Data Science Intern at Data Glacier

Project: Hate Speech Detection using Transformers (Deep Learning)

Week 8: Deliverables

Name: Shubh Goyal

University: Boston University

Email: Shubhg@bu.edu

Country: USA

Specialization: Data Science

Batch Code: LISUM20

Date: 24th May 2023

Submitted to: Data Glacier

Table of Contents:

1. Project Plan

Week	Plan
Week 07	Problem Statement, Data Collection, Data Report
Week 08	Data Preprocessing (Text Cleaning)
Week 09	Data Preprocessing (Preprocessing Operation + Feature Extraction)
Week 10	Building the Model
Week 11	Model Result Evaluation
Week 12	Flask Development + Heroku
Week 13	Final Submission (Report + Code + Presentation)

2. Problem Statement

Hate speech is defined as any type of verbal, written, or behavioral communication that attacks or uses derogatory or discriminatory language against a person or group because of who they are, such as their religion, ethnicity, nationality, race, color, ancestry, sex, or another identity factor. We will walk you through a hate speech detection model using Machine Learning and Python in this challenge.

Hate Speech Detection is a sentiment categorization job. So, for training, a model that can classify hate speech from a specific piece of text may be produced by training it on sentiment classification data. As a result, we will employ Twitter tweets to identify hate speech for the job of hate speech identification model.

3. Data Collection

The data is about Twitter hate speech acquired from Kaggle [1], and it has 3 characteristics and 31962 observations. It was used to explore hate-speech detection using Twitter data. The material is divided into three categories: hate speech, offensive language, and neither. Because of the study's nature, it is necessary to mention that this dataset contains content that might be deemed racist, sexist, homophobic, or objectionable.

Total number of observations	31962
Total number of files	1
Total number of features	3
Base format of the file	CSV
Size of the data	2.95

4. Data Preprocessing

4.1 Text Cleaning

First, we clean our text because it was so messy data.

4.1.1 Lowercase

Converting a word to lower case (NLP -> nlp). Words like Racism and racism mean the same but when not converted to the lower case those two are represented as two different words in the vector space model (resulting in more dimensions). Therefore, we convert all text word into lower case letter.

4.1.2 Remove Punctuation

It is important to remove the Punctuation because is not important. Therefore, we remove that. Punctuation to do that we use regular expression.

4.1.3 Remove URLs

In this part, we remove URLs because we are working on hate speech application which detect the hate and free speech and to get the output, we need to give only text not URLs therefore, we remove the URLs because we need only clean text input.

4.1.4 Remove @tags

In this part, we remove @tags which basically used when we mentioned someone So, it’s doesn’t concern to our application therefore, we remove @tags by using regular expressions.

4.1.5Remove Special Characters

Remove Special Characters is essentially the following set of symbols [!”#$%&’()*+,-./:;<=>?@[]^_`{|}~] which basically don’t have meaning. Therefore, we remove that kind of symbols because we don’t need that. To remove we use python isalnum method.