JAKEBOT: IoT Based and Machine Learning Water Quality Monitoring for Rivers.

Anabel Leynes González
20183L201021@utcv.edu.mx
ORCID: 0000-0002-8590-7757

José Antonio Tepepa García
20183L301024@utcv.edu.mx
ORCID: 0000-0001-9020-5461

Karina Roxana Vázquez Romero
20183L301040@utcv.edu.mx
ORCID: 0000-0003-3602-5026

Jesús Enrique Tepole Illescas
20183L301032@utcv.edu.mx
ORCID: 0000-0001-8496-9828

Ricardo Castro Valdivia
ricardo.castro@utcv.edu.mx
ORCID: 0000-0002-9329-3003

Luis Rolando Guarneros Nolasco*
luis.guarneros@utcv.edu.mx
ORCID: 0000-0001-6379-4969

Programa Educativo de Tecnologías de la Información y Comunicación Área Redes Inteligentes y Ciberseguridad
Universidad Tecnológica del Centro de Veracruz
Av. Universidad No. 350, Carret. Fed. Cuitláhuac-La Tinaja, Loc. Dos Caminos, C.P. 94910, Cuitláhuac, Veracruz, México

Abstract— **According to the data obtained in the analysis of the Conagua Water Treatment Program, in Mexico 70% of lakes, lagoons, rivers, and other water bodies present some type of contamination. This article proposes the design of a prototype for monitoring water quality in rivers based on Internet of Things (IoT) sensors and a predictive model Machine Learning, JAKEBOT. The main components include a microcontroller to process the system, a communication system for inter and intra communication, and pH, turbidity, and temperature sensors. The retrieved data are presented in a visual format on a cloud server for analysis applying a machine learning linear regression algorithm. If the acquired data is above the standard threshold, the model will present an alert in the analyzed time range. The system can monitor water quality automatically, so the components allow a highly mobile and low-cost system that helps environmental and government organizations in the generation of strategies and policies to raise awareness of water quality.**

Keywords—Water monitoring, Internet of Things (IoT), Machine Learning

I. INTRODUCTION

The world is going through an environmental crisis derived from growing pollution that is getting out of control, this has led to global ecosystems being altered and water shortages are becoming more alarming every day, since climate change is drying up the main sources of fresh water and, on the other hand, humans have been responsible for polluting them [1]. According to studies conducted by the World Resources Institute (WRI), Mexico is one of the 25 countries facing the highest water stress worldwide [2], meaning that less water is refilled than is extracted.

On a national level, Mexico is going through an extreme water crisis. According to data obtained from the analysis of Conagua's Water Treatment Program, 70% of lakes, lagoons, rivers and other bodies of water present some type of contamination, but what is most alarming is that the quality of more than half of the country's 653 aquifers is unknown. [3].

Due to the great impact and lack of knowledge of the contamination present in the country's rivers, lagoons and aquifers, water quality monitoring has become more interesting, which is why JAKEBOT has been designed, an efficient and profitable IoT prototype, in which three sensors are used to collect river water parameters such as pH, Turbidity and Temperature. The monitored data will be stored in a cloud server to be analyzed through a predictive model based on a Machine Learning algorithm. The analyzed information can be consulted online on the website by civil associations or private companies in order to generate strategies and policies that contribute to the care and sanitation of the country's water.

The sections that compose this document express the following: section two contains the summary of the works related to this research and the comparative analysis of the contributions and technologies that helped to solve the problem. Section three presents a detailed description of the architecture proposed to solve the detected problem. Section four presents a case study for the detection of water quality levels and finally, section five presents the conclusions and future work.

II. RELATED WORD

In the study of Daigavane, et al [4], they presented a design and development of a low-cost system for real-time monitoring of water quality in IoT. They used parameters such as temperature, pH, and turbidity to be measured. The values collected by the sensors can be processed by the main Arduino-based controller.

