**Experimentation with Three Different NLP Models (Encoder Models Only)**

**Objective:**  
The goal of this is to experiment with three different **encoder-only NLP models**, comparing their performances on the same dataset. **You are not allowed to use models that are already fine-tuned on emotion datasets.** The experiments will focus on data preparation, model fine-tuning, and performance evaluation to gain insights into the effectiveness of various encoder models for a specific NLP task.

**Dataset:**  
Use the dataset Tweet Emotion.

**Key Guidelines:**

* **Encoder-Only Models:** You are restricted to using encoder models only and cannot use models already fine-tuned on emotion datasets.
* **Performance Metrics:** Use consistent evaluation metrics (see the in-class Kaggle Competition).
* **Accounting for Imbalance:** Account for the imbalance to improve your results.

**Experiments:**

**Experiment 1: RoBERTa Base**

* **Model:** roberta-base

**Experiment 2: DistilBERT**

* **Model:** distilbert-base-uncased

**Experiment 3: Similar-Sized Model**

* **Model:** To Choose a model of similar size and architecture to DistilBERT, such as distilroberta-base or albert-base-v2. Ensure the selected model is comparable in size and complexity to distilbert.
* **Here I chose FLAN-**T5