ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 13, Issue, 01, pp.15557-15563, January, 2021

DOI: https://doi.org/10.24941/ijcr.40583.01.2021

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

PREDICTIVE ANALYSIS OF AIR POLLUTION MONITORING SYSTEM USING MACHINE LEARNING

[1]Akilandeswari, A, [2]Thenmozhi, S and [3]Jessilin Percis, T.

[1]Associate Professor, [23]Students, Department of Electronics and Communication Engineering, St. Joseph's Institute of Technology

ARTICLE INFO

Article History:
Received 18th October, 2020
Received in revised form
19th November, 2020
Accepted 17th December, 2020
Published online 30th January, 2021

Key Words:
AVR controller,
Air pollution sensors,
Wi-Fi module, Node JS,
Machine Learning.

ABSTRACT

In recent years, air pollution has acquired critical dimensions and the air quality in most Indian cities that monitor outdoor air pollution fail to meet WHO guidelines for safe levels. In this project we detect the air pollution by using sensor and here we use AVR controller and sensors which will monitor the air pollution. As well as we will also be able to monitor the data wirelessly with the help of a web application developed using Node JS. Enhancement in global concentrations of gasses of greenhouse like carbon dioxide, methane, nitrous oxide is also known as air pollution. We can purify the above statement by considering man-made emissions of hazardous chemicals as air pollution but this consideration has an adverse effect too. A report will be generated by web application for early prediction. Whenever any abnormal level of air pollution is determined then an automatic message will be sent to the person to take immediate actions. With the help of Wi-Fi Module we can able to store the data in cloud. Using machine learning technology, we can predict the air pollution at an earlier stage, the algorithm used in this project is linear regression. Thus, this project helps to early prediction of the air pollution effectively.

Citation: ***Akilandeswari, A., Thenmozhi, S. and Jessilin Percis, T.*** **2021.** "Predictive analysis of air pollution monitoring system using machine learning", *International Journal of Current Research*, 13, (01), 15557-15563.

INTRODUCTION

Earlier the air breathes used to be pure and fresh. But, due to increasing industrialization and concentration of poisonous gases in the environment the air is getting more and more toxic day by day. Air pollution has become a buzz word especially in urban areas due to its high concentration of pollutants. People exposed to this air pollution have been suffering from one or the other acute and chronic respiratory diseases and other health problems aggravated by this indirectly. Heart diseases and strokes are also common in countries where air pollutant levels are beyond the permissible limits. The risk of lung cancer is also increased. Thus air pollution monitoring would help in understanding the nature of pollutants present in the atmosphere, its sources of generation and impact on human health, plants and materials .Governments and citizens are looking for scientific intellect to challenge the common threat of pollution in its many procedures. Currently mobile apps are able to accomplish functions like reporting status of air quality, air quality forecasts, air quality monitoring in a particular area, and risks highlighting connected with threshold breaking quality, etc. Air plays a vital role not only in nature but also in the lives of human beings and all living organisms; around 18,000 distinct species need air.

Corresponding author: *Akilandeswari, A.,
Associate Professor, Students, Department of Electronics and Communication Engineering, St. Joseph's Institute of Technology.

Living organisms can survive without water for some time but cannot survive without air because when we breathe then we intake oxygen from air and releases carbon dioxide, if this procedure does not take place then it is difficult to survive. Well, now we have known the importance of air but due to activities of human beings' contaminants are added to air and this is known as air pollution. Atmosphere what we have today is very much distinct when compared with atmosphere before 1760. If a pure and natural atmosphere is considered to be clean then in the present condition, we cannot find natural atmosphere because chemical contents in atmosphere are enhanced. For example, content of carbon dioxide in natural atmosphere was 280ppm and now in current atmosphere it is 370.0^3ppm, likewise, methane was 0.750ppm and at present it is 1.77^4ppm, nitrous oxide was 0.270ppm and now it is 0.318^5ppm. To give a precise meaning of air pollution is a complex thing and can state that pollution began to include in air when human beings started burning fuels.

