Date of publication Jan. 10, 2022, date of current version Jan. 10, 2022.

Prediction of Air Quality Index in Colombo

RM Fernando [1#], WMKS Ilmini [2] and DU Vidanagama [3]

[1,2,3]Department of Information Technology, General Sir John Kotelawala Defence University, Sri Lanka
[1#]35-cs-0001@kdu.ac.lk

ABSTRACT Air is always considered as the main critical factor on which human survival depends on. The AQI or long firmly air quality index is the index value that illustrates qualitatively the current state of the air. The substantial AQI will further menace the living creatures' health & the living atmosphere. Terrible air quality has been a major concern in Sri Lanka, particularly in urban cities such as Colombo and Kandy. Reliable AQI prediction will assist to decrease the health risks caused by air pollution. The goal of this study has been to find the most suitable machine learning approach for predicting accurate air quality index in Colombo based upon PM2.5 particular concentration. In this study, PM2.5 concentration in Colombo had been predicted using four correlated air pollutant concentrations such as SO2, NO2, PM2.5, & PM10. The obtained dataset was pre-processed via prediction in order to improve prediction accuracy. The gathered dataset Cross-validated as according to 80% for training & 20% for testing the prediction model. Machine learning methods such as K-Nearest Neighboring, Multiple Linear-Regression, Random Forest, and Support Vector Machines were used to train and evaluate the prediction models. In the end, we achieved 83.25% accuracy for the K-Nearest Neighboring algorithm model, 84.68% accuracy for the Support Vector Machines model, 85.17% accuracy for the Random Forest model, and 41.9% accuracy for the Multiple Regression Model. Random Forest was recognized as the best appropriate prediction model after evaluating the models, with over 85% greater accuracy.

KEYWORDS: Air Quality Index (AQI), Correlation, Machine Learning, Model, Pollution

I INTRODUCTION

Air Quality or in other words quality of the breathing air around us is a highly valuable luxury that people should have but unfortunately, more than two-thirds of the people who live on our planet do not have it. By referring to the most recent reports of the WHO, more than half of the people who live on our planet live in extremely air polluted urban areas and almost all the people who live in those areas are not aware of the quality of the air in their atmosphere. When considering the quality of the air in Colombo, it is evident that air pollution in Colombo, increases at a very rapid rate yearly. In the rush hours in the morning and evening, the Air quality Index in Colombo exceeds the 4th category level of the world health organization recommended. Unfortunately most of the people are not aware of AQI level around their living space and as a result of their unawareness, especially in areas such as Colombo, the people have developed an indifferent attitude towards the adverse effects that air pollution might have on their health [1].

This paper mainly explores air quality index prediction using various machine learning algorithms such as regression algorithm, support vector machines algorithm, k-nearest neighboring algorithm & random forest algorithm. Various machine learning models have been implemented & evaluated to find out the best accurate machine learning model. This research mainly focuses on enhancing the methods that have been used to predict air quality index and improve the knowledge on air quality index, and to understand the effects of the bad air quality. In Colombo, the AQI value is always at the average unhealthy level. With some reliable & accurate predictions of the AQI category, the Colombo public can get necessary precautionary measures like increasing the indoor activities & minimize the outdoor activities as much as possible to protect themselves from consequences of bad air quality in Colombo. This chapter presents a very brief description of the AQI, the motivation of this research study and background details about Colombo air quality which is followed by the goal of the research[2]. Minimizing air pollution and ensuring that people have access to clean air, are major responsibilities of human beings as their actions have led to the increased level of air pollution in the world. Some human revolutions such as industrialization have mainly affected the earth's air pollution. An increase in the vehicles and machines which emit carbon dioxide into the air can be identified as another cause for air pollution. Basically, air pollution can be described as mass contamination of the air in a specific geographical area by the emission of enormous amount of harmful gases and chemical substances to the environment. Emissions from productions, industries & vehicles are the main reason for poor air quality. A highly populated urban city such as Colombo most of the time has the poorest air quality