With Moparthi N., et al [5], present a water quality monitoring system using the Internet of Things to check pH content and send messages to the concerned authorities applying it in municipal water tanks and drinking water deposits. For this, they used an Arduino board to find the pH value and the GSM module for the messaging technique. As well as a LED display to have a continuous observation of the water parameters. The collected data is sent to the cloud for global monitoring of water quality.

In the article of Chowdury M. et al [6], they propose a sensor-based water quality monitoring system. Real-time data access can be realized by using remote monitoring and IoT technology. Data collected at the site can be displayed in a visual format on a server PC with the help of Spark streaming analytics through Spark MLlib.

Budiarti R. et al [7] developed an integrated system based on the Internet of Things to measure water quality developing an environmental water management monitoring system using sensors with raspberry pi as an embedded system. The result is that the IoT water quality monitoring system can function

as an automated water monitoring system for surface water and is online in real-time.

In the work of Krishna, et al [8], they propose a system that overcomes the problems of multiple sensor parameter requirements and a robust communication framework that can be extended to water monitoring application domains. They used an industrial multiparameter sensor node, and the sensed data are processed and communicated to the server through a layered architecture. The sensor node uses a file system to store sensor data in case of network loss and the communication technology between nodes and sensors is GSM. On the server, the data is stored and visualized in a web user interface.

Abdulwahid A. [9] uses IoT technology to develop an affordable system to monitor water quality in real-time. It integrates several sensors to measure chemical and physical properties of water, such as conductivity, pH, turbidity, and temperature. The central controller manages the processing of the data captured by the sensor and visualization of the data in the cloud.

Table 1 shows a comparative analysis of the six most relevant papers, to visualize the most notable similarities and differences.

TABLE I.   COMPARATIVE ANALYSIS OF WORKS AND RESEARCH RELATED TO MONITORING WATER.

Work research	Problem	Contribution	Technology
Daigavane V., Gaikwad, M., 2017	Water pollution.	IoT prototype, PH, Temperature, and Turbidity sensor.	Arduino Sensors IoT
Noparthi N, et al, 2018	Water pollution in rivers and lakes.	Water quality monitoring system measuring pH with IoT	Arduino Sensor PH GSM module for messages
Chowdury M. et al, 2019	Manual water quality monitoring.	Water quality control system using IoT sensors.	Arduino Sensor pH, Turbidity, Temperature, LCD, Wi-fi module, IoT platform, Neural network
Budiarti R. et al. 2019	Water pollution in rivers.	Integrated monitoring system with IoT	IoT MariaDB Arduino
Krishna, et al. 2020	Water pollution is caused by people and natural disasters.	The industrial multi-parameter sensor node system	Sensor industrial EXO Sonde Communicatio n GSM
Abdulwahid, A. 2020	Drinking water quality	Real-time water quality control system with IoT sensors	IoT Sensor pH, Turbidity, and Temperature. Internet

In Table 1, there are several solutions related to the design of prototypes to monitor water quality using IoT by collecting data through sensors and presenting them in a remote server, as well as using different technology for central processing such as Arduino and Raspberry. Therefore, our prototype

incorporates some significant improvements, such as the material in which the prototype is built that stores the components to be navigable remotely in the river and the use of artificial intelligence techniques to train the prediction model of how the parameters analyzed in a range of time will be analyzed.

III. ARCHITECTURE DESIGN

The proposed architecture in this research aims to use IoT components to design a prototype that interacts with the analysis of water from rivers without suffering damage. Its purpose is to apply a Machine Learning predictive model with the data obtained, which seeks to raise awareness among the population since it will show the effects over a range of years that will result from not taking measures to preserve clean water. (Figure 1).
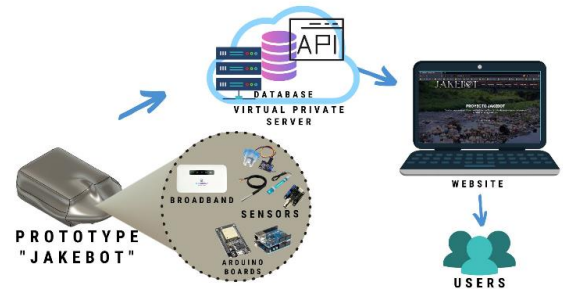

Fig. 1.   The architecture of the proposed system.

