411521104033-PHASE 1

**WATER QUALITY ANALYSIS**

**DATA ANALYTICS WITH COGNOS-GROUP1**

**Problem statement:**

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.

**Abstract:**

Water quality analysis is a critical component of environmental monitoring and management. It is used to assess the quality of water resources, identify and quantify pollutants, and evaluate the effectiveness of water treatment and pollution control measures. Water quality analysis is also used to support research on aquatic ecosystems and the impacts of human activities on water quality.

A wide range of methods and technologies are used for water quality analysis. The specific methods used will depend on the parameters being measured, the desired level of accuracy, and the resources available. Common methods include:

Physical analysis: This involves measuring physical parameters such as temperature, turbidity, color, and odor.

Chemical analysis: This involves measuring the concentrations of chemical constituents in water. This can be done using a variety of methods, such as titration, gravimetry, and spectroscopy.

**Analysis Objectives:**

When conducting water quality analysis using data analytics, the objectives typically revolve around extracting valuable insights and information from large datasets related to water quality. Here are some common analysis objectives for water quality analysis in data analytics:

Pattern Recognition: Identify recurring patterns, trends, and seasonality in water quality data. This can help in understanding natural fluctuations and identifying abnormal events.

Correlation Analysis: Explore relationships between various water quality parameters, environmental factors (e.g., weather conditions), and geographical features to uncover dependencies and potential drivers of water quality changes.

Quality Assurance/Quality Control (QA/QC): Develop data analytics methods to assess data quality, identify measurement errors, and ensure the reliability of collected data.

Spatial Analysis: Perform spatial analysis to understand how water quality varies across different geographical areas, watersheds, or regions. This can support targeted intervention strategies.

Temporal Analysis: Analyze temporal trends in water quality data to assess the impact of long-term changes, such as climate variability or land use alterations.

The specific objective of a water quality analysis using data analytics should be tailored to the goals of the study, the available data, and the challenges or issues being addressed. Defining clear objectives is essential to guide the analysis process effectively and derive actionable insights for improved water quality management.

**Data Collection:**

We discuss the various data sources used in water quality analysis, including remote sensing, in-situ sensors, and laboratory measurements. The integration of data from these sources forms the basis for comprehensive water quality assessment.

Collecting water quality data is an essential step in water quality analysis. You can obtain this data from various sources, including government agencies, environmental organizations, research institutions, or by conducting your own measurements. Here's a guide on gathering water quality data containing parameters like pH, hardness, and total dissolved solids (TDS).

Gather historical water quality data, including parameters such as pH, turbidity, dissolved oxygen, temperature, and pollutant levels. Ensure that the data is accurate and covers a relevant time period.

Preprocess the data by handling missing values, outliers, and formatting issues. Data quality is crucial for building reliable models.

**DATASET LINK:** [**https://www.kaggle.com/datasets/adityakadiwal/water-potability**](https://www.kaggle.com/datasets/adityakadiwal/water-potability)

**Visualization Stratergy:**

1. Data Preparation:

Clean and prepare your water quality data, ensuring it is in a suitable format for analysis within IBM Cognos.

2. Parameter Distribution Visualization:

Histograms: Create histograms to visualize the distribution of each water quality parameter. IBM Cognos offers built-in capabilities for creating histograms.

Box Plots: Utilize box plots to display the central tendency and spread of parameter values. Cognos can generate box plots as part of its visualization tools.

3. Correlation Visualization:

Scatter Plots: Develop scatter plots to show relationships between pairs of water quality parameters. IBM Cognos allows you to create interactive scatter plots.

4. Portability Visualization:

Responsive Dashboards: Design responsive dashboards in IBM Cognos that adapt to various screen sizes, ensuring portability across devices.

**Predictive Modeling:**

Predictive modeling for water quality analysis involves using historical data to build models that can predict future water quality conditions or identify potential issues. These models can be valuable for early warning systems, decision support, and environmental management. Here are the steps to create a predictive model for water quality analysis:

Model Selection:

Choose an appropriate predictive modeling technique. Common methods for water quality prediction include:

Regression Analysis: Linear regression, multiple regression, or nonlinear regression can be used to predict continuous water quality parameters like temperature or pH.

Time Series Analysis: For parameters with a temporal component, time series models like ARIMA or seasonal decomposition can be effective.

Data Splitting:

Split your dataset into training, validation, and test sets. The training set is used to build the model, the validation set helps tune hyperparameters, and the test set is used to evaluate the model's performance.

Model Training:

Train the selected predictive model on the training dataset. Fine-tune model hyperparameters, if necessary, using the validation dataset.

Model Validation:

Validate the model's performance on the test dataset to ensure it generalizes well to unseen data. If the performance is satisfactory, proceed to deploy the model.

**Conclusion:**

water quality analysis using data analytics represents a critical endeavor with far-reaching implications for public health, environmental sustainability, and effective resource management

As we move forward, the synergy between data analytics and water quality analysis will become increasingly pivotal in addressing the complex and dynamic challenges associated with water quality. With ongoing technological advancements and a commitment to data-driven insights, we are better equipped to protect our vital water resources and ensure a healthier and more sustainable future.