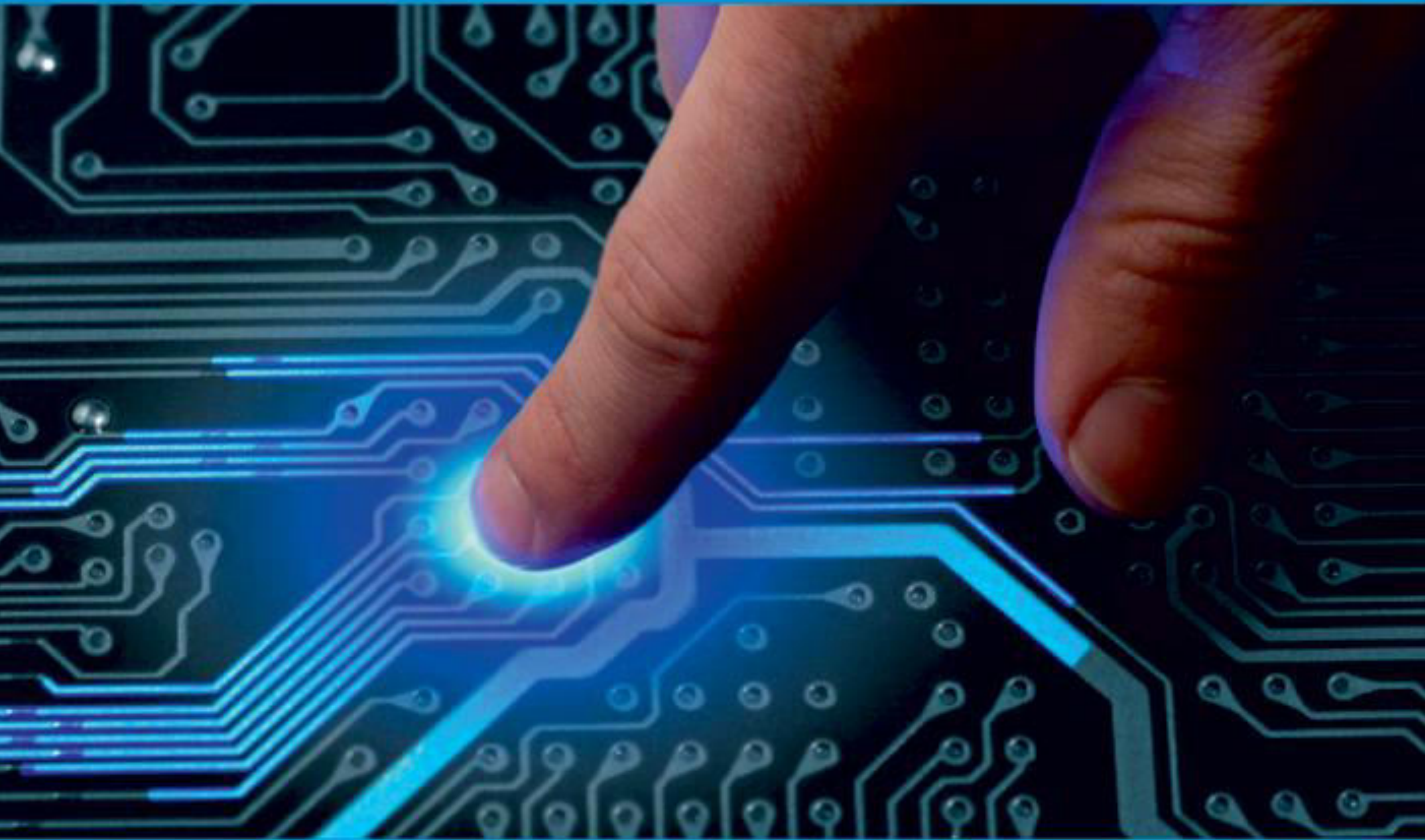
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

A Comprehensive Network Designed for the Continuous Monitoring and Evaluation of Water Quality and Catering to Drinking and Irrigation Requirements

Ashwini B N, Dr.Aravinda T V

PG Student, Department of CS&E., S.J.M. Institute of Technology, Chitradurga, India

Professor, Department of CS&E., S.J.M. Institute of Technology, Chitradurga, India

ABSTRACT: For all living things, including plants and animals, water is essential. The water is not always fit for human consumption, household use, or industrial usage, despite its significance. Water quality may be affected by a multitude of reasons, including industrialization, mining, pollution, and natural disasters. These forces introduce or change different characteristics in the water, which in turn affects its appropriateness for human consumption or other uses.

Water samples intended for irrigation or human consumption must adhere to the criteria established by the World Health Organization. These standards specify the minimum permitted values for several properties. The process of gathering water samples from different locations, evaluating their properties, and comparing the results to established criteria all while following certain protocols for transportation and measurement may be somewhat intimidating. In conclusion, this research suggests a network design for real-time data collection on water characteristics and the autonomous determination of water samples' appropriateness for irrigation and drinking using Machine Learning (ML) techniques. A LoRa-based monitoring network that accounts for land topology has been built. According to Radio Mobile's simulation results, a partial mesh network design is the best option.

Datasets suitable for training ML models had to be created since there were no big, publicly available datasets on irrigation and drinking water. Results from an evaluation of three machine learning models for water classification revealed that Support Vector Machine (SVM) was more appropriate for the purpose of irrigating water and Random Forest (RF) was the most effective for drinking water. Logistic Regression (LR) was shown to have the best overall performance. The water parameter that significantly affects each model's classification performance was determined by simultaneously subjecting three ML models to recursive feature elimination.

KEYWORDS: Decision trees, water quality indices, logistic regression, Support vector machines are among the approaches used.

I. INTRODUCTION

Because it is crucial to human existence, access to clean water has been elevated to the status of a human right. To ensure everyone's continued prosperity in the years to come, the UN established 17 Sustainable Development Goals (SDGs) in 2015. Access to safe drinking water is one of these aims [1]. The importance of guaranteeing universal access to clean water and sanitation is emphasized in our sixth goal [2]. The third SDG is a sustainable development) aims to not only improve access to clean drinking water but also eliminate all preventable diseases, ensuring that everyone may achieve optimum physical and mental well-being. Cholera, typhoid, and diarrhoea are waterborne diseases that have a greater impact on children in undeveloped countries in Africa and Asia [3]. Water is vital to the food and agricultural sectors. Recent estimates indicate that around 10% of the global population suffers from malnutrition, and starvation is the cause of around 45% of infant deaths [5]. This issue is especially severe for developing countries. Therefore, it is of the utmost importance to ensure global food security. By eliminating disparities in food production and distribution, the second Sustainable Development Goal seeks to guarantee that all people have access to sufficient nourishment, as well as by supporting sustainable agriculture. Water is essential for irrigation and animal consumption, which in turn makes it essential for food production and agriculture as a whole. Therefore, it is essential to provide access to and responsible management of water _t for farming purposes.

