PROJECT DOCUMENT – PHASE 2

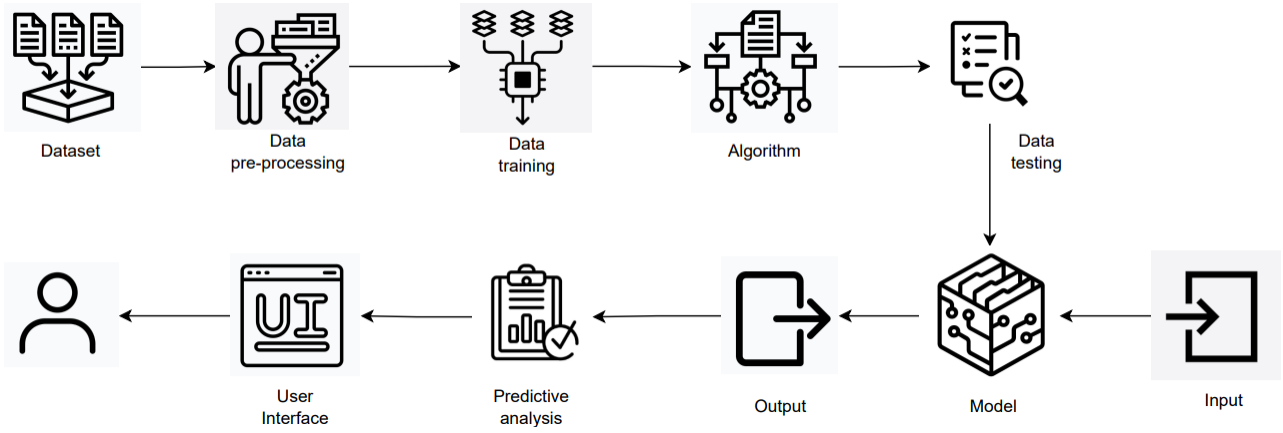
WATER QUALITY ANALYSIS

Madras Institute of Technology, Anna University

Harishma Sabu Ranjana S Ranjini S Roshni N Vithya S

Problem Statement

The ever-expanding process of urbanization has brought forth a pressing issue - ensuring access to safe drinking water. This challenge is amplified by the contamination of water sources due to a multitude of factors, underscoring the importance of water quality analysis and real-time monitoring.



Analytic Flowchart

Solution Description

We have introduced a solution that employs a spectrum of machine learning algorithms to predict the Water Quality Index (WQI). Our predictive model assesses key water quality parameters, including alkalinity, pH levels, temperature, turbidity, dissolved oxygen, and mineral and nutrient content (such as nitrogen and phosphorous). These calculated WQI values are instrumental in determining the suitability of water for various specific applications.

Novelty

This approach stands out by extending beyond the conventional water sample analysis. In addition to this analysis, we harness advanced machine learning techniques to gauge the water's usability and its applicability for specific use cases.

Social Impact

Prioritizing customer satisfaction is paramount, particularly within the framework of total quality management. Given the recent deterioration of water quality attributed to various pollutants, predicting water quality becomes indispensable for managing water pollution and safeguarding the quality of water for consumers. Our evaluation model plays a pivotal role in quantifying and enhancing customer satisfaction.

Business Model (Revenue Model)

Our business model focuses on the enhancement of technology and production processes, translating to increased profitability and streamlined logistics. Furthermore, it offers a revenue stream by aiding users in identifying potential risks associated with water bodies and categorizing nearby water sources according to their suitability for various purposes.

Scalability of the Solution

Our solution boasts impressive scalability, capable of handling vast volumes of data collected from water sources, ranging from smaller water bodies to expansive aquatic ecosystems. This scalable system can adeptly process and analyze the data to meet the real-time needs of a substantial user base.

Data Collection:

- **Web Browsers**: Download data from the dataset.
- **Data Collection Tools**: We need web scraping tools or APIs to collect data if it's not available in a downloadable format.

Data Preprocessing:

- **Python**: For data manipulation and preprocessing.
- **Jupyter Notebook**: For interactive data analysis.
- **Pandas**: A Python library for data cleaning and manipulation.
- **NumPy**: For numerical operations.

Data Visualization:

- **Matplotlib** and **Seaborn**: For creating static visualizations.
- **Plotly** or **Bokeh**: For interactive visualizations.
- **Tableau** or **Power BI**: For advanced data visualization and dashboard creation.

Machine Learning and Predictive Modeling:

- **Python**: For implementing machine learning models.
- **Scikit-Learn**: A Python library for machine learning.
- **XGBoost**, **LightGBM**, or **CatBoost**: For gradient boosting algorithms.

Conclusion

In this second phase, we have defined the project's objectives and outlined the design thinking process for analyzing water quality data. The subsequent phases will involve data preprocessing, exploratory data analysis, model development, and evaluation. This systematic approach will enable us to effectively address the problem and provide valuable insights for water quality assessment and decision-making.