

Intro to NLP

Assignment-2

Name - Malla Sailesh

Roll - 2021101106

FFNN (Hyperparameter Tuning)

Three different configurations i choose for feed forward neural network is:-

1. Activation Function - Tanh and Relu
2. Hidden Layer Size - 50, 100
3. Word Embedding Size - 100, 200

I took the other values constant which are shown below :-

1. Epochs = 5 (These many iterations i used for training the model)
2. $p = 2$
3. $s = 2$

Dev Set Metrics for :-

1.
 - a. Activation Function = "relu"
 - b. Hidden Layer Size = 50
 - c. Embedding Size = 100

Activation Function: relu, Hidden Layer Size: 50, Embedding Size: 100

Epoch: 1, Loss: 1.4223524991969072

Epoch: 2, Loss: 0.569207339508999

Epoch: 3, Loss: 0.343648453783746

Epoch: 4, Loss: 0.23909594144312118

Epoch: 5, Loss: 0.1798851383748335

Validation Accuracy after epoch 5: 96.64%

Validation Metrics

	precision	recall	f1-score	support
ADJ	0.96	0.85	0.90	227
CCONJ	0.97	0.96	0.97	107
PRON	0.99	0.99	0.99	414
PROPN	0.98	0.99	0.98	1551
ADV	0.97	0.53	0.68	59
VERB	0.95	0.96	0.96	653
INTJ	1.00	0.94	0.97	35
PART	0.91	0.84	0.87	73
NUM	0.95	0.89	0.92	131
DET	0.97	0.96	0.97	568
NOUN	0.95	0.98	0.97	1143
ADP	0.96	0.99	0.98	1415
AUX	1.00	0.91	0.95	266
<UNK>	0.00	0.00	0.00	2
accuracy			0.97	6644
macro avg	0.90	0.84	0.86	6644
weighted avg	0.97	0.97	0.97	6644

```
[[ 193    0    0    4    0    1    0    0    4    2   22    1    0    0]
 [   0   103    0    1    0    0    0    0    0    0    0    3    0    0]
 [   0    0   411    0    0    0    0    0    0    2    0    1    0    0]
 [   0    0    0  1534    0    2    0    0    0    1   12    2    0    0]
 [   6    0    1    4   31    6    0    0    0    2    4    5    0    0]
 [   1    0    0    3    0   628    0    0    0    2    8   11    0    0]
 [   0    0    0    0    1    1   33    0    0    0    0    0    0    0]
 [   0    0    0    1    0    1    0   61    0    0    1    9    0    0]
 [   1    0    0    3    0    5    0    0   117    0    5    0    0    0]
 [   0    0    3    4    0    0    0    0    0   544    0   17    0    0]
 [   0    0    0   15    0    5    0    0    1    1  1120    1    0    0]
 [   0    0    0    2    0    1    0    6    0    2    0  1404    0    0]
 [   0    3    1    1    0   12    0    0    1    3    2    1  242    0]
 [   0    0    0    1    0    0    0    0    0    0    1    0    0    0]]
```

2.

a. Activation Function = “relu”

b. Hidden Layer Size = 50

c. Embedding Size = 200

Activation Function: relu, Hidden Layer Size: 50, Embedding Size: 200

Epoch: 1, Loss: 1.246398533237051

Epoch: 2, Loss: 0.38638578939700424

Epoch: 3, Loss: 0.2136154248820646

Epoch: 4, Loss: 0.14239069462474008

Epoch: 5, Loss: 0.10467102263860018

Validation Accuracy after epoch 5: 97.76%

Validation Metrics

	precision	recall	f1-score	support
ADJ	0.95	0.91	0.93	227
CCONJ	1.00	0.98	0.99	107
PRON	0.98	1.00	0.99	414
PROPN	0.98	0.99	0.99	1551
ADV	0.89	0.69	0.78	59
VERB	0.96	0.99	0.98	653
INTJ	1.00	1.00	1.00	35
PART	0.91	0.86	0.89	73
NUM	1.00	0.91	0.95	131
DET	0.99	0.96	0.97	568
NOUN	0.97	0.99	0.98	1143
ADP	0.98	0.99	0.99	1415
AUX	0.99	0.94	0.96	266
<UNK>	0.00	0.00	0.00	2
accuracy			0.98	6644
macro avg	0.90	0.87	0.89	6644
weighted avg	0.98	0.98	0.98	6644

[[207	0	0	2	4	1	0	0	0	0	13	0	0	0]
[[0	105	0	1	0	0	0	0	0	0	0	1	0	0]
[[0	0	412	0	0	0	0	0	0	1	0	0	1	0]
[[1	0	0	1542	0	1	0	0	0	0	6	1	0	0]
[[7	0	0	2	41	0	0	0	0	1	3	4	1	0]
[[1	0	1	2	0	646	0	0	0	1	1	1	0	0]
[[0	0	0	0	0	0	35	0	0	0	0	0	0	0]
[[0	0	0	0	0	0	0	63	0	0	1	8	1	0]
[[2	0	0	4	0	3	0	0	119	0	3	0	0	0]
[[1	0	7	4	0	0	0	0	0	543	0	13	0	0]
[[0	0	0	11	1	4	0	0	0	0	1127	0	0	0]
[[0	0	0	1	0	1	0	6	0	0	1	1406	0	0]
[[0	0	0	1	0	14	0	0	0	0	1	1	249	0]
[[0	0	0	1	0	0	0	0	0	0	1	0	0	0]

3.

a. Activation Function = "relu"

b. Hidden Layer Size = 100

c. Embedding Size = 100

Activation Function: relu, Hidden Layer Size: 100, Embedding Size: 100

Epoch: 1, Loss: 1.462478959152841

Epoch: 2, Loss: 0.5606277207944991

Epoch: 3, Loss: 0.33215319617044603

Epoch: 4, Loss: 0.22762320267574165

Epoch: 5, Loss: 0.169249584206443

Validation Accuracy after epoch 5: 96.73%

Validation Metrics

	precision	recall	f1-score	support
ADJ	0.96	0.85	0.90	227
CCONJ	1.00	0.96	0.98	107
PRON	0.99	0.99	0.99	414
PROPN	0.96	0.99	0.98	1551
ADV	0.82	0.61	0.70	59
VERB	0.96	0.97	0.97	653
INTJ	1.00	1.00	1.00	35
PART	0.92	0.84	0.88	73
NUM	0.99	0.87	0.93	131
DET	0.98	0.95	0.97	568
NOUN	0.95	0.98	0.97	1143
ADP	0.98	0.99	0.98	1415
AUX	0.99	0.92	0.96	266
<UNK>	0.00	0.00	0.00	2
accuracy			0.97	6644
macro avg	0.89	0.85	0.87	6644
weighted avg	0.97	0.97	0.97	6644

```
[[ 194    0    0    4    8    3    0    0    0    2   16    0    0    0]
 [    0   103    0    1    0    0    0    0    0    0    0    3    0    0]
 [    0    0   410    0    0    2    0    0    0    2    0    0    0    0]
 [    0    0    0  1533    0    1    0    0    0    2   15    0    0    0]
 [    5    0    1    9   36    3    0    0    0    0    4    1    0    0]
 [    1    0    0    9    0  635    0    0    1    0    4    2    1    0]
 [    0    0    0    0    0    0   35    0    0    0    0    0    0    0]
 [    0    0    1    0    0    0    0   61    0    1    0    9    1    0]
 [    1    0    0    5    0    0    0    0  114    0   11    0    0    0]
 [    0    0    2    6    0    0    0    0    0  541    1   18    0    0]
 [    1    0    0   18    0    2    0    0    0    2  1120    0    0    0]
 [    0    0    0    5    0    2    0    5    0    1    2  1399    1    0]
 [    0    0    1    3    0   14    0    0    0    1    1    0  246    0]
 [    0    0    0    0    0    0    0    0    0    0    1    1    0    0]]
```

4.
 - a. Activation Function = “relu”
 - b. Hidden Layer Size = 100
 - c. Embedding Size = 200

```

Activation Function: relu, Hidden Layer Size: 100, Embedding Size: 200
Epoch: 1, Loss: 1.2733405955052548
Epoch: 2, Loss: 0.4096704358522787
Epoch: 3, Loss: 0.22302041587313878
Epoch: 4, Loss: 0.14743929887698679
Epoch: 5, Loss: 0.10784528518608728
Validation Accuracy after epoch 5: 97.56%

Validation Metrics

