International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

Real-time Air Pollution Monitoring and Modeling using Machine Learning and statistical Techniques

[1]Obodoeze Fidelis Chukwujekwu, [2]Oliver Ifeoma Catherine, [3]Onuzulike Vincent Chukwuma, [4]Onyemachi George Olisamaka, [5]Udeh Ifeanyi Frank Gideon

[1]*Department of Computer Engineering, Akanu Ibiam Federal Polytechnic Unwana, Ebonyi State, Nigeria*
[2]*Department of Computer Science, Akanu Ibiam Federal Polytechnic Unwana, Ebonyi State, Nigeria*
[3]*Department of Electronic and Computer Engineering, Nnamdi Azikiwe University, Awka, Nigeria*
[4]*Department of Computer Science, Akanu Ibiam Federal Polytechnic Unwana, Ebonyi State, Nigeria*
[5]*Department of Computer Science, Federal College of Agriculture, Ishiagu, Ebonyi State, Nigeria*

Submitted: 15-12-2022 Accepted: 25-12-2022

ABSTRACT:
Environmental and human health are seriously threatened by air pollutants that are released into the immediate area. The detrimental effects of air pollution are more severe in Metropolis or densely populated areas. When air pollutants such microscopic particles or particulate matters like PM2.5, PM10.0 find their it into human respiratory tracts, health concerns such as respiratory, cardiac, and even fatalities might occur. Inhaling these particles can cause illnesses like asthma, COPD, laryngitis, and other secondary lung infections. This study examined how to predict PM10.0 emissions in Nigeria's Anambra State 24 hours in advance using air pollution prediction. About 12,958 historical datasets of air and noise pollution, as well as meteorological parameters captured in Awka from October 25 to December 4 of 2021 using real-time air pollutants and weather wireless sensors, were used as input predictors for the model. These datasets included VOCs, particulate matter, and carbon dioxide, as well as noise, and meteorological parameters like air temperature, relative humidity, pressure, and light intensity. Train-Test data split and statistical comparative analysis techniques were chosen to determine the optimal perfuming model, various experiments were carried out using the study of roughly eight Machine Learning algorithms. The performances of the models' predictions were assessed using three performance metrics, including the Coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Square Mean Error (RMSE). With an R^2 of 0.9834 (98.34%) and the lowest prediction errors (MAE of 0.4225 g/m^3 and RMSE of 1.2413 g/m^3, according to the experimental data, the Multiple Linear Regression (MLR) technique outperformed the other seven ML models in predicting PM10.0 pollution levels in the Awka Metropolis. The experimental testbed used for the tests consists of the Anaconda IDE, Python 3.6.7, and the Keras, Tensorflow, and Scikit-learn libraries for Python. The programming environment and simulation environment used was the Jupyter IDE.

Keywords: Air Pollution, PM10.0, particulate matter, particles, Train-Test method, Comparative Analysis, Awka Metropolis

I INTRODUCTION

Due to its negative impacts on the environment and general well-being of the city's residents, air pollution or a decline in air quality has been a severe concern for Megacities or Metropolises. This is due to the fact that air pollution negatively impacts the surrounding ecosystem in a number of ways, including global warming, ozone layer destruction, which causes cancer and other serious diseases, soil acidification, and a reduction in plant growth. Asthma, pneumonia, bronchitis, laryngitis, and even mortality are some of the respiratory and cardiovascular conditions that are made worse by air pollution in both adults and children.

Particles or particulate matters like PM2.5, PM10.0, and PM1.0 air pollutants as well as other dangerous air pollutants like ozone (O_3), sulphur

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

oxides (SO_X), nitrogen oxides (NO_X), volatile organic compounds (VOCs), etc. are among the many different types of air pollutants thatexist and they can be either primary or secondary types of air pollutants. If a PM2.5 or PM10.0 particulate matter penetrates a person's respiratory system, it poses a serious health risk.

In order to safeguard the lives of people and other living things, it is crucial to monitor air pollution levels in real time or to predict them in advance.

In order to find the best or most appropriate machine learning model for 24 hour advance prediction of PM10.0 particulate matter, this paper will compare and analyze a variety of machine learning algorithms, including Decision Trees, Multi Layer Perceptron Layer Neural Network (MLP ANN), Multiple Linear Regression (MLR), and Single Ensemble Learning Algorithms like Random Forest, Adaptive Boosting algorithm (AdaBoost), Extreme Gradient Boosting (XGBoost), and Extra Trees.