LITERATURE SURVEY

Ruma Ghosh, Julian W. Gardner, This paper describes the different sources of indoor and outdoor pollutants, reviews the current status of gas sensors, and discusses the role of new two dimensional (2-D) materials in detecting these hazardous gases at low power, i.e., close to the ambient temperature. Here, we review different synthesis techniques of 2-D materials and discuss the sensing performances of pristine and functionalized nano materials for some of the important pollutants such as NOx, NH3, SOx, CO, formaldehyde, toluene, and so on. Q. Han, P. Liu, H. Zhang and Z. Cai, we improve the network layout by employing the Zigbee network, which is combined with factory characteristics, and collected data on carbonic oxide, nitrogen dioxide, sulfur dioxide, ozone, particulate matter, temperature, and humidity. And then, to establish the dilution coefficient and diffusion coefficient of pollution diffusion, we adopt air movement as the energy model and, by utilizing the method of pollution traceability, achieve the complete coverage pollution monitoring of the whole city by local monitoring sites.

Finally, we propose an improved long short-term memory (LSTM) method to predict the pollution period of urban air quality. The experimental results show that the improved LSTM prediction model has strong applicability and high accuracy in the period prediction of pollution weather. Meanwhile, by analyzing the specific case in detail, we prove that air pollution in the city is mainly caused by the manufacturing industry S. Dhingra, R.B. Madda, A.H. Gandomi, R. Patan and M. Daneshmand, we propose a three-phase air pollution monitoring system. An IoT kit was prepared using gas sensors, Arduino integrated development environment (IDE), and a Wi-Fi module. This kit can be physically placed in various cities to monitoring air pollution. The gas sensors gather data from air and forward the data to the Arduino IDE. The Arduino IDE transmits the data to the cloud via the Wi-Fi module. We also developed an Android application termed IoT-Mobair, so that users can access relevant air quality data from the cloud. If a user is traveling to a destination, the pollution level of the entire route is predicted, and a warning is displayed if the pollution level is too high. The proposed system is analogous to Google traffic or the navigation application of Google Maps. Furthermore, air quality data can be used to predict future air quality index (AQI) levels.

Ahmed Boubrima, Walid Bechkit, and Hervé Rivano, we focus on using WSN for air pollution mapping and tackle the optimization problem of sensor deployment. Unlike most of the existing deployment approaches that are either generic or assume that sensors have a given detection range, we define an appropriate coverage formulation based on an interpolation formula that is adapted to the characteristics of air pollution sensing. We derive, from this formulation, two deployment models for air pollution mapping using the integer linear programming while ensuring the connectivity of the network and taking into account the sensing error of nodes. We analyze the theoretical complexity of our models and propose the heuristic algorithms based on the linear programming relaxation and binary search. we tackle the deployment issue of heterogeneous sensor networks and propose mixed integer programming models and heuristic algorithms taking into account the network deployment cost and the air pollution mapping quality while ensuring the network connectivity.

Our main contribution is to define an appropriate coverage formulation for air pollution regular mapping and then derive optimal deployment models and approximate resolution algorithms. We applied our models and algorithms on a dataset of the Lyon City, France and evaluated the computational complexity of our proposal.

EXSITING SYSTEM

A new type of outdoor air quality monitoring system is studied and preliminarily practiced and has proven certain feasibility and applicability. The main contributions of this paper are: first, we improve the network layout by employing the Zigbee network, which is combined with factory characteristics, and collected data on carbonic oxide, nitrogen dioxide, sulphur dioxide, ozone, particulate matter, temperature, and humidity. And then, to establish the dilution coefficient and diffusion coefficient of pollution diffusion, we adopt air movement as the energy model and, by utilizing the method of pollution traceability, achieve the complete coverage pollution monitoring of the whole city by local monitoring sites. Finally, we propose an improved long short-term memory (LSTM) method to predict the pollution period of urban air quality. The experimental results show that the improved LSTM prediction model has strong applicability in the period prediction of pollution weather. Meanwhile, by analyzing the specific case in detail, we prove that air pollution in the city is mainly caused by the manufacturing industry.

