gaussian dispersion model[3], 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 i.e. the concentration of the pollution will be maximum at the source and will gradually disperse where this dispersion follows a gaussian curve. The maximum concentration is in the direction of the wind and there will be lateral dispersion on the yz plane. The more the dispersion area, the more diluted will be the pollutant. This model has some basic assumptions that have to be taken into consideration. The emission should be in a steady state (emitted at a constant rate) The wind speed has to be constant Terrain is flat Pollutant is conservative (no gravity fallout) No ground absorption Turbulent diffusion in x-direction is neglected. (NOTE: you may remove the assumptions. And can also include a point on stack height...taller the stack the better as the dispersion area will be more)

3.2.2 Lagrangian Model

The lagrangian model determine the trajectory of air pollutants. The pollutants are tracked as air parcels which move along trajectories determined by wind field, the buoyancy and turbulence effects. The trajectory calculations are based on ordinary differential equations(ODE) instead of partial differential equations which are used in the gaussian dispersion model. The estimation of the concentration field is given by the final distribution of a large number of particles. The particles can be tracked from the source area to the area of reception. The computational cost of these model is independent of the output grid resolution and hence they are very efficient for short-range modelling. For long range simulations, however, the computational cost increases as there is a need to calculate a large number of single trajectories. Also a large amount of computational power is required to handle a large number of puffs.

3.2.3 Puff-Plume Model

The puff-plume model maintains the advantages of the puff models and also of the plume models. The pollutants particles are grouped in clusters. These clusters are treated as Gaussian puffs which are dispersed using the concept of relative diffusion. The center of mass of each puff follows a stochastic trajectory. The particle trajectories of Lagrangian stochastic dispersion model gives the trajectory of the puffs.

Why Gaussian model?

We chose the Gaussian Dispersion model for the following reasons: Amount of computation time is less The gaussian dispersion model has a relatively less computational time as compared to the Lagranian model in which the computational time increases significantly with increase in distance. Amount of resources required for computation is less and fast response time The gaussian model requires lesser amount of resources as compared to other models as they only calculate a single formula. Calculating using a single formula means that these systems have a fast response time. The calculation is fast even on common computers.

4 RESULTS

The prediction module and air dispersion module was implemented in python programing language using Spyder Integrated development environment.

The prediction module was tested on the dummy data set generated then the precision was measured by finding the deviation of the predicted value from the actual value using mean square error. In an attempt to optimize the models the parameters of the models were tuned and the response (mean squared error) was graphed out.

The air dispersion module was implemented to calculate the concentration from a stationary steady state point source i.e. stack and the results were mapped onto a grayscale bitmap image where the pixel color represents the concentration.

4.1 Machine Learning module

Different machine learning models were studied and implemented on the same dataset and the results were compared. The following table shows the range of error in the respective models while the parameters were tuned

Sr no.	Name of ML model	Parameters tuned	range of error
1	Random Forest	• maximum depth • maximum features • minimum sapmle split • minimum sample leaf • n estimators	5-24$\frac{m}{hr}^3$
2	Multi-linear perceptron regression	• activation function • number of neurons • solvers	2.2-2.7$\frac{m}{hr}^3$
3	Support vector regression	• kernel	10-10.8$\frac{m}{hr}^3$
4	K-nearest neighbor	• weight • algorithm	3-14$\frac{m}{hr}^3$
5	Multi-linear Regression	none	10$\frac{m}{hr}^3$

The following graphs shows the mean squre error while the parameters are being tuned for the respective models

- Random forest The random forest model was tuned on its parameters that are n_estimators, min_sample_leaf, max_depth, min_sample_split, max_features

