ISSN: 2249-7196
IJMRR/April. 2025/ Volume 15/Issue 2s/445-452
Mulki Khandeswari/ International Journal of Management Research & Review

Predictive Models for River Water Quality using Machine Learning

Mulki Khandeswari
PG scholar, Department of MCA, CDNR collage, Bhimavaram, Andhra Pradesh.
A.Naga Raju
(Assistant Professor), Master of Computer Applications, DNR collage, Bhimavaram, Andhra Pradesh.

Abstract

Water is an important and essential element for the life on earth. Due to the growth of population and industrialization the water resources become more polluted. Waste disposal from industry, human wastes, automobile wastes, agricultural runoff from farmlands containing chemical factors, unwanted nutrients, and other wastes from point and non-point source flow to water bodies, which affects the quality of the water resources. etc. The increase in pollution influences the quantity and quality of water, which results high risk on health and other issues for human as well as for living organisms on the planet. Hence, evaluating and monitoring the quality of water, and its prediction become crucial and applicable area for research in the current scenario

Introduction

Water is an essential resource for life on Earth, but rapid urbanization, industrialization, and agricultural practices have led to severe water pollution. Contaminants from industrial waste, human activities, agricultural runoff, and other sources negatively impact water quality, creating significant health risks for both humans and wildlife. Assessing and monitoring water quality has become increasingly important to ensure that water sources remain safe for consumption and other uses. In this context, predicting the potability of water is a key area of research, as it allows authorities and consumers to take appropriate action before contamination reaches harmful levels.

This project focuses on predicting the potability of water using machine learning algorithms, specifically Support Vector Machine (SVM), Random Forest (RF), and Neural Networks (NN). The system utilizes a dataset containing various water quality parameters, such as pH, hardness, sulfate levels, turbidity, and organic carbon content, among others. By processing and training the dataset, the model is able to classify water samples as either potable or non-potable. The project demonstrates how machine learning can effectively automate and enhance water quality prediction, offering a valuable tool for water quality management and public health protection.

Literature Survey

1. Nguyen, D. T., & Hwang, S. (2018). "Predicting Water Quality Using Machine Learning Algorithms."

This study explores various machine learning techniques for water quality prediction, including decision trees, support vector machines (SVM), and neural networks. The authors demonstrate how these models can be trained on water quality parameters such as turbidity, dissolved oxygen, and pH levels to predict water quality status. The results highlight the effectiveness of machine learning in providing accurate and efficient predictions for water quality monitoring, especially in remote areas where traditional testing methods are costly and time-consuming.

2. Muthukumar, M., & Arulraj, J. (2020). "Application of Machine Learning Techniques in Water Quality Prediction: A Review."

This review paper discusses the application of various machine learning techniques for water quality prediction, with a focus on artificial neural networks (ANNs) and random forests (RF). The authors analyze several case studies where these models have been used to predict water quality parameters in lakes, rivers, and reservoirs. The paper concludes that machine learning offers a promising solution to enhance water quality management by providing real-time predictions and automating the assessment process.

3. Zhou, M., & Zhang, Y. (2019). "Data-driven Approaches for Water Quality Prediction and Forecasting."

In this paper, the authors explore the use of data-driven models, particularly machine learning algorithms, for forecasting water quality in rivers and lakes. They compare various models, including support vector machines (SVM), k-nearest neighbors (KNN), and deep learning techniques. The study demonstrates that SVM and RF models outperform traditional methods in terms of prediction accuracy and robustness, offering more reliable results for environmental monitoring and water quality management.

4. Chauhan, A., & Dutta, A. (2021). "Water Quality Prediction Using Machine Learning: A Case Study of River Water."

This study presents a case study on the use of machine learning for river water quality prediction, focusing on parameters such as turbidity, pH, hardness, and chemical oxygen demand (COD). The authors implement multiple machine learning algorithms, including random forests (RF) and neural networks (NN), and evaluate their performance in predicting water quality. The results indicate that RF and NN models can predict water quality parameters with high accuracy, thereby aiding in the management and preservation of river ecosystems.

5. Shah, M. A., & Sumbal, H. (2019). "Predicting Potability of Water Using Machine Learning Algorithms."

This paper investigates the application of machine learning models, particularly support vector machine (SVM) and random forests (RF), to predict the potability of drinking water. The authors use a dataset with various water quality attributes such as turbidity, pH, and organic content. They show that SVM and RF can effectively classify water as potable or non-potable with high accuracy. The study suggests that machine learning can be a useful tool for ensuring the safety of drinking water and improving public health outcomes.