```

	precision	recall	f1-score	support
ADJ	0.96	0.89	0.92	227
CCONJ	1.00	0.98	0.99	107
PRON	0.99	0.99	0.99	414
PROPN	0.98	0.99	0.99	1551
ADV	0.84	0.69	0.76	59
VERB	0.96	0.98	0.97	653
INTJ	1.00	1.00	1.00	35
PART	0.91	0.82	0.86	73
NUM	1.00	0.92	0.96	131
DET	0.99	0.96	0.98	568
NOUN	0.97	0.99	0.98	1143
ADP	0.98	0.99	0.99	1415
AUX	0.99	0.95	0.97	266
<UNK>	0.00	0.00	0.00	2
accuracy			0.98	6644
macro avg	0.90	0.87	0.88	6644
weighted avg	0.98	0.98	0.98	6644

```

[[ 202    0    0    3    7    2    0    0    0    0    13    0    0    0]
 [    0   105    0    1    0    1    0    0    0    0    0    0    0    0]
 [    0    0   411    1    0    0    0    0    0    2    0    0    0    0]
 [    0    0    0  1542    0    1    0    0    0    0    7    1    0    0]
 [    6    0    0    4   41    2    0    0    0    0    3    2    1    0]
 [    1    0    0    7    1  640    0    0    0    0    4    0    0    0]
 [    0    0    0    0    0    0   35    0    0    0    0    0    0    0]
 [    0    0    0    0    0    0    0   60    0    0    2   11    0    0]
 [    1    0    0    3    0    6    0    0   120    0    1    0    0    0]
 [    1    0    6    1    0    0    0    0    0   544    0   16    0    0]
 [    0    0    0   12    0    4    0    0    0    1  1126    0    0    0]
 [    0    0    0    1    0    1    0    6    0    0    2  1403    2    0]
 [    0    0    0    1    0   12    0    0    0    0    0    0  253    0]
 [    0    0    0    1    0    0    0    0    0    0    1    0    0    0]]

```

5.
 - a. Activation Function = “tanh”
 - b. Hidden Layer Size = 50
 - c. Embedding Size = 100

```

Activation Function: tanh, Hidden Layer Size: 50, Embedding Size: 100
Epoch: 1, Loss: 1.5833509154103447
Epoch: 2, Loss: 0.998641802753609
Epoch: 3, Loss: 0.8373002408759364
Epoch: 4, Loss: 0.7551689524935861
Epoch: 5, Loss: 0.7038899771403827
Validation Accuracy after epoch 5: 96.84%

Validation Metrics

```

	precision	recall	f1-score	support
ADJ	0.95	0.90	0.92	227
CCONJ	0.99	0.96	0.98	107
PRON	0.99	0.99	0.99	414
PROPN	0.96	0.99	0.98	1551
ADV	0.97	0.59	0.74	59
VERB	0.96	0.96	0.96	653
INTJ	1.00	0.94	0.97	35
PART	0.87	0.84	0.85	73
NUM	1.00	0.83	0.91	131
DET	0.99	0.96	0.97	568
NOUN	0.95	0.98	0.97	1143
ADP	0.97	0.99	0.98	1415
AUX	0.99	0.93	0.96	266
<UNK>	0.00	0.00	0.00	2
accuracy			0.97	6644
macro avg	0.90	0.85	0.87	6644
weighted avg	0.97	0.97	0.97	6644

```

[[ 204    0    0    3    0    1    0    0    0    2   17    0    0    0]
 [   0  103    0    2    0    0    0    0    0    0    0    2    0    0]
 [   0    0  410    0    0    0    0    0    0    2    0    1    1    0]
 [   1    1    0  1534    0    1    0    2    0    0   11    1    0    0]
 [   7    0    0    8   35    3    0    0    0    1    2    3    0    0]
 [   1    0    1   13    0  627    0    0    0    2    6    3    0    0]
 [   0    0    0    0    0    1   33    0    0    0    1    0    0    0]
 [   0    0    0    0    0    0    0   61    0    0    0   10    2    0]
 [   1    0    0    8    0    2    0    0   109    0   11    0    0    0]
 [   0    0    1    7    0    0    0    0    0    544    1   15    0    0]
 [   1    0    1   12    0    2    0    1    0    1  1124    1    0    0]
 [   0    0    0    1    1    2    0    6    0    0    3  1402    0    0]
 [   0    0    1    1    0   13    0    0    0    0    2    1  248    0]
 [   0    0    0    1    0    0    0    0    0    0    1    0    0    0]]

```

6.
 - a. Activation Function = “tanh”
 - b. Hidden Layer Size = 50
 - c. Embedding Size = 200

Activation Function: tanh, Hidden Layer Size: 50, Embedding Size: 200

Epoch: 1, Loss: 1.3215358708737164
 Epoch: 2, Loss: 0.7624783048027836
 Epoch: 3, Loss: 0.6270256878356256
 Epoch: 4, Loss: 0.5589396399663201
 Epoch: 5, Loss: 0.5175893023796257
 Validation Accuracy after epoch 5: 97.56%

Validation Metrics

	precision	recall	f1-score	support
ADJ	0.96	0.93	0.94	227
CCONJ	1.00	0.96	0.98	107
PRON	1.00	0.99	0.99	414
PROPN	0.98	1.00	0.99	1551
ADV	1.00	0.71	0.83	59
VERB	0.97	0.97	0.97	653
INTJ	1.00	1.00	1.00	35
PART	0.90	0.84	0.87	73
NUM	0.99	0.89	0.94	131
DET	0.99	0.96	0.98	568
NOUN	0.97	0.98	0.98	1143
ADP	0.97	0.99	0.98	1415
AUX	0.98	0.95	0.96	266
<UNK>	0.00	0.00	0.00	2
accuracy			0.98	6644
macro avg	0.91	0.87	0.89	6644
weighted avg	0.98	0.98	0.98	6644

```
[[ 210    0    0    2    0    0    0    0    0    1   14    0    0    0]
 [   0  103    0    1    0    0    0    0    0    0    0    3    0    0]
 [   0    0  410    0    0    0    1    0    2    0    0    1    0    0]
 [   0    0    0 1544    0    1    0    0    0    0    6    0    0    0]
 [   7    0    0    3   42    1    0    0    0    0    3    2    1    0]
 [   0    0    0    5    0  632    0    0    0    0    6    8    2    0]
 [   0    0    0    0    0    0   35    0    0    0    0    0    0    0]
 [   0    0    0    0    0    0    0   61    0    0    0   11    1    0]
 [   1    0    0    3    0    1    0    0  117    0    9    0    0    0]
 [   1    0    2    3    0    0    0    0    0  546    0   16    0    0]
 [   0    0    0   13    0    3    0    0    1    1 1125    0    0    0]
 [   0    0    0    2    0    1    0    6    0    0    0  1405    1    0]
 [   0    0    0    1    0   11    0    0    0    0    0    2  252    0]
 [   0    0    0    1    0    0    0    0    0    0    1    0    0    0]]
```

7.

- a. Activation Function = "tanh"
- b. Hidden Layer Size = 100
- c. Embedding Size = 100

Activation Function: tanh, Hidden Layer Size: 100, Embedding Size: 100

Epoch: 1, Loss: 1.4921653481943682

Epoch: 2, Loss: 0.8510891043536369

Epoch: 3, Loss: 0.6855070125813707

Epoch: 4, Loss: 0.6010680509820571

Epoch: 5, Loss: 0.548747336018015

Validation Accuracy after epoch 5: 96.15%

Validation Metrics

	precision	recall	f1-score	support
ADJ	0.94	0.85	0.89	227
CCONJ	0.96	0.95	0.96	107
PRON	0.97	0.99	0.98	414
PROPN	0.97	0.98	0.98	1551
ADV	0.89	0.53	0.66	59
VERB	0.96	0.95	0.96	653
INTJ	1.00	1.00	1.00	35
PART	0.90	0.74	0.81	73
NUM	1.00	0.80	0.89	131
DET	0.98	0.96	0.97	568
NOUN	0.93	0.98	0.96	1143
ADP	0.97	0.99	0.98	1415
AUX	0.98	0.95	0.97	266
<UNK>	0.00	0.00	0.00	2
accuracy			0.96	6644
macro avg	0.89	0.83	0.86	6644
weighted avg	0.96	0.96	0.96	6644

```
[[ 192    0    1    1    4    2    0    0    0    3   24    0    0    0]
 [    0  102    0    1    0    0    0    0    0    0    0    4    0    0]
 [    0    0  409    1    0    1    0    0    0    2    1    0    0    0]
 [    0    0    1 1523    0    1    0    0    0    3   18    5    0    0]
 [    7    0    3    7   31    4    0    0    0    2    3    1    1    0]
 [    0    0    0    4    0  619    0    1    0    2   14   11    2    0]
 [    0    0    0    0    0    0   35    0    0    0    0    0    0    0]
 [    0    0    0    1    0    1    0   54    0    0    0   16    1    0]
 [    3    0    0    6    0    0    0    0  105    0   17    0    0    0]
 [    0    3    3    7    0    0    0    0    0  543    0   12    0    0]
 [    2    0    3   15    0    4    0    0    0    1 1118    0    0    0]
 [    0    0    1    3    0    2    0    5    0    0    0 1403    1    0]
 [    0    1    0    1    0    9    0    0    0    0    0    1  254    0]
 [    0    0    0    1    0    0    0    0    0    0    1    0    0    0]]
```


8.

