Can Machines Understand Emotion?

DTSA-5511 Final Project

Genesis Zhong

August 16th, 2023

Introduction

- Natural Language Processing
- Sentiment Analysis

Large Movie Review Data Set

50,000 MOVIE REVIEWS ON IMDB

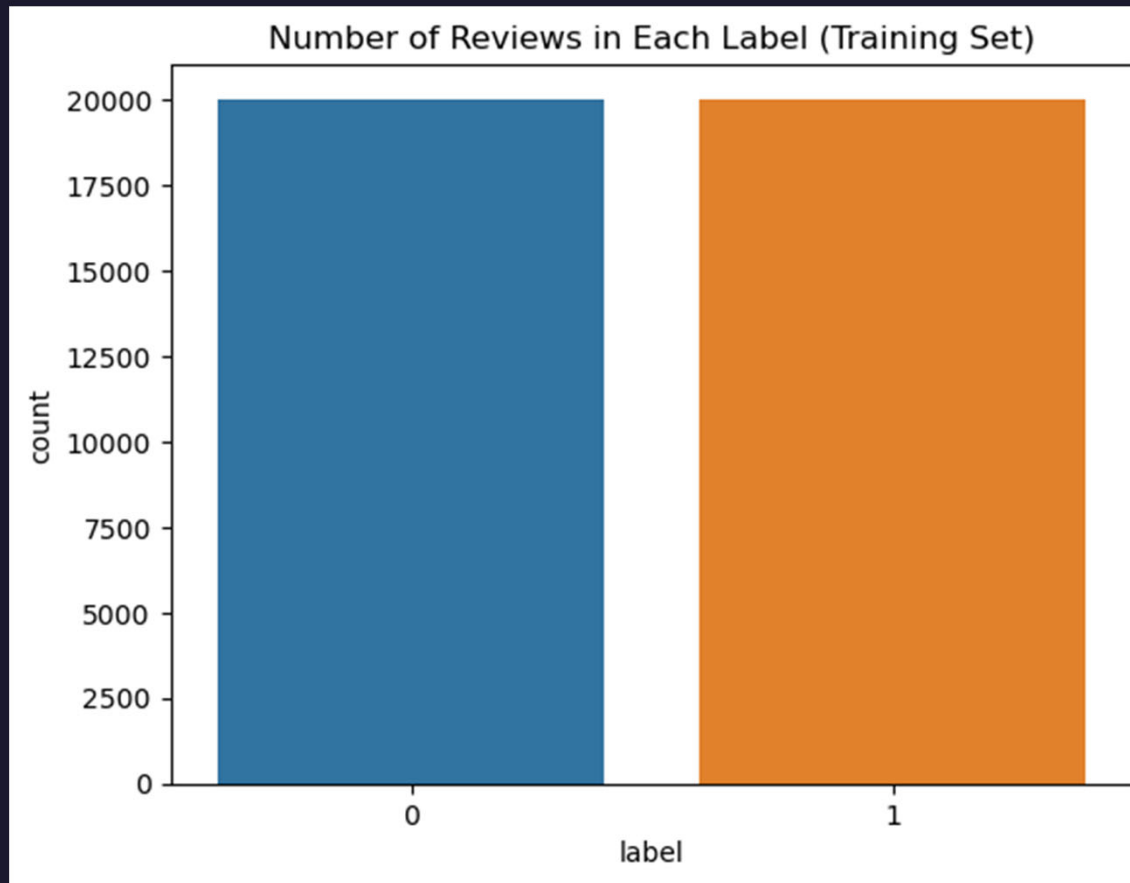https://ai.stanford.edu/~amaas/data/sentiment/
https://www.kaggle.com/datasets/columbine/imdb-dataset-sentiment-analysis-in-csv-format

Data set	Filename	Rows
Train	Train.csv	40000
Test	Test.csv	5000
Validation	Valid.csv	5000

Column	Description
text	Movie review texts
label	Sentiment label (1: positive, 0: negative)

