 **GRT INSTITUTE OF ENGINEERING AND**

**TECHNOLOGY, TIRUTTANI - 631209**

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

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**WATER QUALITY ANALYSIS**

**PROJECT REPORT**

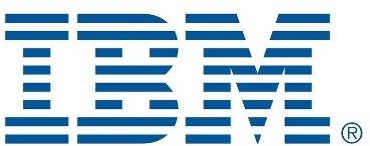
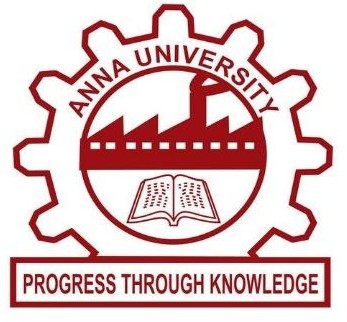
**SUBMITTED BY**

G.DEEPAK KUMAR

3RD YEAR 5TH SEM

110321104302

ronaldodeepak710@gmail.com



**ANNA UNIVERSITY:CHENNAI 600 025**

# BONAFIDE CERTIFICATE

Certified that this project report **“WATER QUALITY ANALYSIS”**

is the bonafide work of **“DEEPAK KUMAR.G [110321104302]”** who carried out the project work under my our supervision.

## **SIGNATURE SIGNATURE**

### **Dr.N. Kamal M.E.,Ph.D**., **Mr.T.A. Vinayagam M.Tech.,**

**HOD Assistant professor**

Department of Computer Science And Department of Computer Science And

Engineering Engineering

GRT Institute of Engineering and GRT Institute of Engineering and

Technology Technology

Tiruttani Tiruttani

Certified that the candidates were examined in Viva-voce in the Examination Held on

**INTERNAL EXAMINER EXTERNAL EXAMINER**

### **ACKNOWLEDGEMENT**

We thank our Management for providing us all support to complete this project successfully.

Our sincere thanks to honorable **Chairman**, **Shri. G. RAJENDRAN and Managing Director, Shri.G.R. RADHAKRISHNAN** for creating a wonderful atmosphere inside the campus.

We are very grateful to **Dr.S. ARUMUGAM, M.E., Ph.D., Principal**, for providing us with consistent guidance and motivation to execute a real time project to learn and experience the project work in an environments to complete our project successfully.

Our sincere thanks to **Dr.N. KAMAL, M.E., Ph.D., Professor and Head, Department of Computer Science and Engineering** for giving me this wonderful opportunity to do the project and providing the require facilities to fulfill our work.

We are highly indebted and thankful to our project Evaluators **Mrs V. Priya M.E**., and **Mrs. Edith Esther M.E., Assistant Professor, Department of Computer Science and Engineering** for his immense support in doing the project.

We are very grateful to our internal guide **Mr. T.A. VINAYAGAM, M.Tech., Assistant Professor, Department of Computer Science and Engineering** for guiding us with her valuable suggestions to complete our project.

We also dedicate equal and grateful acknowledgement to all the **respectable members of the faculty and lab in-charges** of the Department of Computer Science and Engineering, friends and our families for their motivation, encouragement and continuous support.

Our sincere thanks **to IBM and Skill Up Team members** for giving me this wonderful opportunity to do the project and providing the require guidance and valuable online sessions.

|  |  |  |
| --- | --- | --- |
|  | **TABLE OF CONTENTS** |  |
| **CHAPTER**  **No** | **TITLE** | **PAGE**  **No** |
|  | **ABSTRACT** |  |
| **1.** | **PHASE 1** |
|  | 1.0 INTRODUCTION |
|  | 1.1 PROBLEM DEFINITION |
|  | 1.2 DESIGN THINKING |
|  | 1.3 OBJECTIVES |
|  | 1.4 SYSTEM DESIGN AND THINKING |
|  | 1.5 SYSTEM ARCHITECTURE |
|  | 1.6 E – R DIAGRAM |
|  | 1.7 USE CASE DIAGRAM |
|  | 1.8 ARCHITECTURE |
|  | 1.9 SEQUENCE DIAGRAM |
| **2.** | **PHASE 2** |
|  | 2.1 SHORT EXPLAINATION ABOUT |
|  | WATER QUALITY ANALYSIS |
|  |  |