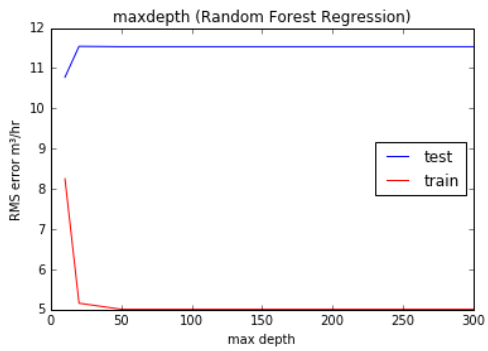

Figure 4: parameter tuned max depth

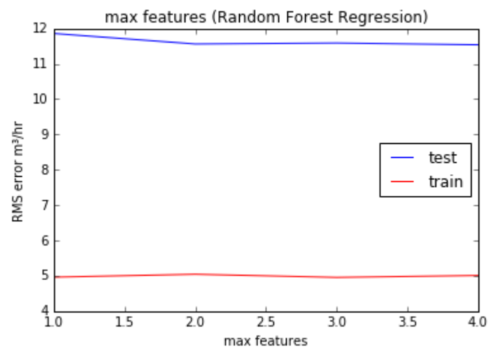

Figure 5: parameter tuned max features

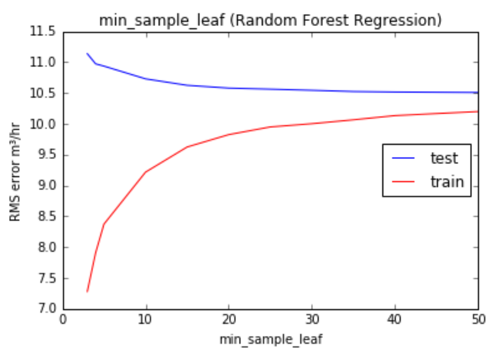

Figure 6: parameter tuned min sample leaf

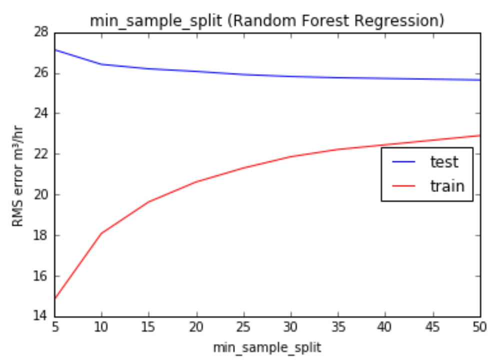

Figure 7: parameter tuned min sample split

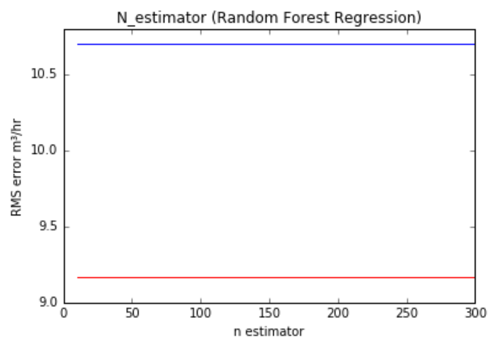

Figure 8: parameter tuned n estimators

- Multi-linear perceptron regression

Here the model is tuned on activation function, number of neurons and type of solvers

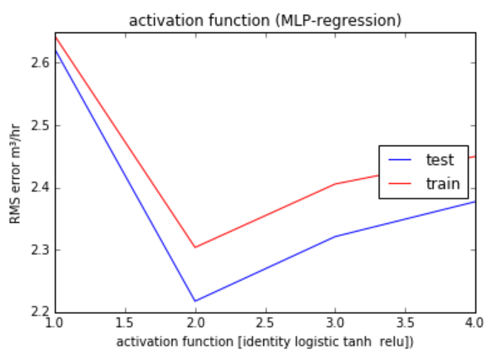

Figure 9: parameter tuned activation function

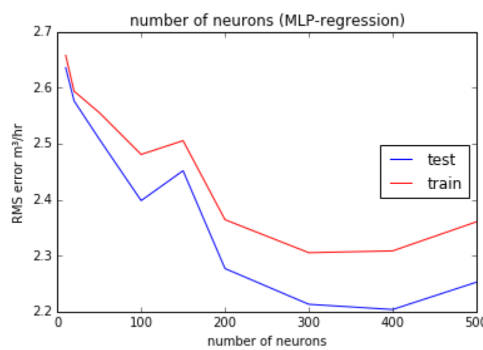

Figure 10: parameter tuned number of neurons

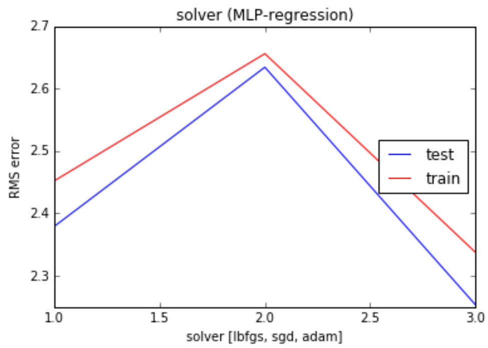

Figure 11: parameter tuned solver

- Support vector regression

This model is only tuned on its kernel property

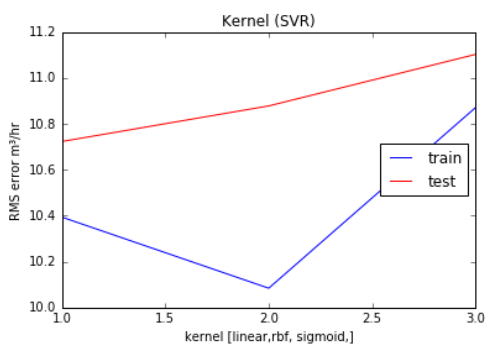

Figure 12: parameter tuned kernel

- K-nearest neighbor

This was tuned on its weight and algorithm parameter, which are used for clustering

