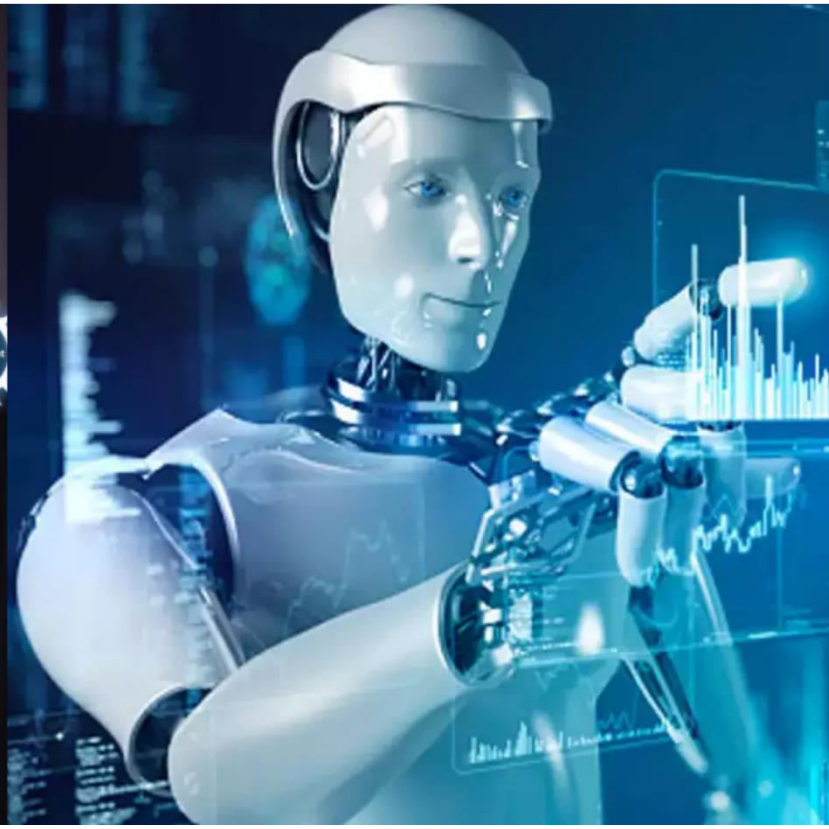
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 7, July 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 7, July 2025

|DOI: 10.15680/IJIRSET.2025.1407080|

Waterborne Disease Prediction using Machine Learning

Ms. J. Jones Miriam[1], Mrs. S. Vinitha Mary[2]

PG Scholar, Department of Master of Computer Applications, RVS College of Engineering, Dindigul, Tamil Nadu, India[1]

Assistant Professor, Department of Master of Computer Applications, RVS College of Engineering, Dindigul, Tamil Nadu, India[2]

ABSTRACT: Waterborne diseases continue to be a major concern worldwide, particularly in areas with limited access to clean drinking water and sanitation facilities. Contaminated water, often laced with harmful chemical and microbial pollutants, is a leading cause of illnesses such as cholera, dysentery, typhoid, and diarrhea. Traditional water testing methods, while accurate, are often expensive, time consuming, and require specialized equipment and expertise. To address these limitations, this project titled "Waterborne Disease Prediction Using Machine Learning" proposes a novel, data-driven approach to detect water contamination and predict related diseases. Using physicochemical parameters like pH, turbidity, hardness, sulfate, chloramines, and conductivity, machine learning models are trained to classify water quality and assess associated health risks. A Random Forest Classifier is used to determine the potability of water, while an XGBoost Classifier is employed to predict possible waterborne diseases if contamination is detected. This dual-layered prediction system enhances accuracy, reduces diagnosis time, and helps in early intervention. The model is built using a labeled dataset, preprocessed and normalized to improve training efficiency and prediction reliability. The system is integrated with a web-based front end, making it accessible to the general public, especially in remote areas. Additionally, the application includes modules for preventive measures, purification techniques, and health tips to guide users on how to respond to unsafe water conditions. A disease mapping system is also included, connecting specific contaminants to potential diseases affecting different parts of the human body. This empowers communities to take immediate action in safeguarding their health. This dual-layered prediction system enhances accuracy, reduces diagnosis time, and helps in early intervention. The model is built using a labeled dataset, preprocessed and normalized to improve training efficiency and prediction reliability. The system is integrated with a web-based front end, making it accessible to the general public, especially in remote areas. Additionally, the application includes modules for preventive measures, purification techniques, and health tips to guide users on how to respond to unsafe water conditions. Water contamination is a significant public health concern, making water quality monitoring essential. Traditional testing methods are costly, time-consuming, and inaccessible in many regions. This project, "Water Quality Detection and Disease Prediction Using Machine Learning," aims to develop an automated system that detects water quality and predicts potential health risks using machine learning models.The system first determines whether water is safe or unsafe using a Random Forest Classifier (RF), trained on physicochemical properties such as pH, hardness, turbidity, and conductivity. If water is deemed unsafe, an XGBoost Classifier (XGB) predicts potential diseases linked to contaminants. By integrating data-driven predictions with disease risk assessment, this system offers an efficient, cost-effective, and scalable solution for water safety analysis.This approach helps users make informed decisions about water consumption, promoting public health awareness and preventive measures.

