DATA ANALYSIS WITH COGNOS

Group 2

Project 11 – WATER ANALYSIS

KINGSTON
ENGINEERING COLLEGE

COLLEGE CODE:5113

TEAM 10

TEAM LEADER: INIYAN P – au511321104032

TEAM MATES:

KAMALESH L - au511321104039

DHANUSH M – au511321104021

MUKESH C - au511321104057

KUGAN K - au511321104044

<u>Water Quality Analysis</u>

Phase 1

Project Definition and Design Thinking

Project overview:

Access to clean and safe water is a fundamental human right and a critical component of public health and environmental sustainability. The project "Water Analysis Using

Data Analysis" aims to utilize modern data analysis techniques to assess, monitor, and improve the quality of water resources. This project recognizes the importance of data-driven decision-making in addressing challenges related to potability, environmental impact, and resource management. Through the comprehensive analysis of water quality data, this initiative seeks to contribute to the protection of public health, the preservation of ecosystems, and the responsible use of this invaluable resource.

Objectives:

Regulatory standards:

This project seeks to identify and evaluate potential issues or deviations from regulatory standards governing water quality. It involves the collection and analysis of data related to various water quality parameters, ensuring that water sources meet the requirements set forth by relevant environmental agencies and regulatory bodies.

Water potability:

Another key objective is to determine the potability of water based on a comprehensive analysis of multiple parameters. The project will assess the suitability of water sources for human consumption by comparing measured data against established drinking water standards, such as those outlined by the World Health Organization (WHO) or local environmental agencies.

Definition:

Water is one of the basic resources for human survival. Most traditional water quality analysis systems, however, generally focus only on water quality data collection, data analysis and data mining, ignoring various properties of water such as pH, hardness. In addition, some dirty data and data loss may occur due to power failures or transmission failures, further affecting data analysis and its application.

The project involves analyzing water quality data to assess the suitability of water for specific purposes, such as drinking. The objective is to identify potential issues or deviations from regulatory standards and determine water potability based on various parameters. This project includes defining analysis objectives, collecting water quality data, designing relevant visualizations, and building a predictive model.

Data analytics techniques are applied to the collected data to identify trends, anomalies, and patterns, enabling informed decision-making and the implementation of measures whether the water is potable or not.

Project approach:

Design Thinking:

Design thinking for water analysis involves approach to ideate solutions and basic structures to solve and improving the analysis in a more efficient and user centric way.

Analyzing water quality data is a critical task with a range of specific objectives that help ensure the safety and sustainability of water resources. Design thinking in water analysis has four core components analysis objectives, data collections, visualization strategy, predictive modelling.

Analysis Objectives:

1.Assessing Potability:

The primary objective of water quality analysis is to determine whether the water meets the standards necessary for safe human consumption. This involves assessing parameters such as pH, turbidity, total dissolved solids (TDS), and concentrations of contaminants like bacteria, heavy metals, and organic compounds. The specific objective here is to compare the measured values to established drinking water standards, such as those set by the World Health Organization (WHO) or the Environmental Protection Agency (EPA) in the United States. If any parameter exceeds the maximum allowable concentration, it signifies a potential health risk, and corrective actions must be taken to ensure the water is safe to drink.

2.Identifying Deviations from standards:

Water quality data analysis aims to identify any deviations from established standards or regulatory limits. This involves detecting spikes or unusual trends in parameters that may indicate contamination or deterioration in water quality. For example, sudden increases in fecal coliform counts might suggest a sewage leak, while elevated levels of nitrates may indicate agricultural runoff. Identifying these deviations is essential for early warning and rapid response, allowing authorities to address potential threats to public health and the environment. The specific objective is to pinpoint the source and cause of these deviations and take appropriate corrective actions to restore water quality.

3.Understanding Parameter Relationships:

To manage and protect water resources effectively, it's crucial to understand the complex relationships between various water quality parameters. For instance, the pH level can impact the solubility of metals in water, while temperature fluctuations can affect the growth of aquatic organisms. The objective here is to use statistical and analytical methods to identify correlations and trends among different parameters. This helps in predicting how changes in one parameter may affect others and allows for more informed decision-making. For instance, understanding the relationship between temperature and dissolved oxygen levels can be vital in managing aquatic ecosystems.

Data Collection:

Access to safe drinking-water is essential to health, a basic human right and a component of effective policy for health protection. This is important as a health and development issue at a national, regional and local level. In some regions, it has been shown that investments in water supply and sanitation can yield a net economic benefit, since the reductions in adverse health effects and health care costs outweigh the costs of undertaking the interventions.so we gather water quality data containing parameters.

1. pH value:

PH is an important parameter in evaluating the acid–base balance of water. It is also the indicator of acidic or alkaline condition of water status. WHO has recommended maximum permissible limit of pH from 6.5 to 8.5. The current investigation ranges were 6.52–6.83 which are in the range of WHO standards.