Disadvantages

- Only predict the pollution in the period of pollution weather not early.
- Project didn't contain cloud storage.
- Real time implementation is less.
- This methodology focuses on only industries.

PROPOSED SYSTEM

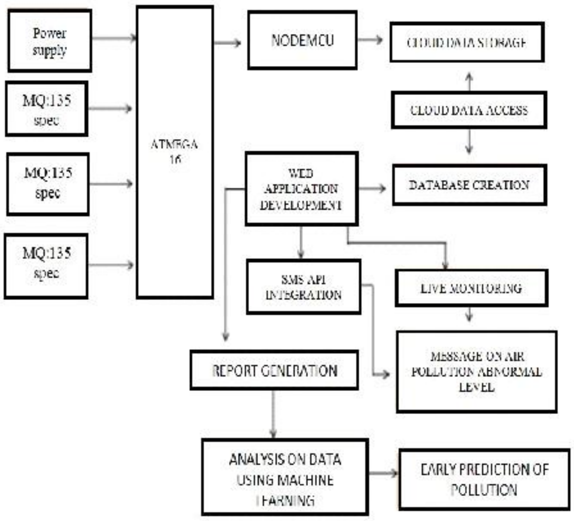
In recent years, air pollution has acquired critical dimensions and the air quality in most Indian cities that monitor outdoor air pollution fail to meet WHO guidelines for safe levels. In this project we detect the air pollution by using sensor and here we use AVR controller and sensors which will monitor the air pollution. As well as we will also be able to monitor the data wirelessly with the help of a web application developed using NodeJS. A report will be generate by web application for early prediction. Whenever any abnormal level of air pollution is determined then an automatic message will be sent to the person to take immediate actions. With the help of Wi-Fi Module we can able to store the data in cloud. Using machine learning technology, we can be able to predict the air pollution at an earlier stage. Thus, this project helps to early prediction of the air pollution effectively.

Advantages

- This project helps in monitoring and predicting air pollution at an earlier stage
- Cheapest method for monitoring the pollution in air.
- Avoid serious diseases spreading by air.
- Provides efficient way for providing best and fast solution of air pollution.
- Safe guard nature from air pollution.

BLOCK DIAGRAM

Sequences of learning experiences, also known as a curriculum, to cumulatively acquire new skills through self-guided exploration and social interaction with humans.



These robots use guidance mechanisms such as active learning, maturation, motor synergies, and imitation.

RANDOMFOREST REGRESSOR: In this project, we are going to predict the air pollution by using sensors. It will help to save the earth from air pollution. The air pollution sensor is used to sense the polluted air. Then after sensing those details by the controller and using Wi-Fi module the data is transmitted and stored in cloud using the MQTT protocol. The data in the cloud is transferred wirelessly to the web application monitoring. We can be able to lively monitor through web application if any abnormal level of air pollution is detected automatically. SMS will be sent to the person by SMS API Integration module in order to take immediate actions. AVR Controller is used to control this whole process based on the reading obtain an automatically report is generated which can be analysis for air pollution occurrence in coming future. This is achieved using machine learning implementation. Thus, the project helps in monitoring the air pollution and early prediction of air pollution effectively.

