

# CS 613: NLP Assignment 3 Report

## Fine Tuning & Evaluation of Pre-trained Models

**Course Code:** CS 613  
**Assignment:** Assignment 3  
**Submission Date:** November 20, 2024

# Contents

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>Introduction</b>                                       | <b>2</b> |
| <b>2</b> | <b>Model Details</b>                                      | <b>2</b> |
| 2.1      | Pre-trained Model Selection . . . . .                     | 2        |
| 2.2      | Parameter Calculation . . . . .                           | 2        |
| <b>3</b> | <b>Fine-Tuning Details</b>                                | <b>2</b> |
| 3.1      | Task 1: Classification (SST-2) . . . . .                  | 2        |
| 3.2      | Task 2: Question-Answering (SQuAD) . . . . .              | 3        |
| <b>4</b> | <b>Results and Analysis</b>                               | <b>3</b> |
| 4.1      | Results . . . . .   | 3        |
| 4.2      | Performance Scores Before and After Fine-tuning . . . . . | 3        |
| 4.3      | Analysis of Results . . . . .                             | 4        |
| <b>5</b> | <b>Model Parameters After Fine-Tuning</b>                 | <b>4</b> |
| <b>6</b> | <b>Model Upload to Hugging Face</b>                       | <b>4</b> |
| <b>7</b> | <b>Conclusion</b>   | <b>5</b> |
| 7.1      | Question Answering Task: Predicted vs Actual . . . . .    | 13       |

# 1 Introduction

This report presents the fine-tuning and evaluation of the pre-trained model **meta-llama/Llama-3.2-1B** for two NLP tasks: Classification on the SST-2 dataset and Question-Answering on the SQuAD dataset. The tasks were performed on Kaggle Notebooks, and the results include both zero-shot and fine-tuned performance scores. For ease of computation, we have frozen the parameters of all layers and added an appropriate trainable layer at the end of the model.

## 2 Model Details

### 2.1 Pre-trained Model Selection

Selected model: **meta-llama/Llama-3.2-1B**.

The Meta Llama 3.2-1B is a compact, transformer-based language model with 1 billion parameters, designed for efficient and versatile natural language processing.

### 2.2 Parameter Calculation

The calculated parameters of the model are as follows:

|  | SST-2         | SQuAD         |
|--|---------------|---------------|
| Total Parameters (base Model)              | 1,235,814,400 | 1,235,814,400 |
| Total Parameters (after adding one layer)  | 1,235,818,496 | 1,235,818,498 |
| Trainable Parameters                       | 4,096         | 4,098         |
| Freezed Parameters (all except last layer) | 1,235,814,400 | 1,235,814,400 |
| Reported Parameters (from paper)           | 1.23B         |               |

Table 1: Parameter Counts for SST-2 and SQuAD Using Base Llama Model

Please refer Appendix 1 for more details

## 3 Fine-Tuning Details

### 3.1 Task 1: Classification (SST-2)

- Dataset: SST-2
- Train-test split: 80:20 (seed=1)
- Metrics: Accuracy, Precision, Recall, F1

Fine-tuning process and hyperparameters:

- Learning rate:  $2 \times 10^{-5}$
- Batch size: 16
- Epochs: 3

### 3.2 Task 2: Question-Answering (SQuAD)

- Dataset: SQuAD
- Train-test split: 80:20 (seed=1)
- Metrics: squad\_v2, F1, BLEU, ROUGE, exact-match

Fine-tuning process and hyperparameters:

- Learning rate:  $5 \times 10^{-5}$
- Batch size: 32
- Epochs: 2

## 4 Results and Analysis

### 4.1 Results

Refer Appendix 2 and Appendix 3 for model results

### 4.2 Performance Scores Before and After Fine-tuning

Table 2: Performance Metrics for Classification (SST-2)

| Metric    | Zero-Shot | Fine-Tuned |
|-----------|-----------|------------|
| Accuracy  | 0.473     | 0.8451     |
| Precision | 0.4369    | 0.8503     |
| Recall    | 0.1171    | 0.8445     |
| F1        | 0.1847    | 0.8474     |

Table 3: Performance Metrics for Question-Answering (SQuAD)

| Metric      | Zero-Shot | Fine-Tuned |
|-------------|-----------|------------|
| F1          | 75.0      | 2.748      |
| BLEU        | 0.00416   | 0.00419    |
| ROUGE-1     | 0.02885   | 0.02889    |
| ROUGE-2     | 0.00930   | 0.00938    |
| ROUGE-L     | 0.02859   | 0.02873    |
| Exact-Match | 0.0       | 7.2        |

Ideally the F1 score should be less than 1, but the values here suggest otherwise. This are the values the code gave. Refer to the github links provided for details, we have used code provided in the official PyTorch documentation.

