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:
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

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).

Data Cleansing and Transformation:

Techniques used for data cleansing:

- **Handling NA values**: We observed no NA values in the dataset during the initial inspection.
- **Outlier Handling:** Since the dataset is primarily textual, traditional outlier detection methods are not applicable. However, we reviewed tweet lengths to identify any extreme cases, such as unusually short or long tweets, that could potentially affect model performance. No formal outliers were detected, and as a result, no truncation or padding of tweets was necessary.
- **Class Imbalance:** Michael addressed the class imbalance using SMOTE (Synthetic Minority Over-sampling Technique) to generate synthetic samples for the minority class, while Nonye handled the imbalance by applying class weights during model training to give higher importance to the minority class.

NLP Featurization Techniques:

- **Featurization**: We applied several techniques to convert raw text into numeric features for model input:
    - **BERT Tokenizer**: For the deep learning model, we used the BERT tokenizer to break the tweet text into tokens, which were then fed into a transformer model.

- **Data Cleaning**: The following methods were applied to clean the tweet text data:
    - **Regular Expressions (Regex)**:
        - Removal of URLs, mentions (@user), special characters, and non-alphanumeric characters.
        - Conversion to lowercase and removal of stop words.

Collaboration and Code Merging:

All code contributions were merged into a single codebase to create a unified NLP pipeline. We combined the best practices from each member's work, resulting in a robust pre-processing pipeline and a model training pipeline for hate speech detection.