

INLP - Assignment - 2 Report

Ishan Kavathekar

2022121003

Feed Forward Neural Network POS Tagger:

1. Model 1

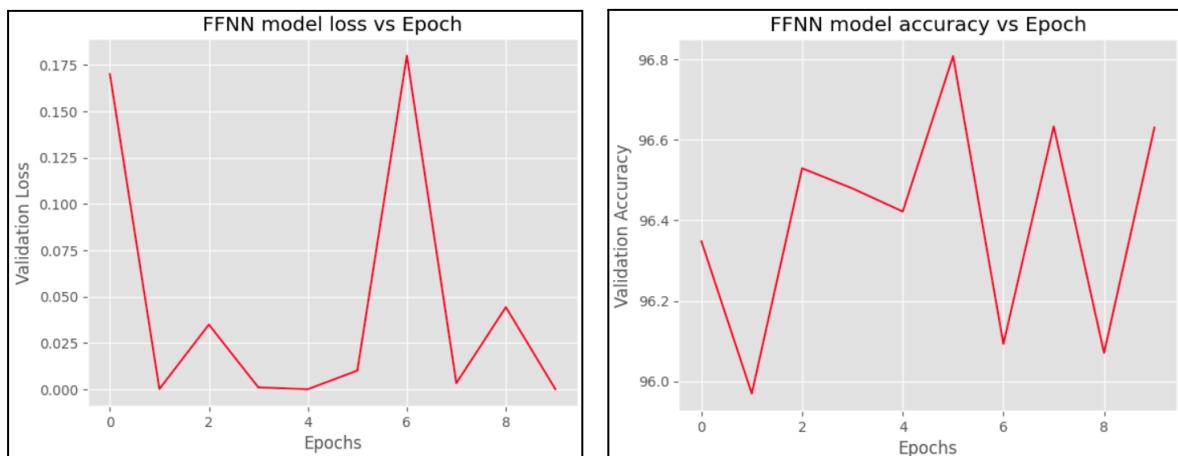
Hyperparameters:

- Embedding dimension = 300 (Glove embeddings)
- Number of hidden layers = 3
- Size of hidden layers = 1024, 512, 256
- Activation function = ReLU
- Learning rate = 0.001
- Epochs = 10

- P = 3
- S = 3

Validation accuracy: 97.344%

Test accuracy: 97.568%

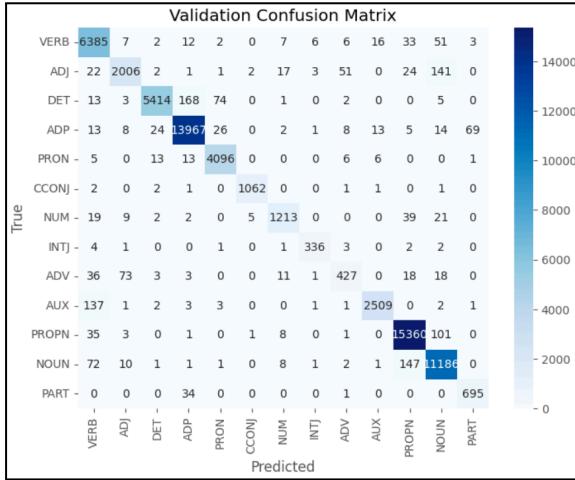


Validation accuracy and loss across epochs

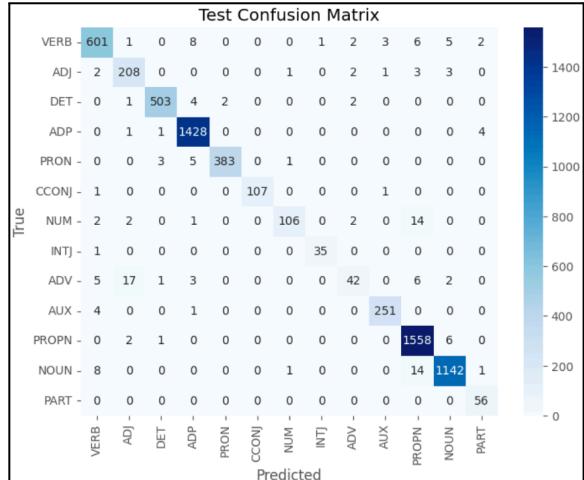
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.95	0.98	0.96	6530	0	0.96	0.96	0.96	629
1	0.95	0.88	0.91	2270	1	0.90	0.95	0.92	220
2	0.99	0.95	0.97	5680	2	0.99	0.98	0.99	512
3	0.98	0.99	0.99	14150	3	0.98	1.00	0.99	1434
4	0.97	0.99	0.98	4140	4	0.99	0.98	0.99	392
5	0.99	0.99	0.99	1070	5	1.00	0.98	0.99	109
6	0.96	0.93	0.94	1310	6	0.97	0.83	0.90	127
7	0.96	0.96	0.96	350	7	0.97	0.97	0.97	36
8	0.84	0.72	0.78	590	8	0.84	0.55	0.67	76
9	0.99	0.94	0.96	2660	9	0.98	0.98	0.98	256
10	0.98	0.99	0.99	15510	10	0.97	0.99	0.98	1567
11	0.97	0.98	0.97	11430	11	0.99	0.98	0.98	1166
12	0.90	0.95	0.93	730	12	0.89	1.00	0.94	56
accuracy			0.97	66420	accuracy			0.98	6580
macro avg	0.96	0.94	0.95	66420	macro avg	0.96	0.93	0.94	6580
weighted avg	0.97	0.97	0.97	66420	weighted avg	0.98	0.98	0.98	6580

Validation performance metrics

Test performance metrics



Validation Confusion Matrix

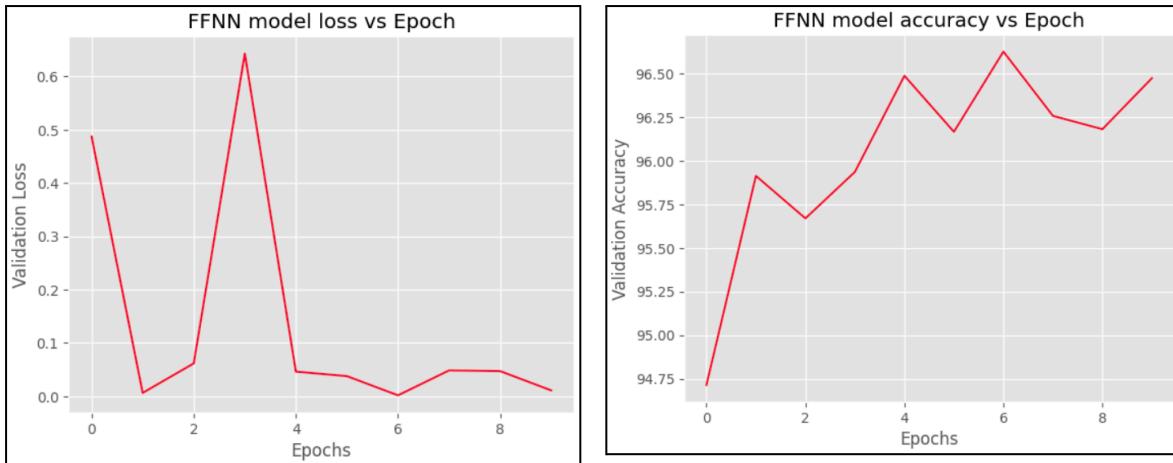


Test Confusion Matrix

- $P = 4$
- $S = 4$

Validation accuracy: 97.076%

Test accuracy: 97.629%



Validation accuracy and loss across epochs

	precision	recall	f1-score	support
0	0.95	0.97	0.96	6530
1	0.92	0.88	0.90	2270
2	0.99	0.95	0.97	5680
3	0.98	0.98	0.98	14150
4	0.95	0.99	0.97	4140
5	1.00	0.99	0.99	1070
6	0.96	0.90	0.93	1310
7	0.99	0.97	0.98	350
8	0.75	0.71	0.73	590
9	0.97	0.94	0.96	2660
10	0.98	0.99	0.99	15510
11	0.97	0.98	0.97	11430
12	0.91	0.97	0.94	730
accuracy			0.97	66420
macro avg	0.95	0.94	0.94	66420
weighted avg	0.97	0.97	0.97	66420

	precision	recall	f1-score	support
0	0.98	0.94	0.96	629
1	0.92	0.93	0.92	220
2	0.99	0.98	0.99	512
3	0.98	0.99	0.99	1434
4	0.97	0.98	0.98	392
5	0.99	0.98	0.99	109
6	0.98	0.88	0.93	127
7	0.97	0.94	0.96	36
8	0.83	0.71	0.77	76
9	0.97	0.99	0.98	256
10	0.98	1.00	0.99	1567
11	0.98	0.98	0.98	1166
12	0.98	1.00	0.99	56
accuracy			0.98	6580
macro avg	0.96	0.95	0.95	6580
weighted avg	0.98	0.98	0.98	6580

