

Data Intake Report

Name: Advance NLP : Hate Speech detection using Transformers (Deep Learning)

Report date: 4/19/2023

Internship Batch: LISUM19

Version: 1.0

Data intake by: Snigdha Chigurupati

Data intake reviewer: NA

Data storage location: NA

Tabular data details:

train_E6oV3lV.csv

Total number of observations	31963
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	3.031 MB

test_tweets_anuFYb8.csv

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	1.598 MB

PROBLEM DESCRIPTION:

The term hate speech is understood as any type of verbal, written or behavioral communication that attacks or uses derogatory or discriminatory language against a person or group based on what they are, in other words, based on their religion, ethnicity, nationality, race, color, ancestry, sex or another identity factor. In this problem, We will take you through a hate speech detection model with Machine Learning and Python.

Hate Speech Detection is generally a task of sentiment classification. So, for training, a model that can classify hate speech from a certain piece of text can be achieved by training it on a data that is generally used to classify sentiments. So, for the task of hate speech detection model, We will use the Twitter tweets to identify tweets containing hate speech.

BUSINESS UNDERSTANDING:

The problem statement involves developing a hate speech detection model using machine learning and Python, with a focus on identifying hate speech in Twitter tweets. This is important for businesses and organizations that rely on social media platforms to engage with customers and stakeholders, as hate speech can lead to negative publicity, damage to brand reputation, and even legal repercussions. By leveraging a hate speech detection model, businesses can

proactively monitor and filter out harmful content on social media, creating a safer and more inclusive online environment for their audience. Additionally, such a model can help businesses comply with regulations related to hate speech, thereby avoiding potential penalties and legal issues.