### 4.3 Analysis of Results

- We froze the model's parameters and only trained the final layer which is same as a training a NN for classification. We leverage the pre-trained model's general language understanding and fine tuned it to the specific requirements of binary sentiment classification.
- For SQuAD, there is not much improvement, this is because question answering is a difficult task, and just training the last layer is not enough. The model captures biases from the original train dataset (of the Llama-3.2-1B Model). However, due to time and computational constraints, we could only train the last layer.

### Data and Fine-Tuning Issues

The minimal improvements in BLEU, ROUGE, and Exact Match suggest that the fine-tuning process was suboptimal. Possible reasons include:

- Insufficient or low-quality training data.
- A mismatch between the pre-trained model's capabilities and the target task.

### Model's Poor Task Adaptation

The consistently low scores across all metrics indicate the model might not be well-suited for the task without significant adjustments. This might require:

- Architecture modifications tailored to the task.
- More comprehensive or task-specific training data.

## 5 Model Parameters After Fine-Tuning

We add an extra layer for specific tasks (Classification and Question Answering). We freeze all the other layers and just train the last (fine-tuning) layers.

|                               | SST-2      | SQuAD      |
|-------------------------------|------------|------------|
| Parameters before Fine Tuning | 1235814400 | 1235814400 |
| Additional Parameters         | 4096       | 4096       |
| Additional Biases             | 0          | 2          |
| Parameters after Fine Tuning  | 1235818496 | 1235818498 |
| Freezed-Layes Parameters      | 1235814400 | 1235814400 |
| Trainable Parameters          | 4096       | 4098       |

After adding the relavant last layer, fine tuning the model does not change the number of parameters.

## 6 Model Upload to Hugging Face

The fine-tuned model is available on Hugging Face - SST2 Link and Hugging Face - SQuAD Link.

The code notebooks are available on Github - SST2 Link and Github - SQuAD Link.

## 7 Conclusion

### Part A

#### **Classification Task (Accuracy, Precision, Recall, F1)**

Higher scores indicate better performance Expectations: Higher scores for fine-tuned models compared to pre-trained (zero-shot) models Rationale: Fine-tuned models are specifically trained on the task-specific dataset, learning relevant features and patterns, leading to improved predictions. Pre-trained models lack task-specific adaptation, so their performance is generally lower in zero-shot settings.

#### **Question Answering Task (SQuAD v2, F1, BLEU, ROUGE, Exact Match)**

Higher scores indicate better performance Expectations: Higher scores for fine-tuned models compared to pre-trained (zero-shot) models Rationale: Metrics like BLEU, ROUGE, and METEOR directly measure textual overlap and semantic similarity. These values improve after fine-tuning since the model has learned to predict answers closer to the reference. Exact Match (EM) and F1 improve as fine-tuning helps the model align predictions more closely with ground truth.

### Part B

The number of parameters remains the same for both pre-trained and fine-tuned models because fine-tuning does not alter the model's architecture (e.g., number of layers, hidden units, or attention heads). Instead, it just updates the weights of final layers to adapt the model to the specific task.

### Part C

Fine-tuned models trained on task-specific datasets outperform zero-shot models because they are additionally trained for the specific task. The zero-shot model, limited by its lack of task-specific optimization, demonstrates poor performance on task specific datasets as compared to finetuned models but displays strength of the pre-trained representations for general tasks.

## Work Distribution

- **Aditya Mehta (Roll Number: 22110017)**

- Developed the code for SQuAD, with the notebook and trained model available in the links above.
- Computed various matrices for SQuAD.
- Compiled and wrote the entire report.

- **Daksh Jain (Roll Number: 22110066)**

- Developed the code for SQuAD, with the notebook and trained model available in the links above.
- Conducted research on the Llama-3.2-1B model and calculated its parameters.

