

Data Collection and Preprocessing Phase

Date	24 April 2024
Team ID	Team-738169
Project Title	Rainfall Prediction Using Machine Learning
Maximum Marks	2 Marks

Data Quality Report:

The Data Quality Report Template 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 Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing values in the 'Evaporation', 'Sunshine'	High	Using Random sample imputation.
Kaggle Dataset	Missing values in the 'MinTemp', 'MaxTemp', 'WindGustSpeed', 'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm', 'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am', 'Temp3pm'	Moderate	Using median imputation.
Kaggle Dataset	Missing values in the 'RainToday', 'RainTomorrow'	Low	Using One Hot Encoding
Kaggle Dataset	Missing values in the "WindGustDir", "WindDir9am" and "WindDir3pm"	High	Using Label Encoding