II LITERATURE REVIEW

In their study, Shishegaran et al.(2020) used ensemble models such Random Forest, XGBoost, AdaBoost, LightGBM, CatBoost, and Extra Trees to forecast the air quality index (AQI) in Tehran, Iran. Daily air pollution data for five years, from 2012 to 2016, including ground-level ozone (O_3), NO_2, PM2.5, and PM10, served as predictors for the created model. Additionally included as input predictors in the study were historical meteorological or weather datasets such as radiation, visibility, pressure, wind speed, and sunshine hours. The experimental findings demonstrated that the ensemble model, with an R^2 score of 0.983, outperformed other models in AQI prediction.

In a study report, Cortina-Januchs et al.(2015) discussed the creation of a model for forecasting PM10 concentrations in Salamanca, Mexico. The proposed model predicted the city's highest hourly daily PM10 concentrations for the following day using a combination of the Multi Layer Perceptron (MLP) Artificial Neural Network and the Clustering algorithm. The source dataset combined PM10 concentrations with historical time series of meteorological variables such wind direction, wind speed, air temperature, and relative humidity. In comparison to utilizing either a conventional analytical modeling tool or an Artificial Neural Network (ANN) alone, experimental results indicated that combining an Artificial Neural Network (MLP) and clustering algorithm together increased the accuracy of PM10 forecasting.

Alexander Trenchevski et al (2020) presented a research paper on the use of weather historical dataset and various machine learning regression models or algorithms to model and predict the future possible PM_{10} concentrations in the cities of Karpos and Kumanovo Municipalities of Macedonia. Four years of historical data from 2015-2018 was used comprising of ten input variables of PM_{10} concentrations and weather dataset. Experimental setups were conducted using various machine learning algorithms such as XGBoost, LightGBM, Decision Trees, Regression algorithm, Dummy regression, Random Forest, Support Vector Regression (SVR). Experimental results showed that XGBoost and LightGBM followed by Random Forest outperformed other algorithms in terms of the performance metrics – R^2, MAE and RMSE. XGBoost is the best, followed by LightGBM, and then Random Forest.

Khoshand et al (2017**)** presented a study on the prediction of ground-level air pollution using Multi-Layer Perceptron (MLP) Artificial Neural Network (ANN) using Back Propagation (BP) algorithm with MATLAB program for training the model. The developed model was used to manage and predict daily concentrations of various air pollutants such as Ozone, PM_{10}, NO_2, CO and $PM_{2.5}$ in Tehran city of Iran within a four-year period from 2012 to 2015.

The experimental results showed appropriate agreement between the observed and predicted concentrations. This shows that ANN can be used to improve prediction accuracy of air pollution.

III METHODS AND MATERIALS

About 12,958 rows of historical datasets were trained using eight machine learning algorithms, including Multiple Linear Regression (MLR), Multi Layer Perceptron Artificial Neural Network (MLP ANN), Decision Tree, Support Vector Regression (SVR), Random Forest, XGBoost, AdaBoost, and Extra Trees. These datasets were obtained from an air pollution monitoring system installed in the Awka Metropolis and were collected between October 25 and December 4, 2021. Visit http://www.mtrackers.com.ng/awka-pollution-monitor/AWKA-POLLUTION-2022NEW.csv to access the dataset used in this study.

The split between the training and testing datasets was 70% to 30%, respectively. While the testing dataset was used to assess the effectiveness of the machine learning models, the training dataset

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

was used to recognize patterns in the air pollution dataset in order to fit the models and generalize to upcoming or new datasets.Fig.1 shows the division of the historical dataset into two parts.

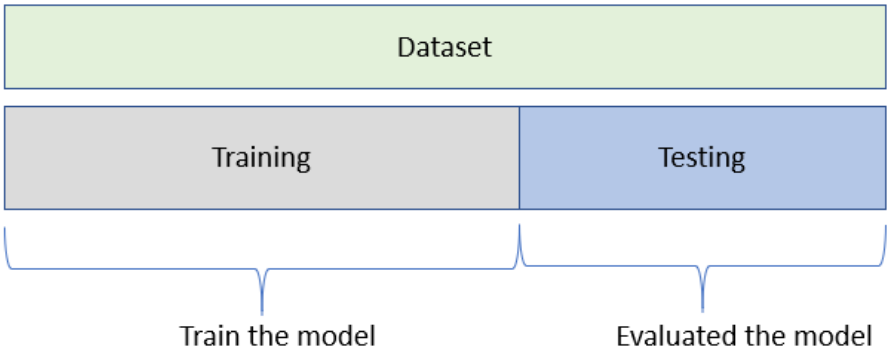

Fig. 1: Dataset split into training and testing data sets for Train-Test Split method

