Assignment Instructions: Experimentation with Three Different Models

In this assignment, you will conduct three experiments, each using a distinct language model. The objective is to understand and compare the performance of each model on the same dataset. For this HW, you will use the same dataset as in the previous HW. Here are the specific instructions for each experiment:

Experiment 1: Using DistilBERT
Model: Utilize the "distilbert-base-uncased" model.
Task: Load and prepare the "distilbert-base-uncased" model for training on the provided dataset. After training, evaluate its performance to gather insights on its effectiveness.

Experiment 2: Choosing a Similar-Sized Model
Model Suggestions: Choose a model that is similar in size and architecture to DistilBERT. Examples include "distilroberta-base" or models like "albert-base-v2". You are not limited to these suggestions but ensure the chosen model is of a comparable size.
Task: After selecting a suitable model, replicate the process you followed in Experiment 1. Train the chosen model on the same dataset and evaluate its performance.

Experiment 3: Using FLAN-T5
Model: Employ "google/flan-t5-base" for this experiment.
Task: With FLAN-T5, you will need to adapt your dataset to match the input expectations of T5 models. Train "google/flan-t5-base" on the dataset, then assess and compare its performance against the other models.

For each experiment, you should:

- Data Preparation: Appropriately format and preprocess your dataset for each specific model, considering any unique input requirements.
- Model Training: Conduct fine-tuning on the dataset, ensuring you track the training process for effectiveness and potential issues.
- Performance Evaluation: After training, evaluate each model using relevant metrics to determine their performance levels. Compare these outcomes to analyze the advantages or limitations of each model concerning your task.
- Documentation: Thoroughly document your methodologies, results, and observations. Your documentation should detail the performance metrics, any encountered challenges, and your interpretative conclusions from the experiments.

Submission Requirements:

- Jupyter Notebook(s): Submit the Jupyter notebooks containing all code, comments, and results for your experiments. The notebooks should be clear, well-commented, and reproducible.
- Report: Provide a detailed report summarizing your findings from each experiment. This report should compare the models' performances, discuss any significant observations, and include your conclusions. Your report should have link to W&B Project: Ensure your

- Weights & Biases project is public. This project should document all your experiments and results

By conducting these experiments and documenting your findings, you will gain valuable insights into the comparative effectiveness of different NLP models for a specific task. This will deepen your understanding of model selection and performance evaluation in the field of NLP.