- a. Activation Function = “tanh”
- b. Hidden Layer Size = 100
- c. Embedding Size = 200

```
Activation Function: tanh, Hidden Layer Size: 100, Embedding Size: 200
Epoch: 1, Loss: 1.2360339557787527
Epoch: 2, Loss: 0.6123031798669337
Epoch: 3, Loss: 0.4647368296520401
Epoch: 4, Loss: 0.3910566715720764
Epoch: 5, Loss: 0.3463024226146494
Validation Accuracy after epoch 5: 97.23%

Validation Metrics
```

	precision	recall	f1-score	support
ADJ	0.94	0.90	0.92	227
CCONJ	1.00	0.97	0.99	107
PRON	0.99	0.99	0.99	414
PROPN	0.97	0.99	0.98	1551
ADV	0.92	0.61	0.73	59
VERB	0.98	0.97	0.98	653
INTJ	1.00	1.00	1.00	35
PART	0.91	0.82	0.86	73
NUM	1.00	0.85	0.92	131
DET	0.99	0.95	0.97	568
NOUN	0.96	0.99	0.97	1143
ADP	0.97	0.99	0.98	1415
AUX	0.98	0.95	0.97	266
<UNK>	0.00	0.00	0.00	2
accuracy			0.97	6644
macro avg	0.90	0.86	0.88	6644
weighted avg	0.97	0.97	0.97	6644

```
[[ 205  0  0  2  3  1  0  0  0  2  14  0  0  0]
 [  0 104  0  1  0  0  0  0  0  0  0  2  0  0]
 [  0  0 408  0  0  0  0  0  0  3  1  1  1  0]
 [  1  0  0 1539  0  1  0  0  0  0 10  0  0  0]
 [  7  0  2  6 36  0  0  0  0  1  4  3  0  0]
 [  1  0  0  9  0 636  0  0  0  1  3  1  2  0]
 [  0  0  0  0  0  0 35  0  0  0  0  0  0  0]
 [  0  0  0  0  0  0  0 60  0  0  0 12  1  0]
 [  3  0  0  6  0  0  0  0 111  0 10  1  0  0]
 [  1  0  2  4  0  0  0  0  0 541  0 20  0  0]
 [  0  0  1 13  0  2  0  0  0  0 1127  0  0  0]
 [  0  0  0  2  0  1  0  6  0  0  0 1405  1  0]
 [  0  0  0  1  0 10  0  0  0  0  0  2 253  0]
 [  0  0  0  1  0  0  0  0  0  0  1  0  0  0]]
```

1. Relu is generally better to use than tanh activation function because tanh usually has vanishing gradient problem .

Note :- For relu also on the negative x axis , vanishing gradient problem persist. So leaky relu is sometimes preferable.

2. Word Embedding Size - 200 is preferable over Word Embedding Size - 100 because if the word embedding size is more it can capture most of the features of the word. So it is better to predict the pos tag .

Note :- If the word embedding size is too much , it may lead to overfitting. So too much word embedding size is also not good.

3. Possible reasons why a hidden neuron size of 50 might outperform a size of 100 in a neural network:

1. Overfitting

- **Memorization vs. Generalization:** A larger hidden layer provides the network with more capacity to fit complex relationships in the data. However, if your dataset is relatively small or contains noise, the model with 100 neurons might start "memorizing" the specifics of the training data rather than learning generalizable patterns. This leads to overfitting, where performance suffers on unseen data.

2. Insufficient Training Data

- **Data vs. Parameters:** When the number of parameters in your model (which increases with hidden neuron size) is large compared to the amount of training data, the network might not have enough examples to learn the optimal weights. This can lead to subpar results.

3. Task Complexity

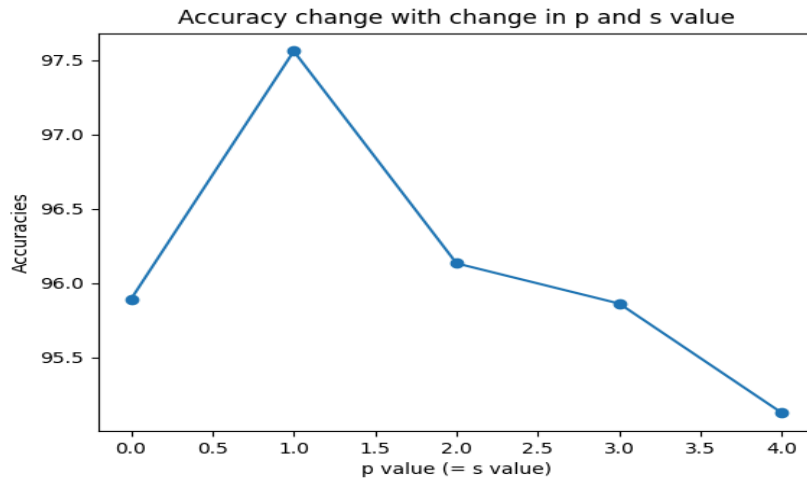
- **The Right Fit:** Sometimes, the underlying patterns in your data might be well-represented by a smaller hidden layer. Increasing the size

Best Metrics for FFNN tuning are :-

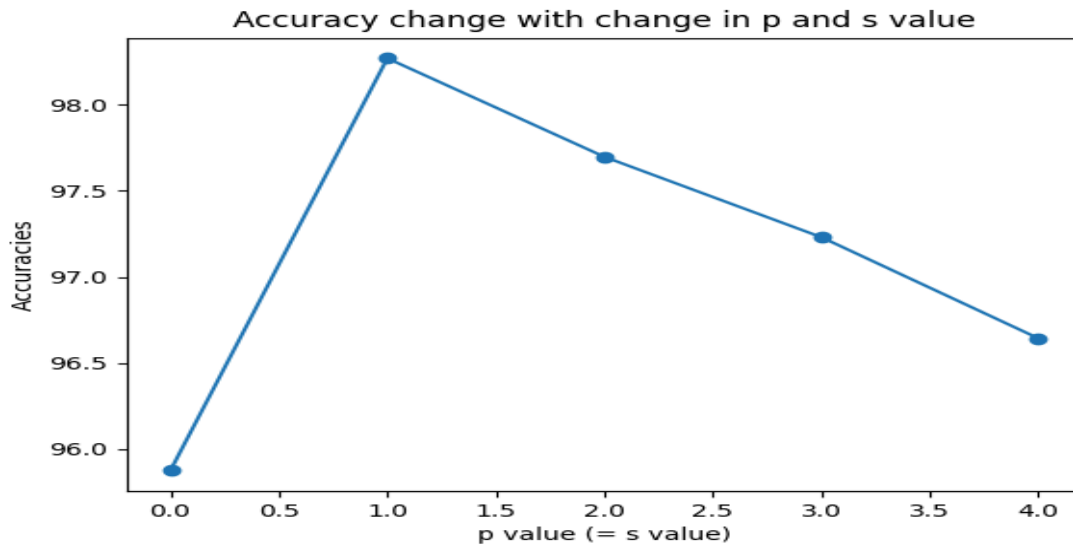
- ## Test Set Metrics for Best hyperparameters :-

Plots for change in $p(=s)$ affects the accuracy :-

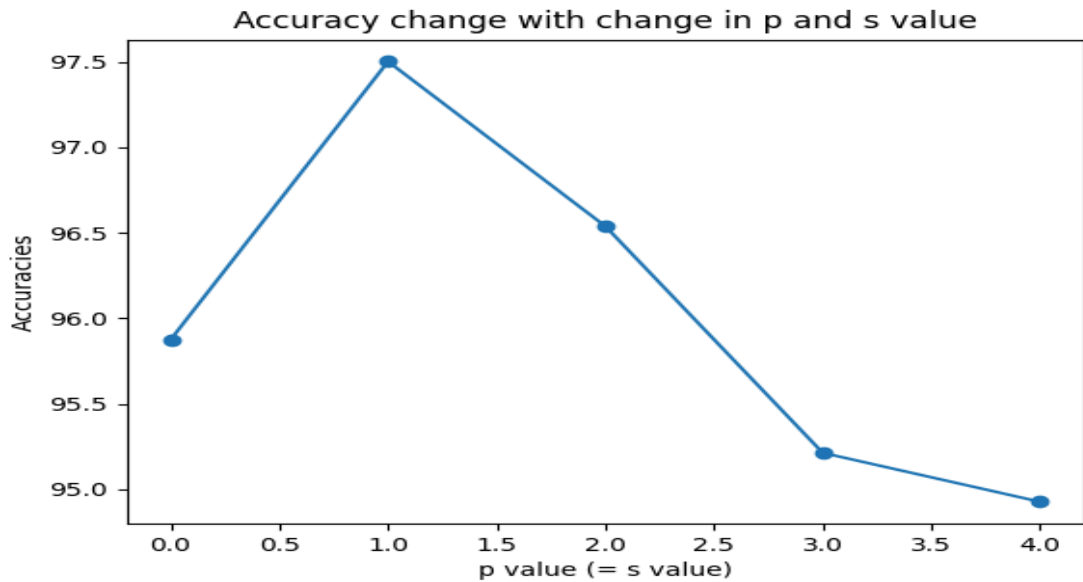
1. Hidden Layer Size - 50 , Embedding Size - 100



