**Assignment 1**

Deadline:

11-August-2024 Sunday 11 :59 pm

**Q1: Hugging Face**

a. Objective:

I. To familiarize yourself with Hugging Face and implement a simple project using pre-

trained models from the Hugging Face library.

**Steps to be completed:**

**a. Part 0:** Project Explanation

*I. Project Idea:*

SDG goal is Quality Education.

Project Title: distilbert-base-multilingual-cased-sentiments-student.

*II. Project Explanation:*

The aim of the project is to develop a model for analyzing sentiment in multilingual textual data relating to health and well-being.

This model will identify positive, negative or neutral sentiments, which can be used to measure public opinion, track changes in sentiment over time or identify areas of concern.

**b. Part 1:**

*I. Hugging Face Transformers Library Research:*

Learn more about the Hugging Face Transformers library, including key features such as the ability to easily load pre-trained models, transfer learning and support for a wide range of NLP tasks.

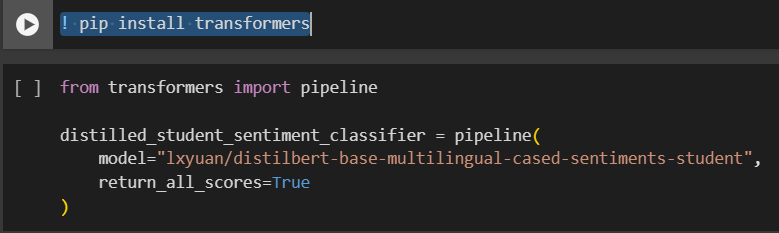
*II. Model Selection:*

After analyzing several models on Hugging Face, I chose the “lxyuan/distilbert-base-multilingual-cased-sentiments-student” model, which is a smaller and faster version of BERT, DistilBERT, adapted to sentiment analysis in several languages.

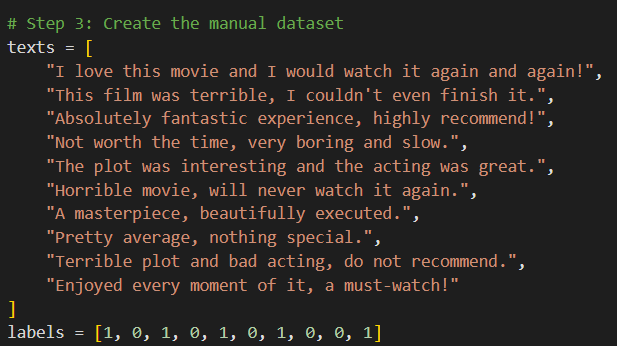
**c. Part 2:**

*I. Loading the Model :*

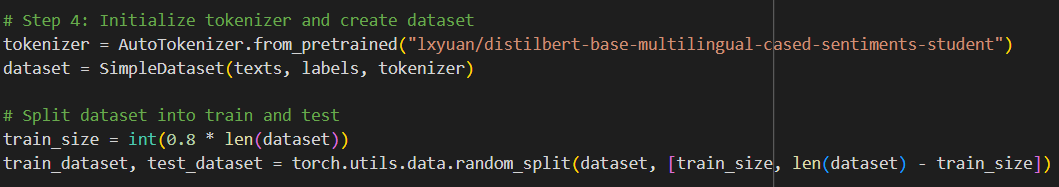
* Load the lxyuan/distilbert-base-multilingual-cased-sentiments-student model using the Hugging Face library in your Python environment.

****

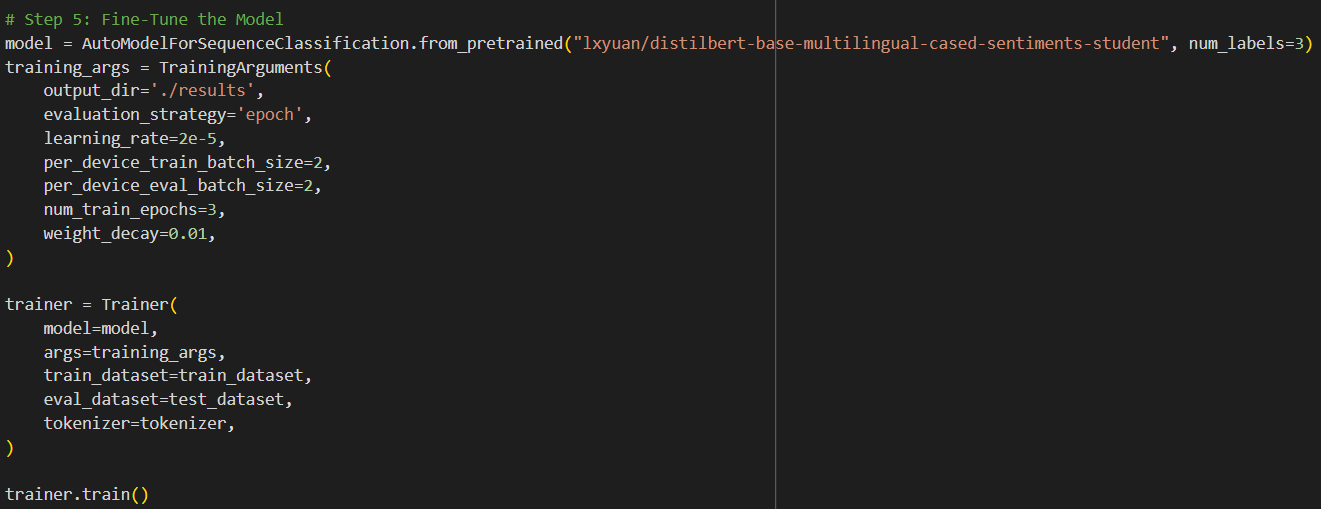
*II. Data Preparation:*



*IV. Test the model before fine-tuning.*



*V. Preprocess and Fine Tuning the model.*



**d. Part 3:**

*I. Evaluation*

**Performance Before Fine-Tuning:**

* **English Sentiment Analysis:**
  + Positive: 71.80%
  + Neutral: 17.99%
  + Negative: 10.20%
* **Malay Sentiment Analysis:**
  + Positive: 84.95%
  + Neutral: 11.80%
  + Negative: 3.24%
* **Japanese Sentiment Analysis:**
  + Positive: 77.95%
  + Neutral: 14.07%
  + Negative: 7.99%

**Performance After Fine-Tuning:**

* **English Sentiment Analysis:**
  + Positive: 97.54%
  + Neutral: 1.56%
  + Negative: 0.90%
* **Malay Sentiment Analysis:**
  + Positive: 97.60%
  + Neutral: 1.80%
  + Negative: 0.59%
* **Japanese Sentiment Analysis:**
  + Positive: 93.42%
  + Neutral: 4.02%
  + Negative: 2.56%

**Performance Evaluation:**

* **Improvement in Positive Sentiment Detection:**
  + **English:** Increased from 71.80% to 97.54%
  + **Malay:** Increased from 84.95% to 97.60%
  + **Japanese:** Increased from 77.95% to 93.42%
* **Reduction in Neutral and Negative Sentiment Scores:**
  + Fine-tuning significantly reduced the neutral and negative scores across all languages, indicating a better differentiation between positive and non-positive sentiments.