GLA University, mathura – 2020

Sentiments Analysis

(machine learning project)

SYNOPSIS

Under the Supervision

Of

Mr. ABHISHEK VERMA

GLA UNIVERSITY

MATHURA

Efforts By:

Group-3

- **JANVI PANGORIYA (TEAM LEAD)**
- **NIDHI GUPTA**
- **KAJOL GUPTA**
- **PRANAV TOMAR**

OBJECTIVE

- The aim of this project is to develop a machine learning algorithm which allow us analyze the sentiments expressed through the tweets and classify them as racist and non racist.

- This model hence can be used to distinguish between the positive tweets and the negative tweets and prevent the chaos beforehand.

MOTIVATION

- Social media has opened a whole new world for the people around the globe. People are just a click away from the whole bunch of information. People from all age groups use social media for their source of information in different fields.

- With information comes people's opinion and with this their comes positive and negative outlook.

- Sometimes this leads to bullying and passing hate comments on someone and something which may further lead to sense of outrage among the people and may cause chaos ,so with this we could recognize those tweets beforehand and delete it from the server side of social media.

DATASET

- The dataset that we have chosen consists of training samples of tweets and labels where the label '1' denotes the tweet is racist/sexist and the label '0' denotes the tweet is not sexist/racist , our objective is to predict the labels on the given dataset.
- There are three columns:-
    - id: The id is associated with the tweets in the given dataset
    - A tweet with the label '0' is of the positive sentiment and of the label '0' is of negative sentiment.

Link of the dataset :

https://www.kaggle.com/arkhoshghalb/twitter-sentiment-analysis-hatred-speech

Technology

- Technology Implemented: Machine Learning
- Software used: Anaconda
- Language used : Python
- Frontend Design : HTML,CSS
- Backend Design : Flask
- Deployed on: Heroku

EXPECTED OUTCOME

- We will be starting our project by the data cleaning procedure where we will be removing the twitter handles as they do not contribute anything significant and then we will be removing punctuations, numbers and special characters that do not help much. And at last removing the stop words.

- Then we'll be applying a number of machine learning algorithm that can be used in text-processing and check the accuracy of the model.

- By the end we will be able to classify the tweets into the two separate classes racist or non racist and the model will work efficiently on testing set as well .