Water Quality Analysis

Phase 1: Problem Definition and Design Thinking

Problem Definition:

The quality of drinking water is a crucial factor for human health. The objective of this study is the assessment of physical, chemical and bacteriological quality of the drinking water. This document provides the procedures to be used to analyse the quality of water. This involves setting analysis objectives, gathering relevant chemicals data, creating informative visualizations, and extracting valuable insights from the data. Ultimately, the project seeks to provide a comprehensive understanding of side effects of poor water quality and provide effective solution for the same.

Design Thinking:

Monitoring of drinking water quality is an important component of water management, while data analysis is necessary for the identification and characterization of water quality problems. Assessment is the process by which water quality data is transformed into information. The information gained from monitoring is essential for analysing water quality.

1.Analysing Objectives:

The assessment of water quality is done in various ways. A very powerful tool for this purpose is the Water Quality Index (WQI). A water quality index is a means to summarize large amounts of water quality data into simple terms (e.g., good, bad) for reporting to management and the public in a consistent manner. Researchers use different types of indices. The objective of an index is to turn multifaceted water quality data into simple information that is comprehensible and useable by the public.

Physical, physical-chemical and bacteriological analysis: Analyses are performed taking into account requirements (testing methods, sampling frequencies, precision and required accuracy, the maximum allowed value -MAA of parameters) for drinking water .To assess the quality of drinking water physical, physical-chemical and bacteriological parameters are measured .
To assess the quality of drinking water we used the Drinking Water Quality Index (DWQI) developed by the Canadian Council of Ministers of the Environment, which is widely used. The DWQI includes three measures of variance from the selected drinking water quality objectives-scope (F1), frequency (F2) and amplitude (F3) .

Excellent: (DWQI Value 95-100)-Water protected with a virtual absence of impairment; conditions are very close to pristine levels; these index values can only be obtained if all measurements meet recommended guidelines virtually all of the time.

Very good: (DWQI Value 89-94)-Water Quality Is protected with a slight presence of impairment; conditions are close to pristine levels.

Good: (DWQI Value 80-88)-Water Quality Is protected with only a minor degree of impairment; conditions rarely depart from desirable levels.

Fair: (DWQI Value 65-79)-Water Quality Is usually protected but occasionally impaired; conditions sometimes depart from desirable levels.

Marginal: (DWQI Value 45-64)-Water Quality Is frequently impaired; conditions often depart from desirable levels.

Poor: (WQI Value 0-44)-Water Quality Is almost always impaired; conditions usually depart from desirable levels.

2.Data Collection:

We obtain the provided water quality data containing parameters like pH, Hardness, Solids, etc.

3.Visualization strategy:

The water quality index (WQI) model is a popular tool for evaluating surface water quality. WQI models involve four consecutive stages; these are (1) selection of the water quality parameters, (2) generation of sub-indices for each parameter (3) calculation of the parameter weighting values, and (4) aggregation of sub-indices to compute the overall water quality index.

- Selection of the water quality parameters: one or more water quality parameters are selected for inclusion in the assessment
- Generation of the parameter sub-indices: parameter concentrations are converted to unit less sub-indices
- Assignment of the parameter weight values: parameters are assigned weightings depending on their significance to the assessment
- Computation of the water quality index using an aggregation function: the individual parameter sub-indices are combined using the weightings to give a single overall index. A rating scale is usually used to categorise/classify the water quality based on the overall index value.

4.Predictive modeling:

An initial evaluation was carried out on the accessible data in order to filter, normalize, and execute classification algorithms steps to improve water quality in order to find that smallest portion of interest which enables for great level of accuracy at a cheap price. As a result, future identical investigations can avoid costly and time-consuming lab analyses with specific sensors.

On the dataset, a number of supervised prediction (classification and regression) techniques are chosen as examples. In the context of numerical water quality analysis, the entire approach is proposed.

XGB (XGBoost), RF (Random Forest), DTC (Decision Tree), ADA (Adaptive Boosting), and SVC (Support Vector Classifier). From these 5 models, we choose the 2 models to find the best accuracy of water. So those 2 models are XGB (XGBoost) and RF (Random Forest). These above models are deployed which are assured to give maximum accuracy.

Excel and Sigma Plot software's were used for statistical analysis of data and to construct graphs.

Dataset Link: **https://www.kaggle.com/datasets/adityakadiwal/water-potability**

This project focused on analyzing water quality data to assess the suitability of water for specific purposes, such as drinking. This study investigated the machine learning performance of approaches as a result of XGB, RF, SVC, ADA, and Decision Trees in predicting the components of a water quality dataset. For this objective, variables in the most well-known datasets, such as pH, hardness, solids, EC, and turbidity, were acquired.