compared to the other rural areas mainly due to the actions of human population. There is a strong correlation between air quality index value and threats to human health, these threats include both short-term & long-term side effects on the health of living beings and their environment. The people who have suffered from diseases such as asthma, and pneumonia are more prone to suffer from heart and lungs related diseases when they are exposed to polluted air. It is stated that once contaminated air is inhaled, PM2.5 particles and PM10 particles which have entered to the human body are either extremely difficult or nearly impossible to be self-purified in the immune system of the human [3]. As mentioned in many studies lack of public awareness in Colombo about bad air quality is the main problem. Therefore, with the rapid increase in air pollution, air quality index has become a very important factor to predict and to make people aware of the effects of air pollution (AQI levels) may have on health and the environment in which they live [4]. Moreover, the adverse impact which is caused by air pollution on human health and the surrounding environment can be significantly reduced/minimized by predicting the AQI value. To identify the most suitable method to predict AQI value from available techniques, and the available dataset to achieve the highest accurate prediction are important accomplishments of this study. The continual human existence is based on air. The Air Quality Index is the index value that illustrates qualitatively the current state of the air. A substantial AQI will further threaten the health of all living beings and the atmosphere. From Central Environment (CEA) Authority, a historical air concentration dataset was gathered for this research, which includes hourly concentration levels of various air pollution factors & weather factors such as PM10, PM2.5, SO2, NO2, CO, O3, wind speed, wind direction, average temperature, relative humidity, and solar radiation. To guarantee the prediction outcome accuracy, a variety of data preparation approaches were used. An already preprocessed dataset was utilized to to cross-validation technique by dividing it into 80 percent for model training & 20 percent for testing. Random Forest, K Nearest Neighbors, Support Vector Machine, & Multiple Linear Regression are the machine algorithms that have been deployed as prediction models. The best-suited model for AQI prediction is chosen based upon the performance of the machine learning model & accuracy.

The paper is structured as follows. The second chapter of this study is the literature review of the research. The third and the fourth chapters of this study highlights the methodology and the findings of this research respectively. Finally, the conclusion of this study is presented in the final section.

II LITERATURE REVIEW

There are several AQI prediction solutions which are available. Out of these predictions, very few solutions provide general guidance to overcome the adverse effects air pollution has on human health which depends on the amount of PM10 & PM2.5 fine particulates in the atmosphere. A comprehensive literature review has been done to provide a brief overview on the existing literature on AQI predictions which ae made using different machine learning models due to the unavailability of one precise methodology to predict the within the domain of research AQI prediction. The study field differs not only in terms of methodologies and methods, but also in the terms of accessible datasets, which are frequently varied owing to the traffic, environmental factors, and climate of the chosen geographical region. Bad air quality is mostly caused by PM particulates. For example PM2.5 & PM10 like air pollutants cause air pollution in certain cities, whereas COx, SOx, and NOx are primarily responsible for polluted air in others. Due to these drawbacks, the comprehensive literature review of this study is carried out as an effort to gain a thorough grasp on the AQI prediction research scope and to locate relevant research studies that serves the same purpose as this study. The literature review chapter of this study consists of a collection of the most recent and relevant researches on AQI predictions. Here some of the current AQI prediction approaches are discussed. It's also crucial to consider which approach is best for predicting air pollution. AQI predictions which are made using a deep learning approach is one of the most widely utilized approaches out of the existing models. When it comes to predicting AQI, the most often used technique is machine learning. Machine learning approaches use a lot of data & machine learning-based algorithms to train the model. The deep learning-based neural network technique is another approach that may be used to forecast the AQI. With a basic neural network, correct predictions of the AQI can be made, and the model may be further modified using various testing settings and input parameters[7].

As per the literature of the study, there are a few flaws in the existing air quality index prediction methods, such as issues with dataset collection. The low accuracy of the predictions made by the available air quality index prediction models in Sri Lanka is a result of inaccurate and null data. Data preprocessing is another aspect that affect the decreased level of impacts accuracy decreases. In the light of these considerations, it is evident that the majority of Sri Lanka's existing systems have failed to provide accurate predictions. In view of the foregoing with other countries, they have been able to overcome these disadvantages and attain high accuracy.

