GROUP NAME: NLP_PROJECT_SNIGDHA

NAME: SNIGDHA

EMAIL: snigdha.150419@gmail.com

COUNTRY: UNITED STATES OF AMERICA

COLLEGE: UNIVERSITY OF MARYLAND BALTIMORE COUNTY

SPECIALIZATION: NLP

PROBLEM DESCRIPTION:

The term hate speech is understood as any type of 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 certain piece of text can be achieved by training it on a data that is generally used to classify sentiments. So, for the task of hate speech detection model, We will use the Twitter tweets to identify tweets containing hate speech.

DATA UNDERSTANDING AND TYPE OF DATA THAT IS UP FOR ANALYSIS:

The data involves the tweets and the labels given to each type of tweet. Basically, through this tweets, one can analyze the attitudes, behaviors, and opinions of various twitter users on different topics or viral events on social media platform.

PROBLEMS IN DATA:

Too many special characters in the tweets while no NA values.

APPROACHES:

Using some functions like regex, isalpha etc. to remove the non-alphanumeric characters from the tweets.