PHASE 2:

**INNOVATION**

**Water Quality Anomaly Detection:**

Anomaly detection techniques are valuable for identifying unusual or abnormal data points in water quality parameters, which can help detect issues or contaminants in water sources. Here are some common anomaly detection techniques you can explore:

**Statistical Methods:**

**Z-Score:** Calculate the z-score for each parameter and flag data points with z-scores beyond a certain threshold as anomalies.

**Modified Z-Score:** Similar to the z-score, but robust to outliers.Percentile-Based Methods: Identify anomalies based on percentiles, such as the 95th or 99th percentile.

**Machine Learning Algorithms:**

**Isolation Forest:** It’s effective for isolating anomalies by building a random forest of decision trees.

One-Class SVM: Suitable for situations with limited anomalous data.

Autoencoders: Deep learning models that can capture complex patterns and identify anomalies.

**Time-Series Analysis:**

Exponential Smoothing: Detect anomalies by comparing observed values with exponentially weighted moving averages.

Seasonal Decomposition: Decompose time series data into trend, seasonal, and residual components to identify anomalies in the residuals.

**Clustering Methods:**

K-Means Clustering: Detect anomalies as data points that don’t belong to any cluster or belong to a small cluster.

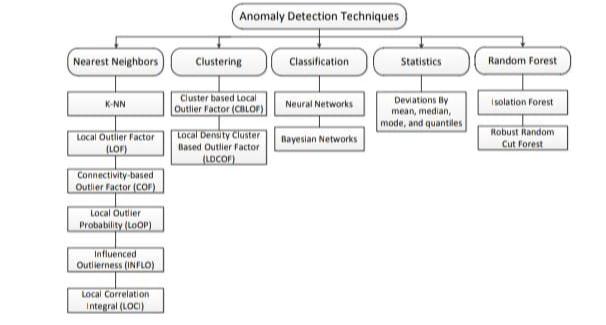
DBSCAN: Identifies anomalies as points that are not part of any dense cluster.

**Visualization Techniques:**

Box Plots: Visualize data distribution and identify outliers.Time-Series Plots: Plotting the in time series data can help identify sudden deviations from the norm.

**Domain-Specific Approaches:**

Consult experts in water quality management for domain-specific knowledge that can guide anomaly detection.Ensemble Methods: Combining multiple anomaly detection techniques can enhance accuracy and reduce false positives.Remember to preprocess and normalize your data appropriately before applying these techniques, and fine-tune parameters based on the specific characteristics of your water quality dataset. Additionally, regularly update and retrain your models to adapt to changing waterquality.

conditions.******

**Python program:**

**import numpy as np**

**import matplotlib.pyplot as plt**

**from sklearn.ensemble import IsolationForest**

**np.random.seed(42)**

**data = np.random.randn(100, 2)**

**model = IsolationForest(contamination=0.1)**

**model.fit(data)**

**anomalies = model.predict(data)**

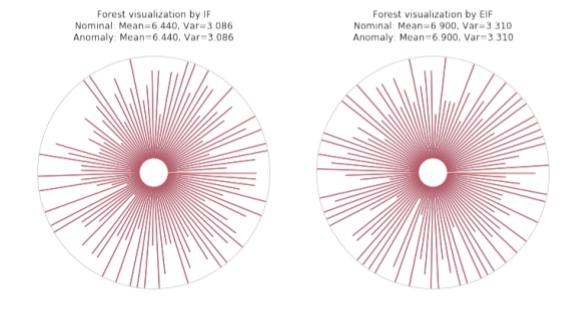
**plt.scatter(data[:, 0], data[:, 1], c=anomalies, cmap='viridis')**

**plt.title('Anomaly Detection with Isolation Forest')**

**plt.xlabel('Feature 1')**

**plt.ylabel('Feature 2')**

**plt.show()**

****

