**Name:** Nishant Golakiya

**SAP**: 60009220150

**Roll Number**: D078

D2-1

Title of the Project**: Predictive Model for Water Quality**

## Aim of the Project

This project aims to create a predictive model capable of assessing water safety based on concentrations of contaminants like aluminum, ammonia, arsenic, and others. By leveraging machine learning techniques, the model will provide quick and accurate evaluations, aiding in proactive water quality management and public health protection.

## Data Description

The dataset comprises various water quality parameters, including concentrations of contaminants such as aluminum, ammonia, arsenic, and others. Each entry includes numerical values representing the levels of these contaminants in water samples. Additionally, there is a binary class attribute "is\_safe," indicating whether the water sample is deemed safe (1) or not safe (0) based on predefined thresholds for each contaminant. This dataset is designed to facilitate the training of a predictive model for water safety assessment, with the goal of accurately categorizing water samples as safe or unsafe for consumption or other uses.

## Data Preprocessing

## For data preprocessing, missing values were dropped from the dataset. Relevant columns were converted to appropriate data types, with 'ammonia' converted to float64 and 'is\_safe' to int64. Numerical columns were selected for further analysis, excluding the target variable. Correlation among features was visualized using a heatmap to identify potential relationships. Additionally, pair plots and histograms were used to visualize the distributions and relationships between selected features.

## Data Modeling:

For data modeling, an XGBoost classifier was chosen due to its effectiveness in handling classification tasks. Hyperparameters were optimized using RandomizedSearchCV with 5-fold cross-validation. The best model was selected based on accuracy score, and its parameters were saved for future use. The model was then trained on the training data and evaluated on both training and testing datasets. Finally, the trained model was saved using pickle for deployment and further analysis.

## Performance Evaluation:

The performance of the XGBoost classifier model was evaluated using accuracy, precision, recall, and F1-score metrics:

Accuracy: 0.93

Precision: 0.94

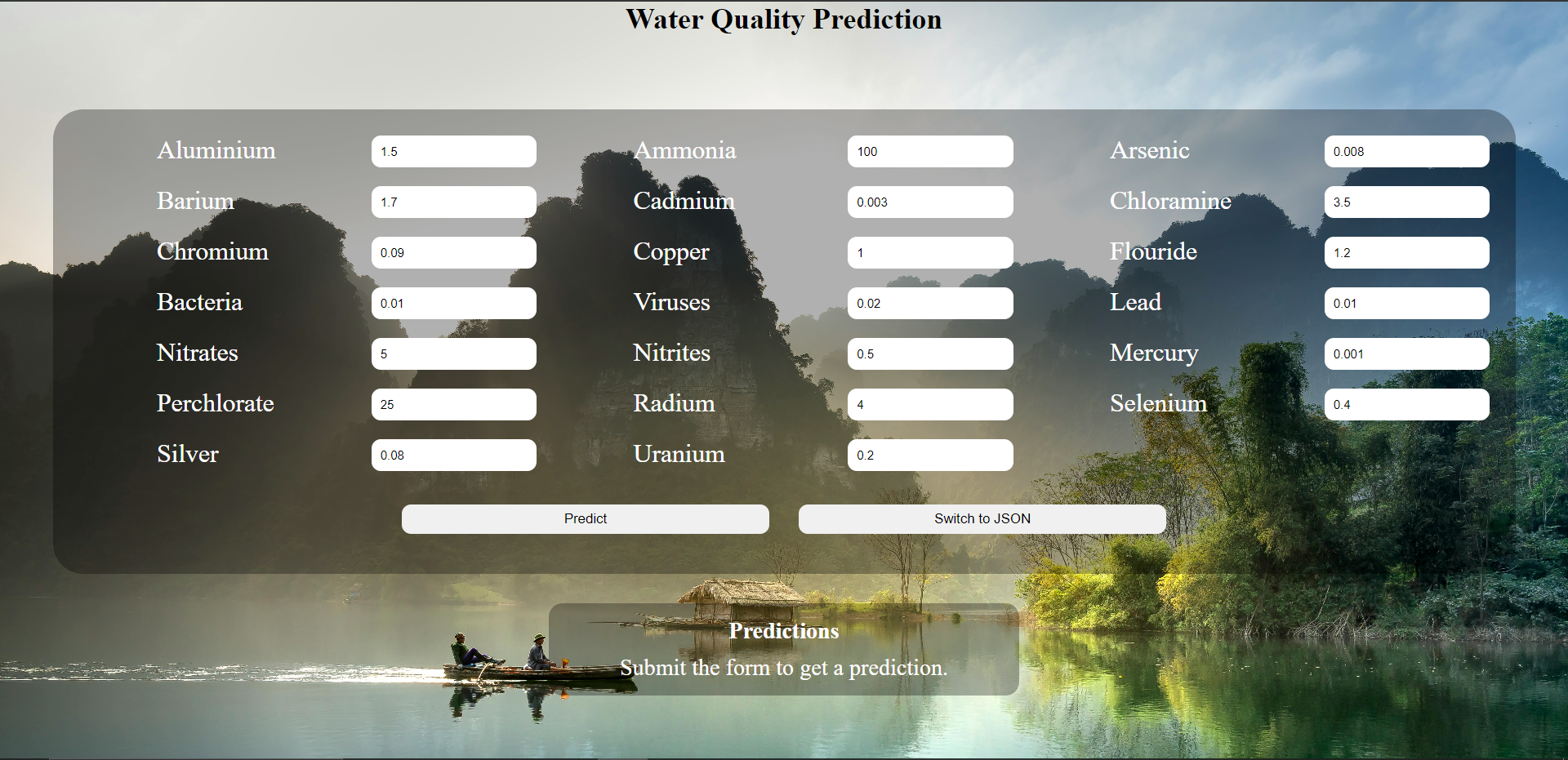
Recall: 0.92

F1-score: 0.93

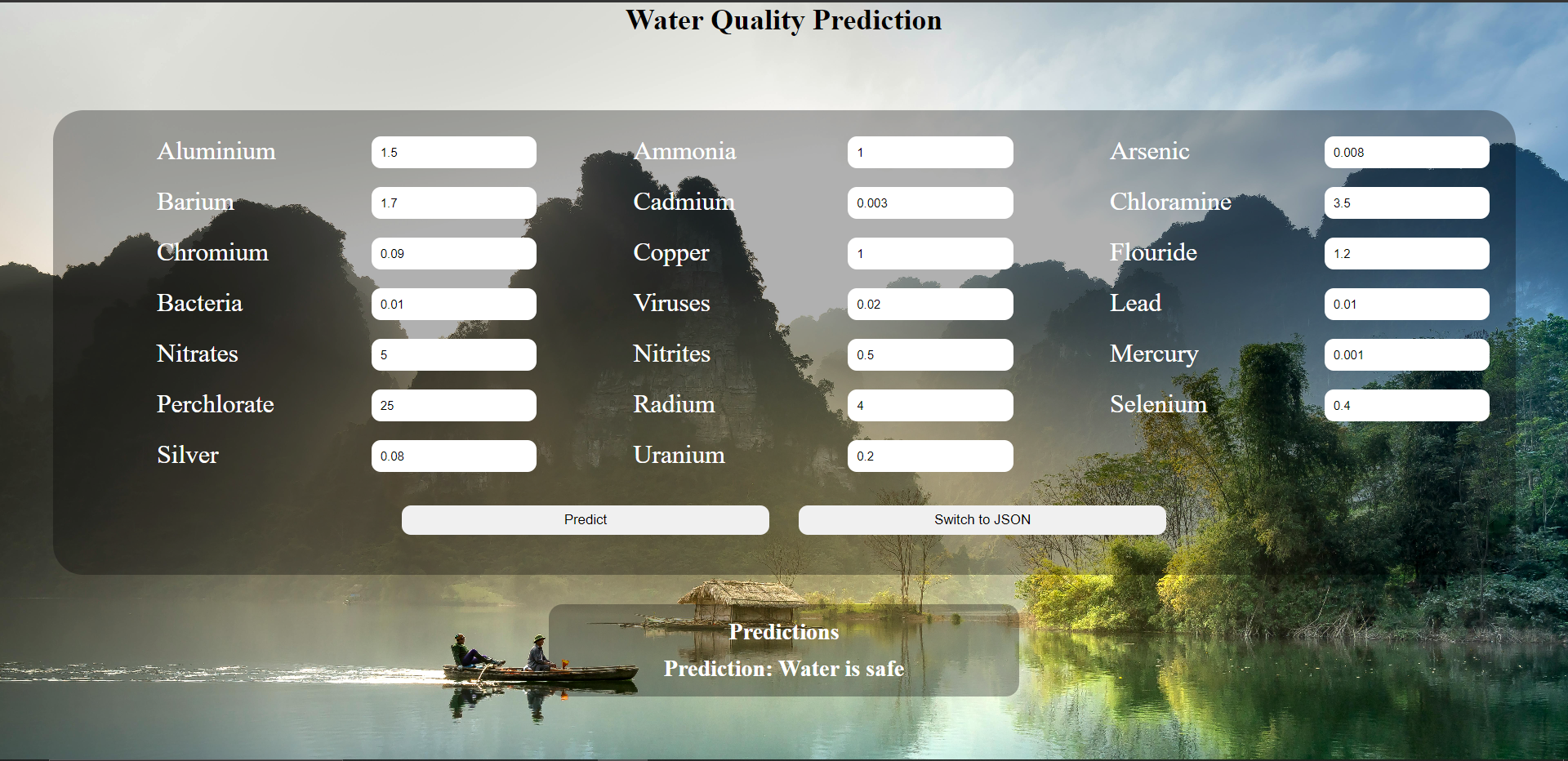
These metrics indicate that the model has high accuracy and performs well in correctly classifying both safe and unsafe water samples. Precision represents the proportion of true positive predictions among all positive predictions, while recall represents the proportion of true positive predictions among all actual positive instances. The F1-score provides a balance between precision and recall, considering both false positives and false negatives.

Overall, the developed predictive model demonstrates strong performance in assessing water safety based on contaminant levels, showing promising results for proactive water quality management and public health protection.

## Flask Deployment:







## Github Link