EDA
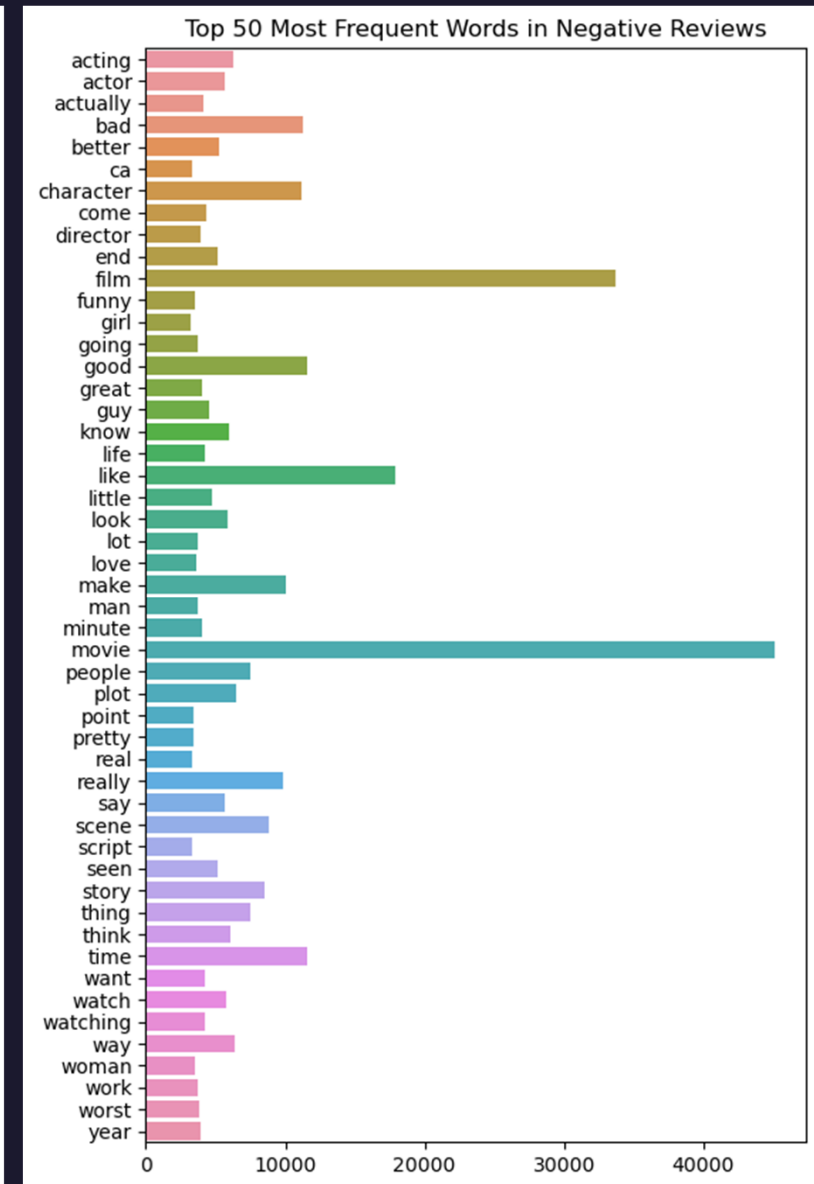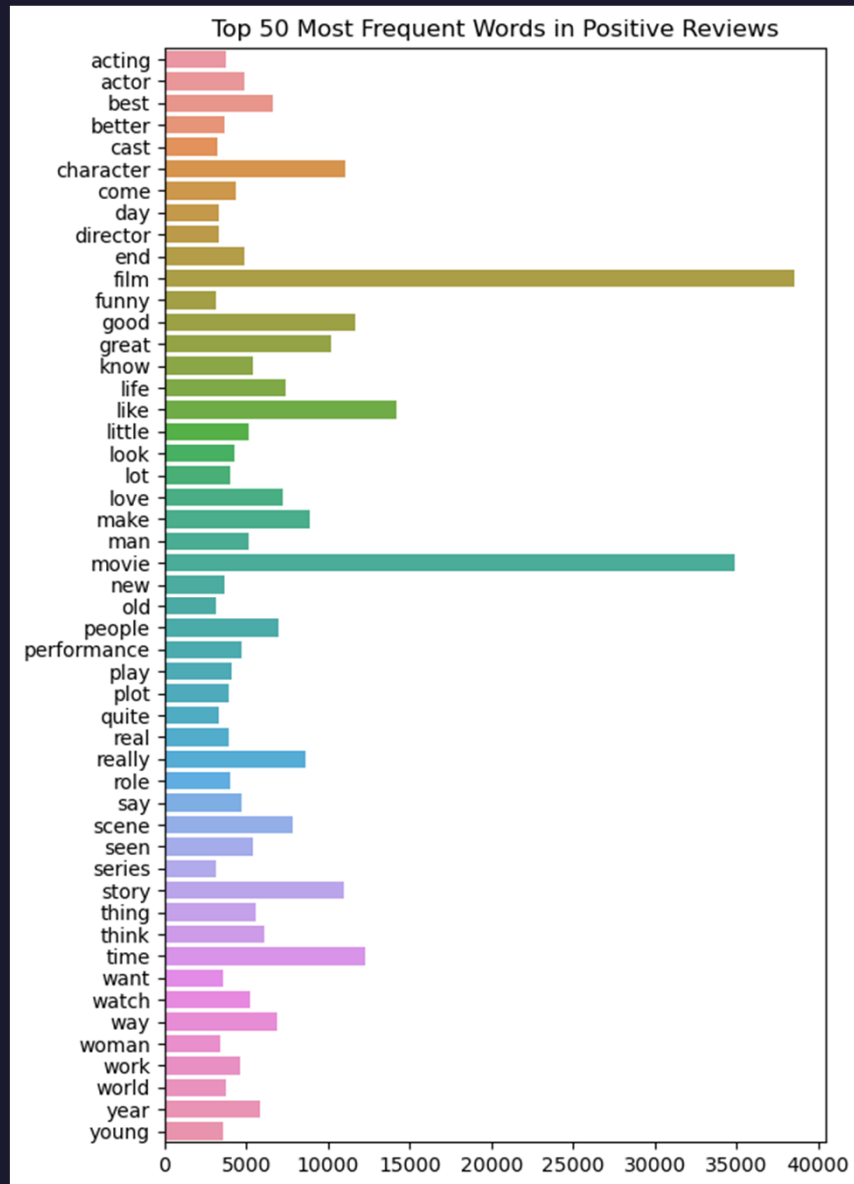
- Distribution of Label

