JHANVI PAREKH

60009210033

Experiment No 9

Aim: To fine-tune a BERT model on custom data for performing a question-answering task.

Introduction:

Question Answering (QA) systems aim to provide precise answers to questions posed in natural language. Fine-tuning BERT (Bidirectional Encoder Representations from Transformers) on custom datasets allows for developing QA models tailored to specific domains. By training on relevant question-answer pairs, the model learns to locate and provide precise answers based on context.

BERT is especially effective for this task due to its ability to understand contextual embedding's, making it well-suited for handling the complexities of language. Fine-tuning involves training the model on a labeled dataset where each entry includes a passage, a question, and the corresponding answer span within the passage.

Fine-Tuning BERT for Question Answering

1. **Model Selection**: Use the pre-trained BERT model as the base, available through Hugging Face's Transformers library. Models such as bert-large-uncased are commonly used for QA tasks.
2. **Data Preparation**: Format the custom data into question-passage-answer triplets. The passage contains the context, and the answer is a span within this context.
3. **Tokenization**: Tokenize the data using the BERT tokenizer, which splits text into tokens compatible with BERT's vocabulary. Additionally, align each token with its position in the original text to aid in identifying the answer span.
4. **Fine-Tuning Process**: Train the BERT model on the formatted dataset, optimizing it to locate the start and end tokens of the answer span within the context passage.

Lab Experiment

Step 1: Install required libraries.

Step 2: Load a pre-trained BERT model for QA and tokenizer.

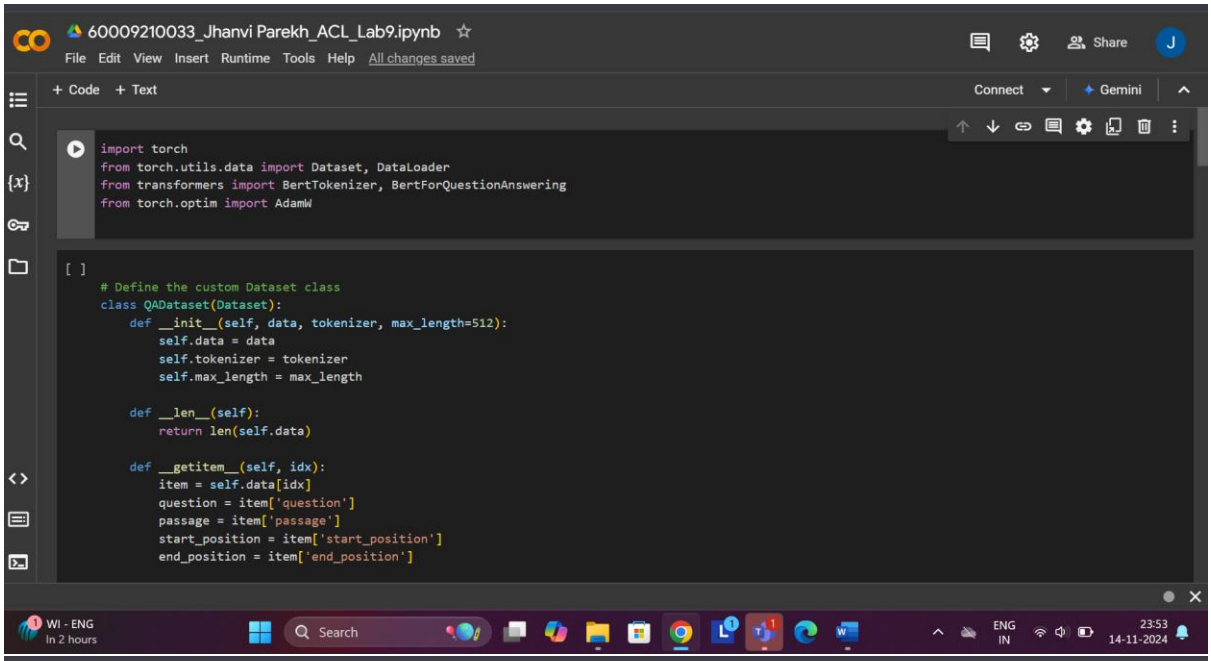
Step 3: Prepare the custom data in question-passage-answer format, ensuring each answer is marked with its start and end positions in the passage.