Fig. 2 shows the following steps taken to actualize the experiments in this research work using Test-Train Split Method:

1. *Load the Awka Metropolis Pollution historical Dataset into Python Pandas or memory storage*
2. *Split the loaded dataset into training samples (70%) and the remaining 30% as testing dataset*
3. *Fit the models on the training dataset*
4. *Perform pollutant's concentration prediction using the test dataset*
5. *Plot the model (measured data versus predicted data)*
6. *Print the accuracy and prediction errors scores*
7. *Compare the scores of accuracy and prediction errors of each model*

Fig. 2: Train-Test Split method steps used in the experiments

The experimental runs or simulations for each of the eight machine learning models were executed in Scikit-learn machine learning module in Python 3.
The results of the experimental runs carried in this experiment are presented in section IV of this research paper.

3.1 The Datasets

In order to fit the models, 12,958 rows of data were collected, spanning a period of 43 days, from the deployed newly constructed Air and Noise Pollution Monitoring station within Awka Metropolis. These datasets, consisting of one-minute historical data of meteorological and pollution concentrations, were used in the various experiments in this paper. TVOC, PM10, PM2.5, and PM1.0 are only a few of the minute pollutant concentrations that are recorded on the ground. Other weather or meteorological datasets include air temperature, pressure, relative humidity, and light intensity.
Visit http://www.mtrackers.com.ng/awka-pollution-monitor/AWKA-POLLUTION-2022NEW.csv to download and access the dataset used in the study.

3.2 Data Normalization

The data used for the experiments are normalized or scaled to remove any irregularities in

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

the values or weights of the models. The range normalization function used is mathematically given as follows:

$$X_{normalisation} = \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \quad (3.1)$$

where $X_{normalisation}$ is the normalized value, X_i is the i_{th} value passed, and X_i and X_i are the minimum and maximum value for X_i value respectively.

The data normalization was implemented using MinMax Scalar() inbuilt Python 3 function.

3.3 Experimental Testbed

Software tools like Python 3.7.3 and Jupyter Notebook version 3.2 were used as a simulation and programming environment for the studies in this article. In the Python Jupyter Notebook and Spyder Integrated development environments, all the machine learning programs for the implementation of the pollution prediction models were created and evaluated. Scikit-learn, a Python-based alternative to MathLab, is the machine learning module used in this study. The following additional required libraries or packages are used: Pandas, Tensorflow, Numpy, Mathplotlib, and Keras. The integration environment that connects all of these programs and packages in Python 3 running on Windows 7 was Anaconda Navigator version 3.

To collect data, analyze data, and predict pollution in this study, a variety of tools and frameworks were used.The Anaconda and Tensorflow Integrated Development Environments utilized in this study are shown in Figs. 3 and 4. The experimental runs in this study used Keras version 2.9.0 and Tensorflow version 2.8.0, respectively.

