

DataGlacier NLP Interns Group

MEMBER DETAILS

➤ Member 1:

- Name – Gabriel Aluede
- Email – ghabyaluede@gmail.com
- Country – United Kingdom
- College – Teesside University
- Specialization – NLP

➤ Member 2:

- Name – Yosha Udaya Shriyan
- Email – yosha.shriyan@gmail.com
- Country – United Kingdom
- College – Royal Holloway University of London
- Specialization – NLP

PROBLEM DESCRIPTION

The term hate speech is understood as any type of verbal, written or behavioural 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, colour, ancestry, sex or another identity factor. In this problem, we will take you through a hate speech detection model with Machine Learning and Python.

BUSINESS UNDERSTANDING

It is essential to identify hate speech on social media. Unchecked hate speech has the potential to destroy our society and cause great harm to marginalised people or groups. Social media is a significant platform for the online propagation of hate speech. Hate Speech Recognition Models can effectively transform the social media platform into a space where everyone can express themselves without offending anyone's feelings by blocking posts, comments, and accounts that promote hate speech.

PROJECT LIFECYCLE ALONG WITH DEADLINE

	Week	7	8	9	10	11	12	13
Task	Connect with team members							
	Select project Topic							
	Research on Topic							
	Distribute Work							
	Gather data							
	Explore the data in hand							
	Make report on data cleaning							
	Identify any issues in data							
	Come up with solutions to problems							
	Make Report							

	Perform EDA on ready data							
	Make EDA report							
	EDA Presentation							
	Model Selection							
	Final Submission							

DATA INTAKE REPORT

Name: Hate Speech Recognition

Internship Batch: LISUM11: 30

Version: 1.0

Data Intake By: Yosha Shriyan

Data Intake Reviewer: Gabriel Aluede

Data Storage Location: [Hate Speech Data Kaggle](#)

Train_tweets

Total Number of Observations	31962
Total Number of Files	1/2
Total Number of Features	2
Base Format of the File	.csv
Size of the Data	2.95 MB

Test_tweets

Total Number of Observations	17197
Total Number of Files	2/2
Total Number of Features	2
Base Format of the File	.csv
Size of the Data	1.55 MB

GITHUB REPO LINK

<https://github.com/YoshaRHUL/DataGlacierProject>