IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Lockdown effect on Air Quality Index Prediction based on Machine Learning

Shailesh Munge[1],
Department of Computer Engineering,
Wagholi, Pune

Sagar Kharche[1],
Department of Computer Engineering,
Wagholi, Pune

Piyush Joshi[1]
Department of Computer Engineering
Wagholi, Pune

Kirti Bathe[1]
Department of Computer Engineering
Wagholi, Pune

Prof. Santosh Waghmode
Department of Computer Engineering
Wagholi, Pune

Abstract: Due to human activities, industrial enterprise, and urbanization air is obtaining contaminated? The most important air pollutants square measure CO, NO, C6H6, etc. The concentration of air pollutants in close air is ruled by meteorologic parameters like part wind speed, wind direction, relative humidity, and temperature. Earlier techniques like chance, Statistics, etc. were wont to predict the standard of air, however, those strategies square measure terribly complex to predict, Machine Learning (ML) is that the higher approach to predict the air quality. With the requirement to predict air ratio by considering varied parameters like CO, Tin oxide, non-metallic hydrocarbons, Benzene, Titanium, NO, Tungsten, atomic number 49 compound, Temperature, etc., the approach uses regression (LR), Support Vector Machine (SVM), call Tree (DT), Random Forest technique (RF) to predict the Relative humidity of air and uses Root Mean sq. Error to predict the accuracy

Keywords: **Air Pollution, Decision Tree, Linear Regression, Machine Learning, Random Forest, Supervised Learning, SVM.**

INTRODUCTION

Air pollution which is detrimental to people's health is a widespread problem across many countries around the world. With the development of the economy and society all over the world, most metropolitan cities are experiencing elevated concentrations of ground-level air pollutants, especially in fast-developing countries like India and China. Exposure to air pollution can affect everyone, but it can be particularly harmful to people with heart disease or a lung condition both short and long-term exposure to air pollutants has been associated with health impacts. More severe impacts affect people building an early warning system, which provides precise forecasts and also alerts health alarms to local inhabitants will provide valuable information to protect humans from damage by air pollution. The combined effects of ambient (outdoor) and household air pollution cause about 7million premature deaths every year. This research aims to predict the level of Air Pollution with a set of data used to make predictions. Through them and to obtain the best prediction using several models and compare them to find the Appropriate solutions.To develop robust application using Machine learning algorithm and different techniques using large datasets and find out the optimum solution for Air Quality that helps the human being and It is used to predict the future concentrations of air pollutants in accordance with methodological variable. The major pollutants area unit oxide (NO), monoxide (CO), stuff (PM), SO2 etc. monoxide is made thanks to the deficient Oxidization of propellant like rock oil, gas, etc. Nitrogen Oxide is made thanks to the ignition of thermal fuel; Carbon monoxide causes headaches, vomiting; aromatic hydrocarbon is made due to smoking, it causes metabolic process problems; gas oxides cause vertigo, nausea; stuff with a diameter of 2.5 micrometres or but that affects additional to human health. Measures should be taken to reduce air pollution within the atmosphere. Air Quality Index (AQI), is used to measure the standard of air. Earlier classical ways like probability, statistics were accustomed predict the standard of air, but those ways area units terribly complicated to predict the standard of air. Due to the advancement of technology, currently, it's terribly straightforward to fetch the data regarding the pollutants of air exploitation sensors. Assessment
of data to notice the pollutants wants vigorous analysis. Convolution Neural networks, algorithmic Neural networks, Deep Learning, Machine learning algorithms assures in accomplishing the prediction of future AQI in order that measures can be taken befittingly. Machine learning that comes under computing has 3 sorts of learning algorithms, they're the supervised Learning, unsupervised learning, Reinforcement learning. within the projected work we tend to have used the supervised learning approach. There area

unit several algorithms underneath supervised learning algorithms like Linear Regression, Nearest Neighbor, SVM, kernel SVM, Naïve mathematician and Random Forest. Compared to all or any different algorithms Random forest provides higher results, so our the approach selects Random Forest to predict the correct air pollution.

