



Data Science Internship at Data Glacier

Project: Hate Speech Detection using Transformers
(Deep Learning)

Week 7: Deliverables

Name: Amrapali Mhaigawali

University: -

Email: amrapali10@gmail.com

Country: United Kingdom

Specialization: Data Science

Batch code: LISUM14

Date: 18/11/2022

Submitted to: Data Glacier

Index

1.	Problem Statement	3
2.	Business Understanding	3
3.	Project Lifecycle	3

1. 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.

Our goal is to classify tweets into two categories, hate speech or non-hate speech. Our project analyzed a dataset CSV file from Kaggle containing 31,962 tweets.

2. Business Understanding:

Social media has experienced incredible growth over the last decade, both in its scale and importance as a form of communication. The nature of social media means that anyone can post anything they desire, putting forward any position, whether it is enlightening, repugnant or anywhere between. Depending on the forum, such posts can be visible to many millions of people. Different forums have different definitions of inappropriate content and different processes for identifying it, but the scale of the medium means that automated methods are an important part of this task. Hate-speech is an important aspect of this inappropriate content.

3. Project Lifecycle

ACTIVITY	WK-7							WK-8							WK-9						WK-10							WK-11							WK12							Wk13							
	13-Nov	14-Nov	15-Nov	16-Nov	17-Nov	18-Nov	19-Nov	20-Nov	21-Nov	22-Nov	23-Nov	24-Nov	25-Nov	26-Nov	27-Nov	28-Nov	29-Nov	30-Nov	01-Dec	02-Dec	03-Dec	04-Dec	05-Dec	06-Dec	07-Dec	08-Dec	09-Dec	10-Dec	11-Dec	12-Dec	13-Dec	14-Dec	15-Dec	16-Dec	17-Dec	18-Dec	19-Dec	20-Dec	21-Dec	22-Dec	23-Dec	24-Dec	25-Dec	26-Dec	27-Dec	28-Dec	29-Dec	30-Dec	
Problem Statement, Data Collection, Data report																																																	
Data Preprocessing																																																	
Feature Extraction																																																	
Building the model																																																	
Model Result Evaluation Flask Development and Heroku																																																	
Final Submission(Report + Code + Presentation)																																																	