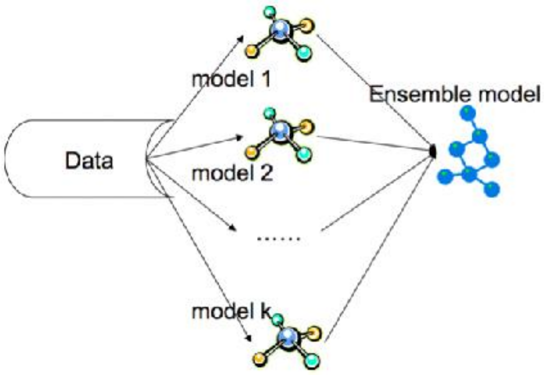
WORKING DOMAIN: Our project is based on the internet of things. The internet of things, or IoT, is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. The definition of the Internet of things has evolved due to the convergence of multiple technologies, real-time analytics, machine learning, commodity sensors, and embedded system. Traditional fields of embedded systems, wireless sensor networks, control systems, automation (including home and building automation) , and others all contribute to enabling the internet of things. In the consumer market, IoT technology is most synonymous with product pertaining to the concept of "Smart home", covering devices and appliances(such as lighting fixtures, thermostats, home security systems and cameras, and other home appliances) that support one or more common ecosystems, and can be controlled via devices associated with that ecosystem, such as smartphones and smart speakers.

MACHINE LEARNING ALORITHM

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers.

The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics. Machine learning tasks are classified into several broad categories. In supervised learning, the algorithm builds a mathematical model from a set of data that contains both the inputs and the desired outputs. For example, if the task were determining whether an image contained a certain object, the training data for a supervised learning algorithm would include images with and without that object (the input), and each image would have a label (the output) designating whether it contained the object. In special cases, the input may be only partially available, or restricted to special feedback Semi-supervised learning algorithms develop mathematical models from incomplete training data, where a portion of the sample input doesn't have labels. Classification algorithms and regression algorithms are types of supervised learning. Classification algorithms are used when the outputs are restricted to a limited set of values. For a classification algorithm that filters emails, the input would be an incoming email, and the output would be the name of the folder in which to file the email. For an algorithm that identifies spam emails, the output would be the prediction of either "spam" or "not spam", represented by the Boolean values true and false. Regression algorithms are named for their continuous outputs, meaning they may have any value within a range. Examples of a continuous value are the temperature, length, or price of an object.

In unsupervised learning, the algorithm builds a mathematical model from a set of data that contains only inputs and no desired output labels. Unsupervised learning algorithms are used to find structure in the data, like grouping or clustering of data points. Unsupervised learning can discover patterns in the data, and can group the inputs into categories, as in feature learning. Dimensionality reduction is the process of reducing the number of "features", or inputs, in a set of data. Active learning algorithms access the desired outputs (training labels) for a limited set of inputs based on a budget and optimize the choice of inputs for which it will acquire training labels. When used interactively, these can be presented to a human user for labeling. Reinforcement learning algorithms are given feedback in the form of positive or negative reinforcement in a dynamic environment and are used in autonomous vehicles or in learning to play a game against a human opponent. Other specialized algorithms in machine learning include topic modeling, where the computer program is given a set of natural language documents and finds other documents that cover similar topics. Machine learning algorithms can be used to find the unobservable probability density function in density estimation problems. Metal earning algorithms learn their own inductive bias based on previous experience. In developmental robotics, robot learning algorithms generate their own. The goal of the blog post is to equip beginners with the basics of the Random Forest algorithm so that they can build their first model easily. Ensemble methods are supervised learning models which combine the predictions of multiple smaller models to improve predictive power and generalization.



Ensemble Model: The smaller models that combine to make the ensemble model are referred to as base models. Ensemble methods often result in considerably higher performance than any of the individual base models.

Bagging: Several estimators are built independently on subsets of the data and their predictions are averaged. Typically, the combined estimator is usually better than any of the single base estimator. Bagging can reduce variance with little to no effect on biasex: Random Forests. The ensemble method we will be using today is called bagging, which is short for bootstrap aggregating. Bagging builds multiple base models with resampled training data with replacement. We train k base classifiers on k different samples of training data. Using random subsets of the data to train base models promotes more differences between the base models. We can use the Bagging Regressor class to form an ensemble of regressors. One such Bagging algorithms are random forest regressor. A random forest regressor is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement if bootstrap=True (default).Random Forest Regressors uses some kind of splitting criterion to measure the quality of a split. Supported criteria are "MSE" for the mean squared error, which is equal to variance reduction as feature selection criterion, and "Mean Absolute Error" for the mean absolute error.