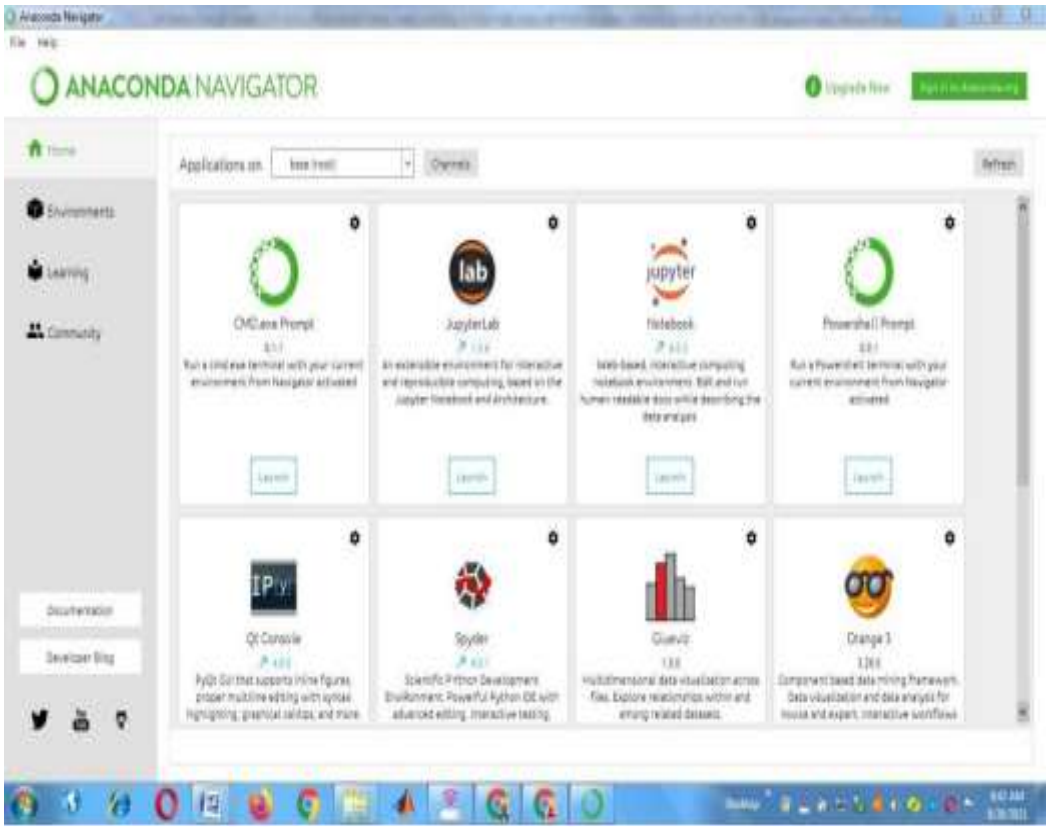

Fig.3: Anaconda Integrated Development and Management Environment for Python 3.7.3 and Machine Learning Modules

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

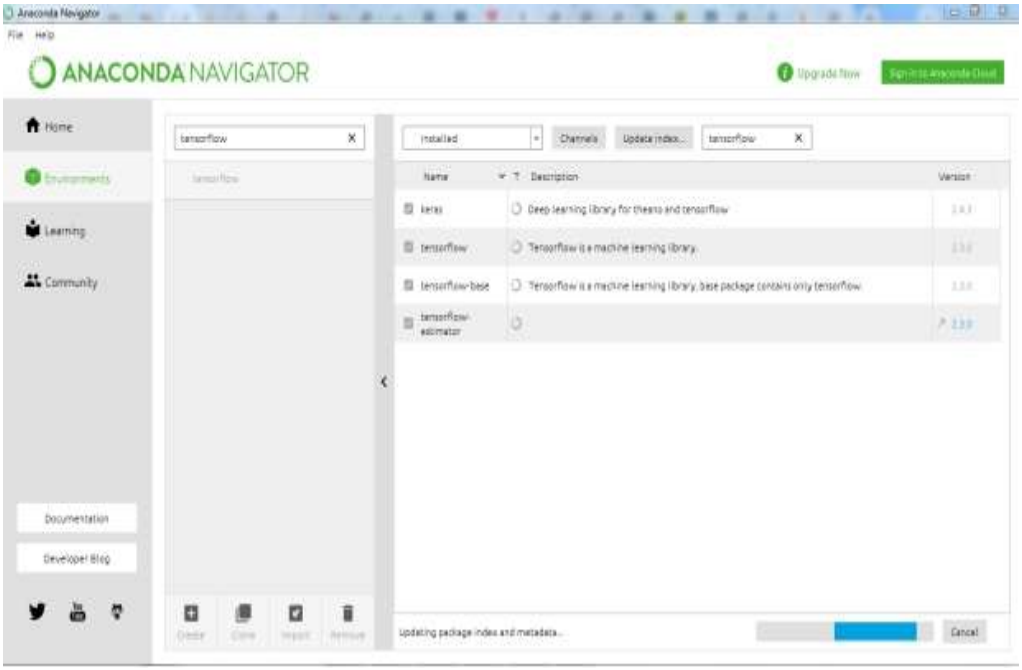

Fig.4: Tensorflow Environment in Anaconda Navigator created for the research experiments using Python 3

3.4 Performance Evaluation Metrics

The best prediction machine learning model was chosen based on model accuracy and residual prediction errors using the performance evaluation metrics R-squared (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

- **Mean Absolute Error (MAE),**
- **Root Mean Square Error (RMSE) and**
- **R^2 (Coefficient of Determination)**

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |P_i - M_i| \quad (3.2)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |P_i - M_i|^2} \quad (3.3)$$

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_{i=1}^{n}(P_i - M_i)^2}{\sum_{i=1}^{n}(M_i - \bar{M}_i)^2} \quad (3.4)$$

SS_{RES} = Sum of Residual or regression Errors

SS_{TOT} = Sum of Total Errors $\bar{M}_i$ = Mean of all measured values

M_i = measured or observed value

P_i = Predicted Value

Where n is the number of data in the test dataset, P_i and M_i are the predicted and measured value for the i^{th} hour.

