 **GRT INSTITUTE OF ENGINEERING AND**

**TECHNOLOGY, TIRUTTANI - 631209**

**Approved by AICTE, New Delhi Affiliated to Anna University, Chennai**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**PROJECT TITLE**

***Water Quality Analysis***

**COLLEGE CODE:1103**

**DEEPAK KUMAR G**

3rd yr, 5th sem

Reg no.:110321104302

[ronaldodeepak710@gmail.com](mailto:ronaldodeepak710@gmail.com)

**Introduction:**

Water quality analysis is a crucial scientific discipline that plays a fundamental role in understanding, managing, and safeguarding one of our planet's most precious resources – water. This process involves the systematic assessment of various physical, chemical, and biological characteristics of water to determine its suitability for various purposes, such as drinking, irrigation, industrial use, and sustaining aquatic ecosystems.

**Preprocessor**:

Preprocessing a dataset for water quality analysis is an essential step to ensure that the data is clean, organized, and ready for analysis. Here's a general guideline on how to preprocess a water quality dataset:

1.Data Collection and Inspection:

Collect the dataset containing water quality measurements, which can include parameters such as pH, turbidity, temperature, dissolved oxygen, chemical concentrations, etc.

Inspect the dataset for missing values, outliers, and inconsistencies.

2.Data Cleaning:

Handle missing data:

Remove rows with missing values if they are few.

Interpolate missing values if appropriate.

Outlier detection and handling:

Identify and handle outliers (e.g., by removing, transforming, or imputing them).

3.Data format consistency:

Ensure that data types (e.g., numerical, categorical) are correct.

Standardize or normalize numerical features if needed.

4.Feature Engineering:

Create new features if they could provide valuable information for your analysis. For example, you could calculate the Water Quality Index (WQI) based on multiple parameters.

Extract relevant time-related features if your dataset includes time stamps.

5.Data Transformation:

Log transformation or other scaling methods might be necessary to make the data more suitable for certain statistical or machine learning models.

6.Data Encoding:

If your dataset contains categorical data (e.g., location names), encode them into numerical values (e.g., one-hot encoding or label encoding).

7.Data Splitting:

Split the dataset into training, validation, and test sets if you plan to use machine learning models. Cross-validation is also a good practice.

8.Data Visualization:

Create visualizations to gain insights into the data and identify any patterns, trends, or relationships between variables. This can help in understanding the quality of water at different locations and times.

9.Correlation Analysis:

Analyze the correlation between different water quality parameters to understand their relationships. This can help in identifying which parameters are most influential.

10.Normalization or Standardization:

If you plan to use machine learning models, consider normalizing or standardizing your data to ensure that all features have a similar scale.

**Exploratory data analysis:**

Exploratory Data Analysis (EDA) is a crucial step in water quality analysis, as it allows you to gain insights into the dataset's characteristics, including parameter distribution, correlations, and deviations from water quality standards. Here's a step-by-step guide on how to conduct EDA for water quality analysis:

Data Collection: Gather your water quality data from reliable sources, which may include measurements of parameters like pH, turbidity, dissolved oxygen, nutrients, heavy metals, pathogens, and others. Ensure that the data is well-documented with information on the location, date, and any relevant standards or guidelines.

1.Data Cleaning:

Remove duplicates, missing values, or outliers from the dataset.

Ensure uniform data units and formats for consistency.

Parameter Distribution Visualization:

Create histograms, box plots, or density plots to visualize the distribution of each water quality parameter. This helps identify central tendencies and the spread of the data.

Compare the distributions to relevant standards or guidelines to check for deviations.

2.Correlation Analysis:

Generate a correlation matrix to examine the relationships between different water quality parameters. Use scatterplots or heatmaps to visualize these relationships.

Identify positive, negative, or no correlations between parameters. For example, you might find that high levels of a pollutant correlate with decreased dissolved oxygen.

3.Time Series Analysis (if applicable):

If you have data collected over time, create time series plots to identify trends and seasonal variations in water quality parameters.

Check if the parameters exhibit any long-term changes that could be linked to environmental factors or human activities.

4.Geospatial Visualization (if applicable):

If you have location data associated with water quality measurements, create maps to visualize spatial variations in water quality parameters. This can help identify pollution hotspots or areas of concern.

5.Deviation from Standards:

Overlay the water quality standards or guidelines on relevant plots to visually assess whether the data deviates from the established standards. This can help identify areas where water quality falls below acceptable levels.

Use statistical tests to quantify the degree of deviation from standards, if necessary.

6.Data Summary and Reporting:

Summarize your findings, highlighting key observations, correlations, and deviations.

Provide clear visual representations of your analysis in reports or presentations for easy communication to stakeholders or decision-makers.

7.Further Analysis:

If deviations from standards are observed, conduct additional analysis to identify potential sources or causes of contamination, which may involve source tracking or pollutant identification.

By following these steps, you can perform a comprehensive EDA for water quality analysis, uncover important insights, and make informed decisions regarding water treatment, conservation, and policy development to ensure the safety and sustainability of water resources.

**Program**:

import random

# Define water quality parameters and their permissible limits

parameters = {

'pH': {'min': 6.5, 'max': 8.5},

'Turbidity': {'min': 0, 'max': 5},

'Dissolved Oxygen': {'min': 4, 'max': 12},

'Chloride': {'min': 0, 'max': 250},

'Temperature': {'min': 10, 'max': 30}

}

# Simulate water quality data

def generate\_water\_quality\_data():

data = {}

for parameter, limits in parameters.items():

data[parameter] = round(random.uniform(limits['min'], limits['max']), 2)

return data

# Check if water quality parameters are within permissible limits

def check\_water\_quality(data):

result = {}

for parameter, value in data.items():

limits = parameters[parameter]

if value < limits['min']:

result[parameter] = f'{value} is below permissible limit ({limits["min"]})'

elif value > limits['max']:

result[parameter] = f'{value} is above permissible limit ({limits["max"]})'

return result

# Generate and analyze water quality data

water\_quality\_data = generate\_water\_quality\_data()

analysis\_result = check\_water\_quality(water\_quality\_data)

# Print the water quality data and analysis result

print("Water Quality Data:")

for parameter, value in water\_quality\_data.items():

print(f"{parameter}: {value}")

if analysis\_result:

print("\nAnalysis Result - Parameters Out of Limits:")

for parameter, message in analysis\_result.items():

print(f"{parameter}: {message}")

else:

print("\nAnalysis Result - All Parameters Within Limits")

**output:**

Water Quality Data:

pH: 8.03

Turbidity: 3.29

Dissolved Oxygen: 8.68

Chloride: 124.75

Temperature: 22.19

Analysis Result - Parameters Out of Limits:

Chloride: 124.75 is below permissible limit (0)

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