

# 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: <https://github.com/UdonnaM/Advance-NLP-Hate-Speech-detection-using-Transformers-Deep-Learning->

We have two datasets – train and test data.

## Tabular data details (train):

<b>Total number of observations</b>	31962
<b>Total number of files</b>	1
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	3,031 KB

## (Test)

<b>Total number of observations</b>	17197
<b>Total number of files</b>	1
<b>Total number of features</b>	2
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	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