Water for agriculture and human consumption may be found in a variety of places, including rivers, streams, rainfall, and groundwater (which can be reached by wells and boreholes). The components of water samples taken from a certain source are often heavily impacted by the source's nature and traits. Water supplies may undergo changes in composition and quality due to natural causes as well as chemical byproducts of human operations including mining,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

crude oil extraction, and industrial waste. Domestic uses, animal feeding, crop watering, and other uses for these waters occur in households and farms. This water is very dangerous and may even be fatal if consumed. There must be a system in place to track the water from its point of origin all the way to its final destination. By collecting water samples at each monitoring site, we can assess their quality, often known as their "fitness for use," serves many purposes including as watering crops, drinking by humans and animals, and carrying out residential and industrial activities.

Hydrological indicators (such as sulphate levels, pH, calcium, and oxygen levels), physical characteristics (such as temperature and clarity), and microbiological indicators (such as E. coli, rotaviruses, and Entamoeba) are all part of the water quality evaluation models that have been developed. Water quality specialists from all around the globe have collaborated to develop a set of standards that are used to determine the Water Quality Index (WQI). The National Sanitation Foundation Water Quality Index (NSFWQI) and the Canadian Council of Ministers of the Environment (CCMEWQI) are two of the most widely used water quality indices in North America. The BCWQI and the Scottish Research and Development Department (SRDD) are two water quality evaluation tools used by many European nations. In Africa, noteworthy standards include SANS 241-1 and the Kenya Bureau of Standards (KEBS), whereas in Asia, including India, the Bureau of Indian Standards (BIS) is well-known. A number of these gadgets were reviewed in [6]. It should be emphasized that a large number of these national standards are essentially regional revisions of the criteria established by the WHO [7]. Both South African and World Health Organization standards served as the basis for this effort.

There are a lot of rules that need to be followed when taking water samples, getting them to the lab, evaluating them, and controlling quality in general, Consequently, the process of assessing water characteristics for various samples may be challenging and time-consuming. Some of these processes and associated instructions are provided in references [8] and [9]. If the water sample passes these tests, it means it is safe to drink. In this paper, we provide an alternative model for water potability testing based on machine learning and a cyber-physical network architecture that can monitor a city's water quality in real-time. Comparable to previous research, Our study only examines the physical and chemical properties of water, disregarding any biological factors [10]_[13] [14]. A constraint of our sensor-based methodology inside the Internet of Things framework is our lack of knowledge about any physical sensors with the ability to quantify biological factors, To include microbiological water factors into our model, There are a variety of physical and virtual sensors that might be useful, including the one proposed in [15], which we recognize as having substantial impact.

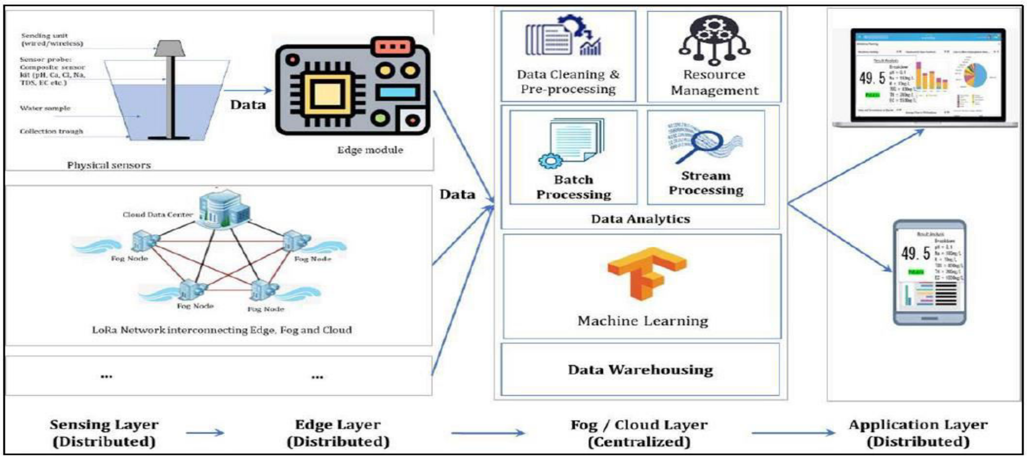


Fig 1.1 Water quality monitoring research based on a theoretical framework

A high-level representation of our suggested 4-layer design is shown in Figure 1.1. The following is a description of the parts that make up this design.

1) The sensor layer, shown in the picture, communicates directly with water samples collected from various sources (rivers, streams, dams, etc.) in order to measure water parameters. Inside a vertical pole called a "sensor probe," it has a bevy of sensors that are all linked to each other. These sensors may be able to monitor temperature, conductivity, turbidity, pH, residual chlorine, and so on, much like what Libelium provides [16]. The transmitting device sends the telemetry data wirelessly or by cables to the Fog Nodes (FNs). An alternative to placing sensors in the water source(s)

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

or in locations where they would be inconvenient is to collect data on water parameter values from the associated treatment facilities.

2) Edge modules make up the second layer; they include low-end processors like ESP32 and Arduino, as well as microcontrollers, as well as single-board computers like Raspberry Pi and Nvidia Jetson. As data pre-processing units, their main function is to, these devices collect, consolidate, filter, and shape data that comes from the sensing layer. In their secondary role as network gateways, they "ferry" telemetry data to the FNs over low-powered long-range network options, such 3G/4G/5G cellular.

3) Layer of Fog or Clouds:
_ In order to reduce latency caused by transmission delays to and from the faraway cloud, fog nodes (FNs) bring compute and storage closer to the data source. These nodes are distributed and small-scale [17]. The aforementioned ML models are used by the FN to classify water samples. In contrast to the Cloud, the Fog has limited computing capability, thus while categorizing water samples, only the most relevant characteristics should be considered. as not all metrics are being monitored, this might be advantageous as it means fewer sensors and, by extension, less computer resources are needed for categorization. Not only that, but FNs are also capable of handling tasks like scheduling and resource management. The data is sent to the Cloud data center when its storage needs are more than what the Fog can handle, or when more complex calculations are needed.

The Cloud is an off-site, highly-efficient computer network that makes computing resources available whenever needed [18]. The cloud hosts the necessary services and software, acts as a data repository, and allows us to do complex data analytics and dashboard boarding.

4) The application layer connects users (This includes entities like customers, end users, and water management agencies) to cloud-based software and services. At this tier, consumers have access to water parameter monitoring software via online and mobile platforms.

Data from dams and treatment plants around Cape Town that store and distribute water, Western Cape, South Africa, will be monitored by the water monitoring network that is planned to be installed in this operation. Data acquired by the network is run using Machine Learning (ML) algorithms to determine whether it is suitable for human consumption or irrigation. A synopsis of the work's main points is as follows:
1) Gather and monitor water quality data in real-time from several storage dams in Cape Town using a pre-existing system. This system accounts for Cape Town's specific topography, which includes hills and other factors that might block radio waves from reaching their destinations.

2) Build and test machine learning models that can automatically determine whether a water sample is fit for its intended purpose by collecting enough data on the quality of irrigation and drinking water.

3) Create models to identify the variables that impact the precision of machine learning algorithms in assessing the excellence of irrigation or drinking water.

II. RELATED WORK

Water quality monitoring and management have become increasingly crucial due to the growing concerns over environmental sustainability and public health. This literature survey aims to explore various studies, reports, and publications related to water quality assessment, monitoring technologies, and SDGs, provide helpful information for creating an all-encompassing water quality monitoring system.