Validation performance metrics

Test performance metrics

Validation Confusion Matrix														
True	VERB	ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART	
VERB	6357	13	1	37	2	1	4	3	8	31	27	46	0	
ADJ	8	1994	3	2	1	0	10	1	109	4	18	120	0	-14000
DET	-1	7	5408	138	122	0	0	0	0	0	1	3	0	-12000
ADP	-19	31	23	13892	77	0	3	0	7	10	4	16	68	-12000
PRON	-7	3	15	16	4088	0	0	0	5	4	0	2	0	-10000
CCONJ	-1	0	1	2	0	1063	0	0	1	2	0	0	0	-8000
NUM	-25	18	0	2	1	2	1183	0	2	0	39	38	0	-8000
INTJ	-6	0	0	0	0	0	341	2	1	0	0	0	0	-6000
ADV	-39	57	2	7	3	1	8	1	420	13	15	24	0	-6000
AUX	-138	0	3	4	0	0	0	0	5	2508	0	1	1	-4000
PROPN	-18	2	0	1	0	0	15	0	1	0	15348	125	0	-4000
NOUN	-54	41	3	8	1	0	12	0	0	2	139	1170	0	-2000
PART	-0	0	0	23	0	0	0	0	1	0	0	706	0	-0

Validation Confusion Matrix

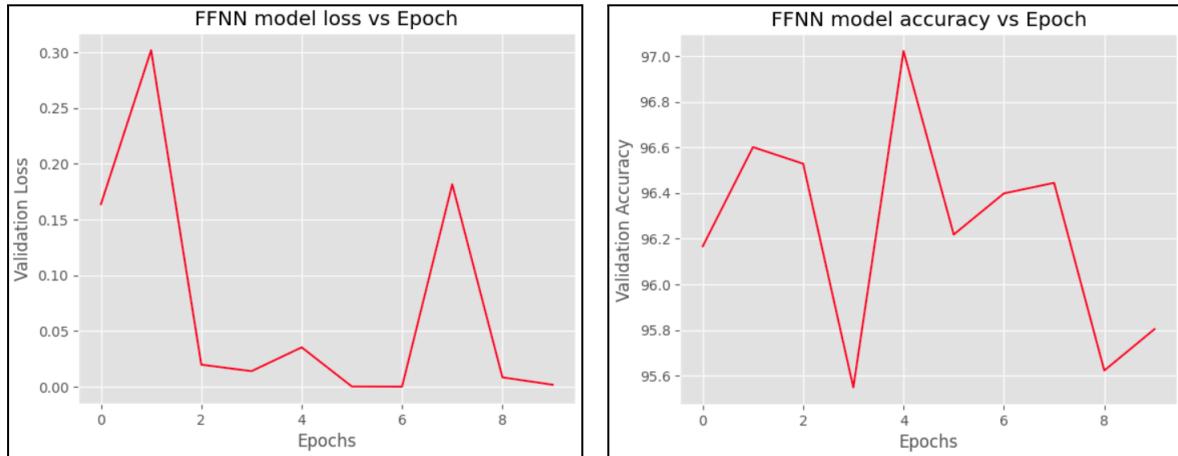
Test Confusion Matrix														
True	VERB	ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART	
VERB	592	1	0	21	2	0	0	0	0	1	4	2	6	0
ADJ	-	204	0	1	1	0	1	1	4	0	2	6	0	-1400
DET	-	0	1	504	2	3	0	0	0	1	0	0	1	-1200
ADP	-	1	1	3	1423	2	0	0	0	0	3	0	0	-1200
PRON	-	0	0	3	1	386	0	1	0	1	0	0	0	-1000
CCONJ	-	0	0	0	0	107	0	0	0	0	0	0	0	-800
NUM	-	1	1	0	0	0	1	112	0	2	0	6	4	-800
INTJ	-	2	0	0	0	0	0	0	34	0	0	0	0	-600
ADV	-	2	10	0	3	1	0	0	0	54	0	3	3	-600
AUX	-	2	0	0	0	0	0	0	0	1	253	0	0	-400
PROPN	-	1	1	0	0	0	0	0	0	0	0	1560	5	-400
NOUN	-	3	3	0	2	0	0	0	0	1	0	18	1139	0
PART	-	0	0	0	0	0	0	0	0	0	0	0	56	-0

Test Confusion Matrix

- P = 2
- S = 2

Validation accuracy: 97.163%

Test accuracy: 97.0516%



Validation accuracy and loss across epochs

	precision	recall	f1-score	support
0	0.96	0.96	0.96	6530
1	0.95	0.86	0.90	2270
2	0.99	0.95	0.97	5680
3	0.98	0.99	0.98	14150
4	0.97	0.99	0.98	4140
5	1.00	0.99	0.99	1070
6	0.91	0.93	0.92	1310
7	0.93	0.97	0.95	350
8	0.79	0.79	0.79	590
9	0.98	0.94	0.96	2660
10	0.98	0.99	0.99	15510
11	0.97	0.98	0.97	11430
12	0.91	0.97	0.94	730
accuracy			0.97	66420
macro avg	0.95	0.95	0.95	66420
weighted avg	0.97	0.97	0.97	66420

Validation performance metrics

	precision	recall	f1-score	support
0	0.98	0.89	0.93	629
1	0.94	0.95	0.94	220
2	0.98	0.99	0.99	512
3	0.97	0.98	0.98	1434
4	0.98	0.99	0.99	392
5	0.99	0.97	0.98	109
6	0.94	0.91	0.92	127
7	0.97	0.97	0.97	36
8	0.80	0.83	0.81	76
9	0.96	0.98	0.97	256
10	0.99	0.98	0.98	1567
11	0.97	0.98	0.98	1166
12	0.92	1.00	0.96	56
accuracy			0.97	6580
macro avg	0.95	0.96	0.95	6580
weighted avg	0.97	0.97	0.97	6580

Test performance metrics

Validation Confusion Matrix														
True	VERB	ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART	Predicted
VERB	6292	6	3	109	1	1	12	6	6	16	28	50	0	-14000
ADJ	-	14	1957	11	3	1	1	19	5	97	2	32	128	0
DET	-	1	4	5417	164	78	0	6	0	2	0	4	4	0
ADP	-	10	6	13	13992	24	0	6	3	6	8	2	15	65
PRON	-	5	0	25	6	4090	0	1	0	7	2	1	2	1
CCONJ	-	4	0	0	2	0	1058	0	0	0	5	0	1	0
NUM	-	7	9	2	1	2	0	1222	0	1	1	33	32	0
INTJ	-	3	0	0	0	2	0	1	339	1	0	2	2	0
ADV	-	20	44	0	8	0	0	11	1	466	4	16	20	0
AUX	-	123	0	1	8	3	0	0	7	2	2508	0	7	1
PROPN	-	31	9	2	3	1	0	46	1	2	0	1528	127	0
NOUN	-	47	23	3	9	1	0	24	3	1	0	121	1198	0
PART	-	0	0	0	18	0	0	0	0	0	3	0	0	709

Test Confusion Matrix														
True	VERB	ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART	Predicted
VERB	558	1	1	46	2	0	2	0	2	4	1	12	0	-1400
ADJ	-	2	208	0	0	0	0	0	1	7	0	1	1	0
DET	-	1	0	509	0	1	0	0	0	0	0	0	1	0
ADP	-	0	2	6	1412	2	0	0	0	5	3	0	1	3
PRON	-	0	0	2	1	388	0	1	0	0	0	0	0	0
CCONJ	-	1	0	0	0	0	106	0	0	0	2	0	0	0
NUM	-	1	3	1	0	0	0	115	0	1	0	3	3	0
INTJ	-	0	0	0	0	0	0	0	35	0	0	1	0	0
ADV	-	1	4	0	0	0	0	0	0	63	1	4	1	2
AUX	-	6	0	0	0	0	0	0	0	0	250	0	0	0
PROPN	-	0	2	1	1	0	0	3	0	1	0	1538	21	0
NOUN	-	1	1	1	0	1	1	0	0	0	0	12	1148	0
PART	-	0	0	0	0	0	0	0	0	0	0	0	56	-

