Report

This project focuses on binary sentiment classification using a range of Recurrent Neural Network (RNN) architectures. A total of 16 models were implemented and evaluated, including uni- and bi-directional variants of RNNs and LSTMs, both with and without attention mechanisms. Three types of attention: additive, multiplicative, and concatenative, were integrated to study their impact on model performance and interpretability. To further enhance analysis, a custom heatmap visualizer was developed to display attention weights across all the 16 models, providing insights into which parts of the input text contributed most to the model decisions. The project highlights how architectural choices and attention mechanisms influence classification performance and explainability.

Dataset

- [IMDB Sentiment Dataset]
- 25,000 Train set and 25,000 Test set
- Labels: Positive(1) and Negative(1)

Text Cleaning

- Removed HTML tags, URLs
- Removed "\n" and "\r" escape keys
- Removed all non-alphanumeric characters
- Removed any extra space present
- Lowered the case of the text, and striped any white space present in front and end of the text.

Data Analysis

I intended to use a BERT tokenizer to tokenize the text, while using a custom embedding layer instead of BERT's pretrained embeddings. So, the model is designed to accept a maximum of 512 time steps. Since the BERT tokenizer produces approximately 1.4 tokens per word on average, this corresponds to about $512/1.4 \approx 365$ words. So I have plotted histograms on no. of words on both train set and test set. And I found around 15–16% of the dataset contains samples with at least 365 words, meaning these samples will be truncated to fit the input limit. For the majority of the dataset, however, the model will process the full input sequence without truncation.

Implementation Approach

Implementing 16 separate classes for each model variant is highly inefficient. Instead, my approach is to create just 4 modular classes— `AdditiveAttention`, `MultiplicativeAttention`, `AttentiveRNN`, and `AttentiveLSTM` —which can be flexibly combined to construct all 16 model variants.

AdditiveAttention: Implements the standard Bahdanau attention mechanism. It is compatible with both unidirectional and bidirectional single-layer sequential models.

MultiplicativeAttention: Implements the Luong attention mechanism, where the scoring function ("dot", "general", or "concat") is passed as an initialization argument. It supports both unidirectional and bidirectional single-layer sequential models.

AttentiveRNN: A configurable RNN wrapper that accepts an attention mechanism (or None) and the direction (uni or bi) as arguments. Passing None allows implementation of the vanilla RNN (without attention).

AttentiveLSTM: Similar to `AttentiveRNN`, but based on LSTM layers. It also accepts an attention mechanism (or None) and a direction flag. Passing None results in a standard LSTM without attention.

Results

Model			Accuracy	Macro			Micro			TP	TN	FP	FN	Train Accuracy	Train Time
Type	Dir.	Attention		Precision	Recall	F1	Precision	Recall	F1						
RNN	uni	none	49.14%	49.14%	49.14%	49.13%	49.14%	49.14%	49.14%	1201	1256	1299	1244	53.17%	28m 28s
RNN	bi	none	69.14%	69.14%	69.54%	69.98%	69.14%	69.14%	69.14%	1907	1550	593	950	81.92%	28m 21s
RNN	uni	dot	79.98%	79.98%	79.99%	79.98%	79.98%	79.98%	79.98%	2020	1979	480	521	79.98%	28m 24s
RNN	bi	dot	78.28%	78.28%	78.29%	78.28%	78.28%	78.28%	78.28%	1934	1980	566	520	81.80%	29m 3s
RNN	uni	general	84.18%	84.18%	84.19%	84.18%	84.18%	84.18%	84.18%	2084	2125	416	375	95.67%	30m 54s
RNN	bi	general	79.32%	79.32%	79.66%	79.26%	79.32%	79.32%	79.32%	2117	1849	383	651	80.47%	30m 8s
RNN	uni	concat	83.26%	83.26%	83.27%	83.26%	83.26%	83.26%	83.26%	2059	2104	441	396	97.94%	28m 18s
RNN	bi	concat	88.26%	88.26%	88.41%	88.25%	88.26%	88.26%	88.26%	2285	2128	215	372	99.56%	30m 30s
LSTM	uni	none	59.60%	59.60%	62.18%	57.34%	59.60%	59.60%	59.60%	915	2065	1585	435	61.63%	29m 33s
LSTM	bi	none	86.56%	86.56%	86.57%	86.56%	86.56%	86.56%	86.56%	2181	2147	319	353	98.28%	34m 23s
LSTM	uni	dot	86.40%	86.40%	86.40%	86.40%	86.40%	86.40%	86.40%	2164	2156	336	344	96.01%	29m 56s
LSTM	bi	dot	87.64%	87.64%	87.94%	87.64%	87.64%	87.64%	87.64%	2302	2080	198	420	99.74%	35m 3s
LSTM	uni	general	88.90%	88.90%	88.92%	88.90%	88.90%	88.90%	88.90%	2250	2195	250	305	97.52%	29m 56s
LSTM	bi	general	87.32%	87.32%	87.33%	87.32%	87.32%	87.32%	87.32%	2208	2158	292	342	99.47%	37m 50s

