Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 7, S. 276-282, 6, 2023

**Araştırma Makalesi**

IJANSER

https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 7, pp. 276-282, 6, 2023

**Research Article**

# DEVELOPMENT OF AN INTERNET OF THINGS BASED AIR QUALITY MONITORING SYSTEM USING MACHINE LEARNING

Abdulrasheed O. Abdulganiyu [1*], Jonathan G. Kolo [1] and Abraham U. Usman [2]

*[1]Department of Electrical and Electronics Engineering, Federal University of Technology, Minna, Nigeria*
*[2]Department of Telecommunications Engineering, Federal University of Technology, Minna, Nigeria.*

*[*](yemic002@gmail.com) Email of the corresponding author*

*(Received: 10 July 2023, Accepted: 24 July 2023)*

(5th International Conference on Applied Engineering and Natural Sciences ICAENS 2023, July 10 - 12, 2023)

**ATIF/REFERENCE:** Abdulganiyu, A. O., Kolo, J. G. & Usman, A. U. (2023). DEVELOPMENT OF AN INTERNET OF THINGS BASED AIR QUALITY MONITORING SYSTEM USING MACHINE LEARNING. *International Journal of Advanced Natural Sciences and Engineering Researches*, 7(6), 276-282.

***Abstract*** – Air pollution and its negative impacts on human health have become serious concerns in many places throughout the world. The traditional methods of monitoring air quality, such as manual sampling and laboratory analysis, are time-consuming, expensive, and may not provide real-time information. In this study, an IoT-based Air Quality Monitoring System that uses Machine Learning to provide accurate and timely analysis of air quality data is presented. The system collects data from a network of sensors measuring various air quality parameters, processes the data using ML algorithms to identify patterns and predict future conditions, and provides insights into the current state of the environment. The findings showed that the emissions had an inversely proportional impact on air quality in the study region and the system achieved an accuracy of 0.978. This study has the potential to provide accurate and timely analysis of air quality data and regulate air quality in real-time.

*Keywords –Air Quality Monitoring, Decision Tree, Internet of Things, Machine Learning, ThingSpeak*

## I. INTRODUCTION

The purity of the Earth's atmosphere cannot be overstated because it adds to the planet's survival. Air pollution, a problem that trends with industrialization and civilization [1], has been identified as a key cause of worry over the years, as it impacts a deterioration in human health and adds to the earth's unbalanced nature [2], [3] Air quality monitoring refers to the process of measuring and tracking the levels of pollutants in the air. It plays a critical role in understanding the state of the environment and ensuring public health [4]. Air quality monitoring is typically performed using a network of sensors that measure various parameters such as temperature, humidity, and levels of pollutants such as particulate matter, ozone, nitrogen dioxide, and sulphur dioxide. The data collected is analysed to determine the levels of pollutants in the air and identify trends and patterns over time [5]. The insights obtained from air quality monitoring can be used to inform air quality management decisions, including the implementation of control measures to reduce pollution levels and protect public health. Effective air quality monitoring requires accurate and reliable data, and the use of advanced technologies such as the Internet of Things (IoT) and Machine Learning (ML) algorithms can help to achieve this goal.

In a study by [6] a system that uses semiconductor sensors was designed to monitor CO, $CO_2$, $NO_2$, and $SO_2$. These sensors include the MQ7, which detects CO, and the MQ135, which detects NH3, CO, and benzene. The ESP8266 was utilized in the design to send data, while the Raspberry Pi was employed as the main processing unit to display the outcome of the pollutant in the environment. The MEAN stack was utilized to present the data obtained on the WEB. While the model presented here is a beautiful approach of presenting data obtained on the web, these data weren't uploaded in real-time.

Reference [7] harnessed latent knowledge using machine learning. MQ7 and MQ135 were utilized in the IoT deployment to collect data on carbon monoxide and air quality, respectively. The findings of their research provided the government of India with justification to ban one-stroke and two-stroke engines in order to limit the emission of dangerous gases. As compared to the air quality index chart presented in Table 2.3, they observed air quality to be greater than 300 parts per million (PPM) on average. The air quality was predicted using linear regression. The Arduino Uno controller and the ESP8266 gateway were utilized in the design. Everything was saved on ThingSpeak. The demerit noted in this work was the inability of the authors to monitor more pollutants such as particulate matter or LPG, this would have given a more comprehensive result about the air quality index in the environment.

In order to track indoor air pollution, [1] developed and IoT based indoor air quality monitoring system. The cloud-based system was created utilizing IoT technology to detect the concentration of aerosol, CO, $CO_2$, volatile organic compound (VOC), temperature, and humidity in order to evaluate air quality. Data were examined using cloud computing, which is incorporated on the web. The emphasis on interior air quality stems from the United States Environmental Protection Agency's confirmation that indoor air is more polluted than outdoor air. Moreover, toxic indoor air can induce respiratory illness 1000 times more than contaminated outside air. DHT11 was used to monitor temperature and humidity. An indoor air quality monitoring system will mostly find application among laymen as such the authors could have used ML to classify the level of safety of these indoor pollutants to enable give uninformed users better understanding of their air quality.

To extract necessary information and translate it into pollution concentrations, a mathematical model with sensor fusion was applied [8]. The concentrations of CO, $NO_2$, and NOx were calculated using machine learning in this study. SVM, DNN, and RN are among the approaches employed. Concentrates were produced from the acquired multi-sensor data at the conclusion of the investigation. The desired outputs of the model are provided by reliable but expensive measurements of the pollutants. Mathematical models may not always be accurate in calculation of concentration models

Reference [9] focused on monitoring hazardous gases produced by companies to ensure the decrease of these harmful gases. $CO_2$, $NH_3$, Benzene, Sulphur Dioxide, Nitrogen Oxide, and Carbon Monoxide were among the gases studied. MQ-3, MQ-6, MQ-2, and MQ-135 sensors were used to monitor the amount of these gases in the environment. Additionally, the Arduino was utilized to compute these values, with the Raspberry Pi serving as the gate way. In addition, VB.net was utilized to construct the graphical user interface. The authors could have taken a step further to classify these gases based on their pollutant concentration using ML, this would have added novelty to the literature.

After reviewing the aforementioned works and carefully studying their limitations, there is need to proffer solutions to the gaps identified in literature, this study models a simple and cost-effective system that provides solutions to the aforementioned challenges by integrating Internet of Things (IoT) and Machine Learning (ML) technologies in the development of a real-time Air Quality Monitoring System able to monitor numerous pollutants at the same time by collecting, processing, and analyzing data from a network of sensors and classifying the level of safety of the gases based on their pollutants concentration. To this aim, this study introduces an IoT-based Air Quality Monitoring System using ML.

## II. Materials and Method

### *A. System Overview*

Fig. 1 depicts the architecture of the system used in this study. The first layer comprises various sensors required to read the pollution concentration in the air. This layer also includes the device used to accomplish the desired network (GSM module) and the communication protocol. In other words, the first layer is made up of hardware, often known as nodes, a gateway that facilitates internet connections, and the communication protocol. As a result, the layer facilitates the collection and processing of data that describes each pollutant. The following layer, the second layer, is made up of data collecting, often known as the cloud. Here is where the information will be saved. The cloud was selected over all other data storage media because it was limitless and available at any time without regard for proximity. In other words, access might be granted regardless of the distance from the node. The third layer contains methods that enable the system to be smart. This entails using artificial intelligence to help forecast whether or not the studied environment is safe. This also provides the prediction's percentage accuracy.

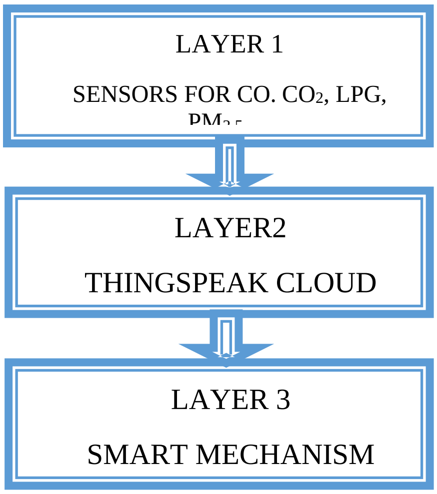


Fig. 1 Architecture of the system

### *B. IoT platform development*

The storage of data in the ThingSpeak platform was initiated through the creation of a Gmail account. Subsequently, access to the ThingSpeak web interface was obtained via the website ThingSpeak.com. The aforementioned Gmail account was employed to establish and authenticate an account on ThingSpeak.com. In the newly created account, a channel comprising four fields was generated and preserved for future use. The complete sequence of these events is illustrated in Fig. 2.