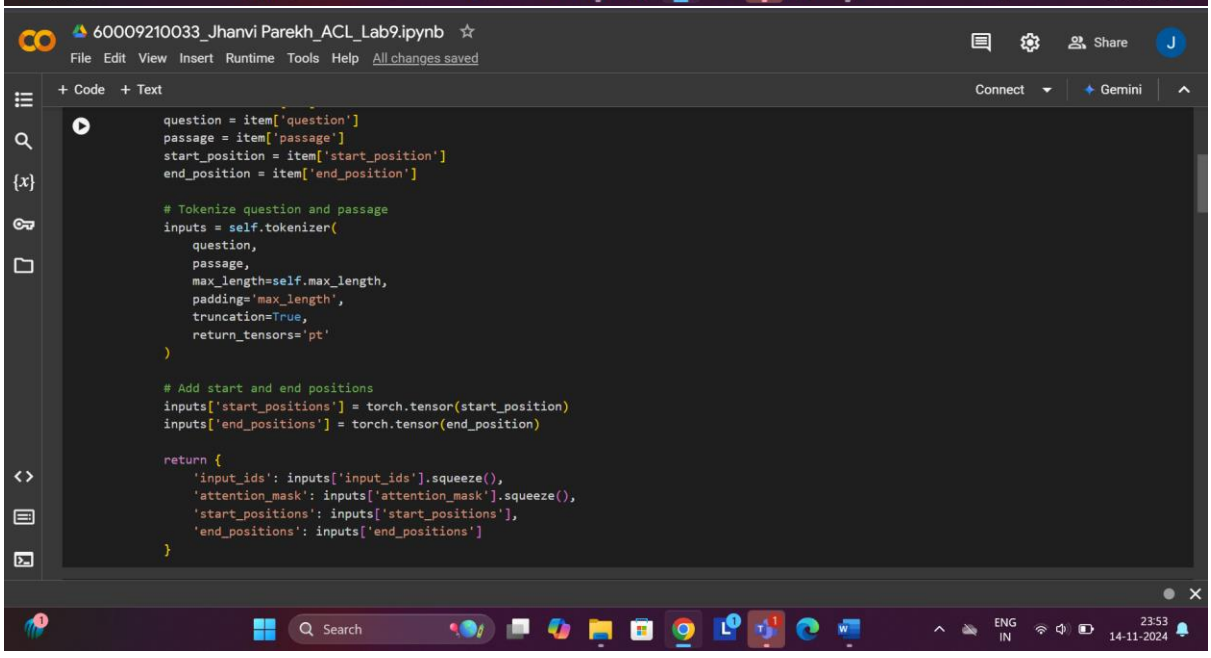
Step 4: Tokenize the data and format it for BERT.

Step 5: Fine-tune the model.

Step 6: Test the model with new questions and passages.

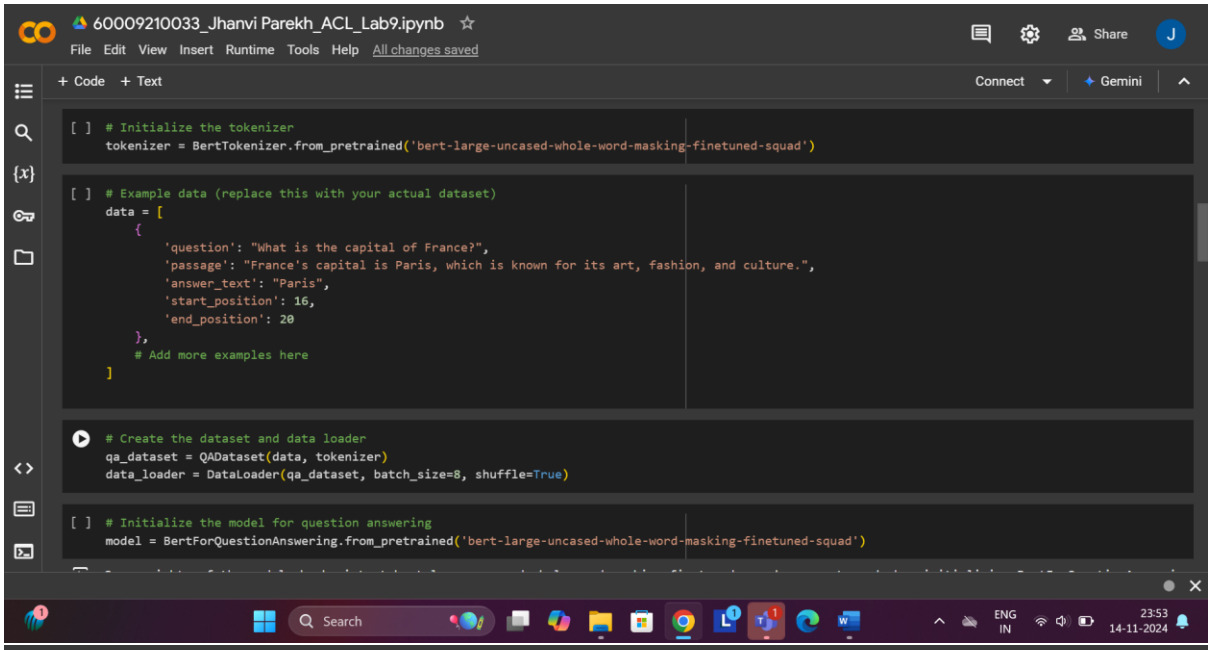
Link: https://colab.research.google.com/drive/17BYqWlPkU82cjDPiBc4IvQqOXtnBVsrf?usp=sharing

