NLP

Assignment5 Report

RNN

Prepared by

Name	ID
Ahmed Ali Ali	20200031
Ibrahim Hisham	20200005
Ahmed Ashraf	20200807

NLP
Assignment Report

Page titles used in this Assignment

“Machine learning”	“Deep learning”	“Natural Language Processing”	“Recurrent neural network”	“Convolutional neural network”
"Data science"	"Supervised learning"	"Reinforcement learning"	"Support vector machine"	"Neural network"
"Gradient boosting"	"Cluster analysis"	"Random forest"	"Dimensionality reduction"	"Transfer learning"

Each title has **10** pages to fetch, totaling **598254** char and **4860** words

- Sample of Text fetched after preprocessing :

'deep learn subset machin learn method base neural network represent learn the adject deep refer use multipl layer network method use either supervis semi supervis unsupervis deep learn architectur deep neural network deep belief network recurr neural network convolut neural network transform appli field includ comput vision speech recognit natur languag process machin translat bioinformat drug design medic imag analysi climat scienc materi inspect ....’

1- Model for predicting the next word :

Model inputs:

length of input sequence : 5
Step size for creating input sequences : 3

EX :

X_train[0] : ['deep', 'learn', 'subset', 'machin', 'learn’]
Y_train [0] : method

NLP
Assignment Report

2- Model for predicting the next char :

Model inputs:

length of input sequence : 5
Step size for creating input sequences : 3

EX :

X_train[0] : [‘d’, ‘e’, ‘e’, ‘p’, ‘ ’]
Y_train [0] : l

- Model Architecture

LSTM

```
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_4 (LSTM)               (None, 5, 256)            290816

 lstm_5 (LSTM)               (None, 5, 256)            525312

 lstm_6 (LSTM)               (None, 256)               525312

 dense_2 (Dense)             (None, 27)                6939

=================================================================
Total params: 1,348,379
Trainable params: 1,348,379
Non-trainable params: 0
```

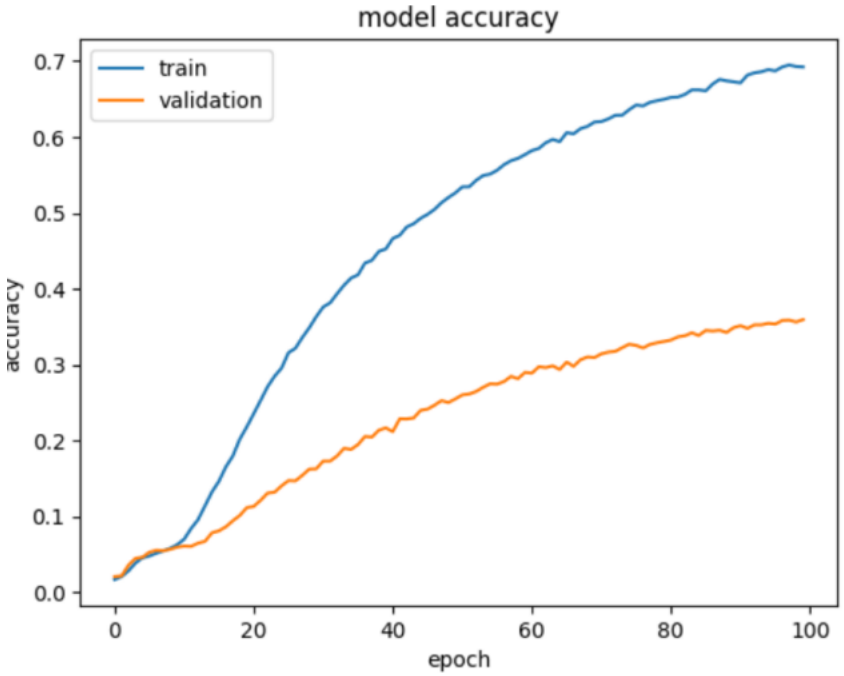
SimpleRNN

```
 Layer (type)                Output Shape              Param #
=================================================================
 simple_rnn_5 (SimpleRNN)    (None, 5, 256)            72704

 simple_rnn_6 (SimpleRNN)    (None, 5, 256)            131328

 simple_rnn_7 (SimpleRNN)    (None, 256)               131328

 dropout_5 (Dropout)         (None, 256)               0

 dense_4 (Dense)             (None, 27)                6939

=================================================================
Total params: 342,299
Trainable params: 342,299
Non-trainable params: 0
```

