Text Sentiment Analysis

Deep Learning Semester Project

Alexandros Chantzaras

Nikolaos Kyriakakis

Objectives

Experiment with different textual representations

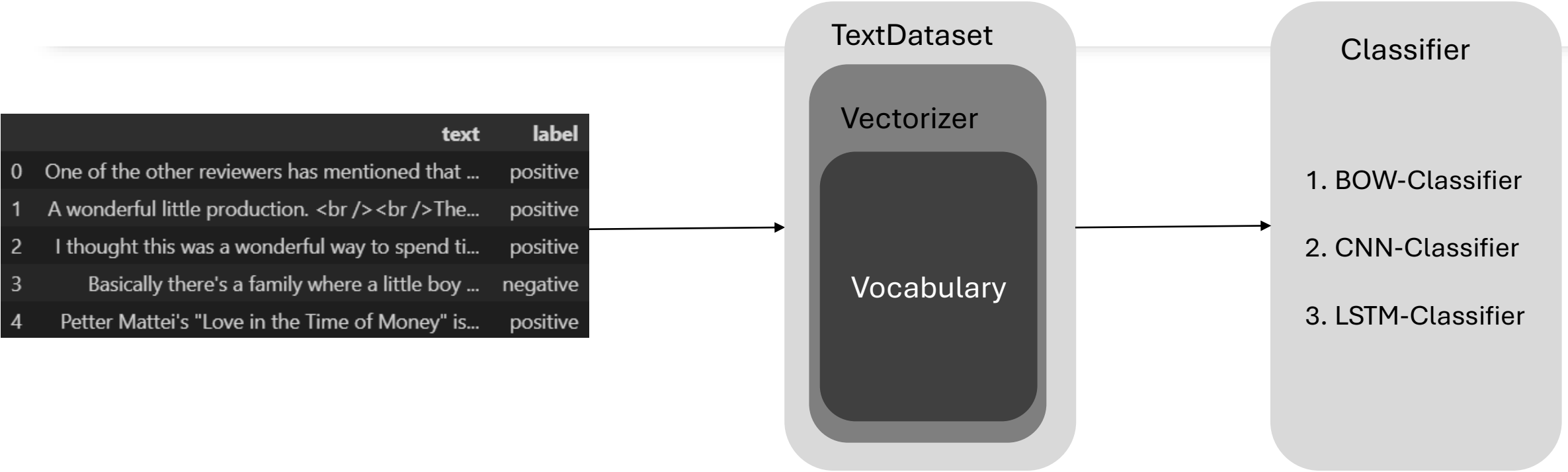
Apply text pre-processing techniques

Use pre-trained word embeddings

Train and evaluate different types of neural networks on textual data

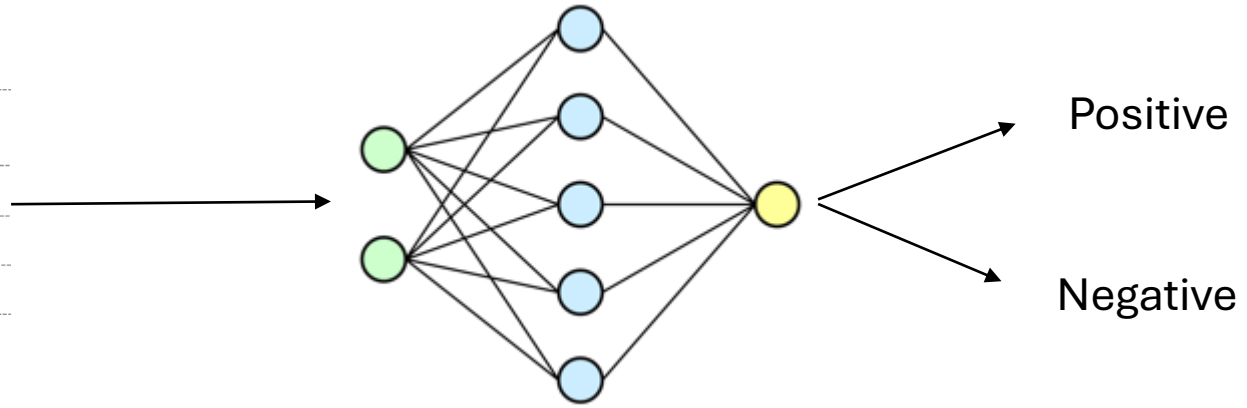
Measure the performance of those models on test data that were never seen before by the model

