Air Pollution Mapping with Sensor-based Methodology

W.G.C.N. Egodagama
Department of Information Technology
Sri Lanka Institute of Information
Technology
New Kandy Road, Malabe, Sri Lanka
it18187960@my.sliit.lk

A.A.M. Zahid
Department of Information Technolog
Sri Lanka Institute of Information
Technology
New Kandy Road, Malabe, Sri Lanka
it18042320@my.sliit.lk

G.P.T.S. Pathirana
Department of Information Technology
Sri Lanka Institute of Information
Technology
New Kandy Road, Malabe, Sri Lanka
it18095418@my.sliit.lk

Dissanayaka D.S.
Department of Information Technology
Sri Lanka Institute of Information
Technology
New Kandy Road, Malabe, Sri Lanka
it18112092@my.sliit.lk

Bhagyani Chathurika
Department of Information Technology
Sri Lanka Institute of Information
Technology
New Kandy Road, Malabe, Sri Lanka
bhagyani.c@sliit.lk

Ravi Supunya
Department of Information Technology
Sri Lanka Institute of Information
Technology
New Kandy Road, Malabe, Sri Lanka
ravi.s@sliit.lk

Abstract—the purpose of this study is to develop a sensor-based methodology(S-BM) for mapping air pollution (AP) related to Gaseous Composition of the Atmosphere in a specific area. It uses a drone equipped with sensors to identify the current composition of the air. After self-identifying the locations with specific distances in a specific area, the drone can go to those locations automatically and obtain sensor readings related to the gas percentages at those locations. After that the data is then transmitted to a computer program which analyzes (cluster analysis methodology), the data and then maps the air pollution in that specific area. Our results provide important informa-tion on how to measure, manage and atmospheric pollution mapping (APM). It also helps to identify air-polluted areas that need to be addressed quickly, and, thereby, it helps to save the atmosphere. We hope to program to get the sensor reading sand analyze the data with a suitable methodology and predict the condition of the atmosphere in the specific area. We hope to use cluster analysis and other analysis methodologies and technologies to this function. We need a dataset to train the model that can do the air quality prediction (AQP) of the relevant area. For that, we surfed the internet and found some datasets regarding the air pollution level of some major countries and their capitals. We think we will be able to make the model by using these datasets and predict the air pollution level of a specific area clearly. In addition, we are going to predict the future AP levels in a specific area by analyzing the current gas percentages of some specific gas components in the atmosphere like CO, CO2, SO2 and NH3 etc.

Keywords—**AP-Air Pollution, S-BM-Sensor Based Methodology, APM-Atmospheric Pollution Mapping, AQP-Air Quality Prediction**

I. INTRODUCTION

A. Background

Air pollution (AP), is a major environmental health problem infect people in the world. Although air pollutants are many, the most important are particle pollution, ground-level ozone (O3), carbon monoxide (CO), sulfur oxides (SO2, SO3), nitrogen oxides (NO, NO2), and lead (Pb) which are found in the ambient air. PM, CO, SOx, NOx, Environmental Tobacco Smoke, formaldehyde (aldehyde category) and poly cyclic organic matter are found indoors[1, 2].

Uncovering to air pollutants leads to a variety of health effects depending on the type of pollutant, amount of the pollutant Uncover to, duration and frequency of exposure, and effect toxicity of the specific pollutant. Air pollution is glowingly documented as an ultimatum to public health in developing countries. Estimation of current air quality levels, regulatory standards and scientific literature on outdoor and indoor pollution of air, and health effects are very much important to identify the burden, develop and implement mediation and to fill knowledge gaps in Sri Lanka. It effect to adverse health effects such as cancer, cardiovascular diseases and high mortality rates. As well as scientists want to go to uncommon areas for find new invention. If they go there without knowing the area pollution then they faced many problems even death. If they know about the drone invention they can check whether the area is polluted or not from our drone.

There are several population groups are more vulnerable to the effects of AP. Those who are innately susceptible more than others, those who become susceptible because of environmental, social and personal behaviors, and those who are simply exposed to unusually large amounts of AP[3]. There are some groups that are more sensitive to air pollutants include unborn and young children, elderly people, and those with a history of cardio-respiratory diseases [3], as well as even death. Hence, the demographic profile of a given population is very important. Because of these problems, we have to get a solution for this AP. First of all, we have to identify what are the polluted areas. In this case, faced how to deal with restrict areas without going. As a solution for this problem, it is important to create a sensor-based drone with an autonomous system.

