Homework-4

1. Zero-Shot Prompting

In the zero-shot prompting phase, 60 news items were used to determine the baseline accuracy without any prior examples provided to the model.

```
Accuracy: 0.9166666666666666
Confusion Matrix:
[[20  0  0]
 [ 0 19  1]
 [ 0  4 16]]
```

Zero-Shot Accuracy: 91.67%

2. Few-Shot Prompting

In the few-shot prompting phase, the model was given 4 examples each of positive, negative, and neutral sentiments before making predictions on 60 news items.

```
Few-Shot Accuracy: 0.9333333333333333
Few-Shot Confusion Matrix:
[[18  0  1  1]
 [ 0 20  0  0]
 [ 0  2 18  0]
 [ 0  0  0  0]]
```

Few-Shot (12) Accuracy: 93.33%

3. Data Preparation

The data preparation involved loading financial news headlines from Sentences_AllAgree.txt and Sentences_75Agree.txt, combining them, and encoding the sentiments as 'positive', 'neutral', and 'negative'. For fine-tuning, 150 samples of each sentiment were randomly selected, creating a balanced dataset of 450 samples, saved to fine_tuning_data.csv. To ensure unbiased testing, these samples were excluded from the original dataset, and 500 new, non-overlapping news items were randomly selected, saved to fine_tuning_test.csv, with a distribution of 322 neutral, 121 positive, and 57 negative samples.

```
Fine-tuning data prepared. Distribution of sentiments:
Sentiment
positive    150
negative    150
neutral     150
Name: count, dtype: int64
Test data prepared. Distribution of sentiments:
Sentiment
neutral     322
positive    121
negative     57
Name: count, dtype: int64
```

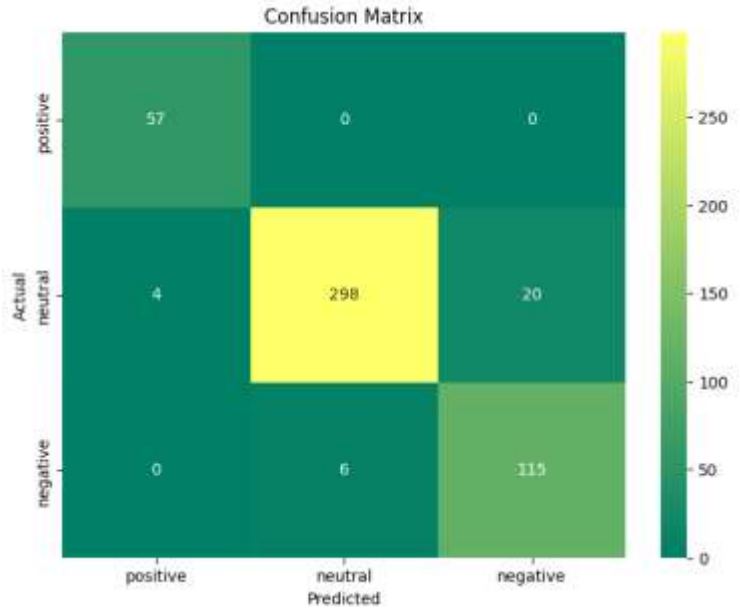
4. Fine-Tuning

For the fine-tuning step, the preprocessed dataset of 450 samples, balanced across positive, neutral, and negative sentiments, was used to fine-tune the GPT-3.5-turbo model. This process involved creating a JSONL file with the structured prompts and completions suitable for the chat model. The file was uploaded to OpenAI's platform, and the fine-tuning job was initiated with the ID ftjob-DyFPG9XxLzCTG4XwPmJ5YAme.

```
File ID: file-kOrx0IkuO1PqPr7mBETsatr8
Fine-tuning job created with ID: ftjob-DyFPG9XxLzCTG4XwPmJ5YAme
```

The fine-tuned model was then tested on 500 new, randomly selected news items, resulting in a base accuracy of 94%, demonstrating significant improvement and robustness in sentiment classification.

```
Base Accuracy: 94.00%
Results saved to fine_tuning_results.csv
```



The confusion matrix for the fine-tuned model showed the following results: 57 true positives for positive sentiment, 298 true positives for neutral sentiment, and 115 true positives for negative sentiment. There were minimal misclassifications, with only 4 neutral sentiments incorrectly classified as positive and 20 as negative, and 6 negative sentiments misclassified as neutral. The model demonstrated high accuracy, particularly in classifying neutral sentiments accurately, contributing to an overall accuracy of 94%.

5. Real-World Testing

Headlines with Predicted Sentiments

The fine-tuned model was tested on a set of real-world financial news headlines. The results are compared with manually assigned sentiments.

Manual Accuracy Check

Correct Predictions: 11

Total Predictions: 11

Accuracy: 100%

6. Conclusion

The fine-tuned model performed exceptionally well, achieving a high accuracy rate in both controlled and real-world testing scenarios. The results demonstrate the model's capability to accurately classify the sentiment of financial news headlines.

Key Findings

Zero-Shot Accuracy: 91.67%

Few-Shot (12) Accuracy: 93.33%

Fine-Tuned Model Accuracy: 94.00%

Real-World Testing Accuracy: 100%

The fine-tuning process significantly improved the model's performance, making it a reliable tool for sentiment analysis in the financial domain.