Lee et al. [1] explore the use of sustainable development goal indicators to track progress toward the 2030 Agenda's stated objectives, with an emphasis on the significance of reaching these targets, particularly Goal 6 related to water and sanitation. This study highlights the significance of integrated approaches for planning SDGs, laying the groundwork for effective water quality management.

Another name for the Economic and Social Commission for Asia and the Pacific is ESCAP, is a United Nations agency. is studying Goal 6, which deals with water and sanitation) [2], which explores integrated methods for plans to achieve the sustainable development objectives. This report provides valuable insights into planning strategies and methodologies for achieving sustainable water management objectives.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

In order to ensure the health of both humans and the environment, the World Health Organization has stressed the need of water conservation [3]. For the primary objective of preserving public health and ensuring the availability of clean water, the World Health Organization has established guidelines for the quality of drinking water.

Ho et al. [4] present a case study on water research supporting SDG 6, with a focus on Belgium. This study demonstrates the application of water quality research in advancing sustainable development goals, providing practical insights into water management practices.

The Global Nutrition Report [5] addresses the importance of ending malnutrition by 2030, emphasizing the role of safe drinking water in promoting public health and well-being. This report underscores the interconnectedness between water quality, nutrition, and sustainable development.

Water quality index (WQI) adjustments for simpler computation utilizing multicriteria decision-making methodologies are reviewed by Akhtar et al. [6]. This study contributes to the enhancement of water quality assessment methodologies, facilitating informed decision-making in water resource management.

WHO's guidelines for drinking-water quality [7] provide comprehensive standards and recommendations for ensuring safe drinking water sources. These guidelines serve as a reference for regulatory authorities and water quality practitioners in maintaining water safety and quality.

Water and wastewater analysis according to the standards laid forth by the APH Association and the Federation WE [8, 9], offer standardized protocols for water quality analysis. These methods are essential for accurate and reliable assessment of water parameters in compliance with regulatory standards.

Howladar et al. [10] Assess the water quality around mining-related industrial sites in Bangladesh using multivariate statistics and the Water Quality Index (WQI). This research showcases the real-world use of WQI in assessing environmental harm and guiding restoration initiatives.

Finotti et al. [11, 12] discuss the development of monitoring networks for water resources management in urban areas. These studies highlight the importance of monitoring infrastructure in supporting municipal environmental management and decision-making.

III. METHODOLOGY

3. 1Algorithms

3.1.1 Classes based on decision trees

Decision tree classifiers are useful in many different fields. Their primary strength is in their ability to glean useful descriptive information for decision-making from the provided data. Decision trees may be built using training sets. This kind of generation may be achieved by implementing the following technique taking use of a set of objects (S), whereby every item is a member of the classes C1, C2,..., Ck:

The first step: is to name a leaf with the class that all the objects in S share, such as Ci, if that class is common.

Second Step: Hypothesize a test T with On, O1, O2,... as its viable options. Each element in the set Si has a single potential outcome Oi for the variable T, since every object in the set S has a unique result for the variable T. As a first step, The test extracts subsets S1, S2,..., Sn from the superset S. To build a decision tree, for each outcome, one must repeatedly apply the same method to the set Si. Oh, and T is the root.

3.1.2 Improving gradient

Gradient boosting is widely used in machine learning for two primary purposes: classification and regression. The approach involves combining many weak prediction models, often decision trees, to create a novel prediction framework.There are two locations where you may locate Gradient-boosted trees often achieve better performance than when dealing with poor learners using a decision tree, random forests are employed. There are several similarities between building a boosting model and building a gradient-boosted trees model. However, it surpasses conventional approaches by enabling the optimization of any loss function that is differentiable.

3.1.3 For K-Nearest Neighbors, it is an acronym.

- A categorization method that is both simple and powerful
- 1Uses a measure of similarity to make classifications
- Less-than-parametric

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

- Unmotivated study
- "Learns" only when presented with a test example
- With each new piece of data that needs categorization, we first locate its K-nearest neighbors in the training set.

Example
- The k-closest samples in feature space make up the training dataset.
- The term "feature space" refers to a spatial model that incorporates non-metric factors for classification.
- Instance-based learning is time-consuming and wasteful since it waits for testing or prediction input vectors from the training dataset to be close by.

3. 1. 4 Categorization of logistic regression
Finding the connection between a set of independent variables and a categorical dependent variable is the main focus of logistic regression analysis. When working with a dependent variable that can only take on two potential values, such "yes" or "no," logistic regression becomes necessary. For example, in multinomial logistic regression, the dependent variable may be "Married," "Single," "Divorced," or "Widowed." In practice, the approach is fairly similarly to multiple regression, with the exception of the dependent variable data type.

Logistic regression and discriminant analysis are competitors in the realm of analyzing categorical-response variables. According to most statisticians, logistic regression is generally more appropriate and flexible than discriminant analysis in most situations. The primary rationale for this is because logistic regression does not make the assumption of regularly distributed independent variables, unlike discriminant analysis.

This application offers two distinct approaches for computing logistic regression: one tailored for numerical variables and another specifically designed for categorical variables. In addition to the regression equation, many metrics are provided, such as odds ratios, deviance, probability, confidence bounds, and quality of fit. The thorough residual analysis involves the development of diagnostic reports and visualizations. The tool offers confidence intervals for anticipated values and ROC curves to assist in selecting the optimal classification threshold. By having it automatically classify rows that weren't in the research, you can verify your results.

3.1.5 Unbiased Bayes
Naïve Bayes approach, a kind of supervised learning, relies on a single naïve assumption as its basic concept: that all characteristics are independent of each other, regardless of whether a specific class characteristic is present or absent.

However, this does not influence its apparent robustness and efficiency. The outcomes are similar to those achieved by alternative supervised learning approaches. The literature has put up a number of possible explanations. This lecture is based on an explanation that is mostly based on representation bias. A few examples of linear classifiers include Naive Bayes, logistic regression, linear support vector machines, and linear discriminant analysis. The learning bias, which decides the classifier's parameters, is where the different techniques reside.

The Naive Bayes classifier has a large user base in academia, but it's not that popular among practitioners who are looking for practical results. Among its many advantages, researchers have noted that it is simple to code and put into practice, has easily estimable parameters, learns quickly even on massive datasets, and outperforms competing methods in terms of accuracy. Users don't understand the benefit of this technique as they don't get a model that is easy to understand and apply.

Therefore, In a fresh way, Here are the outcomes of our educational endeavors. Because of its user-friendliness and natural intuition, the classifier is ideal. A thorough examination of the theoretical foundations of the naive Bayes classifier is presented and discussed in this course. After that, a dataset is used to run the procedure using Tanagra. The results of other linear approaches are assessed, in comparison to our own model's parameters. Significantly, the outcomes exhibit a high level of consistency. Primarily, this is the reason why the procedure surpasses other methods. The subsequent step of the assignment involves using several tools on the identical dataset. The package contains the following: R2.9.2, Weka 3.6.0, Knime 2.1.1, Orange 2.0b, and RapidMiner 4.6.0. First and foremost, we must understand the data.

