Week8 Report

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 Understanding

This is a tweet data set that is comprised of tweets from Twitter users. These tweets contain lots of hashtags, punctuations, as well as numbers. We may delete these things in our data cleansing procedure before running the machine learning model on this data set. Besides, the distribution of labels is imbalanced and this may affect the robustness of our model, so this is a problem to be solved. In order to remove the punctuation as well as hashtags, we can use popular NLP tools like NLTK to clean the text data. We can also apply tokenization and lemmatization on text data to make our analysis more convenient.