Vol. 06, No. 2 (2024) 505-512, doi: 10.24874/PES06.02.008

**Proceedings on Engineering Sciences**

www.pesjournal.net

# REAL TIME AIR QUALITY SURVEILLANCE & FORECASTING SYSTEM (RTAQSFS) IN PUNE CITY USING MACHINE LEARNING-BASED PREDICTIVE MODEL

Tanuja Satish Dhope[1]
Ahmed Shaikh
Dina Simunic
Prashant Pandurang Patil
Kishor S. Wagh
Sharmila K. Wagh

Received 25.09.2023.
Received in revised form 03.11.2023.
Accepted 12.12.2023.
UDC – 502.3:613.15

Keywords:

*Air Quality, Machine Learning, Random Forest. Ridge, Linear Regression, IOT, Sensors*

A B S T R A C T

*Increasing urbanisation and industrialisation produce major environmental challenges such as air pollution, which endangers human, animal, and vegetation life. Reliable measurement, monitoring, and prediction of Air Quality (AQ) have emerged as key global concerns. The State Government-Municipal Corporation is working on policy reforms to fight the deterioration of air quality in Pune and other Indian cities. In this paper, Real Time Air Quality Surveillance & Forecasting System (RTAQSFS) has been developed, which work in the cascaded model incorporating electronics hardware as well as machine learning algorithms. The presence of air pollutants is measured using sensors like MQ135, MQ7, MQ131 etc. The performance of machine learning algorithms viz. Linear regression, Ridge regression, Lasso regression, Decision tree and Random Forest has been evaluated wrt. RMSE and R2. The experimental results show that Random Forest outperforms the other algorithm providing less RMSE and R2 as 99.9% for all the parameters.*



## 1. INTRODUCTION

Air pollution is frequently caused by anthropogenic activities and natural phenomena like disasters and forest fires and its levels remain dangerously high in many parts of the world. Numerous health implications of air pollution. The elderly and young children are most impacted by it. The health impacts include respiratory and cardiac disorders such as asthma, pneumonia and lung cancer. Other devastating repercussions include acid rain, eutrophication, ozone layer thinning, and global warming. Nine out of ten individuals breathe air with high levels of pollution, according to recent WHO data (Annual report, 2018) alarming 7 million people per year die from ambient (outdoor) and residential air pollution, according to updated estimates It is also growing with the increase in traffic especially in big cities. In India nearly 12.4 lakhs people lost their life due to air pollution in 2017(Annual report 2017).The survey indicates that in globe around

[1] Corresponding author: Tanuja Satish Dhope
Email: tanuja_dhope@yahoo.com

18% people loses their life due to air pollution related diseases. In India this is 26% (Annual report, 2017). As shown in Figure 1, the Air Quality (AQ) in Pune city, Maharashtra, India is also deteriorating day by day especially in Sub-Urban regions like Shivaji Nagar, Kothrud, Mandai, Hadapsar. The environment status report (Environment status report, 2017) 2016–17 of the Pune Municipal Corporation (PMC), makes it abundantly evident that air pollutants like nitrogen oxide, nitrogen dioxide, sulphur dioxide, carbon monoxide, and particle matter are constantly rising, Hadapsar and Navipeth are two of the most polluted areas in the city. The State Government-Municipal Corporation is working on policy reforms to fight the deterioration of air quality in Pune and other Indian cities. The U.S. Environment Protection Agency (EPA) established the Air Quality Indicator (AQI) to determine the regional air quality representing pollutant allowable levels that have been inflated by time, place, and other factors (Technical assistance document, 2018).In paper (Gulia et al,202), Respirable Suspended Particulate Matter (RSPM) has been analyzed for Delhi, India from 2003 to 2019,which shows the annual average increment in the range of 0.98–3.19% and for NO2 ,it is of 5.21–6.07%. This alarming condition needs to address by developing a systematic mechanism to monitor/assess the improvement in air quality (pollutant wise) through implemented control strategies.

With technological advancements such as artificial intelligence and machine learning, AQ categorization and prediction can be more efficiently monitored, anticipated, and required measures can be done if AQ exceeds the established levels. In this field, a lot of research has been done. Prior to now, many pollutants were analyzed and anticipated using statistical and cutting-edge approaches. In this paper, our research work is focused on Hadapsar, Pune, Maharashtra, India.

The objectives are:

- Choosing the best statistical model for air-pollution prediction.
- Design and Development of Real Time Air Quality Surveillance & Forecasting System (RTAQSFS)
- Evaluation of empirical analysis using a real-time dataset for Hadapsar, Pune, and comparison to baseline techniques
- Assessment of correlation among air pollutants.
- Addressing the most important factors in air pollution forecast on an hourly basis.