Project Overview



Bag Of Words Classifier

	1 This	2 movie	3 is	4 very	5 scary	6 and	7 long	8 not	9 slow	10 spooky	11 good
Review 1	1	1	1	1	1	1	1	0	0	0	0
Review 2	1	1	2	0	0	1	1	0	1	0	0
Review 3	1	1	1	0	0	0	1	0	0	1	1



CNN Classifier

sequence before padding

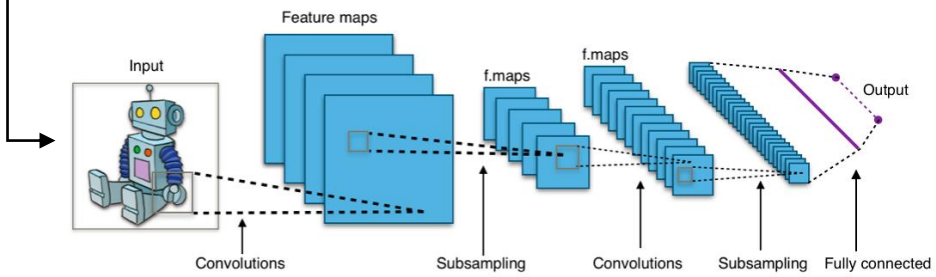
[21, 4, 2, 12, 22, 23, 13, 2, 24, 6, 2, 7, 2, 4, 25],
[2, 26, 7, 27, 14, 9, 1, 4, 28],
[15, 25, 1, 29, 6, 15, 30],
[1, 16, 17, 27, 30, 1, 5, 2],
[31, 2, 28, 6, 32, 9, 33],

sequence after padding
(padding and truncate in front/pre)

[23, 13, 2, 24, 6, 2, 7, 2, 4, 25],
[0, 2, 26, 7, 27, 14, 9, 1, 4, 28],
[0, 0, 0, 15, 25, 1, 29, 6, 15, 30],
[0, 0, 1, 16, 17, 27, 30, 1, 5, 2],
[0, 0, 0, 31, 2, 28, 6, 32, 9, 33],