EDA

- Distribution of Length

Distribution of Review Length in # of Words

EDA

- Word Cloud

Positive reviews

Negative reviews

EDA

- Top 50 Words

Top 50 Most Frequent Words in Positive Reviews
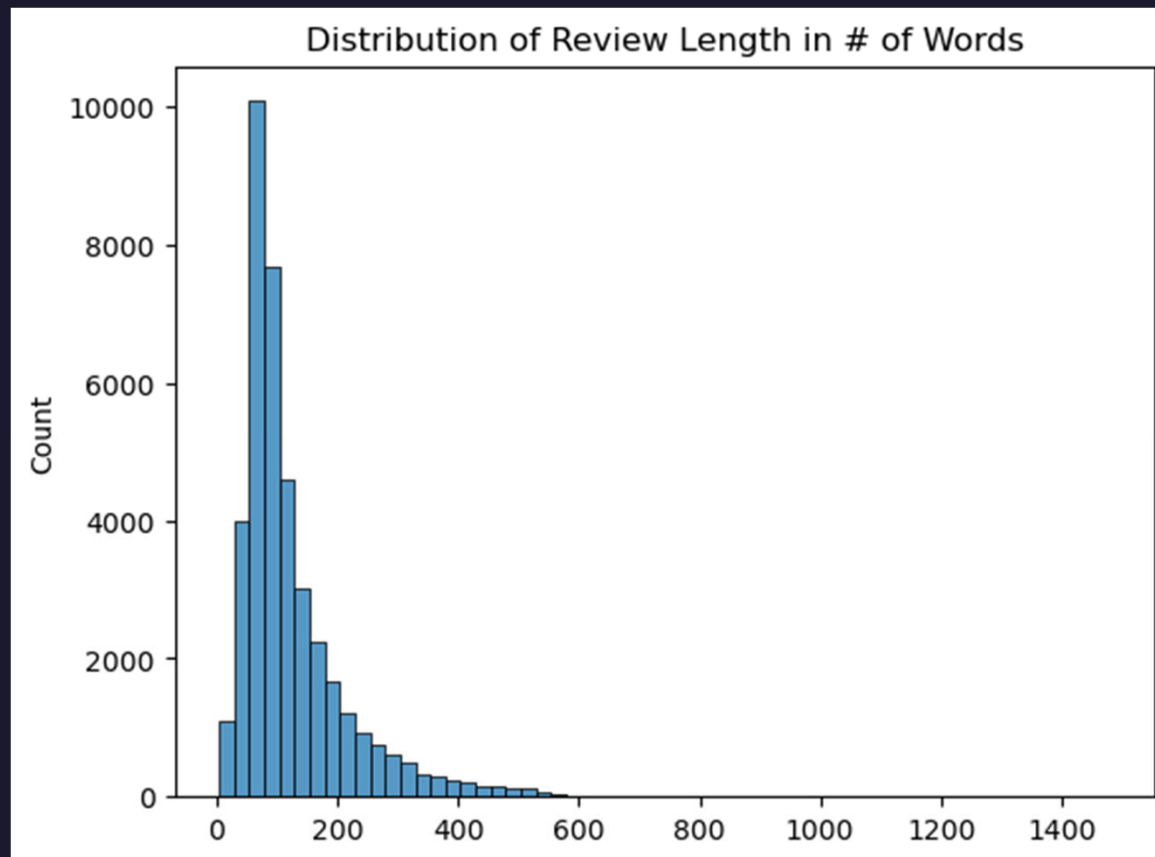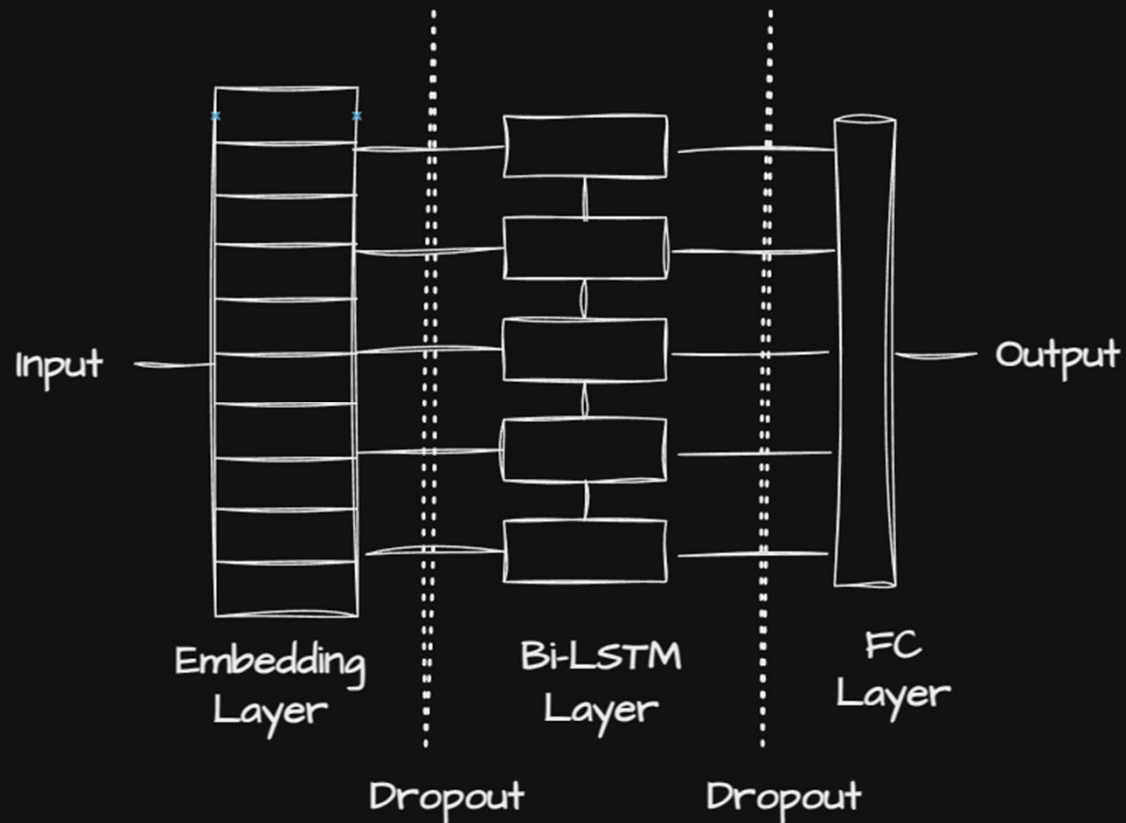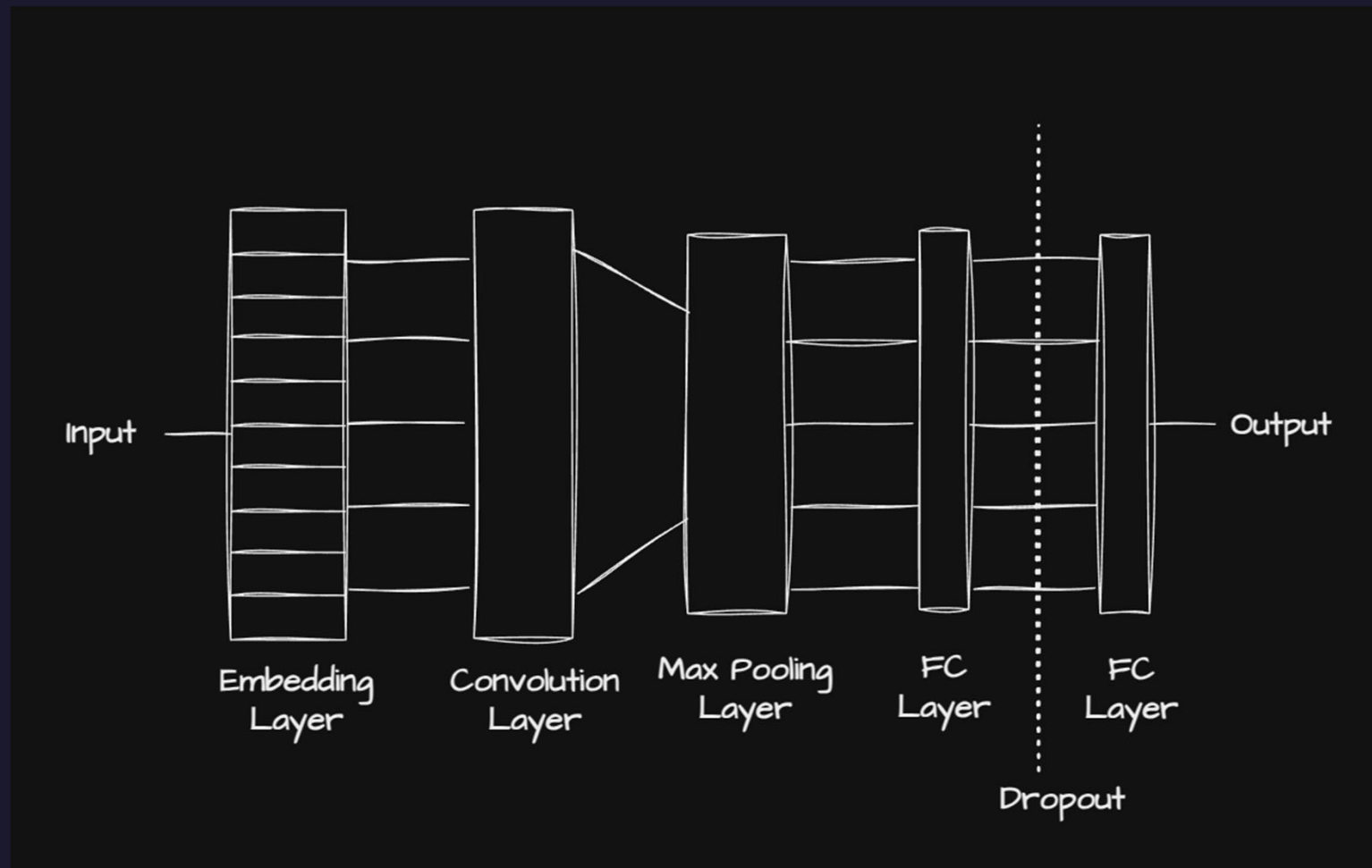
Top 50 Most Frequent Words in Negative Reviews