NLP
Assignment Report

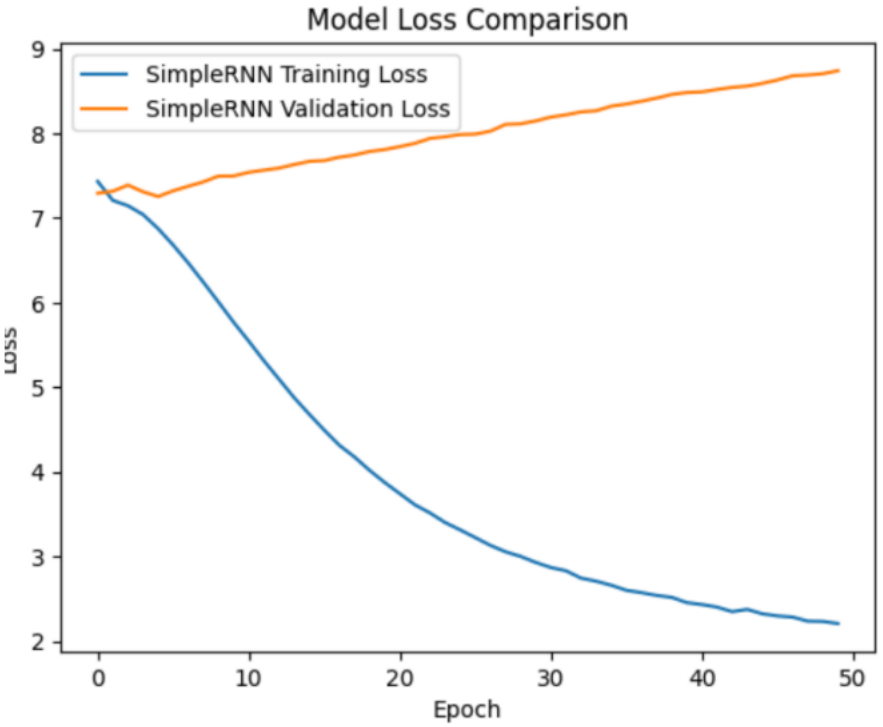
- Evaluation & Results :

1.1- Model LSTM for predicting the next word :