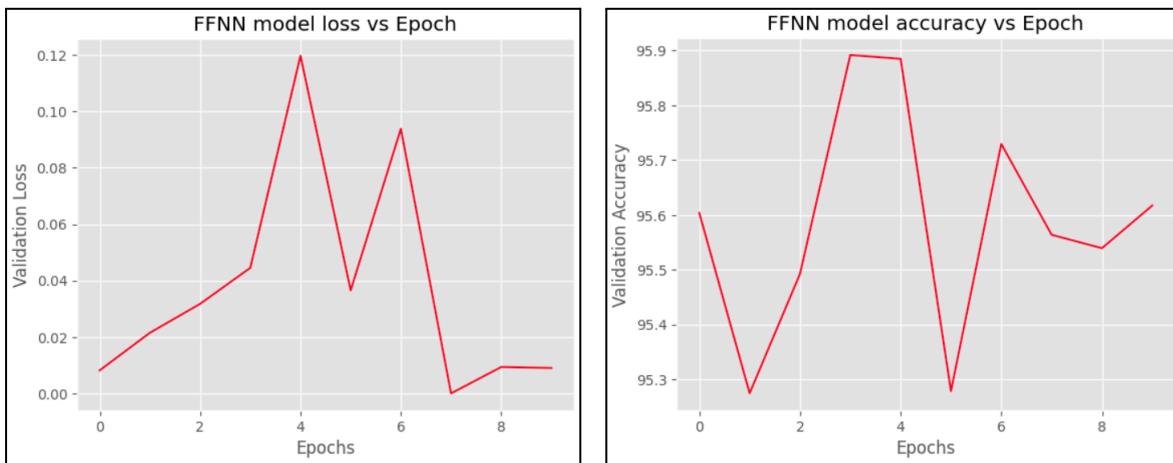
Validation Confusion Matrix

Test Confusion Matrix

- P = 1
- S = 1

Validation accuracy: 96.796%

Test accuracy: 97.188%

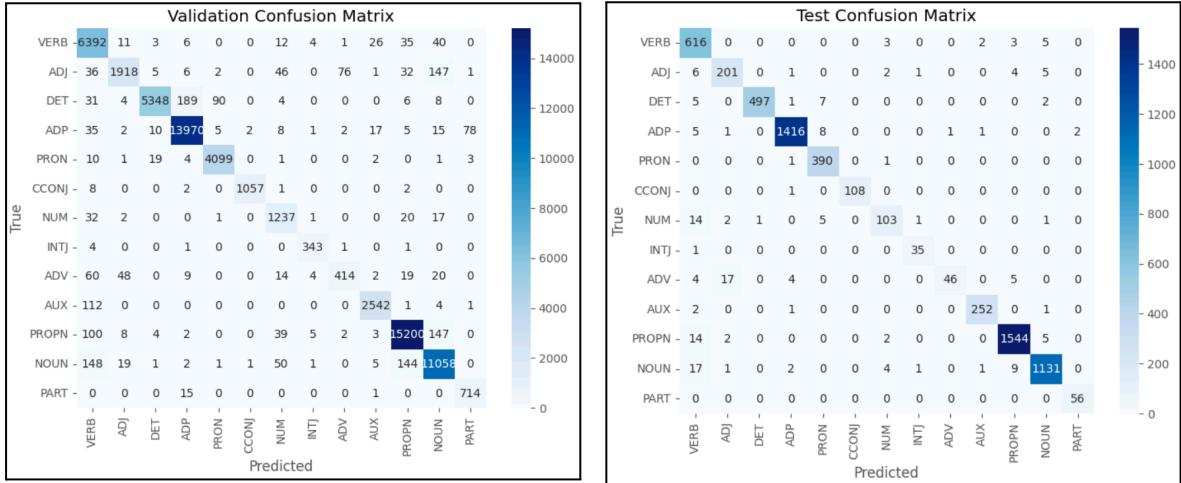


Validation accuracy and loss across epochs

	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.92	0.98	0.95	6530	0	0.90	0.98	0.94	629
1	0.95	0.84	0.90	2270	1	0.90	0.91	0.91	220
2	0.99	0.94	0.97	5680	2	1.00	0.97	0.98	512
3	0.98	0.99	0.99	14150	3	0.99	0.99	0.99	1434
4	0.98	0.99	0.98	4140	4	0.95	0.99	0.97	392
5	1.00	0.99	0.99	1070	5	1.00	0.99	1.00	109
6	0.88	0.94	0.91	1310	6	0.90	0.81	0.85	127
7	0.96	0.98	0.97	350	7	0.92	0.97	0.95	36
8	0.83	0.70	0.76	590	8	0.98	0.61	0.75	76
9	0.98	0.96	0.97	2660	9	0.98	0.98	0.98	256
10	0.98	0.98	0.98	15510	10	0.99	0.99	0.99	1567
11	0.97	0.97	0.97	11430	11	0.98	0.97	0.98	1166
12	0.90	0.98	0.94	730	12	0.97	1.00	0.98	56
accuracy			0.97	66420	accuracy			0.97	6580
macro avg	0.95	0.94	0.94	66420	macro avg	0.96	0.94	0.94	6580
weighted avg	0.97	0.97	0.97	66420	weighted avg	0.97	0.97	0.97	6580

Validation performance metrics

Test performance metrics



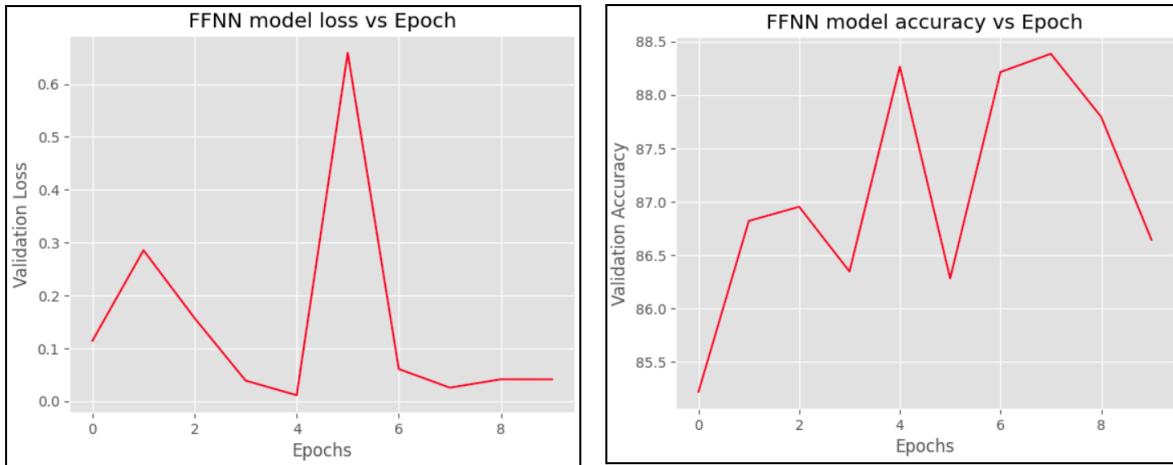
Validation Confusion Matrix

Test Confusion Matrix

- P = 0
- S = 0

Validation accuracy: 90.382%

Test accuracy: 90.106%



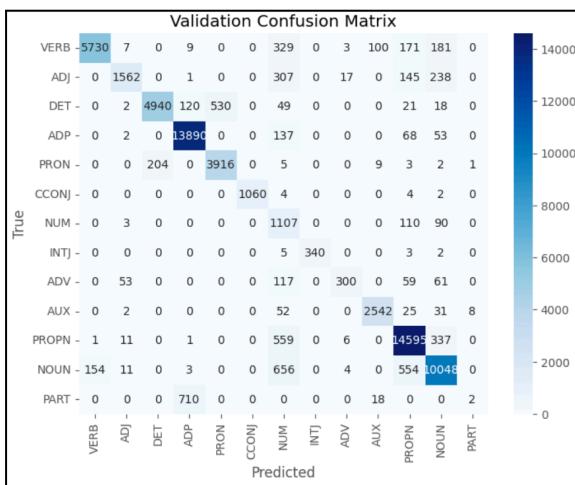
Validation accuracy and loss across epochs

	precision	recall	f1-score	support
0	0.97	0.88	0.92	6530
1	0.94	0.69	0.80	2270
2	0.96	0.87	0.91	5680
3	0.94	0.98	0.96	14150
4	0.88	0.95	0.91	4140
5	1.00	0.99	1.00	1070
6	0.33	0.85	0.48	1310
7	1.00	0.97	0.99	350
8	0.91	0.51	0.65	590
9	0.95	0.96	0.95	2660
10	0.93	0.94	0.93	15510
11	0.91	0.88	0.89	11430
12	0.18	0.00	0.01	730
accuracy			0.90	66420
macro avg	0.84	0.80	0.80	66420
weighted avg	0.91	0.90	0.90	66420

