



Tanglis
Tanglis is a code-mixed Tamil-English chat dataset developed to support sentiment analysis of informal digital conversations. It contains text messages that combine Tamil and English expressions,...
Dataset Type: Structured
File Size Assessed: 322.60 KB/322.60 KB



1. Data Hygiene and Integrity

This section evaluates the accuracy, completeness, and correctness of the dataset.



Criteria	Sub-Criteria	Score	Justification
Data Hygiene	Missing Values	15/15	✔ There are no missing values in the dataset
	Schema Consistency	5/5	✔ The metadata includes Schema Consistency
Data Completeness	Geographical Coverage	5/5	✔ The metadata includes Geographical Coverage
Data Variance and Correctness	Data Uniqueness	5/5	✔ The datasets have no duplicates
	Standard Deviation	5/5	✔ The dataset columns have sufficient variance

2. Data Consistency & Usability

This section checks the dataset's standardization and update frequency.



Criteria	Sub-Criteria	Score	Justification
Data Standardization	File Format	15/15	✔ The dataset is in standard file format
Regular Refresh & Longitudinal View	Regular Updates	15/15	✔ The dataset is frequently updated

3. Data Maintenance & Documentation

This section reviews metadata completeness and compliance.



Criteria	Sub-Criteria	Score	Justification
Documentation	Tags	5/5	✔ The dataset metadata has the tags present
	Source/Provenance	5/5	✔ The dataset metadata has the source present
	Cover Image	5/5	✔ The dataset metadata has the cover image present
	Short Description	5/5	✔ The dataset metadata has the short description present
	Long Description	5/5	✔ The dataset metadata has the long description present
	Year Range	0/5	✘ The dataset metadata does not have year range present
	License	5/5	✔ The dataset metadata has the license present