IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Water Quality Prediction Using Machine Learning

Mrs.Priyanka Gupta, Mansi Adhav, Sonal Chidrawar, Dhairyashil Pawar, Hrutik Chaudhari

Dept. of Information Technology, Dept. of Information Technology, Dept. of Information Technology, Dept. of Information

Technology, Dept. of Information Technology, Dept. of Information Technology, Pune, India.

Abstract: The project aims to predict water quality based on nine factors: pH, hardness, solids, chloramines, sulfates, conductivity, organic carbon, trihalomethanes and turbidity. It estimates water by separating it into drinkable and non-drinkable water. If the water is deemed undrinkable, further evaluation will determine whether the water is suitable for agricultural or commercial use. This project uses Python and Flask for front-end integration and Jupyter Notebook for modeling. The random forest algorithm was validated after comparison with SVM, logistic regression, and decision tree algorithms, which proved to be more accurate. The user interface includes a home page with a set of parameters and a parameter value map. Additionally, PDF reports allow direct parsing of parameter values from uploaded files. Once the product is delivered, the model predicts water quality and provides results indicating potential or recommended use.

Key Words – Water quality, potable and non-potable, python, random forest, accuracy, crucial parameters.

I. INTRODUCTION

Access to clean and safe water is a human right and the foundation of public health and environmental sustainability. However, ensuring water quality requires a complex assessment that is affected by many factors. To overcome this challenge, the "Water Quality Prediction Using Machine Learning" project was created, which aims to use the power of artificial intelligence to reliably predict water quality. The project focuses on nine key parameters (pH, hardness, solids, chloramines, sulphates, conductivity, organic carbon, trihalomethanes and turbidity) to provide a better insight into the capital's drinking water. suitability for consumption. and usage. Combining aspects of Python, Flask's front-end integration, and Jupyter Notebooks' development and evaluation model, this project uses a method to solve the challenge of measuring water quality. Moreover, the adoption of the Random Forest algorithm followed by a rigorous comparison with SVM, logistic regression, and decision tree algorithms shows the promise of the technology to achieve the best performance. Its effect is well documented. For example, pH indicates how acidic or alkaline water is, while hardness indicates mineral content, which affects taste and use. Measurement of impurities in the waste, chloramines and sulphates, indicates the possibility of contamination and provides information on the behavior and composition of organic carbon, water. Trihalomethanes work as an indicator of the disinfection of products, while turbidity measures the clarity of water, an important factor affecting beauty and health. accuracy and the importance of user accessibility and participation. The user interface has been carefully designed to provide a consistent experience with a simple website that provides intuitive plans and easily accessible information for deployment. Additionally, integration of PDF reporting simplifies data entry, allowing users to easily remove negative water quality standards for forecasting. and broader resource management. The program aims to support decision-making by providing stakeholders with an understanding of the quality of water quality and promote equitable access to clean and safe water.

II. LITERATURE SURVEY

In [1] authors Among water quality components, measuring the dissolved oxygen (DO), chemical oxygen demand (COD), biochemical oxygen demand (BOD), electrical conductivity (EC), pH, temperature, K, Na, Mg, etc.

In [2] Authors evaluated the quality of rivers two approaches are considered, including measuring the water quality components and defining the mechanism of pollution transmission.

In [3] Authors has researched about hydrometry stations, the water quality components are measured and the stage-discharge relation is defined. Obtained values from hydrometry stations contain basic information for feasibility studies and development of water conservation projects. Evaluation of water quality is a basic stage for development of agriculture projects in terms of determination of cropping pattern, type of irrigation system, and systems of water purification for industries.

In [4] author analyzed physical, chemical, and biological parameters of water quality are reviewed in terms of definition, sources, impacts, effects, and measuring methods. The classification of water according to its quality is also covered with a specific definition for each type.

.In [5] water quality management that covers timely topics such as new methods of water and wastewater treatment, groundwater modeling and quality. Offers creative solutions to water management problems. Substantially supported by hundreds of discussion questions, references, tables, and appendices.

