

Data Intake Report

Name: Advance NLP: Hate Speech detection using Transformers

Report date: 17.01.2024

Internship Batch: LISUM28

Version: 1.0

Data intake by: Parthkumar Sutariya (Group: Solitude Ensemble)

Data intake reviewer: intern who reviewed the report

Data storage location: <https://www.kaggle.com/datasets/vkrahul/twitter-hate-speech>

Tabular data details:

Total number of observations	31962
Total number of files	2
Total number of features	3
Base format of the file	train_data.csv, test_data.csv
Size of the data	4.74 MB

Proposed Approach:

- Firstly, I read data using pandas
- Identify Whitespaces, commas, irrelevant info., etc.