Transformers For NLP

CSE 705 - Recent Advances in Deep Learning and Reinforcement Learning

Instructor: - Prof. Alina Vereshchaka

Presented by: - Rishi Joshi, Aastha Sood

University at Buffalo The State University of New York

What is a transformer?

What is a transformer?

- Consists of an encoder and a decoder that can process words simultaneously.
- Consists of an attention mechanism without using RNNs.
- Unlike RNNs, transformers process the entire input at once.
- Introduced in 2017 by a team at Google Brain.
- Quickly replacing LSTMs and RNNs which are much slower to train.
- Led to the development of pre-trained models like BERT and GPT

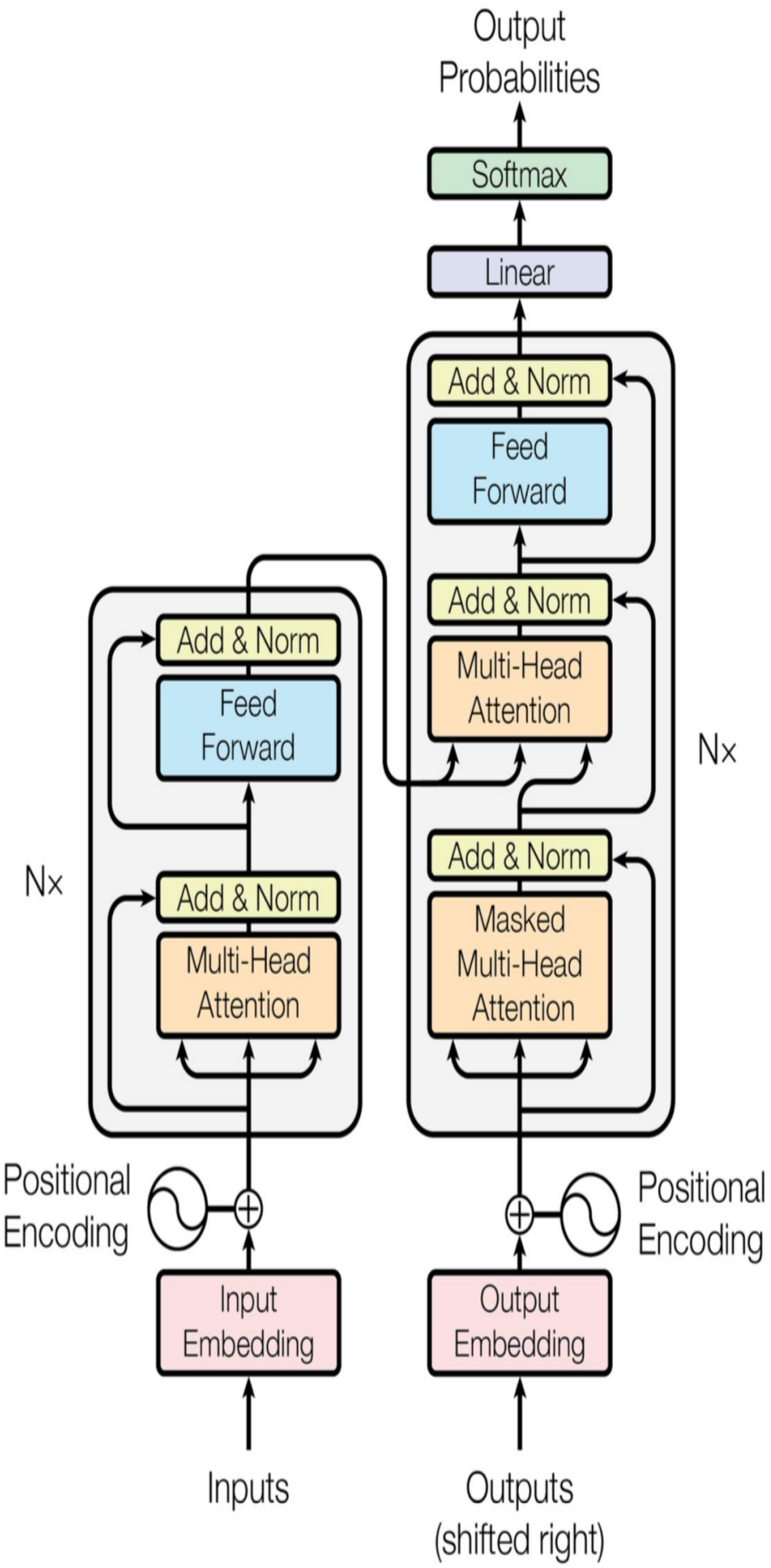


Figure 1: The Transformer - model architecture.