II- LITERATURE SURVEY

[1] Yuelai Su2020 Prediction of air quality based on Gradient Boosting Machine Method In this Paper work on Air Quality Using Gradient Boosting Method of ML and limitation of this paper has Light GBM does not split the eigenvalues one by one, so it is necessary to calculate the splitting benefits of each eigenvalue [02] Yu Jiao1 , Zhifeng Wang1*, Yang Zhang1 Prediction of Air Quality Index Based on LSTM in this paper focus on air quality Using LSTM Method and limitation of this system are is predicted by adding the highest temperature, the lowest temperature and the wind[3] Shajulin Benedict Revenue Oriented Air Quality Prediction Micro Services for Smart Cities in this research paper focus on Using simple Random Forest Algorithm and limitation of this system are Not handled imbalanced data and error prone into the data Under fitting problem[04] Wang Zhenghua1 , Tian Zhihui Prediction of air quality index based on improved neural network in this research work mainly focus on the air quality using Neural Network and disadvantages of this system are Feature are not well handled and use to complex model on small set of datasets. Ishan et.al [5] described the benefits of the Bidirectional Long - Short Memory[BiLSTM] method to forecast the severity of air pollution. The proposed technique achieved better prediction which models the long term, short term, and critical consequence of PM2.5 severity levels. In the proposed method prediction is made at 6h, 12h, 24h. The results obtained for 12h is consistent, but the result obtained for 6h, and 24h are not consistent. Chao Zhang et.al [6] proposed web service methodology to predict air quality. They provided service to the mobile device, the user to send photos of air pollution. The proposed method includes 2 modules a) GPS location data to retrieve the assessment of the quality of the air from nearby air quality stations. b) they have applied dictionary learning and convolution neural network on the photos uploaded by the user to predict the air quality. The proposed methodology has less error rate compared to other algorithms such as PAPLE, DL, PCALL but this method has a disadvantage in learning stability due to this the results are less accurate.

III – BACKGROUND

Machine learning is used in various applications to find out the best solution of real world problem Machine learning algorithm learn without being explicitly programmed. In the machine learning three types of machine learning algorithms are used in various application.

1. Supervised Machine Learning algorithm
2. Unsupervised Machine Learning
3. Reinforcement Machine Leaning

1. Linear Regression:

Linear Regression is used to predict the real values using continuous variables. It is used in many areas such as Economics, Finance, Healthcare, etc.

Assumption in Linear Regression:

There are four assumption are required to execute the linear regression or find out the relationship between one or more independent and dependent variable

1. Homogeneity of variance
2. Independence
3. Linearity
4. Normality

2. Support Vector Machine:

SVM is a SL algorithm in which it divides the plane into 2 parts by drawing a line between the 2 different classes. The line which separates the plane into different parts is called hyperplane. It always gives a perpendicular distance from the data point to the line of separation. It can do both linear and nonlinear classification. It is mainly used to do the classification and regression.

3. Decision Tree

Decision Tree is one of the supervised learning algorithms which it is used to represent the decision that is made based on the condition. It is used for both classification and regression. The Decision tree is always constructed from top to bottom. The first node from the top is called as root node. The last nodes are called as a leaf node. Internal nodes are present in between the root node and leaf nodes. Based on some condition the internal nodes are split and finally, the decisions are made. In the real time as the number of variables increases tree grows larger and algorithm becomes complex. In Decision tree we have two types; they are classification and regression trees. Classification tree is used to classify the dataset, so that it is easy to analyze the data. But using this algorithm we cannot make a prediction. The Regression tree is a tree mainly used to predict continuous values. Growth of tree depends on factors like: • the attribute which is chosen to make a prediction; • condition used for the split the tree; • deciding when to stop or terminate the growth of the tree;

4. Randorm Forest:

Random Forest It is defined as a set of decision trees to do regression and classification. Classification is used to find out the majority voting. Regression is used to calculate the mean value. This algorithm is more accurate, robust, and can handle a variety of data such as binary data, categorical data, and continuous data. Random Forest is nothing but multiple decision trees. 75% of the dataset is considered for the training. The training data is subjected to sampling and based on attribute sampling different decision trees are constructed by applying the Random Forest.

IV- PROPOSED SYSTEM

The Air pollutants data is retrieved from the sensors which square measure processed during a unified schema and hold on as a dataset. This dataset is preprocessed with completely different functionalities like standardization, attribute choice, and discretization. Once the dataset is prepared, it's split into a coaching dataset and check dataset. And any supervised Machine Learning Algorithms square measure applied on the coaching dataset. The obtained results square measure matched with the testing dataset and results square measure analyzed. Fig. one describes the design of the proposed model. The pollution prediction mistreatment supervised Machine Learning approach considers four machine learning algorithms like LR, SVM, DT, and RF