|  |  |  |
| --- | --- | --- |
| **3.** | * 1. WHERE I GOT THE DATASETS AND ITS DETAILS   2. DETAILS ABOUT COLUMNS   3. DETAILS OF LIBRARIES TO BE USED AND WAY TO DOWNLOAD   4. HOW TO TRAIN AND TEST THE DATASET   5. REST OF EXPLAINATION   6. WHAT METRICS USED FOR THE ACCURACY CHECK   **PHASE 3**   * 1. DATASET AND ITS DETAIL EXPLANATION AND IMPLEMENTATION OF   WATER QUALITY ANALYSIS   * 1. BEGIN BUILDING THE PROJECT BY LOAD THE DATASET CUSTOMERIDS   2. PREPROCESS DATASET   3. PERFORMING DIFFERENT ANALYSIS NEEDED |  |

|  |  |  |
| --- | --- | --- |
| **4.** | **PHASE 4**   * 1. : IN THIS TECHNOLOGY YOU WILL CONTINUE BUILDING YOUR PROJECT BY PREPROCESSING YOUR DATASET   2. : IN THIS TECHNOLOGY YOU WILL CONTINUE BUILDING YOUR PROJECT BY PERFORMING FEATURE ENGINEERING 4.3:MODEL TRAINING AND EVALUATION   4.4: PERFORM DIFFERENT ANALYSIS AS NEEDED |  |

**PHASE:1**

An Introduction to Water Quality Analysis

Abstract:

Water is perhaps the most precious natural resource after air. Though the surface of the earth is mostly consists of water, only a small part of it is usable, which makes this resource very limited. This precious and limited resource, therefore, must be used with prudence. As water is required for different purposes, the suitability of it must be checked before use. Also, sources of water must be monitored regularly to determine whether they are in sound health or not. Poor condition of water bodies are not only the indictor of environmental degradation, it is also a threat to the ecosystem. In industries, improper quality of water may cause hazards and severe economic loss. Thus, the quality of water is very important in both environmental and economic aspects. Thus, water quality analysis is essential for using it in any purpose. After years of research, water quality analysis is now consists of some standard protocols. There are guidelines for sampling, preservation and analysis of the samples. Here the standard chain of action is discussed briefly so that it may be useful to the analysts and researchers

**1. INTRODUCTION**

**1.1 What is Water Quality?**

Water Quality can be defined as the chemical, physical and biological characteristics of water, usually in respect to its suitability for a designated use.Water can be used for recreation, drinking, fisheries, agriculture or industry. Each of these designated uses has different defined chemical, physical and biological standards necessary to fulfil the respective purpose. For example, there are stringent standards for water to be used for drinking or swimming compared to that used in agriculture or industry.

**1.2 What is Water Quality Analysis?**

After many years of research, water quality standards are put in place to ensure the suitability of efficient use of water for a designated purpose. Water quality analysis is to measure the required parameters of water, following standard methods, to check whether they are in accordance with the standard.

**1.3 Why Water Quality Analysis is required?**

Water quality analysis is required mainly for monitoring purpose. Some importance of such assessment includes:

(i) To check whether the water quality is in compliance with the standards, and hence, suitable or not for the designated use.

(ii) To monitor the efficiency of a system, working for water quality maintenance

(iii) To check whether upgradation / change of an existing system is required and to decide what changes should take place

(iv) To monitor whether water quality is in compliance with rules and regulations.Water quality analysis is of extremely necessary in the

sectors of:

· Public Health (especially for drinking water)

· Industrial Use

**2. PROCEDURES OF WATER QUALITY ANALYSIS**

The steps for water quality analysis in general is mentioned

**2.1 Selection of Parameters**

The parameters of water quality are selected entirely according to the need for a specific use of that water. Some

examples are:

Drinking: As per WHO/CPCB Standards

**Irrigation:**

pHConductivity

Sodium & Potassium

Nutrients

Specific compounds

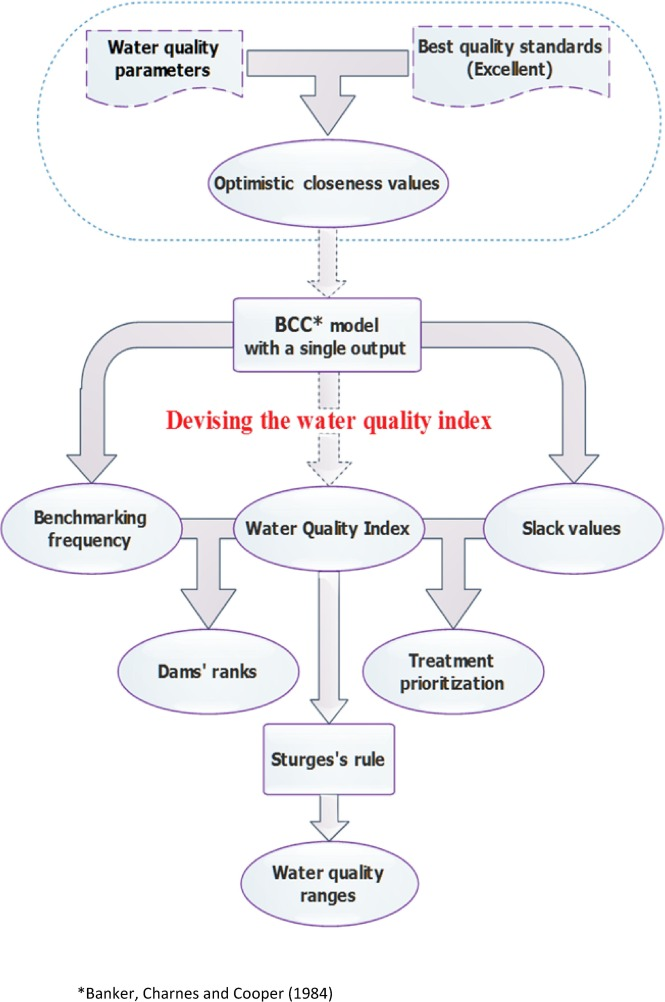
Industries: As per specific requirement

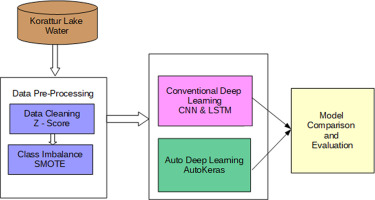
Domestic Consumption: As per BIS Standards

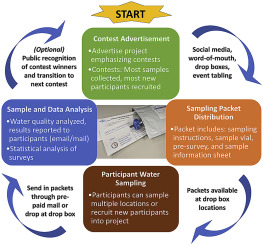
**Water Bodies**:

As per CPCB guidel









**PHASE:2**

**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:

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.

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).

Data format consistency:

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

Standardize or normalize numerical features if needed.

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.

Data Transformation:

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

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).

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.

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.

Correlation Analysis:

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

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.

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.

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.

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.

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.

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.

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.

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.

**PHASE:3**

**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

**PHASE:4**

**Histograms**

Scatterplots are a fundamental graph type—much less complicated than histograms and boxplots. As such, we might use the Mathplotlib library instead of the Seaborn library. But since we have already used Seaborn, I will stick with it here. Just know that there are many ways to create scatterplots and other basic graphs in Python.

To create a bare-bones scatterplot, we must do four things:

1.Load the seaborn library

2.Specify the source data frame

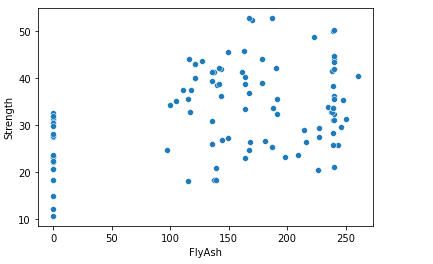
3.Set the x axis, which is generally the name of a predictor/independent variable

4.Set the y axis, which is generally the name of a response/dependent variable

import seaborn as sns

sns.scatterplot(x="FlyAsh", y="Strength", data=con);