B. Literature review

Air pollution has become a growing threat to human health around the world, particularly for those who live in urban areas. Because it is invisible, it is a dangerous and quiet hazard in contrast to other dangers. Air quality levels were understood as an Air Quality Index (AQI) and analyzed using GIS software to calculate air pollution levels and assess the risk of poisoning. The level of pollution has increased year by year lot of factors like the increase in population, increased vehicle use, industrialization and urbanization which results in harmful effects on human lives by directly affecting health of population [4]. We recommend creating a sensor-based approach for mapping air pollution in relation to the Gaseous Composition of the Atmosphere in a given region. It also makes use of a drone with sensors to determine the current composition of the air. After self-identifying locations in a particular area with specific distances, the drone will automatically go to those locations and collect sensor readings related to those locations.

Harmful effects of pollution include mild allergic reactions such as irritation of the eyes and nose as well as some serious problems. Bronchitis, heart diseases, pneumonia, lung and aggravated asthma are the diseases that caused by air pollution. According to survey, due to air pollution 50,000 to 100,000 premature deaths per year occur in some countries [6]. The main objective of this project is to monitor the air quality in industrial and urban areas. The proposed outline includes a set of gas sensors (CO, and Methane and Propane) that are positioned on masses and structure of an IOT (Internet of things) and a dominant server to support both short-range real-time incident management and a continuing deliberate planning [3].

Air pollution can be defined as an alteration of air quality (AQ) that can be characterized by measurements of chemical, biological or physical pollutants in the air. Therefore, AP means the undesirable presence of impurities or the abnormal rise in the proportion of some constituents of the atmosphere [16]. AP is now the biggest environmental risk for early death, responsible for more than 6 million premature deaths each year from heart attacks, strokes, diabetes and respiratory diseases. It effects more than the deaths from AIDS, tuberculosis and malaria combined. There are some group range effect the AP with huge ways. Such as, Children, the elderly, people with existing diseases, and minority and low income communities are particularly vulnerable to adverse health outcomes and economic impacts. As examples missed workdays, from exposure to AP. In addition, recent studies show AP can affect mental health, worker productivity and even stock market performance. We can understand the best way to develop solutions, it is important to best understand this invisible threat.

There is currently no way for human sensors to detect air pollution. In addition, we do not have the senses to do that. Therefore, there is a need to create a sensor-based drone. The best way to detect air pollution is to send it autonomously Through a drone like this without any risk. Air pollution is harmful to the environment and living beings. It is difficult for us to reach in to the places where there is air pollution. There is a risk of going to new places without knowing the air condition is good or harmful.

When taking these points under consideration it is useful to have a sensor Base drone to check the condition of the air whether it is polluted or not. The main process of the proposed system of Air pollution mapping with sensor based drone system is to go to places where human cannot reach and check for air pollution and giving out an analyzed data. For analyzing data that were taken from the sensors, the collected data will undergo through a process and give out clean data. The clean data will go through a process of an algorithm to give out a report of the data as a percentage of all necessary gases in the atmosphere to identify whether the air is polluted or not. The details of the data is predicted and analyzed through a developed equation, which is used in the algorithm. To bring awareness among the public about the air pollution and the system a mobile application is developed. The public can get the information about the place and whether the air is polluted in that place as a percentage, which will be helpful to the user that like to travel for new places.

Air pollution has caused economic, environment and mankind severe damage which cannot be measured. This has lead to measure the air quality using environmental drones where this could help to monitor 24*7 air quality within given location. The E-drone flies up to a programmed location height and measure air quality aka air pollutants within the given location. Whenever Drones give details about the air pollution in a selected area visualize those data using data visualization tool [4], because Data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs. Conversion of data's into visual context would be my roles as this makes the data more familiar to people where they could understand easily. Another task is to create a mobile application that allows the public to easily monitor and understand the level of air pollution in the environment in which they live, using the air pollution color patterns and graphs obtained by previously visualized data. It aims to create a program that can raise public awareness of air pollution and current conditions. When the user enter the location or given the current location, the map locate the location and show the data about the air quality based on that location.