Setup

- Dataset: tf.data.Dataset

- Vectorization: tf.keras.layer.TextVectorization

- Optimizer: AdamW

- Callbacks: EarlyStopping
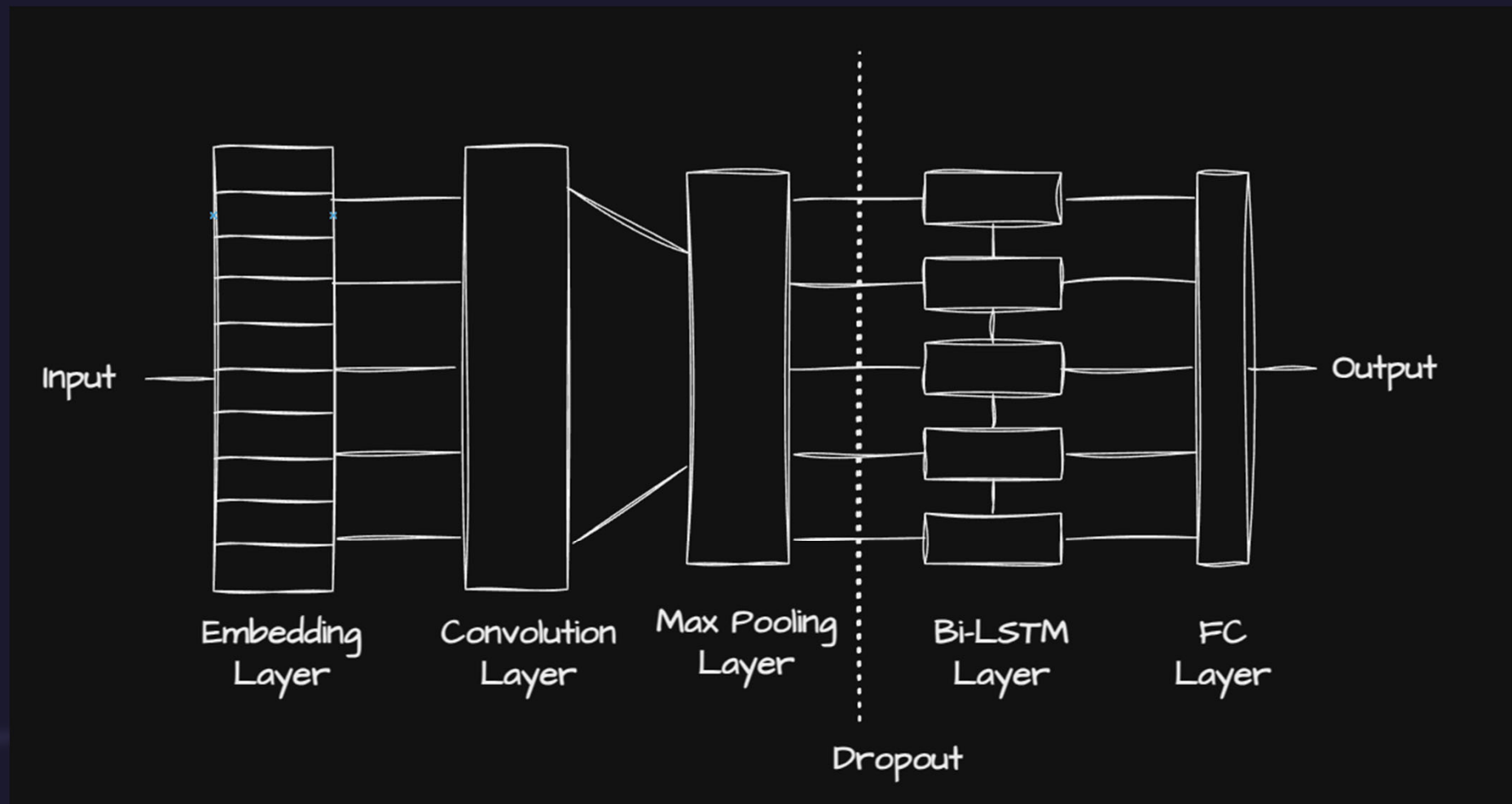
Model 1: LSTM