\_images/08\_correlation\_11\_0.png

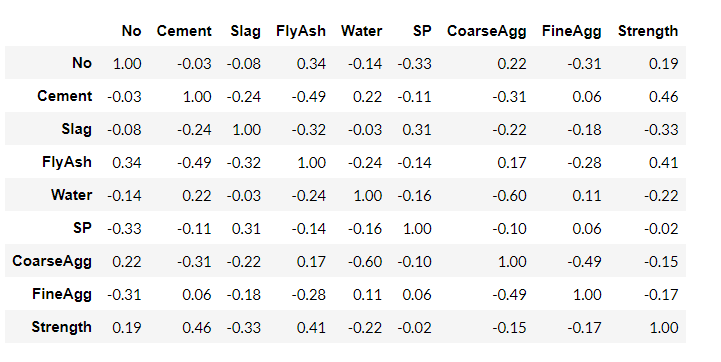


**Corrleation matrix**

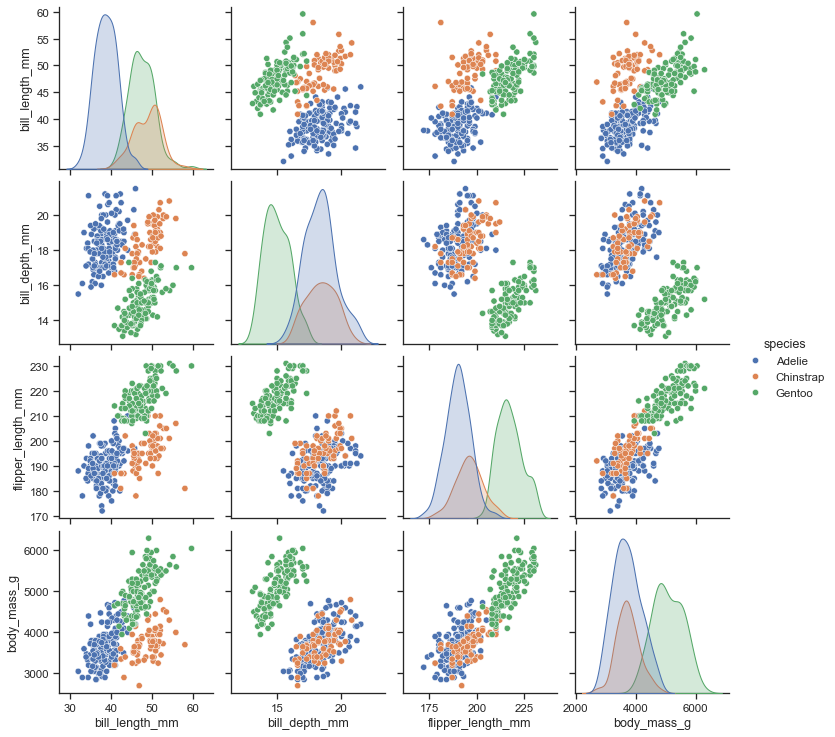
A correlation matrix is a handy way to calculate the pairwise correlation coefficients between two or more (numeric) variables. The Pandas data frame has this functionality built-in to its corr() method, which I have wrapped inside the round() method to keep things tidy. Notice that every correlation matrix is symmetrical: the correlation of “Cement” with “Slag” is the same as the correlation of “Slag” with “Cement” (-0.24). Thus, the top (or bottom, depending on your preferences) of every correlation matrix is redundant. The correlation between each variable and itself is 1.0, hence the diagonal.

cormat = con.corr()

round(cormat,2)



**Scatterplot Matrix**



import seaborn as sns

sns.set\_theme(style="ticks")

df = sns.load\_dataset("penguins")

sns.pairplot(df, hue="species"

**Random Forest**

For classification and regression problems, many people use supervised machine learning, and Random Forest is one of them. Breiman (2001) proposed the Random Forest algorithm, which was extremely successful as a general-purpose classification and regression technique. The approach, which shuffles numerous randomized selection trees and aggregates their predictions by averaging, has shown an excellent overall performance in settings where the set of variables is much larger than the number of observations. In addition, it is flexible enough to be applied to large-scale problems, easily adaptable to various ad hoc study tasks, and returns

measures of different meanings.