II. METHODOLOGY

The proposed system is consist of four main functions. First function is developing a drone and creating a flying program. Second function is to develop a computer program to read data and clean data to send it for analysis. Then the clean data will undergo through an algorithm to analyze and predicting. Then, the analyzed data will used to develop mobile application to visualize the predicted data. Finally, hoped to create chat bot using AI. Here we use to develop the research project based on agile methodology as software development methodology. The Agile model is a project management methodology purposely adopted for the development of sophisticated software as we learnt.

    A. Create a hardware device (drone), get sensor data and clean the data.

    B. Go through an algorithm and predict the data.

    C. Visualize the predicted data using mobile application.

    D. Create a chat bot using Artificial Intelligence (AI). We developed this process step by step.

A. Create a hardware device (drone), get sensor data and clean the data

That is, to create a drone related to this project and to create its flying program. When the drone is built, it must be fully automated and equipped with sensors for specific gaseous components. These sensors should be able to obtain sensor readings for the relevant gaseous component's percentage at that specific location. Also, the aircraft should be designed in such a way that it can identify and avoid obstacles in the environment. Also, once the user has assigned a specific area for the drone through its program, it must automatically fly to that area and automatically locate and fly to a certain three-dimensional data retrieval location in that area. Four phases of work have been used to accomplish this part until the implementation of the system in real environment. It starts with the development of sensor system for sensing the MQ7 (Carbon Monoxide), MQ4 (Methane), MQ2 (Propane) gas sensors concentration and followed by controlling and inter-grating the sensor with an Arduino UNO micro-controller [3]. The third phase focused on the process of data collection and the final phase of work emphasis on the data analysis measured and recorded using the proposed system in the real environment [6].

Arduino is an open-source electronics platform. Arduino hardware board designs use a variety of microprocessors and controlling functions. The hardware boards are equipped with sets of digital and analog input and output pins that may be interfaced to various expansion boards and other circuits [8]. The boards feature serial communications interfaces, including Universal Serial Bus on some models, which are also used for loading programs from personal computers. Arduino IDE is the software part that supplies many very good libraries for basic functionalities and could run on Mac, Windows, and Linux [9].

In here Arduino with the processing software, the program is made to run and reads serial port output. The entire read sensor values have been uploaded into the database in MySQL and the current values of sensor is monitored through the processing software. The report generation in excel (CSV) or pdf format has been obtained. XAMPP and Apache server used for databases connectivity with Arduino [8].Various sensors are connected with Arduino and interfaced with the computer through serial port interface. The c coding or the program is uploaded in Arduino after compilation. Settings of terminal software is configured and going to port option, the right comport has been chosen and set baud rate to 9600 bps and various bound rates. Again going to Capture option and check the flow of sensor values in the computer [9].
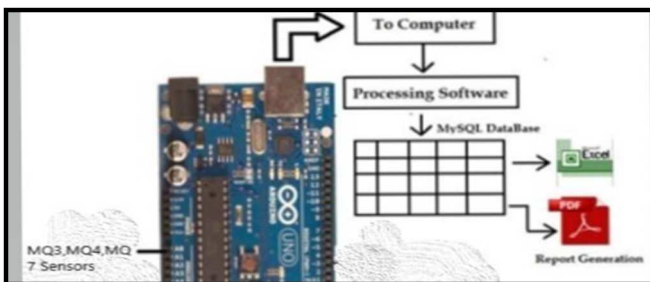

Figure 1: Arduino with processing software and MySQL database

Sensor Name	Application	Power Consumption	Picture of sensor
MQ-2	CH₄ (Methane gas)	≤900mW	
sMQ-7	CO (Carbon monoxide)	≤350mW	
MQ-4	Butane	≤900mW	

Figure 2: Gas sensors that use for data gathering

B. Go through an algorithm and predict the data.

In the analysis and predicting part, I thought to do create an algorithm for build an equation [5]. It will contain percentage of air pollution. Then everyone can get a knowledge about the area polluted or not. For now, it creates many AQI calculator [6]. However, no equation for get strictly air pollution percentage. Therefore, should have a calculating equation. To build a calculation equation I use Python language. When we consider about what software platform we want to use for air pollution the best method is python language with Jupyter notebook platform. As well as in our research project we have both java and python languages. In this case, I want to know about a middle language for combine these two languages. I will use Jython language for fix this problem. For predication using regression technics and time series analysis.