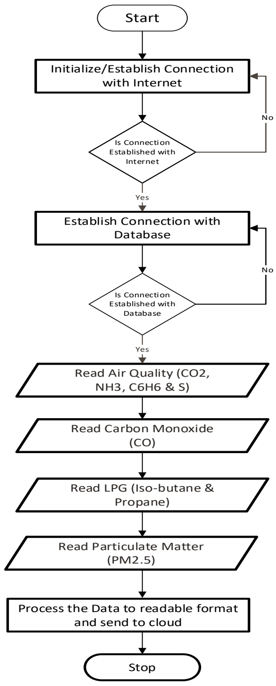


Fig. 2 Flowchart of IoT platform development

### *C. Machine learning model development*

The study focuses on ensuring the safety of individuals within their environment through the classification of data obtained from the system. This classification process involves the implementation of a decision tree algorithm using Python as the programming language. The sequence of steps employed in this process is depicted in Fig. 3. The initial phase of the process involves the reading of data using the pandas library. The data is subsequently processed using the same library and then split into training and test sets. The training set comprises 70% of the total data, while the test set is 30% of the overall data. The decision tree classifier is then trained using the training set. Finally, the accuracy of the classification is evaluated by testing the classifier on the test set.

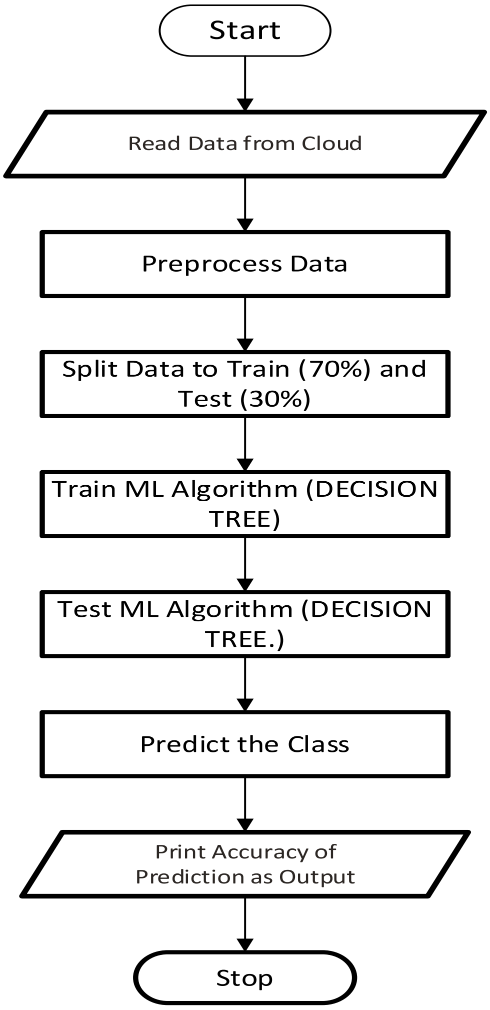


Fig. 3 Flowchart of machine learning model

The block diagram of the system prototype is presented in Figure 4. The system uses a network of sensors to collect data on various air quality parameters such as Carbon dioxide (CO2), Carbon monoxide (CO), Liquid Petroleum Gases (LPG) and Particulate Matter less than 2.5 microns (PM2.5). The data collected is processed and analysed using the ML algorithm resident on the Arduino Uno microcontroller. The data is also transmitted to the cloud via the gateway. The circuit diagram showing the interconnectivity of these devices is presented in Fig. 5.