University at Buffalo The State University of New York

NLP Task and Dataset

NLP Task and Dataset

NLP Task: -

- Perform sentiment analysis on tweets using the BERT transformer and compare the results with LSTM.

Dataset: -

- Financial Tweets
- Contains 5322 unique tweets divided into positive, negative and neutral.

What is LSTM?

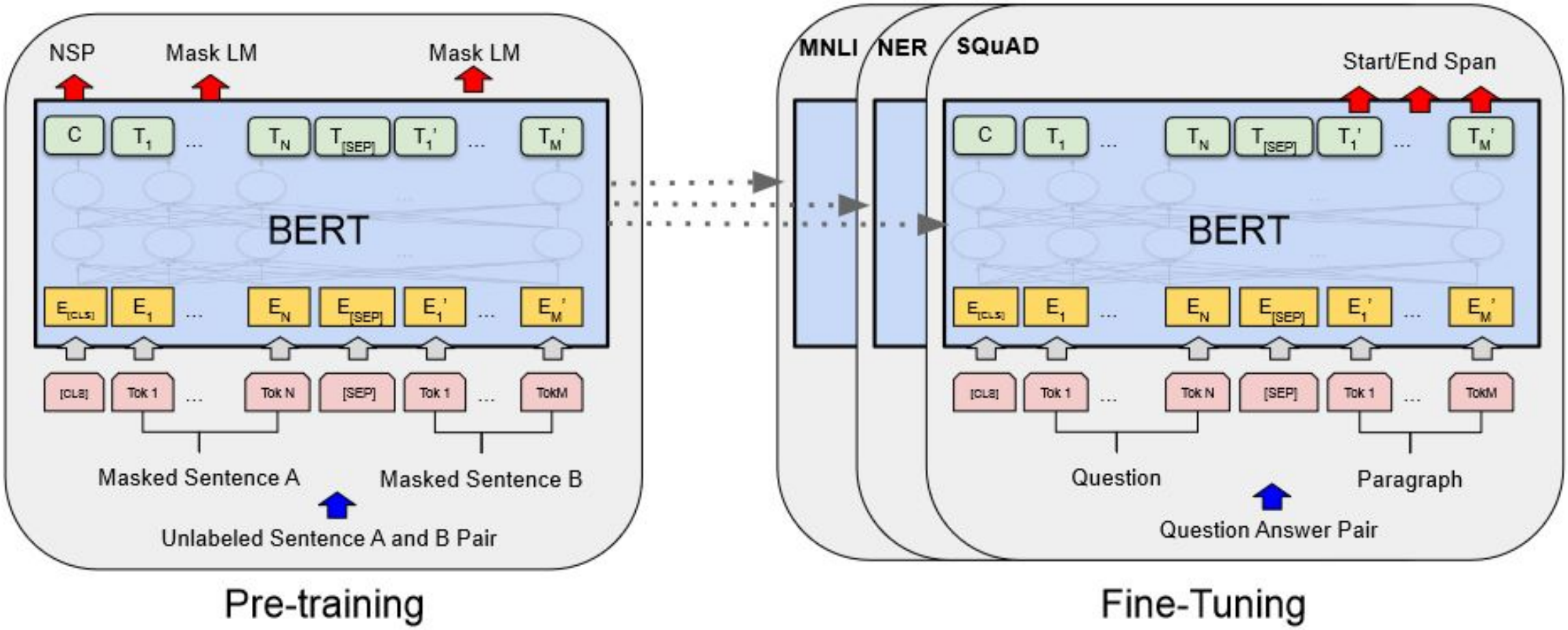
What is LSTM?

- They are a type of RNN, capable of learning long-term dependencies.
- They are explicitly designed to avoid the long-term dependency problem
- LSTMs have a chain like structure, but the repeating module has a different structure.
- Instead of having a single neural network layer, there are four, interacting in a very special way.