The organization of paper is as follows. Literature survey is discussed in section 2, section 3 deals with our proposed RTAQSFS, section 4 elaborates machine learning (ML) algorithms. The results using RTAQSFS have been discussed in section 5 which is followed by conclusions in section 6.

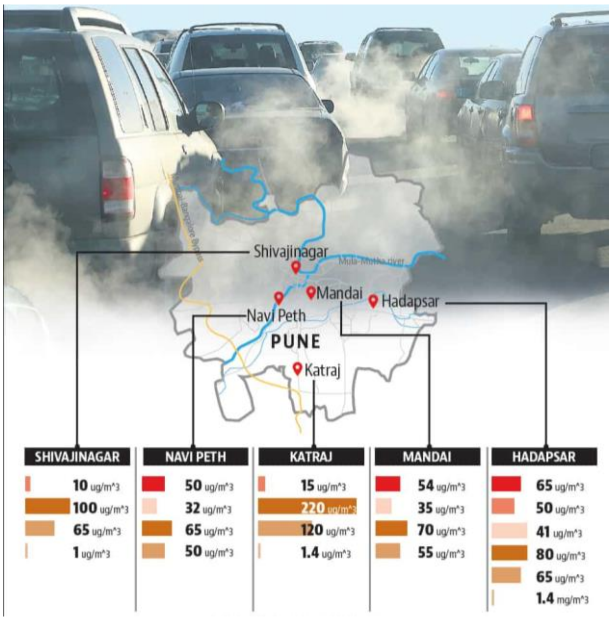


**Figure 1.** Air pollution (Hindustan Times 2017)

## 2. LITERATURE SURVEY

Five distinct machine learning methods, including random forest, multilayer perceptron, K-Nearest Neighbour, SVR, and multi-linear regression, were deployed for the air dispersion model and compared. A set of 1000 dummy points was processed by a professional. Multi-layer Perceptron Algorithm provides the lowest mean squared error (MSE) as 2.1when compared to the other implemented algorithm viz. K-Nearest Neighbour, SVR, Random Forest and Multi-Linear Regression having 3, 9 and 10 respectively (Simu et al, 2020). Various methods such as SVM, linear regression, XGboost, and others, have been examined in order to accurately estimate the incoming pollution level. The neural network and boosting model outperform all other algorithms (Madan et al, 2020).

(Bisht et al, 2018) suggested Extreme Learning Machine (ELM) based system predicts better than existing air pollution prediction systems. (Bekkar et al, 2021) designed a deep learning method utilizing CNN-LSTM with a spatial-temporal feature to predict the hourly forecast of PM 2.5 in Beijing, China, by combining historical data of pollutants, meteorological data, and PM 2.5 in surrounding stations. The experimental results show that "hybrid CNN-LSTM multivariate" strategy produces better accurate predictions than traditional LSTM, Bi-LSTM, GRU, Bi-GRU, CNN.In (Vong et al, 2012),metrological and pollutant data were collected daily at Macau monitoring stations and experimented with SVM different kernels. The prediction results of the SVM-Linear model and the SVM-RBF model demonstrated a relative good fit to the observed test set of over a year of data, particularly for SO2 and NO2. SVM-Linear and SVM-RBF models outperformed other SVM-polynomial, sigmoid, and wavelet tested models in seasonal experiments, with some lagging and underestimate of these two models occurring in the winter experiment.

In (Guo et al, 2021) ,the quarterly compound accumulation grey model QCGM(1, 1) model is proposed to investigate 22 Chinese towns with poor air quality for SO2,NO2 ,PM2.5 and PM10 . The grey model's (GM) accumulation is optimised by incorporating two parameter variables to regulate the accumulation sequence. A seasonal element is incorporated into the model to address issue of seasonal variations of pollutants. When compared to the traditional GM(1, 1) model, the model's initial value is effective, and the fitting accuracy is substantially improved.

The findings of (Castelli et al, 2020) show that SVR with RBF kernel allows author to reliably predict hourly pollutant concentrations such as CO, SO2, NOx, ground-level ozone, and PM2.5 as well as the hourly AQI for the state of California. It is shown in (Dun et al, 2020) that the hybrid model (FGM (0,m)-SVR) performs better at forecasting in Shijiazhuang and Chongqing cities. The three air pollutants (PM10, PM2.5, and NO2) are forecasted using the hybrid model, and it is shown that the hybrid model's prediction accuracy is significantly higher than that of the single model. (Chen et al, 2020) describes an intelligent environmental pollution monitoring system based on a novel semiconductor gas sensor. The system combines data collection, processing, and application into one, enabling real-time monitoring, data storage, and equipment control of polluted gas. The measurement findings demonstrate that the system can successfully measure CO gas at concentrations ranging from 0 to 50 ppm, with a relative measurement error that can essentially be kept at less than 2%, thereby achieving the system's performance index. The AQ monitoring and prediction in Skopje consists of 18 monitoring stations for 7 pollutants reveals that PM10 and PM2.5 are the most hazardous air pollutants in Skopje ( Mijakovski et al,2020).