In [6] author considered Physical, Chemical and Biological Parameters of water determine its quality. These water quality characteristics throughout the world are characterized with wide variability. Therefore the quality of natural water sources used for different purposes should be established in terms of the specific water-quality parameters that most affect the possible use of water. Physical Characteristics of Water Physical characteristics of water (temperature, color, taste, odor etc.) are determined by senses of touch, sight, smell and taste. For example temperature by touch, color, floating debris, turbidity and suspended solids by sight, and taste and odur by smell.

In [7], the application of machine learning models, including Support Vector Machines (SVM) and Regressive Neural Networks, was explored for predicting water quality across different states of India. An accuracy of 97.01% was achieved using SVM. The conclusion emphasized SVM and Regressive Neural Networks as effective models for predicting water quality in diverse regions of India. Water resources are often polluted by human intervention. Water pollution can be defined in terms of its quality which is determined by various features like pH, turbidity, electrical conductivity dissolved oxygen (DO), nitrate, temperature and biochemical oxygen demand (BOD). This paper presents a comparison of water quality classification models employing machine learning algorithms viz., SVM, Decision Tree and Naïve Bayes. The features considered for determining the water quality are: pH, DO, BOD and electrical conductivity. The classification models are trained based on the weighted arithmetic water quality index (WAWQI) calculated. After assessing the obtained results, the decision tree algorithm was found to be a better classification model with an accuracy of 98.50%.

III. PARAMETERS AND THEIR SIGNIFICANCE

3.1 pH

Meaning and importance:

pH is a measure of acidic or alkaline water, ranging from 0 to 14, with 7 being neutral. It is an important negative factor for water quality because it affects many chemical and biological processes. Water with a pH below 7 is considered acidic, and water with a pH above 7 is considered alkaline. The pH of drinking water is important for human health and affects the effectiveness of antibiotics. Acidic water (low pH) can cause corrosion in pipes and leaching of metals such as aluminum and copper, causing problems during consumption. Alkaline water (high pH), on the other hand, can cause a bitter taste and form scale in equipment and pipes. It is important to maintain pH balance because extreme pH values can disrupt the body's acid-base balance and cause health problems. Additionally, pH affects the solubility and toxicity of chemicals and heavy metals in water. For example, the solubility of aluminum increases at low pH, making it easier but potentially more toxic. In summary, the pH of water not only affects the taste and use of water, but also affects health and affects all chemicals in water.

3.2 Hardness

Definition and importance:

Water hardness generally refers to the concentration of dissolved calcium and magnesium ions. These minerals occur naturally and enter water through the weathering of rocks and soil. Hard water is often caused by soap not lathering and tends to form lime deposits in plumbing and heating systems. Although hard water is generally harmless to human health, it can cause many problems. For example, limescale accumulation in water pipes will reduce the efficiency of your water heater and increase energy costs. In manufacturing facilities, hard water can cause serious maintenance problems and affect the life of equipment. On the plus side, the calcium and magnesium in the drink help absorb these important nutrients from the diet and may aid bone health. However, very hard water can cause skin irritation and dry hair. Conversely, soft water low in calcium and magnesium can be more corrosive and cause metal to leach from pipes and appliances. The ideal hardness level balances the benefits of mineral content with the need to reduce scaling and corrosion.

3.3 Solids (Total Dissolved Solids - TDS)

Meaning and Importance:

Total dissolved solids (TDS) represent the total content of all inorganic and organic matter present as suspended molecules, ions, or particles in a liquid. TDS in water includes various salts, minerals and organic matter dissolved in water. Components in TDS include calcium, magnesium, potassium, sodium, bicarbonate, chloride and sulfate. TDS measurement is an important indicator of water quality as it shows the purity of water and the amount of dissolved substances. High TDS levels can indicate that water may be contaminated or high in minerals, which can affect water taste, hardness and usability. Water with too high a TDS will have a bitter or salty taste and can damage pipes and appliances, similar to the effects of hard water. For example, TDS-free water is generally considered purer and tastier, but low levels can also indicate nutrient deficiencies. The presence of high TDS can interfere with industrial and agricultural processes, making water unsuitable for certain uses. TDS analysis is important to maintain the balance between beneficial nutrients and harmful bacteria, making water safe and pleasant to drink.

