SYNOPSIS

ON

Water Potability Detection

Submitted By:

Hitansh Mangla - F – 2115000473

Ayush - M – 2115000248

Nikunj - K – 2115000672

Achintya-J-2115000043

Submitted To:

Dr.Sanjay Madan

Technical Trainer

Dept. of T&D

.

Water Potability Detection

Water Potability Detection using Machine learning :

A Water Potability detection using Machine Learning aims to assess the safety of a specific water source for drinking by employing various ML algorithms to analyze the quality of water.

This project aims to automate the assessment process by developing a model that can predict water potability based on water quality data.

Objective and Scope: -

Objective: -

The primary objective of this project is to utilize machine learning techniques to assess and confirm the potability of a specific water source, such as a local well, tap water, or a small scale water supply system. This project aims to automate the assessment process by developing a model that can predict water potability based on water quality data.

Scope: -

The project's scope is focused on a specific water source or a localized area, and it involves collecting water quality data, developing an ML model and assessing the potability of the water source.

Key Activities: -

- Data Collection
- Data Preprocessing
- Feature Selection
- Model Development
- Model Training
- Model Evaluation
- Potability Prediction

Key Steps

- ## Data Collection:

Gather a dataset of water quality measurements, which should include various parameters such as pH, turbidity, hardness, chloride, sulfate, dissolved solids, and more.

- ## Data Preprocessing:

Clean and preprocess the dataset, handling missing values, outliers, and data normalization

- ## Feature Selection/Engineering:

Choose relevant features (parameters) that are most likely to influence water potability.

- ## Data Splitting:

Split the dataset into training, validation, and testing sets. The training set is used to train the model, the validation set to fine-tune hyper parameters, and the testing set to evaluate the model's performance.

- ## Model Selection:

Select an appropriate machine learning algorithm for binary classification. Common choices include logistic regression, decision trees, random forests, support vector machines, or neural networks.

- ## Model Training:

Train the selected model on the training data, using the water quality parameters as input and the potability labels as the target variable.

- **Model Evaluation:**

  Assess the model's performance on the validation set, considering metrics like accuracy, precision, recall, F1 score, and the receiver operating characteristic (ROC) curve.

- **Hyperparameter Tuning:**

Optimize the model's hyperparameters to improve its performance. Techniques like grid search or random search can be used.

- **Model Testing:**
  - Evaluate the final model on the testing dataset to assess its real-world performance.

- **Model Deployment:**

  Once the model meets the desired performance criteria, it can be deployed in a real-world setting, such as a water quality monitoring system.

Technology Used:

Python, Machine Learning Algorithm

Supporting Technology:

Pandas, Numpy, Matplot.

Implementation Plan

- Week 1 – Data Collection
- Week 2 – Data Preprocessing
- Week 3 **–** Model Training and Model Development
- Week 4 -Model Evaluation
- Week 5- Potability Prediction

Team Members:

Ayush – Data Analyst

Nikunj – Data Analyst

Hitansh Mangla –Model Trainer

Achintya Gupta – Model Trainer

Resources Required:

1. Integrated Development Environments (IDEs) like Visual Studio Code or Jupyter Notebook.

2. Machine Learning Algorithm for training the model.

3. Kaggle for the data to train the model.

4. Version control system like Git for code management.

Expected Outcomes:

By the end of the project, the primary objective is to deliver a model that gives the most accurate prediction for the water potability and optimizes operational efficiency for administrators. The expected outcome is to have a model that determines whether the water is safe for consumption. The outcome should provide clear and concise

information, backed by data and evidence, to help make informed decisions regarding water quality and public health.

Project Supervisor:

Dr.Sanjay Madaan

Technical Trainer

Dept. of T&D

Conclusion:

The primary output of this mini-project is a report that includes the model's predictions of water potability, the model's performance metrics, and recommendations for actions to ensure potable water. The report can be shared with relevant authorities and stakeholders. This project demonstrates the use of machine learning to automate the assessment of water potability making it a valuable tool for monitoring and ensuring the safety of water sources.