MAX_SEQUENCE_LENGTH = 10

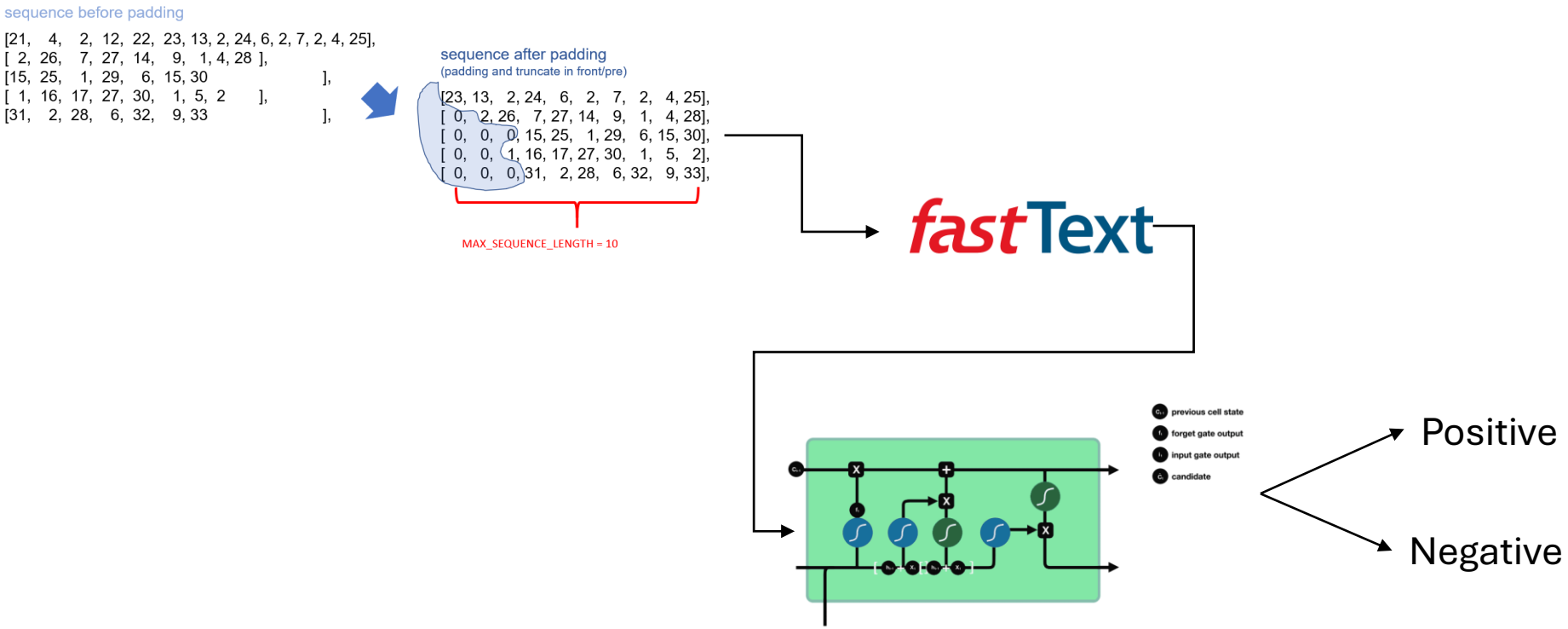
fastText



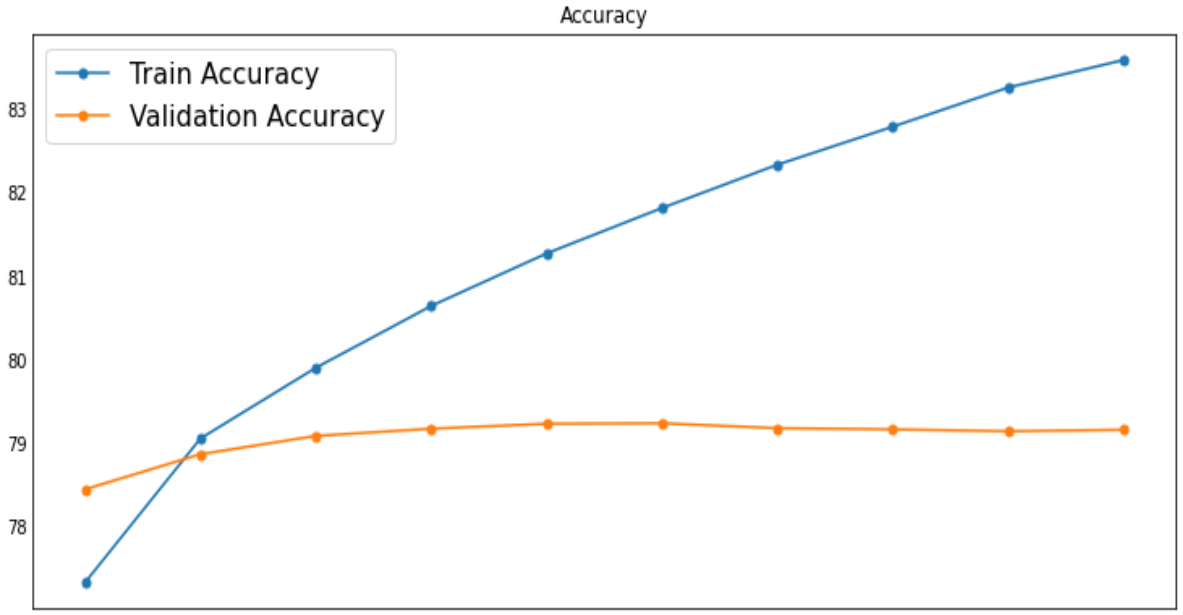