University at Buffalo The State University of New York

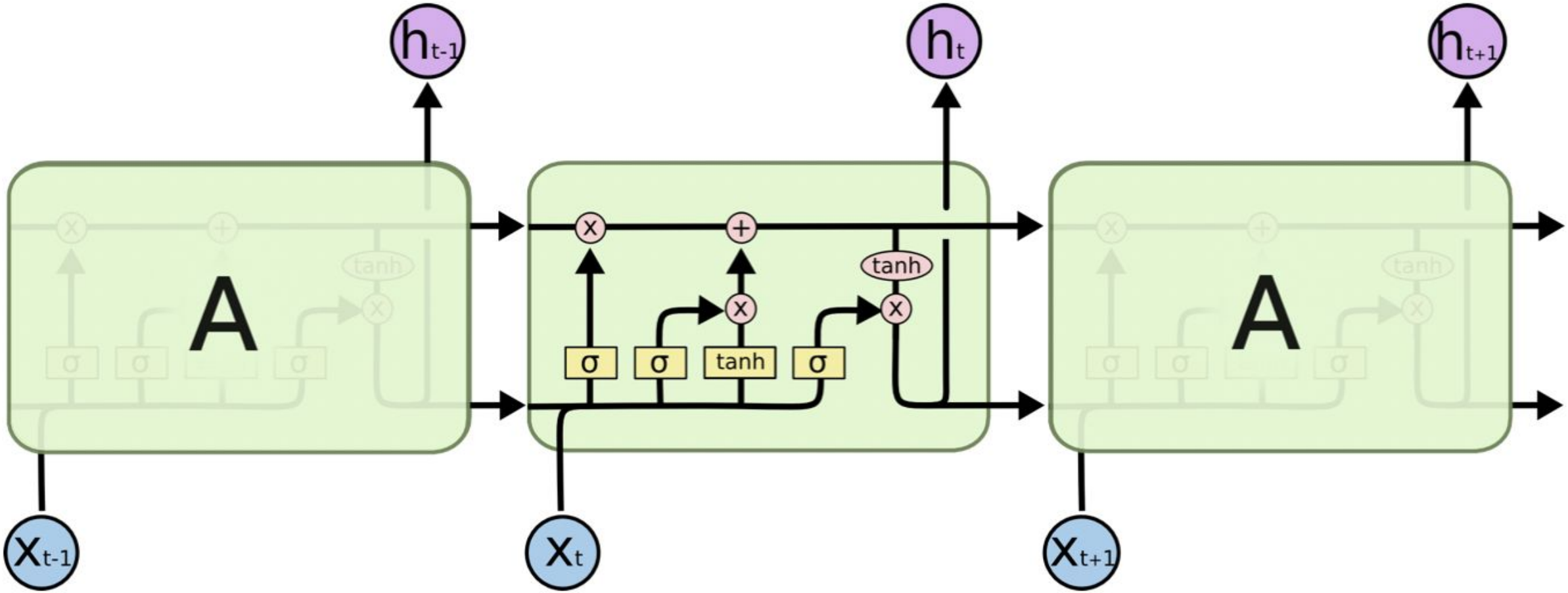
BERT - Bidirectional Encoder Representations from Transformers