A. Description of the architecture

- Prototype: The prototype shield is made of PLA (polylactic acid) filament, which is biodegradable, durable, and resistant plastic. Three versions of the prototype were designed. The initial version was made of metal, but the implementation was not relevant because there was no cover for the electronic devices. In the second version, it was decided to redesign it by implementing a cover to protect the devices, but it was not feasible due to the type of material used. In the third version of the prototype, it was decided to use PLA material, which was compact, light and easy to transport, thus obtaining better results for monitoring. (Figure 2).

- Sensors: Three sensors (pH, Turbidity, and Temperature), allows real-time monitoring of the water quality level. (Figure 3)

- Virtual Private Server (VPS). Through broadband Internet connection, the parameters of pH, Turbidity and Temperature collected as well as the altitude and longitude through a GPS are sent in real time to be stored in a database and managed in an API which is hosted in a VPS.

- Website: A website is available where users can register to access the data collected by the sensors and

in which the prediction model based on a machine learning algorithm is implemented.

- Users: Civil or governmental organizations with access to the website for control and analysis of the data.

Fig. 2.   Prototype shield desing 25x30x20 .

The prototype has a portable WiFi device that provides a connection to a NodeMCU development board that sends data wirelessly and an Arduino Uno board on which is mounted an Ethernet Shield controller that allows connecting it via an Ethernet cable, both boards are loaded with codes of the configuration of each of the sensors which make a connection with the database in the cloud. (Figure 3)
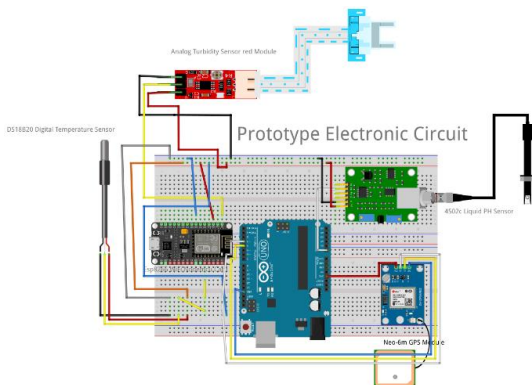

Fig. 3.   Prototype electronic and sensors circuit.

The prototype is designed with scalability thought to add different types of sensors when required, as well as the ease of transportation and displacement in the water is considered by remote control management (Figure 4).
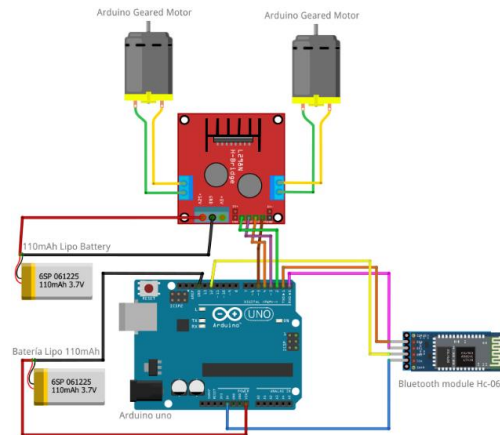

Fig. 4.   Motor electronic circuit

IV. CASE STUDY

This section describes the use case analyzed with the prototype for the monitoring of water quality in a river in Naranjal, Veracruz.