Positive

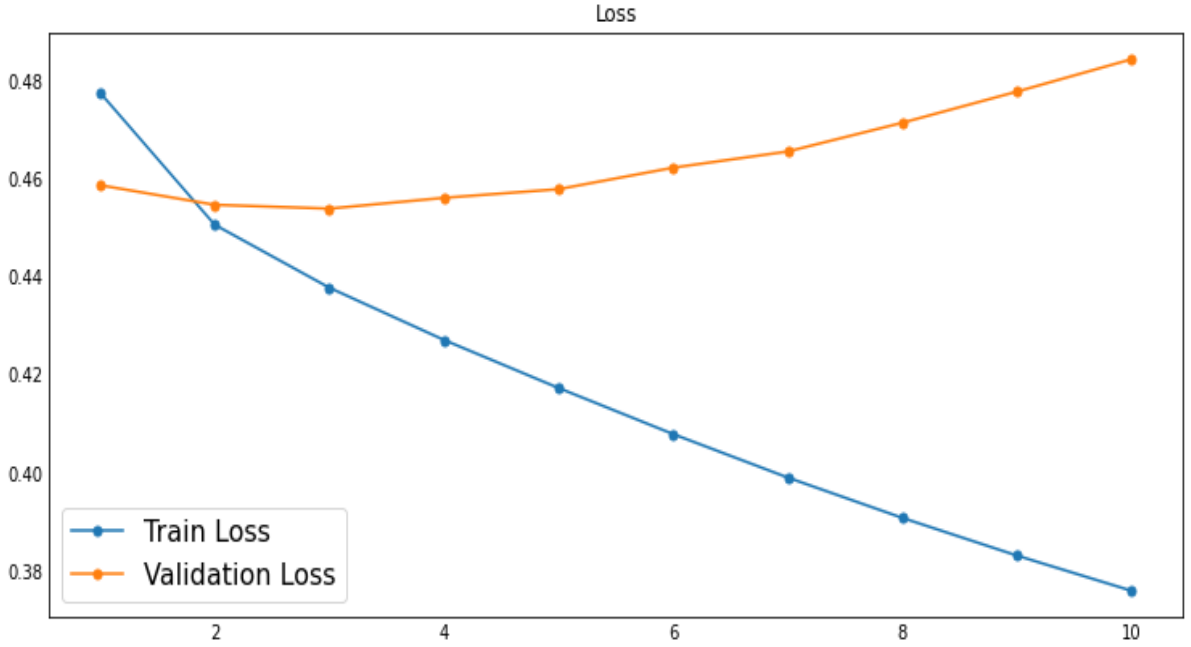
Negative

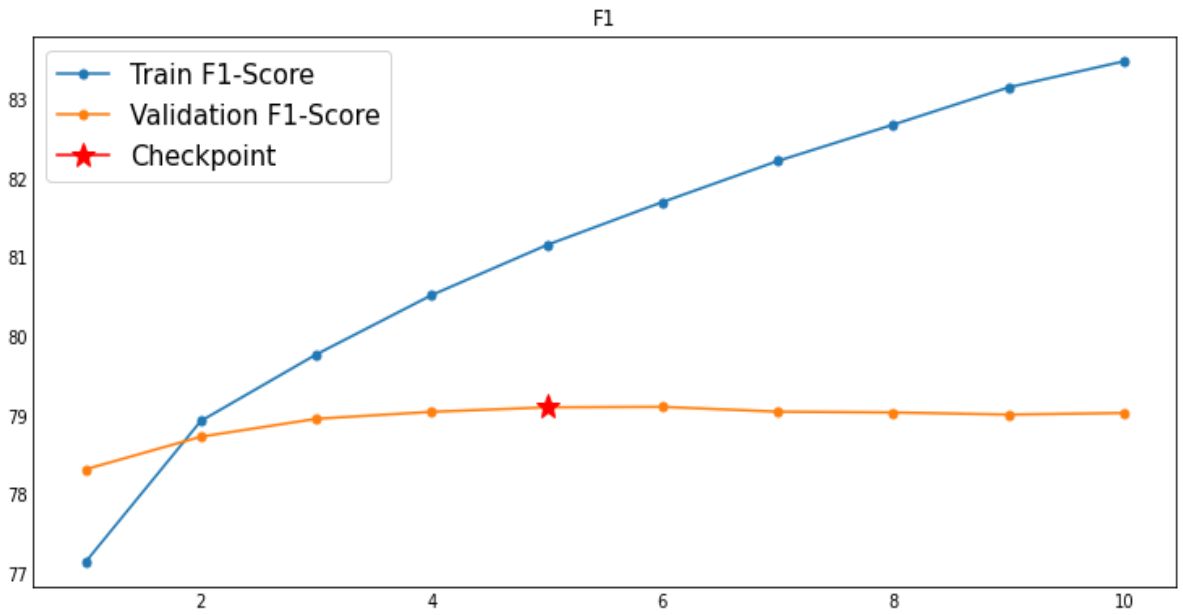
LSTM Classifier



