Data Collection and Preprocessing Phase

Date	23 September 2024
Team ID	LTVIP2024TMID24992
Project Title	Rainfall Prediction Using Machine Learning
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	Rainfall prediction using machine learning entails examining historical weather data to predict future precipitation. By employing advanced algorithms such as Decision Trees, Random Forest, and Neural Networks, we achieve remarkable accuracy in forecasting rainfall patterns. This significantly supports agricultural planning, water resource management, and disaster preparedness, leading to more informed and effective decision-making.
Data Collection Plan	• Searching for Datasets: Look for datasets related to rainfall occurrence from reliable sources like meteorological departments, online databases (e.g., NOAA, OpenWeatherMap), and research institutions. Prioritize datasets that include comprehensive weather metrics over an extended period. • Prioritize dataset with various demographic information

Raw Data Sources Identified	Gather extensive historical weather data, including temperature, humidity, wind speed, and past rainfall records, from reliable sources like local meteorological stations, national meteorological databases, and online platforms such as NOAA or OpenWeatherMap. Ensure data spans multiple years to capture seasonal and annual variations.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Smart Internz Platform	https://docs.google.com/spreadsheets/d/1RA2OO0LZTeQykI_mvnensAjp6LM4YzWI1Tz0SUG5-Ao/edit?usp=sharing	CSV	13.5 MB	Public
Dataset 2	Kaggle	https://www.kaggle.com/datasets/rajanand/rainfall-in-india	CSV	192 KB	Public