Each decision stump will be built with the following criteria:

- A bootstrap will be created by randomly sampling the training set with replacement. The size of the bootstrap is set to equal the size of the training set.
- To form the root node or any node, the best split is determined by searching in a subset of randomly selected features of size sqrt(number of features). In our case, each decision stump is allowed to inspect two out of the four features.
- The maximum depth of the decision stump is one.
- Random Forest uses decision trees with a depth of one or greater. The term random stems from the fact that we randomly sample the training set, and since we have a collection of trees, it's natural to call it a forest — hence Random Forest. To build the root node or any node in the tree, a random subset of features is selected. For each of these selected features, the algorithm searches for the optimal cutting point to determine the split for the given feature. The feature from the randomly selected subset that produces the purest split is then used to create the root node. The tree is grown to a depth of one, and the same process is repeated for all other nodes in the tree, until the desired depth of the tree is reached. Finally, it's important to note that each tree is built separately using a different bootstrap, which introduces variation among the trees.

Extratrees Regressor: Extra Trees is like Random Forest, in that it builds multiple trees and splits nodes using random subsets of features, but with two key differences: it does not bootstrap observations (meaning it samples without replacement), and nodes are split on random splits, not best splits. So, in summary,

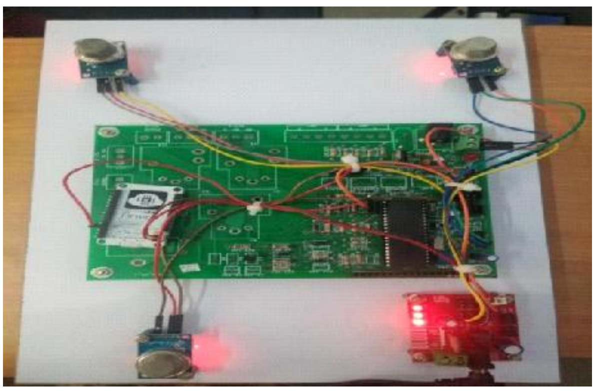
- builds multiple trees with bootstrap = False by default, which means it samples without replacement
- nodes are split based on random splits among a random subset of the features selected at every node

In Extra Trees, randomness doesn't come from bootstrapping of data, but rather comes from the random splits of all observations. The Extra Trees classifier — also known as Extremely Randomized Trees. To introduce more variation into the ensemble, we will change how we build trees. Each decision stump will be built with the following criteria:

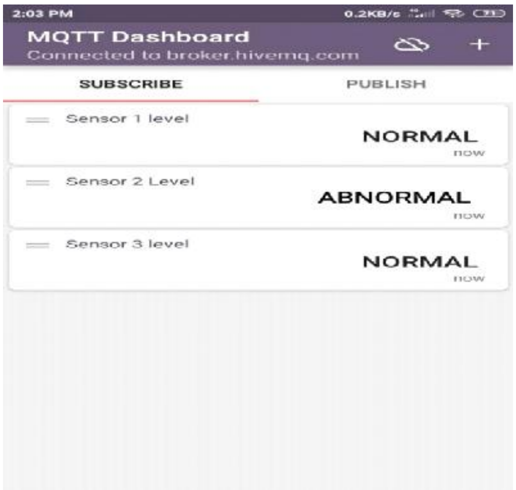
- All the data available in the training set is used to built each stump.
- To form the root node or any node, the best split is determined by searching in a subset of randomly selected features of size sqrt(number of features). The split of each selected feature is chosen at random.
- The maximum depth of the decision stump is one.

