Data Glacier Internship Project
Batch LISUM36: 30 July – 30 Oct 24
Project: Advance NLP: Hate Speech detection using Transformers (Deep Learning) - Group Project

Team:
Team Name: Team Trailblazers

Members:

Team member one:	Team member two:	Team member three:
Michael Udonna Egbuzobi egbuzobi.michael@gmail.com United Kingdom University of Wolverhampton Data Science	**Nweke Nonye** nonyenweke22@gmail.com United Kingdom University of Wolverhampton Data Science	**Sreedhar Rongala** rongalasreedhar@gmail.com Italy University of Naples Federico ii Data science

Problem Description:

Hate speech is a form of communication that uses derogatory language to attack or discriminate against individuals based on aspects like religion, ethnicity, nationality, race, colour, ancestry, or other identity factors. Detecting hate speech online is crucial for maintaining healthy social interactions, particularly on platforms like Twitter, where information spreads quickly. The aim of this project is to develop an advanced hate speech detection model using transformer-based deep learning architectures. The model will classify text (tweets) into hate speech or non-hate speech (binary classification).

Business Understanding:

Hate speech on social media can lead to significant harm, including inciting violence, increasing polarization, and damaging the mental well-being of targeted individuals or groups. For platforms like Twitter, where users can post millions of messages daily, automated hate speech detection is essential to moderate content, ensure compliance with platform policies, and create a safe online environment. A robust model that can accurately classify tweets as hate speech or non-hate speech enables platforms to take proactive measures, reducing harmful content while maintaining free speech.

This project aims to provide an effective solution to this problem by leveraging deep learning techniques, specifically transformer models, which have proven to be highly effective in natural language processing tasks. The solution can be integrated into content moderation systems to enhance efficiency and scalability. It offers value not only to social media platforms but also to policymakers, advocacy groups, and businesses concerned with online safety and reputation management.

Project lifecycle

Six weeks.