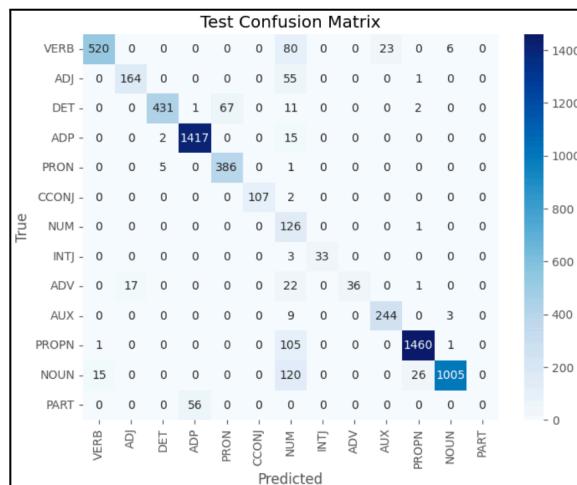
	precision	recall	f1-score	support
0	0.97	0.83	0.89	629
1	0.91	0.75	0.82	220
2	0.98	0.84	0.91	512
3	0.96	0.99	0.97	1434
4	0.85	0.98	0.91	392
5	1.00	0.98	0.99	109
6	0.23	0.99	0.37	127
7	1.00	0.92	0.96	36
8	1.00	0.47	0.64	76
9	0.91	0.95	0.93	256
10	0.98	0.93	0.95	1567
11	0.99	0.86	0.92	1166
12	0.00	0.00	0.00	56
accuracy			0.90	6580
macro avg	0.83	0.81	0.79	6580
weighted avg	0.94	0.90	0.91	6580

Validation performance metrics

Test performance metrics



Validation Confusion Matrix



Test Confusion Matrix

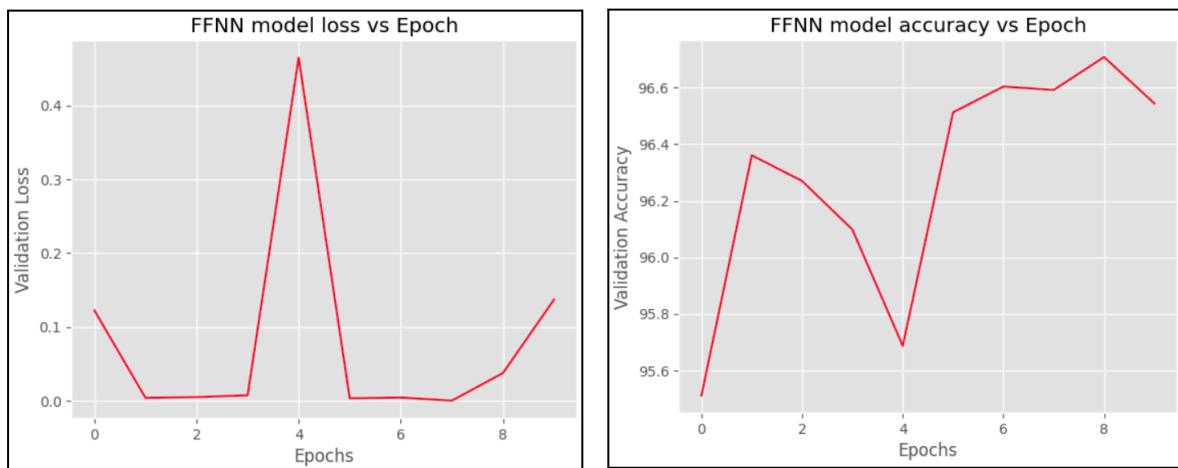
2. Model 2

Hyperparameters:

- P = 3
- S = 3
- Embedding dimension = 300 (Glove embeddings)
- Number of hidden layers = 6
- Size of hidden layers = 1024, 512, 256, 128, 64, 32
- Activation function = ReLU
- Learning rate = 0.001
- Epochs = 10

Validation accuracy: 97.174%

Test accuracy: 98.0547%



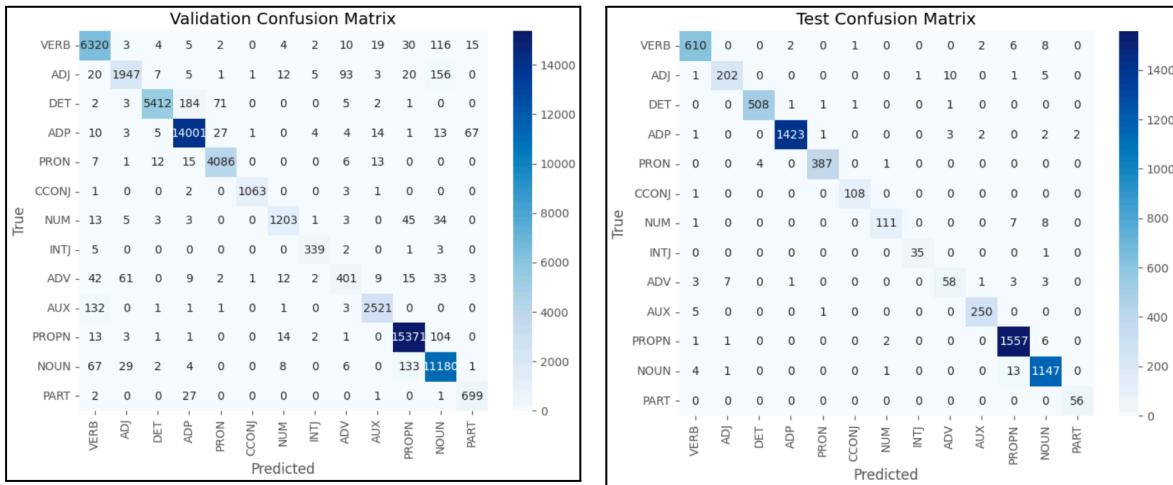
Validation accuracy and loss across epochs

	precision	recall	f1-score	support
0	0.95	0.97	0.96	6530
1	0.95	0.86	0.90	2270
2	0.99	0.95	0.97	5680
3	0.98	0.99	0.99	14150
4	0.98	0.99	0.98	4140
5	1.00	0.99	1.00	1070
6	0.96	0.92	0.94	1310
7	0.95	0.97	0.96	350
8	0.75	0.68	0.71	590
9	0.98	0.95	0.96	2660
10	0.98	0.99	0.99	15510
11	0.96	0.98	0.97	11430
12	0.89	0.96	0.92	730
accuracy			0.97	66420
macro avg	0.95	0.94	0.94	66420
weighted avg	0.97	0.97	0.97	66420

Validation performance metrics

	precision	recall	f1-score	support
0	0.97	0.97	0.97	629
1	0.96	0.92	0.94	220
2	0.99	0.99	0.99	512
3	1.00	0.99	0.99	1434
4	0.99	0.99	0.99	392
5	0.98	0.99	0.99	109
6	0.97	0.87	0.92	127
7	0.97	0.97	0.97	36
8	0.81	0.76	0.78	76
9	0.98	0.98	0.98	256
10	0.98	0.99	0.99	1567
11	0.97	0.98	0.98	1166
12	0.97	1.00	0.98	56
accuracy			0.98	6580
macro avg	0.96	0.95	0.96	6580
weighted avg	0.98	0.98	0.98	6580

Test performance metrics



Validation Confusion Matrix

Test Confusion Matrix

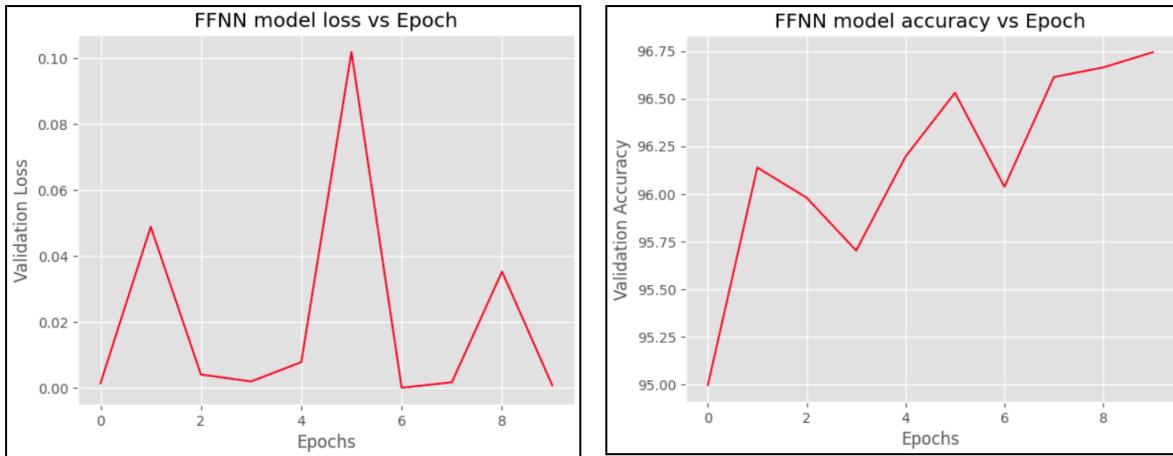
3. Model 3

Hyperparameters:

- P = 3
- S = 3
- Embedding dimension = 300 (Glove embeddings)
- Number of hidden layers = 4
- Size of hidden layers = 512, 1024, 512, 256,
- Activation function = ReLU
- Learning rate = 0.001
- Epochs = 10

Validation accuracy: 97.040%

Test accuracy: 97.6447%



Validation accuracy and loss across epochs

	precision	recall	f1-score	support
0	0.95	0.98	0.96	6530
1	0.96	0.85	0.90	2270
2	0.99	0.95	0.97	5680
3	0.98	0.99	0.98	14150
4	0.98	0.99	0.98	4140
5	0.99	0.99	0.99	1070
6	0.93	0.92	0.93	1310
7	0.94	0.98	0.96	350
8	0.77	0.71	0.74	590
9	0.98	0.95	0.97	2660
10	0.98	0.99	0.98	15510
11	0.96	0.98	0.97	11430
12	0.91	0.86	0.89	730
accuracy			0.97	66420
macro avg	0.95	0.93	0.94	66420
weighted avg	0.97	0.97	0.97	66420

	precision	recall	f1-score	support
0	0.99	0.94	0.97	629
1	0.90	0.95	0.93	220
2	0.99	0.99	0.99	512
3	0.99	0.99	0.99	1434
4	0.98	0.98	0.98	392
5	0.99	0.99	0.99	109
6	0.92	0.88	0.90	127
7	0.92	0.97	0.95	36
8	0.83	0.58	0.68	76
9	0.98	0.99	0.98	256
10	0.99	0.99	0.99	1567
11	0.96	0.99	0.97	1166
12	0.97	1.00	0.98	56
accuracy			0.98	6580
macro avg	0.95	0.94	0.95	6580
weighted avg	0.98	0.98	0.98	6580

Validation performance metrics

Test performance metrics

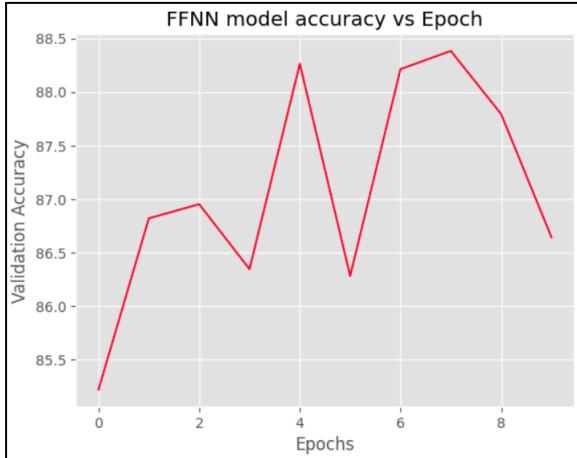
Validation Confusion Matrix														
True	VERB	ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART	
VERB	6380	6	5	5	3	1	8	2	6	16	36	62	0	
ADJ	-23	1926	10	5	4	1	20	4	89	0	31	157	0	-14000
DET	-5	1	5418	188	61	0	1	0	3	0	2	1	0	-12000
ADP	-21	12	15	13976	16	4	2	0	6	15	7	17	59	-12000
PRON	-5	1	19	14	4090	0	0	0	7	3	0	0	1	-10000
CCONJ	-3	0	0	1	0	1062	0	0	4	0	0	0	0	-8000
NUM	-26	7	2	4	0	1	1203	1	4	0	37	25	0	-8000
INTJ	-1	0	0	0	0	0	1	344	0	0	3	1	0	-6000
ADV	-45	43	5	12	4	0	11	3	421	6	12	28	0	-6000
AUX	-131	0	0	1	4	0	0	0	3	2521	0	0	0	-4000
PROPN	-23	3	0	3	0	0	25	9	1	1	15298	147	0	-4000
NOUN	-67	11	3	3	0	0	17	3	6	2	133	11185	0	-2000
PART	-1	0	1	96	1	0	0	0	0	0	1	630		-0
VERB		ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART	
True														
Predicted														

Validation Confusion Matrix

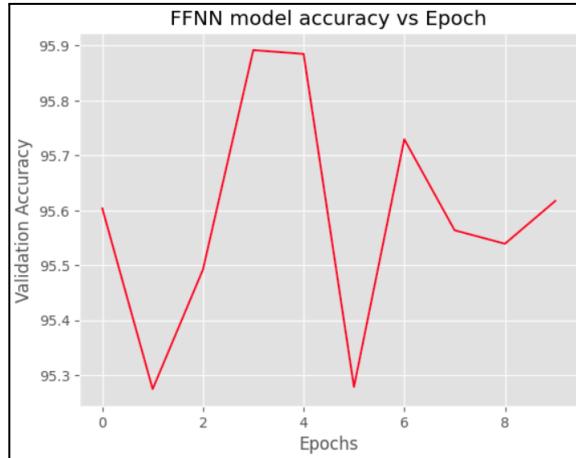
Test Confusion Matrix															
True	VERB	ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART		
VERB	594	1	0	2	1	0	1	1	3	2	3	21	0		
ADJ	0	210	0	0	0	0	0	1	3	0	2	4	0	-1400	
DET	0	0	509	1	1	0	1	0	0	0	0	0	0	-1200	
ADP	0	1	2	1422	2	0	0	0	1	3	0	2	1	-1000	
PRON	0	0	4	2	383	0	1	1	1	0	0	0	0	-800	
CCONJ	0	1	0	0	0	108	0	0	0	0	0	0	0	-600	
NUM	1	0	0	0	0	0	112	0	1	0	4	9	0	-400	
INTJ	0	0	0	0	0	0	0	35	0	0	1	0	0	-200	
ADV	0	18	0	5	3	1	0	0	44	1	3	1	0	-100	
AUX	3	0	0	0	0	0	0	0	0	253	0	0	0	-60	
PROPN	2	1	0	0	0	0	6	0	0	0	1544	14	0	-400	
NOUN	0	1	0	0	0	0	1	0	0	0	0	8	1155	1	-200
PART	0	0	0	0	0	0	0	0	0	0	0	0	56		-0
VERB		ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART		
True															
Predicted															

Test Confusion Matrix

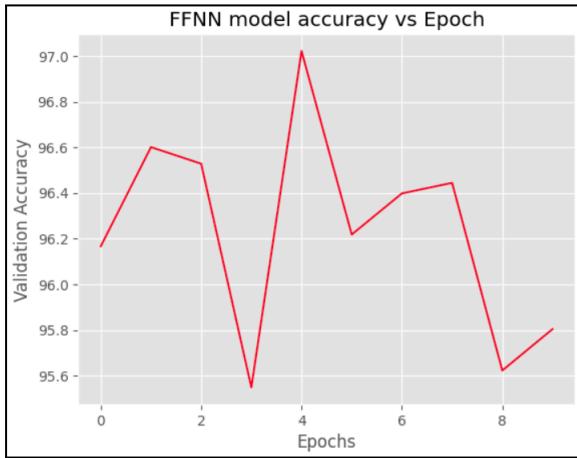
Validation accuracy vs Number of epochs for various p and s values:



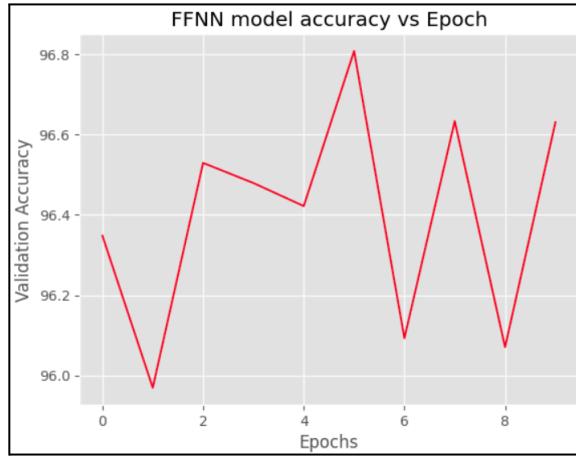
p=s=0



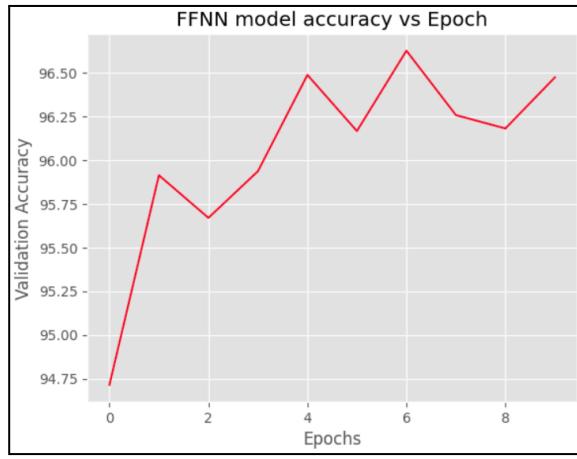
p=s=1



p=s=2



p=s=3



p=s=4

Validation accuracy vs Epoch for p=s= [0,4]