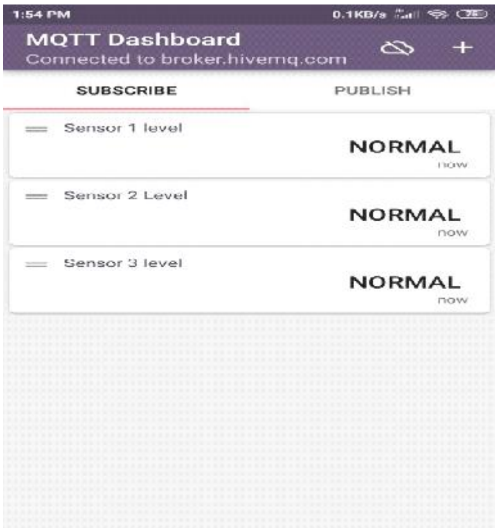
RESULTS AND CONCLUSION

HARDWARE OUTPUT: We have connected the three mq130 air quality sensor with the Atmega16 microcontroller .Voltage supply is connected to convert the 12v supply to 5v supply cause the micro controller needs only 5v or 3v for Working. The analog values from the sensor is sent as a input to the controller. Atmega 16 has a inbuilt analog to digital converter so the analog values from the sensor is converted to digital value. These values are send to the local server through the wifi module.



The output can also be seen through the mobile application "MQTT Dashboard". It works on the basis of MQTT protocol through the broken.hivemq.com:1883.





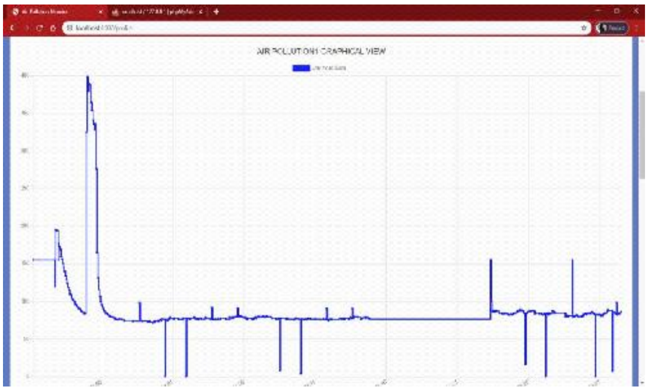
SOFTWARE OUTPUT

By using the HTML and CSS the webpage is created for the Front end. In the web page we can view the Front end for the Air pollution monitoring

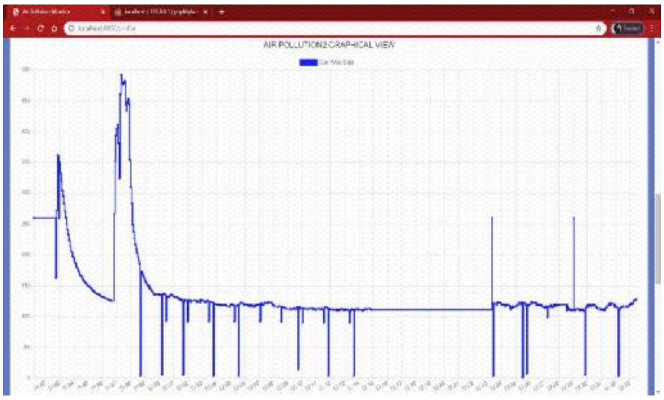
"http:localhost:4007"

Then the login page appears we can sign in as well as login in The user login data get stored at their First login Using the MYSQL in the XAMPP Panel Control. If u don't have a account we can create a new one by giving a username, phone number and a password.

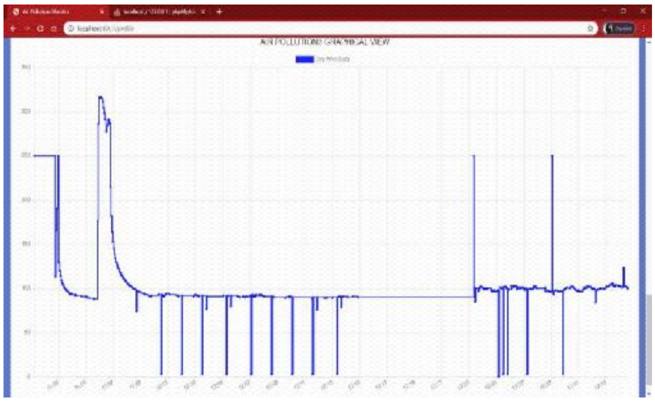
Webpage for air pollution monitoring: The air pollution is continuously monitored From the three location where the sensorsare kept .They are represented in a chart graphically.