KEYWORDS: Machine learning, Waterborne disease Prediction, Random Forest Classifier , XG Boost.

I. INTRODUCTION

"Waterborne Disease Prediction Using Machine Learning", addresses this critical need by leveraging data-driven technologies to assess water quality and predict potential health threats. By analyzing key physicochemical parameters such as pH, turbidity, hardness, sulfate, chloramines, and conductivity, machine learning algorithms are employed to determine the potability of water and identify possible waterborne diseases. The project utilizes a dual-model system: a Random Forest Classifier for water quality classification and an XGBoost Classifier for disease prediction, ensuring high accuracy and quick results. Access to clean and safe drinking water is a fundamental human necessity, yet millions of

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 7, July 2025

|DOI: 10.15680/IJIRSET.2025.1407080|

people around the world still face daily challenges due to contaminated water sources. In regions lacking adequate sanitation and water purification infrastructure, the risk of contracting waterborne diseases remains alarmingly high. Illnesses such as cholera, typhoid, dysentery, and diarrhea continue to cause significant health and economic burdens, particularly in low-resource communities. Traditional methods of water quality testing, while accurate, often involve costly laboratory equipment, trained personnel, and time-intensive procedures, making them impractical for widespread or real-time use especially in remote or underdeveloped areas. In response to these limitations, there is an urgent need for smarter, faster, and more accessible solutions to detect water contamination and prevent associated health risks..

This layered analysis provides deeper insight and serves as a proactive health measure, guiding both individuals and authorities in taking preventive actions swiftly. The dataset used in this project includes multiple samples with labeled classifications for potability and corresponding disease associations, where applicable. Preprocessing steps such as handling missing values, normalization, and feature selection are applied to ensure high-quality input for model training. The Random Forest model, known for its robustness and ensemble decision-making capabilities, is trained to assess water quality. Upon detecting contamination, the system transitions to the XGBoost Classifier, which excels in speed and performance for multiclass disease prediction. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to validate both models, ensuring the reliability of predictions.

The possibility of the calculated capability that it utilizes is the wellspring of the expression "strategic around the world still face daily challenges due to contaminated water sources relapse." The essential ability is generally called the sigmoid capacity. The value of this essential ability lies some place in the scope of nothing and one.

II. LITERATURE SURVEY

Xu dong Jia et al. [1] have used both descriptive analysis and machine learning to examine the quality of the water. They start by obtaining the data source from the Kaggle website. After data processing, we perform data mining using the Python sklearn module. The first step is the selection of the machine learning data mining technique using description analysis. The decision tree, Bayesian algorithm, and KNN are the methods we ultimately use to analyze the water data from the Kaggle website. By using a machine learning technique, the data will be divided into available and unavailable categories. Finally, using these three approaches, they have obtained the outcomes of three approaches and conduct a matching comparison and analysis.

K Abirami et al. [2] worked on Water Quality Index (WQI), which serves as a single number to identify the quality of water, would be used in the proposed study to evaluate the water quality. A unique class called the Water Quality Class (WQC) will be created based on the WQI result. pH, temperature, conductivity, dissolved oxygen (DO), biological oxygen demand (BOD), nitrate, and total coliform are the variables used to calculate the water quality index (WQI). While there are numerous machine learning algorithms available for categorization, it is essential to pick the best one. In order to compare the performance of the K-Nearest Neighbor (K-NN), Naive Bayes, Support Vector Machine (SVM), Decision Tree, and Random Forest algorithms, various evaluation measures, including Accuracy score, Confusion Matrix, Precision, Recall, and f1- score, were used.