Model 2: CNN

Model 3: CNN_LSTM
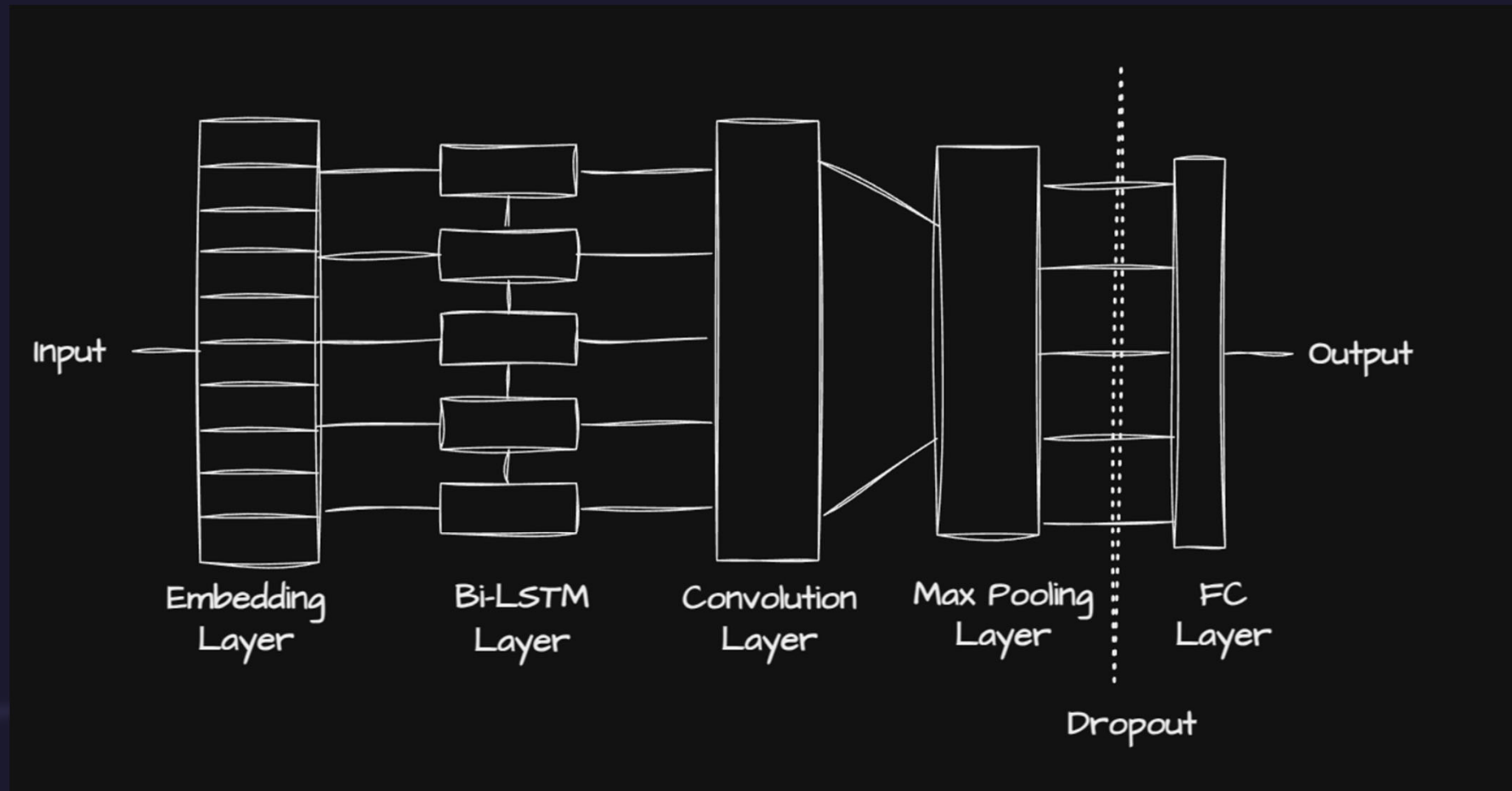
Model 3: CNN_LSTM
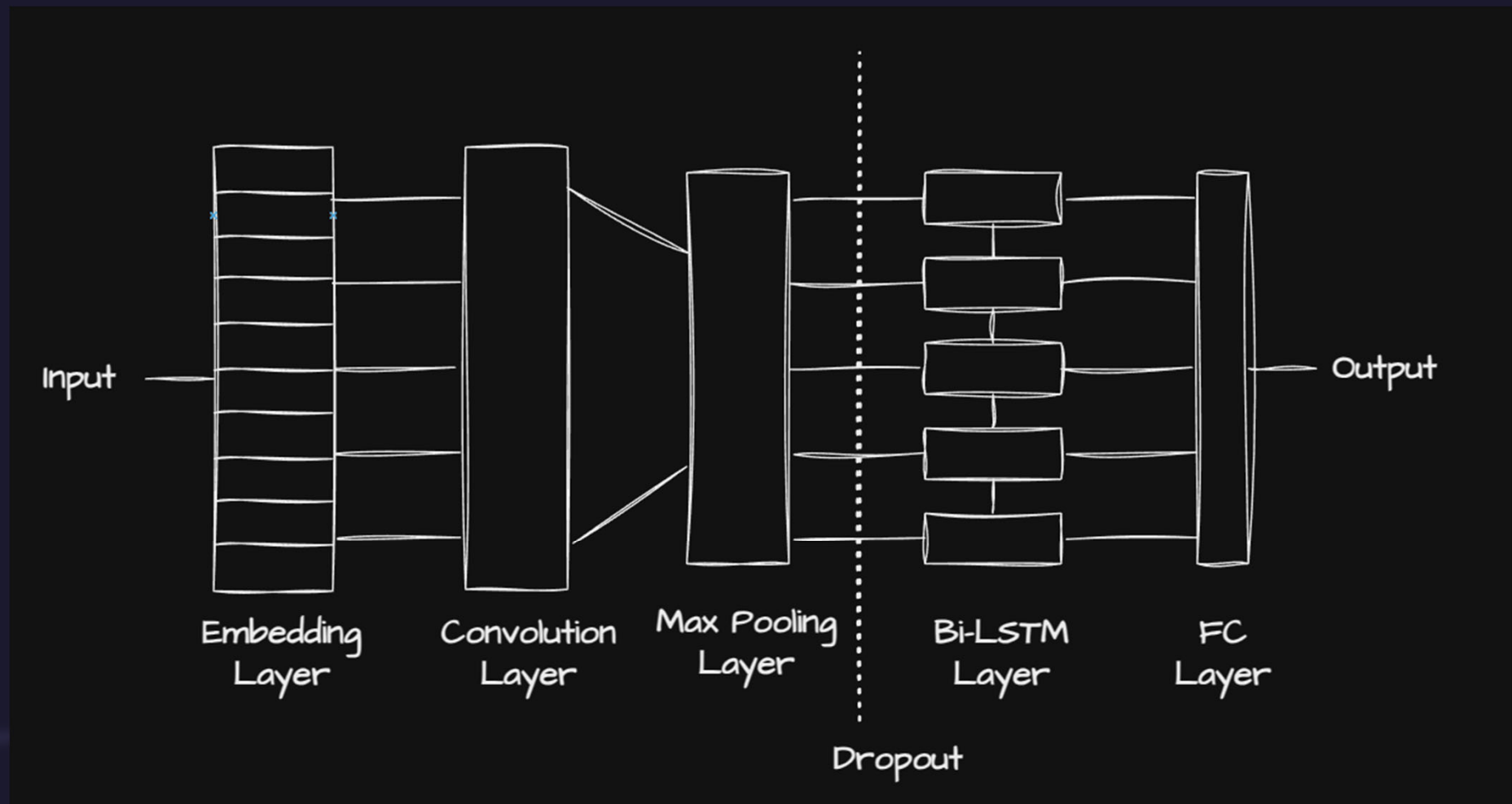
Model 4: LSTM_CNN