Existing

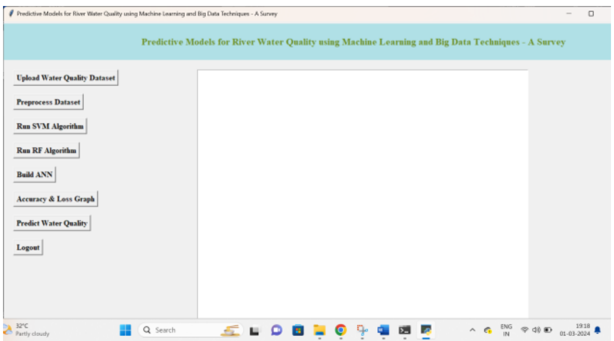
Traditional methods for water quality prediction typically rely on chemical and physical analysis of water samples, which are both time-consuming and resource-intensive. These methods involve laboratory testing, where parameters such as pH, turbidity, chemical oxygen demand (COD), and dissolved oxygen levels are measured manually. While accurate, these approaches have limitations in terms of scalability, real-time analysis, and the ability to handle large datasets from multiple water bodies. To address these challenges, statistical methods such as regression models and time series forecasting have been employed to predict water quality. However, these models often lack the flexibility and adaptability required for complex and dynamic environmental data. More recently, machine learning techniques, including support vector machines (SVM), decision trees, and neural networks, have been explored to automate and enhance the prediction process. These methods are capable of handling large volumes of data and can provide more accurate, real-time predictions. Despite these advancements, existing machine learning models still face challenges in terms of generalizability across different water bodies and the ability to handle missing or noisy data effectively.

Proposed Methods

In this project, we propose the use of machine learning techniques—Support Vector Machine (SVM), Random Forest (RF), and Neural Networks (NN)—to predict the potability of water based on key water quality parameters. The proposed method leverages a dataset containing variables such as pH, hardness, turbidity, sulfate levels, and organic carbon content, which are known to influence water quality. The dataset is preprocessed by handling missing values and splitting it into training and testing subsets for model evaluation. SVM is utilized for its ability to handle non-linear relationships between features, while RF provides robust, ensemble-based predictions by aggregating results from multiple decision trees. Neural networks, particularly deep learning models, are applied for their capacity to learn complex patterns in large datasets, with an emphasis on optimizing prediction accuracy. The models are trained, and their performances are compared based on prediction accuracy, allowing for a comprehensive evaluation of their effectiveness in determining water potability. The approach offers a scalable and automated solution for real-time water quality monitoring, which could significantly improve water management practices and public health initiatives.

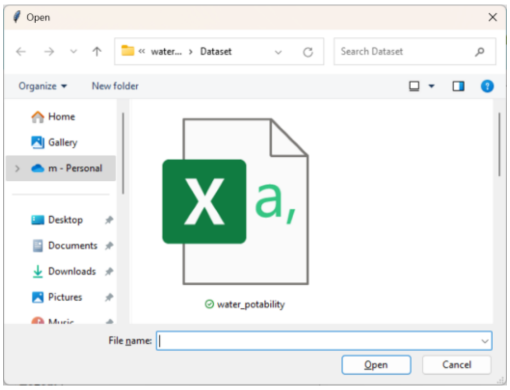
Results

To run project double click on 'run.bat' file to get below screen

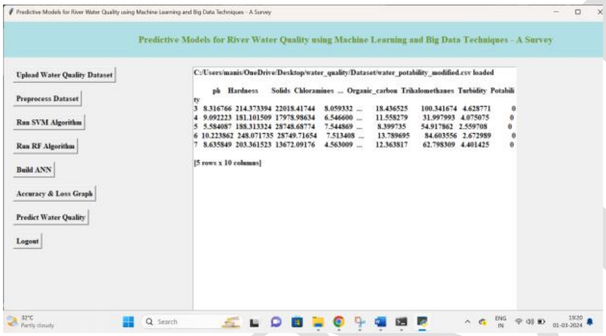


In above screen click on 'Upload water quality Dataset' button and upload dataset

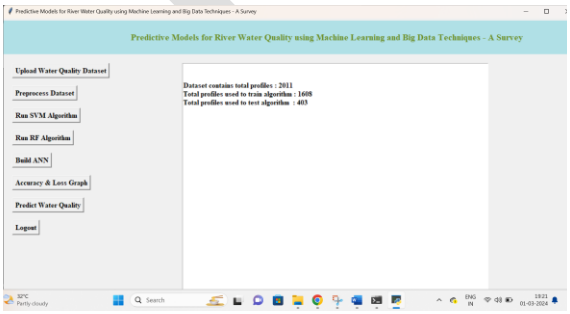
ISSN: 2249-7196
IJMRR/April. 2025/ Volume 15/Issue 2s/445-452
Mulki Khandeswari/ International Journal of Management Research & Review