Regression analysis (RA) is a statistical technique for investigating and modeling the relationship between variables. It is also one of the most common methods of prediction. The technique concerned with predicting some variables by knowing others. We can distinguish our outcome variables into two categories. There are continuous variables and categorical variables. Continuous variables described by numerical values and regression models used to predict them (linear regression). Categorical variables are restricted to a limited number of classes or categories and we use classification models (best method for classify) for their prediction. Perhaps, the outcome has two categories this is referred to as binary classification and typical techniques are decision trees and logistic regression [7].
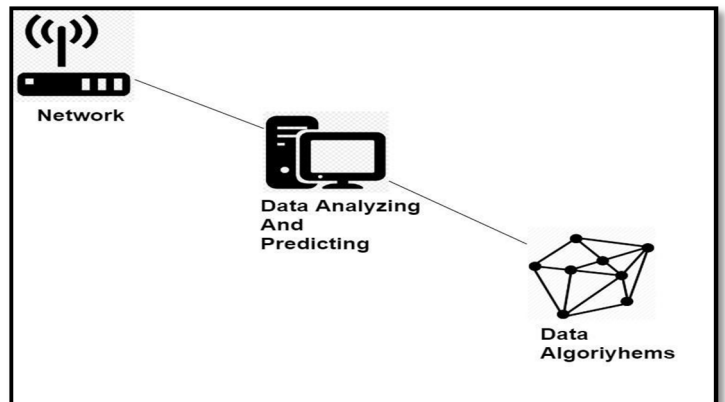

Figure 3: System diagram in predicting data

C. Visualize the predicted data using mobile application

Next part is Data Visualization. All sensor data obtained by the second member in this project should be subject to data visualization. All data analyzed and predicted by the third member should also be subject to data visualization. The task of the fourth member is to create a computer program for this purpose [4], which is expected to use specialized technologies and methods for visualizing data related to geographical locations. In particular, it hopes to create the relevant computer program using the GIS Geo-technical techniques. This is because Data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs.

This makes the data for the man mind to comprehend and makes it easier to identify trends, patterns, and outlets within large data sets. Another task of the fourth member is to create a mobile application that allows the public to easily monitor and understand the level of air pollution in the environment in which they live, using the maps and graphs obtained by previously visualized data. It aims to create a program that can raise public awareness of air pollution and current conditions. It will also allow them to monitor the level of air pollution in the specific area and the future air pollution conditions predicted by the third member. Then use Flutter for mobile application development.

This section will detail the steps that will be followed for the development of the mobile application. The mobile application have to show the level of air pollution in the environment in which they live, using the maps and graphs obtained by previously visualized data. It aims to create a program that can raise public awareness of air pollution and current conditions. Use Fire-base as my database. It is a platform for the development of mobile applications that offers us a wide variety of services as well as hosting [5], Real-time database, authentication, among others. The application will use some of the services offered by Fire Base such as a real-time database, hosting, among others. The Fire-base Real-time Database is a cloud-hosted database. Data is stored as JSON and synchronized in real time to every connected client. When you build cross-platform apps with our iOS, Android, and JavaScript SDKs, all of your clients share one Real-time Database instance and automatically receive updates with the newest data.
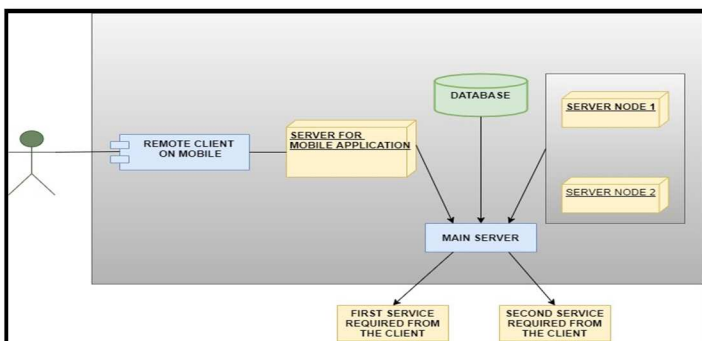

