# Data Intake Report

Name: Advance NLP: Hate Speech detection using Transformers (Deep Learning) - Group Project
Report date: 19/09/2024
Internship Batch: LISUM36
Version:
Data intake by: Michael Udonna Egbuzobi, Nweke Nonye and Sreedhar Rongala
Data intake reviewer:
Data storage location: NonnyCodes/Advance-NLP: Advance NLP: Hate Speech detection using Transformers (Deep Learning) - Group Project (github.com)

We have two datasets – train and test data.

**Tabular data details (train):**

| | |
|---|---|
| **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** | 3,031 KB |

**(Test)**

| | |
|---|---|
| **Total number of observations** | 17197 |
| **Total number of files** | 1 |
| **Total number of features** | 2 |
| **Base format of the file** | .csv |
| **Size of the data** | 1,598 KB |

**Note: Replicate same table with file name if you have more than one file.**

**Proposed Approach:**
- Problem Understanding
- Data Cleaning and Normalisation
- Representation Learning
- Model Building & Training
- Performance Evaluation & Reporting
- Model Deployment
- Model Inference