

Project: Hate Speech detection using Transformers (Deep Learning)

Group Name: Itvn_sergey

Name: Litvinov Sergey

Email: Itvn.sergey.work@gmail.com

Country: Russia

Company: NVI Solutions

Specialization: NLP

Problem description

Hate speech detection is a problem of sentiment classification. The task is to classify certain piece of text as the one with hate speech or not, by training deep learning model on labeled data.

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.

Business understanding

Twitter is a social net where people can share messages with each other using so called 'tweets'. Users are not censored in any way, so they can write whatever they want.

This idea of free speech works great when people are sharing something which has positive impact on the society like knowledge, news, videos with cats and so on. But there is other side of the coin, when it comes to individuals who want to spread violent ideas, mock other people and apply to anything which is referred to hate speech.

Due to the fact that there are millions of users in Twitter, it is hard for service teams to find and moderate tweets with hate speech.

The goal is to automate searching process by using deep learning to detect such tweets, thus make Twitter safer space for everyone, attract more advertisement companies and build overall positive reputation for the net.

Project lifecycle along with deadline

Stages	Due date - dd.mm.yy
Problem Understanding	19.04.23
Data Analysis and preprocessing	02.05.23
Proposed modeling technique	16.05.23
Model Building and Training	23.05.23
Model Deployment and Inference	27.05.23
Final Report	30.05.23

Data Intake report

Name: Hate Speech detection using Transformers (Deep Learning)

Report date: 16.04.2023

Internship Batch: LISUM19

Version:1.0

Data intake by: Litvinov Sergey

Data intake reviewer:

Data storage location:

https://github.com/LtvnSergey/Internship_Project_Twitter-Hate-Speech/tree/main/dataset

Tabular data details:

File Name	train.csv
Total number of observations	29530
Total number of files	1
Total number of features	3
Base format of the file	.csv

Size of the data	3.1 Mb
-------------------------	--------

File Name	test.csv
Total number of observations	16130
Total number of files	1
Total number of features	2
Base format of the file	.CSV
Size of the data	1.64 Mb

Github Repo link

https://github.com/LtynSergey/Internship_Project_Twitter-Hate-Speech