Group Name: Group Juzheng Shi
Name: Juzheng Shi
Email: jshi3@scu.edu, frankshi1997@gmail.com
Country: The United States
College: Santa Clara University
Specialization: NLP
Project: Advance NLP—Hate Speech detection using Transformers (Deep Learning)

Problem Description:

The term hate speech is understood as any 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. In this problem, We will take you through a hate speech detection model with Machine Learning and Python.

Hate Speech Detection is generally a task of sentiment classification. So, for training, a model that can classify hate speech from a particular piece of text can be achieved by training it on data generally used to classify sentiments. Therefore, for the hate speech detection model task, We will use Twitter tweets to identify tweets containing Hate speech.

Data Cleansing and transformation

Using regex to remove URLs, Ascii codes. Lower letters, remove numbers and punctuations. Perform lemmatization on vectorized text data.