Results and Discussion

Model	Architecture	Validation Accuracy	Test Accuracy
lstm	Bi-LSTM	0.8166	0.8264
cnn	1D Convolution	0.8140	0.8252
cnn_lstm	1D Conv + Bi-LSTM	0.8086	0.8180
lstm_cnn	Bi-LSTM + 1D Conv	0.8146	0.8218

Further Modeling: LSTM 2
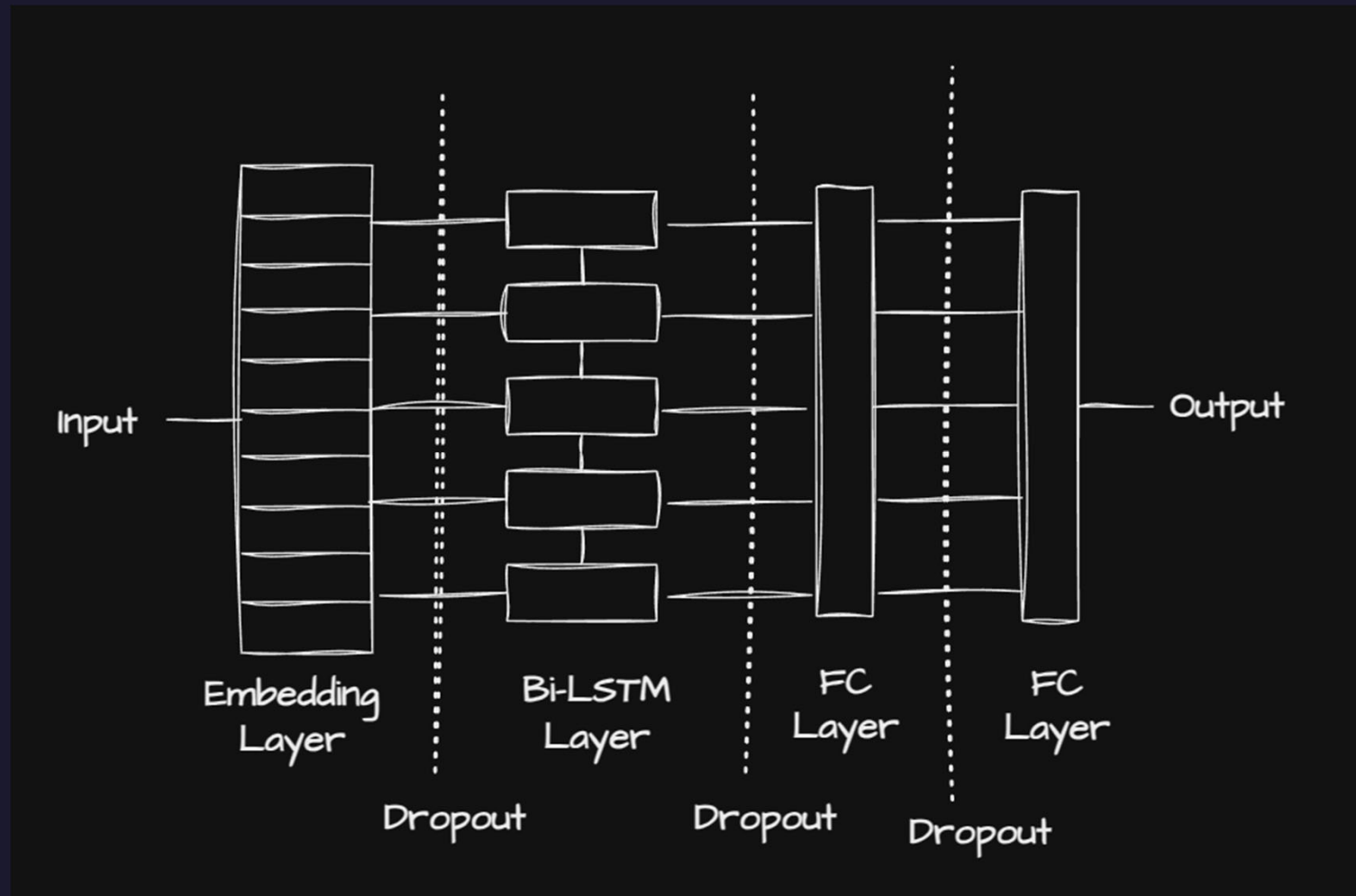
Further Modeling: LSTM 3
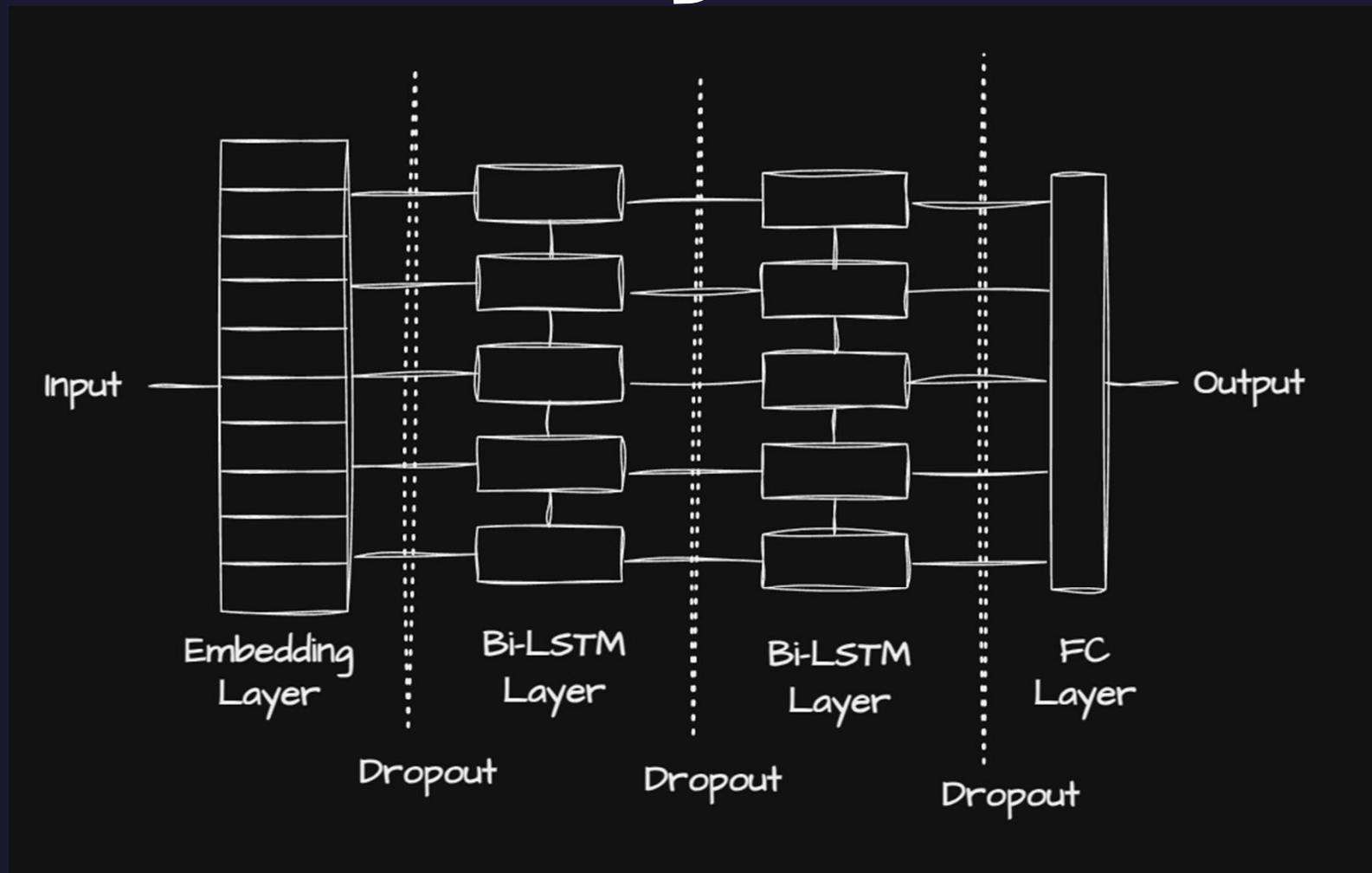
Results and Discussion