Bilal Aslam [3] in this study, water samples were taken from wells in the study area (Northern Pakistan) to develop WQI prediction models. Four independent algorithms, i.e., random trees (RT), random forest (RF), M5P, and reduced error pruning (REPT), were used in this study. In addition, 12 hybrid data mining algorithms (combination of discrete, bagging (BA), cross-validation parameter selection (CVPS) and randomized filtered classification (RFC)) were used. Using a 10-fold cross-validation technique, the data were divided into two groups (70:30) to generate the algorithm. Ten random input permutations were generated using Pearson's correlation coefficients to identify the best possible combination of data sets to improve the algorithm's prediction. Variables with correlations performed poorly, while hybrid algorithms improved the predictive power of multiple independent algorithms.

Vinoth Kumar P et al. [4] has studied that indicates potential improvements after analyzing previous water quality prediction studies. In this study, a new intelligent aquaculture system using a machine intelligence model (SAS-MI) was analyzed and proposed in previous water quality prediction works. The new SAS-MI model uses a deep convolutional neural network (D-CNN) and a k-means clustering technique. The k-means clustering used in SAS-MI aggregates the unlabeled data set used for training and testing. D-CNN predicts water quality for intelligent aquaculture using neural automatic feature technology. The performance of the proposed model was evaluated through a comparative study

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 7, July 2025

|DOI: 10.15680/IJIRSET.2025.1407080|

conducted with existing prediction models such as logistic regression, decision trees, XG boost classifiers, k- nearest neighbors and SVM. The SAS- MI model with D-CNN and k-means clustering provides significant results in terms of prediction accuracy, F1 score, and mean squared error (MSE).

Priyanshu Rawat et al. [5] studied provides a comprehensive analysis of the effectiveness of eight different machine learning algorithms in predicting water quality. Algorithms including Gaussian Naive Bayes, Extreme Gradient Boost classifier, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Logistic Regression, Random Forest and Decision Tree were tested using the potable water dataset. The main goal of this study was to find the best accurate machine learning algorithm for water quality prediction and to provide a comprehensive comparison of these methods. Algorithmic efficiency. The results of the study showed that one algorithm performed better than the others, with the lowest root mean square error and the highest accuracy.

Jaswanth Reddy Vilupuru et al. [6] in paper uses artificial intelligence techniques to predict water quality index (WQI) and water quality classification (WQC). Water quality data from India was used in this article. Neural network models such as long short term memory (LSTM) and regression models such as Ridge Regression, Random Forest Regressor with Randomized search CV have been developed to predict WQI. Machine learning models like KNN, Logistic Regression, Logistic Regression using GridSearchCV, XGBoost, SVM and SVM using Grid SearchCV for train test distributions like 70-30, 80- 20 were used in WQC predictions. WQI prediction results showed that Ridge regression achieved the best R^2 of 95.21% with an MSE of 0.11. In WQCpredictions, XGBoost achieved the highest accuracy (97.48%).

Wildan Azka Fillah et al. [7] studied on a benchmark water quality model using ARIMA, SVR and LSTM. It showed that LSTM algorithm gave the best result with less error. The model of the LSTM method can be used to make predictions, such as a seven-day forecast of the next day's pH value, whether it follows the rules or whether it needs to be checked. The company is not penalized for this preventive maintenance.

Sheng Cao et al. [8] has focused on water quality pollution, builds a water quality assessment model to analyze the water quality level, and provides an objective additional forecast on the development of its factors. In that paper, the genetic algorithm mutation factor is incorporated into the PSO algorithm. A least squares support vector machine (LS-SVM) based on an adaptive particle swarm optimization (PSO) algorithm for hyperparameter optimization creates a single water quality classification evaluation model. The fuzzy data granulation method is combined with LS-SVR (Least Square Support Regression) to create a water quality time series model that can predict the changing trend of water quality data over three days. Thanks to theoretical analysis and experimental data, this estimation model and prediction algorithm is faster in terms of training speed and accuracy compared to the traditional BP neural network.