Table 1. Literature Review Summary

Author	Application	Technique	Remark
S. Silva and others	Air quality prediction for smart cities	• Support vector regression	• Predict PM 2.5 levels variability. • Model is suitable for predict hourly air pollution. • Obtain an accuracy of 94.1%
Usha Mahalingam and others	Air Quality prediction	• Neural Networks • Support vector machine	• Accuracy of 91.62% for neural network • Accuracy of 97.3% for support vector machine
Min Lee and others	Air pollution prediction	• Deep Learning	• Predict against PM 2.5, PM 10 particulars. • Accuracy based on PM 10 is very low. • Accuracy based on PM 2.5 is very high.
Timothy M. A. and others	Air quality monitoring model development	• Naïve Bayesian • KNN • Support Vector Machines • Neural networks • Random Forest	• Highest accuracy was obtained through Neural Networks. • Sometimes Neural Networks leads to slower responses.
C. Zhao and others	Air Quality Index Prediction	• Linear regression	• AQI Prediction based on a year data of PM2.5, PM10, etc. • There is a deviation between predicted results and actual data.
Esmail Almadi	Air pollution prediction	• Data Mining • Decision Tree	• Used Clementine software for data clustering. • Data sample include climate data of 53 years
Colin Bellinger and others	A systematic review based on Machine Learning and data mining for Air Pollution	• Machine Learning Algorithms • Data Mining • Big Data	• Refer 400 research papers & reduce to 47 after the inclusion/exclusion criteria's • Divided research papers into three categories • End of the literature survey that highest accuracy levels always obtain in Machine Learning Algorithms based approaches.

III METHODOLOGY

The proposed approach consists of a sequence of phases for predicting the AQI. The sequence of phases includes collecting the dataset from Central Environmental Authority, pre-processing the collected dataset, analyzing the collected dataset to identifying the correlations among air pollution factors, applying appropriate ml algorithms, & ultimately selecting the most suitable machine learning approach & analyzing the prediction results.

A Data Collection & Pre-processing

Historical datasets containing information regarding air pollution factors' hourly concentration levels in Colombo are obtained from the Central Environment Authority(CEA) and the National Building Research Organization(NBRO). From January 2019 to February 2021, the dataset contains average concentrations of air & weather factors such as humidity, CO, SO2, NO2, PM10 & PM2.5. The obtained dataset is being pre-processed by using various preprocessing approaches to improve accuracy & assure the reliability of the values that have been predicted.

B Data Analysis

Correlation matrix and distribution charts are used to determine the correlations among air pollution variables, as well as to determine the dataset's distribution and nature. The RStudio software[11] is used for the data analysis. The most & least factors that are affected by PM2.5 can be identified using correlation matrix and distribution graphs.

C Evaluation

The Train-Test-Split technique is so far the most frequent approach used in the cross-validation of pre-processed data. The pre-processed dataset is divided into two sample sets, 80 percent of data is used to train the prediction machine learning model and 20 percent of data is used in evaluating the results that have been predicted.

D Training the Model

i. Random Forest
ii. Multiple Linear Regression
iii. Support Vector Machine
iv. K Nearest Neighbors
The dataset is trained using several Machine Learning algorithms. The default parameters have been used in each of these instances. Python[12] libraries such as pandas[13], scikit learn[14], & the PyCharm IDE[15] were used in the implementation.

E Model Evaluation

The machine learning model is utilized to predict the air quality index based on the previously pre-processed dataset. It is predicted once the machine learning model training stage is completed. The most appropriate machine learning algorithm has been identified based on the prediction accuracy of all the used machine learning algorithms. Accuracy has been calculated using the following equation.

i. $0 \leq PM2.5 \leq 30$: AQI Category 1(Good)

ii. 31$\leq PM2.5 \leq$ 60: AQI Category 2(Satisfactory)

iii. 61 $\leq PM2.5 \leq$ 90: AQI Category 3(Moderate)

iv. 91 $\leq PM2.5 \leq$ 120: AQI Category 4(Poor)

v. 121 $\leq PM2.5 \leq$ 250: AQI Category 5(Very Poor)

vi. 251 $\leq$ PM2.5: AQI Category 6(Severe)

