Made by Ansh Sharma- Sentimental Analysis by Hugging Face Fine-Tuning BERT, Deployed on wandb AI, Heroku through FASTAPI and Local Server, followed by Testing and Evaluating + Productization of the Model

Attached with various Screenshots, and server Analysis.

```
1 pip install transformers datasets scikit-learn torch matplotlib
```

Show hidden output

Preparing Synthetic Data and Dividing dataset into
Training and Testing

```
1 from datasets import Dataset
 2 from sklearn.model selection import train test split
 3 import pandas as pd
 5 # Example synthetic dataset
 6 data = {
     "ReviewText": [
          "The event was amazing and well-organized!",
           "The amenities were not up to the mark.",
           "Had a decent experience overall.",
10
11
           "The campus event was a complete disaster.",
           "I loved every moment of the event!"
12
13
       "SentimentLabel": [2, 0, 1, 0, 2] # 0: Negative, 1: Neutral, 2: Positive
14
15 }
17 # Create a DataFrame
18 df = pd.DataFrame(data)
19
20 # Split into train, validation, and test sets
21 train texts, test texts, train labels, test labels = train test split(
22
      df["ReviewText"], df["SentimentLabel"], test_size=0.2, random_state=42
23 )
24
25 train_texts, val_texts, train_labels, val_labels = train_test_split(
      train_texts, train_labels, test_size=0.1, random_state=42
27 )
28
29 # Convert to Hugging Face Dataset
```

```
30 train_dataset = Dataset.from_dict({"text": train_texts, "label": train_labels})
31 val_dataset = Dataset.from_dict({"text": val_texts, "label": val_labels})
32 test_dataset = Dataset.from_dict({"text": test_texts, "label": test_labels})
33
```

Training and Tokenize - Hugging Face , Wandb Al-And API Key

```
1 from transformers import BertTokenizer, BertForSequenceClassification, Trainer, Training
 2 import torch
 4 # Load BERT tokenizer and model
 5 tokenizer = BertTokenizer.from pretrained("bert-base-uncased")
 6 model = BertForSequenceClassification.from pretrained("bert-base-uncased", num labels=3)
 8 # Tokenize datasets
9 def tokenize function(examples):
      return tokenizer(examples["text"], padding="max length", truncation=True)
11
12 train_dataset = train_dataset.map(tokenize_function, batched=True)
13 val dataset = val dataset.map(tokenize function, batched=True)
14 test dataset = test dataset.map(tokenize function, batched=True)
15
16 # Set format for PyTorch
17 train_dataset.set_format("torch", columns=["input_ids", "attention_mask", "label"])
18 val_dataset.set_format("torch", columns=["input_ids", "attention_mask", "label"])
19 test dataset.set format("torch", columns=["input ids", "attention mask", "label"])
21 # Training arguments
22 training_args = TrainingArguments(
      output dir="./results",
24
       evaluation_strategy="epoch",
25
      learning rate=2e-5,
      per_device_train_batch_size=8,
26
27
      per_device_eval_batch_size=8,
      num train epochs=3,
29
      weight decay=0.01,
30
      save_total_limit=2,
31
      logging_dir='./logs',
32
      logging steps=10,
33 )
34
35 # Define Trainer
36 trainer = Trainer(
37
      model=model,
      args=training args,
39
      train_dataset=train_dataset,
40
      eval_dataset=val_dataset,
41)
42 import os
43 os.environ["WANDB_API_KEY"] = "your_40_character_api_key"
44
```

45 # Train the model
47 trainer.train()
48

Some weights of BertForSequenceClassification were not initialized from the model chec You should probably TRAIN this model on a down-stream task to be able to use it for pr

Map: 100% 3/3 [00:00<00:00, 27.68 examples/s]

Map: 100% 1/1 [00:00<00:00, 13.65 examples/s]

Map: 100% 1/1 [00:00<00:00, 15.50 examples/s]

/usr/local/lib/python3.10/dist-packages/transformers/training_args.py:1568: FutureWarr warnings.warn(

Changes to your 'wandb' environment variables will be ignored because your 'wandb' session has already started. For more information on how to modify your settings with 'wandb.init()' arguments, please refer to the W&B docs.