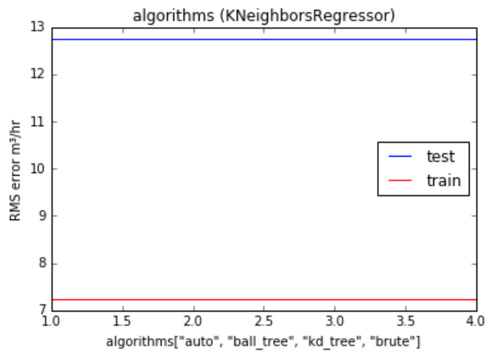

Figure 13: parameter tuned alogorithm

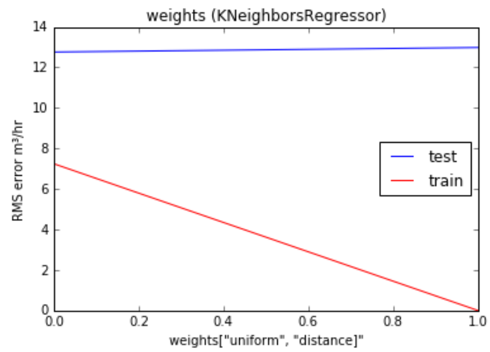

Figure 14: parameter tuned weights

- Multi-linear Regression

This algorithm is an extension of linear regression with added support of higher dimensions in feature space

Figure 15: predicted and actual data using samples form the test data

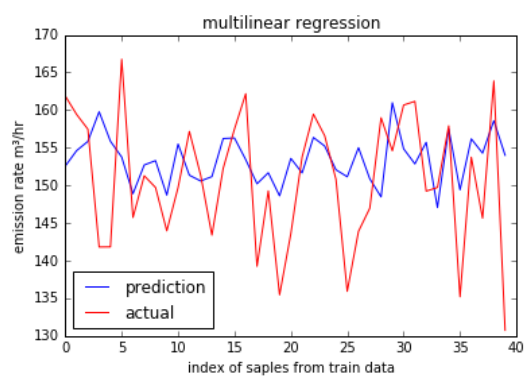

Figure 16: predicted and actual data using samples form the train data

4.2 Air Dispersion module

To calculate the concentration of the pollutant after emission from the stack, Gaussian air dispersion model was used for simulation, this was implemented using python programing language

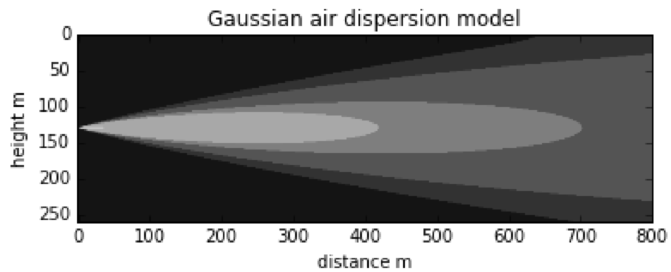

Figure 17: Gaussian air dispersion model

the white region represents regions with high concentration and black with low concentration

5 CONCLUSION

The results of the prediction module clearly shows that the multi-perceptron regression model has the least mean square error out of all the models also we can see that the error using test data loosely follows the error using training data i.e. the area between both plots over parameter tuning does not change much which is a good indication that overfitting is not occurring.

Whereas multi-linear regression has the most error due to its simplistic design

References

[1] R. Falkner, "The paris agreement and the new logic of international climate politics," ***International Affairs*****, vol. 92, no. 5, pp. 1107–1125, 2016.**

[2] C. Xiaojun, L. Xianpeng, and X. Peng, "Iot-based air pollution monitoring and forecasting system," in ***2015 International Conference on Computer and Computational Sciences (ICCCS)*****, pp. 257–260, IEEE, 2015.**

[3] K. Singh, "Air pollution modeling," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 4, no. 3, 2018.

[4] P. De Haan and M. W. Rotach, "A novel approach to atmospheric dispersion modelling: The puff-particle model," *Quarterly Journal of the Royal Meteorological Society*, vol. 124, no. 552, pp. 2771–2792, 1998.

[5]