M Uma Maheswari et al. [9] has investigated various supervised machine learning algorithms for water quality detection. Several variables are important in determining water quality, such as pH, hardness, solids, chloramines, sulfates, conductivity, organic carbon, trihalomethanes, turbidity, and potability. water Random Forest (RF) and Decision Tree (DT) are used to determine the caliber of water suitable for human consumption. The standard laboratory method of testing water quality is time consuming and can sometimes be expensive. The algorithms proposed in this study are able to provide an assessment of drinking water quality in a very short period of time. DT height accuracy F1 is 99% while RF score is 87.86% and accuracy is 82.36%. The difference between these two scores is because the accuracy of DT is lower. The proposed method shows its potential for use in real- time water quality monitoring systems, achieving adequate accuracy with a minimal set of parameters. This is necessary to demonstrate the usefulness of this program.

Suma S et al. [10] has developed predictive model to identify water samples that require further analysis tomake the lab technician's work more efficient. WEKA software was used to implement the model, based on secondary data collected from the Kenya Water Institute. Water samples were classified into clean and polluted categories using a decision tree algorithm. When evaluating water quality, the determining factor is its alkalinity and conductivity. Public health and safety depend on the availability of clean drinking water. The researchers used five decision tree classifiers to evaluate the accuracy of the model: J48, LMT, Random Forest, Hoeffding Tree and Decision Stump.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 7, July 2025

|DOI: 10.15680/IJIRSET.2025.1407080|

III. EXISTING SYSTEM

The existing systems for water quality assessment and disease prevention largely rely on conventional laboratory-based testing methods. These involve the manual collection of water samples followed by analysis in specialized labs using chemical, biological, and physical testing techniques. While such methods provide accurate results, they are often time-consuming, expensive, and require access to high-end equipment and trained professionals. As a result, regular monitoring becomes difficult in rural or remote areas where access to laboratories and skilled resources is limited. In most current public health infrastructures, disease prediction based on water quality is not automated, and there is minimal integration between water quality monitoring and health risk assessment. Furthermore, existing systems lack real-time analysis capabilities and often do not provide immediate feedback or preventive suggestions to users. Data collection and interpretation are typically done manually, increasing the chances of human error and delays in action. In some developed regions, IoT-based water quality monitoring systems have been introduced,. The remaining schemes fail toforetell all possible surroundings of the tolerant**.**

Disadvantages
Time-Consuming Testing Process No Disease Prediction

IV. PROPOSED SYSTEM

The proposed system leverages Machine Learning to predict waterborne diseases based on water quality parameters, offering a fast, intelligent, and cost- effective alternative to traditional testing methods. This system is designed to analyze physicochemical properties such as pH, turbidity, conductivity, hardness, sulfate, and other critical indicators to determine the safety of water. Initially, the system classifies the water as either safe or unsafe using models like the Random Forest Classifier. If the water is found to be unsafe, the system proceeds to predict the possible waterborne disease prediction.

Advantages
Easily analyze the disease User-Friendly Interface
Ready to scope with huge dataset. Improved Accuracy