A. Description of the case.

Our case study was carried out in the river located in Naranjal, Veracruz, which originates in this town and flows into the Blanco River near San Martín community in the same town. When walking along the riverbank, in some areas people use it to wash clothes and junk and dump sanitary waste, which has caused the river's water quality decline. Some of the contamination factors that have been detected are as follows:

1) Turbidity

2) Foam

3) Sanitary waste

4) Garbage

5) Decrease in Flow rate

6) Decrease in flora and fauna

According to the factors detected, the quality of the water flowing in the river has degraded over time, so the Town of Naranjal and the Tequecholapa Environmental Management Unit need to know the levels of water quality in terms of pH, turbidity, and temperature to protect the flora and fauna of the area, as well as the people who use it for consumption. Therefore, the following questions arise:

1) How to measure water quality levels in the river?

2) How to obtain the result of the analysis in the shortest possible time?

3) How could I store that information to be consulted at any time?

4) How to use this information?

To solve the above questions, the architecture previously proposed in this research will be implemented as a solution.

B. Parameter acquisition and display.

To monitor water quality levels in the river, JAKEBOT is implemented as an IoT prototype consisting of a PLA shield (Figure 5), which contains three calibrated sensors inside to detect the levels of pH, Turbidity, Temperature and GPS for its location (Figure 6), and control code was developed on Arduino platform. Data acquisition is done through APIs through a website where they are analyzed (http://www.jakebot.website). The parameters recorded from each sensor (pH can be from 1 to 14, turbidity from 1 to 5, NTU and Temperature -55ºC and 125ºC) are stored in the database, which will allow training the machine learning prediction model.
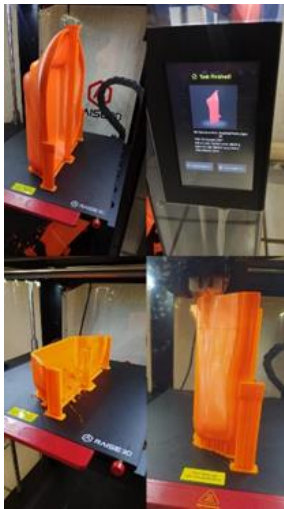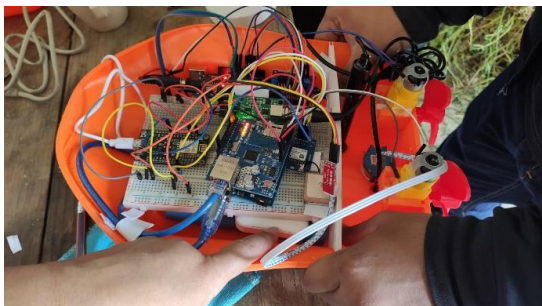

Fig. 5.   3D printed PLA.

Fig. 6.   Prototype JAKEBOT.

C. Data analysis.

The analysis of the parameters collected by sensors of the prototype is recorded and visualized on the implemented website. The website has different visualization modules such as the locations where monitoring has been carried out (Figure 7), pH levels, turbidity, and temperature (Figure 8).
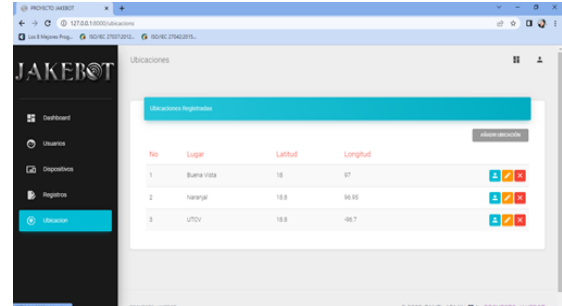

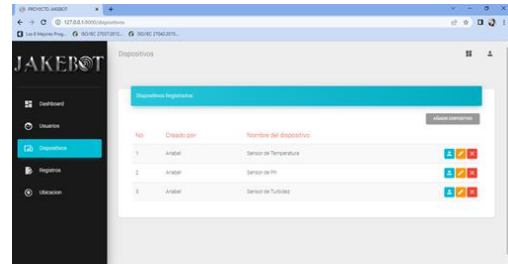Fig. 7.   The location of the rivers is monitored.

Fig. 8.   Logging of sensors installed in the prototype.

