# Data Collection and Preprocessing Phase

| | |
|---|---|
| Date | 20 June 2025 |
| Project Title | Rising Waters: A Machine Learning Approach to Flood Prediction |
| Maximum Marks | 2 Marks |

**Data Quality Report:**

The Data Quality Report will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

**Data Quality Report:**

| Data Source | Data Quality Issue | Severity | Resolution Plan |
|---|---|---|---|
| Kaggle Dataset | Missing values in columns such as Temperature, Humidity, Cloud Cover, Rainfall (seasonal), or Runoff | Moderate | Use mean/median imputation. |
| Kaggle Dataset | Presence of categorical or mixed-type columns (e.g., possibly mislabeled or region-specific identifiers) | Moderate | Apply appropriate encoding or remove unused categorical columns |
| Kaggle Dataset | Outliers in continuous variables like Rainfall and Temperature | Moderate | Use IQR-based outlier removal techniques |
| Kaggle Dataset | Class imbalance between flood and no-flood cases | Moderate | Apply class balancing techniques such as SMOTE or class weights |