Bag of Words - NN Results

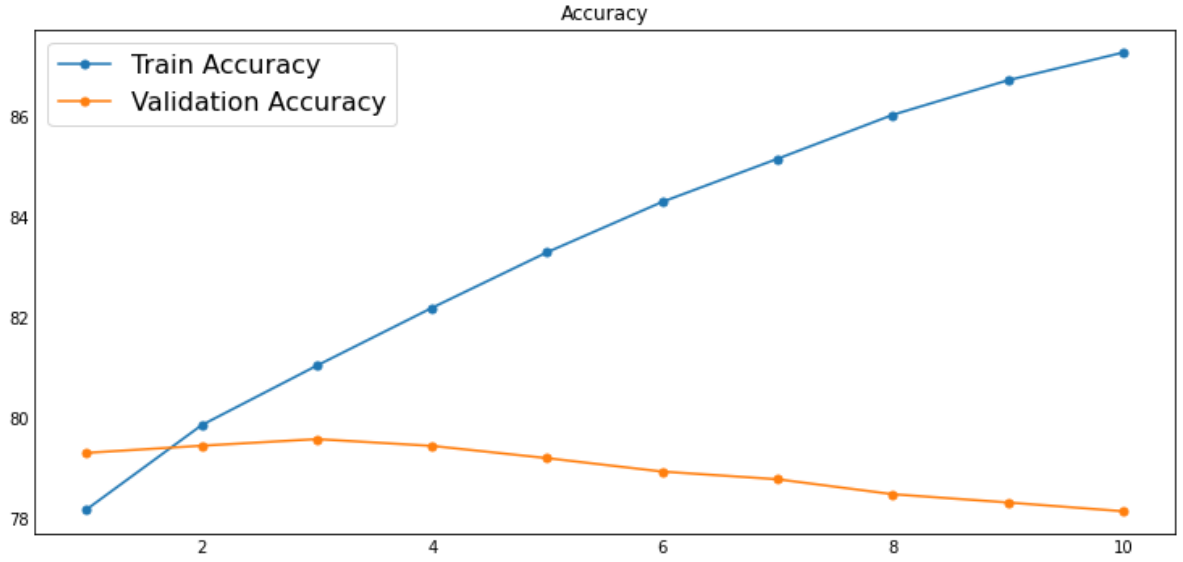


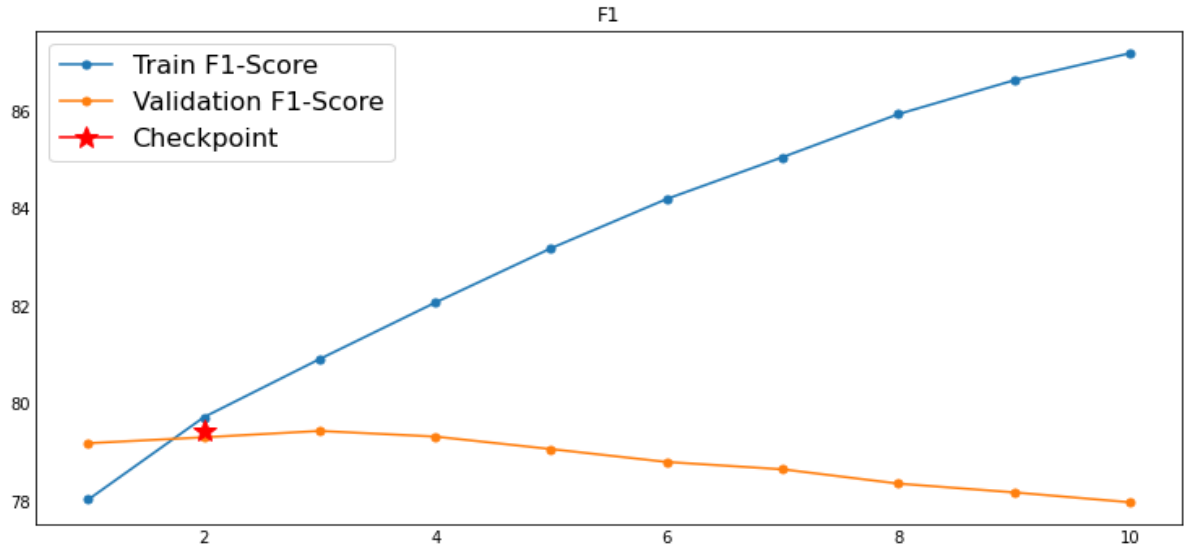