The portable air pollution monitoring system based on a chemical sensor array has been developed in (Chen et al, 2020). Internet of Thing (IoT) has been utilized in (Jiyal et al, 2020) for monitoring and prediction of air pollutant. People can adjust their travel path by knowing the pollution level of a populated place using an android app on their cell phone. Further, alarm systems also sound, and action can be done to reduce air pollution levels. In (Parmar et al, 2017 ), authors analysed up to seven air contaminants in Skopje.

An analysis of the pollutants' inventory suggests that suspended particles (PM10 and PM2.5) have been the most serious air pollutants. Air quality for Colaba region of Mumbai, India using AR, FFNN, univariate LSTM, multivariate LSTM and Bidirectional LSTM discussed in (Vibhute et al, 2020). Bidirectional LSTM model for PM2.5 perform well comparatively the AR, FFNN and univariate LSTM and multivariate LSTM with RMSE 19.54 & R2 value 0.66. (Xiaojun et al, 2015) performs reduction in hardware cost by 1:10 using IoT. (Liang et al, 2020) shows the result for R2 for stacking ensemble and AdaBoost, where stacking ensemble gives best RMSE and AdaBoost gives the good MAE result. SVM gives the worst result except for 1-hr prediction.

## 3. REAL TIME AIR QUALITY SURVEILLANCE & FORECASTING SYSTEM (RTAQSFS)

As shown in figure 2, RTAQSFS consists of power supply circuit with battery backup, sensor, ESP8266/ESP32 controller with Wi-Fi system on chip for monitoring & recording of the environmental parameters, 744067 demultiplexer and various sensors attached to IO ports of controller. RTAQSFS contains ESP8266 for monitoring & recording of the environmental parameters, gas sensors like MQ131, MQ135, MQ7 and GP2Y1010AU0F and environment parameter sensor DHT22.

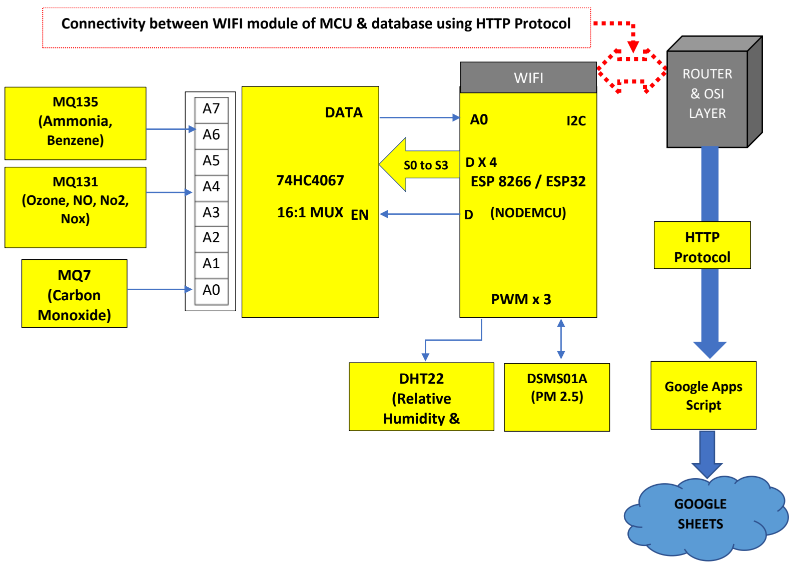


**Figure 2**. Block Diagram for RTAQSFS

MQ131 senses Ozone, NOx, NO2, NO & Benzene. The sensor incorporates an acute sensitivity to ozone and has a sensitivity to strong oxides emissions, CL2. MQ135 is implemented to detect Ammonia (NH4) & greenhouse emission (CO2).MQ7 is used for sensing CO. DHT22 measures temperature and humidity. GP2Y1010AU0F is used to measure PM 2.5-the dust within the environment. The sensed data at the controller is processed and sent using web socket, further the data is forwarded using Google app script then the data is inserted into new row in google spreadsheet. IoT is implemented using HTTP which process the data using URL.