3.4 Chloramine

Meaning and Importance:

Chloramines are disinfectants used in drinking water treatment. They are made from chlorine combined with ammonia and are used as secondary disinfectants in water supplies. Chloramines are particularly valuable for their stability because they persist in the water supply longer than chlorine alone and provide ongoing protection against microbial contamination during the water flow of facilities treating customers. This makes them effective in controlling bacterial growth, which is essential for maintaining drinking water quality. But chloramines also cause problems and health problems. They can interact with organic matter in water to produce antibiotic byproducts, some of which can be harmful to health. Exposure to high levels of chloramines can cause respiratory problems, skin irritation, and other health problems, especially in sensitive groups such as infants, the elderly, and people with respiratory illnesses. Additionally, chloramines affect the taste and odor of water, often providing perfume or fragrance. In aquariums, chloramines are toxic to fish and must be removed from the water. Despite these problems, the use of chloramine is a good way to ensure the microbiological safety of drinking water.

3.5 Sulphates

Meaning and Importance:

Sulphates occur naturally in soil and rock formations and dissolve in water as a result of contact with geological materials. They usually result from the breakdown of minerals such as gypsum and are often found in natural waters. Sulfates are important for water quality because their presence in high concentrations can cause many health and beauty problems.

Sulfates, when taken in large amounts, can have a laxative effect and cause stomach upset such as diarrhea, especially in people who are not used to high sulfates. Sulfates can also affect the taste of water, causing a bitter or chemical taste that may bother consumers. Additionally, high sulfate content can cause corrosion in pipes, allowing metals to enter the water, creating additional health risks. Therefore, monitoring sulfate levels is important to ensure water safety and maintain infrastructure.

3.6 Conductivity

Meaning and Importance:

Conductivity is a measure of water's ability to conduct electricity and is directly related to the concentration of ions such as sodium, chloride, calcium and magnesium dissolved in water. It works as an indirect indicator of total dissolved solids (TDS) in water, which includes many inorganic and some organic substances. High conductivity values may indicate increased levels of dissolved ions, which may result from natural sources such as mineral springs or anthropogenic sources such as agricultural runoff, industrial emissions, and wastewater. Increased activity is often associated with the presence of harmful bacteria, indicating poor water quality. For example, high levels of sodium and chloride can affect the taste of water and can be dangerous for people with certain conditions, such as high blood pressure. Quality monitoring helps identify changes in water chemistry, detect contaminants, and ensure water is safe and fit for human consumption and use.

3.7 Organic Carbon (Total Organic Carbon - TOC)

Meaning and Importance:

Total organic carbon (TOC) measures the amount of carbon in water, from organic compounds such as plant material and microbial residues to man-made chemicals such as pesticides and solvents. TOC is important for measuring water quality because it is an indicator of organic matter in water. High TOC levels can have many effects. First, they can provide nutrients for bacteria, which can lead to microbial growth and biofilm formation in the water supply, affecting the microbiological safety of the water. Second, during water treatment, high levels of TOC will react with disinfectants such as chlorine to produce disinfectant byproducts (DBPs) such as trihalomethanes (THMs) that are hazardous to human health. High TOC can also affect the taste, odor and color of water, making it unpleasant. Therefore, measuring TOC is important for controlling the water purification process and ensuring the safety and quality of drinking water.

