

# Data Intake Report

Name: Advance NLP: Hate Speech Detection with Transformers (Deep Learning)

Report date: September 18, 2023

Internship Batch: LISUM24

Version:1.0

Data intake by: Siobhan Hwang

Data intake reviewer: Andrew O'Drain

Data storage location: [GitHub](#)

## Tabular data details: train\_tweets.csv

<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>	3031KB

## Dedupe Validation:

- Features:
  - id int64
  - label int64
  - tweet object
- 0 duplicate rows

## Tabular data details: test\_tweets.csv

<b>Total number of observations</b>	17196
<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>	1598KB

## Dedupe Validation:

- Features:
  - id int64
  - tweet object
- 0 duplicate rows