Figure 4: Progress in visualizing

D. Create a chat bot using Artificial Intelligence (AI).

Based on the research here described, it was possible to provide an AI-based chatbot application to assist the specialist in the acquisition of air quality sensor data through a WSN. As a proof of concept, the AirCop was developed, a chatbot that uses NLP and AI to interact with the user and search for the desired information in a WSN, adapting to different forms of dialogue to achieve the same intention. Next, the functionalities developed for the chatbot system are demonstrated. The dialogues shown in this section were translated to English for a better understanding of their content. [9]. Voice recognition (voice to text) is widely accepted as the future of interaction with computers and mobile applications in future and now; there is no need to use traditional the input devices such as the mouse, keyboard or touch sensitive screen and is especially useful for users who do not have the ability to use these traditional devices in modern world [14]. In AirCop we can get these sensor readings by using speech recognition. Sentimental analysis and mood control is also added to this part.

The domain specific voice recognition model provides natural language text, which the chatbot assists in extracting intentions and entities from. In this case, the NLU component is utilized to extract the intents and entities. The entity reflects a condition that is important to the purpose, whereas the intent represents sorts of user needs [20]. For a machine to understand, user inputs will be extremely ambiguous. So, first and foremost, the natural language content should be transformed to a machine-readable format known as the purpose and entity. To extract the intents and entities, algorithms like CRF and BOW are employed. For feature extraction, the BOW method has been utilized in several systems and research articles. According to the majority of readings, BOW is one of the finest algorithms for extracting characteristics. BOW is a basic model that is utilized in the situation of word manipulation. This approach is significantly quicker than the Tf-idf algorithm since it requires less computing power. The retrieved intents and entities are passed to a component that creates a response using machine learning models and rules. At each phase of a dialogue, a policy class determines which action should be taken. This chatbot uses a memorization policy to remember a specific amount of conversation history and context, a mapping policy to directly map intents to actions, a TED policy, a form policy to allow the user to fill out some information, and a two-stage fallback policy to handle incorrect dialogue flow.

The chatbot may not always be able to perform services such as getting the details about App. In this casethe chatbot will need to connect to a back-end API (Eg: A Rest API for Air Pollution Mapping Service). As a result, an action file is responsible for all back-end communication as well as linking it to the chatbot in order to provide a comprehensive service. The output of the aforementioned components is assembled into replies, which are displayed to the user. Notification, FAQ-based chatbots, Contextual AI, Personalized Assistants, and Autonomous Organization are the five main types of chatbots. Notifications and chatbots that are based on FAQs are ineffective. The context of a discussion plays an essential part in communication in a typical conversation. Chatbots that are focused on notifications and FAQs are unable to manage these situations. Contextual AI, on the other hand, can perform such tasks. This Air Cop bot can also handle context and make comparisons, allowing it to function as a customized assistant.
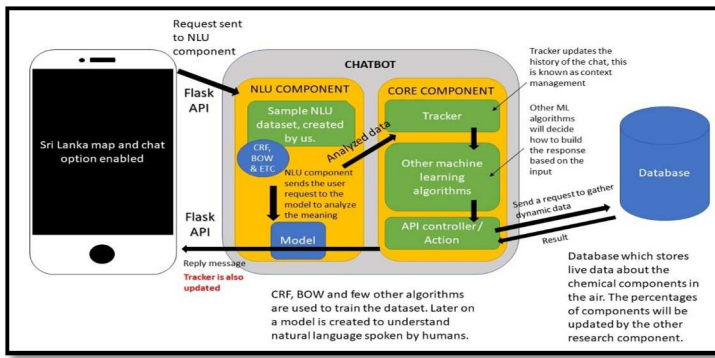
Figure 5: Progress in chat bot

E. Equations

In this research project we should calculate the Air pollution index for analyze the data to get a good prediction. A – Color Range Bmax – Maximum color range   Ex: Maximum value for CO2 gas. (Cumulative value) Bmin – Minimum color range Cmax – Maximum value for related gas in the chart according to the gas. Cmin - Minimum value for related gas in the chart according to the gas.

