**PHASE 2 PROJECT SUBMISSION**

## WATER QUALITY ANAYSIS

**PROJECT DEFINITION:**

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.

**PROJECT APPROACH:**

Water quality analysis using data analytics typically involves a combination of data preprocessing, feature engineering, and the application of machine learning algorithms. Below is a simplified algorithmic outline for water quality analysis:

**1. Data Collection:**

- Gather water quality data from various sources, such as sensors, monitoring stations, or laboratory tests. Ensure that the data includes relevant parameters (e.g., pH, temperature, dissolved oxygen, contaminants).

**2. Data Preprocessing:**

- Clean the data by handling missing values and outliers.

- Normalize or scale the data as needed.

- Perform any necessary data transformations.

**3. Feature Engineering:**

- Select relevant features (water quality parameters) that will be used for analysis.

- Create new features if domain knowledge suggests they might be informative.

**4. Data Splitting:**

- Split the dataset into training and testing sets to evaluate model performance.

**5. Algorithm Selection:**

- Choose appropriate algorithms based on the nature of the problem:

- For regression tasks (e.g., predicting contaminant levels): Use algorithms like Linear Regression, Random Forest Regression, or Gradient Boosting Regression.

- For classification tasks (e.g., classifying water quality as safe or polluted): Use algorithms like Logistic Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), or Neural Networks.

- For time-series data: Consider Time Series Forecasting methods or LSTM-based neural networks.

**6. Model Training:**

- Train the selected algorithm(s) on the training dataset.

- Tune hyperparameters using techniques like grid search or randomized search to optimize model performance.

**7. Model Evaluation:**

- Evaluate the model's performance on the testing dataset using appropriate evaluation metrics:

- For regression: Use metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE).

- For classification: Use metrics like Accuracy, Precision, Recall, F1-Score, and ROC-AUC.

**8. Model Interpretation:**

- Interpret the model to understand which features have the most significant impact on water quality predictions/classifications. Techniques like feature importance analysis can help.

**9. Visualization:**

- Visualize the model's predictions or classification results. You can create plots, graphs, or maps to make the results more accessible and understandable.

**10. Continuous Monitoring:**

- Implement a system for real-time or periodic data collection to monitor water quality over time.

- Update the model as new data becomes available to maintain its accuracy.

**11. Decision Support**:

- Use the model's insights to inform decision-making related to water resource management, pollution control, or treatment strategies.

**12. Reporting and Communication:**

- Create comprehensive reports that communicate the findings, model performance, and recommendations to stakeholders.

**FLOWCHART:**