IV RESULTS AND DISCUSSION

The findings from the many experimental runs included in this study are presented and discussed in this section.

The regression graphs for the evaluation of PM10.0 prediction performances using the eight "machine learning" methodologies are shown in Figs. 5 to 12.

The following results were achieved using the MLP Artificial Neural Network (ANN) method, as shown in Fig. 5 to predict PM10.0 concentrations: R^2 = 0.9828, or 98.28 percent prediction accuracy; RMSE = 1.2637 µg/m^3 and MAE = 0.5885 µg/m^3.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

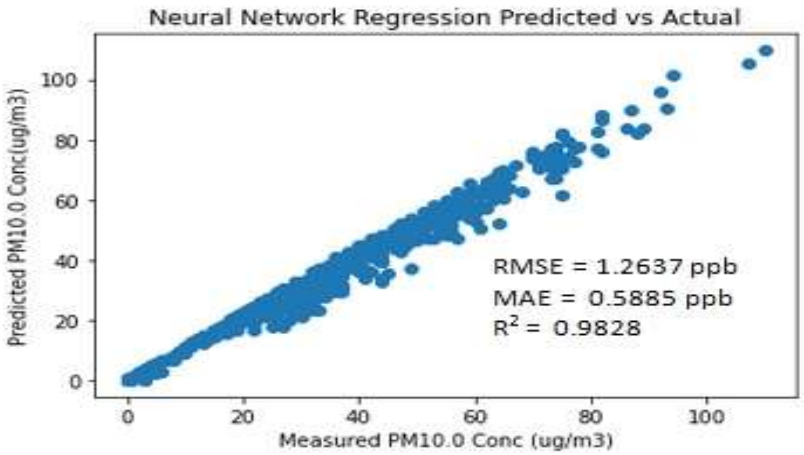

Fig.5: Regression Scatterplot of PM10.0 pollution using MLP Artificial Neural Network (ANN) algorithm

Fig. 6 displays the regression plot result of PM10.0 concentrations prediction using the XGBoost algorithm. The following results were obtained: RMSE= 1.4319 µg/m^3; MAE= 0.4664 µg/m^3, and R^2= 0.9779 or 97.79 percent prediction accuracy.

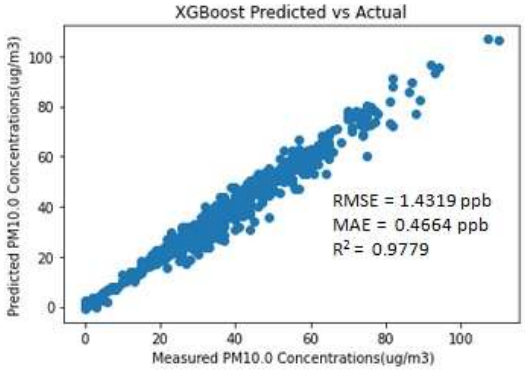

Fig. 6: Regression Scatterplot of $PM_{10.0}$ pollution using XGBoost algorithm

Using the XGBoost method, the regression plot result for the prediction of PM10.0 concentrations as shown in Fig. 6 yielded the following results: RMSE= 1.2413 µg/m^3, MAE= 0.4223 µg/m^3, and R^2= 0.9834, indicating 98.34 percent prediction accuracy.

Fig. 7 shows the regression plot of PM10.0 concentrations prediction using Multiple Linear Regression (MLR) and the following results were obtained: RMSE=1.2415µg/m^3, MAE=0.4223 µg/m^3 and R^2=0.9834.

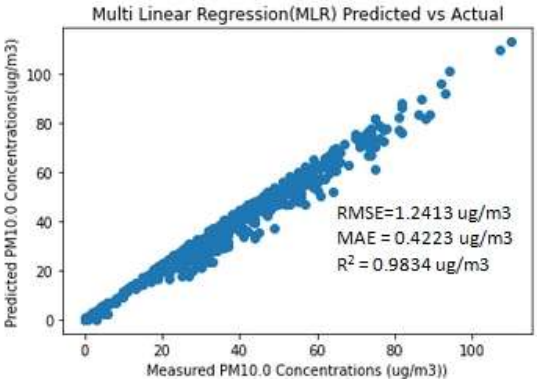

Fig. 7: Regression Scatterplot of $PM_{10.0}$ pollution using Multiple Linear Regression (MLR) algorithm

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

Fig. 8 displays the regression plot result of PM10.0 concentrations prediction using Decision Tree method and the following results were obtained: RMSE= 1.8457 µg/m^3, MAE= 0.5486 µg/m^3, and R^2= 0.9634 or 96.34 percent prediction accuracy.