Train Accuracy : 70%

Test Accuracy : 36%

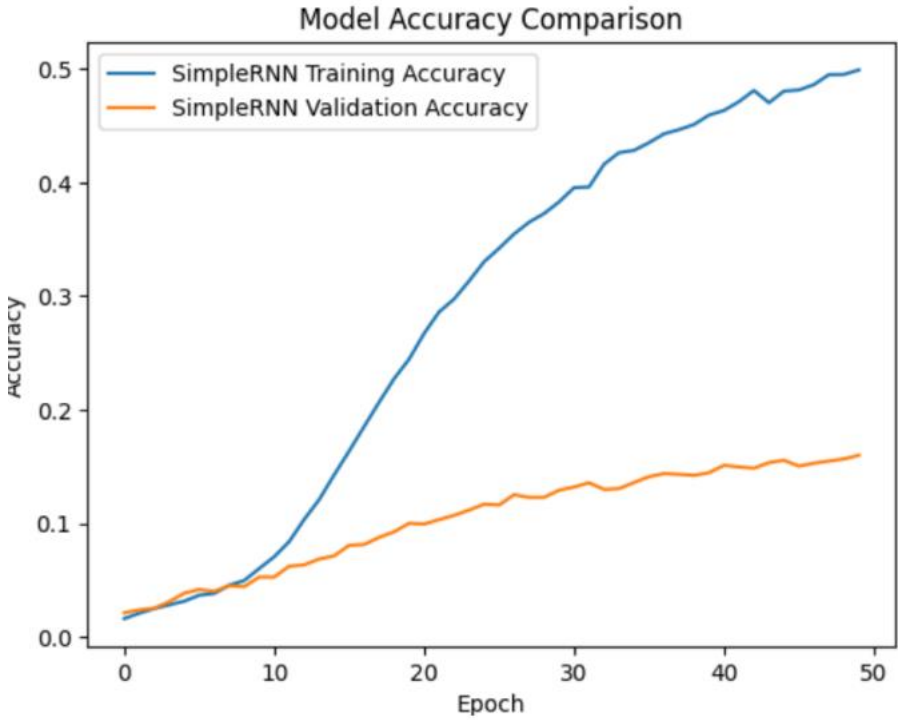


0.5158

NLP
Assignment Report

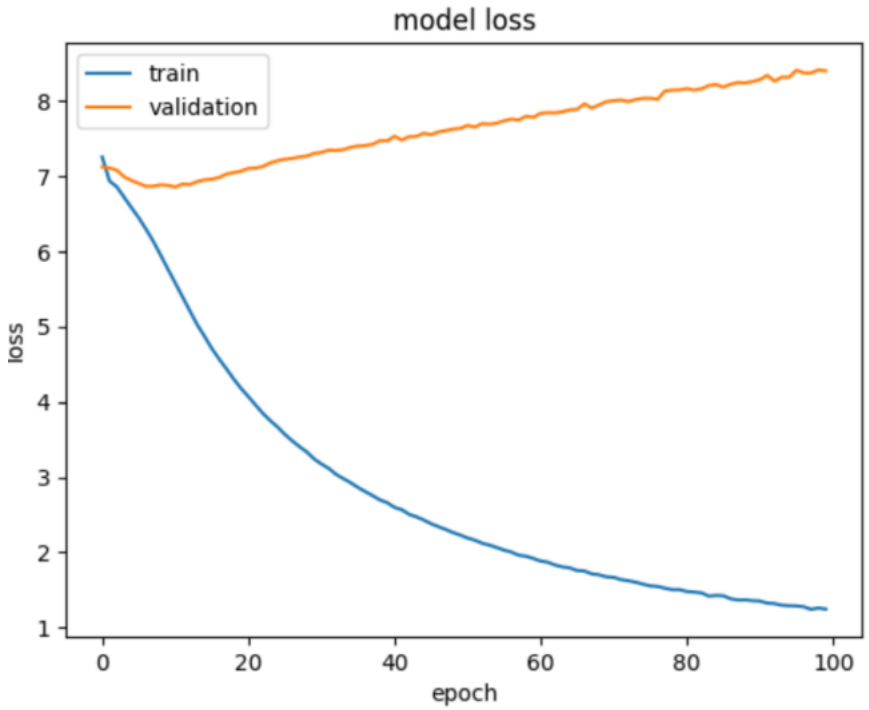
- Evaluation & Results :

1.2- Model SimpleRNN for predicting the next word :



Train Accuracy : 51 %

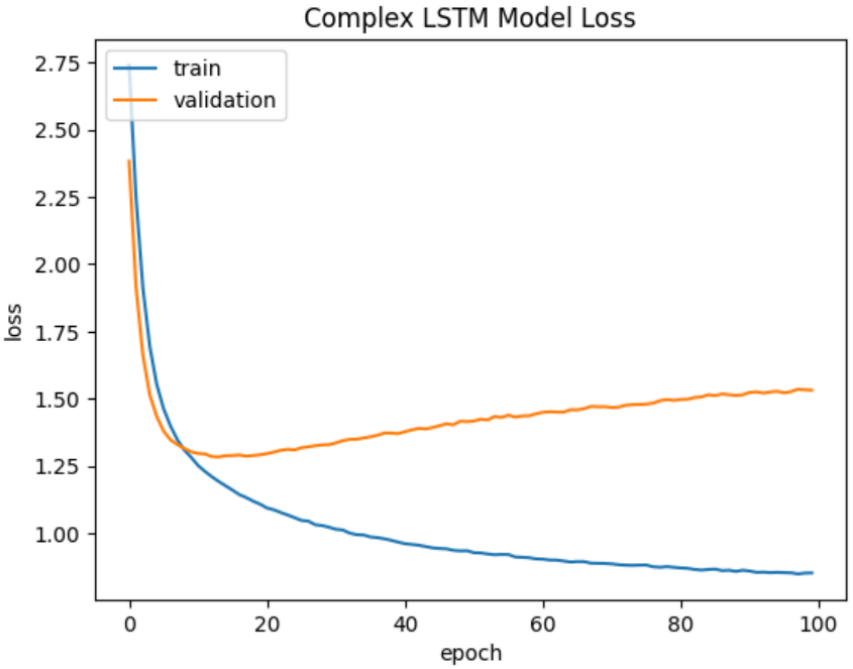
Test Accuracy : 16%



NLP
Assignment Report

- Evaluation & Results :

2.1- Model LSTM for predicting the next char :