V. SYSTEM ARCHITECTURE

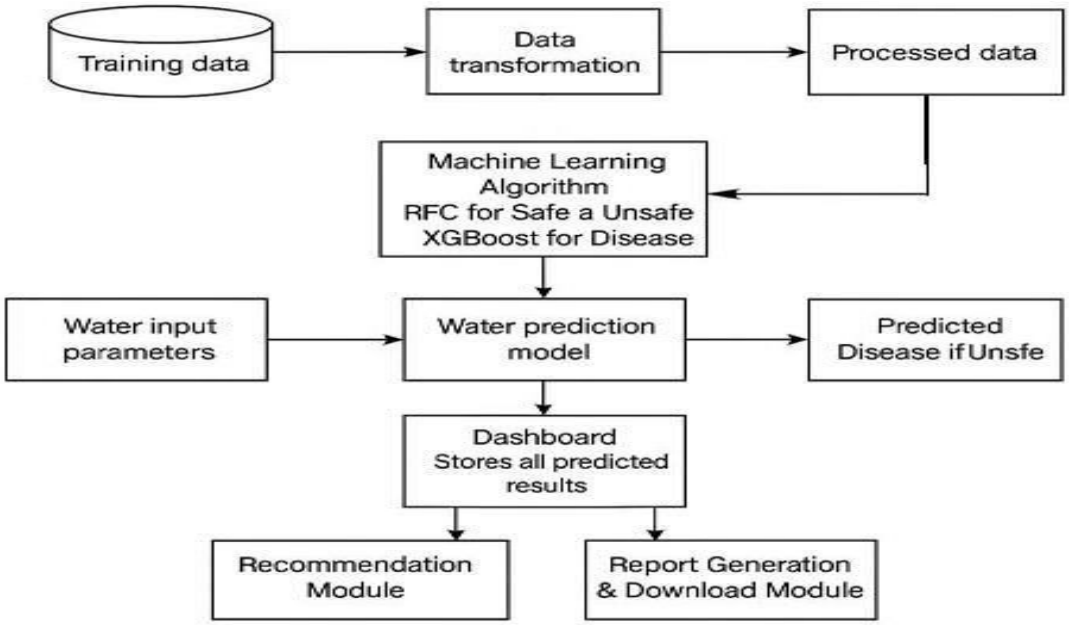


Fig4.1 Architectural Diagram

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 7, July 2025

|DOI: 10.15680/IJIRSET.2025.1407080|

VI. METHODOLOGY

1. Data Gathering,
2. Preprocessing of the data,
3. Feature extraction,
4. Model Building
5. Integration into Prediction
6. User Interface

Data Gathering

Data collection is the first and foundational step in this project. It involves acquiring a comprehensive dataset that includes various water samples along with their physicochemical parameters and associated labels for potability and potential diseases. The dataset features parameters such as pH, turbidity, hardness, sulfate, chloramines, and conductivity.

Pre-Processing of Data

Format, clean, and sample from your chosen data

There are three typical steps in data pre-processing:

i. Designing
ii. Information cleaning
iii. Inspecting

Designing: It's conceivable that the information you've picked isn't in a structure that you can use to work with it. The information might be in an exclusive record configuration and you would like it in a social data set or text document, or the information might be in a social data set and you would like it in a level document.

Information cleaning; is the most common way of eliminating or supplanting missing information. There can be information examples that are inadequate and come up short on data you assume you really want to resolve the issue. These events could should be eliminated.Moreover, a portion of the traits might contain delicate data, and it very well might be important to antonymize or totally eliminate these properties from the information.

Inspecting: You might approach significantly more painstakingly picked information than you want. Calculations might take significantly longer to perform on greater measures of information, and their computational and memory prerequisites may likewise increment. Prior to considering the whole datasets, you can take a more modest delegate test of the picked information that might be fundamentally quicker for investigating and creating thoughts.

Feature Extraction

The following stage is to A course of quality decrease is include extraction. Highlight extraction really modifies the traits instead of element choice, which positions the ongoing ascribes as indicated by their prescient pertinence. The first ascribes are straightly joined to create the changed traits, or elements. Finally, the Classifier calculation is utilized to prepare our models. We utilize the Python Normal Language Tool stash's classify module.

We utilize the gained marked dataset. The models will be surveyed utilizing the excess marked information we have. Pre-handled information was ordered utilizing a couple of AI strategies. Irregular woodland classifiers were chosen. These calculations are generally utilized in positions including text grouping.

Model Building

A **dual-model architecture** is adopted to predict both potability and disease outcomes:

a. Water Quality Classification

b. Disease Prediction

The model building phase employs a dual-model architecture designed to handle both water quality classification and disease prediction effectively. For classifying water samples as either potable or non- potable, a Random Forest Classifier is used. This algorithm is well-suited for structured data, provides high accuracy, and is resistant to overfitting. It also offers interpretability by ranking feature importance, making it a reliable choice for this task. Once a sample is identified as non-potable, the system proceeds to predict the most likely waterborne disease using an XGBoost Classifier.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 7, July 2025

|DOI: 10.15680/IJIRSET.2025.1407080|

VII. ALGORITHMS

1.Random Forest Classifier

The **Random Forest Classifier** is employed for water quality classification, determining whether a given water sample is potable or non-potable based on its physicochemical parameters. Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the class that is the mode of the classes predicted by individual trees. It is particularly effective in handling structured datasets, offers high accuracy, and reduces the risk of overfitting by averaging results across multiple trees. Furthermore, it provides insights into feature importance, helping to understand which parameters most significantly influence water potability.

