

# Data Intake Report

Name: <hate speech detection using transformers (Deep Learning)>

Report date: <19/07/2023>

Internship Batch:<LISUM 2023>

Version:<1.0>

Data intake by:<Zyad Hussein>

Data intake reviewer:<Zyad Hussein>

Data storage location: <Github: >

## Tabular data details:

### Train file:

<b>Total number of observations</b>	31963
<b>Total number of files</b>	2
<b>Total number of features</b>	3 in train csv
<b>Base format of the file</b>	csv
<b>Size of the data</b>	2.95 MB

### Test File:

<b>Total number of observations</b>	17198
<b>Total number of files</b>	2
<b>Total number of features</b>	2 in test.csv
<b>Base format of the file</b>	csv
<b>Size of the data</b>	1.55 MB

## Proposed Approach:

- Identification of data cleaning required.
- Identification of model requirements.
- Research available models.
- Implement a chosen model.
- Test the output.
- Verify the results.
- Refine the models if needed.
- Deploy the model on a web application.
- Submit the deployed model.