3.1.6 Decision Tree Random
As an ensemble learning method, random forests (infrequently referred to as random decision forests). When faced with classification issues, a random forest is the most often used result among trees. Random choice forests are an

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

excellent answer for cases when decision trees tend to overfit their training set. Although random forests are less exact than gradient enhanced trees, they often outperform decision trees. However, the qualities of data1 might have an impact on their effectiveness.

For a random decision forest, Tin Kam Ho came up with the first approach in 1995. When using Eugene Kleinberg's "stochastic discrimination" classification technique, Ho opted for the random subspace approach.

In 2006, Leo Breiman and Adele Cutler obtained a trademark for their algorithm extension called "Random Forests". As of 2019, the trademark is now held by Minitab, Inc. This modification integrates Breiman's "bagging" concept with Ho's and Amit and Geman's unique approaches to random feature selection and variance control to create an assortment of decision-making trees.

Random forests are often used by organizations as "blackbox" models because to their high predictive accuracy across diverse datasets with little training.

3.1.7 SVM

Independent and identically distributed (iid) training datasets are essential for machine learning methods that rely on discriminant analysis to build a discriminant function. This function is then used to correctly assign labels to newly acquired instances in classification tasks. When working with a feature space that has several dimensions and just requiring posterior probabilities, discriminant methods are more economical in terms of training data and processing resources compared to generative approaches, which are often used for outlier identification in predictions. In the field of geometry, In order to train a classifier, Think of it as solving an equation to determine the best way to split the feature space into its numerous classes using a multidimensional surface.

Support vector machines (SVMs) provide a reliable and optimal hyperplane value by analytically solving the convex optimization issue. This sets them apart from perceptrons and genetic algorithms (GAs), which are two commonly used machine learning classification approaches. The perceptron solutions are significantly influenced by the initial and final criteria. Through the use of a unique kernel during training, which converts the input data into a space of features, the parameters of the SVM model may be precisely defined for a particular training set. Conversely, each time training begins, the GA and perceptron classifier models exhibit dissimilarities. Perceptrons and Genetic Algorithms (GAs) are designed to minimize training errors. To do this, many hyperplanes may be used.

3.2 Modules

3.2.1 Supplier of Services

In order to access this module, the Service Provider will need to enter their login and password. He will have access to certain capabilities once he logs in, like logging in, Read more on Water Data Sets, Train and Test, See the Accuracy Results for Trained and Tested Water Data Sets, as well as a Bar Chart Displaying Those Results, See the Ratio of Predicted Water Quality Detection Types, Download Data Sets for Predictions,

Go over the findings of the water quality detection ratio, See Who Is Online From Afar.

3.2 2 Monitor and Permit Users

Here the administrator may get a full list of all users who have signed up. Here, the administrator may see the user's information (name, email, and address) and grant them access.

3.2.3 Work from afar

All all, there are n users in this module. Prior to performing any actions, users are required to register. After a person registers, their details are automatically added to the database. After he successfully registers, he will be asked to provide his authorized username and password. Users may access their profiles, make predictions on the kind of water quality detector, and register and log in when they log in.

3.3 Dataset

This project makes use of the Kaggle data source for its dataset. Data scientists and machine learning experts from all around the globe come together on the Kaggle platform, which has hundreds of datasets from various businesses and subjects. By scraping information about the top 10,000 datasets on Kaggle, we have created a single source of truth for the most popular and useful datasets on the platform. This dataset is not just a list of names and numbers, but a valuable tool for data enthusiasts and professionals alike, providing insights into the latest trends and techniques in data science and machine learning.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