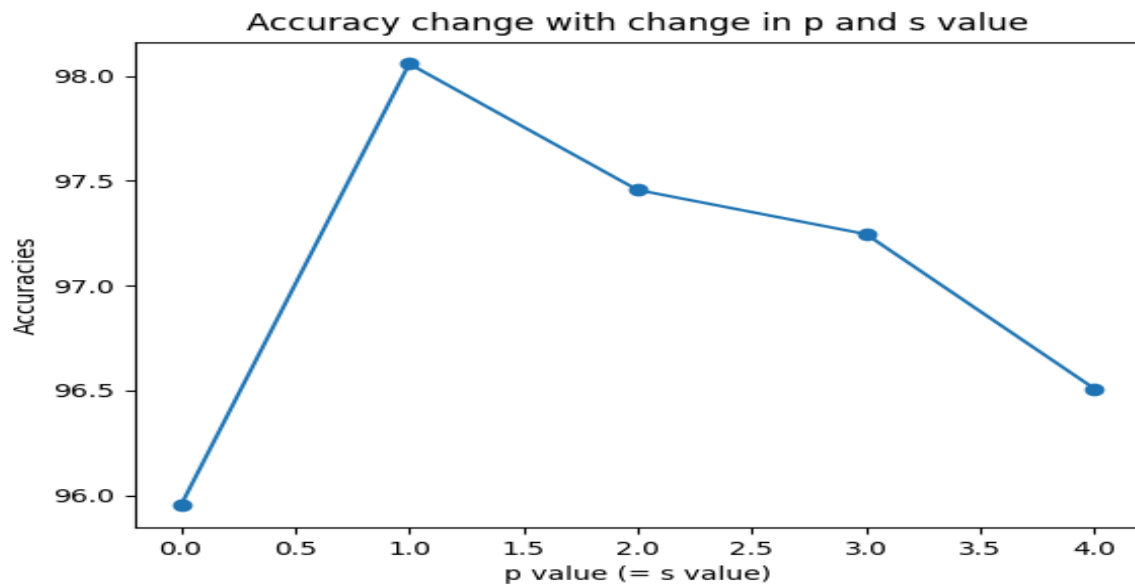
2. Hidden Layer Size - 50, Embedding Size - 200



3. Hidden Layer Size - 100, Embedding Size - 100



4. Hidden Layer Size - 100, Embedding Size - 200



We can observe that for any configuration the accuracy turns out to be highest in the case of $p = s = 1$.

RNN (Hyperparameter Tuning)

Three different configurations i choose for recurrent neural network is:-

1. Hidden Dimension of LSTM - 50, 100
2. Word Embedding Size - 100, 200
3. Number of LSTM Layers - 1 , 2

I took the other values constant which are shown below :-

1. Epochs = 5 (These many iterations i used for training the model)

Dev Set Metrics for :-

1.
 - a. Hidden Dimension of LSTM - 50
 - b. Word Embedding Size - 100
 - c. Number of LSTM Layers - 1

```
Hidden Dimension: 50, Embedding Size: 100, Number of Layers: 1
Epoch: 1, Loss: 1.813390919550814
Epoch: 2, Loss: 1.3373383238430496
Epoch: 3, Loss: 1.1999951038173162
Epoch: 4, Loss: 1.1340316051794204
Epoch: 5, Loss: 1.0960455446612574
Validation Accuracy after epoch 5: 95.85%

Validation Metrics
```

	precision	recall	f1-score	support
NOUN	0.96	0.98	0.97	1143
INTJ	0.97	1.00	0.99	35
CCONJ	1.00	1.00	1.00	107
ADP	0.93	1.00	0.97	1415
PART	0.00	0.00	0.00	73
AUX	0.94	1.00	0.97	266
NUM	1.00	0.89	0.94	131
DET	0.99	0.86	0.92	568
PRON	0.86	1.00	0.92	414
PROPN	0.99	0.99	0.99	1551
VERB	0.99	0.97	0.98	653
ADV	0.83	0.81	0.82	59
ADJ	0.97	0.89	0.93	227
<UNK>	0.00	0.00	0.00	2
accuracy			0.96	6644
macro avg	0.82	0.81	0.81	6644
weighted avg	0.95	0.96	0.95	6644

```
[[1115 0 0 0 0 0 0 2 3 13 9 0 1 0]
 [0 35 0 0 0 0 0 0 0 0 0 0 0 0]
 [[0 0 107 0 0 0 0 0 0 0 0 0 0 0]
 [[0 0 0 1414 0 0 0 0 1 0 0 0 0 0]
 [[0 0 0 71 0 2 0 0 0 0 0 0 0 0]
 [[1 0 0 0 0 265 0 0 0 0 0 0 0 0]
 [[5 0 0 0 7 0 116 0 0 2 0 0 0 1]
 [[0 0 0 16 0 0 0 490 61 0 0 1 0 0]
 [[0 0 0 0 0 1 0 1 412 0 0 0 0 0]
 [[14 0 0 2 0 2 0 1 0 1531 0 1 0 0]
 [[5 0 0 1 0 11 0 0 1 2 633 0 0 0]
 [[1 1 0 2 0 0 0 0 2 0 0 48 5 0]
 [[14 0 0 0 0 0 0 1 0 2 0 8 202 0]
 [[2 0 0 0 0 0 0 0 0 0 0 0 0 0]]
```

2.

a. Hidden Dimension of LSTM - 50

b. Word Embedding Size - 100

c. Number of LSTM Layers - 2

Hidden Dimension: 50, Embedding Size: 100, Number of Layers: 2

Epoch: 1, Loss: 2.458083955216307

Epoch: 2, Loss: 2.3733881501098297

Epoch: 3, Loss: 2.3485556836433927

Epoch: 4, Loss: 2.336937159146446

Epoch: 5, Loss: 2.330189130883045

Validation Accuracy after epoch 5: 76.49%

Validation Metrics

	precision	recall	f1-score	support
NOUN	0.74	0.84	0.79	1143
INTJ	1.00	0.77	0.87	35
CCONJ	0.60	0.03	0.05	107
ADP	0.74	0.91	0.82	1415
PART	0.00	0.00	0.00	73
AUX	0.60	0.86	0.71	266
NUM	0.86	0.57	0.69	131
DET	0.71	0.58	0.64	568
PRON	1.00	0.29	0.45	414
PROPN	0.77	0.98	0.86	1551
VERB	0.99	0.65	0.79	653
ADV	0.00	0.00	0.00	59
ADJ	0.94	0.46	0.62	227
<UNK>	0.00	0.00	0.00	2
accuracy			0.76	6644
macro avg	0.64	0.50	0.52	6644
weighted avg	0.77	0.76	0.74	6644

```
[[ 955    0    0  105    0    2    0    1    0   76    3    0    1    0]
 [    2   27    0    0    1    0    0    0    0    5    0    0    0    0]
 [   23    0    3    1    0    0    4    0    0   75    0    0    1    0]
 [    0    0    0 1294    0    0    0    0    0  120    0    0    1    0]
 [    4    0    0   67    0    0    0    0    0    2    0    0    0    0]
 [    8    0    0   21    0  228    0    1    0    8    0    0    0    0]
 [   16    0    0   37    0    0   75    0    0    3    0    0    0    0]
 [  154    0    0   44    0    1    0  328    0   38    1    0    2    0]
 [   11    0    2   14    0  126    0  125  120   15    0    0    1    0]
 [   16    0    0   11    0    3    1    0    0 1520    0    0    0    0]
 [   16    0    0  130    0   13    0    2    0   64  427    0    1    0]
 [    4    0    0   15    1    2    7    3    0   27    0    0    0    0]
 [   78    0    0   21    0    5    0    3    0   13    2    0  105    0]
 [    2    0    0    0    0    0    0    0    0    0    0    0    0    0]]
```

3.

