Report

Objective :

The main objective of the Task is to generate summaries of product reviews obtained from the Amazon Fine Food Reviews dataset. We'd fine-tuned the GPT-2 model on the review data and evaluate its performance in generating to the point summaries.

1. Data Preprocessing :

First we take out the two columns "Text" and "Summary" from the original Amazon Fine Food Review dataset and then applied text preprocessing by the method preprocess_text() to both the columns "Text" and "Summary" and converted them into a new dataset called `preprocessed_data`. And after that we kept only 10% of "preprocessed_data " data called trimmed_data and used it for further steps(Training, Evaluation, Generating Summaries) because of time and space constraints.

```python
# Trim the DataFrame to its 10% of original size
trimmed_data = data.sample(n=new_size, random_state=42)
trimmed_data
```

[22]:

	Text	Summary
41432	actually tasty pure potatoes great texture nas...	like
209473	realize taste matter personal preference food ...	good subjectively 5 star
247296	one favorite cup soup choices never carried st...	lipton cup soup spring vegetable4 oz
80087	like classic taste good margarita likely youll...	suited purpose quite goal
218571	willing give chance even reading scathing revi...	tastes artificial
...	...	...
518527	bought thing curiosity tackle roach problem en...	impressed
492179	ive used xplod drink mixes like past always en...	energy without stomach aches
299310	im training first marathon order package heari...	great keeping energy
101940	even think baby food tastes good 9 month old a...	great baby food buy somewhere else

2. Model Training :

We initialized a GPT-2 tokenizer and model from HuggingFace and divided the trimmed_data into train and test set in the ratio(75:25).

A custom dataset class was created to prepare the data for training where each input consisted of the review text concatenated with a special token "<BOS>" and summary text.

```
[33]:   # Step-4 : Training

        num_epochs = 10
        for epoch in range(num_epochs):
            model.train()
            for batch in train_loader:
                input_ids, labels = batch
                input_ids = input_ids.to(device)
                labels = labels.to(device)

                optimizer.zero_grad()

                outputs = model(input_ids=input_ids, labels=labels)
                loss = outputs.loss
                loss.backward()

                optimizer.step()

            print(f"Epoch {epoch+1}/{num_epochs}, Loss: {loss.item()}")
```

```
Epoch 1/10, Loss: 0.22095863352109909
Epoch 2/10, Loss: 0.18844905495643616
Epoch 3/10, Loss: 0.35137054324150085
Epoch 4/10, Loss: 0.0789911076426506
Epoch 5/10, Loss: 0.08545723557472229
Epoch 6/10, Loss: 0.0809001624584198
Epoch 7/10, Loss: 0.06948107481002808
Epoch 8/10, Loss: 0.07574597001075745
Epoch 9/10, Loss: 0.05396013334393501
Epoch 10/10, Loss: 0.09450507909059525
```

3. Fine-tuning the Model:

The GPT-2 model was fine-tuned on the review data using a training set. We did some experiments with different hyperparameters such as learning rate, batch size and number of epochs to optimize the model to get best performance.

While training, we chose the AdamW optimizer and back-propagated loss to the model.

```
[30]: train_loader = DataLoader(train_dataset, batch_size=8, shuffle=True)
      test_loader = DataLoader(test_dataset, batch_size=8, shuffle=False)
```

```
[32]: optimizer = torch.optim.AdamW(model.parameters(), lr=1e-4)
```

4. Evaluation :

After fine-tuning the model, we evaluated the model on the test dataset to calculate the model performance in generating summaries.

We determined ROUGE scores (Recall-Oriented Understudy for Gisting Evaluation) to measure the overlap between the generated summaries and the actual summaries in terms of n-gram overlap, recall, precision, and F1 score.

```
num of batches: 889
Average Loss on Test Set: 0.21930818596626428
```

```
[41]: # Print ROUGE scores
      print("ROUGE Scores:")
      print("ROUGE-1: Precision: {:.2f}, Recall: {:.2f}, F1-Score: {:.2f}".format(scores['rouge-1']['p'], scores['rouge-1']['r'], sc
      print("ROUGE-2: Precision: {:.2f}, Recall: {:.2f}, F1-Score: {:.2f}".format(scores['rouge-2']['p'], scores['rouge-2']['r'], sc
      print("ROUGE-L: Precision: {:.2f}, Recall: {:.2f}, F1-Score: {:.2f}".format(scores['rouge-l']['p'], scores['rouge-l']['r'], sc
```

```
ROUGE Scores:
ROUGE-1: Precision: 0.01, Recall: 0.06, F1-Score: 0.01
ROUGE-2: Precision: 0.00, Recall: 0.00, F1-Score: 0.00
ROUGE-L: Precision: 0.01, Recall: 0.06, F1-Score: 0.01
```

