Data Collection and Preprocessing Phase

Date	13 June 2025
Team ID	SWTID1749618778
Project Title	Rising Waters: A Machine Learning Approach To Flood Prediction
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	The project aims to predict flood events using machine learning models based on environmental data. The dataset includes features like temperature, humidity, cloud cover, and seasonal rainfall to forecast flood risks and enable early warnings.
Data Collection Plan	• Search for meteorological and hydrological datasets related to floods • Focus on datasets with seasonal, regional, and environmental diversity • Ensure flood event labels (yes/no) are included
Raw Data Sources Identified	The primary data source for this project is a cleaned Excel dataset comprising historical weather metrics and labeled flood events. It includes variables such as rainfall by season, temperature, humidity, and cloud cover, which are used for building predictive models.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	Contains temperature, humidity, cloud cover, seasonal rainfall totals, and flood labels for each year	https://www.kaggle.com/arbethi/rainfall-dataset?select=flood+dataset.xlsx	XLSX	16 KB	Public