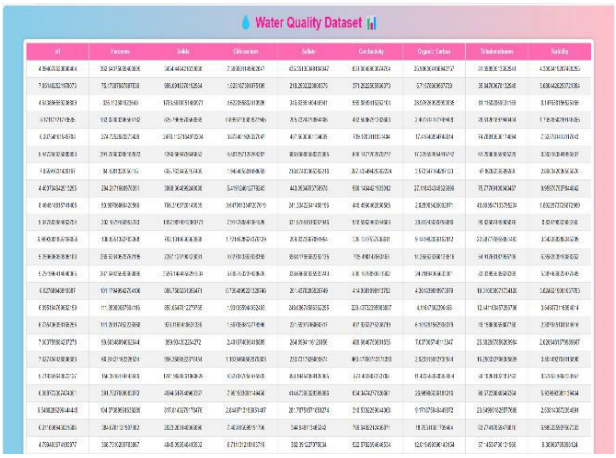
2.XGBoost Classifier

For disease prediction, the project uses the XGBoost Classifier, a powerful and efficient implementation of gradient boosting algorithms. XGBoost is well-suited for multi-class classification tasks and is known for its superior performance, scalability, and computational speed. After identifying that the water is non-potable, XGBoost takes the same input parameters to predict the most likely waterborne disease such as cholera, typhoid, dysentery, or diarrhea. Its ability to handle imbalanced data and prevent overfitting through regularization makes it an ideal choice for accurate disease prediction in this context.

Together, these algorithms form a robust and efficient system that not only assesses water safety but also aids in early detection of potential health risks, ensuring timely preventive actions.

Integration into Prediction

The trained machine learning models are integrated into a user-friendly, web-based prediction system developed using Flask, a lightweight Python web framework. This application allows users to input values for various physicochemical parameters such as pH, turbidity, hardness, sulfate, chloramines, and conductivity. Once the user submits the data, the system performs real-time predictions to determine whether the water is potable or not using the Random Forest Classifier.



If the water is identified as unsafe for consumption, the application then utilizes the XGBoost Classifier to predict the most likely waterborne disease associated with the contamination. In addition to these predictions, the system also provides users with preventive recommendations and health tips based on the results, offering a practical solution for early intervention and risk mitigation. This integration ensures quick, reliable, and accessible water quality and health assessments, especially for communities with limited resources.

User Interface:

The pattern of Information Science and Examination is expanding step by step. From the information science fpipeline, one of the main advances is model sending. We have a ton of choices in python for sending our model. A few well known systems are Carafe and Django. Yet, the issue with utilizing these systems is that we ought to have some information on HTML, CSS, and JavaScript. Remembering these requirements, Adrien Treuille, Thiago Teixeira, and Amanda Kelly made "Streamlit". Presently utilizing streamlit you can send any AI model and any python project easily

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 7, July 2025

|DOI: 10.15680/IJIRSET.2025.1407080|

and without stressing over the frontend. Streamlit is very easy to use.

In this article, we will get familiar with a few significant elements of streamlit, make a python project, and convey the task on a nearby web server. How about we introduce streamlit. Type the accompanying order in the order brief.
pip install streamlit

When Streamlit is introduced effectively, run the given python code and in the event that you don't get a mistake, then streamlit is effectively introduced and you can now work with streamlit. Instructions to Run Streamlit record:
How to Run Streamlit file:

```
You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://192.168.0.115:8501
```