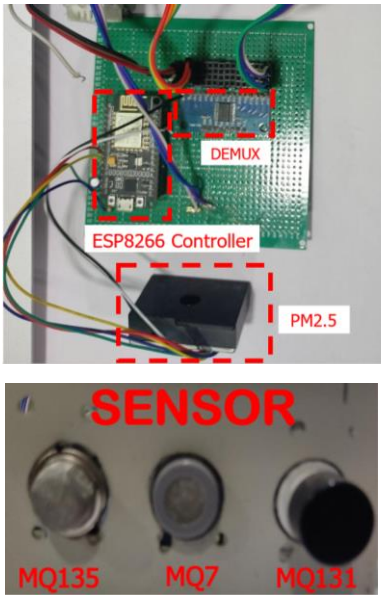


**Figure 3.** Hardware of RTAQSFS

## 4. MACHINE LEARNING

In this research paper supervised machine learning(ML) algorithm has been studied.

### 4.1 Linear Regression (LR)

In LR, independent and dependent variable follows linear relationship of straight line. Practical equation of LR consists of:

$$Y_i = X_i \beta_1 + \beta_0 \quad (1)$$

Where $Y_i$ = Dependent Variable, $X_i$ = Independent Variable, $\beta_0$ = intercept, $\beta_1$ = slope

The sum of squared discrepancies between observed values and anticipated values is decreased by the line.

### 4.2 Regularization

Regularization is an approach in ML algorithms that preserves accuracy by reducing the magnitude of the variable. A generic linear or polynomial regression will fail if there is sufficient co-linearity between the independent variables. Ridge regression can be used to solve these problems.

### 4.2.1 Ridge Algorithm

It is an LR model that reduces over fitting by including a minor penalty of square magnitude of the coefficients, as illustrated in equation 2.

$$\sum_{i=1}^{M}(yi - y'i)^2 = \sum_{i=1}^{M}\left(yi - \sum_{j=0}^{n}\beta j * xij\right)^2 + \lambda\sum_{j=0}^{n}\beta j^2 \quad (2)$$

Where βj =weights, λ =penalty term, yi = actual output y'i = predicted output

### 4.2.2. Lasso Regression

$$\sum_{i=1}^{M}(yi - y'i)^2 = \sum_{i=1}^{M}\left(yi - \sum_{j=0}^{n}\beta j * xij\right)^2 + \lambda\sum_{j=0}^{n}|\beta j| \quad (3)$$

βj = weights, λ = penalty term, yi = actual output, y'i = predicted output

It is least absolute selector operator which works similar to ridge model but the penalty is in only absolute weight rather than square weight like ridge. Lasso also help in feature selection.

### 4.3. Decision Tree

It is akin to a tree structure, with nodes representing datasets where the data is separated, leaves representing final outcomes output, and branches representing decisions. Decision nodes may have several branches, with the leaf node serving as the end point. In DT, Information Gain (I.G.) is used to measure classifier performance. The better the split, the greater the I.G.

$$I.G = Entropy_{Parent} - Entropy_{childern} \quad (4)$$

Each pollutant is measured in µg/m3. Descriptive statistics of each pollutant given in table 1 which clearly indicates that the pollutants values are exceeding the safer limit [D]. Figure 5 reflects the correlation plot among the pollutant and metrological parameters. Temperature and humidity are highly correlated with each other while Ozone and NOx shows correlation between them.

**Table 1.** Statistics of Pollutants for Hadapsar Region, Pune.

| Parameter | Min | Max | Mean | Std | Q1 | Q2 | Q3 |
|---|---|---|---|---|---|---|---|
| Temperature | 26.60 | 31.20 | 28.93 | 0.91 | 28.40 | 28.90 | 29.60 |
| Humidity | 50.60 | 84.60 | 61.82 | 5.16 | 57.23 | 62.30 | 64.88 |
| Ozone | 1.95 | 171.80 | 9.63 | 6.82 | 9.06 | 9.32 | 9.32 |
| CO | 0.22 | 113.51 | 11.16 | 11.47 | 3.91 | 8.01 | 10.98 |
| CO2 | 400.00 | 431.02 | 402.89 | 4.96 | 400.00 | 400.39 | 404.76 |
| PM2.5 | 6.75 | 545.3 | 276.03 | 34.19 | 154.80 | 281.01 | 77.33 |
| Ammonia (NH4) | 0.68 | 21.94 | 6.23 | 4.70 | 2.11 | 5.45 | 8.24 |
| Benzene | 0.18 | 3.63 | 0.20 | 0.14 | 0.19 | 0.20 | 0.20 |
| Nitric Oxide (NO) | 6.64 | 155.08 | 80.86 | 0.89 | 43.75 | 80.86 | 117.97 |
| No2 | 13.02 | 129.58 | 71.3 | 1.49 | 42.16 | 71.3 | 100.44 |
| Nitrogen Oxide (Nox) | 27.13 | 244.36 | 135.74 | 2.53 | 81.43 | 135.74 | 190.05 |