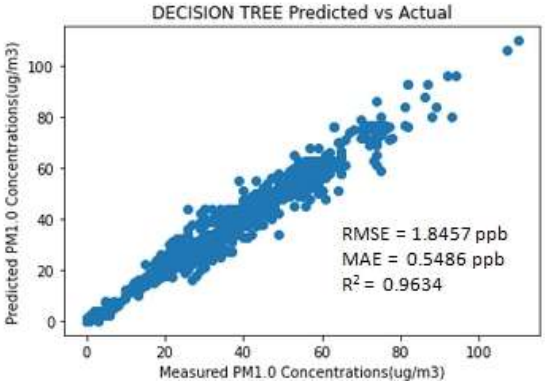

Fig.8: Regression Scatterplot of $PM_{10.0}$ pollution using Decision Tree algorithm

Fig. 9 displays the regression plot result of PM10.0 concentrations prediction using Adaptive Boosting algorithm (AdaBoost) method and the following results were obtained: RMSE= 1.5803 µg/m^3, MAE= 0.4619 µg/m^3 and R^2= 0.9731 or 97.31% prediction accuracy.

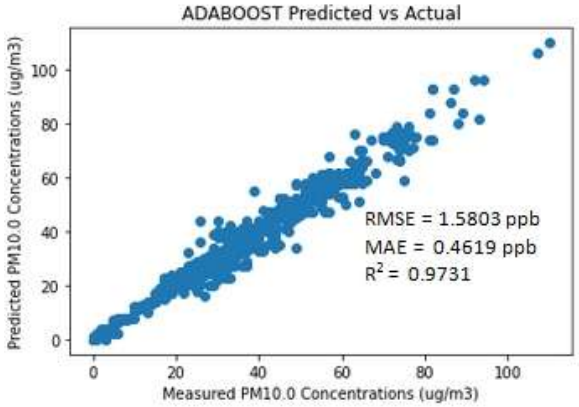

Fig.9: Regression Scatterplot of $PM_{10.0}$ pollution using AdaBoost algorithm

Fig. 10 displays the regression plot of PM10.0 concentrations prediction using Extra Tree method and the following results were obtained: RMSE=1.3652 µg/m^3, MAE= 0.4305 µg/m^3 and R^2= 0.9800 or 98% prediction accuracy.

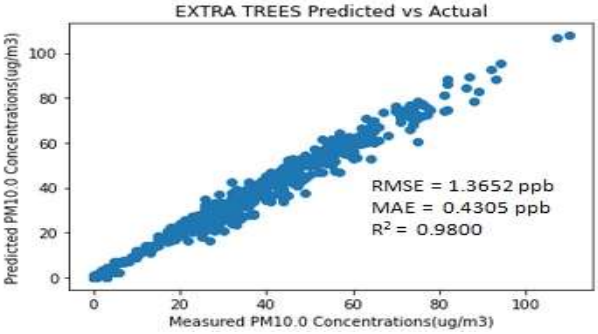

Fig. 10: Regression Scatterplot of $PM_{10.0}$ pollution using Extra Tree algorithm

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

Using the Random Forest technique, the regression plot result in Fig. 11 for the prediction of PM10.0 concentrations yielded the following results: RMSE=1.3531 µg/m^3, MAE= 0.4313 µg/m^3 and R^2= 0.9803 or 98.03% prediction accuracy.

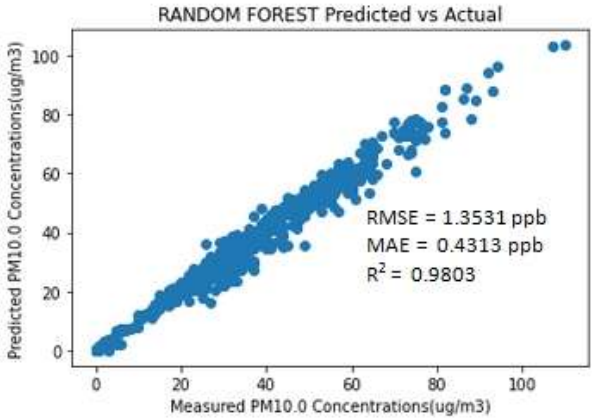

Fig.11: Regression Scatterplot of $PM_{10.0}$ pollution using Random Forest algorithm