Figure 2

VIII. RESULTS

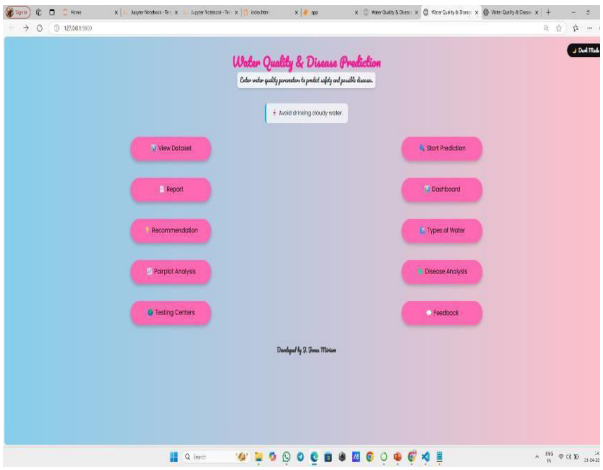


Fig 8.1 Home Screen

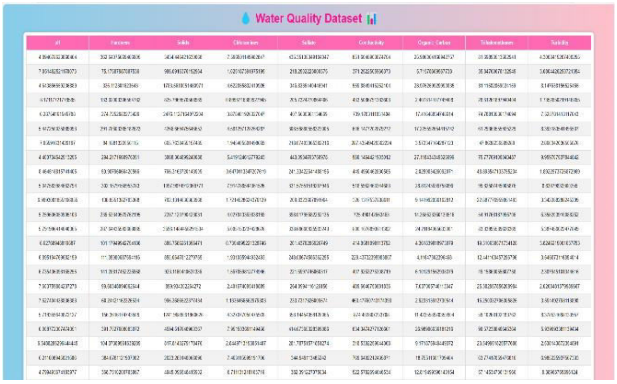


Fig 8.2 Sample Dataset

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 7, July 2025

|DOI: 10.15680/IJIRSET.2025.1407080|

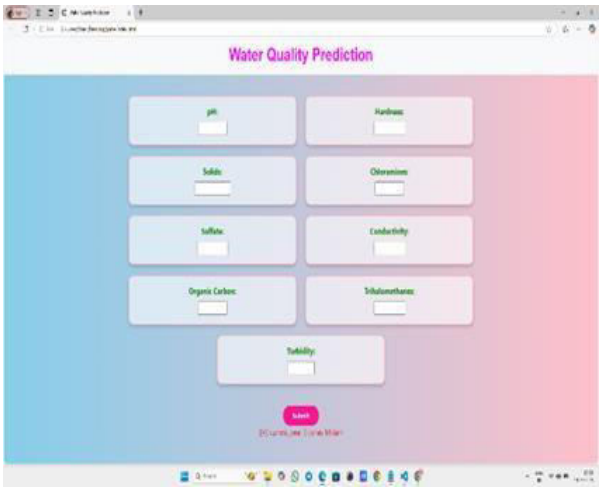
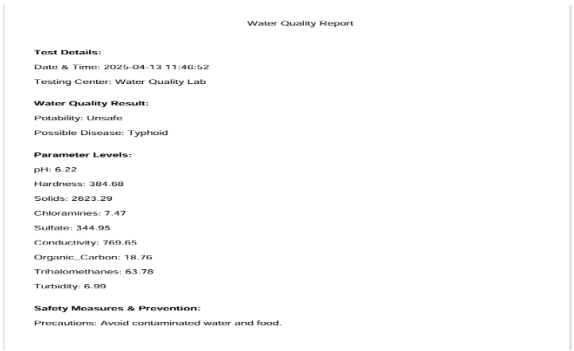


Fig 8.3 Prediction Page

Fig 8.4 Result Screen



Water Quality Report

Test Details:
Date & Time: 2025-04-13 11:46:52
Testing Center: Water Quality Lab

Water Quality Result:
Potability: Unsafe
Possible Disease: Typhoid

Parameter Levels:
pH: 6.22
Hardness: 384.68
Solids: 2623.29
Chloramines: 7.47
Sulfate: 344.95
Conductivity: 769.65
Organic_Carbon: 18.76
Trihalomethanes: 63.78
Turbidity: 6.99

Safety Measures & Prevention:
Precautions: Avoid contaminated water and food.

