IMPLEMENTATION OF HATE-SPEECH USING TRANSFORMERS

Name – Deekshant Nandeshwar
Email – deekshantnandeshwar@gmail.com
Country – United Kingdom
College – University of Surrey
Specialization – Master of Science (Data Science)

Problem Description–

Any form of communication—verbal, written, or nonverbal—that targets or employs disparaging or discriminatory language against an individual or group because of who they are—their religion, ethnicity, nationality, race, color, ancestry, sex, or another identity characteristic—is considered hate speech. We'll walk you through a model that detects hate speech in this issue.

The task of sentiment categorization is typically involved in hate speech detection. Therefore, by using data that is often used to categorize attitudes, a model that can identify hate speech from a specific piece of text may be trained. As a result, in order to identify tweets that include hate speech, we will use the Twitter tweets.

Project lifecycle along with deadline –

Weeks 07	Problem description, Project lifecycle along with deadline, Data Report
Weeks 08	Data Analysis
Weeks 09	different featurization technique
Weeks 10	EDA performed on the data
Weeks 11	EDA presentation for business users
Weeks 12	Model Selection and Model Building
Weeks 13	Final Project Report and Code

Data Collection

The data on Twitter hate speech was extracted from Kaggle and includes 3 characteristics and 31962 observations. It was a dataset created using Twitter data and used to study the identification of hate speech. The text is categorized as either offensive language, hate speech, or neither. It is crucial to be aware that this dataset includes language that could be seen as objectionable, racist, sexist, or homophobic due to the nature of the study.

<u>Data Analysis</u>

- Text Cleaning - We cleaned our text because the data was so disorganized.
- Remove Punctuation - Punctuation should be deleted since it is unnecessary. Because of this, we utilize regular expression to eliminate the punctuation.

- Lowercase - Lowercase word conversion . Although terms like racism and racism have the same meaning when written in lower case, the vector space model represents them as two distinct meanings (resulting in more dimensions). As a result, we change all content to lower case letters.
- Remove URL - In this section, URLs are removed since we are working on a hate speech program that detects hate and free speech and because in order to acquire the result, we can only provide text and not URLs.
- Remove @ and Special Character - We eliminate the @tags, which were essentially used when mentioning someone. Which doesn't matter. The group of symbols known as Remove Special Characters essentially has no meaning

Different Featurization Technique

Preprocessing -

- Tokenization - Tokenization involves cutting the raw text into manageable pieces. We utilise the nltk work tokenize module to turn our text data, which is in paragraph form, into work tokens. These tokens aid in context comprehension or model development for NLP. By examining the word order in the text, tokenization aids in comprehending the text's meaning.
- Removing StopWords - The core of StopWords is "are," etc. The words in our lexicon have no meaning, therefore we don't need them to create an application for detecting hate speech. To remove stop words from a phrase, we first tokenize the text (like we did previously) into words, and if a word appears in the NLTK stop word list, we delete it.
- Lemmatization - Lemmatization is the process of combining a word's several inflected forms into a single unit for analysis. Lemmatization is similar to stemming, except it gives the words context. Thus, it connects words with related meanings into a single term. Like the term intelligently, change intelligent into its base form.
- WordCloud - Often used to display keyword information on websites or to visualise free form text, a worldcloud is a visual representation of text data. Tags are often single words, and the font size or colour of each one indicates its significance.

Feature Extraction

- Splitting the Data into Train into Test - The data was divided into Train. And we allocated 20% for the test and 80% for training. When data is separated into two or more subgroups, this is known as data splitting. A two-part split often involves testing or evaluating the data in one part and training the model in the other. Data splitting is a crucial component in data science, especially when building models from data.
- TF-IDF Model - When the dictionary is complete, we use the Term Frequency-Inverse Document Frequency (TFIDF) model to extract the 2000 most common terms from dictionaries for each Hate/Free Speech of the whole dataset. The frequency of 2000 words is contained in each word count vector throughout the whole dataset file.