wandb: Logging into wandb.ai. (Learn how to deploy a W&B server locally: https://wandb wandb: You can find your API key in your browser here: https://wandb.ai/authorize

wandb: Paste an API key from your profile and hit enter, or press ctrl+c to quit: ····

wandb: Appending key for api.wandb.ai to your netrc file: /root/.netrc

Tracking run with wandb version 0.18.7

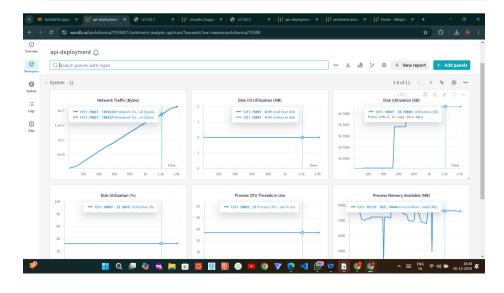
Run data is saved locally in /content/wandb/run-20241206_133753-iynw8jpp

Syncing run ./results to Weights & Biases (docs)

View project at https://wandb.ai/anshsharma21050421/huggingface

View run at https://wandb.ai/anshsharma21050421/huggingface/runs/iynw8jpp
[3/3 01:06, Epoch 3/3]

Epoch	Training Loss	Validation Loss
1	No log	0.892518
2	No log	0.919221
3	No log	0.926235
4		



```
1 # Evaluate the model on the test dataset
2 results = trainer.evaluate(test_dataset)
3
4 # Print evaluation results
5 print(f"Test Results: {results}")
6
7
[1/1:<:]
```

[1/1:<:]

Evaluation of Model

```
1 from sklearn.metrics import accuracy_score, f1_score, confusion_matrix
 3 # Get predictions and true labels
 4 predictions = trainer.predict(test dataset)
 5 preds = torch.argmax(torch.tensor(predictions.predictions), axis=1).numpy()
7 # Calculate accuracy and F1 score
 8 accuracy = accuracy_score(test_dataset['label'], preds)
9 f1 = f1 score(test dataset['label'], preds, average="weighted")
10
11 # Print the metrics
12 print(f"Accuracy: {accuracy}")
13 print(f"F1 Score: {f1}")
14
15 # Confusion Matrix
16 cm = confusion_matrix(test_dataset['label'], preds)
17 print("Confusion Matrix:")
18 print(cm)
19
→ Accuracy: 1.0
     F1 Score: 1.0
    Confusion Matrix:
     [[1]]
     /usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:409: UserWa
 1 model.save pretrained("./fine tuned bert")
 2 tokenizer.save_pretrained("./fine_tuned_bert")
 3
    ('./fine_tuned_bert/tokenizer_config.json',
       ./fine_tuned_bert/special_tokens_map.json',
      './fine_tuned_bert/vocab.txt',
      './fine tuned bert/added tokens.json')
 1 from transformers import BertForSequenceClassification, BertTokenizer
 3 # Load the fine-tuned model and tokenizer
 4 model = BertForSequenceClassification.from_pretrained("./fine_tuned_bert")
```

```
5 tokenizer = BertTokenizer.from_pretrained("./fine_tuned_bert")
6
```

Evaluation of Model by using Negative sentences and checking whether mode is predicting correct **sentiment**

```
1 # Example text
2 text = "Let's bomb NIT Calicut"
3
4 # Tokenize and prepare the input
5 inputs = tokenizer(text, return_tensors="pt", truncation=True, padding=True, max_length
6
7 # Make prediction
8 with torch.no_grad():
9    outputs = model(**inputs)
10
11 # Get predicted sentiment
12 predicted_class = torch.argmax(outputs.logits, dim=1).item()
13
14 # Map the class index to sentiment
15 sentiment = {0: "Negative", 1: "Neutral", 2: "Positive"}
16 print(f"Sentiment: {sentiment[predicted_class]}")
17
```