### 4.4 Random Forest

It combines the output of multiple DTs to make the ultimate decision. The outcome is based on average. So, for X inputs, multiple Y output are available. Based on averaging of multiple Y final output of Random Forest is determined. Machine learning is a part of Artificial Intelligence, it consists of a dataset. Where data is cleaned, algorithms are processed by training & testing. Based on testing & training values error rate of algorithm is delivered. Jupyter IDE is used to performed machine learning algorithm.

## 5. RESULT

RTAQSFS, Figure 4, is installed at Hadapsar region of Pune which has the Latitude 18.51950 and Longitude 73.85540. The region is well settled urbanized region. System monitors for continuously 1 Month for 24 hours. 80% of data is used of training and 20% of data is used for testing the performance of ML algorithm.


**Figure 4.** Map view pointing installation location.

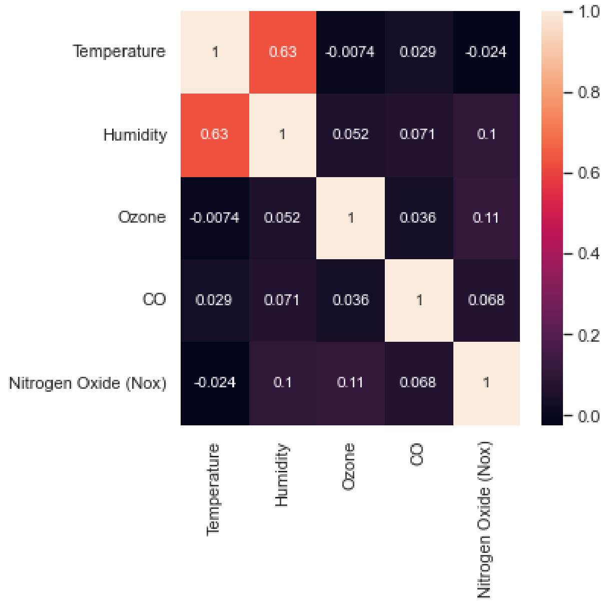


**Figure 5.** Correlation Plot of Pollutants and Metrological Parameters

As shown in Figure 6, when the power is turned on, the hardware is initialized, and the connectivity of all components, including sensors, has been checked. The raw values sensed by sensors with the assistance of algorithms for various sensors has been pre-processed. Once the value of the all the sensor has been collected the controller transforms it into the string and pushes the string to cloud using web socket based URL. Google Apps Script translates the data from string to parameter-wise format, which is subsequently entered into Google Sheets. After the cloud procedure is finished, it sends the status code to the controller. If there is an error status code, it will push the string until a positive status is received. The same process is repeated every 20 minutes.

