

DS02: Water Quality Classification & Public Health Analytics

Project Requirements Document

This document specifies the technical and non-technical requirements needed to execute the DS02 project successfully.

1. Hardware Requirements

- 1 Laptop/Desktop with minimum 8 GB RAM (16 GB recommended)
- 2 At least 20 GB free storage
- 3 Stable internet connection

2. Software Requirements

- 1 Operating System: Windows / Linux / macOS
- 2 Python 3.9 or higher
- 3 Jupyter Notebook or Google Colab
- 4 VS Code (optional)

3. Python Libraries Required

- 1 pandas
- 2 numpy
- 3 matplotlib
- 4 seaborn
- 5 scikit-learn
- 6 imbalanced-learn
- 7 scipy

4. Dataset Requirements

- 1 Source: National Water Monitoring Programme (NWMP), MPCB
- 2 File: NWMP_August2025_MPCB_0.csv
- 3 Rows: ~225, Columns: 54
- 4 Target variable: Use Based Class (A, B, C, E)

5. Knowledge & Skill Requirements

- 1 Python programming fundamentals
- 2 Data preprocessing and EDA
- 3 Machine learning classification concepts

- 4 Understanding of evaluation metrics
- 5 Basic environmental science knowledge

6. Optional Tools

- 1 Streamlit for dashboard deployment
- 2 GitHub for version control
- 3 PowerPoint / Google Slides for presentation

These requirements ensure reproducibility, scalability, and smooth execution of the DS02 project.