College code: 4212
Register number: 421221243019

WATER QUALITY ANALYSIS
PHASE2-DATA ANALYTICS WITH COGNOS: GROUP2

INTRODUCTION:

Water is a fundamental resource essential for sustaining life, supporting ecosystems, and facilitating various human activities. The quality of water, determined by its chemical, physical, and biological characteristics, plays a pivotal role in its suitability for different purposes. Water quality analysis is the systematic assessment of these attributes to ensure the safety, health, and sustainability of water resources.

In this phase we are going to explain about design and ideology that are going to present to solve this problem.

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

To this problem this dataset is given to us so by using this dataset we are going to solve our problem.

In the phase1 we have defined certain steps to solve the problem step by step now we are going to explain which methodology we are going to use to solve this problem in each step.

CLEARLY DEFINE THE PROBLEM:

We have clearly understood the problem that we are going to understand about ensuring the safety and sustainability of water resources

Data collection:

The dataset is already given for us:

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

Preparing of the data:

First we have to understand what was the data we are going to analyse for this we have to clean and process the data by using suitable techniques like dropping the null values, data types, remove the duplicate values, visualize the missing values drop the duplicates, by using the suitable functions like drop, is null etc.…

Exploratory data analyses:

This was the most important step in this project so we have to represent our data in the understandable visualization tools like pie chart, bar graph, histogram to represent the relation and variation .

Model training and validation:

We are using various model validation methods like Statistical Analysis, Time Series Analysis,Regression Analysis,Classification Models,Clustering Analysis,Neural Networks,Spatial Analysis,Ensemble Models,Anomaly Detection,Natural Language Processing (NLP),Dimensionality Reduction,Feature Engineering.,Time-Series Forecasting Models.

College code: 4212
Register number: 421221243019

Model evaluation:

We are divide our data set by using the training and testing, predicting the values by using the 2 dimensional planes as a result the accuracy model will tells us which model do we use.

Result representation:

According to this project can be represented in many ways like using the ROC curve, confusion matrix etc… we are going to use the confusion matrix to solve this.

Reporting and visualization:

To inform stakeholders about the water quality and insights provide the periodical reports and dashboards we are going to do this in a confusion matrix and accuracy of the data.