D. Machine Learning algorithm prediction.

For our analysis we have applied a machine learning algorithm such as linear regression, which is used to predict the relationship between two variables by applying a linear equation to the observed data in which there are two types of variables, one is the independent variable and the other is a dependent variable. This algorithm has been selected because linear regression is commonly used for predictive analysis [10].

At the implemented site the data from each monitored sensor can be retrieved and evaluated, and with this information the dataset (Table 1) for model training is generated (Figure 9).
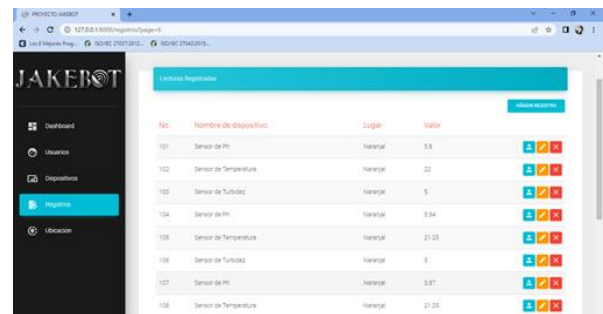

Fig. 9.   Records of the monitored parameters.

TABLE II.   LECTURES MONITORED DATASET ATTRIBUTES.

Attribute	Description
Id	Record ID.
Device name	Stores the applied sensor name.
Location	GPS location of analyzed site.
Value	Value monitored by applied sensor.

An important feature in the monitoring system is that the collected data is used to generate datasets of each sensor to perform a predictive training model through a machine learning linear regression algorithm.

TABLE III.   EXAMPLE PH DATASET.

Id	Valor	Year
454	3.6	2022
968	3.7	2022
573	3.8	2022
798	3.9	2022
133	4	2022
520	4.4	2022
283	4.8	2022
484	5	2022
14	5.2	2022
721	5.3	2022
933	5.5	2022
250	5.6	2022
635	5.7	2022
177	6.1	2022
96	6.5	2022
198	6.9	2022
688	7	2022
659	7.2	2022
149	7.5	2022
516	7.7	2022
23	7.9	2022
670	8.2	2022
847	8.4	2022
429	8.9	2022
741	9	2022
998	9.5	2022
852	9.8	2022
36	10.1	2022

Id	Valor	Year
57	10.6	2022
419	10.8	2022
327	11.2	2022
232	11.3	2022

The model predictions are queried using an API programmed in R [11] language with the Plumber library [12]. This API returns the prediction of a range of years, which will show the results over time of how the values of the analyzed factor will be found. With the previously mentioned, an important support will be provided to implement and improve the strategies of sensibilization for the care of the water of the analyzed river. (Figure 10)

TABLE IV.   EXAMPLE TEMPERATURE DATASET.

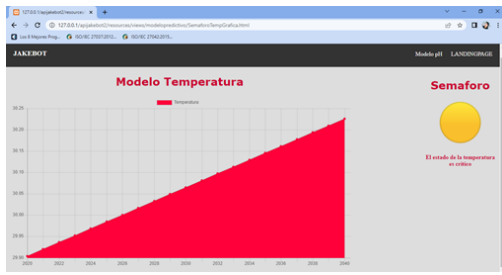| Id	Valor	Year
182	28	2022
375	20	2022
512	19	2022
727	42	2022
619	32	2022
577	34	2022
150	24	2022
454	35	2022
968	15	2022
573	35	2022
798	17	2022
133	25	2022
520	18	2022
283	34	2022
402	36	2022
14	39	2022
721	30	2022
933	18	2022
250	38	2022
635	31	2022
660	24	2022
177	30	2022
96	31	2022
549	22	2022
675	35	2022

Fig. 10. Result of the predictive model consulted over a range of 20 years of the temperature sensor.

