**Water Quality Analysis**

**Phase 2: Innovation**

In this section you need to put your design into innovation to solve the problem. Create a document around it and share the same for assessment as per the instructions mentioned.

Consider exploring anomaly detection techniques to identify unusual patterns in water quality parameters.

**Dataset Link:**

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

**Introduction**

* Water quality analysis is crucial for ensuring the safety and health of communities. Regular monitoring of water quality parameters such as pH, turbidity, dissolved oxygen, and contaminants is essential to detect any anomalies or deviations from the normal range. Anomalies in water quality can indicate contamination, equipment malfunction, or environmental changes that may impact the quality of water.
* This document outlines an innovative approach to water quality analysis by incorporating anomaly detection techniques using IBM Cognos Analytics. By integrating data analytics with anomaly detection, we can enhance our ability to identify unusual patterns in water quality parameters efficiently. This approach aims to improve the early detection of water quality issues, allowing for prompt corrective actions and ultimately ensuring the delivery of safe and clean water to the community.

**Problem Statement**

* The traditional approach to water quality analysis often relies on periodic manual sampling and laboratory testing. This method is not only time-consuming but also lacks real-time monitoring capabilities. In addition, it may not promptly detect anomalies or deviations in water quality parameters, which can have serious consequences for public health.

**Proposed Solution**

1. Data Collection and Integration

To implement anomaly detection for water quality analysis, we need to first collect and integrate relevant data sources:

* Sensor Data: Install water quality sensors at key points in the water supply system to continuously monitor parameters such as pH, turbidity, dissolved oxygen, and temperature.
* Historical Data: Gather historical water quality data to establish baseline patterns and characteristics for each parameter.
* Weather Data: Collect weather data from local weather stations to account for environmental factors that may influence water quality.

2. Data Preprocessing

Before applying anomaly detection techniques, it is essential to preprocess the data:

* Data Cleaning: Handle missing values, outliers, and noisy data points to ensure data quality.
* Feature Engineering: Create relevant features or transformations to improve the effectiveness of anomaly detection.

3. Anomaly Detection Techniques

Several anomaly detection techniques can be applied to the preprocessed data:

* Statistical Methods: Utilize statistical methods like z-scores, mean and standard deviation analysis to identify deviations from the norm.
* Machine Learning Models: Train machine learning models such as Isolation Forests, One-Class SVM, or Autoencoders to identify anomalies in real-time.
* Time Series Analysis: Apply time series analysis techniques to capture temporal patterns and detect anomalies based on historical data.

4. Visualization and Alerts

Use IBM Cognos Analytics to create interactive dashboards for water quality monitoring. These dashboards should include:

* Real-time parameter visualizations.
* Anomaly indicators and alerts triggered when anomalies are detected.
* Historical trend analysis to identify long-term shifts in water quality.

5. Continuous Improvement

Implement a feedback loop for continuous improvement:

* Regularly update anomaly detection models to adapt to changing conditions.
* Incorporate user feedback and domain knowledge to refine the system.
* Conduct periodic audits to ensure the accuracy of anomaly detection.

**Benefits**

The proposed solution offers several benefits:

* Early Anomaly Detection: Promptly identify deviations in water quality parameters, allowing for rapid response to potential issues.
* Real-time Monitoring: Continuously monitor water quality parameters in real-time, reducing the reliance on periodic manual sampling.
* Improved Public Health: Ensure the delivery of safe and clean water to the community, minimizing health risks associated with poor water quality.
* Efficient Resource Allocation: Allocate resources more efficiently by focusing efforts on areas with detected anomalies.

**CODE :**

import pandas as pd

import numpy as np

from sklearn.ensemble import IsolationForest

import matplotlib.pyplot as plt

# Load your water quality data into a DataFrame (assuming CSV format).

data = pd.read\_csv('water\_quality\_data.csv')

# Select the columns you want to analyze (e.g., pH, dissolved oxygen).

selected\_columns = ['ph', 'dissolved\_oxygen']

# Create a sub-dataframe with selected columns.

sub\_data = data[selected\_columns]

# Train the Isolation Forest model on your data.

clf = IsolationForest(contamination=0.05) # You can adjust the contamination parameter.

clf.fit(sub\_data)

# Predict anomalies for each data point.

anomaly\_scores = clf.decision\_function(sub\_data)

anomalies = clf.predict(sub\_data)

# Add the anomaly scores and predictions to the original DataFrame.

data['anomaly\_score'] = anomaly\_scores

data['is\_anomaly'] = anomalies

# Visualize anomalies.

plt.figure(figsize=(12, 6))

plt.scatter(data['timestamp'], data['ph'], c=data['is\_anomaly'], cmap='viridis')

plt.xlabel('Timestamp')

plt.ylabel('pH')

plt.title('Anomaly Detection for pH')

plt.colorbar()

plt.grid(True)

plt.show()

# You can repeat the same process for other water quality parameters and create more visualizations.

# Filter out the anomalies if needed.

anomalous\_data = data[data['is\_anomaly'] == -1]

# Save the results to a CSV file if necessary.

anomalous\_data.to\_csv('anomalous\_water\_quality\_data.csv', index=False)

**Conclusion:**

* Integrating anomaly detection techniques into water quality analysis using IBM Cognos Analytics offers a powerful solution to enhance the monitoring and management of water quality. By leveraging real-time data analytics and visualization, we can proactively address water quality issues, protect public health, and ensure the delivery of high-quality water to the community. This innovative approach represents a significant advancement in water quality management and should be considered for implementation in water supply systems.