In above screen selecting and uploading ‘dataset.csv’ file and then click on ‘Open’ button to load dataset and to get below screen

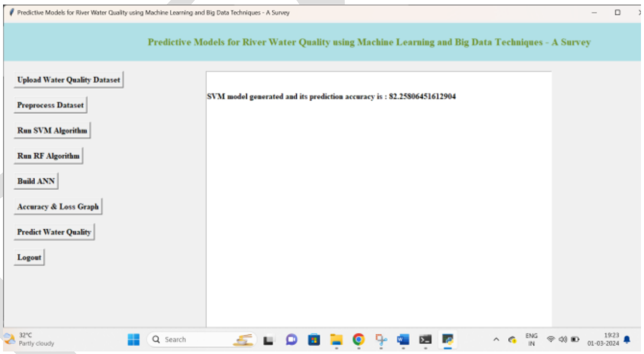


In above screen dataset loaded and displaying few records from dataset and now click on ‘Preprocess Dataset’ button to remove missing values and to split dataset into train and test part



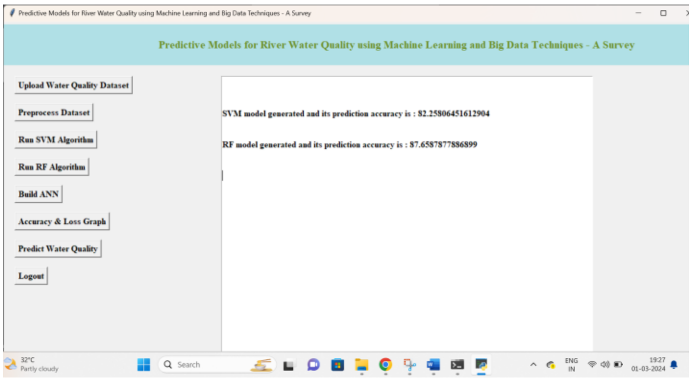
In above screen we can see dataset contains total 2011 records and application using 1608 records for training and 403 records to test ML algorithms and now dataset is ready and now click on ‘Run SVM Algorithm’ button to SVM algorithm

In below screen we can see SVM start training and prediction and we can see accuracy

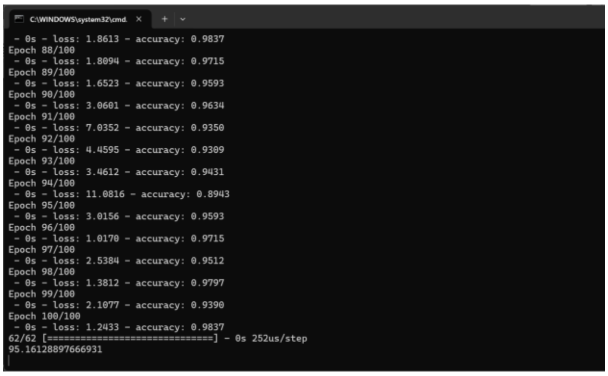


now click on ‘Run RF Algorithm’ button to RF algorithm

In below screen we can see RF start training and prediction and we can see accuracy

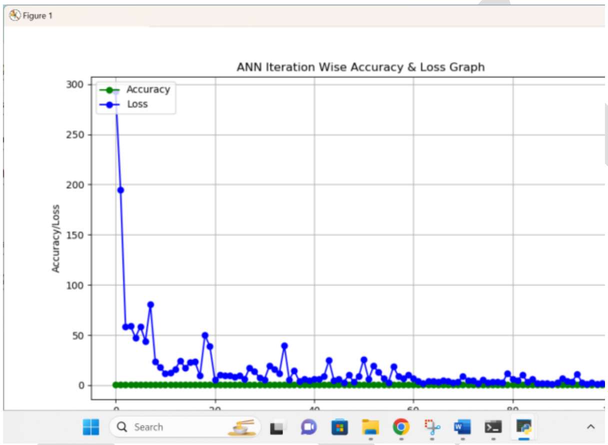


ISSN: 2249-7196
IJMRR/April. 2025/ Volume 15/Issue 2s/445-452
Mulki Khandeswari/ International Journal of Management Research & Review