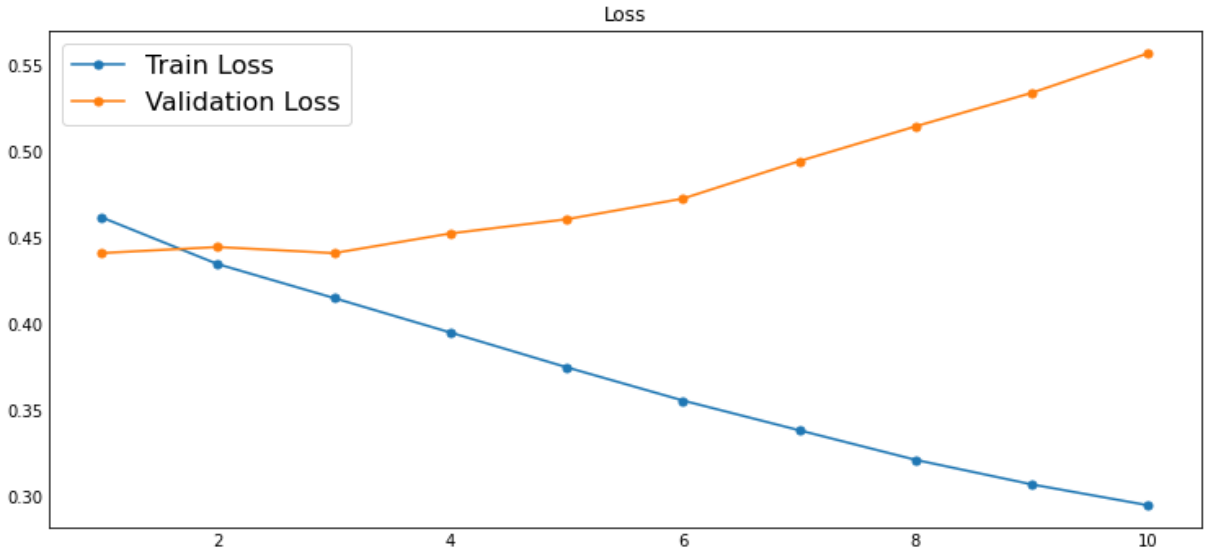


Dataset	Tweets	IMDB
Accuracy Score	79.14%	51.09%
F1 - Score	79.00%	39.94%
Loss	0.45	5.51