Model			Accuracy	Macro			Micro			TP	TN	FP	FN	Train Accuracy	Train Time
Type	Dir.	Attention		Precision	Recall	F1	Precision	Recall	F1						
LSTM	uni	concat	88.46%	88.46%	88.51%	88.46%	88.46%	88.46%	88.46%	2165	2258	335	242	99.38%	30m 7s
LSTM	bi	concat	89.12%	89.12%	89.17%	89.12%	89.12%	89.12%	89.12%	2271	2185	229	315	99.62%	38m 47s

Best Model Performance

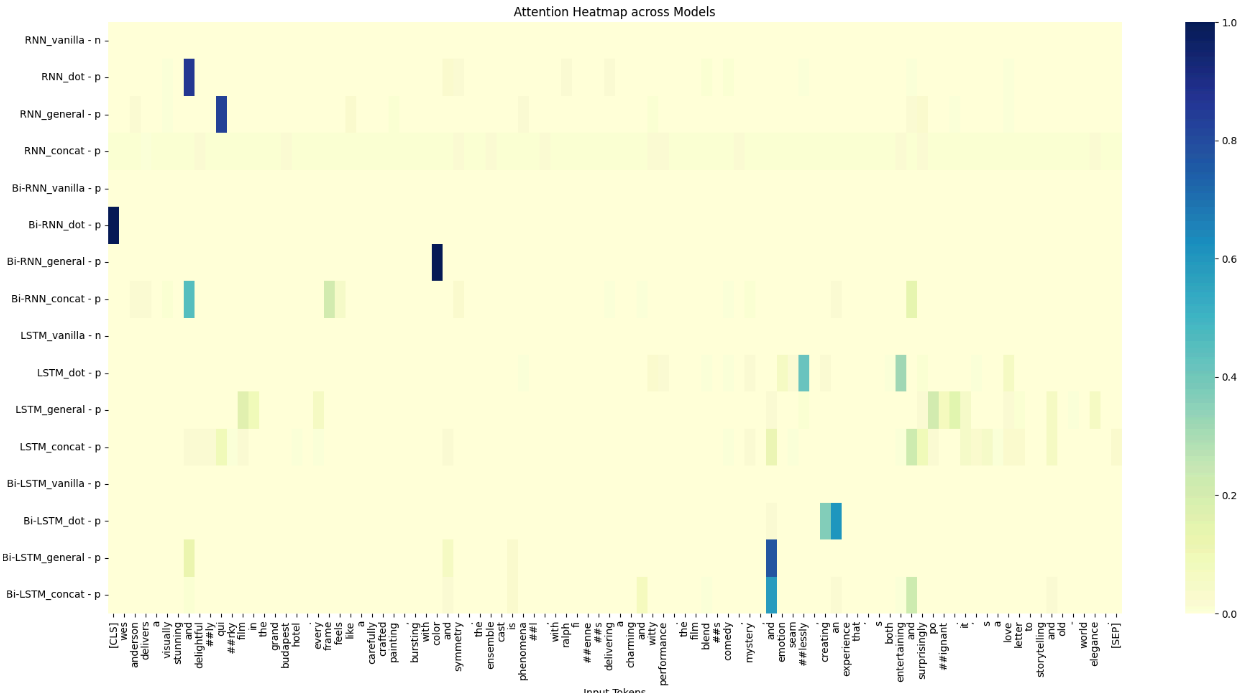
The best-performing model in this project was a **Bidirectional Long Short-Term Memory (BiLSTM)** network integrated with a **concatenative attention mechanism**. This architecture processes input sequences in both forward and backward directions, allowing it to capture richer contextual information. The concatenative attention layer enhances interpretability by learning to focus on the most relevant parts of the input when making predictions. This model achieved a **test accuracy of 89.12%**, outperforming all other RNN and LSTM variants evaluated.

Attention Weights Visualization

For a Positive movie review on "The Grand Budapest Hotel" such as

> Wes Anderson delivers a visually stunning and delightfully quirky film in *The Grand Budapest Hotel*. Every frame feels like a carefully crafted painting, bursting with color and symmetry. The ensemble cast is phenomenal, with Ralph Fiennes delivering a charming and witty performance. The film blends comedy, mystery, and emotion seamlessly, creating an experience that’s both entertaining and surprisingly poignant. It’s a love letter to storytelling and old-world elegance.

The Heatmap plot for attention weights for all the 16 models



For a Negative movie review on "The Last Airbender" such as

> As a fan of the original animated series, *The Last Airbender* was a crushing disappointment. The characters felt flat and unrecognizable, the dialogue was stiff, and the plot rushed through key moments with no emotional payoff. The bending effects were underwhelming, and the lack of chemistry between the cast made everything feel forced. It completely missed the charm, depth, and spirit of the source material, turning a rich world into a dull, joyless spectacle.

The Heatmap plot for attention weights for all the 16 models