IV RESULTS

Nearly 11000 data records of air pollutants and weather variables were obtained for this research, including PM10, PM2.5, SO2, NO2, CO, O3, Wind speed, Wind direction, Average Temperature, Relative Humidity, & Solar Radiation. As per the correlation matrix graph in Figure 2, which was computed using RStudio, the correlation between PM10 & PM2.5 is the best correlation value shared between two factors. When compared with other air and weather parameters, CO, NO2, SO2, and PM10 have the highest correlation values with PM2.5, as shown in the correlation matrix. As a result, we applied PM10, PM2.5, CO, SO2, and After a comprehensive correlation analysis of all the air parameters and weather factors, four air parameters were chosen from the selected dataset. SO2 concentration, NO2 concentration, PM10 concentration, and PM2.5 concentration NO2 parameters to train & evaluate the prediction machine learning model.

are these four parameters. Apart from CO, SO2, NO2, and PM10, correlations among PM2.5 and other weather parameters and air parameter concentrations are average, as shown in Figures 1,2 and 3. To obtain a better correlation value, the correlations between PM2.5 and multiple air parameters concentrations were computed, as shown in Table 2. PM2.5 and the collection of SO2, NO2, CO, and PM10 have a correlation of 0.8644, which is really a great value.

After a comprehensive correlation analysis of all the air parameters and weather factors, four air parameters were chosen from the obtained dataset. SO2 concentration, NO2 concentration, PM10 concentration, and PM2.5 concentration are the four parameters. Apart from CO, SO2, NO2, and PM10, correlations among PM2.5 and other weather parameters and air parameters concentrations are modest, as shown in Figures 1,2 and 3. To obtain a better correlation value, the correlations between PM2.5 and multiple air parameters concentrations were computed, as shown in Table 2. PM2.5 and the collection of SO2, NO2, CO, and PM10 have a correlation of 0.8644, which is really a great value.

Table 2. Multiple Correlation Summary

	PM2.5
PM10 + NO2	0.8635649
PM10 + NO2 + CO	0.8623043
PM10 + NO2 + SO2	0.8642185
PM10 + NO2 + SO2 + CO	0.8644263

A Multiple Regression Model

Using Multiple regression model Mean Absolute error, Mean Squared Error, Root Mean Squared Error & default accuracy score have been computed. All of these scores are average values, as shown in the sheet which is illustrated by Figure 4. The nature & the type of the collected dataset constantly influence these values which are predicted from the regression machine learning model. For this dataset and prediction procedure, multiple regression algorithm is not a suitable machine learning algorithm.

B Support Vector Machines Model

The SVM model has been able to achieve 84.68% accuracy, as shown by the classification details sheet in Figure 5. When compared to the regression method, this Support Vector Machine model has relatively very higher accuracy. The reason for the higher accuracy is because SVM handle input parameters with polynomial properties and SVM is suitable for this prediction work.

C Random Forest Model

The Random Forest machine learning algorithm was also deployed for this prediction. Random Forest is a supervised learning technique that could be used to solve both classification & regression-based problems, and through this module it is simple to calculate the relative significance of each feature that makes up the prediction. As seen in Figure 6 of the paper, the Random Forest model has been able to achieve 85.17% accuracy. When the Random Forest model accuracy is compared to the accuracy of the Multiple Regression approach, this 85.17% accuracy rate is quite a remarkable one.

D KNN Model

In numerous research projects, the KNN machine learning model has been utilized to predict the category of air quality index. When k = 3, the KNN prediction model obtained 83.25% accuracy, as shown by Figure 7 of the paper. It's able to get a higher accuracy from using this KNN prediction model although some air pollutant factors are very weak.

