

Week-9 Deliverables

Team member's details :

Group Name: Data Junkies

Name: Farheen Fatima

Email: Farheen.fatima07@gmail.com

Country: USA

College/Company: California State University, Sacramento.

Specialization (Data Science, NLP, Data Analyst): Data Science

Name: Mohamed Derbeli

Email: derbelimohamed1@gmail.com

Country: Spain

College/Company: Data Glacier

Specialization: Data Science

Name : Mohamed Akrem Ben Jemia

Email: akrem008.benjamie@gmail.com

Country: Spain

College/Company : École National d'ingénieur de Carthage

Specialization: NLP

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.

GitHub Repo link: https://github.com/Mohamed-derbeli/Hate_Speech_Detection_Using_Transformers

Data Cleaning and Transformation:

The following steps were performed for Text processing:

1. Checked for missing/Null values – No null values found.
2. Removed punctuation using regex
3. Removed stop-words using NLTK library
4. Removed numbers, links, and emojis using regex
5. Performed tokenization and Lemmatization