Hyper Parameter tuning (using 1% of original data)

1e-05 2
Epoch 1/3, Loss: 0.10763468950336859
Epoch 2/3, Loss: 0.10264296125704475
Epoch 3/3, Loss: 0.09768123026259286

Hyperparameters: LR=1e-05, Batch Size=2, Epochs=3, Average Loss on Test Set: 0.17511065624413275
Epoch 1/5, Loss: 0.09353177442248582
Epoch 2/5, Loss: 0.08884868070773264
Epoch 3/5, Loss: 0.08446727622314365
Epoch 4/5, Loss: 0.07997531429156801
Epoch 5/5, Loss: 0.07599304376631333
Hyperparameters: LR=1e-05, Batch Size=2, Epochs=5, Average Loss on Test Set: 0.21473581127497549
Epoch 1/8, Loss: 0.07280005656329541
Epoch 2/8, Loss: 0.06836133787764925
Epoch 3/8, Loss: 0.06503848290144279
Epoch 4/8, Loss: 0.061898817195544825
Epoch 5/8, Loss: 0.05885151964323997
Epoch 6/8, Loss: 0.056171942233172636
Epoch 7/8, Loss: 0.05348305890859082
Epoch 8/8, Loss: 0.051170274920195456
Hyperparameters: LR=1e-05, Batch Size=2, Epochs=8, Average Loss on Test Set: 0.2743694307400456
1e-05 4
Epoch 1/3, Loss: 0.04858995489007313
Epoch 2/3, Loss: 0.04696351165165094
Epoch 3/3, Loss: 0.04541790510194051
Hyperparameters: LR=1e-05, Batch Size=4, Epochs=3, Average Loss on Test Set: 0.28949261961107176
Epoch 1/5, Loss: 0.044589667830888816
Epoch 2/5, Loss: 0.04347351418832565
Epoch 3/5, Loss: 0.04234540036147352
Epoch 4/5, Loss: 0.04173196038129909
Epoch 5/5, Loss: 0.040652127632448416
Hyperparameters: LR=1e-05, Batch Size=4, Epochs=5, Average Loss on Test Set: 0.3070501001614533
Epoch 1/8, Loss: 0.040381700041621754
Epoch 2/8, Loss: 0.03945687493417713
Epoch 3/8, Loss: 0.03892290752772486
Epoch 4/8, Loss: 0.038349220405138794
Epoch 5/8, Loss: 0.03760478530764077
Epoch 6/8, Loss: 0.03709731797126819
Epoch 7/8, Loss: 0.036642360488760094
Epoch 8/8, Loss: 0.03630341311451921

Hyperparameters: LR=1e-05, Batch Size=4, Epochs=8, Average Loss on Test Set: 0.3266136610273565

1e-05 8

Epoch 1/3, Loss: 0.03570236867436996
Epoch 2/3, Loss: 0.034856723911543055
Epoch 3/3, Loss: 0.03487640853069625

Hyperparameters: LR=1e-05, Batch Size=8, Epochs=3, Average Loss on Test Set: 0.33421916531377965

Epoch 1/5, Loss: 0.034779092415775856
Epoch 2/5, Loss: 0.03433776617497485
Epoch 3/5, Loss: 0.03417712970729393
Epoch 4/5, Loss: 0.03356366385452408
Epoch 5/5, Loss: 0.033599912798762545

Hyperparameters: LR=1e-05, Batch Size=8, Epochs=5, Average Loss on Test Set: 0.3486149608251754

Epoch 1/8, Loss: 0.03349210304779474
Epoch 2/8, Loss: 0.03314433281447941
Epoch 3/8, Loss: 0.03295854468590547
Epoch 4/8, Loss: 0.03288785447881996
Epoch 5/8, Loss: 0.03240311210433381
Epoch 6/8, Loss: 0.032486338804407815
Epoch 7/8, Loss: 0.032367967596840234
Epoch 8/8, Loss: 0.03207327238166981

Hyperparameters: LR=1e-05, Batch Size=8, Epochs=8, Average Loss on Test Set: 0.3564302914598015