From sensor 1

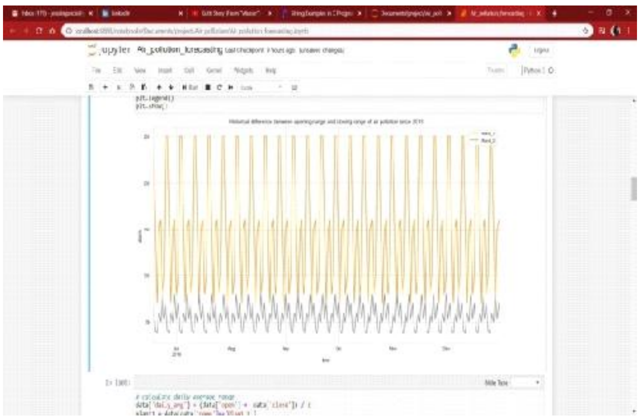


From sensor 2



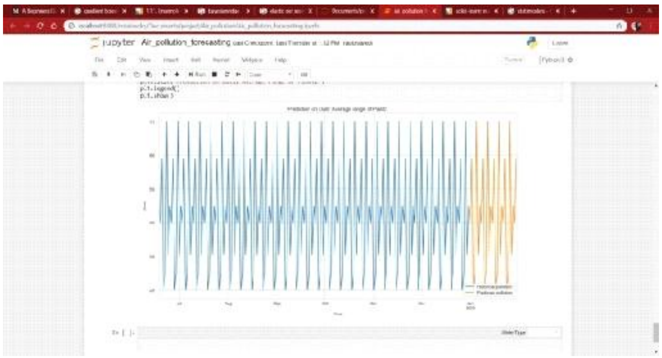
From sensor 3

PREDICTIVE ANALAYSIS: From the data set we got from the kit (Analog values) are trained by the modules . The average range from the start to the end for ever month we have collected the data for one year. Those data are processed and gets trained up.



From the previous data set

By using the randomforest regression and extratree Regression we can predict the air pollution for month.



Predictive analysis for one month: By using the random forest regression and extra tree Regression we can predict the air pollution for month.

SMS API INTERGRATION

A SMS API is well-defined software interface which enables code to send short messages via a SMS Gateway. As the infrastructures for SMS communications and the internet are mostly divided, SMS APIs are often used to 'bridge the gap' between telecommunications carrier networks and the wider web. SMS APIs are used to allow web applications to easily send and receive text messages through logic written for standard web frameworks. We will be using text local SMS API for our integration which enables us to easily integrate our SMS services with your website, software or CRM application in PHP, ASP,. NET, Java or any other language. The integrated solution of Text Local and Optimove makes it easy to plan and execute marketing text message campaigns, based on the advanced customer segmentation available in Optimove.

Optimove connects to the Text Local system for two purposes:

- To instruct Text Local to execute text campaigns, i.e., to send out particular text message templates to specified customer lists
- To receive delivery metrics from Text Local for reporting within Optimove

In order to set up the integration between your Optimove site and your Text Local account, the following steps are required:

Ensure that the following customer attributes are available in the Text Local Customer Contact Group:

- Customer ID – This field is used to match customers between Optimove and Text Local (note that you can also choose to use any alternative unique identifier that appears in both the Optimove and Text Local databases). The attribute used as the key identifier must be saved in one of the custom fields provided in TextLocal (Custom 1, Custom 2 or Custom 3).
- First name
- Last name
- Phone number
- Maintain a daily update of this database. For example, when new customers join, their information must be promptly updated in Text Local.
- Define the Sender name for text messages sent by Optimove in the Text Local interface. Note: If you would prefer that campaigns sent by Optimove use a different sender name, the Optimove Integration Team can configure a different sender name for the Text Local campaigns executed by Optimove.
- Provide the Optimove Integration Team with the name and ID of the particular Text Local Contact Group to which to connect
- Provide the Optimove Integration Team with login credentials to the Text Local API - notes:
- You should create a new administrator-level Text Local API account specifically for use by Optimove.
- These credentials may not contain any of these characters: $ & =