a. Hidden Dimension of LSTM - 50

b. Word Embedding Size - 200

c. Number of LSTM Layers - 1

Hidden Dimension: 50, Embedding Size: 200, Number of Layers: 1

Epoch: 1, Loss: 1.3636470052143876

Epoch: 2, Loss: 0.8952956257939729

Epoch: 3, Loss: 0.8051176607873767

Epoch: 4, Loss: 0.767542255185562

Epoch: 5, Loss: 0.7473262028126277

Validation Accuracy after epoch 5: 96.06%

Validation Metrics

	precision	recall	f1-score	support
NOUN	0.97	0.98	0.97	1143
INTJ	0.97	1.00	0.99	35
CCONJ	1.00	1.00	1.00	107
ADP	0.94	1.00	0.97	1415
PART	0.00	0.00	0.00	73
AUX	0.95	1.00	0.97	266
NUM	1.00	0.89	0.94	131
DET	0.99	0.87	0.93	568
PRON	0.87	1.00	0.93	414
PROPN	0.98	0.99	0.99	1551
VERB	0.99	0.96	0.98	653
ADV	0.86	0.81	0.83	59
ADJ	0.96	0.89	0.93	227
<UNK>	0.00	0.00	0.00	2
accuracy			0.96	6644
macro avg	0.82	0.81	0.82	6644
weighted avg	0.95	0.96	0.95	6644

```
[[1119  0  0  1  0  0  0  2  0 13  7  0  1  0]
 [  0 35  0  0  0  0  0  0  0  0  0  0  0  0]
 [  0  0 107  0  0  0  0  0  0  0  0  0  0  0]
 [  1  0  0 1414  0  0  0  0  0  0  0  0  0  0]
 [  0  0  0  71  0  2  0  0  0  0  0  0  0  0]
 [  1  0  0  0  0 265  0  0  0  0  0  0  0  0]
 [  5  0  0  6  0  0 116  0  0  4  0  0  0  0]
 [  0  0  0 12  0  0  0 494 62  0  0  0  0  0]
 [  0  0  0  0  0  1  0  1 412  0  0  0  0  0]
 [  9  0  0  1  0  2  0  0  0 1539  0  0  0  0]
 [  6  0  0  1  0 10  0  0  1  4 630  0  1  0]
 [  1  1  0  2  0  0  0  0  1  0  0 48  6  0]
 [ 14  0  0  0  0  0  0  0  0  2  0  8 203  0]
 [  1  0  0  0  0  0  0  0  0  1  0  0  0  0]]
```


4.
 - a. Hidden Dimension of LSTM - 50
 - b. Word Embedding Size - 200
 - c. Number of LSTM Layers - 2

b. Word Embedding Size - 200

c. Number of LSTM Layers - 2

```
Epoch: 1, Loss: 2.409273564620279
Epoch: 2, Loss: 2.3099759021651036
Epoch: 3, Loss: 2.2888294754599543
Epoch: 4, Loss: 2.279449697292794
Epoch: 5, Loss: 2.27424480427522
Validation Accuracy after epoch 5: 54.77%
```

Validation Metrics															
					precision	recall	f1-score	support							
NOUN					0.95	0.43	0.59	1143							
INTJ					1.00	0.57	0.73	35							
CCONJ					0.00	0.00	0.00	107							
ADP					1.00	0.02	0.03	1415							
PART					0.00	0.00	0.00	73							
AUX					0.95	0.75	0.84	266							
NUM					1.00	0.15	0.25	131							
DET					0.99	0.59	0.74	568							
PRON					0.73	0.94	0.82	414							
PROPN					0.37	1.00	0.54	1551							
VERB					0.77	0.66	0.71	653							
ADV					0.49	0.56	0.52	59							
ADJ					0.79	0.65	0.71	227							
<UNK>					0.00	0.00	0.00	2							
accuracy								0.55	6644						
macro avg					0.65	0.45	0.46	6644							
weighted avg					0.76	0.55	0.49	6644							
[[491	0	0	0	0	2	0	3	32	559	41	4	11	0]
[[0	20	0	0	0	0	0	0	8	7	0	0	0	0]
[[1	0	0	0	0	0	0	0	0	104	0	0	2	0]
[[6	0	0	23	0	0	0	0	0	1357	21	0	8	0]
[[1	0	0	0	0	0	0	0	0	32	40	0	0	0]
[[0	0	0	0	0	199	0	0	17	40	9	1	0	0]
[[0	0	0	0	0	0	19	0	3	93	0	1	15	0]
[[0	0	0	0	0	0	0	335	76	141	6	9	1	0]
[[0	0	0	0	0	0	0	1	391	19	2	1	0	0]
[[2	0	0	0	0	0	0	0	0	1549	0	0	0	0]
[[1	0	0	0	0	8	0	0	4	208	431	0	1	0]
[[0	0	0	0	0	0	0	0	4	17	3	33	2	0]
[[13	0	0	0	0	0	0	0	0	44	4	18	148	0]
[[0	0	0	0	0	0	0	0	0	2	0	0	0	0]

[491	0	0	0	0	2	0	3	32	559	41	4	11	0]
[0	20	0	0	0	0	0	0	8	7	0	0	0	0]
[1	0	0	0	0	0	0	0	0	104	0	0	2	0]
[6	0	0	23	0	0	0	0	0	1357	21	0	8	0]
[1	0	0	0	0	0	0	0	0	32	40	0	0	0]
[0	0	0	0	0	199	0	0	17	40	9	1	0	0]
[0	0	0	0	0	0	19	0	3	93	0	1	15	0]
[0	0	0	0	0	0	0	335	76	141	6	9	1	0]
[0	0	0	0	0	0	0	1	391	19	2	1	0	0]
[2	0	0	0	0	0	0	0	0	1549	0	0	0	0]
[1	0	0	0	0	8	0	0	4	208	431	0	1	0]
[0	0	0	0	0	0	0	0	4	17	3	33	2	0]
[13	0	0	0	0	0	0	0	0	44	4	18	148	0]
[0	0	0	0	0	0	0	0	0	2	0	0	0	0]

5.

- a. Hidden Dimension of LSTM - 100
- b. Word Embedding Size - 100
- c. Number of LSTM Layers - 1

Hidden Dimension: 100, Embedding Size: 100, Number of Layers: 1

Epoch: 1, Loss: 1.7863093224512288

Epoch: 2, Loss: 1.207481659248046

Epoch: 3, Loss: 1.0090896128687727

Epoch: 4, Loss: 0.9064016580776907

Epoch: 5, Loss: 0.842311381986892

Validation Accuracy after epoch 5: 95.86%

Validation Metrics

	precision	recall	f1-score	support
NOUN	0.97	0.97	0.97	1143
INTJ	0.97	1.00	0.99	35
CCONJ	1.00	1.00	1.00	107
ADP	0.95	1.00	0.97	1415
PART	0.87	0.18	0.30	73
AUX	0.93	1.00	0.96	266
NUM	1.00	0.85	0.92	131
DET	0.98	0.87	0.92	568
PRON	0.87	0.99	0.92	414
PROPN	0.98	0.99	0.98	1551
VERB	0.98	0.96	0.97	653
ADV	0.94	0.76	0.84	59
ADJ	0.97	0.90	0.93	227
<UNK>	0.00	0.00	0.00	2
accuracy			0.96	6644
macro avg	0.89	0.82	0.83	6644
weighted avg	0.96	0.96	0.96	6644

```
[[1109 0 0 0 0 1 0 3 1 16 12 0 1 0]
 [ 0 35 0 0 0 0 0 0 0 0 0 0 0 0]
 [ 0 0 107 0 0 0 0 0 0 0 0 0 0 0]
 [ 0 0 0 1412 2 0 0 0 1 0 0 0 0 0]
 [ 0 0 0 58 13 2 0 0 0 0 0 0 0 0]
 [ 1 0 0 0 0 265 0 0 0 0 0 0 0 0]
 [ 6 0 0 7 0 0 112 2 0 4 0 0 0 0]
 [ 0 0 0 13 0 1 0 493 58 3 0 0 0 0]
 [ 0 0 0 0 0 2 0 1 410 1 0 0 0 0]
 [ 11 0 0 1 0 2 0 1 0 1536 0 0 0 0]
 [ 6 0 0 1 0 10 0 0 1 7 628 0 0 0]
 [ 1 1 0 2 0 1 0 0 2 1 1 45 5 0]
 [ 14 0 0 0 0 1 0 1 0 4 0 3 204 0]
 [ 1 0 0 0 0 0 0 0 0 1 0 0 0 0]]
```

6.

- a. Hidden Dimension of LSTM - 100
- b. Word Embedding Size - 100
- c. Number of LSTM Layers - 2

Hidden Dimension: 100, Embedding Size: 100, Number of Layers: 2

Epoch: 1, Loss: 2.4528607286244393
Epoch: 2, Loss: 2.3423350952917805
Epoch: 3, Loss: 2.303497069284745
Epoch: 4, Loss: 2.2827426488007907
Epoch: 5, Loss: 2.2697409476751917
Validation Accuracy after epoch 5: 75.45%

Validation Metrics

	precision	recall	f1-score	support
NOUN	0.73	0.83	0.77	1143
INTJ	0.85	0.83	0.84	35
CCONJ	1.00	0.04	0.07	107
ADP	0.71	0.87	0.78	1415
PART	0.00	0.00	0.00	73
AUX	0.94	0.70	0.80	266
NUM	1.00	0.02	0.03	131
DET	0.94	0.58	0.71	568
PRON	0.86	0.86	0.86	414
PROPN	0.70	0.98	0.82	1551
VERB	0.98	0.56	0.71	653
ADV	0.66	0.36	0.46	59
ADJ	0.95	0.08	0.15	227
<UNK>	0.00	0.00	0.00	2
accuracy			0.75	6644
macro avg	0.74	0.48	0.50	6644
weighted avg	0.79	0.75	0.72	6644

```
[[ 946    0    0    79    0    1    0    0    0    111    6    0    0    0]
 [    1   29    0    0    0    0    0    0    0    5    0    0    0    0]
 [   18    0    4    1    0    0    0    1    0    83    0    0    0    0]
 [    9    0    0 1231    0    0    0    0    0    175    0    0    0    0]
 [    1    0    0    70    0    0    0    0    0    2    0    0    0    0]
 [   14    4    0    23    0  185    0    9    1    25    0    5    0    0]
 [   66    0    0    40    0    0    2    0    0    23    0    0    0    0]
 [   30    0    0   128    0    0    0  327   53   28    1    0    1    0]
 [   11    0    0    22    0    3    0    4  355   19    0    0    0    0]
 [   13    0    0    7    0    1    0    4    0 1526    0    0    0    0]
 [   29    0    0   118    0    6    0    1    2  129  368    0    0    0]
 [   17    1    0    4    0    0    0    0    3   13    0   21    0    0]
 [  147    0    0   16    0    1    0    3    0   33    2    6   19    0]
 [    1    0    0    0    0    0    0    0    0    1    0    0    0    0]]
```

7.