IV. EXPERIMENTAL RESULTS

The experimental results are shown below,

4.1 Main Page

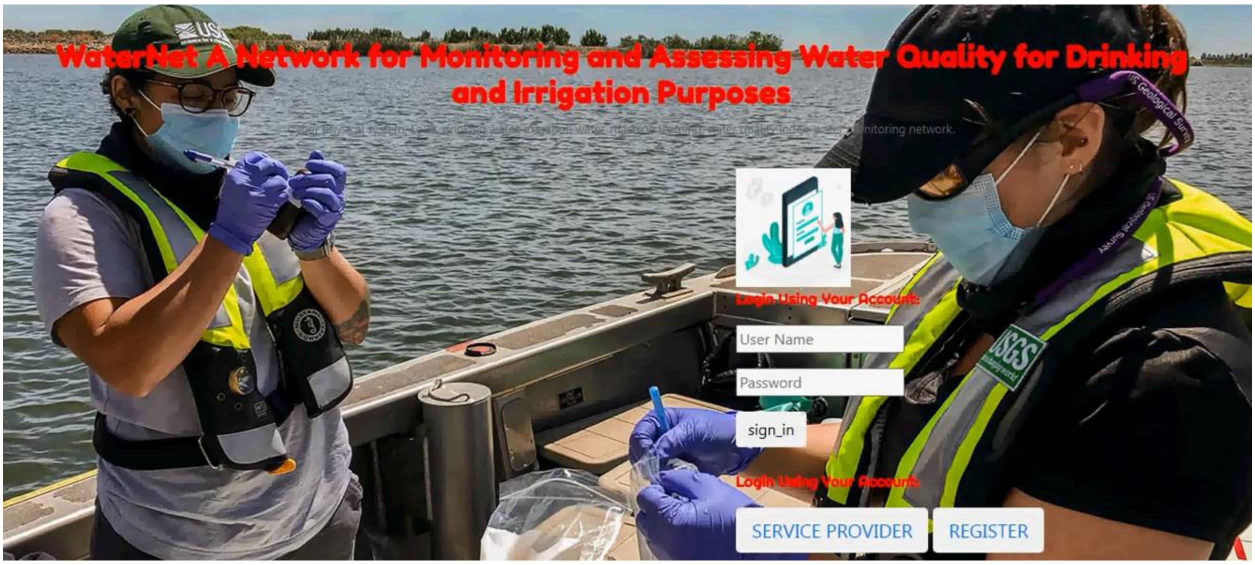


Fig 4.1: Main Page

4.2 Page for logging in

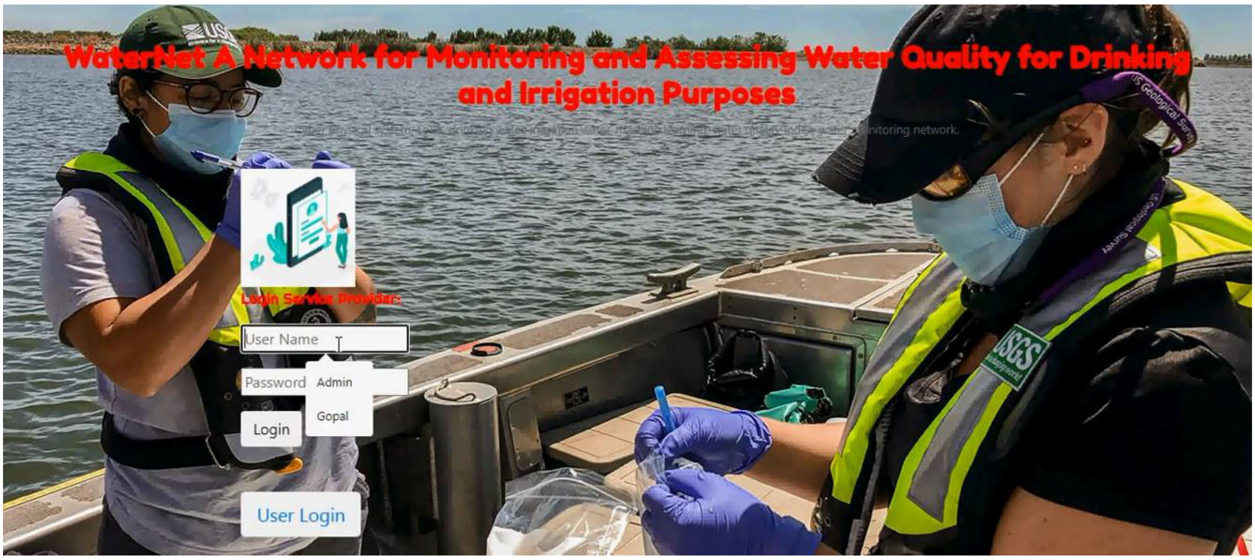


Fig 4.2: Login Page

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

4.3. View all remote users page

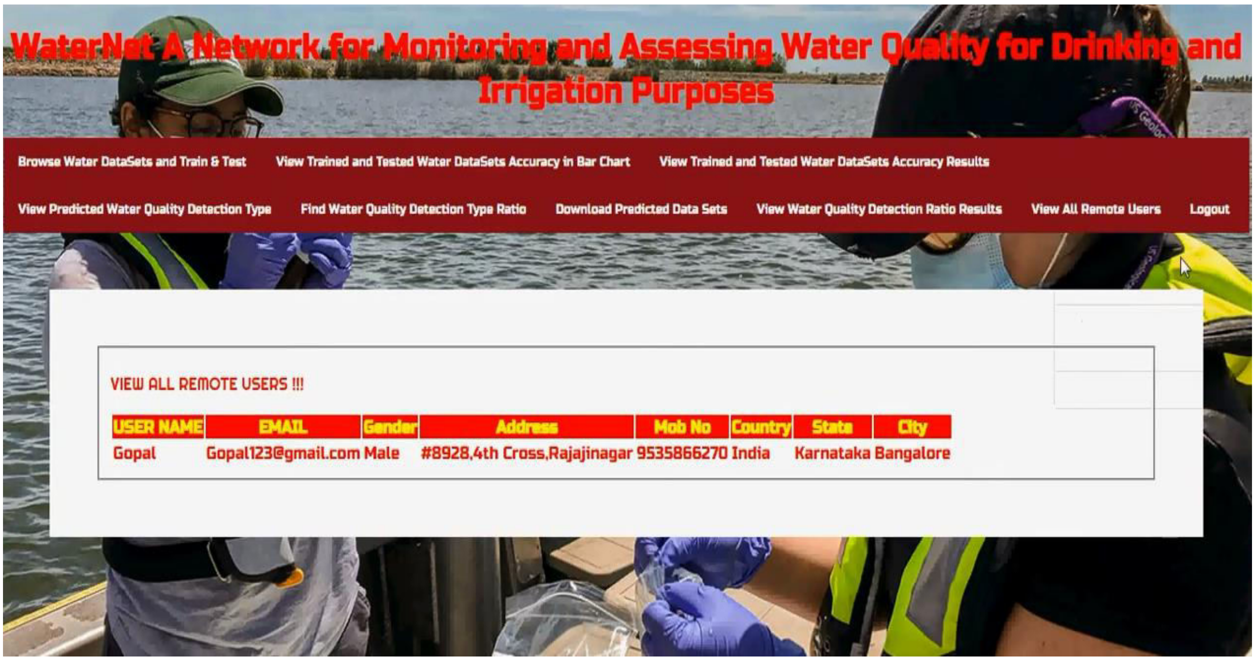


Fig 4.3: View all remote users page

4.4. Register page

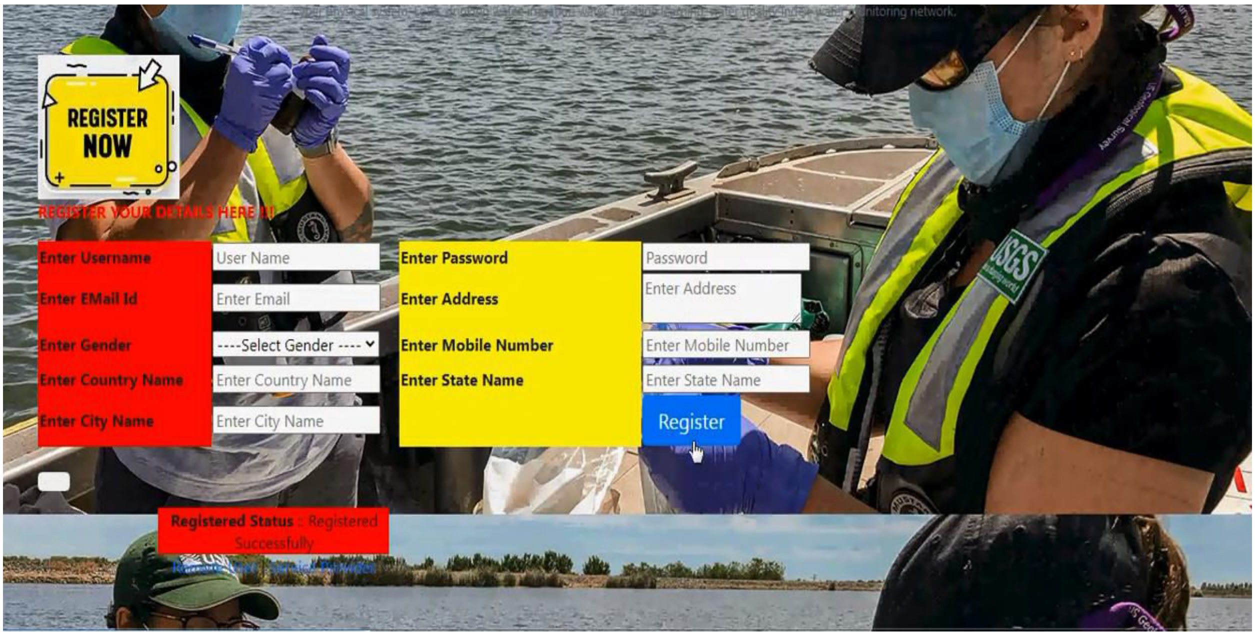


Fig 4.4: Register page

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

4.5. Water dataset trained and tested results

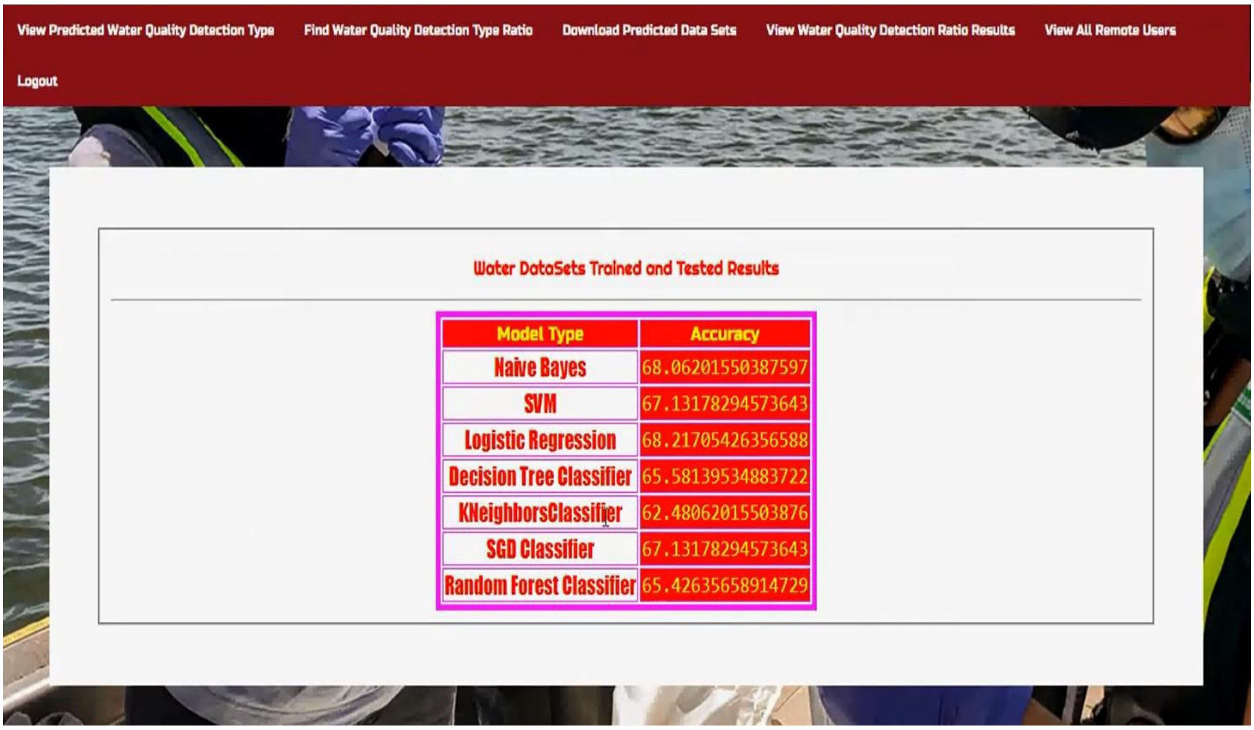


Fig 4.5: Water dataset trained and tested results

4.6. Trained and tested water dataset accuracy result

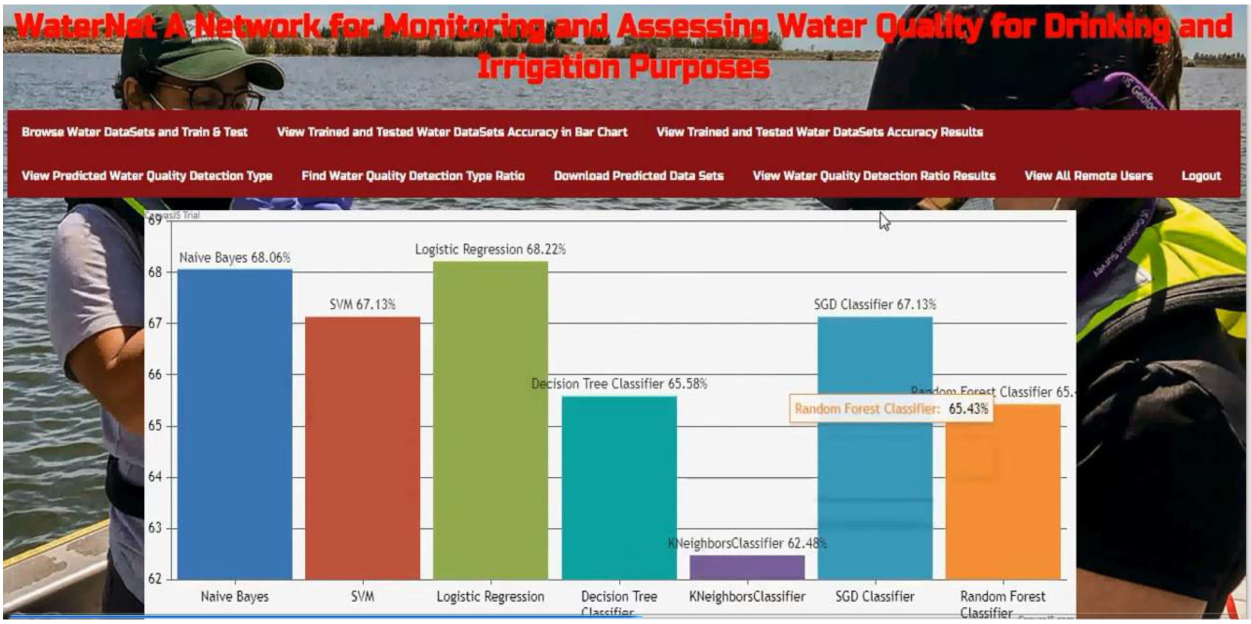


Fig 4.6 a: Trained and tested water dataset accuracy bar chart

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

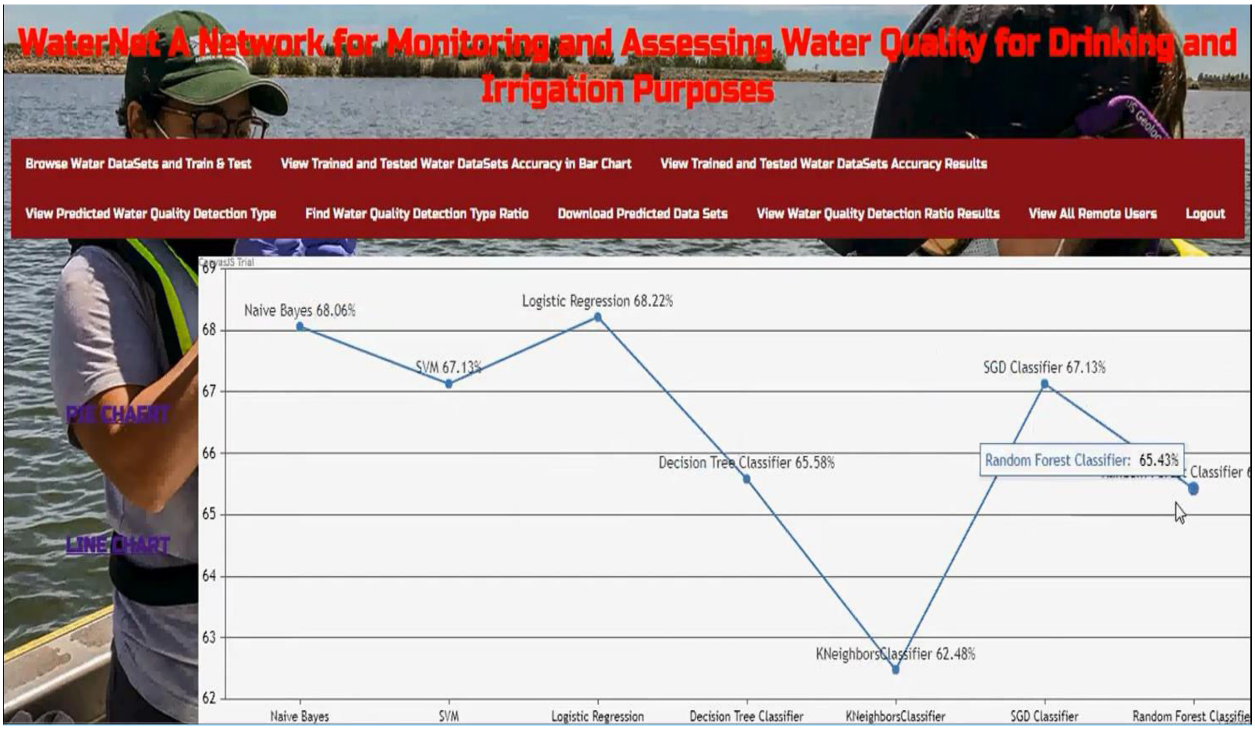


Fig 4.6 b:Trained and tested water dataset accuracy line chart

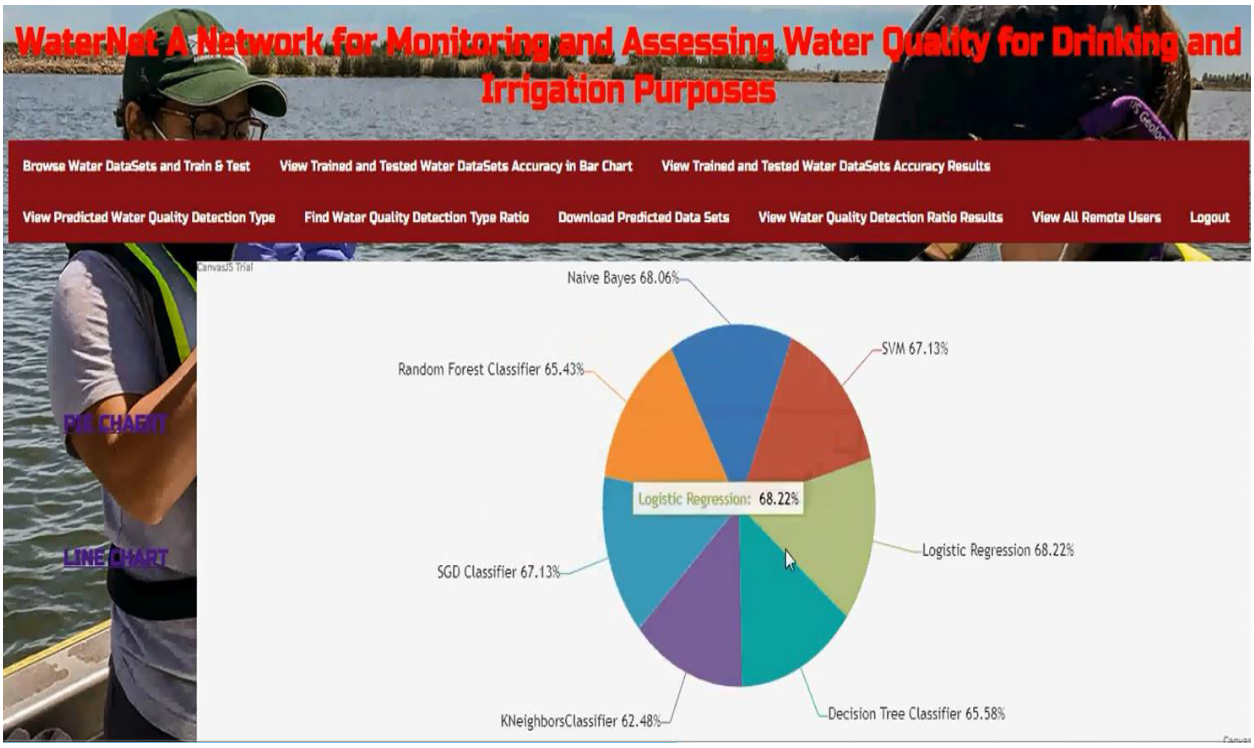


Fig 4.6 c:Trained and tested water dataset accuracy pie chart

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

4.7. Water quality detection ratio results

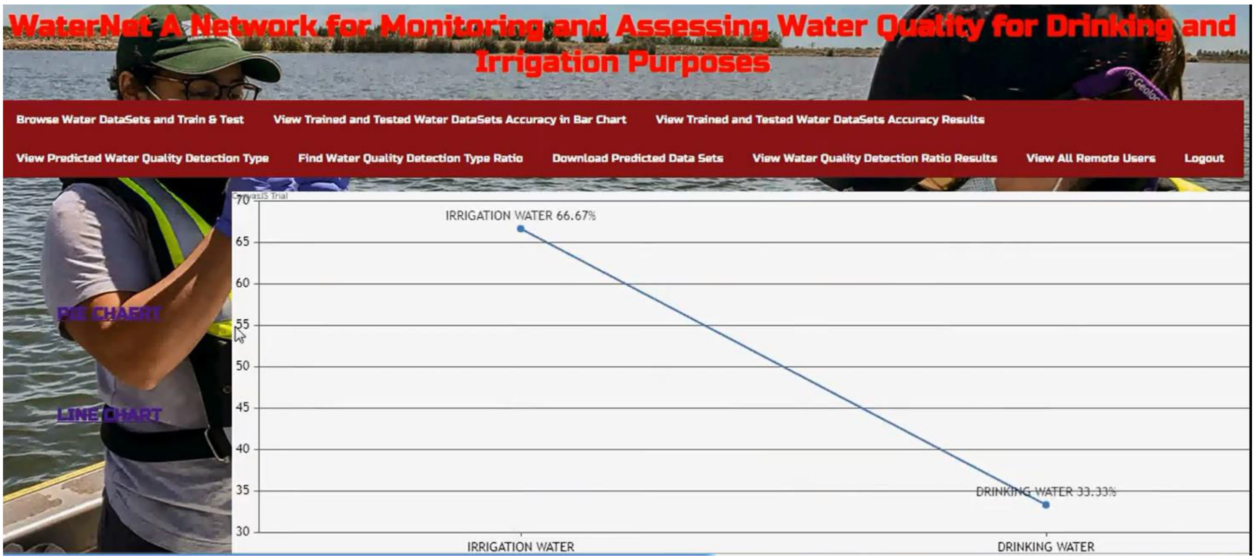


Fig 4.7 a: Water quality detection ratio results in Line chart

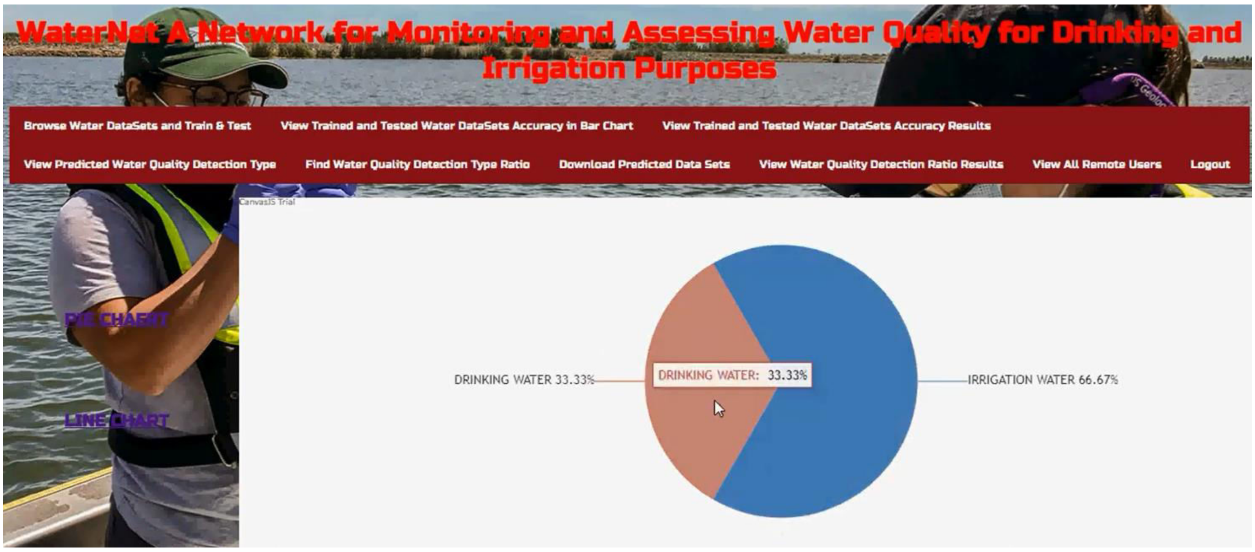


Fig 4.7 b: Water quality detection ratio results in pie chart

V. CONCLUSION AND FUTURE WORK

One of the primary goals of this study was to propose a network that could monitor water bodies in real time and record information about various water characteristics. The next step is to assess the water's condition using ML models. Lo Ra is a low power long range data transmission technology that the creators used to