



# **Data Science Intern at Data Glacier**

**Project:** Hate Speech Detection Using Transformers (Deep Learning)

**Week 8:** Project Deliverable 2

**Team Name:** Fibo

**Team Members:** Mahyar Arani

**Email:** [arani.mahyar@gmail.com](mailto:arani.mahyar@gmail.com)

**Batch code:** LISUM19

**Submission date:** Apr 19, 2023

**Submitted to:** Data Glacier

## **Problem Description:**

The problem at hand is detecting hate speech in Twitter tweets using a machine learning model. Hate speech is a type of communication that attacks or uses derogatory or discriminatory language against a person or group based on their religion, ethnicity, nationality, race, color, ancestry, sex, or other identity factors. The goal is to develop a model that can accurately classify whether a tweet contains hate speech or not.

## **Data Understanding**

The provided dataset contains information related to tweets and their associated labels. There are a total of 32,000 tweets in the dataset, each identified by a unique ID. The tweets have been labeled based on their sentiment, with 0 indicating a negative sentiment and 1 indicating a positive sentiment.

The dataset is further divided into 20 bins based on the length of the tweets, ranging from 1,599 characters to 31,962 characters. Each bin contains 1,598 or 1,599 tweets, and the labels are distributed almost equally among the bins.

The majority of the tweets (29,720) have been labeled with a sentiment score between 0 and 0.05, while only 2,242 tweets have been labeled with a sentiment score between 0.95 and 1.00. There are 29530 unique tweets in the dataset.

Overall, this dataset seems to be suitable for sentiment analysis tasks, particularly for classification of tweets based on their sentiment scores. However, further analysis and preprocessing may be necessary to ensure the quality and consistency of the data.

## **Data Preprocessing**

### **Sentence and word Tokenization**

The text is sentence tokenized to perform sentiment analysis. The sentiment scores and labels is then used as a feature to detect hate speech.

### **Punctuation Removal**

Punctuation is removed to improve modeling and reduce noise in the data.

### **URL Removal**

As part of preprocessing, URLs are removed to better model the analysis. URLs do not provide any relevant information for detecting hate speech and can add noise to the training set.

### **Special Character Removal**

As part of preprocessing, mentions, hashtags, and other special characters and symbols are removed from the text to improve modeling.

## Project Lifecycle

<b>Weeks</b>	<b>Date</b>	<b>plan</b>
Weeks 07	Apr 19, 2022	Problem Statement, Data Collection, Data Report
Weeks 08	Apr 26, 2022	Data Preprocessing
Weeks 09	May 2, 2022	Feature Extraction
Weeks 10	May 9, 2022	Building the Model
Weeks 11	May 16, 2022	Model Evaluation
Weeks 12	May 23, 2022	Flask Development + Heroku
Weeks 13	May 30, 2022	Final Submission (Report + Code + Presentation)