5e-05 2

Epoch 1/3, Loss: 0.04644670335123382
Epoch 2/3, Loss: 0.04616730562917972
Epoch 3/3, Loss: 0.04346038064505297

Hyperparameters: LR=5e-05, Batch Size=2, Epochs=3, Average Loss on Test Set: 0.2611521486475875

Epoch 1/5, Loss: 0.04271872591325143
Epoch 2/5, Loss: 0.03887979833690509
Epoch 3/5, Loss: 0.036991499493241056
Epoch 4/5, Loss: 0.0355270800008622
Epoch 5/5, Loss: 0.03407640685523925

Hyperparameters: LR=5e-05, Batch Size=2, Epochs=5, Average Loss on Test Set: 0.29631856020995834

Epoch 1/8, Loss: 0.03378680743225527
Epoch 2/8, Loss: 0.03189574879324186

Epoch 3/8, Loss: 0.03054238436197579
Epoch 4/8, Loss: 0.029168894763325653
Epoch 5/8, Loss: 0.028377244904068146
Epoch 6/8, Loss: 0.027680464527794896
Epoch 7/8, Loss: 0.026649384099042644
Epoch 8/8, Loss: 0.02575026346156749
Hyperparameters: LR=5e-05, Batch Size=2, Epochs=8, Average Loss on Test Set:
0.2938796127458665
5e-05 4
Epoch 1/3, Loss: 0.024623924197531243
Epoch 2/3, Loss: 0.02325929330109739
Epoch 3/3, Loss: 0.022892354583711926
Hyperparameters: LR=5e-05, Batch Size=4, Epochs=3, Average Loss on Test Set:
0.326121072065127
Epoch 1/5, Loss: 0.023132231835656804
Epoch 2/5, Loss: 0.022790469415256524
Epoch 3/5, Loss: 0.0222262707455905
Epoch 4/5, Loss: 0.021792218390953556
Epoch 5/5, Loss: 0.021838182929857838
Hyperparameters: LR=5e-05, Batch Size=4, Epochs=5, Average Loss on Test Set:
0.31441768223231426
Epoch 1/8, Loss: 0.022010583654185845
Epoch 2/8, Loss: 0.021582246462746365
Epoch 3/8, Loss: 0.02095900726566654
Epoch 4/8, Loss: 0.021129785271363218
Epoch 5/8, Loss: 0.02046058248253679
Epoch 6/8, Loss: 0.020195405327017798
Epoch 7/8, Loss: 0.02036487774162376
Epoch 8/8, Loss: 0.020015041738762528
Hyperparameters: LR=5e-05, Batch Size=4, Epochs=8, Average Loss on Test Set:
0.3233362851099352
5e-05 8
Epoch 1/3, Loss: 0.019490315264359983
Epoch 2/3, Loss: 0.018948607178191752
Epoch 3/3, Loss: 0.018783769916260778
Hyperparameters: LR=5e-05, Batch Size=8, Epochs=3, Average Loss on Test Set:
0.34032257939322613
Epoch 1/5, Loss: 0.019084679431109
Epoch 2/5, Loss: 0.018912580126336182
Epoch 3/5, Loss: 0.01847564858012251

Epoch 4/5, Loss: 0.01841368572174888
Epoch 5/5, Loss: 0.018388857741754566
Hyperparameters: LR=5e-05, Batch Size=8, Epochs=5, Average Loss on Test Set:
0.34681457682941735
Epoch 1/8, Loss: 0.018686698478822655
Epoch 2/8, Loss: 0.018472766895965442
Epoch 3/8, Loss: 0.018259748617295478
Epoch 4/8, Loss: 0.018180884582499012
Epoch 5/8, Loss: 0.01834815284272212
Epoch 6/8, Loss: 0.01792970732008203
Epoch 7/8, Loss: 0.0179684556335714
Epoch 8/8, Loss: 0.01788071472299116
Hyperparameters: LR=5e-05, Batch Size=8, Epochs=8, Average Loss on Test Set:
0.3437500244446015
0.0001 2
Epoch 1/3, Loss: 0.02649111281207843
Epoch 2/3, Loss: 0.025692680213002705
Epoch 3/3, Loss: 0.025332192645181772
Hyperparameters: LR=0.0001, Batch Size=2, Epochs=3, Average Loss on Test Set:
0.2604559545265052
Epoch 1/5, Loss: 0.026409062564894593
Epoch 2/5, Loss: 0.024649096487200865
Epoch 3/5, Loss: 0.02443895482195639
Epoch 4/5, Loss: 0.024012225008624315