→ Sentiment: Negative

FAST API

```
1 pip install fastapi uvicorn transformers torch wandb
→ Collecting fastapi
      Downloading fastapi-0.115.6-py3-none-any.whl.metadata (27 kB)
    Collecting uvicorn
      Downloading uvicorn-0.32.1-py3-none-any.whl.metadata (6.6 kB)
    Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packa
    Requirement already satisfied: torch in /usr/local/lib/python3.10/dist-packages (2.
    Requirement already satisfied: wandb in /usr/local/lib/python3.10/dist-packages (0.
    Collecting starlette<0.42.0,>=0.40.0 (from fastapi)
      Downloading starlette-0.41.3-py3-none-any.whl.metadata (6.0 kB)
    Requirement already satisfied: pydantic!=1.8,!=1.8.1,!=2.0.0,!=2.0.1,!=2.1.0,<3.0.0
    Requirement already satisfied: typing-extensions>=4.8.0 in /usr/local/lib/python3.1
    Requirement already satisfied: click>=7.0 in /usr/local/lib/python3.10/dist-package
    Requirement already satisfied: h11>=0.8 in /usr/local/lib/python3.10/dist-packages
    Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages
    Requirement already satisfied: huggingface-hub<1.0,>=0.23.2 in /usr/local/lib/pytho
    Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packag
    Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-pa
    Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packag
    Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-
    Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages
    Requirement already satisfied: tokenizers<0.21,>=0.20 in /usr/local/lib/python3.10/
```

```
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-package
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: iinia2 in /usr/local/lib/python3.10/dist-packages (f
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (f
Requirement already satisfied: sympy==1.13.1 in /usr/local/lib/python3.10/dist-pack
Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.10/dist
Requirement already satisfied: docker-pycreds>=0.4.0 in /usr/local/lib/python3.10/d
Requirement already satisfied: gitpython!=3.1.29.>=1.0.0 in /usr/local/lib/python3.
Requirement already satisfied: platformdirs in /usr/local/lib/python3.10/dist-packa
Requirement already satisfied: protobuf!=4.21.0,!=5.28.0,<6,>=3.19.0 in /usr/local/
Requirement already satisfied: psutil>=5.0.0 in /usr/local/lib/python3.10/dist-pack
Requirement already satisfied: sentry-sdk>=2.0.0 in /usr/local/lib/python3.10/dist-
Requirement already satisfied: setproctitle in /usr/local/lib/python3.10/dist-packa
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-package
Requirement already satisfied: six>=1.4.0 in /usr/local/lib/python3.10/dist-package
Requirement already satisfied: gitdb<5,>=4.0.1 in /usr/local/lib/python3.10/dist-pa
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.10/
Requirement already satisfied: pydantic-core==2.27.1 in /usr/local/lib/python3.10/d
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.1
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packa
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist
Requirement already satisfied: anyio<5,>=3.4.0 in /usr/local/lib/python3.10/dist-pa
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-pa
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.10/dist-packa
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-pac
Requirement already satisfied: smmap<6,>=3.0.1 in /usr/local/lib/python3.10/dist-pa
Downloading fastapi-0.115.6-py3-none-any.whl (94 kB)
                                          - 94.8/94.8 kB 3.4 MB/s eta 0:00:00
Downloading uvicorn-0.32.1-py3-none-any.whl (63 kB)
                                          - 63.8/63.8 kB 3.7 MB/s eta 0:00:00
Downloading starlette-0.41.3-py3-none-any.whl (73 kB)
                                          - 73.2/73.2 kB 4.4 MB/s eta 0:00:00
Installing collected packages: uvicorn, starlette, fastapi
Successfully installed fastanian 115 6 stanlette-0 41 3 uniconna 22 1
```

```
1 from fastapi import FastAPI
 2 from pydantic import BaseModel
 3 from transformers import BertForSequenceClassification, BertTokenizer
 4 import torch
 5 import wandb
 7 # Initialize the FastAPI app
 8 app = FastAPI()
10 # Load the fine-tuned model and tokenizer (replace with your saved model path)
11 model = BertForSequenceClassification.from_pretrained("./fine_tuned_bert")
12 tokenizer = BertTokenizer.from pretrained("./fine tuned bert")
13
14 # Set up wandb (optional for experiment tracking)
15 wandb.init(project="sentiment-analysis-api", name="api-deployment")
17 # Define the input format using Pydantic
18 class Review(BaseModel):
19
      text: str
20
21 # Define the sentiment prediction endpoint
```

```
22 @app.post("/predict/")
23 def predict sentiment(review: Review):
24
       # Tokenize input text
       inputs = tokenizer(review.text, return_tensors="pt", truncation=True, padding=True,
25
26
27
      # Make prediction
28
      with torch.no grad():
20
           outputs = model(**inputs)
30
31
       # Get the predicted sentiment (choose the max logit)
       predicted class = torch.argmax(outputs.logits, dim=1).item()
32
33
      # Map class index to sentiment label
35
       sentiment = {0: "Negative", 1: "Neutral", 2: "Positive"}
36
       predicted sentiment = sentiment[predicted class]
37
38
      # Log the prediction to wandb (optional)
39
      wandb.log({"review": review.text, "predicted sentiment": predicted sentiment})
40
      return {"review": review.text, "sentiment": predicted sentiment}
41
42
```