3.8 Trihalomethanes (THM)

Meaning and Importance:

Trihalomethanes (THM) are a group of compounds formed when chlorine used in water disinfection destroys organic substances (NOM) such as humic and fulvic acids. Common THMs include chloroform, bromoform, dibromochloromethane, and bromodichloromethane. The presence of trihalomethanes in drinking water is concerning because some trihalomethanes are considered carcinogenic and have been linked to other health problems, such as problems with the liver, kidneys, and lungs in the brain. Prolonged exposure to high levels of THMs increases the risk of cancer and adverse reproductive outcomes. It is important to monitor and control THM levels in water treatment processes to minimize these health risks while maintaining effective disinfection. Strategies to reduce THM production include proper use of chlorine, removal of organic matter prior to disinfection, and use of alternative disinfection methods such as ozone or ultraviolet light.

3.9 Turbidity

Meaning and Importance:

Turbidity is a measure of water clarity and is measured by the degree to which suspended solids disperse in the water and absorb light. High turbidity levels can affect water quality and drinking water. Aesthetic concerns aside, increased turbidity can prevent harmful bacteria from emerging from the disinfection process; because the product may contain bacteria, viruses and protozoa, which may make them resistant to chlorination or other treatments. Additionally, high turbidity can interfere with the operation of water purification systems such as filtration and indicate the presence of contaminants that may be detrimental to clean health. Turbidity can also affect the physical properties of water, such as taste and color, making it less attractive to consumers. It is important to monitor turbidity regularly to ensure the effectiveness of water treatment and the safety of drinking water supplies.

IV. METHODOLOGY

The methodology employed in the "Water Quality Prediction Using Machine Learning" project encompasses several distinct phases, each crucial for the successful development and deployment of the predictive model.

A) Planning and Requirements Analysis:

The first phase of developing a water quality forecasting application includes an overall project plan and a detailed analysis. This phase establishes a solid foundation for the next phase by ensuring that the project meets the user's needs, management standards and operational capabilities**.**

1) Objective Definition

The main goal is to create a machine learning-based application that can predict water quality using nine parameters: pH, hardness, solids, chloramines, sulfates, conductivity, organic carbon, trihalomethanes and turbidity. The aim is to determine whether the water is suitable for drinking or not, and if not, to evaluate whether it is suitable for agriculture or commerce. By clearly defining these goals, the development team can prioritize features and functionality that improve user experience and meet requirements.

2) Market Research

Market research is necessary to understand current water quality testing equipment and identify gaps. This involves analyzing existing applications, assessing their strengths and weaknesses, and understanding user pain points. Insights from this study informed the development of specific features to address unmet needs, making the app an important tool for many stakeholders, including environmental organizations, agriculture leaders, and business owners.

3) Requirements Gathering

Collecting detailed information about functional and non-functional is important to provide a method for improving practice. Functional requirements include data entry of nine parameters, execution of the random forest prediction algorithm, and generation of detailed reports. Non-functional requirements focus on functionality, security, and usability to ensure that the application is functional, secure, and efficient.

B) Design Phase

1) Architecture Design

selected a standard design to ensure usability and security. This system has three main layers: presentation layer (front-end), application layer (back-end) and data layer. The front-end is built using Flask and provides an interactive user interface for data entry and visualization of results. The backend is built in Python and the integration with the machine learning model is built in Jupyter Notebook. Separating these concerns allows each component to be developed and tested independently, thus increasing overall process robustness.

2) UI/UX Design

Creating an intuitive and engaging user interface is critical to user retention and satisfaction. The homepage provides detailed information, while the input form allows for manual or automated data entry through PDF downloads. The user interface features a clean, user-friendly interface to ensure users can easily navigate the app regardless of skill level. Clear, intuitive visuals enhance the user experience and facilitate quick decision-making.

3) Data Model Design

The data structure has been carefully designed to support the functionality of the application. Water quality data is designed to capture nine parameters and metadata related to sample collection. This model makes data efficient and compatible with machine learning models. The data model also includes provisions for storing historical forecasts and user feedback, which facilitates continuous development of forecasting algorithms.

C) Development Phase

1) Model Development and Comparison

The first step involved building and comparing four learning models (SVM, logistic regression, decision tree, and random forest) using water quality data. Jupyter Notebook was used for this comparison. The random forest algorithm was chosen for its accuracy and robustness in handling variable data. The model is then fine-tuned to optimize performance for these parameters.