- Hidden Dimension of LSTM - 100
- Word Embedding Size - 200
- Number of LSTM Layers - 1

Hidden Dimension: 100, Embedding Size: 200, Number of Layers: 1

Epoch: 1, Loss: 1.4038235794649985

Epoch: 2, Loss: 0.8184223527251754

Epoch: 3, Loss: 0.6717969160326519

Epoch: 4, Loss: 0.6059743203763154

Epoch: 5, Loss: 0.568827796370056

Validation Accuracy after epoch 5: 95.95%

Validation Metrics

	precision	recall	f1-score	support
NOUN	0.97	0.98	0.97	1143
INTJ	0.97	1.00	0.99	35
CCONJ	0.99	1.00	1.00	107
ADP	0.94	1.00	0.97	1415
PART	0.00	0.00	0.00	73
AUX	0.94	1.00	0.97	266
NUM	1.00	0.88	0.93	131
DET	0.99	0.87	0.92	568
PRON	0.87	1.00	0.93	414
PROPN	0.98	0.99	0.99	1551
VERB	0.98	0.97	0.98	653
ADV	0.87	0.80	0.83	59
ADJ	0.97	0.89	0.93	227
<UNK>	0.00	0.00	0.00	2
accuracy			0.96	6644
macro avg	0.82	0.81	0.81	6644
weighted avg	0.95	0.96	0.95	6644

```
[[1115    0    0    0    0    1    0    2    1    13    10    0    1    0]
[    0   35    0    0    0    0    0    0    0    0    0    0    0    0]
[    0    0   107    0    0    0    0    0    0    0    0    0    0    0]
[    0    0    0  1414    0    0    0    0    1    0    0    0    0    0]
[    0    0    0    71    0    2    0    0    0    0    0    0    0    0]
[    1    0    0    0    0  265    0    0    0    0    0    0    0    0]
[    2    0    0    7    0    0   115    0    0    7    0    0    0    0]
[    0    0    0   16    0    0    0   493   59    0    0    0    0    0]
[    0    0    0    0    0    1    0    1  412    0    0    0    0    0]
[    7    0    0    1    0    2    0    1    0  1540    0    0    0    0]
[    5    0    0    1    0   10    0    0    1    5   631    0    0    0]
[    1    1    0    2    0    1    0    0    2    0    0   47    5    0]
[   14    0    1    0    0    0    0    1    0    3    0    7   201    0]
[    1    0    0    0    0    0    0    0    0    1    0    0    0    0]]
```

8.

- a. Hidden Dimension of LSTM - 100
- b. Word Embedding Size - 200
- c. Number of LSTM Layers - 2

Hidden Dimension: 100, Embedding Size: 200, Number of Layers: 2

Epoch: 1, Loss: 2.4030366030930916

Epoch: 2, Loss: 2.288190648106141

Epoch: 3, Loss: 2.256006737038751

Epoch: 4, Loss: 2.2404677423295096

Epoch: 5, Loss: 2.2314846256820187

Validation Accuracy after epoch 5: 76.87%

Validation Metrics

	precision	recall	f1-score	support
NOUN	0.76	0.82	0.79	1143
INTJ	1.00	0.71	0.83	35
CCONJ	1.00	0.03	0.05	107
ADP	0.67	0.98	0.79	1415
PART	0.00	0.00	0.00	73
AUX	0.93	0.84	0.88	266
NUM	0.97	0.29	0.45	131
DET	1.00	0.48	0.65	568
PRON	0.64	0.95	0.76	414
PROPN	0.85	0.98	0.91	1551
VERB	0.93	0.41	0.57	653
ADV	0.74	0.39	0.51	59
ADJ	0.96	0.11	0.20	227
<UNK>	0.00	0.00	0.00	2
accuracy			0.77	6644
macro avg	0.75	0.50	0.53	6644
weighted avg	0.80	0.77	0.74	6644

```
[[ 938   0   0  124   0   1   0   0   4  70   6   0   0   0]
 [   0  25   0   5   0   0   0   0   3   2   0   0   0   0]
 [  18   0   3   6   0   0   1   0   0  79   0   0   0   0]
 [   0   0   0 1380   0   0   0   0   2  32   1   0   0   0]
 [   1   0   0   69   0   0   0   0   0   1   2   0   0   0]
 [  24   0   0   9   0 223   0   0   6   3   0   1   0   0]
 [  24   0   0   64   0   0  38   0   1   4   0   0   0   0]
 [  74   0   0  134   0   1   0 273  65  11   9   0   1   0]
 [   1   0   0   7   0   2   0   0 395   9   0   0   0   0]
 [  14   0   0   17   0   2   0   0   2 1516   0   0   0   0]
 [  16   0   0  181   0  10   0   0  134  45 267   0   0   0]
 [   5   0   0   19   0   0   0   0   4   7   1  23   0   0]
 [ 122   0   0   56   0   1   0   0   6   8   1   7  26   0]
 [   2   0   0   0   0   0   0   0   0   0   0   0   0   0]]
```

1. **Word Embedding Size - 200 is preferable over Word Embedding Size - 100** because if the word embedding size is more it can capture most of the features of the word. So it is better to predict the pos tag .

Note :- If the word embedding size is too much , it may lead to overfitting. So too much word embedding size is also not good.

2. **Better accuracy with a single-layer LSTM compared to a two-layer LSTM:**

Possible Reasons:

- **Overfitting:** With more layers, LSTMs become more complex models with a larger number of parameters. This can make them more prone to overfitting, especially if your dataset is relatively small or lacks sufficient variability. A simpler model (one layer) might better generalize to unseen data.
- **Vanishing Gradients:** Even though LSTMs are designed to mitigate vanishing gradients, very deep networks can still suffer from this issue. Information flow over multiple layers can become less effective. This might be a factor if your sequences are long.
- **Nature of the Task:** Sometimes, the underlying patterns in your data may not require the hierarchical complexity offered by multiple LSTM layers. A single layer might adequately capture the necessary sequential dependencies.
- **Insufficient Regularization:** Multi-layered models often benefit from techniques like dropout or weight decay to combat overfitting. If you haven't tuned these techniques appropriately, the two-layer LSTM might suffer.

3. Better accuracy with hidden dimension of 50 than hidden dimension of 100.

Possible Reasons :-

- **Overfitting:** LSTMs with a large hidden dimension have more parameters, increasing their capacity to learn complex patterns. However, this also makes them more susceptible to overfitting, especially with smaller datasets. They might end up memorizing the training data's noise instead of generalizing well.
- **Insufficient Data:** When your dataset is limited, a large hidden dimension might lead the model to discover spurious relationships that don't exist in the overall population of data.
- **Task Complexity:** The optimal hidden dimension depends on the complexity of the patterns within your data. Sometimes, a smaller dimension is sufficient to capture the important dependencies, while a larger one doesn't add significant value.

