

Week7 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.

Business Understanding: This project helps evaluate users' and customers' attitudes to products and services, as simple ratings from 1 to 5 may not reveal users' actual opinions on products. By generalizing the technic of recognizing people's sentiments in their tweets, we may find a way to evaluate customers' opinions on companies' products.

Project Lifecycle along with the Deadline: We can update our database from time to time to make our model more accurate, which means this model can be used for a long time.

Data Intake Report:

Name: <Tweet hate speech>

Report date: <2022/10/18>

Internship Batch:<LISUM 13:30>

Version:<1.0>

Data intake by:<Juzheng Shi>

Data intake reviewer:<intern who reviewed the report>

Data storage location: <https://www.kaggle.com/datasets/vkrahul/twitter-hate-speech?select=train_E6oV3lV.csv>

Tabular data details:

Total number of observations	<32000>
Total number of files	<2>
Total number of features	<3>
Base format of the file	<.csv >
Size of the data	<5MB >