The Random Forest regression algorithm was chosen for the following two key reasons:

Multivariate regression analysis: The target parameter can be dependent on multiple attributes/parameters. This type of many-to-one relationship requires multivariate regression analysis instead of the usual one-to-one linear regression analysis.

Relatively small dataset: As the total number of samples in the used dataset is less than 5,000, it is considered to be a small dataset. Small datasets are difficult to analyse as sufficient samples are required to train a model as well as for the model testing and validation process

****

**Software used for Random Forest**

Python was used for model building and prediction analysis. From the scikit-learn package, Random Forest Regressor algorithm was imported from the ensemble methods available. The dataset was split into train and test samples in a 7:3 ratio using the train\_test\_split method from the sklearn package. For visualization, matplotlib and seaborn packages are used. Pandas package was used in the formatting of the dataset, and pre-processing methods.

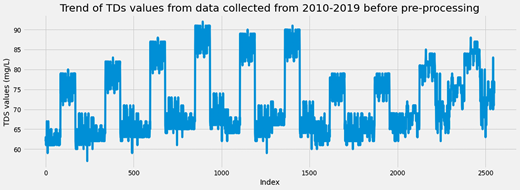
Model validation

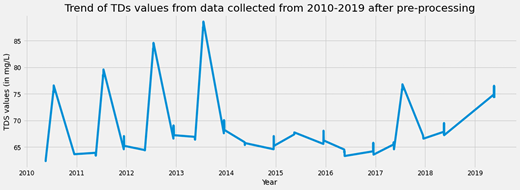
The consequences derived from the version were assessed using several statistical tests. R2 is used to assess the relationship between located values and expected values. The equation for the calculation is as follows:

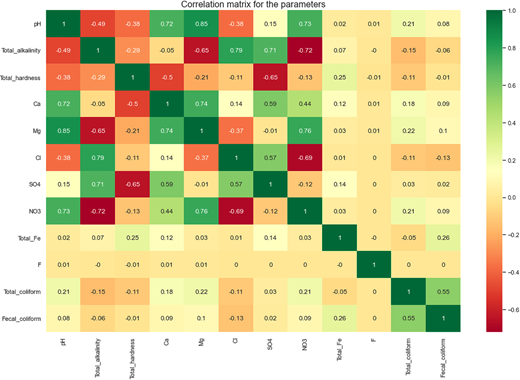
formula

formula

formula







Model accuracy

Here, we have taken two key parameters which are used to measure whether the model can predict the target parameter with high accuracy.

R2 (coefficient of determination): It is a statistical measure for how well the regression line is able to approximate the actual data.

formulawhere yi is the actual ith sample; yi′ is the predicted ith sample; Y is the mean of the target parameter; R2 has a range from 0 to 1. Higher the R2 for a model, better the model can predict for the target parameter.

RMSE (root mean squared error): It gives the standard deviation of the prediction errors(residuals). It measures how spread out the errors are from the main concentration of actual data points.

formula

where N is the total number of samples.

Lower the RMSE for a model, the prediction is more precise with less residuals.

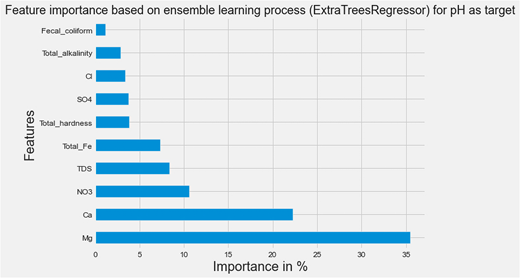
Dataset

The dataset used in this paper is made of parameters that are considered to be vital for a healthy water ecosystem, such as, total iron content, pH, sulfate, and nitrate levels. The breakdown of organic matter is measured in terms of TDS, the total amount of coliforms, and fecal coliform contents. Fluoride content, hardness, and alkalinity of the water are considered for the ergonomic use of the river water.

The data are collected from four different locations on the Kulik River bed. The samples for the parameters are in numerical representation except for the presence of E. coli bacteria.

