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.

Featurization

It is a process of converting raw text data into numerical representations (features) that can be used by machine learning models for analysis and prediction.

In this project we have applied three featurization techniques :

1. **Bag of Words (BoW) :** Converts text into fixed-length vectors by counting the frequency of each word in the text, without considering the order or context of the words.

   When we apply this BoW on the datasets, we get the following dataset shape :

   ```
   BOW featurized train data shape :  (31962, 33950)
   BOW featurized test data shape :  (17197, 33950)
   ```

2. **Term Frequency-Inverse Document Frequency (TF-IDF) :** Enhances the BoW model by weighting each word's frequency in a document against its frequency across all documents, emphasizing unique words that are more informative.

   When we apply this TF-IDF on the datasets, we get the following dataset shape :

   ```
   TF-IDF featurized train data shape :  (31962, 33950)
   TF-IDF featurized test data shape :  (17197, 33950)
   ```

3. **Word2Vec :** Represents words as continuous vectors in a high-dimensional space, capturing semantic meaning by training on word co-occurrence patterns, allowing for context-aware embeddings.

   When we apply this TF-IDF on the datasets, we get the following dataset shape :

   ```
   Word2Vec featurized train data shape :  (31962, 100)
   Word2Vec featurized test data shape :  (17197, 100)
   ```