**CSCE 5290: NATURAL LANGUAGE PROCESSING**

**PROJECT INCREMENT 1**

**Project Description:**

**Introduction**

We would like to do the project on performing the targeted Sentiment analysis mainly on Hate speech classification in order to analyze the hate of particular aspects mainly concentrating on the hate speeches or the tweets that are being published on the sports or hate on sports.

**Project Proposal Description**

**Title**: To perform sentiment analysis on Twitter API data to get the sentiment analysis mainly to analyze the hate of the tweets to categorize the tweets into a neutral hate or a targeted hate. Our idea was to take the data from the twitter tweets and find out whether the tweet contains any hate in it and classify whether the hate is a targeted hate or a general neutral statement.

We wanted to mainly concentrate on the sports related tweets data because majority of the quick hates are being displayed on the sports area.

**Team Members:**

* Harin Jinakala
* Govind Naidu
* Nitin Reddy
* Lewis Edberg

**Goals and Objectives:**

**Motivation**:

The main motivation for doing this project is that, in our real day to day life if we want to get a quick update on what is happening around the world or quick information we browse to the twitter for a get quick understanding on what is happening around the world and we browse the tweets to know the information. While browsing the twitter for updates, mainly in the sports, we find some of the tweets being much targeted and find offensive.

So we wanted to build a model in order to distinguish the tweet on the amount of hate and determine whether the tweet is of a neutral statement or a targeted hate related statement.

So if we could distinguish we could eliminate or minimize the targeted hate generated on the twitter to the specific group of players in sports area.

**Significance**:

The main significance of this model is that if we have successfully achieved a completed working model then we could eliminate the amount of the hate on the players in the sports by the tweets on the twitter.

**Objectives**:

The main objective of this project is to perform the targeted sentiment analysis and perform the hate analysis. The analysis is mainly classifying the tweet hate into a general normal neutral tweet or the tweet that is representing or reflecting the targeted hate in the tweets that are being posted in the sports area. Performing this will help in removal or the reduction of the amount of targeted hate generated on the players of the sports area.

**Features**:

The main features of the project is that we use the twitter api to get the tweets and then we perform the targeted sentiment analysis to analyze the hate present in the given tweet and give the result whether the tweet is a neutral tweet or the hate specific targeted tweet.

So by classifying the tweet on the hate in the sports specific area, we will remove the controversial hate related tweets from the twitter or even we can prevent others from keeping this hate specific like tweets on the twitter itself.

**Design, Implementation, testing and deployment**

The main basic components that are required in order to perform the target sentiment analysis on the sports tweet data are

* Keras
* Tensarflow
* Tweepy
* Vader
* Seaborn

The idea of the project flow is as follows

* First the data set (in the first step with existing data set and then change to get real time data set) and clean the data for any redundencies
* Group the data in an organized format
* Using the text processing functionalities create and split the words and generate the bag of words.
* Get the sentiment by comparing the words with their sensitivity and perform the hate analysis
* Generate the hate and compare the result to predict or group and label the tweet as the neutral tweet or a hate tweet related to the sport.

**Example**: If the tweet is “even though lost or won sportsmen ship is not lost”, will be classified as the neutral tweet and if the tweet is like “Game is lost because the main player did not love his country” is classified as an hate speech

**Tasks by our team:**

Lewis – Working with tweepy

Hareen – Redundancy removal and classifications

Govind and Nitin – Integration, testing and reporting

**Progress:**

We are making a two step approach to achieve the complete working model of our project that is to predict or perform the hate analysis on the twitter data (preferably live dataset data) in two phases as,

* First we perform the basic sentiment analysis on the tweets and classify the tweets in to a neutral or positive or a negative tweet
* Then we add the targeted hate factor to the model to predict whether the given tweet is a target hated tweet or a neutral tweet
* Then we work with the twitter api to get the live dataset and then perform the hate or the targeted hate on the real data to find whether the tweet is a neutral tweet or a hate specific tweet.

**Code and Implementation:**

1. Installing vaderSentiment

**VADER (Valence Aware Dictionary and sEntiment Reasoner)** is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER uses a combination of a sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative. VADER not only tells about the Positivity and Negativity score but also tells us about how positive or negative a sentiment is.

**Scatter chart

Description automatically generated**

1. Implementing the code to find the sentiment of the airline tweets.

We used the tweeter dataset in <https://www.kaggle.com/crowdflower/twitter-airline-sentiment> for testing purposes

Text

Description automatically generated

1. Output – We will get sentiment for each tweet either positive, negative or neutral

Graphical user interface, text, application, email

Description automatically generated

**Issues/Concerns:**

We have to still implement the live tweets to recognize the hate speech. We are waiting for the approval to get the consumer key, consumer secret, access key and access secret from twitter developer. Once we get it we will start our implementation.

**References:**

<https://arxiv.org/abs/1911.00288>

<https://en.wikipedia.org/wiki/Sentiment_analysis>

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0207996>

<https://www.authoritylabs.com/what-sentiment-analysis-can-tell-us-about-our-content-and-seo/>

<https://monkeylearn.com/blog/aspect-based-sentiment-analysis/>

<https://gwu-libraries.github.io/sfm-ui/>

<https://towardsdatascience.com/detecting-hate-tweets-twitter-sentiment-analysis-780d8a82d4f6>

<https://arxiv.org/abs/2108.12521>

<https://www.upi.com/Top_News/World-News/2021/10/28/uk-prince-harry-meghan-markle-twitter-hate/7451635418014/>

<https://www.semanticscholar.org/paper/Analyzing-the-hate-and-counter-speech-accounts-on-Mathew-Kumar/e4834a6d59ef49d6c05ded2c10a4d6105b8120ad>

<https://guides.libraries.psu.edu/c.php?g=796631&p=5698003>

<https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664>

<https://towardsdatascience.com/tweepy-for-beginners-24baf21f2c25>

<https://www.loginradius.com/blog/async/beginners-guide-to-tweepy/>

<https://www.researchgate.net/post/How_to_extract_tweets_between_two_dates_using_tweepy_from_all_users_how_to_do_this_without_giving_search_keyword>

<https://www.freblogg.com/tweeting-with-python-and-tweepy>

**Github link:**

https://github.com/Harin1010/NLP-Twitter-analysis