Provide the Integration Team with the following details

- Text Local API key
- Username
- Password
- API login URL (e.g.,https://api.txtlocal.com)

Conclusion

This project is used to find the presence of air pollution in the surrounding environment sensor can sense NH3, NOx, alcohol, Benzene, smoke, CO2 and some other gases. Then measure the temperature and humidity which can also be lively monitored for sudden changes due to pollution increase. So, we reduce the time required for manual classification and eliminates the human error rate.

Future Work

In the coming future, we review the application of the air pollution technology in the environmental field and it can promote for detecting various type of air pollution with more accuracy. In this field there are more chance to develop or convert this project in many ways. Thus this provideearly prediction of air pollution as well as eliminates the human error rate.

REFERENCES

[1] Ahmed Boubrima, Walid Bechkit, and HervéRivano,”On the Deployment of Wireless Sensor Networks for Air Quality Mapping: Optimization Models and Algorithms”,IEEE/ACM Transactions On Networking, Vol. 27, No. 4, August 2019.

[2] Ahmed Boubrima, Walid Bechkit and Hervé Rivano, ”Optimal WSN Deployment Models for Air Pollution Monitoring”, IEEE Transactions on Wireless Communications, Vol.1536-1276,2016.

[3] Dhingra, S. R. B. Madda, A. H. Gandomi, R. Patan and M. Daneshmand, "Internet of Things Mobile–Air Pollution Monitoring System (IoT-Mobair)," in IEEE Internet of Things Journal, vol. 6, no. 3, pp. 5577-5584, June 2019.

[4] Han, Q. P. Liu, H. Zhang and Z. Cai, "A Wireless Sensor Network for Monitoring Environmental Quality in the

Manufacturing Industry," in IEEE Access, vol. 7, pp. 78108-78119, 2019.

[5] Ke Gu, JunfeiQiao, Member, IEEE, and Weisi Lin, Fellow, IEEE," Recurrent Air Quality Predictor Based on Meteorology- and Pollution-Related Factors",IEEE Transactions on Industrial Informatics,Vol.1551-3203,2017.

[6]Luo X. and J. Yang, "Problems and challenges in water pollution monitoring and water pollution source localization using sensor networks," 2017 Chinese Automation Congress (CAC), Jinan, 2017, pp. 5834-5838.

[7]Ruma Ghosh, Julian W. Gardner , Fellow, IEEE, and Prasanta Kumar Guha Member, IEEE," Air Pollution Monitoring Using Near Room Temperature Resistive Gas Sensors: A Review", IEEE Transactions On Electron Devices, Vol. 66, No. 8, August 2019.

[8]Sun, C. Y. Yu, V. O. K. Li and J. C. K. Lam, "Optimal Multi-type Sensor Placements in Gaussian Spatial Fields for Environmental Monitoring," 2018 IEEE access (ISC2), Kansas City, MO, USA, 2018, pp. 1-8.

[9]Venkatanarayanan, A. A. Vijayavel, A. Rajagopal and P. Nagaradjane, "Design of sensor system for air pollution and human vital monitoring for connected cyclists," in IET Communications, vol. 13, no. 19, pp. 3181-3186, 3 12 2019.

[10]Zikun Deng, Di Weng, Jiahui Chen, Ren Liu, Zhibin Wang,JieBao, Yu Zheng, and Yingcai Wu," AirVis: Visual Analytics of Air Pollution Propagation AirVis: Visual Analytics of Air Pollution Propagation", IEEE TransactionsonVisualizationand Computer Graphics,Vol.1077-2626, 2019

****