$$AP\ I = [(ACmin)(Bmax - Bmin)/(Cmax - Cmin)] + Bmin\gamma$$

(1)

F. System overview architecture
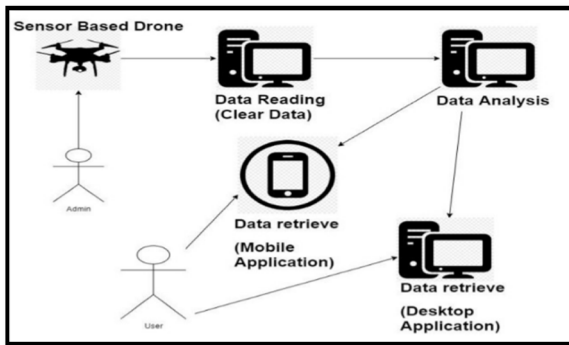
Figure 6: System Overview architecture

III. RESULTS AND DISCUSSIONS

The results obtained from the four main implemented solutions and discussions of the solutions are presented in this section of the paper. The specific four components with the results and discussion are shown below.

A. Program the drone to identify the data points within the given area in 3D space and get the sensor readings

The LPG gas sensor is also used to detect cigarette smoke, toxic gases, combustible, propane, ISO-butane and LNG. This is a data logging system for wind electric generator using Arduino UNO micro controller. Firstly, this system acquires different parameters like speed, velocity, current, voltage etc. After that it give these signals to interface via line driver IC. The signal coming to Arduino micro controller, these signals entered as analog but after processing the output is digital signal. It is observed that all the parameters in time interval of 2ms and stored in the storage drive. These system is used to monitor the air quality of the atmosphere using MQ7, MQ4, MQ2 gas sensor and Arduino. Here the IOT technology is used to monitor the

various gases present in air and the ppm values are displayed in the thinks web server.

B. Sensor data analyzing and predicting

Program to get the sensor readings and analyze the data with suitable methodology and predict the condition of the atmosphere in the specific area. Use cluster analysis and other analysis methodologies and technologies to this function. Use a dataset to train the model that can predict the air pollution level of the relevant area.
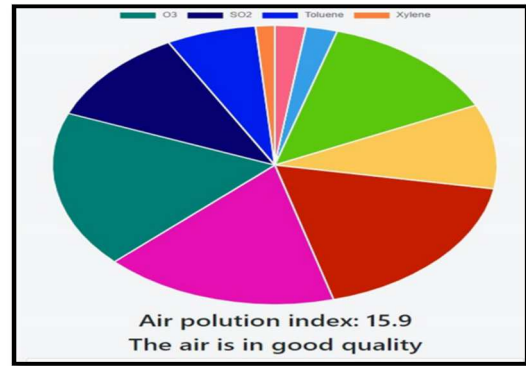

Figure 7: Pie chart of Predicted data

C. Visualize the data using visualization technologies and GIS technologies

Present the results and predictions in a user-friendly manner. For this function, use relevant data visualization technologies and GIS technologies.
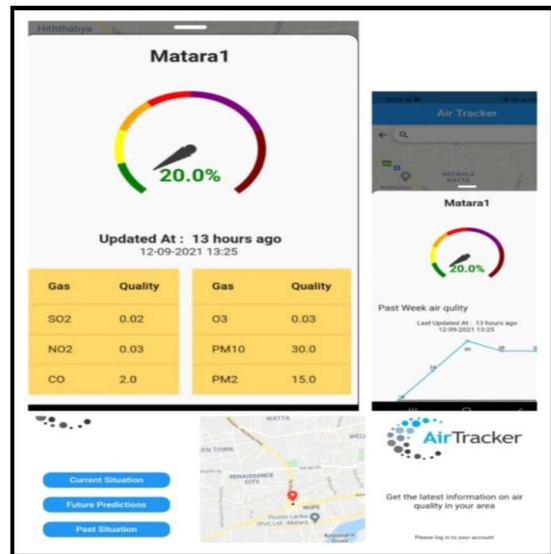

Figure 8: Sample screen shot in visualizing

D. AI chatbot assistant personalized to some extent

A conversation bot may provide users with a wealth of information about this software. As a result, even if users

405

have no prior knowledge of the program, they may acquire knowledge of it. They can directly ask the question to obtain information quickly. The table below is a collection of some of the most popular user chats with which this module has been tested.