Changes to your `wandb` environment variables will be ignored because your `wandb` session has already started. For more information on how to modify your settings with `wandb.init()` arguments, please refer to the W&B docs.

Finishing last run (ID:iynw8jpp) before initializing another...

Run history:

eval/loss eval/runtime eval/samples_per_second eval/steps_per_second test/loss test/runtime test/samples_per_second train/epoch train/global_step

Run summary:

train steps per second

eval/loss	1.02003
eval/runtime	3.4203
eval/samples_per_second	0.292
eval/steps_per_second	0.292
test/loss	1.02003
test/runtime	2.8443
test/samples_per_second	0.352
test/steps_per_second	0.352
total_flos	2368020759552.0
train/epoch	3
train/global_step	3
train_loss	1.08066
train_runtime	212.1883
train_samples_per_second	0.042

0.014

View run ./results at: https://wandb.ai/anshsharma21050421/huggingface/runs/iynw8jpp

View project at: https://wandb.ai/anshsharma21050421/huggingface

Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)

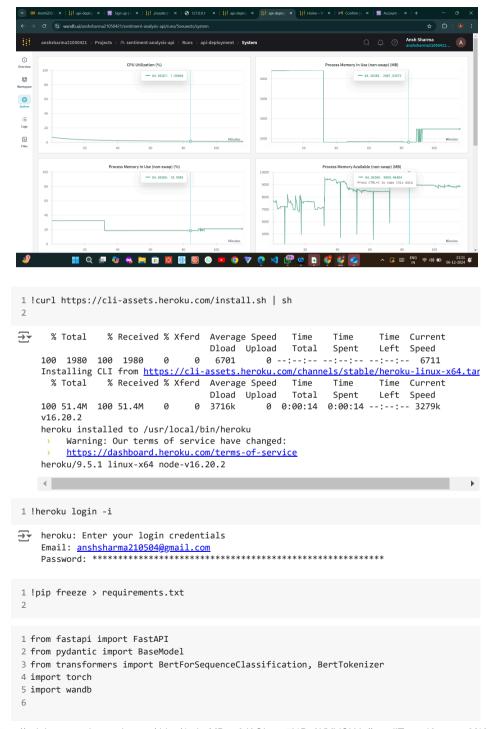
Find logs at: ./wandb/run-20241206_133753-iynw8jpp/logs

Successfully finished last run (ID:iynw8jpp). Initializing new run:

Tracking run with wandb version 0.18.7

Run data is saved locally in /content/wandb/run-20241206_134511-5neuwzts

Syncing run api-deployment to Weights & Biases (docs)



```
7 app = FastAPI()
9 model = BertForSequenceClassification.from pretrained("./fine tuned bert")
10 tokenizer = BertTokenizer.from pretrained("./fine tuned bert")
12 class Review(BaseModel):
13
     text: str
1/
15 @app.post("/predict/")
16 def predict sentiment(review: Review):
      inputs = tokenizer(review.text, return_tensors="pt", truncation=True, padding=True,
12
      with torch.no grad():
19
           outputs = model(**inputs)
      predicted class = torch.argmax(outputs.logits, dim=1).item()
       sentiment = {0: "Negative", 1: "Neutral", 2: "Positive"}
21
      predicted sentiment = sentiment[predicted class]
22
23
      wandb.log({"review": review.text, "predicted_sentiment": predicted_sentiment})
      return {"review": review.text, "sentiment": predicted sentiment}
25
```

