Name of the Project : Water quality parameters.

Problem Statement :

Ensuring safe and sustainable water resources is a critical global challenge. Access to clean and potable water is essential for human health, agriculture, industry, and the environment. The quality of water is influenced by various natural and anthropogenic factors, making it imperative to monitor and manage key water quality parameters. However, there are several pressing issues and challenges related to water quality parameters that need to be addressed:

Problem Definition :

Anomaly detection techniques are valuable for identifying unusual patterns or outliers in water quality parameters, which can help detect contamination events, equipment failures, or other issues in water monitoring systems. Here are some commonly used anomaly detection techniques for water quality parameter analysis:

Statistical Methods:

Z-Score or Standard Score: Calculate the z-score for each parameter and flag data points with z-scores exceeding a threshold as anomalies.

Percentile-Based Methods: Identify anomalies based on percentiles, such as the 95th percentile. Data points above this threshold may be considered anomalies.

Machine Learning Techniques:

1) **Isolation:** This algorithm isolates anomalies by building a tree structure and identifying data points that require fewer splits to be isolated.

2) **One-Class SVM:** One-Class Support Vector Machines can be used fornovelty detection, where the model learns the normal pattern and flags data points that deviate from it.

3) **Autoencoders**:   Use deep learning autoencoders to reconstruct data andidentify data points with high reconstruction errors as anomalies.

Time-Series Analysis:

1) **Exponential Smoothing:** Apply exponential smoothing techniques to detect anomalies by comparing actual values with smoothed predictions.
2) **Seasonal Decomposition**: Decompose time series data into seasonal, trend, and residual components, and identify anomalies in the residual component.

Clustering Techniques:

DBSCAN (Density-Based Spatial Clustering of Applications with Noise):
Cluster data points and consider outliers as anomalies.

Means Clustering:
Detect Anomalies as Data points that don't belong to any cluster or belong to small cluster.

Principal Component Analysis (PCA):
Use PCA to reduce dimensionality while preserving most of the variance. Anomalies can be detected by analyzing the residual error between the originaldata and the reconstructed data.

Time-Series Decomposition:
Decompose time series data into components such as trend, seasonality, andnoise. Anomalies can be detected in the noise component.

Rule-Based Methods:
Define rules based on domain knowledge and flag data points that violate theserules as anomalies. For example, if pH values fall outside a specific range, it could be considered an anomaly.

Deep Learning for Sequences:

If you have sequential data, recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks can be trained to detect anomalies in the sequence.

Ensemble Methods:
Combine multiple anomaly detection techniques to improve overall detectionaccuracy.

Conclusion :

When applying these techniques to water quality parameter data, it's essential to preprocess the data, set appropriate thresholds or hyperparameters, and validate the results using labeled data if available. Additionally, consider the specific characteristics of your water quality data and the context of your monitoring system to choose the most suitable technique(s) for your application. Regular monitoring and updating of anomaly detection models are also important to ensure their continued effectiveness.