Dataset Link

https://drive.google.com/file/d/1qf0JzKWaPCUARXarkASnIkNmu1fE8_PG/view?usp=drive_link

---

Dataset Details

1. **pH value**:
    - pH is an important parameter in evaluating the acid–base balance of water. It is also the indicator of acidic or alkaline conditions of water status. WHO has recommended a maximum permissible limit of pH from 6.5 to 8.5. The current investigation ranges were 6.52–6.83, which are within the WHO standards.
2. **Hardness**:
    - Hardness is mainly caused by calcium and magnesium salts. These salts are dissolved from geologic deposits through which water travels. The length of time water is in contact with hardness-producing material helps determine how much hardness there is in raw water. Hardness was originally defined as the capacity of water to precipitate soap caused by Calcium and Magnesium.
3. **Solids (Total Dissolved Solids - TDS)**:
    - Water has the ability to dissolve a wide range of inorganic and some organic minerals or salts such as potassium, calcium, sodium, bicarbonates, chlorides, magnesium, sulfates, etc. These minerals produce unwanted taste and dilute the color in the appearance of water. Water with high TDS values indicates that water is highly mineralized. The desirable limit for TDS is 500 mg/l and the maximum limit is 1000 mg/l prescribed for drinking purposes.
4. **Chloramines**:
    - Chlorine and chloramine are the major disinfectants used in public water systems. Chloramines are most commonly formed when ammonia is added to chlorine to treat drinking water. Chlorine levels up to 4 milligrams per liter (mg/L) or 4 parts per million (ppm) are considered safe in drinking water.
5. **Sulfate**:
    - Sulfates are naturally occurring substances found in minerals, soil, and rocks. They are present in ambient air, groundwater, plants, and food. The principal commercial use of sulfate is in the chemical industry. Sulfate concentration in seawater is about 2,700 milligrams per liter (mg/L). It ranges from 3 to 30 mg/L in most freshwater supplies, although much higher concentrations (1000 mg/L) are found in some geographic locations.
6. **Conductivity**:
    - Pure water is not a good conductor of electric current; it’s rather a good insulator. The increase in ion concentration enhances the electrical conductivity of water. Generally, the amount of dissolved solids in water determines the electrical conductivity. Electrical conductivity (EC) measures the ionic process of a solution that enables it to transmit current. According to WHO standards, EC value should not exceed 400 μS/cm.
7. **Organic Carbon**:
    - Total Organic Carbon (TOC) in source waters comes from decaying natural organic matter (NOM) as well as synthetic sources. TOC is a measure of the total amount of carbon in organic compounds in pure water. According to US

EPA, TOC should be < 2 mg/L in treated/drinking water, and < 4 mg/L in source water used for treatment.

8. **Trihalomethanes (THMs)**:
    - THMs are chemicals that may be found in water treated with chlorine. The concentration of THMs in drinking water varies according to the level of organic material in the water, the amount of chlorine required to treat the water, and the temperature of the water. THM levels up to 80 ppm are considered safe in drinking water.
9. **Turbidity**:
    - The turbidity of water depends on the quantity of solid matter present in a suspended state. It is a measure of light-emitting properties of water and the test is used to indicate the quality of waste discharge with respect to colloidal matter. The mean turbidity value obtained for Wondo Genet Campus (0.98 NTU) is lower than the WHO-recommended value of 5.00 NTU.
10. **Potability(Target)**:
    - Indicates if water is safe for human consumption where 1 means **Potable** and 0 means **Not Potable**.

Submission Criteria

- **Project Repository**: The project should be uploaded on **GitHub**.
    - The repository name should be the **team name**.
    - The repository must include:
        - A **demo video** explaining the project, its functionality, and how it works.
        - The **IPython Notebook (.ipynb file)** used for training the model and EDA.
        - The trained **model file** (.pkl) and any other necessary files(frontend & backend).

Evaluation Criteria (Total: 100%)

1. **Model Performance + Preprocessing – 25%**
    - Model accuracy, F1-Score, or other relevant metrics.
    - Feature engineering, handling of missing values, scaling, and transformations.
2. **Exploratory Data Analysis (EDA) – 10%**
    - Visualizations, correlation analysis, distribution checks.
    - Identification of data imbalances, outliers, or anomalies.
3. **Backend (FastAPI) – 20%**
    - API development, integration with the model.
    - Scalability, asynchronous handling, and overall architecture.
4. **Frontend (Streamlit, React, etc.) – 20%**
    - UI/UX design and layout.
    - Integration with the backend and user experience.
5. **PDF Report – 15%**
    - Explanation of the approach, insights from EDA, and model rationale.

- Visualizations and challenges faced during the project.

6. **Code & Project Structure – 10%**
    - Clean, modular, and well-commented code.
    - Proper Git usage and environment setup.