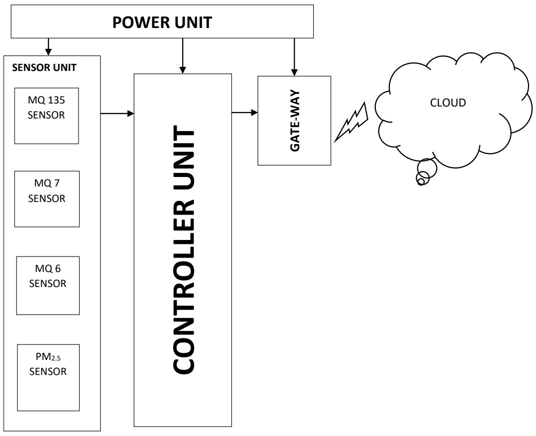


Fig. 4 Block diagram of the system prototype

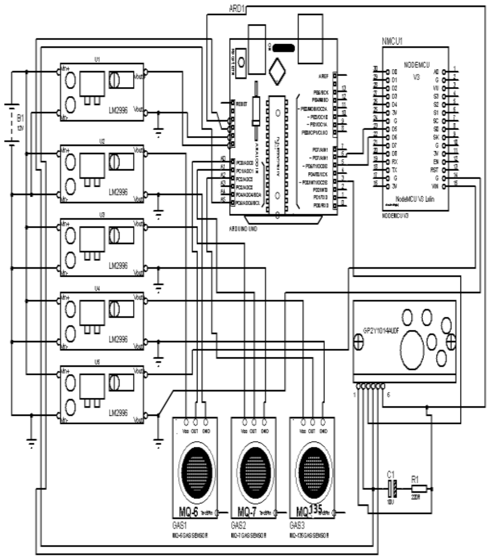

Fig. 5 Circuit diagram of air quality monitoring system

## III. RESULTS

The study was conducted within Minna Metropolis, Niger State, Nigeria. Four major activity centres were selected for this study, namely Kpakungu Roundabout, Mobil Roundabout, Tunga Roundabout, and OBJ Complex Roundabout

In each study location, four parameters (Air Quality, Carbon Monoxide, LPG and Particulate Matter) were continuously measured over a time interval when the centres were busy. These parameters include air quality, carbon monoxide, liquefied petroleum gas and particulate matter. The measurement of these parameters provides a comprehensive understanding of the air quality status in each location. Fig. 6 to Fig. 9 shows a graphical description of Part Per Millions (PPM) against Entry ID of each location.