Pushing GIT Repository and Heroku Deployement

```
1 !git init
\Longrightarrow hint: Using 'master' as the name for the initial branch. This default branch name
    hint: is subject to change. To configure the initial branch name to use in all
    hint: of your new repositories, which will suppress this warning, call:
            git config --global init.defaultBranch <name>
    hint: Names commonly chosen instead of 'master' are 'main', 'trunk' and
    hint: 'development'. The just-created branch can be renamed via this command:
    hint: git branch -m <name>
    Initialized empty Git repository in /content/.git/
1 !git config --global user.email "ansh_b230825mt@nitc.ac.in"
2 !git config --global user.name "Anshsharmacse"
1 !git add .config/ fine_tuned_bert/ logs/ requirements.txt results/ sample_data/ wandb/
2
→ .config/
    fine tuned bert
    requirements.txt
    results
    sample data
1 !git commit -m "Initial commit"
2
```

```
[master (root-commit) 13426eel Initial commit
56 files changed, 84591 insertions(+)
create mode 100644 .config/.last opt in prompt.yaml
create mode 100644 .config/.last survey prompt.yaml
create mode 100644 .config/.last update check.json
create mode 100644 .config/active config
create mode 100644 .config/config sentinel
create mode 100644 .config/configurations/config default
create mode 100644 .config/default configs.db
create mode 100644 .config/gce
create mode 100644 .config/hidden gcloud config universe descriptor data cache con
create mode 100644 .config/logs/2024.12.04/14.22.43.568001.log
create mode 100644 .config/logs/2024.12.04/14.23.06.536794.log
create mode 100644 .config/logs/2024.12.04/14.23.17.977201.log
create mode 100644 .config/logs/2024.12.04/14.23.19.373946.log
create mode 100644 .config/logs/2024.12.04/14.23.31.005371.log
create mode 100644 .config/logs/2024.12.04/14.23.31.672215.log
create mode 100644 fine tuned bert/config.json
create mode 100644 fine tuned bert/model.safetensors
create mode 100644 fine tuned bert/special tokens map.json
create mode 100644 fine_tuned_bert/tokenizer_config.json
create mode 100644 fine_tuned_bert/vocab.txt
create mode 100644 logs/events.out.tfevents.1733491855.223463f26f82.1619.0
create mode 100644 logs/events.out.tfevents.1733492156.223463f26f82.1619.1
create mode 100644 logs/events.out.tfevents.1733492454.223463f26f82.1619.2
create mode 100644 requirements.txt
create mode 100644 results/checkpoint-3/config.json
create mode 100644 results/checkpoint-3/model.safetensors
create mode 100644 results/checkpoint-3/optimizer.pt
create mode 100644 results/checkpoint-3/rng_state.pth
create mode 100644 results/checkpoint-3/scheduler.pt
create mode 100644 results/checkpoint-3/trainer_state.json
create mode 100644 results/checkpoint-3/training args.bin
create mode 100755 sample data/README.md
create mode 100755 sample_data/anscombe.json
create mode 100644 sample_data/california_housing_test.csv
create mode 100644 sample_data/california_housing_train.csv
create mode 100644 sample data/mnist test.csv
create mode 100644 sample data/mnist train small.csv
create mode 120000 wandb/debug-internal.log
create mode 120000 wandb/debug.log
create mode 120000 wandb/latest-run
create mode 100644 wandb/run-20241206 133753-iynw8jpp/files/config.yaml
create mode 100644 wandb/run-20241206 133753-iynw8jpp/files/output.log
create mode 100644 wandb/run-20241206_133753-iynw8jpp/files/requirements.txt
create mode 100644 wandb/run-20241206_133753-iynw8jpp/files/wandb-metadata.json
create mode 100644 wandb/run-20241206_133753-iynw8jpp/files/wandb-summary.json
create mode 120000 wandb/run-20241206 133753-iynw8jpp/logs/debug-core.log
create mode 100644 wandb/run-20241206 133753-iynw8jpp/logs/debug-internal.log
create mode 100644 wandb/run-20241206_133753-iynw8jpp/logs/debug.log
create mode 100644 wandb/run-20241206 133753-iynw8jpp/run-iynw8jpp.wandb
create mode 100644 wandb/run-20241206 134511-5neuwzts/files/output.log
create mode 100644 wandb/run-20241206 134511-5neuwzts/files/requirements.txt
create mode 100644 wandb/run-20241206 134511-5neuwzts/files/wandb-metadata.json
create mode 120000 wandb/run-20241206_134511-5neuwzts/logs/debug-core.log
create mode 100644 wandb/run-20241206_134511-5neuwzts/logs/debug-internal.log
create mode 100644 wandb/run-20241206 134511-5neuwzts/logs/debug.log
```