- **Hrriday Ruparel (Roll Number: 22110099)**

- Fine-tuned the SQuAD model.
- Analyzed the results for Question 7.
- Provided debugging assistance and support as needed.

- **Kishan Ved (Roll Number: 22110122)**

- Developed the code for SST-2, with the notebook and trained model available on Hugging Face (link above).
- Computed various matrices for both SST-2 and SQuAD.
- Wrote scripts to upload the model to Hugging Face.

- **Summet Sawale (Roll Number: 22110234)**

- Developed the code for SST-2, with the notebook and trained model available on Hugging Face (link above).
- Computed various matrices for SST-2.

## Appendix 1

Add all references cited in the report.

### Parameter details for Llama 3.2 1B Base Model

| Modules  | Parameters |
|--|------------|
| model.embed <sub>tokens</sub> .weight                                  | 262668288  |
| model.layers.0.self <sub>attn</sub> .q <sub>proj</sub> .weight         | 4194304    |
| model.layers.0.self <sub>attn</sub> .k <sub>proj</sub> .weight         | 1048576    |
| model.layers.0.self <sub>attn</sub> .v <sub>proj</sub> .weight         | 1048576    |
| model.layers.0.self <sub>attn</sub> .o <sub>proj</sub> .weight         | 4194304    |
| model.layers.0.mlp.gate <sub>proj</sub> .weight                        | 16777216   |
| model.layers.0.mlp.up <sub>proj</sub> .weight                          | 16777216   |
| model.layers.0.mlp.down <sub>proj</sub> .weight                        | 16777216   |
| model.layers.0.input <sub>layernorm</sub> .weight                      | 2048       |
| model.layers.0.post <sub>attn</sub> input <sub>layernorm</sub> .weight | 2048       |
| model.layers.1.self <sub>attn</sub> .q <sub>proj</sub> .weight         | 4194304    |
| model.layers.1.self <sub>attn</sub> .k <sub>proj</sub> .weight         | 1048576    |
| model.layers.1.self <sub>attn</sub> .v <sub>proj</sub> .weight         | 1048576    |
| model.layers.1.self <sub>attn</sub> .o <sub>proj</sub> .weight         | 4194304    |
| model.layers.1.mlp.gate <sub>proj</sub> .weight                        | 16777216   |
| model.layers.1.mlp.up <sub>proj</sub> .weight                          | 16777216   |
| model.layers.1.mlp.down <sub>proj</sub> .weight                        | 16777216   |
| model.layers.1.input <sub>layernorm</sub> .weight                      | 2048       |
| model.layers.1.post <sub>attn</sub> input <sub>layernorm</sub> .weight | 2048       |
| model.layers.2.self <sub>attn</sub> .q <sub>proj</sub> .weight         | 4194304    |
| model.layers.2.self <sub>attn</sub> .k <sub>proj</sub> .weight         | 1048576    |
| model.layers.2.self <sub>attn</sub> .v <sub>proj</sub> .weight         | 1048576    |
| model.layers.2.self <sub>attn</sub> .o <sub>proj</sub> .weight         | 4194304    |
| model.layers.2.mlp.gate <sub>proj</sub> .weight                        | 16777216   |
| model.layers.2.mlp.up <sub>proj</sub> .weight                          | 16777216   |
| model.layers.2.mlp.down <sub>proj</sub> .weight                        | 16777216   |
| model.layers.2.input <sub>layernorm</sub> .weight                      | 2048       |
| model.layers.2.post <sub>attn</sub> input <sub>layernorm</sub> .weight | 2048       |
| model.layers.3.self <sub>attn</sub> .q <sub>proj</sub> .weight         | 4194304    |
| model.layers.3.self <sub>attn</sub> .k <sub>proj</sub> .weight         | 1048576    |
| model.layers.3.self <sub>attn</sub> .v <sub>proj</sub> .weight         | 1048576    |
| model.layers.3.self <sub>attn</sub> .o <sub>proj</sub> .weight         | 4194304    |
| model.layers.3.mlp.gate <sub>proj</sub> .weight                        | 16777216   |
| model.layers.3.mlp.up <sub>proj</sub> .weight                          | 16777216   |
| model.layers.3.mlp.down <sub>proj</sub> .weight                        | 16777216   |
| model.layers.3.input <sub>layernorm</sub> .weight                      | 2048       |
| model.layers.3.post <sub>attn</sub> input <sub>layernorm</sub> .weight | 2048       |
| model.layers.4.self <sub>attn</sub> .q <sub>proj</sub> .weight         | 4194304    |
| model.layers.4.self <sub>attn</sub> .k <sub>proj</sub> .weight         | 1048576    |
| model.layers.4.self <sub>attn</sub> .v <sub>proj</sub> .weight         | 1048576    |
| model.layers.4.self <sub>attn</sub> .o <sub>proj</sub> .weight         | 4194304    |



