Project Tasks Outline

# 1. Dataset Download and Description

Download the dataset and explain each feature briefly, highlighting its importance for water quality analysis.

# 2. Exploratory Data Analysis (EDA)

Perform visualizations, check distributions, and analyze correlations between features to uncover data patterns.

# 3. Data Preprocessing

Handle missing values, apply scaling or normalization, and prepare the dataset for model building.

# 4. Outlier Detection

Identify and treat outliers using statistical methods like boxplots or Z-score to ensure clean data for modeling.

# 5. Train/Test Split

Split the dataset into training and testing subsets to properly evaluate the model’s performance.

# 6. Model Building

Train different classification models, tune hyperparameters, and select the best-performing model.

# 7. Model Evaluation

Evaluate models using Accuracy, F1-score, and Confusion Matrix to assess predictive performance.

# 8. Saving the Model

Save the trained machine learning model into a .pkl file for later use in the backend.

# 9. Backend Development (FastAPI)

Create a FastAPI server that loads the trained model and exposes a prediction API endpoint.

# 10. Frontend Development (Streamlit/React)

Develop a simple web interface where users can input features and receive water potability predictions.

# 11. GitHub Repository Setup

Organize and upload all project files to GitHub, including notebook, backend, frontend, and model files.

# 12. Demo Video Recording

Record a short video explaining the project flow, model usage, frontend interaction, and API working.

# 13. Final Report Writing

Summarize the full project in a PDF report including insights, challenges, and future improvements.