**PHASE 2 PROJECT SUBMISSION**

## WATER QUALITY ANAYSIS

**TEAM MEMBERS:**

1. Aazam Khan - 2021504002
2. Sasmitha A – 2021504038
3. Swetha K – 2021504051
4. Gokul R – 2021504515

**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.

**Dataset Link:**[**https://www.kaggle.com/datasets/adityakadiwal/=water-potability**](https://www.kaggle.com/datasets/adityakadiwal/=water-potability)

**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 (SVM MODEL):**

Support Vector Machine (SVM) is a powerful machine learning algorithm that can be used for water quality analysis, especially for classification tasks where you want to categorize water quality as safe or polluted

**SVM Model Selection**:

Choose the appropriate SVM variant based on your problem:

For linearly separable data (where you can draw a straight line to separate classes): Use a linear SVM.

For non-linearly separable data (where classes are not easily separable by a straight line): Use a kernel SVM, such as the Radial Basis Function (RBF) kernel.

**6. Model Training:**

Train the SVM model on the training dataset using the selected kernel and hyperparameters.

Hyperparameter tuning can be done using techniques like grid search or randomized search to optimize model performance.

**7. Model Evaluation:**

Evaluate the SVM model's performance on the testing dataset using classification metrics such as:

Accuracy: The percentage of correctly classified instances.

Precision: The ratio of true positives to the total predicted positives.

Recall: The ratio of true positives to the total actual positives.

F1-Score: The harmonic mean of precision and recall.

Confusion matrix: Provides a detailed breakdown of true positives, true negatives, false positives, and false negatives.

**8. Model Interpretation:**

Understand which features are most influential in making classification decisions. Some SVM implementations provide feature importance scores.

**9. Visualization:**

Visualize the SVM decision boundary, especially in 2D or 3D space, to see how it separates the classes.

**10. Deployment and monitoring:**

Deploy the trained SVM model in your water quality monitoring system or application for real-time or periodic analysis.

Continuously monitor water quality and update the model as new data becomes available to maintain its accuracy.

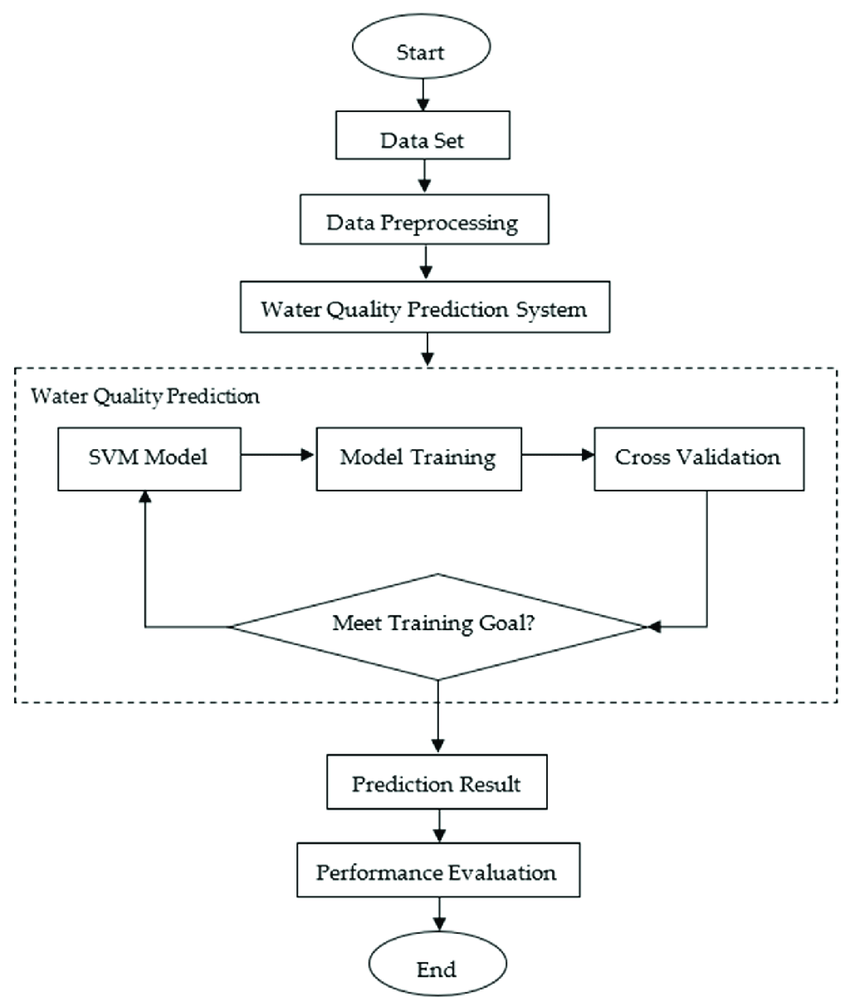
**11. Decision Support**:

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:**



**CONCLUSION:**

In conclusion, SVM models are a valuable tool for water quality analysis, offering accurate classification capabilities, flexibility for complex data relationships, and robustness to outliers. When used appropriately and complemented by data preprocessing, feature selection, and careful hyperparameter tuning, SVMs can contribute significantly to the assessment and management of water quality in various environmental contexts.