Fig 8.5 Report

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 7, July 2025

|DOI: 10.15680/IJIRSET.2025.1407080|

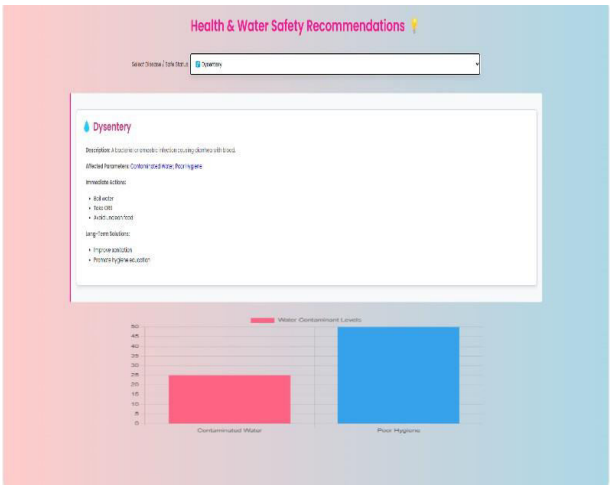


Fig 8.6 Safety Measures Page

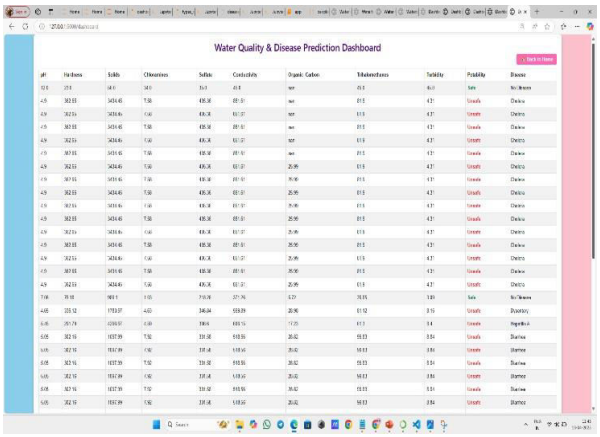


Fig 8.7 Result Storage Dashboard

IX. CONCLUSION

"Waterborne Disease Prediction Using Machine Learning" successfully demonstrates how data-driven technologies can be leveraged to enhance water quality assessment and public health protection. By utilizing physicochemical parameters and applying machine learning algorithms such as Random Forest and XGBoost, the system accurately classifies water potability and predicts potential diseases with the prediction The use of deep learning models and ensemble techniques can further improve prediction performance. In essence, this project serves as a bridge between environmental data science and public health, offering a holistic and tech-driven solution to one of humanity's most critical needs — access to safe, clean water.associated with contaminated water. The integration of these models into a web-based application ensures that users can access real-time predictions and preventive recommendations easily. This approach not only reduces the dependency on traditional laboratory-based testing but also offers a scalable and cost-effective solution, especially beneficial for underdeveloped and rural regions where access to clean water and healthcare infrastructure is limited. The integration of these models into a web-based application ensures that users can access real-time predictions and preventive recommendations easily. This approach not only reduces dependency on traditional laboratory-based testing but also offers a scalable and cost-effective solution, particularly beneficial for underdeveloped and rural regions where access to clean water and healthcare infrastructure is limited. Moreover, the model's adaptability allows it to be trained on region-specific data, making it viable for global application. When integrated with IoT-enabled sensors, it can offer real-time, automated monitoring, greatly enhancing early detection and response capabilities.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 7, July 2025

|DOI: 10.15680/IJIRSET.2025.1407080|

X. FUTURE ENHANCEMENT

In the future, this project can be enhanced by integrating real-time data from IoT-enabled water quality sensors, allowing for continuous and automated monitoring of water sources. Deep learning models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks can be incorporated to further improve prediction accuracy and handle complex temporal patterns in data. Expanding the dataset to include microbiological and meteorological parameters will enrich the analysis and increase the system's reliability. A mobile application version can be developed to ensure wider accessibility, especially in remote and underserved areas. Region-specific customization based on geolocation can provide more accurate, localized predictions. Multilingual support should be added to make the platform user-friendly across diverse populations. An alert and notification system can be implemented to warn users and authorities about potential disease outbreaks in real time. Additionally, incorporating a user.

REFERENCES

[1] X. Jia, "Detecting Water Quality Using KNN, Bayesian and Decision Tree," 2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML), Hangzhou, China, 2022, pp. 323-327, doi:10.1109/CACML55074.2022.00061.

[2] K. Abirami, P. C. Radhakrishna and M. A. Venkatesan, "Water Quality Analysis and Prediction using Machine Learning," 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2023, pp. 10.1109/CSNT57126.2023.10134661.

[3] B. Aslam, A. Maqsoom, A. H. Cheema, F. Ullah, A. Alharbi and M. Imran, "Water Quality Management Using Hybrid Machine Learning and Data Mining Algorithms: An Indexing Approach," in IEEE Access, vol. 10, pp. 119692- 119705, 2022, doi: 10.1109/ACCESS.2022.3221430.

[4] V. K. P, S. K, B. M. D and R. Reshma,"Predicting and Analyzing Water Quality using Machine Learning for Smart Aquaculture," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp.10.1109/ICSCDS56580.2023.10104 677.

[5] P. Rawat, M. Bajaj, V. Sharma and S. Vats, "A Comprehensive Analysis of the Effectiveness of Machine Learning Algorithms for Predicting Water Quality," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 1108-1114, 10.1109/ICIDCA56705.2023.10099968.

[6] J. R. Vilupuru, D. C. Amuluru and G. B. K, "Water Quality Analysis using Artificial Intelligence Algorithms," 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2022, pp. 1193-1199, doi:10.1109/ICIRCA54612.2022.9985650.

[7]W. A. Fillah and D. Purwitasari, "Prediction of Water Quality Index using Deep Learning in Mining Company," 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 2022,10.1109/ICITISEE57756.2022.10057870.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details