PROJECT : Hate-Speech-detection-using-Transformers-Deep-Learning

Group Name: Hate Speech Detective

Members:

Name: VARSHIT MANEPALLI

Email: varshitmanepalli1810@gmail.com

Country: INDIA

College/Company: Stevens Institute of Technology

Specialization: NLP

Batch code : LISUM32

Problem Description

Hate speech detection aims to identify and classify statements that contain offensive, derogatory, or discriminatory language directed towards individuals or groups based on their identity factors such as religion, ethnicity, nationality, race, color, ancestry, sex, or other identity factors. This project involves developing a machine learning model to detect hate speech in Twitter tweets.

Business Understanding

Hate speech can have serious consequences, including perpetuating discrimination, inciting violence, and causing psychological harm. Detecting hate speech on social media platforms like Twitter is crucial for maintaining a safe and inclusive online environment. By identifying and flagging hate speech, we can help prevent the spread of harmful content and protect vulnerable individuals and communities.

Project Life Cycle

Week	Deadline	Tasks
Week 7	19 May 2024	• Define the problem and understand the business requirements. • Outline the project lifecycle and set deadlines for each phase. • Document the details of the dataset, including its source and structure.
Week 8	26 May 2024	• Analyse the data to understand its characteristics • Identify problems in the data (e.g., missing values, outliers, skewness • Plan approaches to handle the Data issues
Week 9	2 June 2024	• Apply at least 2 different techniques to handle missing values • Try different NLP featurization techniques(e.g., tokenization, embeddings) • Cleaning the data using python • Code review
Week 10	9 June 2024	• Perform EDA to understand the data distribution, correlations and insights. • Based on EDA, provide recommendations for the next steps.
Week 11	16 June 2024	• Prepare and present the EDA findings to business users • Ensure all updates are reflected in the Github repo
Week 12	23 June 2024	• Select a base model and explore one model from each family of models such as Linear models, Ensemble models, Boosting models etc., • Ensure the selected models work according to our business requirements
Week 13	01 July 2024	• Final Repost and Code Presentation • Project Submission