JSRT JOURNAL
JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY
PAGES: 51-54
12/08/25
VOLUME-3 ISSUE-8 AUGUST

Water Quality Classification Using SVM And XGboost Method Machine Learning

Naveen Babshette[1], Vijayalaxmi S[2]

[1]*Student, Master of Computer Application, VTU CPGS, Kalaburagi, India, naveenbabshette@gmail.com*

[2]*Asst.Prof, Master of Computer Application, VTU CPGS, Kalaburagi, India.*

ABSTRACT

Various pollutants have been endangering water quality over the past decades. As a result, predicting and modeling water quality have become essential to minimizing water pollution. This research has developed a classification algorithm to predict the water quality classification (WQC). The WQC is classified based on the water quality index (WQI) from 7 parameters in a dataset using Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost). The results from the proposed model can accurately classify the water quality based on their features. The research outcome demonstrated that the XGBoost model performed better, with an accuracy of 94%, compared to the SVM model, with only a 67% accuracy. Even better, the XGBoost resulted in only 6% misclassification error compared to SVM, which had 33%. On top of that, XGBoost also obtained consistent superior results from 5-fold validation with an average accuracy of 90%, while SVM with an average accuracy of 64%. Considering the enhanced performance, XGBoost is concluded to be better at water quality classification.

Key Words: Water Quality, SVM, XGBoost, Machine Learning.

I.INTRODUCTION

Each cell in the body receives its energy mostly from water, which also controls all the body's functions. 80% of the cerebrum is made up of water. Extreme dehydration may result in mental impairments and a loss of the ability to clearly think. One of the most important regular resources for the survival of all species on Earth is water. Water is used for many different things, such as drinking, washing, and water systems, due to its nature. Water is essential for both living things and plants. Simply put, all organic living things require a huge quantity and exceptional quality of water to exist

Freshwater is a fundamental asset to horticulture and industry for its essential presence. Water quality observation is a key stage in the administration of freshwater assets. As indicated by the yearly report of WHO, many individuals are kicking the bucket because of the absence of unadulterated drinking water parti. It is critical to check the nature of water for its expected reason, whether it be animals watering, compound showering, or drinking water

A tool called water quality testing can be used to locate pure drinking water. This means that for the protection of pure and clean water, the proper water testing is quite important. Water testing is crucial in determining the proper operation of water sources, evaluating the safety of drinking water and deducing the measures to curb the menace.

We can respond to questions like whether the water is fit for drinking, washing, or water systems, to name a few applications, by testing the nature of a water body. It can use the results of water quality tests to examine the nature of water in a location, a state, or the entire country, starting with one water body and moving on to the next. Since irresistible illnesses caused by pathogenic bacteria, infections, helminths, and other parasites are the most well- known and pervasive health danger associated with drinking water, microbiological quality is typically the most urgent issue to be addressed during this process.

When certain synthetic compounds are present in drinking water in excess, health risks result. These synthetics contain nitrate, fluoride, and arsenic. To the client should be given safe drinking (consumable) water for drinking, meal preparation, personal hygiene, and cleaning. To ensure purity at the point of client supply, the water must adhere to standard quality standards.

II.SYSTEM ANALYSIS

Existing System:

In existing systems, water quality classification using machine learning methods like Support Vector Machines (SVM) and XGBoost has been widely explored and implemented. SVM is particularly effective in separating classes by finding the optimal hyperplane that maximizes the margin between data points of different classes.
This makes it suitable for tasks where clear boundaries between classes exist, such as classifying water quality based on distinct parameters like pH levels, dissolved oxygen content, and pollutant concentrations.

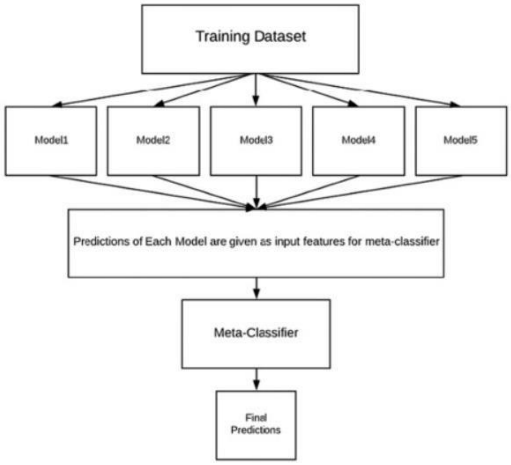
On the other hand, XGBoost (Extreme Gradient Boosting) is a powerful ensemble learning method known for its efficiency and effectiveness in handling complex datasets. It sequentially builds an ensemble of weak decision trees, optimizing the model's performance through boosting, which focuses on instances that previous models have misclassified. In the context of water quality classification, XGBoost can handle nonlinear relationships between variables and capture intricate patterns in water quality data, making it suitable for scenarios where relationships between water quality parameters are not straightforward.

Proposed system
In the proposed system, the classification of water quality is tackled using a combination of Support Vector Machines (SVM) and XGBoost algorithms in a machine learning framework. SVM is employed for its capability to handle high- dimensional data and effectively classify complex datasets by finding an optimal hyperplane that separates different classes. On the other hand, XGBoost, known for its ensemble learning approach, boosts the performance by sequentially improving the weak learners and combining them to create a strong learner. By integrating these methods, the system aims to achieve robust and accurate classification of water quality based on diverse input parameters such as pH levels, dissolved oxygen, turbidity, and chemical concentrations. This approach not only enhances prediction accuracy but also provides insights into the factors contributing to water quality, thereby aiding in effective environmental management and decision-making processes.