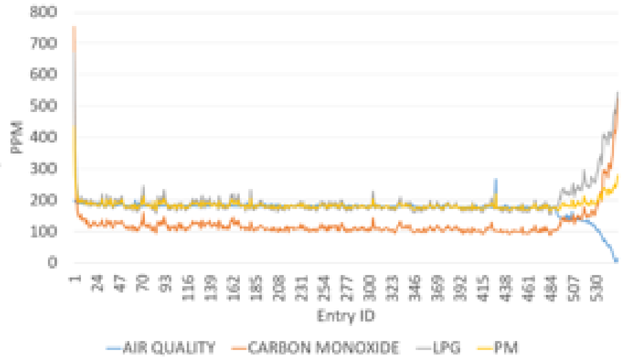


Fig. 6: Air Quality Monitoring in OBJ Complex Roundabout

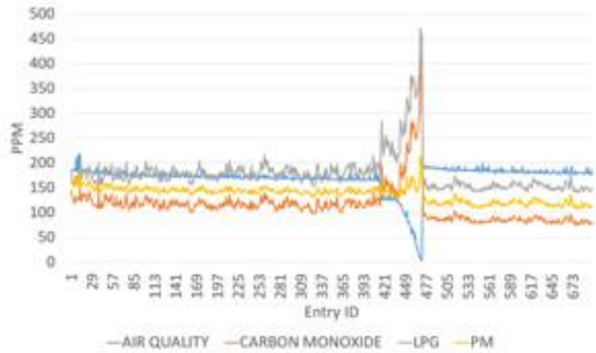


Fig. 7: Air Quality Monitoring in Mobil Roundabout

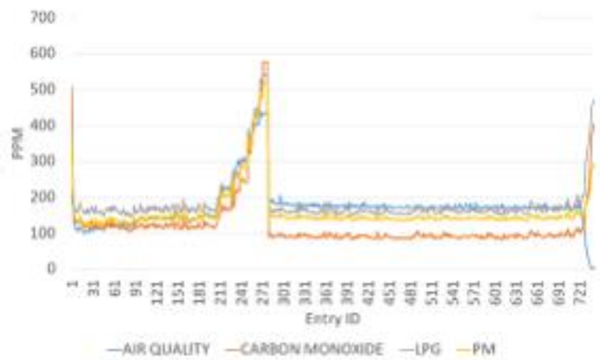


Fig. 8: Air Quality Monitoring in Kpakungu Roundabout

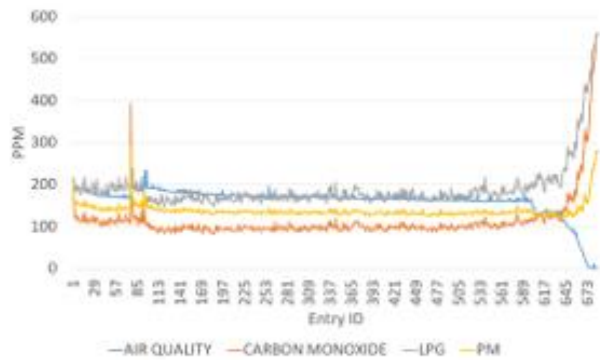


Fig. 9: Air Quality Monitoring in Tunga Roundabout

## IV. DISCUSSION

The dataset consisted of 2682 measured data in Parts Per Million (PPM) of air quality, carbon monoxide, LPG and Particulate Matter. The objective of the study was to evaluate the performance of a decision tree algorithm in predicting air quality based on sensor data. Before applying the decision tree algorithm, the dataset was pre-processed to handle missing values and categorical data. The dataset was then split into a training set (70%) and a test set (30%) for model training and evaluation.

Table 1. Air Quality Index

| Range (PPM) | Status |
|---|---|
| 0-50 | Good |
| 51-100 | Moderate |
| 100-150 | Unhealthy for sensitive people |
| 151-200 | Unhealthy |
| 201-300 | Very unhealthy |
| 301-500 | Hazardous |

A class column with values for 'Good', 'moderate', 'unhealthy for sensitive people', 'unhealthy', 'very unhealthy', and 'hazardous' and its corresponding PPM range is shown in Table 1 above. This index describes the classification air quality based on its pollutants concentration.

The decision tree algorithm was trained on the pre-processed dataset using the training set. Fig. 10 presents the decision tree obtained from the data.

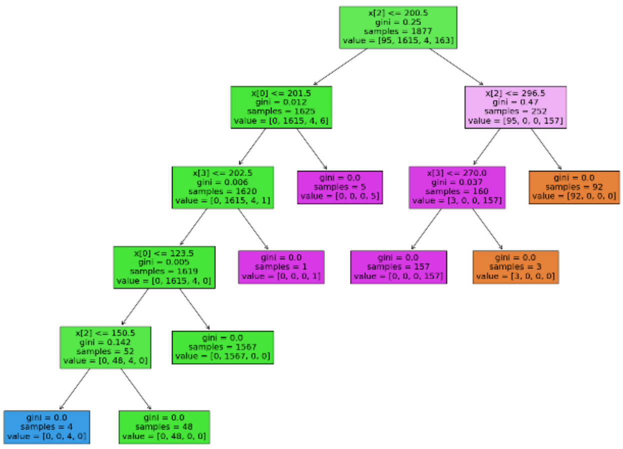


Fig.10: Decision Tree from the training dataset

The performance of the model was evaluated using precision, recall, accuracy, and F1 score metrics. The model achieved an accuracy of 0.9937 on the test set, indicating that the model was able to correctly predict the air quality category for 99.37% of the test instances. The precision and recall of the model were 0.994 and 0.9937, respectively, indicating that the model was able to correctly identify the true positives and avoid false positives and false negatives. The F1 score of the model was 0.9937, which is a harmonic mean of precision and recall, indicating a good balance between precision and recall. A summary of these results is presented in the confusion matrix shown in Fig. 11.

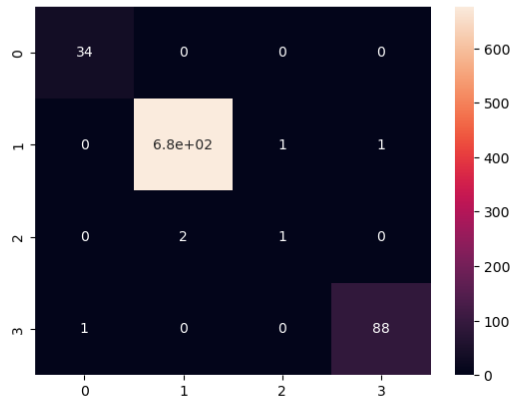


Fig. 11 Confusion Matrix of Decision Tree

From the confusion matrix, it is observed that there were 34 instances correctly classified as 'Good' air quality and there were no instances misclassified as 'Good' air quality. Similarly, for the 'Moderate' air quality class, 680 instances were correctly classified and there were only 4 misclassified instances as 'Unhealthy'. The third class, 'Unhealthy for sensitive people', had only 2 correct classifications while 0 were misclassified as 'Moderate'. The final class, 'Very unhealthy', had 88 correct classifications, while 1 was misclassified as 'Moderate'

