**Advanced Deep Learning (AIGC 5500) Final Project**

**Topic: Sentiment Analysis on Yelp Restaurant Reviews**

**Section: ONA**

**Group Members:**

* **Hozefa Patel** [N01686385]
* **Jenil Pancholi** [N01665133]
* **Dev Ariwala** [N01664568]

**1. Introduction**

This report presents the final project for the Advanced Deep Learning (AIGC 5500) course. The project focuses on sentiment analysis using advanced NLP techniques, specifically comparing the performance of an LSTM (Long Short-Term Memory) model and a transformer-based DistilBERT model. The objective is to classify Yelp restaurant reviews into positive, negative, or neutral sentiment. The study emphasizes understanding the impact of different deep learning architectures on the performance of text classification tasks.

**2. Dataset Description and Preprocessing**

**Dataset:**

The dataset used for this project consists of 19,897 Yelp restaurant reviews, with the following key columns:

* **Review Text:** The textual content of the review, which serves as the primary input for the models.
* **Rating:** The associated star rating of each review, used to generate sentiment labels.

**Sentiment Labeling:**

The sentiment labeling was done based on the star ratings:

* **Positive:** Ratings of 4 or 5.
* **Neutral:** Rating of 3.
* **Negative:** Ratings of 1 or 2.

**Data Preprocessing:**

The following preprocessing techniques were applied:

1. **Text Cleaning:** Removal of HTML tags, punctuation, and conversion of the text to lowercase to standardize the format.
2. **Tokenization:** The text was tokenized using DistilBertTokenizer for DistilBERT and standard tokenization methods for LSTM. The tokenization process converts text into tokens (words or subwords).
3. **Padding and Truncation:** All sequences were padded or truncated to a maximum length of 200 tokens to ensure uniform input dimensions.

**3. Classification Methods**

**LSTM Model:**

The LSTM model is a type of recurrent neural network that is particularly suited for processing sequential data, such as text. The architecture used in this project is as follows:

* **Embedding Layer:** Converts input words into dense vector representations (embeddings).
* **LSTM Layers:** The model consists of two LSTM layers. The first LSTM layer has 128 units with return\_sequences=True, followed by a second LSTM layer with 64 units to capture sequential dependencies in the text.
* **Dense Layer:** A fully connected layer with 64 units and ReLU activation to process the features extracted by the LSTM layers.
* **Dropout Layer:** A dropout layer with a 50% dropout rate to prevent overfitting.
* **Output Layer:** A single neuron with softmax activation, providing the probabilities of the input text belonging to the positive, neutral, or negative classes.

The model was trained using the Adam optimizer with a categorical cross-entropy loss function. The training was conducted over 10 epochs with a batch size of 32, and 10% of the training data was reserved for validation.

**DistilBERT Model:**

DistilBERT is a transformer-based model that has been distilled to reduce size and increase speed while maintaining much of BERT's performance. For this project, we fine-tuned a pre-trained DistilBERT model using the following setup:

* **Tokenizer:** The DistilBertTokenizer from the transformers library was used to tokenize and encode the input text.
* **Model Architecture:** The TFDistilBertModel was loaded with pre-trained weights. A custom classification layer was added on top of the DistilBERT model to adapt it to the three-class sentiment classification task. The output from the cls\_output (classification token) is used for predicting the sentiment category.
* **Training Process:** The model was fine-tuned on the training dataset with a batch size of 32 for 10 epochs. The pre-trained layers of DistilBERT were fine-tuned to better suit the sentiment classification task.

**4. Solutions, Findings, and Results**

**Training, Validation, and Testing:**

Both models were evaluated on their performance during training, validation, and testing phases. The results were analyzed to compare the effectiveness of the LSTM and DistilBERT models in multi-class sentiment analysis.