Random Forest regression algorithm

Random Forest is a type of ensemble learning algorithm as it uses multiple decision trees to estimate a prediction with high accuracy. When multiple attributes are heavily correlated to the target parameter, a decision tree selects the parameter with the highest correlation with the target. From there, it starts the prediction process with a sequence of comparisons with other parameters based on pre-learned threshold values. Starting from the top (parameter with the highest correlation with the target), it works its way to the lowest level nodes (with the least correlation with the target), resulting in a leaf (decision/prediction) at the end of the tree. The comparison is done using MSE (mean squared error) to determine how the data branches from each node. It is given by the MSE equation



**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

**REFERENCES:**

* Brikké F (2000) Operation and maintenance of rural water supply and sanitation systems: A training package for managers and planners. Delft, IRC International Water and Sanitation Centre; and Geneva, World Health Organization ([https://www​.who.int/publications​/i/item/WHO-SDE-WSH-00.2](https://www.who.int/publications/i/item/WHO-SDE-WSH-00.2)).
* Sawyer R, Simpson-Hébert M, Wood S (1998) PHAST step-by-step guide: A participatory approach for the control of diarrhoeal disease. Geneva, World Health Organization (WHO/EOS/98.3; [https://www​.who.int/publications​/i/item/WHO-EOS-98.3](https://www.who.int/publications/i/item/WHO-EOS-98.3)).
* Simpson-Hébert M, Sawyer R, Clarke L (1996) The PHAST initiative: Participatory Hygiene and Sanitation Transformation—A new approach to working with communities. Geneva, World Health Organization, United Nations Development Programme/World Bank Water and Sanitation Program (WHO/EOS/96.11; [http://apps​.who.int/iris​/handle/10665/63260](http://apps.who.int/iris/handle/10665/63260)).
* WHO (1976) Surveillance of drinking-water quality. Geneva, World Health Organization ([https://www​.who.int/publications​/i/item/9241400633](https://www.who.int/publications/i/item/9241400633)). [[PubMed](https://pubmed.ncbi.nlm.nih.gov/826021)]
* WHO (1997) Guidelines for drinking-water quality, 2nd ed. Vol. 3. Surveillance and control of community supplies. Geneva, World Health Organization ([https://www​.who.int/publications​/i/item/9241545038](https://www.who.int/publications/i/item/9241545038)).
* WHO (2006) Guidelines for the safe use of wastewater, excreta and greywater in agriculture and aquaculture, Vol. 1–4. Geneva, World Health Organization ([https://www​.who.int/teams​/environment-climate-change-and-health​/water-sanitation-and-health​/sanitation-safety​/guidelines-for-safe-use-of-wastewater-greywater-and-excreta](https://www.who.int/teams/environment-climate-change-and-health/water-sanitation-and-health/sanitation-safety/guidelines-for-safe-use-of-wastewater-greywater-and-excreta)).
* WHO (2018) Guidelines on sanitation and health. Geneva, World Health Organization ([https://www​.who.int/publications​/i/item/9789241514705](https://www.who.int/publications/i/item/9789241514705)).
* WHO (2021) Guidelines on recreational water quality. Vol. 1: Coastal and fresh waters. Geneva, World Health Organization ([https://www​.who.int/publications​/i/item/9789240031302](https://www.who.int/publications/i/item/9789240031302)).
* UNGA (United Nations General Assembly) (2010a) Resolution 64/292: The human right to water and sanitation. In: 108th Plenary Meeting, 28 July 2010.
* UNGA (United Nations General Assembly) (2010b) Human Rights Council Resolution 15/9: Human rights and access to safe drinking water and sanitation.
* Howard G et al. (2002) Healthy villages: A guide for communities and community health workers. Geneva, World Health Organization ([https://www​.who.int/publications​/i/item/9241545534](https://www.who.int/publications/i/item/9241545534)).
* Prüss-Üstün A et al. (2016) Preventing disease through healthy environments: A global assessment of the burden of disease from environmental risks. Geneva, World Health Organization ([https://www​.who.int/publications​/i/item/9789241565196](https://www.who.int/publications/i/item/9789241565196)).
* WHO (2019) Safer water, better health. Geneva, World Health Organization ([https://www​.who.int/publications​/i/item/9789241516891](https://www.who.int/publications/i/item/9789241516891)).