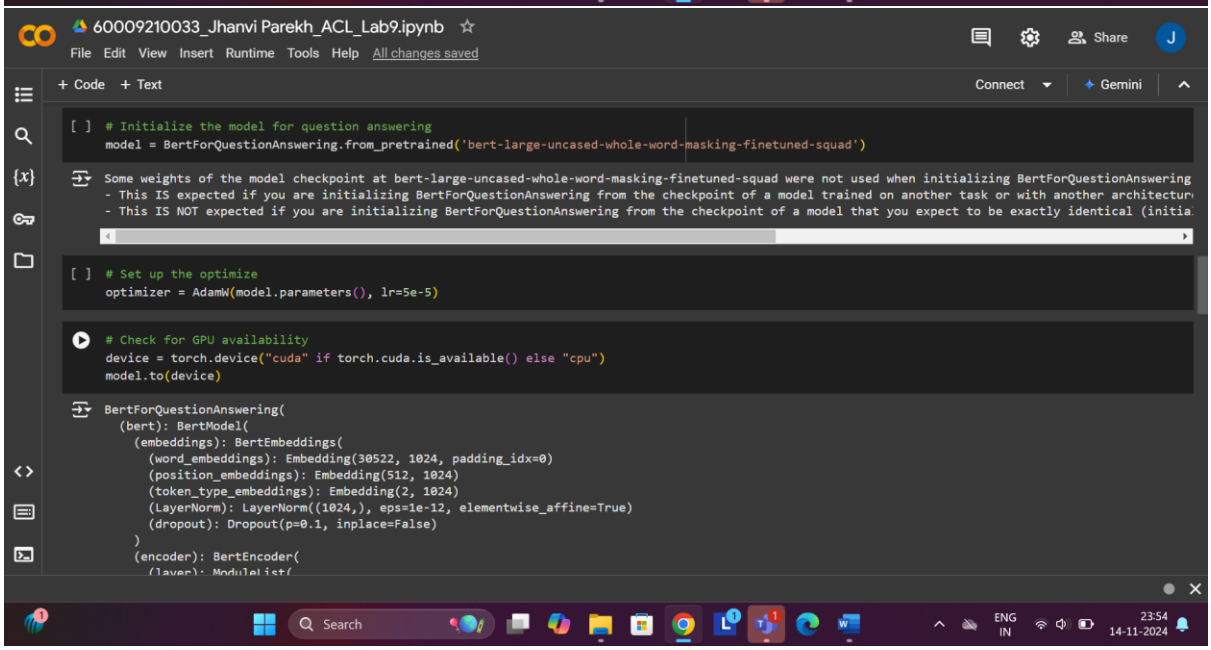




Department of Computer Science and Engineering (Data Science)
Lab Manual

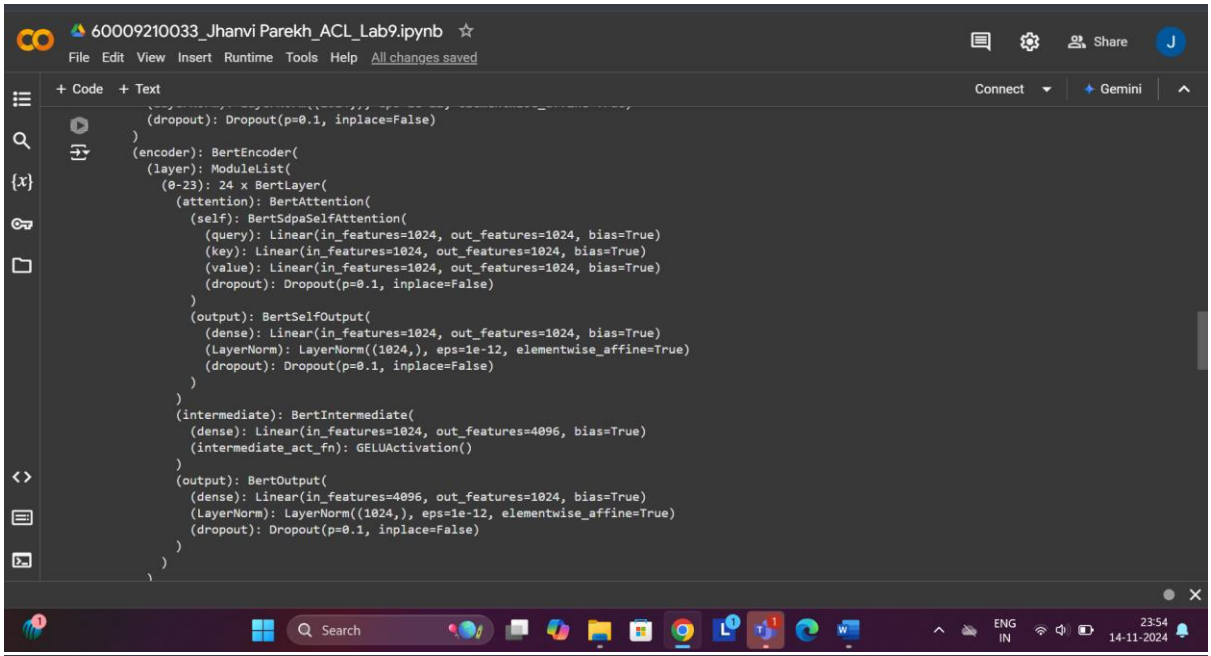
Sub: Advanced Computational Linguistics **Year/Sem: BTech/VII**

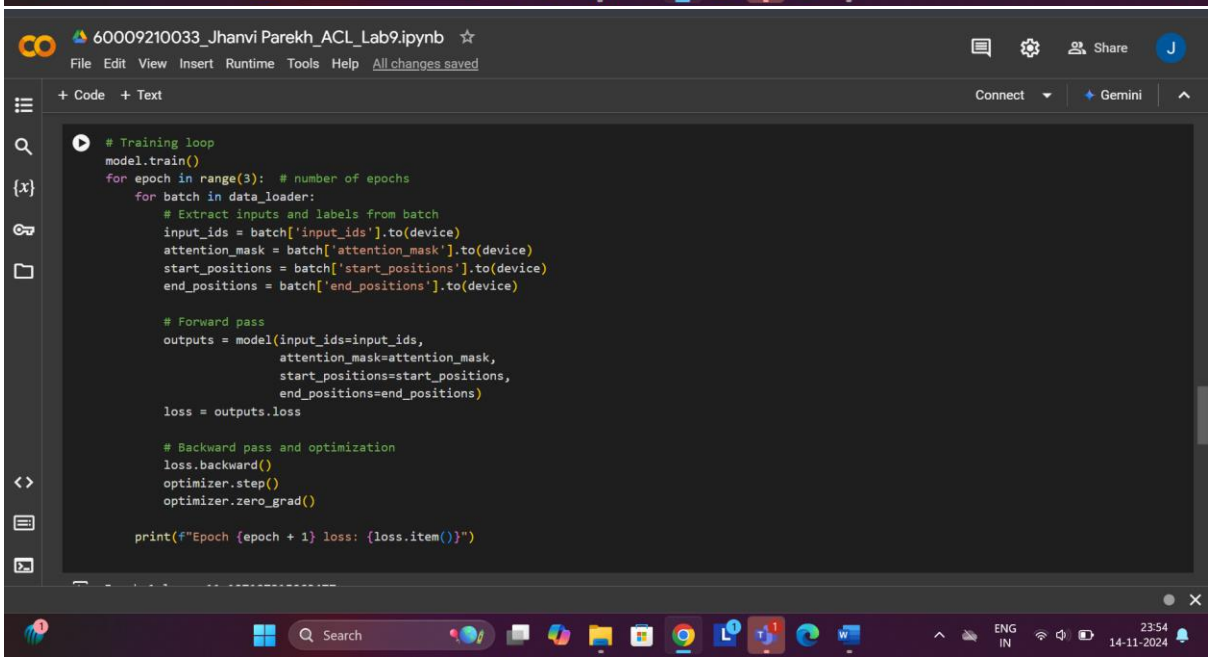




Department of Computer Science and Engineering (Data Science)
Lab Manual

Sub: Advanced Computational Linguistics **Year/Sem: BTech/VII**





Department of Computer Science and Engineering (Data Science)
Lab Manual

Sub: Advanced Computational Linguistics **Year/Sem: BTech/VII**