CNN Results

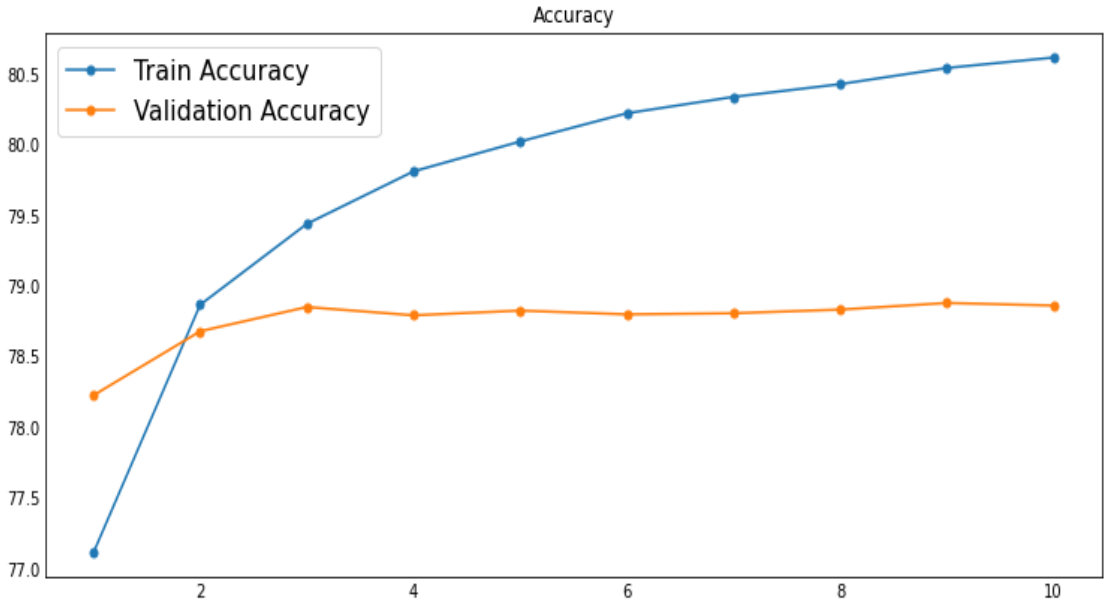


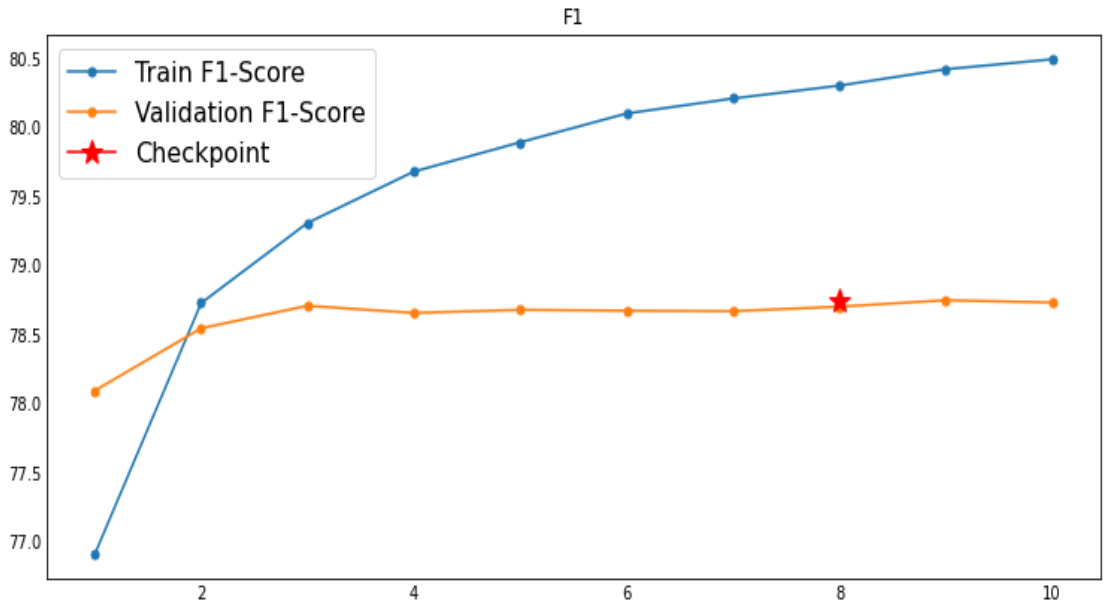


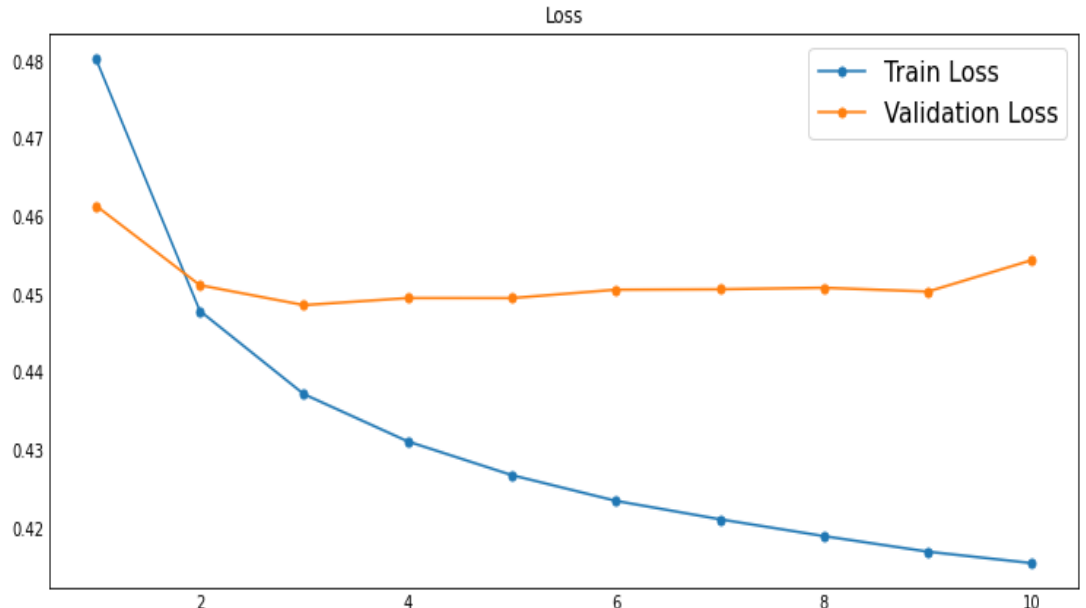


Dataset	Tweets	IMDB
Accuracy Score	79.08%	63.39%
F1 - Score	78.94%	61.08%
Loss	0.46	0.63

LSTM Results







Dataset	Tweets	IMDB
Accuracy Score	78.88%	61.01%
F1 - Score	78.73%	60.34%
Loss	0.44	0.65