1 !heroku create <Ansh Sentimental Analysis>

```
→ /bin/bash: -c: line 1: syntax error near unexpected token `newline'
    /bin/bash: -c: line 1: `heroku create <Ansh Sentimental Analysis>'
1 !heroku apps
→ heroku: Press any key to open up the browser to login or q to exit:
1 !heroku create Anshsharmacse
→ Creating • Anshsharmacse...
    Creating • Anshsharmacse...
    Creating ● Anshsharmacse...?
    heroku: Press any key to open up the browser to login or q to exit: Creating ● Anshsh
1 !git remote add heroku https://git.heroku.com/Anshsharmacse.git
→ error: remote heroku already exists.
1 !git remote -v
→ heroku <a href="https://git.heroku.com/your-app-name.git">https://git.heroku.com/your-app-name.git</a> (fetch)
    heroku https://git.heroku.com/your-app-name.git (push)
1 !heroku login -i
→ heroku: Enter your login credentials
    Email: anshsharma210504@gmail.com
    Password: **************
```

Final Evaluation

Steps

- 1. Write the Review Text in the review box
- 2. Click on Predict Sentiment

Note-->Refresh the cell on every new sentiment, to obtain the accurate result

```
1 # Import libraries
2 from transformers import BertForSequenceClassification, BertTokenizer
3 import torch
4 import ipywidgets as widgets
```

```
5 from IPvthon.display import display
7 # Load your fine-tuned model and tokenizer (replace with your model path if necessary)
8 model = BertForSequenceClassification.from pretrained("./fine tuned bert")
9 tokenizer = BertTokenizer.from pretrained("./fine tuned bert")
10
11 # Create a text box for user input
12 input text = widgets.Textarea(
      value=''.
14
       placeholder='Type your review here...',
15
      description='Review Text:',
16
      disabled=False.
17
      layout=widgets.Layout(width='50%', height='100px')
18)
19
20 # Define a function to predict sentiment
21 def predict_sentiment(text):
22
      # Tokenize input text
23
      inputs = tokenizer(text, return tensors="pt", truncation=True, padding=True, max le
24
25
      # Make prediction
26
      with torch.no grad():
27
           outputs = model(**inputs)
28
29
      # Get predicted sentiment (choose the max logit)
30
      predicted class = torch.argmax(outputs.logits, dim=1).item()
31
      # Map class index to sentiment label
33
      sentiment = {0: "Negative", 1: "Neutral", 2: "Positive"}
34
       predicted sentiment = sentiment[predicted class]
35
36
      # Display the result
37
      print(f"Predicted Sentiment: {predicted_sentiment}")
39 # Create a button to trigger the prediction
40 predict button = widgets.Button(description="Predict Sentiment")
42 # Link the button to the prediction function
43 predict button.on click(lambda x: predict sentiment(input text.value))
45 # Display the input field and the button
46 display(input_text, predict_button)
47
```

Review Text: GDSC is the best club , Ansh Sharma will certainely be selected in GDSC Club.

Predicted Sentiment: Positive

Predict Sentiment