build a water monitoring network in Cape Town. According to Radio Mobile's simulation results, the best network architecture to cover the metropolis is a partial mesh network. Gathered data from this network of sensors should ideally be stored on a server in the cloud. Thereafter, the water's suitability for irrigation or human consumption might be assessed using machine learning algorithms. Train and evaluate three ML models: The methods that were the center of attention in the research were Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR). These models were trained using two datasets that were specifically developed due to the absence of suitable datasets. The test results showed that SVM was more appropriate for irrigation water, whereas LR was superior for drinking water due to its lower false positive and negative values and greatest classification accuracy. Using recursive feature elimination (RFE), Our final step was to look at a model that could tell us which water characteristics affected the ML models' accuracy in classifying data

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207027 |

the most. The results showed that total hardness and pH were the metrics with the least influence on drinking water, whereas SSP was the statistic with the least effect on irrigation water.

Although deep learning models weren't used in this work, the authors acknowledge their potential usefulness. Using other kinds of neural networks and other deep learning models, this study may be further developed in future studies. In addition, the "fitness for use" of water was determined using water quality indicators that were computed by hand; however, unsupervised ML models might be investigated in future works as a possible substitute for these indices. The same logic applies to the identification of important parameters; alternative methods to RFE, similarly to how multi-criteria decision-making has to be thoroughly investigated. Lastly, tracing the origins of water contaminants and integrating consumption prediction models into the water network are two potential ways to further this effort.

REFERENCES

1.B. X. Lee et al., "Transforming our world: Implementing the 2030 agenda through sustainable development goal indicators," J. Public Health Policy, vol. 37, no. S1, pp. 13-31, Sep. 2016.
2.U. ESCAP, "Integrated Approaches for Sustainable Development Goals Planning: The Case of Goal 6 on Water and Sanitation," Bangkok, Thailand, 2017.
3.WHO, "Water. Protection of the Human Environment," Accessed: Jan. 24, 2022. [Online]. Available: www.afro.who.int/health-topics/water
4.L. Ho et al., "Water research in support of the sustainable development goal 6: A case study in Belgium," J. Cleaner Prod., vol. 277, Dec. 2020, Art. no. 124082.
5.International Food Policy Research Institute, "Global Nutrition Report 2016: From Promise to Impact: Ending Malnutrition by 2030," Washington, DC, USA, 2016, doi: 10.2499/9780896295841.
6.N. Akhtar et al., "Modification of the water quality index (WQI) process for simple calculation using the multicriteria decision-making (MCDM) method: A review," Water, vol. 13, no. 7, p. 905, Mar. 2021.
7.World Health Organization, "Guidelines for Drinking-Water Quality," World Health Organization, Accessed: Jan. 12, 2022. [Online]. Available: http://apps.who.int/iris/bitstream/handle/10665/44584/9789241548151-eng.pdf
8.W. E. Federation and A. P. H. Association, "Standard Methods for the Examination of Water and Wastewater," Washington, DC, USA, 2005.