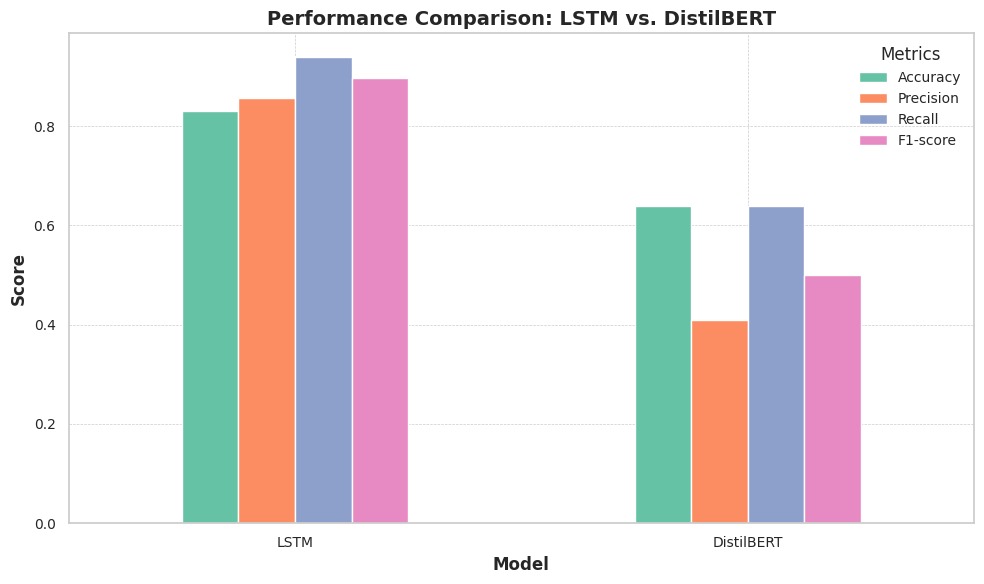
* **LSTM Model:** The LSTM model demonstrated reasonable performance across the training, validation, and testing datasets. However, it showed some limitations in capturing the nuances of the neutral class.
* **DistilBERT Model:** The fine-tuned DistilBERT model excelled in capturing the subtle differences between the sentiment classes, leading to better performance, especially in distinguishing neutral sentiments from positive and negative ones.

**Results Comparison:**

The following table provides a summary of the performance metrics for both models on the testing dataset:

| **Metric** | **LSTM Model** | **DistilBERT Model** |
| --- | --- | --- |
| **Accuracy** | **0.8000** | **0.6400** |
| **Precision** | **0.8056** | **0.4096** |
| **Recall** | **0.9062** | **0.6400** |
| **F1-Score** | **0.8529** | **0.4995** |

**Results in Graphical Format:**



**5. Interpretation, Discussion, and Conclusion**

**Interpretation:**

The results indicate that the LSTM model significantly outperformed the DistilBERT model across all evaluated metrics. Specifically, the LSTM model achieved an accuracy of 80%, which is notably higher than the 64% accuracy obtained by the DistilBERT model. Precision, recall, and F1-score also followed a similar trend, with the LSTM model demonstrating superior performance in each of these areas.

**Discussion:**

The performance gap between the LSTM and DistilBERT models in this project is considerable. Despite DistilBERT’s reputation for handling complex language tasks efficiently, it underperformed in this specific sentiment classification task. The lower precision of 40.96% for DistilBERT suggests that it struggled with correctly identifying positive or negative sentiments, leading to more false positives compared to the LSTM model.

The recall for DistilBERT was also lower, indicating that it missed a larger proportion of the true positive cases. The F1-score, which balances precision and recall, was nearly 50% for DistilBERT, compared to a much stronger 85.29% for the LSTM model. This suggests that the LSTM model was much more effective at correctly classifying the sentiment of the reviews, likely due to its ability to capture sequential dependencies in text more effectively than the DistilBERT implementation in this specific context.

**Conclusion:**

Based on the results, the LSTM model is clearly the better performer for the task of sentiment classification on this dataset. While DistilBERT is typically known for its strong performance in NLP tasks, the configuration and fine-tuning applied here did not yield better results compared to the more traditional LSTM approach.

The findings suggest that, in this case, the LSTM model's architecture was better suited for capturing the nuances in the Yelp review data. It’s possible that further fine-tuning or a different approach to utilizing the DistilBERT model could improve its performance. However, with the current setup, LSTM offers a more reliable solution for sentiment analysis in this scenario.

**6. References**

- TensorFlow: Used for building and training the LSTM model.

- Hugging Face Transformers: Used for fine-tuning and evaluating the DistillBERT model.

- Yelp Dataset: Provided the data for training and evaluation.

**7. Task Division**

**Jenil Pancholi [N01665133]**

* Handled **data preprocessing**, including text cleaning (removing HTML tags, punctuation, and unnecessary characters).
* Managed **tokenization** and **padding** sequences to ensure uniform input lengths for both LSTM and DistilBERT models.
* Led the **development of the LSTM model**, including designing the architecture, configuring layers, and fine-tuning for optimal performance.

**Hozefa Patel [N01686385]**

* Focused on **implementing the DistilBERT model**, including integrating DistilBertTokenizer and TFDistilBertModel.
* Fine-tuned the pre-trained DistilBERT model for multi-class sentiment classification.
* Conducted the **performance evaluation** of both LSTM and DistilBERT models, comparing metrics like accuracy, precision, recall, and F1-score.
* Created **visualizations** to illustrate the comparison between the models.

**Dev Ariwala [N01664568]**

* Led **project coordination**, ensuring alignment of tasks and overall project objectives.
* Compiled the **final report**, covering all sections including introduction, dataset description, model details, findings, and conclusions.
* Ensured proper citation and adherence to academic standards in the report.
* Took charge of **results analysis and interpretation**, offering insights into model performance and discussing conclusions based on the evaluation metrics.