1. System Architecture:

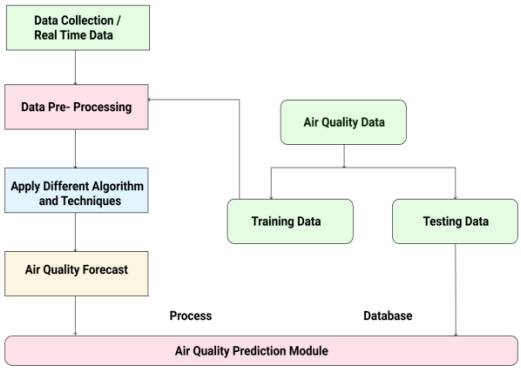

Figure: architecture of Air Quality Prediction

V- CONCLUSION

This project aims to develop a robust module for predicting of Pollution and Risk prediction on Air Quality. The features used for prediction are considered.. The prediction model has been built for prediction of Air Quality with a maximum accuracy. Few highly correlated features are used for analyzing and prediction of risk factor and calculating Air Quality using Machine Learning Algorithm and Techniques

VI- BIBLIOGRAPHY

.

[1]Verma, Ishan, Rahul Ahuja, HardikMeisheri, andLipikaDey. ”Air pollutant severity rediction using Bi-directional LSTM Network.” In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 651-654. IEEE, 2018.

[2] Figures Zhang, Chao, Baoxian Liu, Junchi Yan, Jinghai Yan, Lingjun Li, Dawei Zhang, XiaoguangRui, and RongfangBie. ”Hybrid Measurement of Air Quality as a 5 Fig. 8. RH w.r.t tin oxide Fig. 9. RH w.r.t C6H6 Mobile Service: An Image Based Approach.” In 2017 IEEE International Conference on Web Services (ICWS), pp. 853-856. IEEE,2017.

[3] Yang, Ruijun, Feng Yan, and Nan Zhao. ”Urban air quality based on Bayesian network.” In 2017 IEEE 9th Fig. 10. RH w.r.t NO Fig. 11. RH w.r.t NO2 International Conference on Communication Softwareand Networks (ICCSN), pp. 1003-1006. IEEE,2017.

[4] Ayele, TemeseganWalelign, and RutvikMehta.”Air pollution monitoring and prediction using IoT.” In 2018 Second International Conference on Inventive Communication 6 Fig. 12. RH w.r.t Temperature Fig. 13. RH w.r.t CO and Computational Technologies (ICICCT), pp. 1741-1745. IEEE,2018.

[5] Djebbri, Nadjet, and MouniraRouainia. ”Artificial neural networksbased air pollution monitoring inindustrial sites.” In 2017 International Conference on Engineering and Technology (ICET), pp. 1-5. IEEE,2017.

[6] Kumar, Dinesh. ”Evolving Differential evolution method with random forest for prediction of Air Pollution.” Procedia computer science 132 (2018): 824-833.

[7] Jiang, Ningbo, and Matthew L. Riley. ”Exploring the utility of the random forest method for forecasting ozone pollution in SYDNEY.” Journal of Environment Protection and Sustainable Development 1.5 (2015): 245-254.

[8] Svetnik, Vladimir, et al. ”Random forest: a classification and regression tool for compound classification and QSAR modeling.” Journal of chemical information and computer sciences 43.6 (2003): 1947-1958.

[9] Biau, GA˜ Srard. ”Analysis of a random forest model.” ˇJournal of Machine Learning Research 13.Apr (2012): 1063- 1095.

[10] Biau, Gerard, and ErwanScornet. ”A random forest ´ guided tour.” Test 25.2 (2016): 197-227.

[11] Grimm, Rosina, et al. ”Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis.” Geoderma 146.1- 2 (2008): 102-113.

[12] Strobl, Carolin, et al. ”Conditional variable importance for random forests.” BMC bioinformatics 9.1 (2008): 307.

[13] Svetnik, Vladimir, et al. ”Random forest: a classification and regression tool for compound classification and QSAR modeling.” Journal of chemical information and computer sciences 43.6 (2003): 1947-1958.

[14] Verikas, Antanas, AdasGelzinis, and MarijaBacauskiene. ”Mining data with random forests: A survey and results of new tests.” Pattern recognition 44.2 (2011): 330-349.

[15] Ramasamy Jayamurugan,1 B. Kumaravel,1 S. Palanivelraja,1 and M.P.Chockalingam2 International Journal of Atmospheric Sciences Volume 2013, Article ID 264046, 7 pages http://dx.doi.org/10.1155/2013/264046