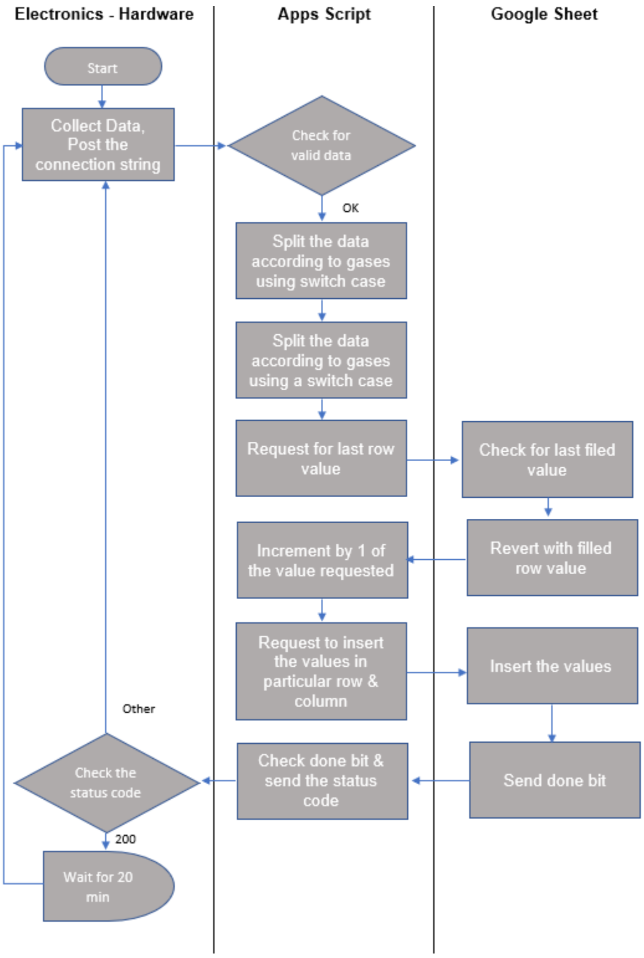


**Figure 6.** Flow Chart of the system.

The data collected for 31 days consists of total 2232 reading of the pollutant and environment parameters which forms the dataset. From total dataset 80% data is taken for training and 20% for testing various ML algorithm. The performance of ML algorithm has been compared by root mean square error (RMSE) and coefficient of determination ($R^2$).

For Random Forest, the parameters like Maximum depth = 8, N estimators = 115, Minimum sample leaf = 2 and Minimum sample split = 2, have been optimized and were used to calculate the final error.

Comparing the results from Table 2 and from figure 7 to 9, with respect to $R^2$ the best performing algorithm for Humidity, Ozone, CO, $NO_2$ is Lasso where the score is 99.9% and for ammonia it is 99.8% whereas for PM2.5 & Benzene it is Linear Regression with the score of 99.6% & 99.4% respectively. Random forest gives the accuracy of 99.9% almost for all parameters.

**Table 2.** Performance Comparisions

| Algorithm Parameter | Linear Regression | | Ridge | | Lasso | | Decision Tree | | Random Forest Regression | |
|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | $R^2$ | RMSE | $R^2$ | RMSE | $R^2$ | RMSE | $R^2$ | RMSE | $R^2$ |
| **Humidity** | 1.64 | 0.996 | 6.9 | 1.0 | 0.2 | 0.998 | 0.74 | 0.997 | 0.56 | 0.998 |
| **Ozone** | 6.56 | 0.969 | 2.61 | 1.0 | 0.01 | 0.999 | 0.22 | 0.936 | 0.65 | 0.998 |
| **CO** | 1.11 | 0.997 | 2.05 | 1.0 | 0.09 | 0.999 | 0.5 | 0.972 | 0.71 | 0.998 |
| **CO2** | 1.33 | 0.97 | 5.85 | 1.0 | 0.26 | 0.997 | 0.78 | 0.981 | 0.14 | 0.999 |
| **PM2.5** | 1.01 | 0.996 | 8.05 | 1.0 | 0.27 | 0.001 | 2.21 | 0.7 | 0.167 | 0.998 |
| **Ammonia (NH4)** | 7.25 | 0.995 | 4.17 | 1.0 | 0.18 | 0.998 | 0.09 | 1.0 | 0.33 | 0.998 |
| **Benzene** | 1.16 | 0.994 | 1.25 | 1.0 | 1.4 | 0.828 | 3.35 | 0.881 | 0.3 | 0.998 |
| **Nitric Oxide (NO)** | 9.53 | 0.997 | 4.87 | 1.0 | 2.5 | 0.999 | 0.23 | 1.0 | 0.08 | 0.999 |
| **No2** | 1.48 | 1.0 | 3.48 | 1.0 | 2.47 | 0.999 | 1.51 | 1.0 | 0.33 | 0.999 |
| **Nox** | 3.14 | 1.0 | 3.12 | 1.0 | 9.2 | 0.999 | 0.05 | 1.0 | 0.1 | 0.999 |