BERT Architecture



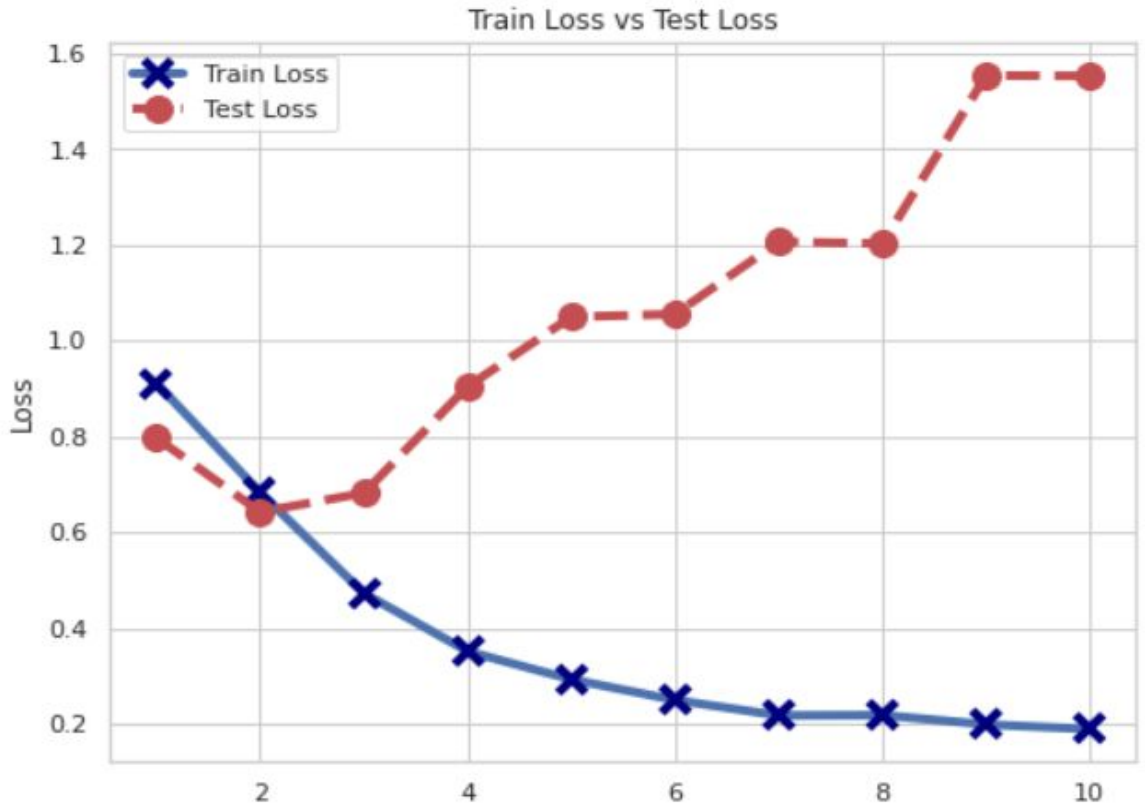
LSTM - Long Short Term memory

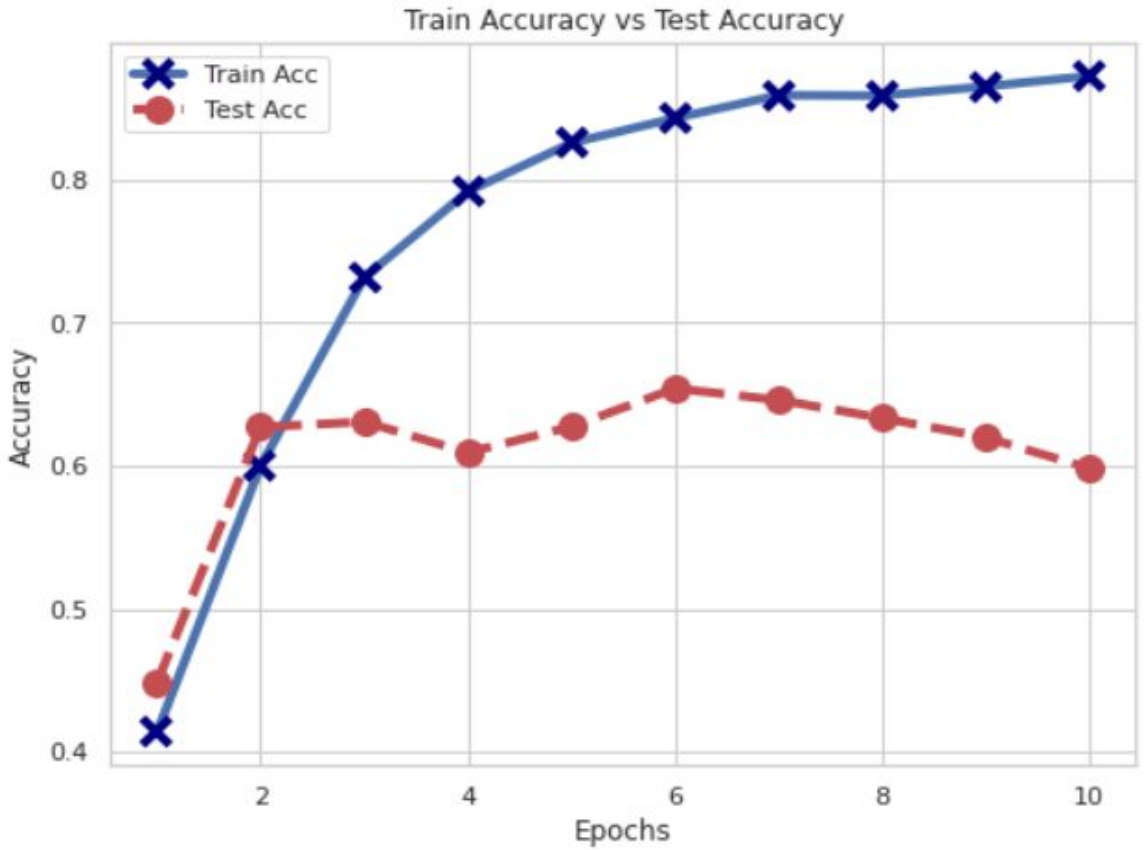
LSTM Architecture



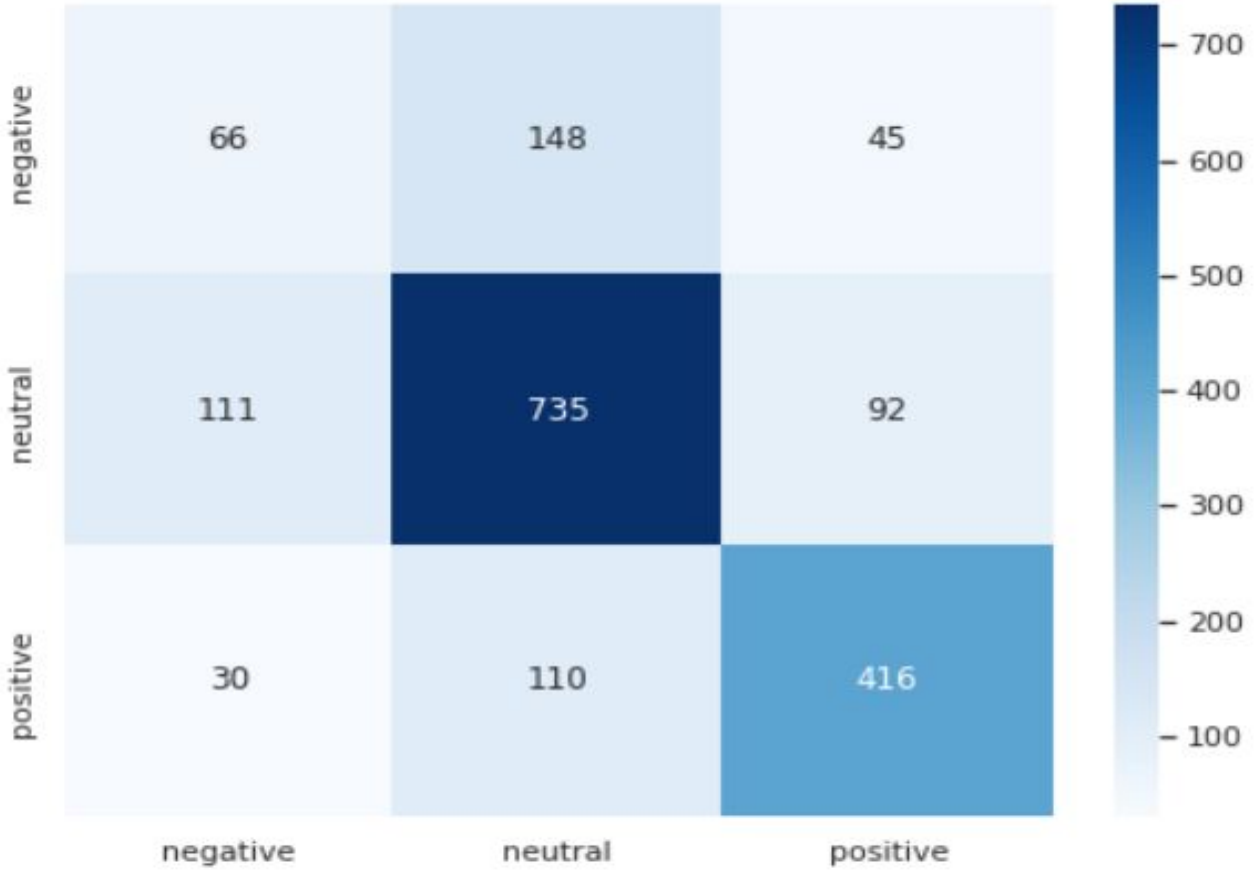
BERT Results

Training Loss / Accuracy vs Testing Loss / Accuracy



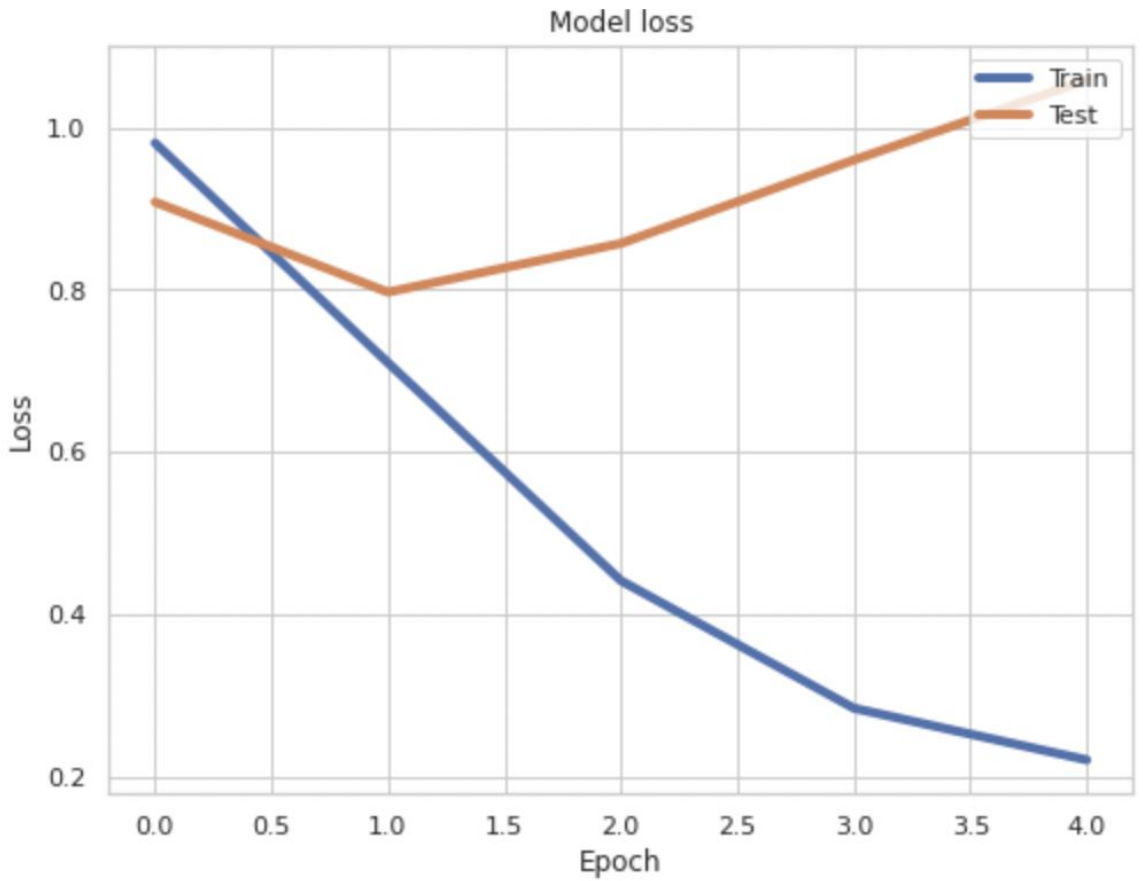