Best Metrics for RNN tuning are :-

1. Hidden Dimension of LSTM - 50
2. Word Embedding Size - 200
3. Number of LSTM Layers - 1

Test Set Metrics for Best hyperparameters :-

Test Accuracy: 96.22%

Test Metrics

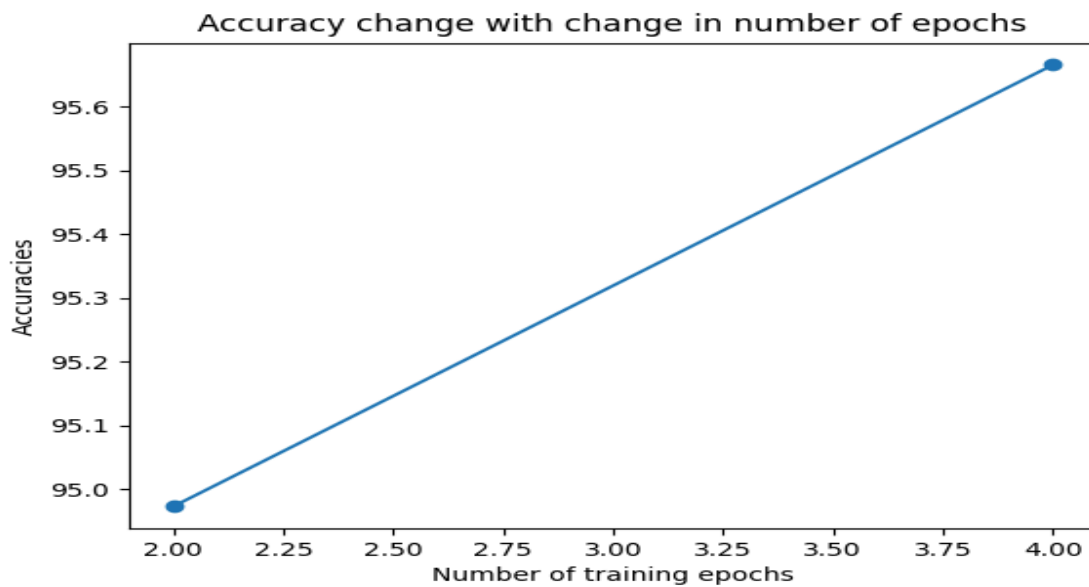
	precision	recall	f1-score	support
NOUN	0.97	0.98	0.97	1143
INTJ	0.97	1.00	0.99	35
CCONJ	1.00	1.00	1.00	107
ADP	0.95	1.00	0.97	1415
PART	1.00	0.30	0.46	73
AUX	0.94	1.00	0.97	266
NUM	0.99	0.89	0.94	131
DET	0.99	0.86	0.92	568
PRON	0.86	0.99	0.92	414
PROPN	0.99	0.99	0.99	1551
VERB	0.98	0.97	0.97	653
ADV	0.89	0.80	0.84	59
ADJ	0.97	0.89	0.92	227
<UNK>	0.00	0.00	0.00	2
accuracy			0.96	6644
macro avg	0.89	0.83	0.85	6644
weighted avg	0.96	0.96	0.96	6644

```
[[1119    0    0    0    0    1    0    2    0   10   10    0    1    0]
 [    0   35    0    0    0    0    0    0    0    0    0    0    0    0]
 [    0    0  107    0    0    0    0    0    0    0    0    0    0    0]
 [    1    0    0 1414    0    0    0    0    0    0    0    0    0    0]
 [    0    0    0   49   22    2    0    0    0    0    0    0    0    0]
 [    1    0    0    0    0  265    0    0    0    0    0    0    0    0]
 [    4    0    0    6    0    0  117    0    0    4    0    0    0    0]
 [    0    0    0   13    0    0    0  490   64    1    0    0    0    0]
 [    0    0    0    0    0    1    0    0  411    1    1    0    0    0]
 [   11    0    0    2    0    2    0    1    0 1535    0    0    0    0]
 [    6    0    0    1    0   11    0    0    1    1  632    0    1    0]
 [    1    1    0    1    0    0    1    0    2    0    1   47    5    0]
 [   14    0    0    1    0    1    0    1    0    3    0    6  201    0]
 [    1    0    0    0    0    0    0    0    0    1    0    0    0    0]]
```


Plots for change in number of epochs affects accuracy :-

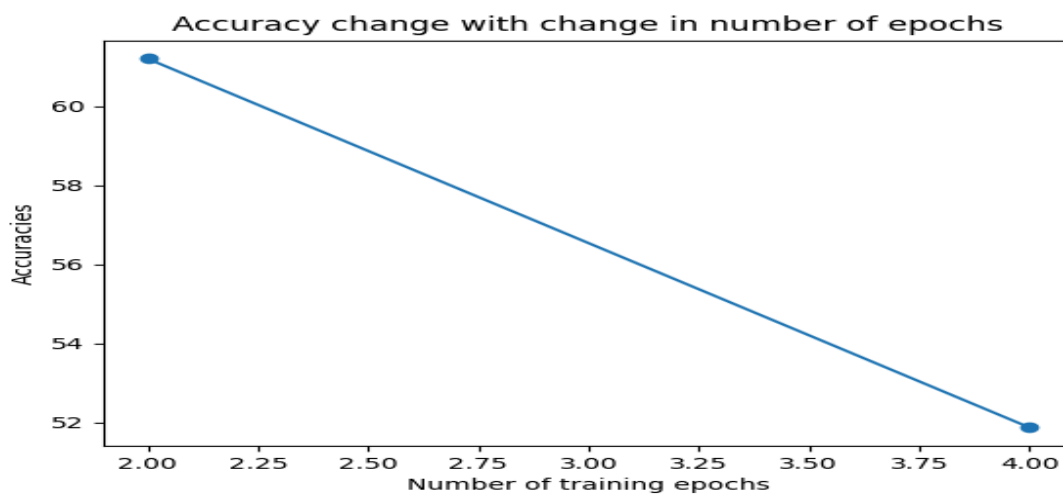
1. Hidden Dimension of LSTM - 50 , Word Embedding Size - 100 ,

Number of LSTM Layers - 1



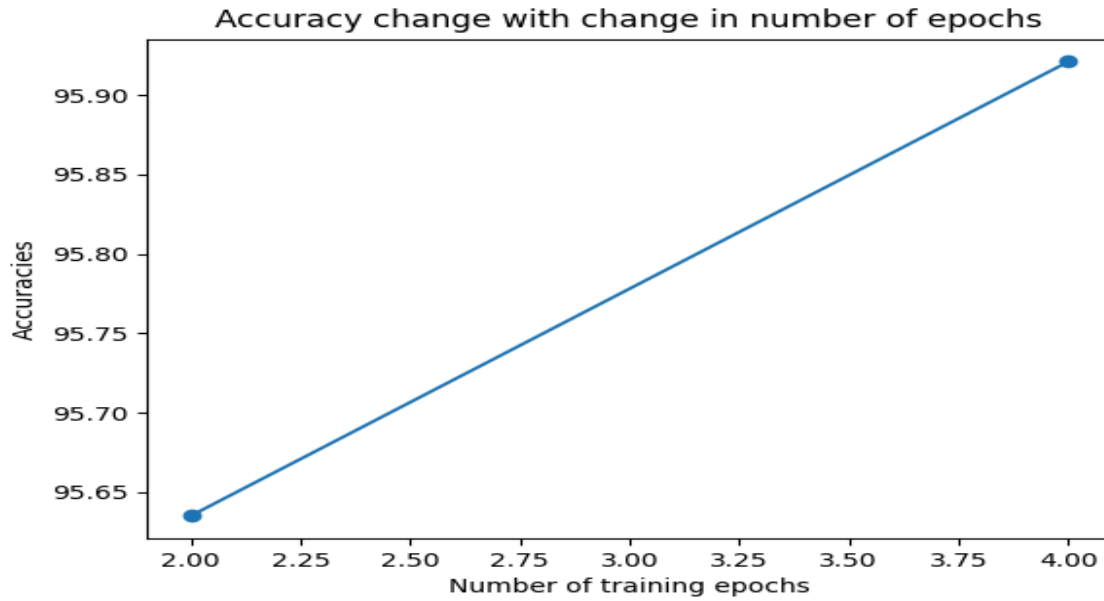
2. Hidden Dimension of LSTM - 50 , Word Embedding Size - 100 ,

Number of LSTM Layers - 2



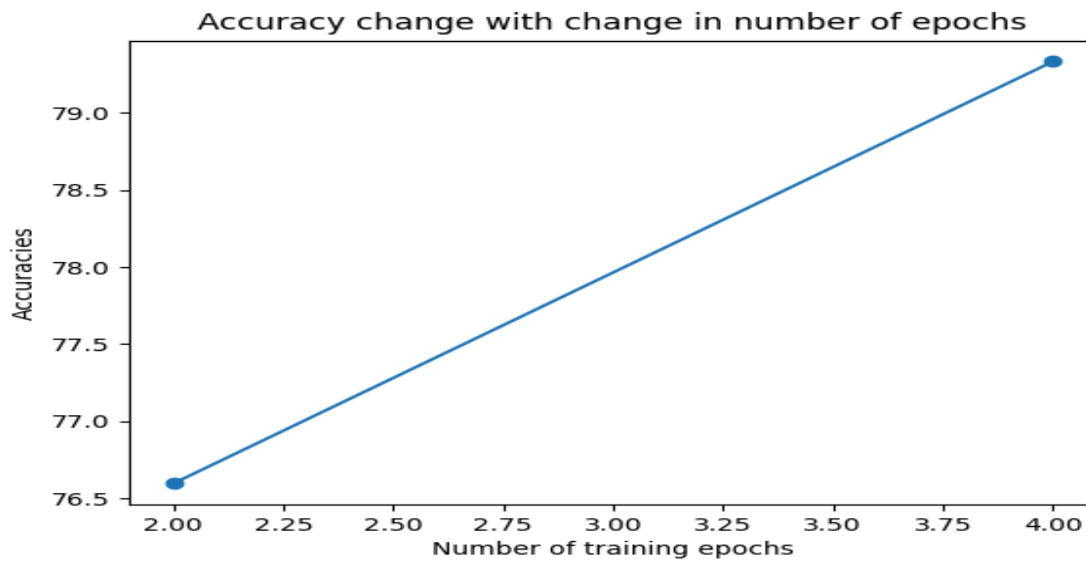
3. Hidden Dimension of LSTM - 50 , Word Embedding Size - 200 ,