2) Frontend Development

Using Flask to create a frontend is to make user interaction easier. This includes creating home pages, data entry, and results pages. The PDF upload function is used to simplify data entry and uses a Python library to identify and extract relevant data in the uploaded file.

3) Backend Development

The backend is built in Python and is responsible for data processing, prediction, and integration with the frontend. This involves setting up an API to take data input, call the machine learning model, and return predictions. Implement security features such as data encryption and secure communication protocols to protect user data.

4) Integration and Testing

The combination ensures seamless communication between frontend and backend. Perform a variety of testing procedures, including build tests, integration tests, and user acceptance tests, to verify application performance, functionality, and security. Feedback from these tests is used to tune the system to provide better user experience and accurate predictions.

D) Continuous Improvement and Iteration

1) Feedback Incorporation

Monitor user feedback and performance metrics regularly to identify areas for improvement. Use feedback strategies such as in-app surveys, user reviews, and analytics tools to gather insights into user satisfaction and performance patterns.

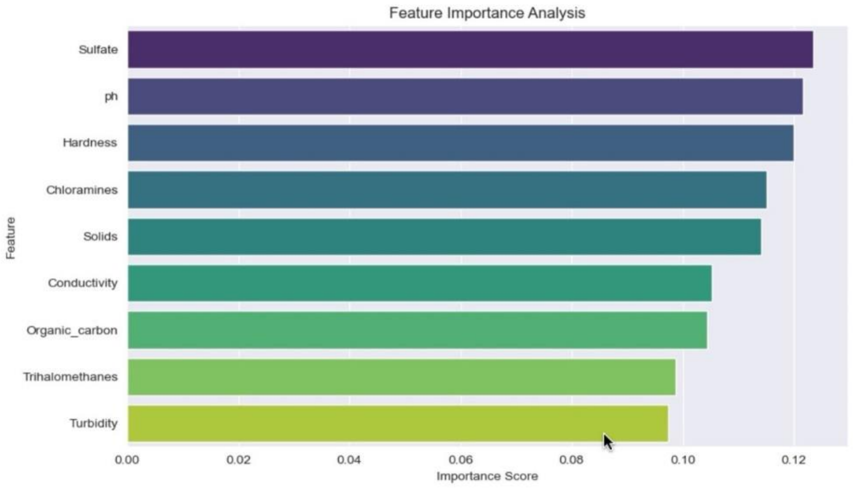
2) Model Refinement

Based on feedback and evaluation results, the prediction model is improved to increase accuracy, validity, and reliability. This may involve retraining the model using updated data, fine-tuning hyperparameters, or investigating alternative algorithms.

3) Feature Enhancement

The front-end interface is constantly updated to include new features, improve user experience and take user feedback into account. This optimization ensures that the application continues to respond to changing customer needs and technological advances. Demonstrates judgment regarding water quality and use.

V. PARAMETERS IMPORTANTANCE ANALYSIS



The shows that 'Sulfates' Contribute most to the result followed by ph, hardness and we can work on it to make the model more accurate.

