Data Science Intern at Data Glacier

Project: Hate Speech Detection using Transformers (Deep Learning)

Data Intake Report

Name: Twitter Hate Speech
Report date: 06/19/2024
Internship Batch: LISUM33
Version: 1.0
Data intake by: Manhui Zhu
Data Intake reviewer: Data Glacier
Data Storage location: https://github.com/Manhui-z/Data-Glacier-Internship/tree/0083a551094656a2b96e6b2b64fd353394d34756/Week%207

Tabular data details:

Name of data	hate_speech.csv
Total number of observations	31962
Total number of features	3
Base format of the file	.csv
Size of the data	2.95 MB

Proposed Approach:

- The full dataset is consisting of 3 features: `id` with data type int64, `label` with data type int 64, and `tweet` with data type object.
- There is no missing values in the dataset.

Project Plan

Weeks	Date	Deliverables
Week 7	June 19, 2024	Problem Statement, Data Intake Report, Project Plan
Week 8	June 26, 2024	Data Preprocessing
Week 9	July 2, 2024	EDA (Exploratory Data Analysis)
Week 10	July 9, 2024	Feature Extraction
Week 11	July 16, 2024	Model Building and Training
Week 12	July 23, 2024	Model Performance Evaluation
Week 13	July 30, 2024	Final Submission (Slides + Report + Code)

Problem Statement

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. 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. For the task of hate speech detection model, we will use the Twitter tweets to identify tweets containing Hate speech.