■ Water Potability Prediction using AI/ML

AICTE–Edunet–Shell Skills4Future Virtual Internship Program (Green Skills & AI Domain)

Author: Harsh Chauhan

B.E. Information Technology, GEC Modasa

GitHub Repository: Project Link

Abstract

Access to clean drinking water is a major global challenge. This project applies machine learning techniques to predict water potability (safe for drinking or not) using real-world datasets. Models such as Logistic Regression, Random Forest, and XGBoost were evaluated to achieve accurate predictions and gain insights into key water quality parameters.

Introduction

Ensuring safe drinking water is vital for public health. Traditional testing methods can be slow and resource-intensive. Machine learning offers a faster, more scalable solution. The aim of this project is to analyze water quality parameters and predict whether water is potable or non-potable using AI/ML algorithms.

Dataset Description

The dataset used in this project was sourced from Kaggle (Water Potability Dataset). It includes features such as pH, hardness, solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, and turbidity. The target variable indicates water potability: 1 (potable) or 0 (non-potable).

Methodology

The project followed a structured 3-week milestone approach: - Week 1: Data cleaning, preprocessing, Exploratory Data Analysis (EDA), baseline model using Logistic Regression. - Week 2: Implementation of advanced models such as Random Forest and XGBoost; performance comparison. - Week 3: Hyperparameter tuning, feature importance analysis, and final model optimization.

Results & Analysis

The baseline Logistic Regression model achieved an accuracy of approximately 65%. With advanced models, performance improved significantly. Random Forest and XGBoost achieved accuracy levels above 75%, with Random Forest providing the best balance between precision and

recall. Feature importance analysis revealed that pH, sulfate, and trihalomethanes were key factors influencing potability.

Conclusion & Future Scope

This project successfully demonstrated the use of machine learning in predicting water potability. Future enhancements could include integrating the model with IoT-based real-time monitoring systems, deploying the solution as a web or mobile application, and expanding the dataset with regional water quality data. Such developments would support sustainable water management and ensure better access to safe drinking water.

Acknowledgment

This project was completed under the AICTE–Edunet–Shell Skills4Future Virtual Internship Program. Special thanks to the program mentors for their guidance and AICTE for providing this platform.