Number of LSTM Layers - 1



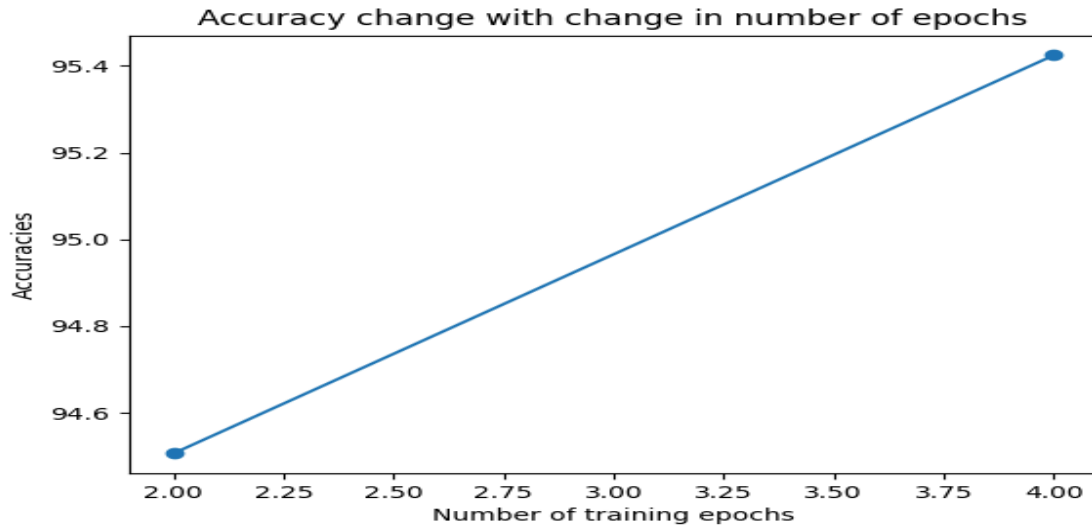
4. Hidden Dimension of LSTM - 50 , Word Embedding Size - 200 ,

Number of LSTM Layers - 2



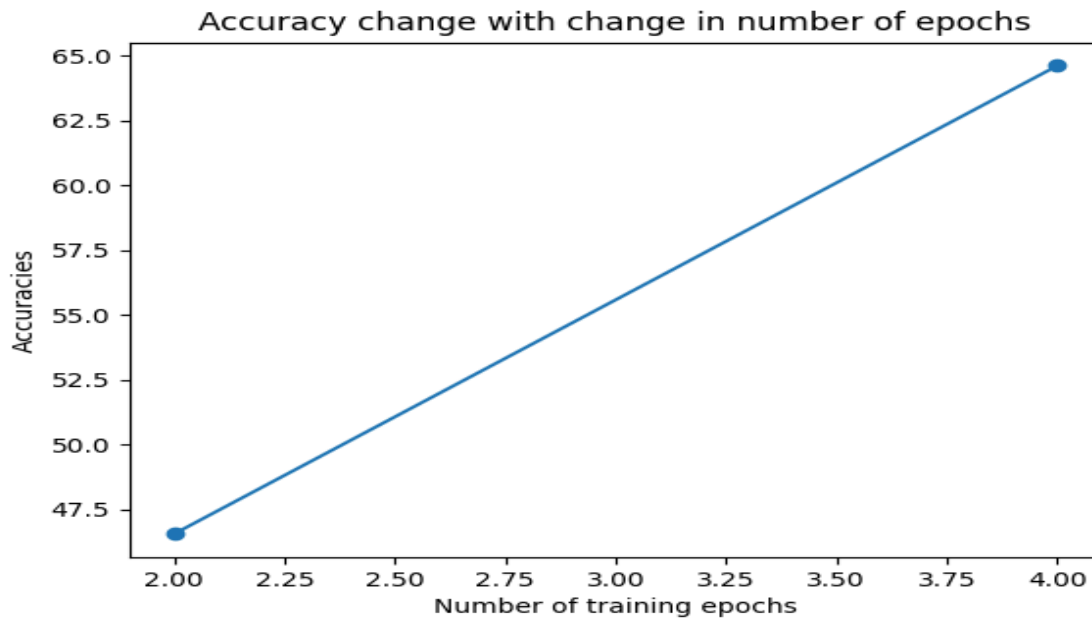
5. Hidden Dimension of LSTM - 100 , Word Embedding Size - 100 ,

Number of LSTM Layers - 1



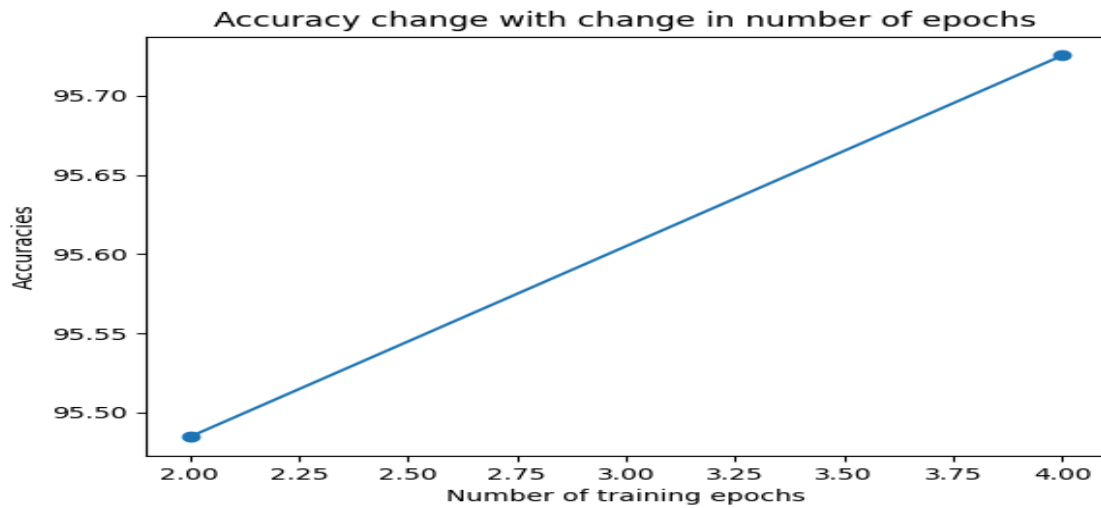
6. Hidden Dimension of LSTM - 100 , Word Embedding Size - 100 ,

Number of LSTM Layers - 2



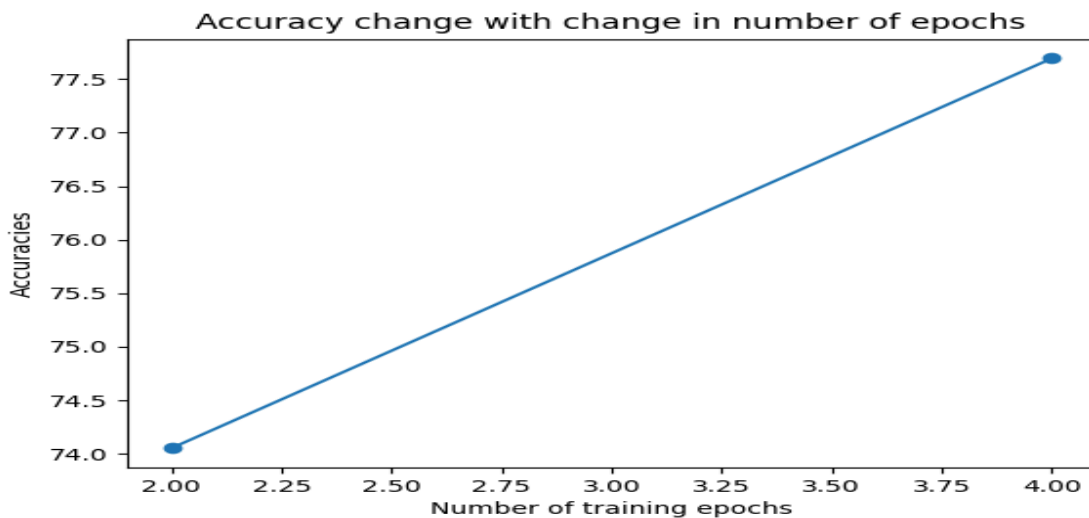
7. Hidden Dimension of LSTM - 100 , Word Embedding Size - 200 ,

Number of LSTM Layers - 1



8. Hidden Dimension of LSTM - 100 , Word Embedding Size - 200 ,

Number of LSTM Layers - 2



Here , we can clearly see that if epochs are increased , then accuracy increases because it trains well if the number of iterations increases.