Assignment 2 : Sentiment Classification Model using Transformers

Objective:

In this assignment, you'll modify the provided Colab notebook to use a new dataset (IMDB) instead of the Twitter dataset. You will then deploy your Sentiment Classification app using **Streamlit** and submit your work as per the guidelines below.

Instructions:

1. **Access the Colab File**:
   - Here is the link to the Colab file:
     [Param_Soni_Fine_Tune_BERT_for_Multi_Class_Sentiment_Classification_with_Twitter_Data_|_Python_&_Streamlit.ipynb](make a copy)
   - The Colab notebook contains all the necessary code for a sentiment classification app using a Twitter dataset.
2. **Change the Dataset**:
   - Replace the **Twitter dataset** in the Colab notebook with the **IMDB dataset** provided below:
     - **IMDB Dataset Link**:
       [https://drive.google.com/file/d/1Xm3HWAnrHjLmvPTMVIMW4q624SBr_SCH/view?usp=sharing]
   - Ensure that the new dataset is correctly loaded, preprocessed, and used in the model training and evaluation process.
3. **Host the App on Streamlit**:
   - Use **Streamlit** to deploy your app.
   - Follow the deployment instructions in the Colab file.
   - Ensure that your app works seamlessly with the IMDB dataset for **sentiment classification**.
4. **Optional (Recommended)**:
   - Use **VS Code** (or another local IDE) to edit the notebook and save the files locally. This will make it easier to manage files and configurations.
   - After ensuring the app works locally, deploy it using Streamlit.

Submission Guidelines:

1. **GitHub Repository**:
   - Push your modified code to a GitHub repository.
   - Share the link to your GitHub repository in the submission form.
2. **Screen Recording**:

- ○ Record a short video (screen recording) demonstrating your app's working functionality. Ensure the recording shows the following:
  - ■ App loading on Streamlit.
  - ■ At least 2-3 examples of sentiment classification results for different inputs.
3. **Submission Form**:
   - ○ Submit the following items using the provided Google Form:
     - ■ Link to your GitHub repository.
     - ■ Link or upload of your screen recording.
   - ○ **Google Form Link**:
     [https://docs.google.com/forms/d/e/1FAIpQLSe_qfqgIZdEqCTDLlUOqpaaU4Qllu5Pqz2WCPWvnujhtRXPNQ/viewform?usp=dialog]

Deadline:

- **Sunday EOD** (Exact Date: [19 Jan 2025]).

Grading Criteria:

1. Successful implementation of the IMDB dataset in the notebook.
2. Functional and correctly deployed Streamlit app.
3. Proper submission via Google Form (GitHub link and working app demonstration video).

Resources:

- **Colab File**:Param_Soni_Fine_Tune_BERT_for_Multi_Class_Sentiment_Classification_with_Twitter_Data_|_Python_&_Streamlit.ipynb
- **IMDB Dataset**:
  https://drive.google.com/file/d/1Xm3HWAnrHjLmvPTMVIMW4q624SBr_SCH/view?usp=sharing
- **Streamlit Documentation**: https://docs.streamlit.io/

Important Notes:

- All the base code is already provided; your task is to **replace the dataset**, change the labels as the IMDB dataset only has two/three classification i.e positive/negative.
- If you face any issues, feel free to try running the code locally using an IDE like **VS Code**, which will also save your work.

- Late submissions will not be accepted unless prior approval is granted.

---

2. Optional

Complete the given model for twitter tweets multiclass sentiment (seven emotions) classification and host on streamlit. (The given data has everything completed for this project, you just have to run the code for this task2)