The following results were obtained using the Support Vector Regression (SVR) algorithm, as shown in Fig.12, to predict PM10.0 concentrations: R^2 = 0.9582, or 95.82 percent prediction accuracy, and RMSE = 1.9719 µg/m^3, MAE = 0.5876 µg/m^3,

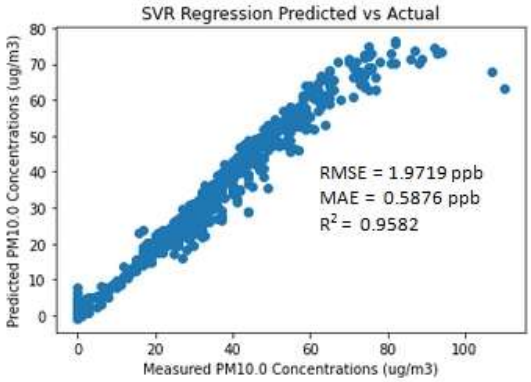

Fig.12: Regression Scatterplot of $PM_{10.0}$ pollution using Support Vector Regression (SVR) algorithm

Table 1 shows the performance evaluation results for the various eight (8) "machine learning" algorithms used to estimate PM10.0 Pollution using the Train-Test Split approach. For PM10.0 prediction using the MLP ANN algorithm, the printout of the actual values versus the predicted values is shown in Table 2; for PM10.0 prediction using the Random Forest algorithm, the printout is shown in Table 3. The actual versus predicted values for PM10.0 using the Extra Trees technique are shown in Table 4. The actual versus predicted values for SVR algorithm is shown in Table 5. The printouts for the actual versus predicted values for Decision Trees and AdaBoost algorithms are shown in Table 6 and Table 7 respectively.

Table 1: PM10 Pollution prediction results using various machine learning algorithms and Train-test Split data method

ML Model or Algorithm	RMSE (µg/m^3)	MAE (µg/m^3)	R^2	Ranking
MLR	1.2413	0.4225	0.9834	1st
SVR	1.9719	0.5876	0.9582	8th
MLP ANN	1.2637	0.5885	0.9828	2nd

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

Decision Tree	1.8457	0.4313	0.9634	7th
Random Forest	1.3531	0.4313	0.9803	3rd
AdaBoost	1.5803	0.4619	0.9731	6th
XGBoost	1.4319	0.4664	0.9779	5th
Extra Trees	1.3652	0.4305	0.9800	4th

Table 2: MLP ANN PM10.0 prediction: Actual versus Predicted Values

	Actual-Values PM10 (ug/m3)	Predicted-Values PM10 (ug/m3)
0	25.0	25.021405
1	25.0	24.944486
2	25.0	25.051864
3	25.0	24.908995
4	25.0	25.056245
5	26.0	26.669581
6	25.0	25.069622
7	25.0	24.973060
8	25.0	25.114676
9	25.0	25.020091
10	25.0	25.028941
11	25.0	24.962883
12	25.0	26.925041
13	21.0	22.073756
14	25.0	24.977969
15	25.0	24.988212
16	25.0	24.991902
17	25.0	25.018644
18	25.0	25.029647
19	16.0	16.945953

Table 3: Random Forest PM10.0 prediction: Actual versus Predicted Values

	Actual-Values PM10 Concentrations(ug/m3)	Predicted-Values PM10 Concentrations(ug/m3)
0	25.0	25.000
1	25.0	25.000
2	25.0	25.000
3	25.0	25.000
4	25.0	25.000
5	26.0	27.285
6	25.0	25.000
7	25.0	25.000
8	25.0	25.000
9	25.0	25.000
10	25.0	25.000
11	25.0	25.000
12	25.0	26.585
13	21.0	22.575
14	25.0	25.000
15	25.0	25.000

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

Table 4: Extra Trees PM10.0 prediction: Actual versus Predicted Values

	Actual-Values PM10.0 Concentrations(ug/m3)	Predicted-Values Concentrations(ug/m3
0	25.0	25.000
1	25.0	25.000
2	25.0	25.000
3	25.0	25.000
4	25.0	25.000
5	26.0	26.835
6	25.0	25.000
7	25.0	25.000
8	25.0	25.000
9	25.0	25.000
10	25.0	25.000
11	25.0	25.000
12	25.0	26.580
13	21.0	22.035
14	25.0	25.000
15	25.0	25.000