2. Hardness:

Hardness is mainly caused by calcium and magnesium salts. These salts are dissolved from geologic deposits through which water travels. The length of time water is in contact with hardness producing material helps determine how much hardness there is in raw water. Hardness was originally defined as the capacity of water to precipitate soap caused by Calcium and Magnesium.

3. Solids (Total dissolved solids - TDS):

Water has the ability to dissolve a wide range of inorganic and some organic minerals or salts such as potassium, calcium, sodium, bicarbonates, chlorides, magnesium, sulfates etc. These minerals produced un-wanted taste and diluted color in appearance of water. This is the important parameter for the use of water. The water with high TDS value indicates that water is highly mineralized. Desirable limit for TDS is 500 mg/l and maximum limit is 1000 mg/l which prescribed for drinking purpose.

4. Chloramines:

Chlorine and chloramine are the major disinfectants used in public water systems. Chloramines are most commonly formed when ammonia is added to chlorine to treat drinking water. Chlorine levels up to 4 milligrams per liter (mg/L or 4 parts per million (ppm)) are considered safe in drinking water.

5. Sulfate:

Sulfates are naturally occurring substances that are found in minerals, soil, and rocks. They are present in ambient air, groundwater, plants, and food. The principal commercial use of sulfate is in the chemical industry. Sulfate concentration in seawater is about 2,700 milligrams per liter (mg/L). It ranges from 3 to 30 mg/L in most freshwater supplies, although much higher concentrations (1000 mg/L) are found in some geographic locations.

6. Conductivity:

Pure water is not a good conductor of electric current rather’s a good insulator. Increase in ions concentration enhances the electrical conductivity of water. Generally, the amount of dissolved solids in water determines the electrical conductivity. Electrical conductivity (EC) actually measures the ionic process of a solution that enables it to transmit current. According to WHO standards, EC value should not exceeded 400 μS/cm.

7. Organic_carbon:

Total Organic Carbon (TOC) in source waters comes from decaying natural organic matter (NOM) as well as synthetic sources. TOC is a measure of the total amount of carbon in organic compounds in pure water. According to US EPA < 2 mg/L as TOC in treated / drinking water, and < 4 mg/Lit in source water which is use for treatment.

8. Trihalomethanes:

THMs are chemicals which may be found in water treated with chlorine. The concentration of THMs in drinking water varies according to the level of organic material in the water, the amount of chlorine required to treat the water, and the temperature of the water that is being treated. THM levels up to 80 ppm is considered safe in drinking water.

9. Turbidity:

The turbidity of water depends on the quantity of solid matter present in the suspended state. It is a measure of light emitting properties of water and the test is used to indicate the quality of waste discharge with respect to colloidal matter. The mean turbidity value obtained for Wondo Genet Campus (0.98 NTU) is lower than the WHO recommended value of 5.00 NTU.

10. Potability:

Indicates if water is safe for human consumption where 1 means Potable and 0 means Not potable.

The real time water potability data is available in the csv file below:

water_potability.csv

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

Visualization strategies:

Visualizing parameter distributions, correlations, and potability in water quality data is essential for gaining insights and making informed decisions. Here's a plan on how to visualize these aspects using suitable tools:

1.patrameter distibutions:

Visualize the distribution of individual water quality parameters like pH, hardness, solids, and contaminants to understand their variability and identify potential outliers.

Histogram**:** Use histograms to display the frequency distribution of each parameter. This helps in identifying common ranges and potential anomalies.

Box plots: Box plots provide a visual summary of the distribution's central tendency, spread, and outliers, making them useful for comparing parameters.

Density plots: Create density plots to visualize the probability density of continuous parameters, revealing patterns and variations.

2.Correlations:

Explore the relationships between different water quality parameters to identify dependencies and potential causations.

Scatter plots: Create scatter plots to visualize the pairwise relationships between parameters. Scatter plots are particularly useful for identifying linear or non-linear correlations

Correlation Matrices: Use heatmaps to display correlation matrices, showing the strength and direction of correlations between parameters. Tools like Python's Seaborn or R's ggplot2 are excellent for this purpose.

3.Potability Assessment:

Visualize the assessment of water potability based on the water quality parameters and regulatory standards.

Bar charts: Create bar charts to compare measured parameter values against regulatory standards. Parameters exceeding the standards can be highlighted to indicate non-potability.

Pie charts: Use pie charts to represent the percentage of potable and non-potable water samples in your dataset, providing a clear overview.

Map Visualization: If analyzing water sources across geographical locations, consider geographic information system (GIS) tools to create maps. Color-coded markers can indicate potable and non-potable water sources on the map.

Predictive modelling:

logistic regression is a machine learning algorithm primarily used for binary classification tasks, where the goal is to predict one of two possible outcomes based on input features. In your case, you want to predict water potability, which is typically a binary classification problem (potable or non-potable). Here's an explanation of logistic regression and features you can use for predicting water potability: