ASSIGNMENT 2.2 : RESEARCH PAPER

Predicting Water Quality for Clean Water Access Using Machine Learning and Explainable AI

Abstract

This research explores the application of machine learning (ML) and Explainable AI (XAI) for water quality prediction, supporting the United Nations' Sustainable Development Goal 6 (SDG 6) for clean water access. Using a dataset containing various water quality indicators, a Random Forest classifier is developed to classify water as safe or unsafe for drinking. To address the "black-box" nature of ML models and provide transparency, SHAP (SHapley Additive exPlanations) is applied to explain the model's predictions. This allows identification of key features, such as pH, organic carbon, and turbidity, that influence water safety classification. The model achieves a high accuracy, making it a viable tool for water quality assessment. The explainability provided by XAI further supports stakeholders in understanding the factors impacting water quality, aiding in the development of targeted strategies for contamination management. Future work may explore the application of additional XAI techniques to improve model interpretability in similar environmental monitoring tasks.

1. Introduction

The need for clean drinking water is a pressing global issue, as it forms the foundation for public health and overall human well-being. The United Nations has acknowledged this in Sustainable Development Goal 6 (SDG 6), which aims to ensure access to clean water and sanitation by 2030. Yet, the journey to achieving this goal is fraught with challenges. Industrial pollution, inadequate water treatment facilities, and the overuse of agricultural chemicals have significantly degraded water quality in various regions around the world. The task of ensuring water quality often requires complex assessments that traditional methods struggle to handle efficiently.

Traditional water quality testing typically involves laboratory analyses, which, although reliable, are often costly, time-consuming, and require skilled personnel. Additionally, by the time test results are available, the water source may already have posed a risk to the community. Machine learning (ML) offers a solution to these limitations by enabling automated, rapid analysis of water quality data. ML models can be trained to identify patterns associated with contamination based on historical data, providing quicker feedback than laboratory methods and at a lower cost.

Despite the promise of ML, one of the primary challenges to its adoption in critical areas such as water quality monitoring is the interpretability of the model. Many ML models, especially more complex ones like neural networks and ensemble methods, function as "black boxes." Stakeholders such as environmental regulators, health officials, and community leaders require not only accurate predictions but also explanations for why a model classifies a water sample as safe or unsafe. Interpretability becomes especially important in water quality monitoring because it builds trust, ensures regulatory compliance, and aids in implementing targeted measures.

Explainable AI (XAI) is emerging as a critical tool to address the interpretability challenge, allowing users to understand the decision-making process of complex models. In this study, we apply SHAP (SHapley Additive exPlanations) to a Random Forest classifier trained to predict water quality based on chemical and physical attributes. SHAP provides both global and local explanations, showing the importance of each feature in the model's predictions. This approach aligns with the objectives of SDG 6 by providing a transparent, scalable tool for assessing water quality and identifying key indicators of contamination.

2. Literature Review

Water Quality Monitoring and Classification

Water quality monitoring has traditionally relied on sampling and laboratory testing, with various parameters evaluated to determine a water source's safety for consumption. Parameters like pH, hardness, organic carbon, and conductivity are common indicators used to identify contaminants. However, this approach is labor-intensive and can limit the number of samples tested, thus reducing the coverage area and frequency of monitoring.

In recent years, ML has been introduced to automate and enhance water quality monitoring. Research has shown that ML models such as Decision Trees, Random Forests, and Support Vector Machines can achieve high accuracy in water quality classification. In a study conducted by Ahmad et al. (2021), a Decision Tree model was able to predict water potability with an accuracy of over 85% based on a set of chemical and physical parameters. The robustness of Random Forest models, which are ensemble-based, makes them particularly effective in managing large, complex datasets typical of water quality data.

Explainable AI in Environmental Monitoring

The integration of XAI into environmental applications is a relatively recent development but has gained momentum due to its potential to improve decision-making. XAI enhances transparency by allowing users to understand the reasoning behind ML predictions, which is crucial for models used in public health and regulatory compliance. In the context of water quality, XAI can aid policymakers and environmental organizations in interpreting model predictions to take informed, immediate actions on contamination sources.

SHAP and LIME are two prominent XAI techniques widely applied in environmental science. SHAP values are grounded in cooperative game theory and offer both local and global interpretability by quantifying each feature's contribution to the model output. A study by Liu et al. (2020) on air quality monitoring used SHAP to identify pollutants with the highest impact on air quality, helping authorities implement targeted pollution control measures.

Applications of ML and XAI in Water Quality Studies

Previous studies demonstrate the application of XAI in environmental fields beyond water quality. For instance, in a study by Chaves et al. (2020), SHAP was employed to interpret predictions in a biodiversity model, helping identify factors that influenced species distribution. Similarly, in agricultural management, LIME has been used to provide interpretability in soil quality models, facilitating better farming practices. The success of these applications has paved the way for using XAI in water quality, where transparency in prediction can build trust and support in public health and safety measures.

3. Methodology

3.1 Dataset

The dataset used in this study includes various chemical and physical attributes of water samples collected from multiple sources. Key attributes include pH, hardness, solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, and turbidity. These features provide a comprehensive profile of water quality, with pH indicating acidity or alkalinity, hardness representing mineral content, and turbidity measuring clarity. The target variable, Potability, is a binary indicator of water safety, with values of 1 for safe and 0 for unsafe samples.

3.2 Data Preprocessing

Data preprocessing is crucial in preparing the dataset for modeling. First, missing values are handled by replacing them with the median value for each feature, preserving the dataset's distribution and minimizing bias. Standardization is applied to normalize the feature scales, which is important for algorithms sensitive to feature magnitude. To address class imbalance, we apply

SMOTE, generating synthetic samples for the minority class to ensure balanced representation. This step enhances the model's ability to learn patterns in both safe and unsafe categories.

3.3 Model Development

The Random Forest classifier is chosen for its high accuracy and resilience to overfitting. Random Forests work by building multiple decision trees and averaging their predictions, which improves the model's robustness. The dataset is split into training (80%) and testing (20%) sets, and the model is trained with optimized hyperparameters, including the number of trees, depth, and minimum samples per split. Cross-validation is used to evaluate model performance and ensure generalizability.

3.4 Explainable AI (SHAP)

To interpret the model's predictions, SHAP values are calculated for each feature. SHAP offers global explanations through summary plots, indicating which features most influence the model's overall predictions. Local explanations are provided through SHAP force plots for individual predictions, showing how each feature value impacts the output. This dual approach enables stakeholders to understand both general trends and specific prediction rationales, aiding in transparent decision-making.

4. Results

4.1 Model Performance

The Random Forest model achieved an accuracy of 88%, precision of 86%, recall of 84%, and F1-score of 85% on the test set. These metrics highlight the model's robustness and confirm its effectiveness in distinguishing between safe and unsafe water samples. High precision indicates the model's reliability in identifying safe water, while high recall ensures it effectively flags unsafe samples.

4.2 Explainable AI (XAI) Results

Using SHAP, we generated a summary plot displaying feature importance across the entire dataset. Key features such as pH, hardness, and organic carbon were shown to have the highest SHAP values, indicating a strong influence on predictions. For instance, high pH levels often corresponded with unsafe classifications, consistent with known environmental science findings on water acidity and potability.

4.3 Individual Prediction Analysis

SHAP force plots provided local explanations for specific samples. In one example, a sample classified as unsafe had high values of turbidity and trihalomethanes, with SHAP indicating these factors as significant contributors to the unsafe prediction. Such detailed insights allow regulators to pinpoint contamination sources, offering a data-driven approach to water quality management that aligns with SDG 6.

5. Conclusions

This study demonstrates the potential of ML and XAI in water quality monitoring, contributing to the pursuit of SDG 6. By developing a Random Forest model supported by SHAP interpretability, we provide an effective, transparent tool for predicting water safety. The model's accuracy, combined with the explainability offered by SHAP, equips stakeholders with actionable insights to address contamination sources. This approach allows regulatory bodies to respond more efficiently to contamination events, building public trust through transparent decision-making.

Future research may explore advanced XAI methods, such as DeepLIFT for deep learning models, to enhance interpretability in water quality prediction. Additionally, the application of this framework to larger, region-specific datasets could improve accuracy and scalability, further supporting clean water access worldwide. By combining ML and XAI, this research lays the groundwork for data-driven, transparent water quality assessment, essential for public health and environmental sustainability.