Table 5: PM10.0 prediction using SVR: Actual versus Predicted Values

	Actual-Values PM10 Concentration(ug/m3)	Predicted-Values PM10 Concentration(ug/m3)
0	25.0	25.086575
1	25.0	25.066703
2	25.0	25.089693
3	25.0	25.012889
4	25.0	25.090789
5	26.0	25.203136
6	25.0	25.071770
7	25.0	25.092434
8	25.0	25.096902
9	25.0	25.094918
10	25.0	25.078559
11	25.0	25.085950
12	25.0	24.436370
13	21.0	20.900206
14	25.0	25.090160
15	25.0	25.080989
16	25.0	25.100456
17	25.0	25.093950
18	25.0	25.096322
19	16.0	14.595528

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 697-708 www.ijaem.net ISSN: 2395-5252

Table 6: PM10.0 prediction using Decision Trees: Actual versus Predicted Values

	Actual-Values PM10 Concentrations (ug/m3)	Predicted-Values PM10 Concentrations (ug/m3)
0	25.0	25.0
1	25.0	25.0
2	25.0	25.0
3	25.0	25.0
4	25.0	25.0
5	26.0	26.0
6	25.0	25.0
7	25.0	25.0
8	25.0	25.0
9	25.0	25.0
10	25.0	25.0
11	25.0	25.0
12	25.0	26.0
13	21.0	21.0
14	25.0	25.0
15	25.0	25.0

Table 7: PM10.0 prediction using AdaBoost: Actual versus Predicted Values

	Actual-Values PM10.0 Concenrations(ug/m3)	Predicted-Values PM10.0 Concenrations(ug/m3)
0	25.0	25.0
1	25.0	25.0
2	25.0	25.0
3	25.0	25.0
4	25.0	25.0
5	26.0	27.0
6	25.0	25.0
7	25.0	25.0
8	25.0	25.0
9	25.0	25.0
10	25.0	25.0
11	25.0	25.0
12	25.0	25.0
13	21.0	23.0
14	25.0	25.0
15	25.0	25.0

The results obtained from Tables 1 to 7 and Figs 5 to 12 show that Multiple Linear Regression (MLR) algorithm scored the highest accuracy score (R^2= 0.9834 or 98.34%) and lowest errors of prediction (MAE=0.4225µg/m^3 and RMSE= 1.2413 µg/m^3) compared to other seven Machine Learning algorithms and hence MLR is selected for future prediction of PM10.0 for Awka Metropolis. This also means that the predictor

variables in the dataset have a very high linearity with the target variable PM10.0.

V CONCLUSION

It is impossible to overstate the importance of clean air in megacities and Metropolis. The survival of humans and all living creatures on earth depends heavily on air. Humans are vulnerable to a wide range of respiratory illnesses and infections that can even be fatal or seriously unwell. The challenge of air pollution has been brought on by the fast industrialization and urbanization, and Nigerian cities are not exempt. In order to prevent the escalation of pollution thresholds that may negatively impact human health, it is essential to monitor the concentrations or levels of air pollution in large cities around-the-clock and predict the emission levels of these air pollutants in advance.

This study was successful in implementing a system to predict or anticipate air pollution levels for the Awka Metropolis 24 hours in advance using Train-Test and a technique for comparative analysis for PM10.0 monitoring utilizing eight well-known machine learning algorithms.

REFERENCES

[1]. Cortina-Januchs, M.G., Quintanilla-Dominguez, G., Vega-Corona, A., and Andina, O. (2015). Development of a model for forecasting of PM10 concentrations in Salamanca Mexico. Atmospheric Pollution Research 6(2015).pp.626-634.

[2]. Khoshand, A., Sehrani, M.S., Kamalan, H. and Bodaghpour, S. (2017). Prediction of Ground-Level Air Pollution Using Artificial Neural Network in Tehran. Anthropogenic Pollution, Vol. 1(1), 2017, pp.61-67. DOI: 10.22034/apj.2017.1.1.6167

[3]. Shishegaran, A., Saeedi, M., Kumar, A., Ghiasinejad, H. (2020). “Prediction of air quality in Tehran by developing the nonlinear ensemble model”. Journal of Cleaner Production, 259, 120825.

[4]. Trenchevski, Alexander.; Marija Kalendar, Hristijan Gjoreski and Danijela Efnusheva (2020).

[5]. "Prediction of Pollution Concentration using weather data and Regression Analysis Models". Proc. of the 8th International Conference on Applied Innovations in IT (ICAIIT), March 2020. pp.55-61.