	AT	RH	SolarRad	RainGauge	WSRaw	WDRaw	O3	CO	NO2	SO2	PM2.5	PM10
AT	1.00	-0.90	0.79	-0.07	0.60	0.01	0.48	0.08	0.22	0.25	0.04	0.27
RH	-0.90	1.00	-0.79	0.12	-0.68	0.13	-0.59	-0.01	-0.23	-0.24	-0.10	-0.31
SolarRad	0.79	-0.79	1.00	-0.05	0.66	-0.15	0.40	0.03	0.21	0.19	0.05	0.23
RainGauge	-0.07	0.12	-0.05	1.00	-0.01	0.02	-0.04	0.04	0.05	0.00	-0.05	-0.07
WSRaw	0.60	-0.68	0.66	-0.01	1.00	-0.33	0.52	-0.05	0.16	0.14	-0.07	0.10
WDRaw	0.01	0.13	-0.15	0.02	-0.33	1.00	-0.30	0.07	0.07	0.10	0.02	0.05
O3	0.48	-0.59	0.40	-0.04	0.52	-0.30	1.00	-0.24	-0.19	-0.03	0.01	0.01
CO	0.08	-0.01	0.03	0.04	-0.05	0.07	-0.24	1.00	0.50	0.19	0.48	0.48
NO2	0.22	-0.23	0.21	0.05	0.16	0.07	-0.19	0.50	1.00	0.40	0.54	0.63
SO2	0.25	-0.24	0.19	0.00	0.14	0.10	-0.03	0.19	0.40	1.00	0.24	0.32
PM2.5	0.04	-0.10	0.05	-0.05	-0.07	0.02	0.01	0.48	0.54	0.24	1.00	0.86
PM10	0.27	-0.31	0.23	-0.07	0.10	0.05	0.01	0.48	0.63	0.32	0.86	1.00

Figure 1 : Correlation Matrix Chart

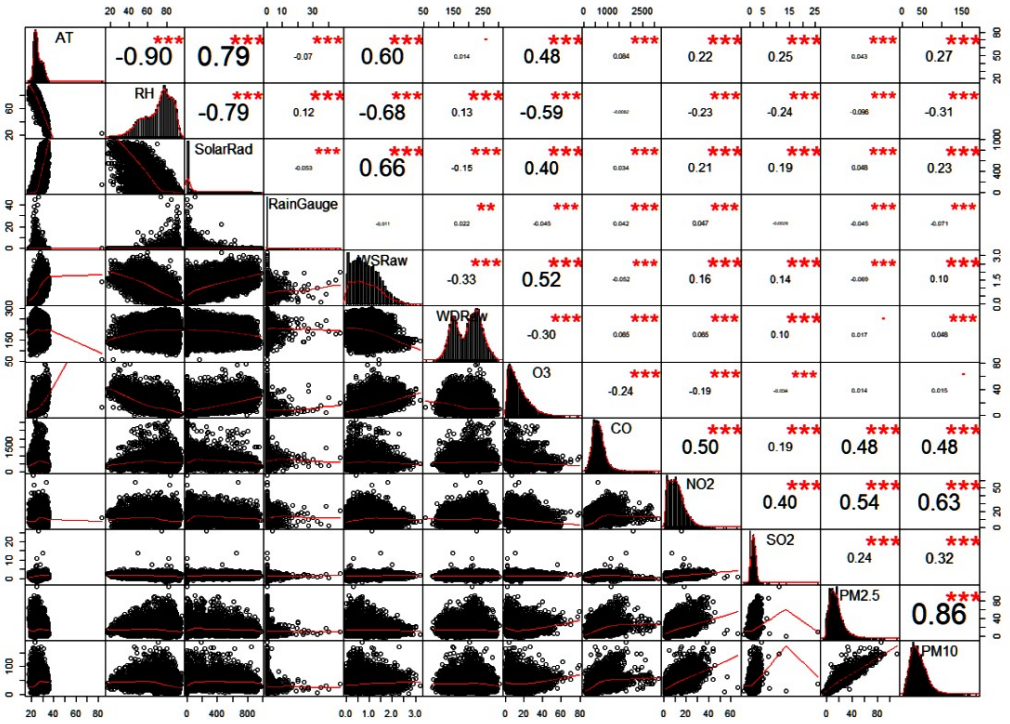

Figure 2 : Correlation Matrix

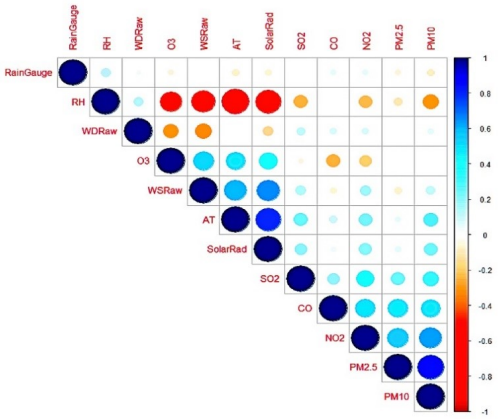

Figure 3 : Correlogram

Mean Absolute Error: 0.2694807929045523
Mean Squared Error: 0.15263856450175747
Root Mean Squared Error: 0.39068985717798904
Default Accuracy Score: 41.90964213392034