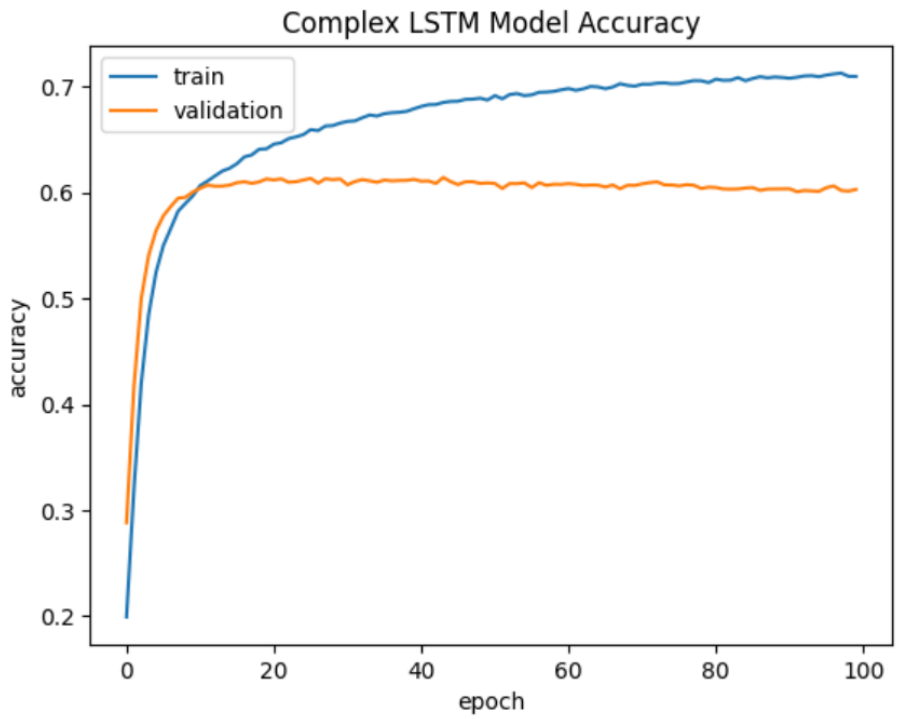
Train Accuracy :

71%

Test accuracy:

60%

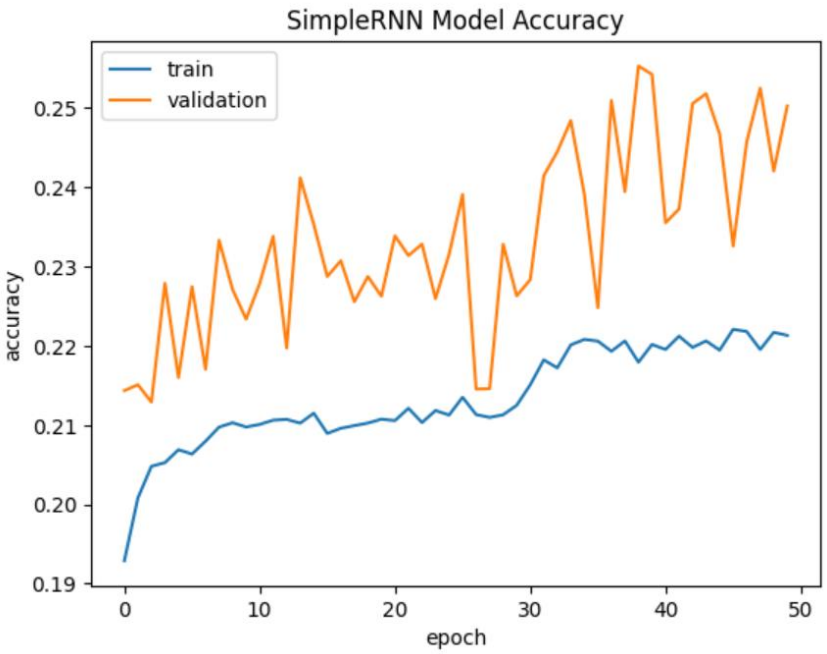
```
The next 10 characters after 'machin ' are predicted to be 'machin learn algo'
```



NLP
Assignment Report

- Evaluation & Results :

2.2- Model SimpleRNN for predicting the next char :



Train Accuracy :

57 %

Test :

accuracy: 0.59%

```
The predicted sequence after 'machin ' is 'machin tere sert '
```