Data Science Intern at Data Glacier

Project: Hate Speech Detection using Transformers (Deep Learning)

Week 7: Deliverables

Name: Harshad Chormare

University: University of Mumbai

Email: chormareharshad@gmail.com

Country: India

Specialization: Data Science(NLP)

Batch Code: LISUM19

Date: 19 April 2023

Submitted to: Data Glacier

Table of Contents:

1. Project Plan

Weeks	Date	Plan
Weeks 07	April 19, 2023	Problem Statement, Data Collection, Data Report
Weeks 08	April 26, 2023	Data Preprocessing (Text Cleaning)
Weeks 09	May 2, 2023	Data Preprocessing (Preprocessing Operation + Feature Extraction)
Weeks 10	May 9,2023	Building the Model
Weeks 11	May 16,2023	Model Result Evaluation
Weeks 12	May 23,2023	Build ML APP Based on Flask
Weeks 13	May 30,2023	Final Submission (Report + Code + Presentation + Application)

2. Problem Statement

The term hate speech is understood as any type of verbal, written or behavioural 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, colour, 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.