| Modules  | Parameters |
|--|------------|
| model.layers.4.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.4.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.4.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.4.input <sub>l</sub> ayernorm.weight                      | 2048       |
| model.layers.4.post <sub>a</sub> ttention <sub>l</sub> ayernorm.weight | 2048       |
| model.layers.5.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.5.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.5.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.5.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |
| model.layers.5.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.5.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.5.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.5.input <sub>l</sub> ayernorm.weight                      | 2048       |
| model.layers.5.post <sub>a</sub> ttention <sub>l</sub> ayernorm.weight | 2048       |
| model.layers.6.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.6.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.6.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.6.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |
| model.layers.6.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.6.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.6.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.6.input <sub>l</sub> ayernorm.weight                      | 2048       |
| model.layers.6.post <sub>a</sub> ttention <sub>l</sub> ayernorm.weight | 2048       |
| model.layers.7.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.7.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.7.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.7.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |
| model.layers.7.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.7.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.7.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.7.input <sub>l</sub> ayernorm.weight                      | 2048       |
| model.layers.7.post <sub>a</sub> ttention <sub>l</sub> ayernorm.weight | 2048       |
| model.layers.8.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.8.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.8.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.8.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |
| model.layers.8.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.8.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.8.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.8.input <sub>l</sub> ayernorm.weight                      | 2048       |
| model.layers.8.post <sub>a</sub> ttention <sub>l</sub> ayernorm.weight | 2048       |
| model.layers.9.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.9.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.9.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.9.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |

| Modules   | Parameters |
|---|------------|
| model.layers.9.mlp.gate <sub>p</sub> proj.weight                        | 16777216   |
| model.layers.9.mlp.up <sub>p</sub> proj.weight                          | 16777216   |
| model.layers.9.mlp.down <sub>p</sub> proj.weight                        | 16777216   |
| model.layers.9.input <sub>i</sub> ayernorm.weight                       | 2048       |
| model.layers.9.post <sub>a</sub> ttention <sub>i</sub> ayernorm.weight  | 2048       |
| model.layers.10.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.10.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.10.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.10.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |
| model.layers.10.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.10.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.10.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.10.input <sub>i</sub> ayernorm.weight                      | 2048       |
| model.layers.10.post <sub>a</sub> ttention <sub>i</sub> ayernorm.weight | 2048       |
| model.layers.11.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.11.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.11.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.11.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |
| model.layers.11.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.11.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.11.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.11.input <sub>i</sub> ayernorm.weight                      | 2048       |
| model.layers.11.post <sub>a</sub> ttention <sub>i</sub> ayernorm.weight | 2048       |
| model.layers.12.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.12.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.12.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.12.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |
| model.layers.12.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.12.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.12.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.12.input <sub>i</sub> ayernorm.weight                      | 2048       |
| model.layers.12.post <sub>a</sub> ttention <sub>i</sub> ayernorm.weight | 2048       |
| model.layers.13.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.13.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.13.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.13.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |
| model.layers.13.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.13.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.13.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.13.input <sub>i</sub> ayernorm.weight                      | 2048       |
| model.layers.13.post <sub>a</sub> ttention <sub>i</sub> ayernorm.weight | 2048       |
| model.layers.14.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.14.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.14.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.14.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |

| Modules   | Parameters |
|---|------------|
| model.layers.14.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.14.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.14.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.14.input <sub>l</sub> ayernorm.weight                      | 2048       |
| model.layers.14.post <sub>a</sub> ttention <sub>l</sub> ayernorm.weight | 2048       |
| model.layers.15.self <sub>a</sub> ttn.q <sub>p</sub> proj.weight        | 4194304    |
| model.layers.15.self <sub>a</sub> ttn.k <sub>p</sub> proj.weight        | 1048576    |
| model.layers.15.self <sub>a</sub> ttn.v <sub>p</sub> proj.weight        | 1048576    |
| model.layers.15.self <sub>a</sub> ttn.o <sub>p</sub> proj.weight        | 4194304    |
| model.layers.15.mlp.gate <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.15.mlp.up <sub>p</sub> proj.weight                         | 16777216   |
| model.layers.15.mlp.down <sub>p</sub> proj.weight                       | 16777216   |
| model.layers.15.input <sub>l</sub> ayernorm.weight                      | 2048       |
| model.layers.15.post <sub>a</sub> ttention <sub>l</sub> ayernorm.weight | 2048       |
| model.norm.weight   | 2048       |

All the above parameters were frozen while fine tuning, to reduce computation.

**Trainable parameter details for fine tuned model for sentiment analysis**

| Modules      | Parameters |
|--------------|------------|
| score.weight | 4096       |

**Trainable parameter details for fine tuned model for question answering**

| Modules      | Parameters |
|--------------|------------|
| score.weight | 4096       |
| score.bias   | 2          |

## Appendix 2

### Sentiment analysis: Model Predictions and Actual Sentiments

The following are examples of model predictions compared to the actual sentiments for various sentences:

- **Sentence 1:** "it 's a charming and often affecting journey ."
  - **Predicted:** Positive (1)
  - **Actual:** Positive (1)
- **Sentence 2:** "unflinchingly bleak and desperate"
  - **Predicted:** Positive (1)
  - **Actual:** Negative (0)
- **Sentence 3:** "allows us to hope that nolan is poised to embark a major career as a commercial yet inventive filmmaker ."
  - **Predicted:** Positive (1)
  - **Actual:** Positive (1)
- **Sentence 4:** "the acting , costumes , music , cinematography and sound are all astounding given the production 's austere locales ."
  - **Predicted:** Positive (1)
  - **Actual:** Positive (1)
- **Sentence 5:** "it 's slow – very , very slow ."
  - **Predicted:** Negative (0)
  - **Actual:** Negative (0)
- **Sentence 6:** "although laced with humor and a few fanciful touches , the film is a refreshingly serious look at young women ."
  - **Predicted:** Positive (1)
  - **Actual:** Positive (1)
- **Sentence 7:** "a sometimes tedious film ."
  - **Predicted:** Negative (0)
  - **Actual:** Negative (0)
- **Sentence 8:** "or doing last year 's taxes with your ex-wife ."
  - **Predicted:** Negative (0)
  - **Actual:** Negative (0)
- **Sentence 9:** "you do n't have to know about music to appreciate the film 's easygoing blend of comedy and romance ."

- **Predicted:** Positive (1)
- **Actual:** Positive (1)
- **Sentence 10:** "in exactly 89 minutes , most of which passed as slowly as if i 'd been sitting naked on an igloo , formula 51 sank from quirky to jerky to utter turkey ."
  - **Predicted:** Negative (0)
  - **Actual:** Negative (0)

## Appendix 3

### 7.1 Question Answering Task: Predicted vs Actual

The following shows the predicted and actual results for the first ten samples in the Question Answering task:

- **Sample 1:**
  - **Predicted:** ""
  - **Actual:** France
- **Sample 2:**
  - **Predicted:** 10
  - **Actual:** 10th and 11th centuries
- **Sample 3:**
  - **Predicted:** Normans
  - **Actual:** Denmark, Iceland and Norway
- **Sample 4:**
  - **Predicted:** The Normans
  - **Actual:** Rollo
- **Sample 5:**
  - **Predicted:** and 11th centuries gave
  - **Actual:** 10th century
- **Sample 6:**
  - **Predicted:** ""
  - **Actual:** ""
- **Sample 7:**
  - **Predicted:** Rollo
  - **Actual:** Rollo
- **Sample 8:**
  - **Predicted:** to Normandy, a
  - **Actual:** ""
- **Sample 9:**
  - **Predicted:** The
  - **Actual:** William the Conqueror