Figure 4 : Multiple Regression Classification

V DISCUSSION

Using Random Forest, Support Vector Machines, and K-Nearest Neighboring algorithms, several existing international researches have achieved high accuracies, as shown in Table 1. The incompleteness in some records within the collected dataset is the major factor that has contributed to the poor accuracy which is produced by the Multiple Regression machine learning model. Another reason that affected low accuracy is the nature of the

Default Accuracy Score:
84.68021814576103

	precision	recall	f1-score	support
0	0.89	0.96	0.92	1601
1	0.60	0.48	0.53	350
2	0.69	0.16	0.26	57
3	0.00	0.00	0.00	9
accuracy			0.85	2017
macro avg	0.54	0.40	0.43	2017
weighted avg	0.83	0.85	0.83	2017

Figure 5 : SVM Classification

Accuracy Score:
85.17600396628656

	precision	recall	f1-score	support
0	0.91	0.94	0.92	1601
1	0.60	0.57	0.59	350
2	0.47	0.25	0.32	57
3	0.00	0.00	0.00	9
accuracy			0.85	2017
macro avg	0.49	0.44	0.46	2017
weighted avg	0.84	0.85	0.84	2017

Figure 6 : Random Forest Classification

Accuracy: 0.8325074331020813

	precision	recall	f1-score	support
0	0.91	0.93	0.92	807
1	0.49	0.53	0.51	161
2	0.50	0.18	0.27	38
3	0.00	0.00	0.00	3
accuracy			0.83	1009
macro avg	0.48	0.41	0.43	1009
weighted avg	0.83	0.83	0.83	1009

Figure 7 : KNN Classification

regression machine learning algorithms. The performance of the algorithms is affected by the lack of data records and some noisy factors in the pre-processed dataset. In an effort to improve the quality & the performance of the obtained dataset, several data pre-processing approaches were used.

The Random Forest machine learning model is the best-fitted model for this air quality index prediction process since Random Forest obtained the best accuracy rate when compared with the Support Vector Machines, Multiple Regression, and K-Nearest Neighboring models.

Table 3. Model Evaluation Summary

Model	Accuracy
Multiple Regression	41.90\%
SVM	84.68\%
Random Forest	85.17\%
KNN	83.25\%

VI CONCLUSION

Air is always considered as the main critical factor on which human survival depends on. The AQI or long firmly air quality index is the index value that illustrates qualitatively the current state of the air. The substantial AQI will further threaten the living creatures' health & the living atmosphere. Terrible air quality has been a major concern in Sri Lanka, particularly in urban cities such as Colombo and Kandy. Reliable Air Quality Index prediction will assist in decreasing the health risks caused by air pollution. To determine the most affected air pollutant concentrations air quality index prediction correlation analysis has been done. In this study after the comprehensive correlation analysis, PM2.5 concentration is predicted in Colombo using four correlated air pollutant concentrations such as SO2, NO2, PM2.5, & PM10. The obtained dataset was pre-processed via prediction in order to improve prediction accuracy. The gathered dataset Cross-validated as according to 80% for training & 20% for testing the prediction model. Machine learning methods such as K-Nearest Neighboring, Multiple Linear-Regression, Random Forest, and Support Vector Machines were used to train and evaluate the prediction models. In the end, we achieved 83.25% accuracy for the K-Nearest Neighboring algorithm model, 84.68% accuracy for the Support Vector Machines model, 85.17% accuracy for the Random Forest model, and 41.9% accuracy for the Multiple Regression Model. Random Forest was recognized as the best and the most appropriate prediction model after evaluating all the models. Under the circumstances with limited data, the model had over 85% greater accuracy.