## V. CONCLUSION

This study aimed to design and implement a comprehensive air pollution monitoring system that measures various pollutants, transmits data in real-time and classifies the level of safety in a location based on pollutant concentrations. To achieve this, the first objective was to design a system that measures Carbon Monoxide (CO), Air Quality (Carbon Dioxide (CO2), Particulate Matter (PM) and Liquefied Petroleum gas (LPG) in the atmosphere. This objective was met by developing a system that utilized sensors to measure the pollutants and generate data. The second objective was to design an IoT-based system that transmits the measured pollutant concentration in real-time to an online platform. This objective was achieved by integrating the system developed in objective one with an IoT module and Thingspeak platform. The third objective was to design a system that classifies the level of safety of a location based on the pollutant concentration using Machine learning. This objective was accomplished by using the decision tree algorithm to classify pollutant concentrations into safe, fairly safe, and unsafe levels. The fourth objective was to implement the systems designed in objectives one, two, and three. This objective was achieved by developing a prototype of the air pollution monitoring system and testing it in four study locations within Minna Metropolis in Niger State. The final objective was to appraise the performance of the system by evaluating its recall, F1-score, accuracy, and precision. The evaluation showed that the system was stable and had high recall, F1-score, accuracy, and precision in measuring pollutant concentrations and classifying the level of safety in a location based on the pollutant concentration.

In summary, this study successfully designed and implemented a comprehensive air pollution monitoring system that measures various pollutants, transmits data in real-time, and classifies the level of safety in a location based on pollutant concentrations. The system was found to be stable and had high sensitivity, accuracy, and precision in measuring pollutant concentrations and classifying the level of safety in a location based on the pollutant concentration. The developed system has the potential to contribute to efforts aimed at mitigating air pollution and improving air quality. Future works will focus on the development of a central monitoring dashboard based on IoT that can observe a wider geographical location and the use of hybrid machine learning techniques to ascertain the performance in air quality monitoring.

## ACKNOWLEDGMENT

The first author acknowledges the role played by the second and third authors and everyone who made this study a reality.

## REFERENCES


[1] Jo, J., Jo, B., Kim, J., Kim, S., & Han, W. (2020). Development of an IoT-based indoor air quality monitoring platform. *Journal of Sensors*, *2020*, 1-14.

[2] Gupta, K., & Rakesh, N. (2018). IoT Based Automobile Air Pollution Monitoring System. 2018 *8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 14–15.

[3] Salam, A., & Salam, A. (2020a). Internet of things for environmental sustainability and climate change. *Internet of Things for sustainable community development: Wireless communications, sensing, and systems*, 33-69.

[4] Rollo, F., Bachechi, C., & Po, L. (2023). Anomaly Detection and Repairing for Improving Air Quality Monitoring. *Sensors,* 23(2), 640-661.

[5] Wu, C. L., Song, R. F., Zhu, X. H., Peng, Z. R., Fu, Q. Y., & Pan, J. (2023). A hybrid deep learning model for regional $O_3$ and $NO_2$ concentrations prediction based on spatiotemporal dependencies in air quality monitoring network. *Environmental pollution*, 121075

[6] Chattopadhyay, M. K. (2019). An IoT Based Low-Cost Air Pollution Monitoring System. *Proceeding International conference on Recent Innovations is Signal Processing and Embedded Systems (RISE -2017)*, 27-29

[7] Ramasubbareddy, S., Bharath, K., & Sai, K. (2019). IOT based air quality monitoring system using mq135 and mq7 with machine learning analysis. *Scalable Computing: Practice and Experience*, 20(4), 599–606. https://doi.org/10.12694/scpe.v20i4.1561

[8] Vidnerová, P. (2021). Air Pollution Modelling by Machine Learning Methods. *Modelling, 2(4)* 659–674.

[9] Mir, A., & Ajitkumar, K. (2018). Sensing Harmful Gases in Industries using IOT and WSN. *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 1–3.