	Table Column Head			
	User Text	Entity	Confidence	Times Taken
1	"Tell me the summary of whole map"	Requirements	0.9999	1 seconds
		Location	0.9854	
2	"What is the composition level of Carbon Dioxide"	Requirements	0.9309	1 seconds
		Words	0.6968	
3	"I want to know the summary of Colombo on 2021-09-22"	Location	0.9333	1 seconds
		Date	0.9874	

Figure 9: Working chatbot

IV. CONCLUSION

Finally we can conclude that the proposed objective was achieved, which is to carry out the mobile application to make people aware of the damage they cause to the environment. The scientific contribution provided by this research work is very important in the science environment since it is the basis for future projects that will have a great impact on society, projects that will be totally necessary to reduce environmental pollution and thus to save the planet. For future work it is proposed to improve the subject, apply technologies such as artificial intelligence or some other that makes the idea evolve. There are other alternatives to this solution to make people aware of the state of the environment, however those that exist are not available to people.

REFERENCES

[1] Bruce N, Perez-Padilla R, Albalak R: The health effects of indoor air pollution exposure in developing countries. Geneva: World Health Organization; 2002:11.

[2] Senevirathne SRDA: Air pollution: a case study of environmental pollution. Journal of College of Community Physicians of Sri Lanka 2003, 8(1):1-9.

[3]WHO: Health Aspects of Air Pollution: Results from the WHO project " Systematic Review of Health Aspects of Air Pollution in Europe". Europe: World Health Organization; 2004

[4] D. Vallero, Air Pollution Calculations, 1st ed. Elsevier, 2019.

[5] "Air Quality Health Index – calculation," Alberta.ca. [Online]. Available: https://www.alberta.ca/air-quality-health-index–calculation.aspx. [Accessed: 25-Feb-2021].

[6]Researchgate.net. [Online]. Available: https://www.researchgate.net/publication/329873035. R: The

[7] R Project for Statistical Computing. [Online]. Available: https://www.rproject.org/. Python: The official home of the

[8] Python Programming Language. [Online]. Available: https://www.python.org/

[9] Matlab: The easiest and most productive software environment for engineers and scientists. [Online]. Available: https://www.mathworks.com/products/matlab.html.

[10]SPSS: The world's leading statistical software used to solve business and research problems by means of ad-hoc analysis, hypothesis testing, geospatial analysis and predictive analytics. [Online]. Available: https://www.ibm.com/analytics/spss-statistics-software

[11]SAS: SAS/STAT State-of-the-art statistical analysis software for making sound decisions. [Online]. Available: https://www.sas.com/enus/software/stat.html.

[12]Orange: Open source machine learning and data visualization for novice and expert. [Online]. Available: https://orange.biolab.si/

[13]Weka: Data mining software in Java. [Online]. Available: https://www.cs.waikato.ac.nz/ml/ index.html.

[14]RapidMiner Studio: Visual workflow designer for data scientists. [Online]. Available: https:// rapidminer.com/products/studio/.

[15]Solar Impulse, "Solutions to air pollution," Slarimpulse.com. [Online].

[16]O. E. G. O. Godall Rohi,"Autonomous monitoring, analysis, and countering of air pollution using environmental drones.," ScienceDirect, Racett Canada Inc, 404 St George Street, Moncton, New Brunswick, E1C 1X0, Canada, 2020

[17]N. R. Z. E. N. C. T. C. Oscar Alvear, "Using UAV-Based Systems to Monitor Air Pollution in Areas with Poor Accessibility," Journal of Advanced Transportation, Compiegne Cedex, France, 2017.

[18]P. A. Ozcan, "Mapping Air Pollution from A Drone," University of California, California, 2021

[19] F. S. K. M. L. M. F. G. Tommaso Francesco Villa, "Development and Validation of a UAV Based System for Air Pollution Measurements," Queensland University of Technology (QUT), Tasmania 7000, Australia, 2016

[20] Air Pollution http://www.climate.lk/Nuraicholai/index.html