Group Name: Hate Speech Detection

Name	Email	Country	College/Company	Specialization
Sarah Littell	sarahlit@me.com	USA	Tulane University	NLP

Problem description:

Hate speech is any type of verbal, written and/or behavioral communication that attacks, uses derogatory/discriminatory language against a person or group based on inherent characteristics. For example: based on their religion, ethnicity, nationality, race, color, ancestry, sex or any other identity. In this problem, we will implement a hate speech detection model in order to optimize the speed at which hate speech is identified and removed from a social media platform.

Hate Speech Detection is generally a task of sentiment classification. A model that can classify hate speech from non-hate speech can be achieved by training it on data that is generally used to classify sentiments. For this task of hate speech detection, we will use pre-labeled tweets to train a ML model to identify tweets containing hate speech.

Data cleansing/transformation:

Type of data: the data comes in a csv file with 3 columns; ID, label, and the tweet itself

Problems in data: the tweet can feature useless information when determining if it’s hate speech such as uppercase letters, punctuation, and special characters

I cleaned and transformed the data by making all the tweets lowercase, removing usernames, punctuation, and special characters. I transformed the data through tokenization, removing stop words, and lemmatization. I then performed feature extraction using Tfidvectorization.

Github Repo link: https://github.com/sarahlittell/NLP_project