IBM Data Analytics Internship

Group Name - The Insight Crew

Group Leader - Raj Nanasaheb Patil

Team Members -

1. Niraj Vikram More
2. Swayam Vijay Bhosale
3. Bhavesh Avinash Gadekar
4. Om Raju Golesar
5. Siddhesh Balasaheb Kankrale
6. Neha Sanjay Gaikwad
7. Ishwari Shyam Yadav
8. Chaitali Pradip Patil

CONCEPT NOTE

Water Quality Monitoring and Prediction (SDG 6: Clean Water and Sanitation)

Concept of the Project

Water quality is a critical issue that affects ecosystems, human health, and economic activities. This project aims to monitor and predict water quality in urban water bodies to ensure safe and clean water, addressing the challenges of pollution and resource management. By leveraging data analysis tools and methodologies, the project seeks to propose actionable solutions that align with Sustainable Development Goal 6 (SDG 6): Clean Water and Sanitation. This SDG aims to ensure availability and sustainable management of water and sanitation for all.

Problem Statement

Urban water bodies are facing increasing levels of pollution due to rapid industrialization, urban runoff, and inadequate wastewater treatment. Poor water quality poses significant risks to public health, contributing to waterborne diseases, and adversely affects aquatic ecosystems. Furthermore, contaminated water can disrupt local biodiversity and negatively impact the economy, particularly in sectors reliant on clean water such as agriculture and tourism. Despite efforts to manage water pollution, challenges persist due to the lack of real-time monitoring and comprehensive data analysis. This project aims to address these challenges by leveraging advanced technologies to monitor and predict water quality, identifying pollution sources and trends, and proposing targeted interventions to ensure safe and clean water in urban areas.

Objective of the Project

The primary objective of this project is to monitor and predict water quality in urban water bodies to ensure safe and clean water, and to propose data-driven solutions that can help improve water quality. The specific objectives are:

- To collect and analyse water quality data from reliable sources
- To identify the primary sources of water pollution in urban areas.
- To understand the temporal and spatial trends of water quality.
- To develop predictive models for future water quality levels based on current data.
- To propose actionable solutions and policy recommendations to mitigate water pollution.
- To assess the potential impact of the proposed solutions on achieving SDG 6: Clean Water and Sanitation.

Data Sources Used

The project will use air quality datasets from the following sources:

1. Kaggle: Various air quality datasets are available on Kaggle, such as the "Water Quality Data" and "India Water Quality Data."
2. Government Websites: Datasets from governmental organisations like the Environmental Protection Agency (EPA) in the USA, the European Environment Agency (EEA), and Central Pollution Control Board (CPCB) in India.
3. OpenAQ: While primarily an air quality platform, OpenAQ can be cross-referenced for air-water pollution studies.
4. World Health Organization (WHO): Water quality guidelines and global reports.

Features

The key features of the dataset will include:

- Location: Geographic coordinates of the monitoring stations
- Pollutants: Levels of various pollutants such as NO3-, PO4 3-.
- Time: Temporal data indicating the exact date and time of each recording.
- Weather Conditions: Temperature, Precipitation, humidity, wind speed and DIrection, and Solar Radiation.

Tool for Analysis

The following tools and technologies will be used for data analysis:

1. Python: For data cleaning, analysis, and visualization, using libraries such as Pandas, NumPy, Matplotlib, and Seaborn.
2. Jupyter Notebooks: For documenting the analysis process and visualizations.
3. Scikit-learn: For developing predictive models and machine learning algorithms.
4. QGIS: For spatial analysis and creating geographic visualizations of air quality data.
5. Tableau: For creating interactive dashboards and visualizations to present the findings.

Hypothesis

The implementation of advanced water quality monitoring systems and predictive analytics will lead to a significant improvement in the management of urban water bodies. Specifically, the continuous real-time monitoring and predictive modelling will enhance the early detection of contamination events and enable targeted interventions. This approach will result in a notable reduction in pollution levels, improved water quality, and better compliance with environmental regulations over the next decade. Additionally, identifying specific temporal and spatial patterns in water quality data will allow for more effective pollution control strategies and resource allocation.

Methodology

The project will be conducted in the following phases:

Data Collection:

- Gather water quality data from the aforementioned sources.
- Compile weather and other relevant data to support the analysis.

Data Cleaning and Preprocessing:

- Handle missing values, outliers, and inconsistencies in the data.
- Standardize data formats and integrate datasets from different sources.

Exploratory Data Analysis (EDA):

- Perform descriptive statistical analysis to understand the distribution and variability of pollutants.
- Visualize changes in water quality over time (daily, monthly, seasonal) using time series plots.

Source Identification:

- Use statistical methods to identify correlations between water quality metrics and potential pollution sources.
- Analyze the impact of different factors (e.g., rainfall, temperature, industrial discharge) on pollution levels.

Predictive Modelling:

- Develop machine learning models (e.g., linear regression, random forest, LSTM) to predict future water quality based on historical data and influencing factors.
- Validate models using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.

Solution Development:

- Based on the analysis, propose solutions such as stricter emissions regulations, promotion of public transportation, and green infrastructure.
- Evaluate the feasibility, cost, and potential impact of the proposed solutions on urban water bodies.

Reporting and Presentation:

- Compile the findings into a comprehensive report.
- Create visualizations and interactive dashboards to present the results.
- Develop policy briefs and recommendations for stakeholders.

Probable Outcome

The expected outcomes of the project are:

- Comprehensive Analysis: A thorough analysis of water quality data from various urban water bodies to identify key sources of contamination, trends, and patterns affecting water quality.
- Predictive Models: Reliable models for predicting future water quality, identifying potential contamination events, and forecasting trends based on historical and real-time data.
- Actionable Solutions: Data-driven recommendations for reducing water pollution, such as optimizing industrial discharge practices, improving waste management, and enhancing stormwater management.
- Impact Assessment: Evaluation of progress towards achieving SDG 6: Clean Water and Sanitation by monitoring improvements in water quality and overall environmental health.
- Awareness and Engagement: Increased awareness among policymakers, local authorities, and the public about the current state of water quality, sources of pollution, and the benefits of proposed interventions.

By addressing water quality through comprehensive data analysis and predictive modelling, this project will contribute to creating safer and cleaner urban water systems, aligning with the objectives of SDG 6: Clean Water and Sanitation