Figure 5.1: Feature Importance Analysis

VI. DATA FLOW DIAGRAM

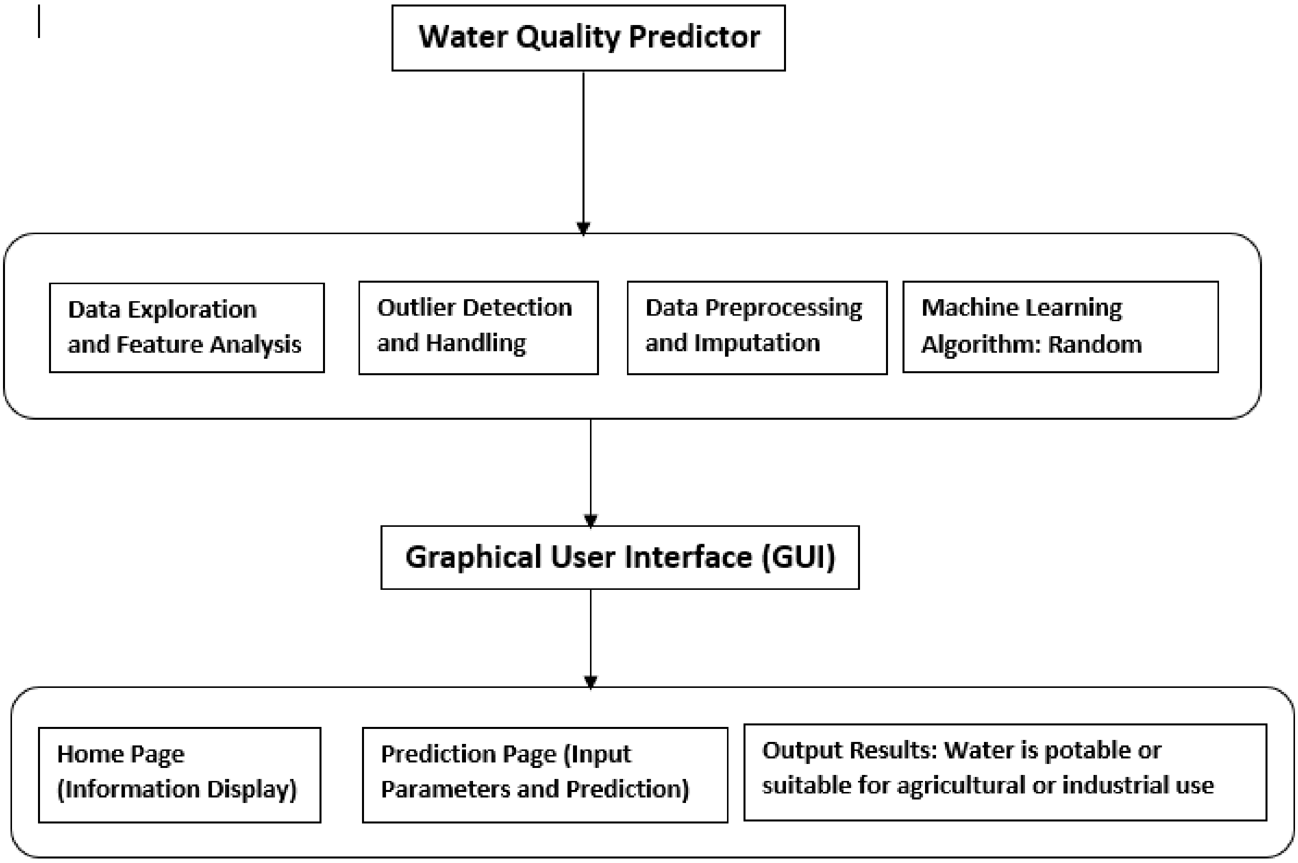

Water Quality Predictor
Data Exploration and Feature Analysis
Outlier Detection and Handling
Data Preprocessing and Imputation
Machine Learning Algorithm: Random
Graphical User Interface (GUI)
Home Page (Information Display)
Prediction Page (Input Parameters and Prediction)
Output Results: Water is potable or suitable for agricultural or industrial use

Figure 6.1: Data Flow Diagram

VII. RESULT

The success of "Water Quality Machine Learning" is a powerful predictor that can identify water quality with high accuracy and confidence. This project uses a carefully selected database of nine parameters including pH, hardness, solids, chloramines, sulphates, conductivity, organic carbon, trihalomethanes and turbidity to provide information on the potability and diversity of water, providing a good understanding of the use for each application. Through rigorous model development and evaluation, the random forest algorithm emerged as the best choice and demonstrated better prediction performance compared to other algorithms. Front-end development with Flask integration simplifies user interaction by providing important content and an intuitive parameter submission form on the home page. Additionally, the new PDF sharing function makes it easier to enter data and can be deleted from the uploaded file, enhancing user convenience. Once the values are submitted, the model processes the data to make accurate predictions about drinking water and recommendations for agricultural or business use. The forecast provides key information for stakeholders to make informed decisions, ensuring access to clean and safe water while supporting environmental sustainability. Based on user feedback and guidance from performance indicators, continuous study and improvements should be made to increase the effectiveness and impact of the project and ultimately support the advancement of water quality measurement and management worldwide.

