

## Data Collection and Preprocessing Phase

Date	23 September 2024
Team ID	LTVIP2024TMID25030
Project Title	FAKE NEWS ANALYSIS IN SOCIAL MEDIA
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
Twitter Dataset	Missing values in fields like user_location, tweet_language, and hashtags.	Moderate	Use mean/median imputation for continuous data, or default values for categorical data.
General Social Media Data	Categorical data, such as post_type and user_verification_status.	Moderate	Apply encoding (e.g: One-hot encoding or Label encoding)
Twitter and Facebook Dataset	Presence of outliers in engagement metrics such as likes, shares, retweets	High	Remove or normalize outliers based on statistical analysis