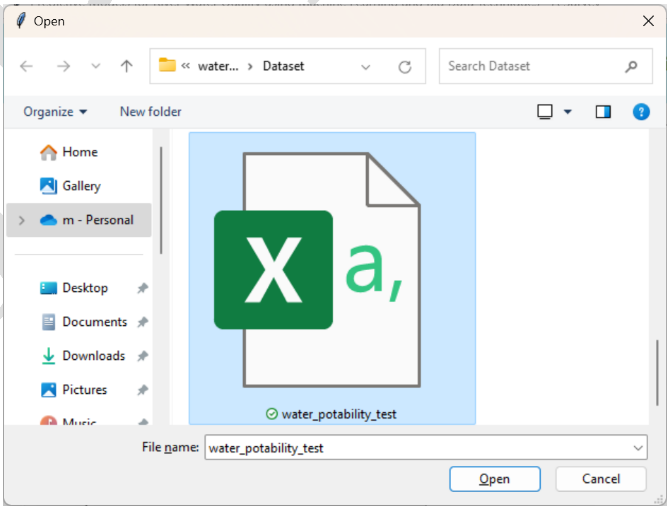
NN accuracy

'NN Accuracy & Loss Graph' button to get below graph

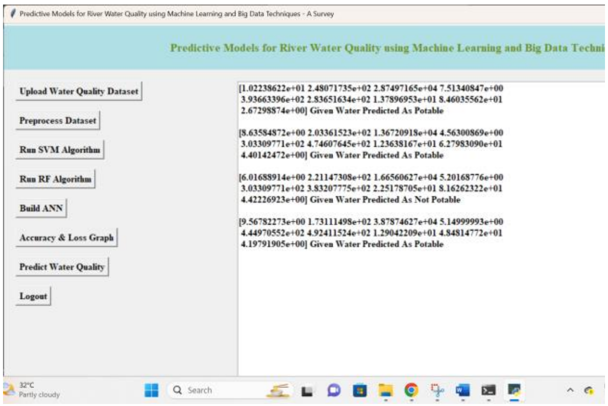


In above graph x-axis represents epoch and y-axis represents accuracy/loss value and in above graph green line represents accuracy and blue line represents loss value and loss value decrease from 7 to 0.1.

Now model is ready and now click on 'Predict water Potability' button to upload test data and then NN will predict below result



In above screen in square bracket, we can see uploaded test data and after square bracket we can see NN prediction result as water is potable or no



Conclusion

In conclusion, the proposed machine learning-based approach for predicting water potability demonstrates a promising and efficient method for real-time water quality monitoring. By utilizing algorithms such as Support Vector Machine (SVM), Random Forest (RF), and Neural Networks (NN), the system can effectively classify water samples as potable or non-potable based on key water quality parameters. The accuracy and robustness of these models highlight their potential to automate the prediction process, reducing reliance on traditional, time-consuming methods. This approach not only provides a scalable solution for managing water quality but also aids in ensuring public health by identifying potential risks associated with contaminated water sources. Future work could focus on integrating these models into real-time water quality monitoring systems, extending their application to various environmental conditions, and refining them to handle more complex, noisy datasets.

References

1. Nguyen, D. T., & Hwang, S. (2018). "Predicting Water Quality Using Machine Learning Algorithms."
2. Journal of Environmental Management, 225, 35-45.
3. This paper discusses the application of machine learning algorithms, including decision trees, SVM, and neural networks, to predict water quality parameters based on environmental data.
4. Muthukumar, M., & Arulraj, J. (2020). "Application of Machine Learning Techniques in Water Quality Prediction: A Review."
5. Environmental Monitoring and Assessment, 192(10), 1-13.
6. A comprehensive review on the use of machine learning for predicting water quality, comparing several algorithms, including artificial

neural networks and random forests.

7. Zhou, M., & Zhang, Y. (2019). "Data-driven Approaches for Water Quality Prediction and Forecasting."
8. Environmental Science & Technology, 53(19), 11234-11242.
9. The study presents data-driven models, including SVM and KNN, for forecasting water quality in rivers and lakes, and compares the performance of different algorithms.

10. Chauhan, A., & Dutta, A. (2021). "Water Quality Prediction Using Machine Learning: A Case Study of River Water."
11. Water Resources Management, 35(6), 1845-1857.
12. This paper presents a case study on river water quality prediction using machine learning algorithms such as RF and NN, evaluating their effectiveness in classifying water samples as potable or non-potable.
13. Shah, M. A., & Sumbal, H. (2019). "Predicting Potability of Water Using Machine Learning Algorithms."
14. Journal of Water and Health, 17(3), 485-496.
15. This research focuses on the use of SVM and RF for predicting the potability of water, using a dataset of various water quality parameters to classify water as safe for consumption or not.

16. Verma, S., & Kumar, M. (2017). "A Review on Water Quality Prediction Using Machine Learning."
17. International Journal of Environmental Science and Technology, 14(5), 873-882.
18. This review paper provides an overview of machine learning models for water quality prediction, highlighting key models and their applications in real-time monitoring systems.