

Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID728136330
Project Title	Fake news analysis on social media using NLP
Maximum Marks	2 Marks

Data Quality Report Template

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
Dataset	Language Variability: Posts in multiple languages; requires uniform language for NLP..	Moderate	Use language detection to filter or translate posts to the target language, ensuring uniformity.
Dataset	Duplicate Records: Duplicate posts or similar entries in the dataset.	High	Use record deduplication techniques, e.g., hashing text content and removing duplicates, to retain only unique records.