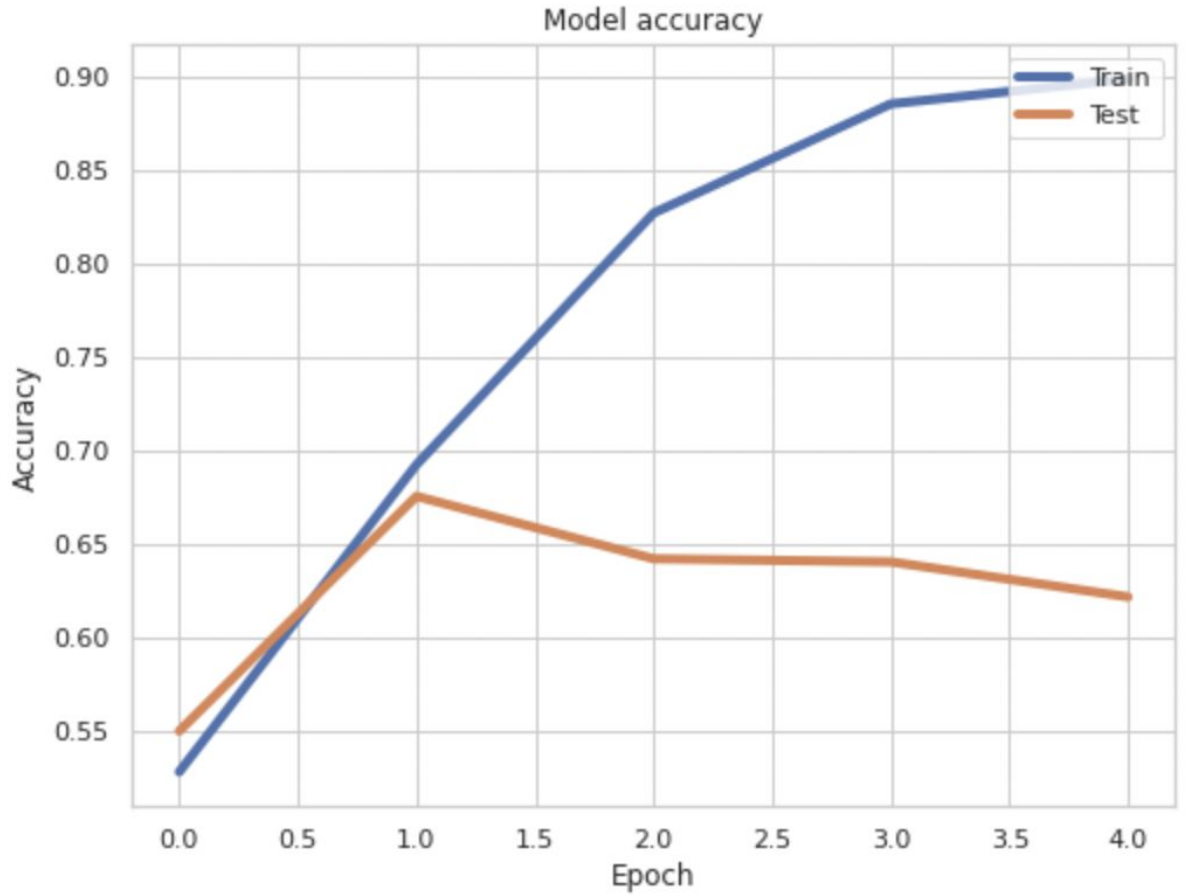
Confusion Matrix



LSTM Results

Training Loss / Accuracy vs Testing Loss / Accuracy





Contribution Summary

Team Member	Presentation / Project Part	Contribution (%)
Rishi Joshi	BERT	50%
Aastha Sood	LSTM	50%

References

References

- https://towardsdatascience.com/understanding-lstm-and-its-quick-implementation-in-keras-for-sentiment-analysis-af410fd85b47
- https://colah.github.io/posts/2015-08-Understanding-LSTMs/
- https://neptune.ai/blog/how-to-code-bert-using-pytorch-tutorial
- https://mccormickml.com/2019/07/22/BERT-fine-tuning/#4-train-our-classification-model
- Alex Graves, et al., Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures, 2005.
- CodeEmporium - Transformer Neural Networks - EXPLAINED! (Attention is all you need) - https://www.youtube.com/watch?v=TQQlZhbC5ps
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, et al., Attention Is All You Need, 2017

Thank You!