VII FUTURE WORK

In the future, more datasets in Colombo from ambient air quality parameters monitoring stations inside Sri Lanka are expected to be collected and more appropriate pre-processing techniques for the dataset will be used. Since the Multiple Regression machine learning algorithm is inaccurate, the study team intends to develop a deep learning-based prediction model to calculate the air quality index prediction. Moreover, since the current prediction is based only on Colombo, the research team plans to predict the Air Quality Index in other districts as well in the future.

REFERENCES

[1] S. Mahanta, T. Ramakrishnudu, R. R. Jha, and N. Tailor, "Urban Air Quality Prediction Using Regression Analysis," in TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, Oct. 2019, pp. 1118–1123. doi: 10.1109/TENCON.2019.8929517.

[2] M. Castelli, F. M. Clemente, A. Popovič, S. Silva, and L. Vanneschi, "A Machine Learning Approach to Predict Air Quality in California," Complexity, vol. 2020, pp. 1–23, Aug. 2020, doi: 10.1155/2020/8049504.

[3] S. Zhong, Z. Yu, and W. Zhu, "Study of the Effects of Air Pollutants on Human Health Based on Baidu Indices of Disease Symptoms and Air Quality Monitoring Data in Beijing, China," IJERPH, vol. 16, no. 6, p. 1014, Mar. 2019, doi: 10.3390/ijerph16061014.

[4] Y. L. S. Nandasena, A. R. Wickremasinghe, and N. Sathiakumar, "RAesierarpchoarltilculetion and health in Sri Lanka: a review of epidemiologic studies," p. 14, 2010.

[5] T. Xayasouk and H. Lee, "AIR POLLUTION PREDICTION SYSTEM USING DEEP LEARNING," Naples, Italy, Jun. 2018, pp. 71–79. doi: 10.2495/AIR180071.

[6] D. Iskandaryan, F. Ramos, and S. Trilles, "Air Quality Prediction in Smart Cities Using Machine Learning Technologies based on Sensor Data: A Review," Applied Sciences, vol. 10, no. 7, p. 2401, Apr. 2020, doi: 10.3390/app10072401.

[7] S. Sampath, "Air Quality Analysis & Prediction," 2019, doi: 10.13140/RG.2.2.14624.23040.

[8] C. Bellinger, "A systematic review of data mining and machine learning for air pollution epidemiology," p. 19, 2017.

[9] U. Mahalingam, K. Elangovan, H. Dobhal, C. Valliappa, S. Shrestha, and G. Kedam, "A Machine Learning Model for Air Quality Prediction for Smart Cities," in 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Chennai, India, Mar. 2019, pp. 452–457. doi: 10.1109/WiSPNET45539.2019.9032734.

[10] T. M. Amado and J. C. Dela Cruz, "Development of Machine Learning-based Predictive Models for Air Quality Monitoring and Characterization," in TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), Oct. 2018, pp. 0668–0672. doi: 10.1109/TENCON.2018.8650518.

[11] RStudio. Boston, MA: RStudio, Inc., 2015. Accessed: Aug. 17, 2021. [Online]. Available: http://www.rstudio.com/

[12] python. Scotts Valley, CA: Van Rossum, Guido and Drake Jr, Fred L, 1995. Accessed: Sep. 21, 2021. [Online]. Available: https://www.python.org/

[13] Pandas. The pandas development team, 2020. Accessed: Sep. 21, 2021. [Online]. Available: https://pandas.pydata.org/

[14] scikit Learn. Pedregosa, F. and others, 2011. Accessed: Sep. 21, 2021. [Online]. Available: https://scikit-learn.org/stable/

[15] Pycharm. JetBrains IntelliJ, 2017. Accessed: Sep. 21, 2021. [Online]. Available: https://www.jetbrains.com/pycharm/

ACKNOWLEDGMENT

The corresponding author wishes to express gratitude to the supervisors for their dedication and guidance during this research project.

AUTHOR BIOGRAPHIES

Mr. RM Fernando is a Computer Science undergraduate at the Department of Computer Science, Faculty of Computing, KDU.

Ms. WMKS Ilmini is a Lecturer at the Department of Computer Science, Faculty of Computing, KDU.

Ms. DU Vidanagama is a Lecturer at the Department of Computer Science, Faculty of Computing, KDU.