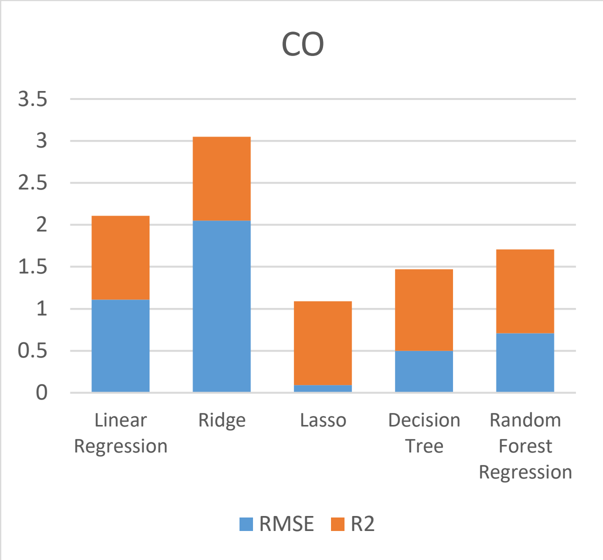


**Figure 7.** Performance of ML Algorithms for CO

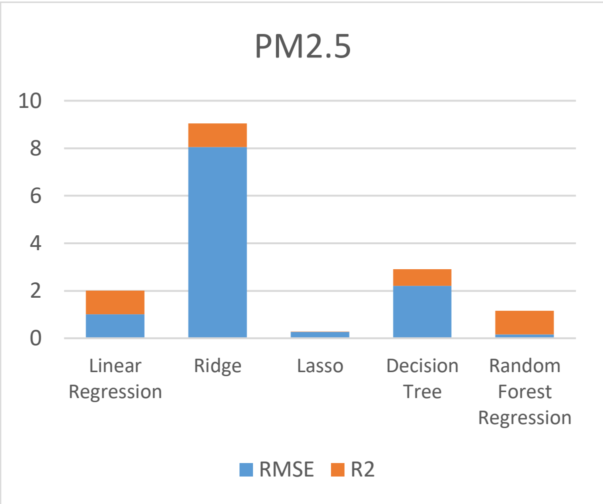


**Figure 8.** Performances of ML Algorithms for PM 2.5

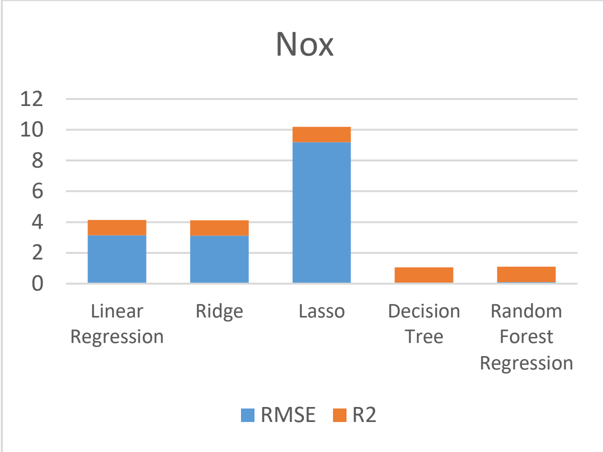


**Figure 9.** Performance of ML Algorithms for NOX

## 6. CONCLUSION

Predicting air quality is difficult due to environmental uncertainty, unpredictability, and the temporal and spatial fluctuation of contaminants. Air pollution is a severe global issue that has a negative impact on human life as well as animals and plants. Our proposed RTAQSFS collects real-time data from Hadapsar, Pune, which is utilized in the modelling of air quality prediction models. The collected data of pollutants and metrological parameters is transferred to the cloud using HTTP protocol. Pollutant data were statistically analyzed before being preprocessed. Correlation analysis displays the relationship between contaminants. For air pollution prediction 5 different ML algorithms viz. linear regression, Lasso regression, Ridge regression, Decision tree and random forest were implemented. The results show that Random Forest has the highest accuracy, followed by Lasso regression and linear regression. Further work could include developing more robust hybrid models with deep learning approach.

## References:


Annual report on Air Pollution Impact on Human life. Hindustan Times (2017). Thousand-People-Died-in-India-Due-to-Air-Pollution-in-2017.Retrieved from: https://www.livehindustan.com /national/story-nearly-12-lakh-40-thousand-people-died-in-india-due-to-air-pollution-in-2017-2301221.html.

Annual report on World Health Statistics WHO 2018(2018). Retrieved from: https:www.who.int/docs/default-source/gho-documents/world-health-statistic-reports/6-june-18108-world-health-statistics-2018.pdf.

Bekkar, A., Hssina,B. ,Samira Douzi, S., & Douzi,K.(2021). Air-pollution Prediction in Smart City-Deep Learning Approach. *Journal of Big Data*, 8(161), 1–21. doi:10.1186/s40537-021-00548-1.

Bisht, M., & Seeja, K. R. (2018). Air Pollution Prediction Using Extreme Learning Machine: A Case Study on Delhi (India). *Proceedings of First International Conference Smart Innovation, Systems and Technologies,* 181–189. doi:10.1007/978-981-10-5828-8_18

Castelli, M., Clemente,F., Popovic, A., Silva, S., & Vanneschi, L.(2020). A Machine Learning Approach to Predict Air Quality in California. *Hindawi Complexity*. 1–23.doi:10.1155/2020/8049504.

Chen,Y., Song,T., Chen,D., Song,K,.(2020). An Intelligent and Portable Air Pollution Monitoring System Based on Chemical Sensor Array. *IEEE 4th International Conference on Frontiers of Sensors Technologies.*21–25, doi: 10.1109/ICFST51577.2020.9294761.

