

# Final Project – Data Glacier Virtual Internship 2021 Project

## Title: NLP - Twitter Hate Speech Detection with Transformer (Deep Learning)

---

**Name:** Ankita Manna

**E-mail:** ankitamanna1@gmail.com

**Country:** UK

**Specialization:** Natural Language Processing

---

**Problem Statement:** Hate Speeches are taking over every media, social platform etc., 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. Designing a model to detect such speeches or posts on these platforms are getting complicated these days with an increase in the usage of different languages.

---

**Data cleaning and Featurization:** Data cleansing has been done using the BeautifulSoup and regex libraries, eliminating stopwords and doing stemming. Removed unnecessary words, symbols, and HTML tags. Also, I lowered all texts using the regex library.