Model	Architecture	Validation Accuracy	Test Accuracy
lstm	Bi-LSTM	0.8166	0.8264
lstm2	Bi-LSTM + additional FC layer	0.8162	0.8276
lstm3	Stacked Bi-LSTM	0.8158	0.8254

Further Analysis

- Looking at the mis-labeled reviews

| 3912 | Isabelle Huppert portrays a talented female piano teacher who is staid, unfriendly and distant in public, and bitter towards her students. Privately, she seethes with violence and frustration, and her sexual life is solitary and perverse. She lives with her overbearing mother, who obsessively drives her to become noticed (and so advance in life) as a talented pianist. The key to the characters of both mother and daughter is 'obsession.' These characters cannot change their impulses anymore than a rabbit caught in headlights can avoid death.

The piano teacher meets a young, attractive, talented pianist who from the beginning is attracted to her. They start a relationship in the most unconventional way, but from the outset she makes perverse and violent terms that he must perform on her, which sickens him enough to want to terminate the relationship before it has really begun.

The film ties itself to the female lead. Isabelle Huppert amazes with a brutal, completely convincing performance as the piano teacher. She cleverly shows a woman who is drawn to beauty and perversion, but her violence is fed by her perverted impulses. As a film that is so character driven, you know it would not work half as well, had she acted poorly.

This is powerful, intelligently acted, and intelligently and sensitively adapted from the novel. The camera work also suits the film. There are what I can only think to call, framing shots where the director holds a scene and forces the eye to dart about. This is done extremely effectively against a blank bathroom wall, and is a further testament to the director's mastery.

Expect to be disturbed and sickened by this film. But, be brave - have the guts to go and see it. This is a very private look into essentially one person's life, but do not expect to be entertained in the Hollywood sense....there are no car chases in this film! | 1 | 0 |
| 858 | The "all I have is 5 dollars and my wedding ring..." scene was a riot. I also guffawed at the scene in the bar where Hyde snorted the horse radish and flipped the bird to the Japanese guy and said "Pearl Harbor buddy". I think my IQ is higher than 115, but I'm not sure because I can't count that high.

Funny thing, this 10 line requirement. Seems as though they would bash you for making your comments too long, not too short. I hope I don't make it to 1,000 words before I get to line 10.

I'm still two lines short. Pardon me while I think and drink, or drink and think. It depends on whether my hands are faster than my mind. Good, I think I've made it to the 10 line limit. Thanks for reading! | 1 | 0 |

Conclusion

- Similar performance from all models (~82% accuracy)

- Limiting factor:

  - Confusing reviews in the data set

  - 2011 data set, modern neural network

- Future work:

  - More challenging data set

  - Pre-trained models like BERT

Thank you