NLP -Project Proposal_Group 20

Project Title: Hate Speech Detection

Group-20

Team Members:

Vishnu Manoj Deepala(vishnumanojdeepala@my.unt.edu)-**UNT ID:11519490**

Seemaparvez Shaik(seemaparvezshaik@my.unt.edu)-**UNT ID:11512343**

Sai Anjali Potula (saianjalipotula@my.unt.edu)-**UNT ID-11519812**

Goal and Objectives

Motivation:

Hate speech is a difficult issue that plagues online social media. As the availability of internet information expands, so does the propagation of hate speech. We identify and investigate the obstacles that online automated systems for hate speech identification in text confront. Among these challenges include linguistic nuances, varying definitions of what constitutes hate speech, and data availability constraints for training and testing these algorithms. We present a Logistic Regression Model for detecting hate speech in Twitter messages.

Significance:

On the Internet these days, something unusual is happening. Hate speech directed at persons of a certain gender or ethnicity is prevalent on social media, and these online abominations have real-world effects, such as the spread of fear and hatred across communities. Twitter is one of the most prominent social media platforms, and it has altered the way people communicate and exchange information, ideas, and opinions. The Twitter platform, however, has been increasingly misused for the propagation of hostile and hateful content due to its dynamic, democratic, and unregulated nature. While there is no official definition of hate speech, researchers and service providers

generally agree that it is any speech that criticizes a person or a group based on a trait such as race, color, ethnicity, gender, or religion.

As a result of this phenomenon, We began to wonder if we could tackle the problem by developing a system that could detect hateful and divisive comments. Hate speech is only effective because it can instill discriminatory beliefs without being observed. What if we could devise a method for detecting hate speech that is hidden from plain view?
We decided to take on this challenge because I thought it was worth a shot.

Objectives:

We're going to employ Python-based NLP machine learning techniques to construct a hate-speech detection program. Machine learning is the process of teaching machines to do specific tasks by using data to train them.

- In this scenario, we'll utilize data from Kaggle or extract data from Twitter using a tool called "Twint".
- We'll then extract terms that indicate prominence inside hate speech using an NLP (or Natural Language Processing) approach called Tf-Idf vectorization.
- Finally, using data retrieved from Kaggle or by using "Twint", we'll train the computer to identify hate speech using a machine learning approach called logistic regression, which is commonly used for probability estimates (or any kind of data you wish to utilize for training).

Github Link:

https://github.com/Seemaparvez17/NLP-Project

Inspiration Source -:

Inspiration https://github.com/t-davidson/hate-speech-and-offensive-language