9.L. S. Clesceri et al., "Standard methods for the examination of water and wastewater," Amer. Public Health Assoc. (APHA), Washington, DC, USA, Tech. Rep. 21, 2005.
10.M. F. Howladar et al., "An application of water quality index (WQI) and multivariate statistics to evaluate the water quality around Maddhapara granite mining industrial area, Dinajpur, Bangladesh," Environ. Syst. Res., vol. 6, no. 1, pp. 1-8, Jan. 2018.
11.A. R. Finotti et al., "Use of water quality index as a tool for urban water resources management," Int. J. Sustain. Develop. Planning, vol. 10, no. 6, pp. 781-794, Dec. 2015.
12.A. R. Finotti et al., "Development of a monitoring network of water resources in urban areas as a support for municipal environmental management," WIT Trans. Ecol. Environ., vol. 182, pp. 133-143, May 2014.
13.M. Chilundo et al., "Design of a water quality monitoring network for the limpopo river basin in Mozambique," Phys. Chem. Earth, A/B/C, vol. 33, nos. 8-13, pp. 655-665, Jan. 2008.
14.M. Karamouz et al., "Design of water quality monitoring network for river systems," in Critical Transitions in Water and Environmental Resources Management. London, U.K.: IWA, 2004, pp. 1-9.
15.J. Foschi et al., "Soft sensor predictor of E. Coli concentration based on conventional monitoring parameters forwastewater disinfection control," Water Res., vol. 191, Mar. 2021, Art. no. 116806.
16.Libelium.com, "IoT Solution for Water Management," Accessed: Jan. 27, 2022. [Online]. Available: https://www.libelium.com/iot-solutions/smart-water/
17.K. Ma et al., "An IoT-based fog computing model," Sensors, vol. 19, no. 12, p. 2783, Jun. 2019.
18.I. Odun-Ayo et al., "Cloud computing and quality of service: Issues and developments," in Proc. Int. Multi-Conf. Eng. Comput. Scientists (IMECS 2018), Hong kong, Mar. 2018, pp. 14-16.