1. Data Preprocessing:

Performed - lowercasing and sequencing of the data such that window of 40 and stride of 3 and performed one hot encoding.

Number of training Samples = 31213

Vocab Size 38

**Sample Sequence:**

Input : b'the sonnets\n\nby william shakespeare\n\nfro'

Target: b'he sonnets\n\nby william shakespeare\n\nfrom'

Length of above texts are 40 chars individually.

One Hot encoding of Input(

[[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]

...

[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]], shape=(40, 38), dtype=float32)

One Hot encoding of Target(

[[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]

[0. 0. 1. ... 0. 0. 0.]

...

[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]], shape=(40, 38), dtype=float32)

1. **Model Architecture:**

**Model:** "lstm\_model"

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

embedding (Embedding) multiple 4992

lstm (LSTM) multiple 394240

dense (Dense) multiple 10023

=================================================================

Total params: 409,255

Trainable params: 409,255

Non-trainable params: 0

Tried with embedding layer, one lstm layer and finally a fully connected output layer.

**Hyperparameters:**

Learning rate: 0.001 default

Embedding Dimension = 128

No of lstm units = 256

No of Lstm layer = 1

Batch\_Size = 32

Epochs = 1000

3)

**Initial Prediction before the model training**:

**Input**:

b'om nature hath not made for store,\nharsh'

**Next Char Predictions:**

b"vrxl:jytl(dtgnu'('xg(ib\ncn'.t.o,-\neoxmxa"

4)

**After model training and predicting 400 characters**:

**Input Text**:

inspired by you

**Output Text:**

inspired by youth,

and sweets grown common lose their dear victories once foiled,

is from the book of honour razed quite,

and all the rest forgot for which he toiled:

then happy i that love and am beloved

where i may not remove nor be removed.

lord of my love, to whom in vew shore away,

death's second self that seals up all in rest.

o surn dear head, else true, love twinness shaken

as i by yo

**Observation:**

As we can see the model can predict the next character and completes the words and sentences perfectly, eventhough its not a continuation of the seed word and the words generated from the model is present in the original learning data, by this we can see that model learned the pattern and word probability distribution.

**Model loss and Perplexity:**

5/5 - 0s - loss: 7.4239 - 92ms/epoch - 18ms/step

Test Perplexity: 1675.514115646483

**Challenges:**

Training with high batch size results in crashing in-between execution. Introducing an additional layer increases the training time a lot.

After 513 Epochs the model reached is saturation and the loss value doesn’t decrease much further. Due to this the model overfits the learning data and performs poorly in hold out data, but it learned the character distribution very well.