III. METHODOLOGY

After understanding the data, processing some attributes, andanalyzing the correlations and predictive potential of the attributes, the major goal of any data science project is model construction. like it was explained in the earlier chapters. Creating a model using the decision tree technique is one of the most straightforward and effective ways of predicting information based on test values. A categorization paradigm called a decision tree, which resembles a flowchart, is frequently employed. Each internal node (non-leaf node) of a decision tree represents a test on an attribute, each branch a test result, and each leaf node (or terminal node) a class label. The root node is the topmost node in a tree. Tree induction, which is the learning or creation of decisiontrees from a class-labeled training dataset, is a method for creating decision trees. Deduction is the process of classifying a test dataset using a decision tree that has already been built. The method of deduction involves applying the test condition to a record or data sample starting at the root node of a decision tree, then, depending on the results of the test, the appropriate branch is proceeded to. This step leads to either a leaf node or to another internal node for which a new test condition is applied. The record or data sample is subsequently given the class label associated with the leaf node.
Decision trees facilitate decision-making under certain conditions and enhance communication. The idea that different actions can result in different operational nature of the situation is easier for computational purposes. Making the best choice possible is beneficial. When instances are represented by attribute values and training data contains errors, the method performs well. In cases where the target function contains discrete output values, it is also relevant. It automatically screens variables, and prepares data with comparatively little user work. Non-linear relations are simple to comprehend and have little impact on the performance of trees. The decision tree is helpful for exploring data and highly suggested when the requirement topredict data is based on expectations.

ARCHITECTURE



IV. RESULTS AND ANALYSIS

Future Works

Decision trees' relative instability in comparison to other decision predictors is one of their drawbacks. A minor change in the data can have a significant impact on the decision tree's structure, which can express a different outcome than what users would receive in a typical event. Hence better prediction models can replace for more robust result One of the most valuable natural resources ever given to humans is water. The ecosystem and human health are directly impacted by the water quality. Water is used for many different things, including drinking, farming, and industrial uses. Over the years, numerous pollutants have put water quality in danger. Predicting and estimating water quality are now crucial to reducing water pollution as a result. Real-time monitoring is unsuccessful because conventionally, water quality is assessed using expensive laboratory and statistical processes. Low water quality calls for a more workable and economical solution. The proposed system builds a model that can forecast the water quality index and water quality class by utilizing the advantages of machine learning techniques. This proposed system is to develop a novel approach for water quality classification using Gradient Boosting Classifier. The method includes the calculation of the Water Quality Index, which is used as a measure of water quality. The proposed approach achieves a high Train Accuracy of 98% and Test Accuracy of 94%. The approach uses various water quality parameters and features such as pH, dissolved oxygen, temperature, and electrical conductivity to classify water into different categories. The model developed in this study is capable of predicting the water quality as Excellent, Good, Poor and Very Poor, which can be used for real-time monitoring and management of water quality. The results demonstrate the effectiveness and accuracy of the proposed approach in predicting water quality, highlighting the potential of machine learning techniques for water quality monitoring and management. The proposed approach can be used in various applications such as water treatment, environmental monitoring, and aquatic life management.

V. CONCLUSION

We are all aware of how vital water is to human health. Knowing the water's quality is crucial because if we consumewater without first making sure it is safe to do so, we run the risk of getting sick. Numerous illnesses that are transmitted through water exist and if we consume non- drinkable water, we risk contracting hazardous diseases. Consequently, the most crucial factor is understanding the water's quality. But this is where the real issue is. We must test the water at a lab, which is expensive and time-consuming in addition to being necessary for determining the water's quality. In this study, we therefore provide a different strategy for predicting waterquality using artificial intelligence.

REFERENCES

1. Jayalakshmi, T.; Santhakumaran, A. Statistical normalization and back propagation for classification. Int. J.Comput. Theory Eng. 2011.
2. Park, J.; Kim, K.T.; Lee, W.H. Recent advances in information and communications technology (ICT) and sensor technology for monitoring water quality.
3. Kangabam, R.D.; Bhoominathan, S.D.; Kanagaraj, S.; Govindaraju, M. Development of a water quality
4. The Environmental and Protection Agency, "Parameters of water quality," Environ. Prot., p. 133, 2001.
5. Kangabam, R.D.; Bhoominathan, S.D.; Kanagaraj, S.; Govindaraju, M. Development of a water quality.
6. Jiang, J.; Tang, S.; Han, D.; Fu, G.; Solomatine, D.; Zheng, Y. A comprehensive review on the design and optimization of surface water quality monitoring
7. Manish Kumar Jha,Rajni Kumari Sah, M.S. Rashmitha, Rupam Sinha, B. Sujatha. Smart Water Monitoring System for Real-Time Water Quality and Usage Monitoring,2018
8. Ashwini K, D. Diviya, J.Janice Vedha, M. Deva Priya. Intelligent Model For Predicting Water Quality
9. Priya Singh,Pankaj Deep Kaur. Review on Data Mining Techniques for Prediction of Water Quality,2017
10. Hadi Mohammed, Ibrahim A. Hameed, Razak Seidu. Machine Learning: Based Detection of Water Contamination in Water Distribution systems,2018.