Analysis:

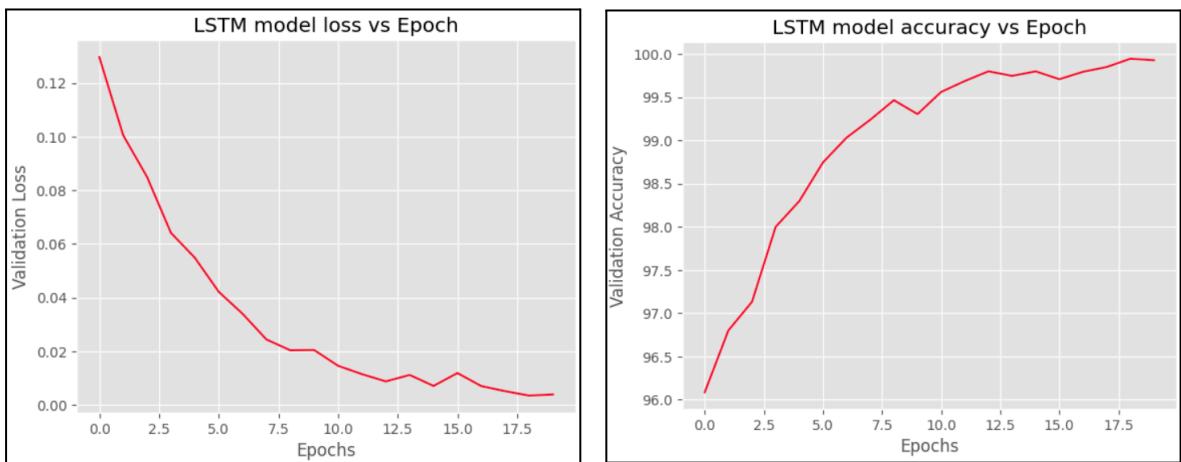
We observe that model 1 performs the best on the dev set. We keep this model constant and check the validation accuracy vs epoch for various values of $p=s=[0,4]$. The performance of the model with $p=s=3$ is superior to that of others. POS tagging is a task wherein the context of the sentence also matters. Hence, taking the previous and successive words into consideration is of great importance. It should be noted that as the values of p and s decreases, the accuracy of the model also declines. This can be confirmed from the accuracy of the model with p and s value of 0. Since, this model does not have access to the context of target word while training, it performs poorly after deployment. However, considering a larger context window is redundant and can take into account words which are of no use to the task. These words will again reduce the accuracy of the model. Hence, an optimal set of p and s values are necessary for maximising the model accuracy. In these experiments we train the models for 10 epochs. The performance of the model can be increased by experimenting with different neural net architectures and training the model for more epochs.

Recurrent Neural Network POS Tagger:

1. Model 1

Hyperparameters:

- Embedding dimensions = 200
- Hidden dimensions = 200
- Epochs = 20
- Learning rate = 0.001
- Bidirectional = True
- Batch first = True

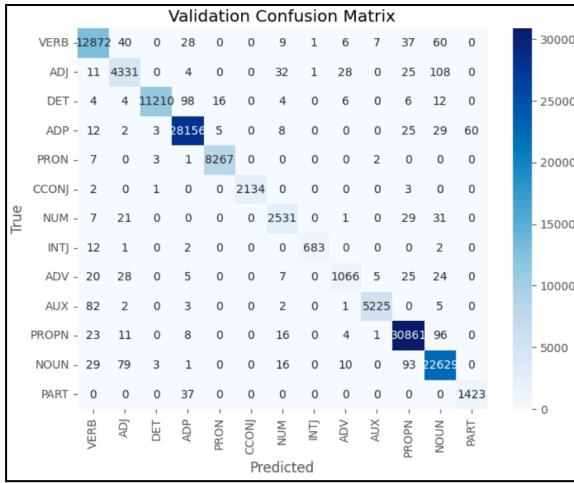


Validation accuracy and loss across epochs

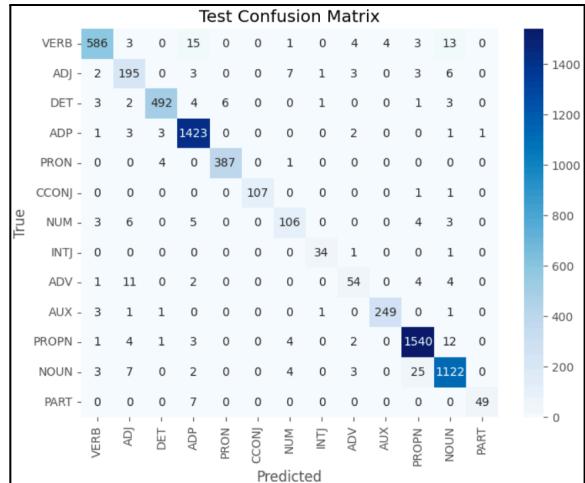
	precision	recall	f1-score	support		precision	recall	f1-score	support	
0	0.99	0.98	0.98	13081		0	0.93	0.97	0.95	603
1	0.95	0.96	0.96	4519		1	0.89	0.84	0.86	232
2	0.99	1.00	0.99	11220		2	0.96	0.98	0.97	501
3	0.99	0.99	0.99	28343		3	0.99	0.97	0.98	1464
4	1.00	1.00	1.00	8288		4	0.99	0.98	0.99	393
5	1.00	1.00	1.00	2134		5	0.98	1.00	0.99	107
6	0.97	0.96	0.97	2625		6	0.83	0.86	0.85	123
7	0.98	1.00	0.99	685		7	0.94	0.92	0.93	37
8	0.90	0.95	0.93	1122		8	0.71	0.78	0.74	69
9	0.98	1.00	0.99	5240		9	0.97	0.98	0.98	253
10	0.99	0.99	0.99	31104		10	0.98	0.97	0.98	1581
11	0.99	0.98	0.99	22996		11	0.96	0.96	0.96	1167
12	0.97	0.96	0.97	1483		12	0.88	0.98	0.92	50
accuracy			0.99	132840	accuracy			0.96	6580	
macro avg	0.98	0.98	0.98	132840	macro avg	0.92	0.94	0.93	6580	
weighted avg	0.99	0.99	0.99	132840	weighted avg	0.96	0.96	0.96	6580	

Validation performance metrics

Test performance metrics



Validation Confusion Matrix

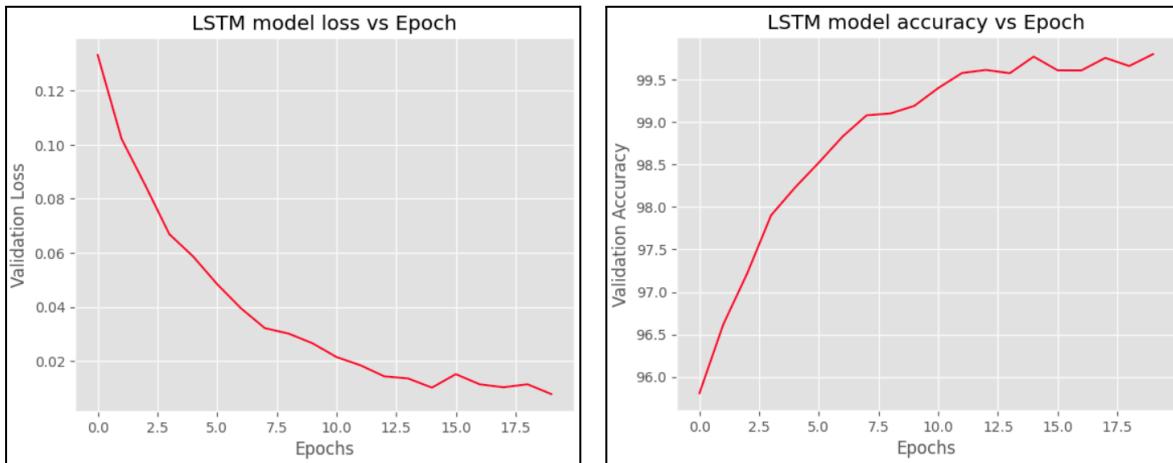


Test Confusion Matrix

2. Model 2

Hyperparameters:

- Embedding dimensions = 300
- Hidden dimensions = 128
- Epochs = 20
- Learning rate = 0.001
- Bidirectional = True
- Batch first = True



Validation accuracy and loss across epochs

	precision	recall	f1-score	support
0	0.98	0.98	0.98	13093
1	0.94	0.95	0.95	4486
2	0.98	1.00	0.99	11179
3	0.99	0.99	0.99	28339
4	1.00	1.00	1.00	8294
5	1.00	1.00	1.00	2135
6	0.96	0.96	0.96	2629
7	0.98	1.00	0.99	687
8	0.89	0.95	0.92	1105
9	0.98	1.00	0.99	5226
10	0.99	0.99	0.99	31120
11	0.99	0.98	0.98	23032
12	0.98	0.95	0.96	1515
accuracy			0.99	132840
macro avg	0.97	0.98	0.98	132840
weighted avg	0.99	0.99	0.99	132840

Validation performance metrics

	precision	recall	f1-score	support
0	0.94	0.93	0.93	636
1	0.88	0.73	0.80	264
2	0.96	0.98	0.97	503
3	0.99	0.98	0.98	1446
4	0.99	0.98	0.99	396
5	0.98	1.00	0.99	107
6	0.78	0.91	0.84	109
7	0.94	0.97	0.96	35
8	0.62	0.76	0.68	62
9	0.98	0.98	0.98	256
10	0.98	0.98	0.98	1576
11	0.95	0.97	0.96	1143
12	0.82	0.98	0.89	47
accuracy			0.96	6580
macro avg	0.91	0.93	0.92	6580
weighted avg	0.96	0.96	0.96	6580

Test performance metrics

Validation Confusion Matrix																													
	True	VERB	ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART	VERB	ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART		Predicted
VERB -	12856	47	0	34	0	0	12	1	5	10	38	57	0	-	30000	-	-	-	-	-	-	-	-	-	-	-	-		
ADJ -	26	4265	0	2	0	0	36	0	29	0	13	169	0	-	-	-	-	-	-	-	-	-	-	-	-	-			
DET -	3	7	11169	117	26	0	0	0	7	0	10	21	0	-	-	-	-	-	-	-	-	-	-	-	-	-			
ADP -	15	2	1	28132	2	0	10	0	0	0	26	35	77	-	-	-	-	-	-	-	-	-	-	-	-	-			
PRON -	5	0	6	0	8266	0	0	0	0	0	0	0	3	-	-	-	-	-	-	-	-	-	-	-	-	-			
CCONJ -	0	0	0	0	0	2135	0	0	0	0	0	5	0	0	-	-	-	-	-	-	-	-	-	-	-	-			
NUM -	1	16	0	1	0	0	2516	0	1	1	39	45	0	-	-	-	-	-	-	-	-	-	-	-	-	-			
INTJ -	4	5	0	3	0	0	0	686	2	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-			
ADV -	23	35	0	2	0	0	0	11	0	1052	3	31	23	0	-	-	-	-	-	-	-	-	-	-	-	-			
AUX -	86	1	0	9	0	0	4	0	1	5212	1	6	0	-	-	-	-	-	-	-	-	-	-	-	-	-			
PROPN -	34	14	0	11	0	0	16	0	6	0	30851	88	0	-	-	-	-	-	-	-	-	-	-	-	-	-			
NOUN -	40	94	3	3	0	0	24	0	2	0	106	22588	0	-	-	-	-	-	-	-	-	-	-	-	-	-			
PART -	0	0	0	25	0	0	0	0	0	0	1435	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
VERB -	ADJ -	DET -	ADP -	PRON -	CCONJ -	NUM -	INTJ -	ADV -	AUX -	PROPN -	NOUN -	PART -	-	-	-	-	-	-	-	-	-	-	-	-	-	-			

Validation Confusion Matrix

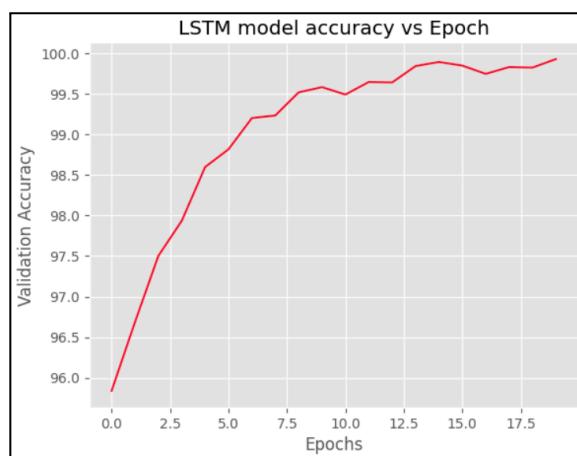
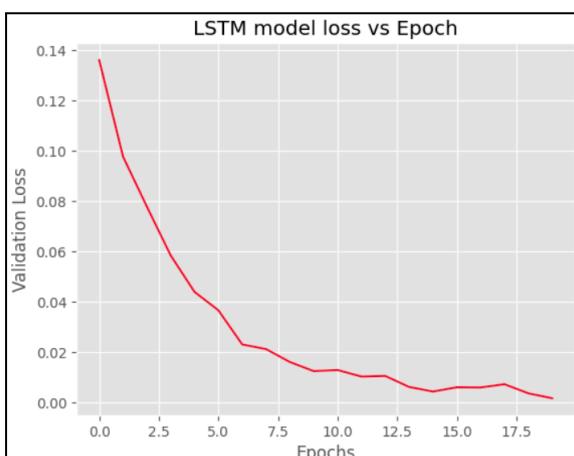
Test Confusion Matrix																													
	True	VERB	ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART	VERB	ADJ	DET	ADP	PRON	CCONJ	NUM	INTJ	ADV	AUX	PROPN	NOUN	PART		Predicted
VERB -	591	10	0	6	0	0	1	0	6	4	3	8	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
ADJ -	8	194	0	1	0	0	5	0	2	0	2	8	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
DET -	4	2	494	2	5	0	0	1	1	0	1	2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
ADP -	4	2	4	1416	2	0	0	0	0	2	0	1	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-		
PRON -	0	0	2	0	389	0	1	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
CCONJ -	1	0	0	0	0	107	0	0	0	0	0	1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
NUM -	3	10	1	3	0	0	99	0	2	1	6	2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
INTJ -	0	1	0	0	0	0	0	34	0	0	0	0	1	0	-	-	-	-	-	-	-	-	-	-	-	-	-		
ADV -	5	17	0	1	0	0	0	0	0	47	0	4	2	0	-	-	-	-	-	-	-	-	-	-	-	-	-		
AUX -	2	4	0	0	0	0	0	0	0	0	250	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-		
PROPN -	7	8	2	5	0	0	1	0	0	0	0	1537	7	0	-	-	-	-	-	-	-	-	-	-	-	-	-		
NOUN -	11	16	0	2	0	0	2	0	2	1	21	1111	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
PART -	0	0	0	10	0	0	0	0	0	0	0	0	46	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
VERB -	ADJ -	DET -	ADP -	PRON -	CCONJ -	NUM -	INTJ -	ADV -	AUX -	PROPN -	NOUN -	PART -	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		

Test Confusion Matrix

3. Model 3

Hyperparameters:

- Embedding dimensions = 200
- Hidden dimensions = 256
- Epochs = 20
- Learning rate = 0.001
- Bidirectional = True
- Batch first = True
- Num_layers = 2