Dun, M., Xu, Z.,Chen, Y., & Wu, L.(2020). Short-Term Air Quality Prediction Based on Fractional Grey Linear Regression and Support Vector Machine. *Hindawi Mathematical Problems in Engineering*, 1–13. doi:10.1155/2020/8914501.

Environment status report| Pune Municipal corporation (2017). Retrieved from:https://www.pmc.gov.in/en/environment-status-report-2016-17.

Gulia, S., Goyal, P., Prakash, M., & Kumar, R. (2021). Policy Interventions and Their Impact on Air Quality in Delhi City — an Analysis of 17 Years of Data. *Water, Air & Soil Pollution -International Journal of Environmental Pollution*, *232*(465). doi:10.1007/s11270-021-05402-x

Guo, J., Chen, Y., & Wu, L.(2021) .Using the Quarterly Compound Fractional Grey Model to Predict the Air Quality in 22 Cities of China. *Journal of Mathematics*, pp. 1–14, doi.org/10.1155/2021/4959457

Jiyal, S., & Saini, R.(2020). Prediction and Monitoring of Air Pollution Using Internet of Things (IoT). *Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 57–60, doi:10.1109/PDGC50313.2020.9315831.

Liang, Y., Maimury, Y., Chen, A., & Juarez, J.(2020). Machine Learning-Based Prediction of Air Quality. *Applied Science*, *10*(9151), pp. 1–17, doi:10.3390/app10249151.

Madan, T., Sagar, S., & Virmani, D.(2020). Air Quality Prediction Using Machin Learning Algorithms –A Review.*2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 140–45. doi: 10.1109/ICACCCN51052.2020.9362912.

Mijakovski, V., & Lutovska, M. (2020). Monitoring of Air Pollution in Skopje in Order to Determine Possible Causes. *International Scientific Conference on Information, Communication and Energy Systems and Technologies*, 239–243:doi: 10.1109/ICEST49890.2020.9232702.

Simu, S., Turkar, V., Martires, R., Asolkar, V., Monteiro, S., Fernandes, V., & Salgaoncar, V. (2020).Air Pollution Prediction Using Machine Learning. *IEEE Bombay Section Signature Conference*, 231–35. doi: 10.1109/IBSSC51096.2020.9332184.

Parmar, G., Lakhani, S., & Chattopadhyay, M. (2017). An IoT Based Low Cost Air Pollution Monitoring System. *International Conference on Recent Innovations Is Signal Processing and Embedded Systems*, 524–28.doi:org/10.1109/RISE.2017.8378212.

Vibhute, S., & Dhope.T. (2020). Air Quality Prediction in Mumbai City Using Machine Learning-Based Predictive Models. *International Journal of Future Generation Communication and Networking.* 13(4), 4731–4740. sersc.org/journals/index.php/IJFGCN/article/view/35071.

Vong, C.-M., Ip, W.-F., Wong, P., & Yang, J. (2012). Short-Term Prediction of Air Pollution in Macau Using Support Vector Machines. *Journal of Control Science and Engineering*, 2012, 1–11. doi:10.1155/2012/518032

Technical Assistance Document for the Reporting of Daily Air Quality-the Air Quality Index (AQI). (2018). Retrieved from https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf.

Xiaojun, C., Xianpeng, L., & Peng, X. (2015). IOT-based air pollution monitoring and forecasting system. *2015 International Conference on Computer and Computational Sciences (ICCCS)*. 257-260. doi:10.1109/iccacs.2015.7361361

**Tanuja Satish Dhope**
Electronics and Communication,
Bharati Vidyapeeth (Deemed to be University) College of Engineering,
Pune,
India,
tanuja_dhope@yahoo.com
ORCID 0000-0003-2907-8509

**Ahmed Shaikh**
Electronics and Communication,
Bharati Vidyapeeth (Deemed to be University) College of Engineering,
Pune,
India,
shaikahmed.8598@gmail.com
ORCID 0009-0001-6640-5727

**Dina Simunic**
Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia.
dina.simunic@fer
ORCID 0000-0002-2406-774X

**Prashant Pandurang Patil**
Bharati Vidyapeeth (Deemed to be University) Y. M. Institute of Management,
Karad.
India,
prashant.patil@bharatividyapeeth.edu
ORCID 0000-0002-8797-3715

**Kishor S. Wagh**
Associate Professor in Computer Engineering, AISSMS IOIT,
Pune,
India,
waghks@gmail.com
ORCID 0000-0002-0309-0705

**Sharmila K. Wagh**
Professor in Computer Engineering, Modern Education Society's College of Engineering,
Pune,
India,
sharmila.wagh123@gmail.com
ORCID 0000-0001-7232-3819