VIII. CONCLUSION

In conclusion, the "Water Quality Prediction Using Machine Learning" project is a major step forward in the field of water management and provides effective solutions for water quality measurement and decision-making. Through quality data collection, prioritization and modelling, the project successfully leveraged the power of machine learning to predict the drinkability and suitability of water for agricultural and commercial use as nine key points. By leveraging algorithms such as Random Forest and using Flask for seamless front-end integration, the project achieved a high level of user-friendliness and usability, making it useful for the different groups of people used. Integration of new features such as PDF report analysis further improves usability, simplifies user data entry and increases interactivity. Going forward, continuous iteration and improvement through user feedback and performance evaluation will be critical to maintaining the program's impact and effectiveness in solving evolving problems in water quality measurement. The program contributes to public health, environmental sustainability, and equitable access to clean and safe water by enabling stakeholders to better understand and support decision-making processes. As we continue to work together to manage water quality around the world, the "Using Machine Learning for Water Quality Prediction" project demonstrates the evolution of technology being used to solve important problems around the world.

IX. REFERENCES

[1] Şener, Ş., Şener, E. & Davraz, A. 2017 Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). Sci. Total Environ. 584–585, 131–144.

[2] Qishlaqi, A., Kordian, S. & Parsaie, A. 2016 Hydrochemical evaluation of river water quality – a case study. Appl. Water Sci. 7 (5), 2337–2342.

[3] Chen, X., Chen, Y., Shimizu, T., Niu, J., Nakagami, K. i., Qian, X., Jia, B., Nakajima, J., Han, J. & Li, J. 2017 Water resources management in the urban agglomeration of the Lake Biwa region, Japan: an ecosystem services-based sustainability assessment. Sci. Total Environ. 586 (Suppl. C), 174–187.

[4] Nayla Hassan Omer Department of Environmental Engineering, College of Water and Environmental Engineering, Sudan University for Science and Technology, Khartoum, Sudan.

[5] Tchobanoglous G, Schroeder E. Water Quality: Characteristics, Modeling, Modification. 1985.

[6] Chirag Ramesh Shah : Jain Irrigation Systems Ltd. | jains · on demand irrigation projects phd research scholer (environmental engg).

[7] Aldhyani, T.H.; Al-Yaari, M.; Alkahtani, H.; Maashi, M. Water quality prediction using artificialintelligence algorithms. Appl. Bionics Biomech. 2020, 2020. [CrossRef]

[8] Zhang, Q., Liu, J., & Han, D. (2015). Machine Learning Approaches for Modeling Water Quality. Water Research, 47(17),6010-6026. https://doi.org/10.1016/j.watres.2013.09.002

[9] Singh, K. P., Basant, A., Malik, A., & Jain, G. (2012). Artificial Neural Network Modeling of the River Water Quality—A Case Study. Ecological Modelling, 220(6), 888-895. https://doi.org/10.1016/j.ecolmodel.2008.11.015

[10] Liu, W., Liu, G., & He, Y. (2014). Analyzing and Modeling of Water Quality Parameters for Water Supply Systems. Journal Of Water Resources and Protection, 6(14), 1330-1339. https://doi.org/10.4236/jwarp.2014.614122

[11] Petković, D., Sekuloski, P., & Jovanović, V. (2016). Comparison of Different Machine Learning Algorithms in Water Quality Prediction. Journal of Hydrology, 540, 570-579. https://doi.org/10.1016/j.jhydrol.2016.06.057

[12] Tran, T. (2019). Building Machine Learning Web Applications with Flask. Towards Data Science. https://towardsdatascience.com/building-machine-learning-web-application-with-flask-part-1-c3d3bd4e48de