V. CONCLUSIONS AND FUTURE WORK

With the knowledge that water quality will serve as support for some organizations that do not know the levels found in rivers for their use and conservation, with this prototype it was possible to obtain real-time records of water quality, which allowed training a predictive model based on linear regression machine learning showing a prediction in a range of time, allowing to observe that if appropriate strategies for the care and sanitation are not generated, water quality may be lost in the analyzed river, thus contributing to the damage of flora and fauna that depend on it. Some sensor parameter values that could seriously affect the health of the population and the flora and fauna were observed during the monitoring of the river, for example, low values were detected in the pH sensor that could cause eye and skin irritation. The system can monitor water quality automatically, so the components allow for a highly mobile and low-cost system. In future work, we intend to improve the prototype by adding oxygenation and conductivity measurement sensors, as well as to incorporate the sending of alert messages from the API when identifying ranges outside the established limits, which will make it possible to determine whether the water flowing through these tributaries is suitable for agriculture, fauna and human consumption in the area.

REFERENCES

[1] Bernabe, I., Elizondo, P. L., Strong, C., Kuzma, S., Bosch, A., Robles, E. H., Pool, J., Martínez, A. M., López, S., Aparicio, E. S. S. & Morales, G. (s. f.). Agua | WRI Mexico. Recuperado 18 de octubre de 2022, de https://wrimexico.org/our-work/topics/water

[2] Morales, G. (s. f.). La infraestructura natural como herramienta para enfrentar el estrÃ©s hÃdrico en MÃ©xico | WRI Mexico. Recuperado 18 de octubre de 2022, de https://wrimexico.org/bloga/la-infraestructura-natural-como-herramienta-para-enfrentar-el-estr%C3%A9s-h%C3%ADdrico-en-m%C3%A9xico

[3] Agua, C. N. del. (s. f.). PROSAN. gob.mx. Recuperado 18 de octubre de 2022, de https://www.gob.mx/conagua/acciones-y-programas/prosan

[4] Daigavane, V. V., & Gaikwad, M. A. (2017). Water quality monitoring system based on IoT. Advances in wireless and mobile communications, 10(5), 1107-1116.

[5] N. R. Moparthi, C. Mukesh and P. Vidya Sagar, "Water Quality Monitoring System Using IOT," 2018 Fourth International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2018, pp. 1-5, doi: https://doi.org10.1109/AEEICB.2018.8480963

[6] Mohammad Salah Uddin Chowdury, Talha Bin Emran, Subhasish Ghosh, Abhijit Pathak, Mohd. Manjur Alam, Nurul Absar, Karl Andersson, Mohammad Shahadat Hossain, IoT Based Real-time River Water Quality Monitoring System, Procedia Computer Science, Volume 155, 2019, Pages 161-168, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2019.08.025

[7] R. P. N. Budiarti, A. Tjahjono, M. Hariadi and M. H. Purnomo, "Development of IoT for Automated Water Quality Monitoring System," 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), 2019, pp. 211-216, DOI: https://doi.org/10.1109/ICOMITEE.2019.8920900

[8] K. S., S. T.V., M. S. Kumaraswamy and V. Nair, "IoT based Water Parameter Monitoring System," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 1299-1303, DOI: https://doi.org/10.1109/ICCES48766.2020.9138001

[9] Abdulwahid A., IoT Based Water Quality Monitoring System for Rural Areas, 9th International Conference on Renewable Energy Research and Application (ICRERA), 2020, pages 279-282. https://doi.org/10.1109/ICRERA49962.2020.9242798

[10] Guarneros-Nolasco, L.R.; Cruz-Ramos, N.A.; Alor-Hernández, G.; Rodríguez-Mazahua, L.; Sánchez-Cervantes, J.L. Identifying the Main Risk Factors for Cardiovascular Diseases Prediction Using Machine Learning Algorithms. Mathematics 2021, 9, 2537. https://doi.org/10.3390/math9202537

[11] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

[12] Schloerke B, Allen J (2022). plumber: An API Generator for R. https://www.rplumber.io, https://github.com/rstudio/plumber.