Data Intake Report

Name: **NLP - Twitter Hate Speech Detection with Transformer (Deep Learning)**

Report date: 14.05.2025

Internship Batch: LISUM 45

Version:<1.0>

Data intake by: Ankita Manna

Data intake reviewer: Data Glacier

Data storage location: Github

**Training file details:**

|  |  |
| --- | --- |
| **Total number of observations** | 31962 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | csv |
| **Size of the data** | 3mb |

**Test file details:**

|  |  |
| --- | --- |
| **Total number of observations** | 17198 |
| **Total number of files** | 1 |
| **Total number of features** | 2 |
| **Base format of the file** | csv |
| **Size of the data** | 1.55mb |

**Proposed Approach:**

The training dataset is used for training the model, testing, and validation. The test dataset is sent through the model and get the results of those tweets.

Assumption:

The tweets are not the raw text. It comes with noise.