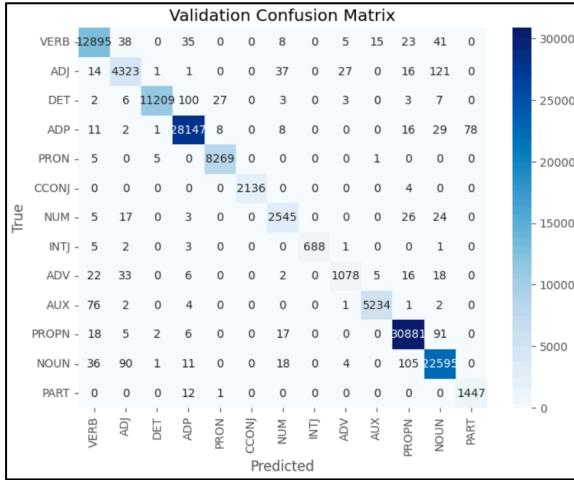


Validation accuracy and loss across epochs

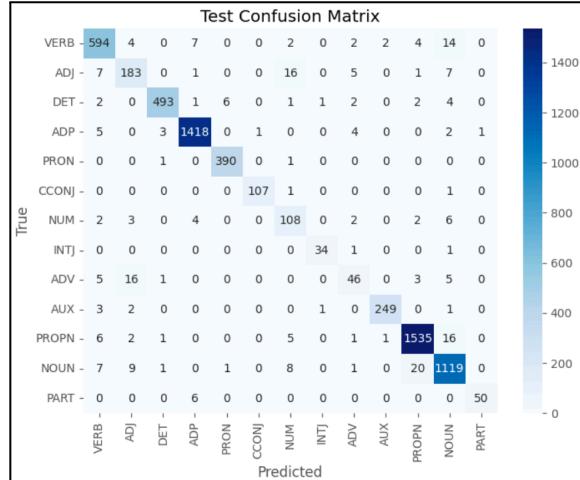
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.99	0.99	0.99	13089	0	0.94	0.94	0.94	631
1	0.95	0.96	0.95	4518	1	0.83	0.84	0.83	219
2	0.99	1.00	0.99	11219	2	0.96	0.99	0.97	500
3	0.99	0.99	0.99	28328	3	0.99	0.99	0.99	1437
4	1.00	1.00	1.00	8305	4	0.99	0.98	0.99	397
5	1.00	1.00	1.00	2136	5	0.98	0.99	0.99	108
6	0.97	0.96	0.97	2638	6	0.85	0.76	0.80	142
7	0.98	1.00	0.99	688	7	0.94	0.94	0.94	36
8	0.91	0.96	0.94	1119	8	0.61	0.72	0.66	64
9	0.98	1.00	0.99	5255	9	0.97	0.99	0.98	252
10	1.00	0.99	0.99	31091	10	0.98	0.98	0.98	1567
11	0.99	0.99	0.99	22929	11	0.96	0.95	0.96	1176
12	0.99	0.95	0.97	1525	12	0.89	0.98	0.93	51
accuracy			0.99	132840	accuracy			0.96	6580
macro avg	0.98	0.98	0.98	132840	macro avg	0.92	0.93	0.92	6580
weighted avg	0.99	0.99	0.99	132840	weighted avg	0.96	0.96	0.96	6580

Validation performance metrics

Test performance metrics



Validation Confusion Matrix

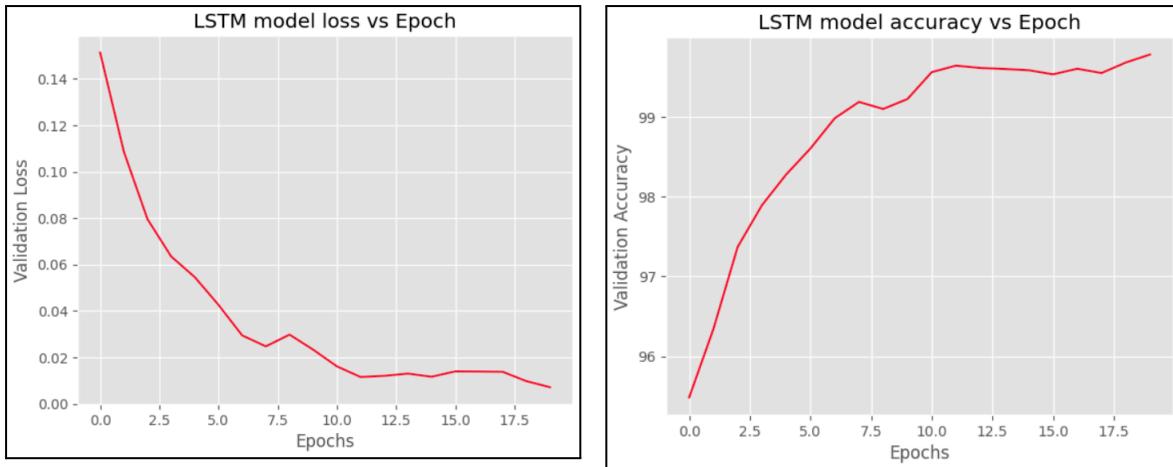


Test Confusion Matrix

4. Model 4

Hyperparameters:

- Embedding dimensions = 300
- Hidden dimensions = 128
- Epochs = 20
- Learning rate = 0.001
- Bidirectional = True
- Batch first = True
- Num_layers = 3



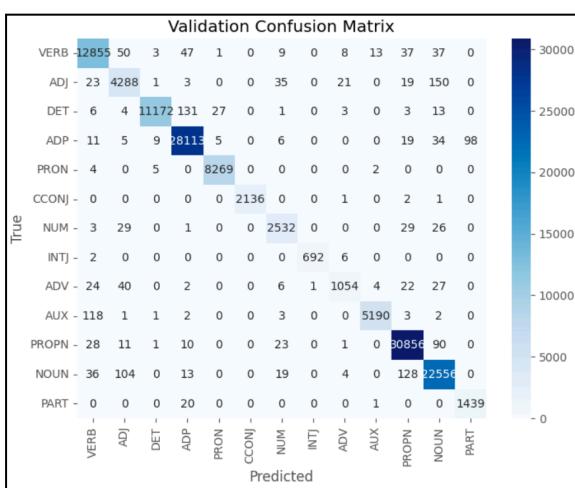
Validation accuracy and loss across epochs

	precision	recall	f1-score	support
0	0.98	0.98	0.98	13110
1	0.94	0.95	0.95	4532
2	0.98	1.00	0.99	11192
3	0.99	0.99	0.99	28342
4	1.00	1.00	1.00	8302
5	1.00	1.00	1.00	2136
6	0.97	0.96	0.96	2634
7	0.99	1.00	0.99	693
8	0.89	0.96	0.93	1098
9	0.98	1.00	0.99	5210
10	0.99	0.99	0.99	31118
11	0.99	0.98	0.99	22936
12	0.99	0.94	0.96	1537
accuracy			0.99	132840
macro avg	0.98	0.98	0.98	132840
weighted avg	0.99	0.99	0.99	132840

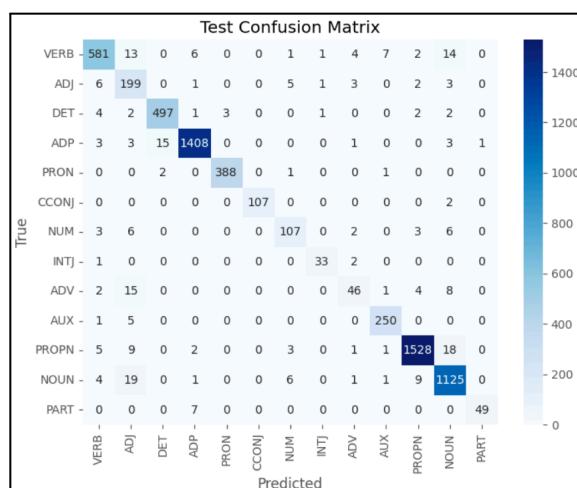
	precision	recall	f1-score	support
0	0.92	0.95	0.94	610
1	0.90	0.73	0.81	271
2	0.97	0.97	0.97	514
3	0.98	0.99	0.98	1426
4	0.99	0.99	0.99	391
5	0.98	1.00	0.99	107
6	0.84	0.87	0.86	123
7	0.92	0.92	0.92	36
8	0.61	0.77	0.68	60
9	0.98	0.96	0.97	261
10	0.98	0.99	0.98	1550
11	0.96	0.95	0.96	1181
12	0.88	0.98	0.92	50
accuracy			0.96	6580
macro avg	0.92	0.93	0.92	6580
weighted avg	0.96	0.96	0.96	6580

Validation performance metrics

Test performance metrics



Validation Confusion Matrix



Test Confusion Matrix

Analysis:

We choose LSTM as the RNN model for this Part of Speech tagging task. Specifically the report includes results from the BiLSTM model because it has been known to be used for POS tagging. The BiLSTM model has superior performance to the FFNN model with almost every dev set accuracy reaching 99%. Different parameters such as embedding dimension, hidden dimension and number of layers were changed to check the efficiency of the models. Since most of the models have the same accuracy, we also consider various other metrics such as precision and recall. We use the F1 score because it encompasses precision as well as recall. We use the embedding layer to embed the words into vectors. The embedding dimension is a very important hyperparameter because a large embedding dimension captures the nuances of the word relations, but is complex and might be compute heavy. On the other hand, too small embedding layers will not capture all the intricacies of the words. The hidden dimension also plays a vital role in promoting or avoiding overfitting. A large hidden dimension can cause overfitting.