Introduction

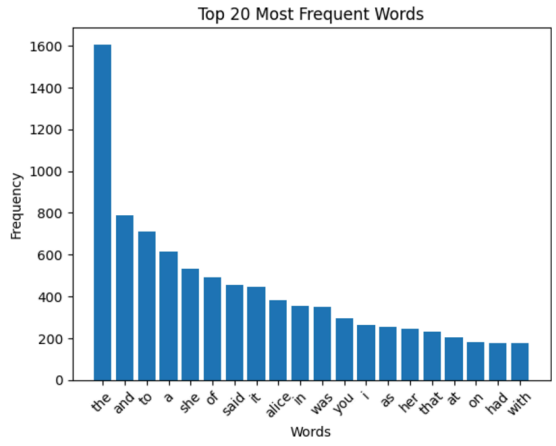
This assignment explores sequential models, specifically LSTMs, GRUs, and SimpleRNNs in my case. The goal is to generate new lines of text based on patterns learned from a chosen book, The Three Little Kittens by an anonymous author. Additionally, to build it from scratch using Keras/tensorflow without using any pre-trained models or transfer learning. The book was sourced from Project Gutenberg as a raw text file. The text will be preprocessed and converted to tokens to be inputted in each sequential model I mentioned. It will then be trained for a reasonable number of epochs and compared between the three types of models and hyperparameters for the most accurate. Finally, the best model used to generate new text sequences that reflect the style and content of the inputted book.

Analysis

The raw data was first loaded anc cleaned using a series of preprocessing steps Unwanted characters such as “--” were replaced with spaces, and punctuation was removed to ensure a clean tokenization process. The text was then split into individual tokens(words), and any non-alphabetic tokens were filtered out, leaving only meaningful words for analysis. To ensure consistency, all tokens were converted to lowercase. The additional text in the text file that was not a part of the book was removed. The total number of tokens in the dataset was calculated, and the number of unique tokens was determined by converting the list of tokens into a set. The results are given in the following table:

Total Tokens	Unique Tokens	Total Sequences
651	250	600

To prepare the data for sequential mode, sequences of length 51 were generated using a sliding window approach, where each sequence consists of 50 input tokens and the following token as the target. This generated a large number of sequences, forming the basis for training the model. The vocabulary size was found by fitting a tokenizer on the text, which mapped each unique token to a unique index. Additionally, a histogram of the 20 most frequent words in the dataset was found to give deeper understanding to potential outputs of the generated sequences from the model.



Methods

The first model I built was the LSTM model. I started when a blank sequential model. I added a LSTM layer with 128 nodes, input shape of the sequences, l2 kernel regularizer with a strength value of 0.01, and tanh activation function. An additional regularization of 0.3 dropout followed. Next, another LSTM layer with 64 nodes that didn't return the sequences. Finally a fully connected dense layer with softmax activation. The model was compiled with categorical cross entropy loss function and Adam optimizer with a learning rate of 0.001 using accuracy as the metric. The summary of the model is:

Layer (type)	Output Shape	Param #
lstm_27 (LSTM)	(None, 50, 128)	66,560
dropout_16 (Dropout)	(None, 50, 128)	0
lstm_28 (LSTM)	(None, 64)	49,408
dense_16 (Dense)	(None, 2418)	157,170

Total params: 273,138 (1.04 MB)
Trainable params: 273,138 (1.04 MB)
Non-trainable params: 0 (0.00 B)

The model was fit on the dataset with a validation split of 0.2 and batch size of 128 over 20 epochs.

The next model I built was a basic RNN model. I started with a blank sequential model. I created an embedding layer with 100 dimensions per word. Next, I added a simpleRNN layer with 128 nodes and an elastic net kernel regularizer with 0.01 strength values for both l1 and l2. Another simpleRNN layer with 128 nodes. Then a 0.2 dropout regularization. 2 more simpleRNN layers with 64 nodes. All simpleRNN layers had tanh activation functions. A final fully connected dense layer with softmax activation was added. The model was compiled using categorical cross entropy loss function and adam optimizer with 0.001 learning rate using accuracy as the metric. The summary of the model is:

Layer (type)	Output Shape	Param #
embedding_12 (Embedding)	(None, 50, 100)	241,800
simple_rnn_16 (SimpleRNN)	(None, 50, 128)	29,312
simple_rnn_17 (SimpleRNN)	(None, 50, 128)	32,896
dropout_18 (Dropout)	(None, 50, 128)	0
simple_rnn_18 (SimpleRNN)	(None, 50, 64)	12,352
simple_rnn_19 (SimpleRNN)	(None, 64)	8,256
dense_18 (Dense)	(None, 2418)	157,170

A learning rate schedule that monitored the validation accuracy for plateau and would reduce the learning rate by a factor of 0.2. Additionally an early stopping was added that monitored validation accuracy that restored the best weights, and had a patience of 15 epochs. The model was fit with a validation split of 0.2 and batch size of 128 over 50 epochs(stopped at 19).

Results

The LSTM model obtained a final peak validation accuracy of 10.89% and training accuracy of 8.14%. The simpleRNN model performed worse with a final validation accuracy of 9.48% and training accuracy of 7.6%. The outputted sequences for the LSTM model was:

```
1. Seed Text: it very nice it had in fact a sort of mixed flavour of cherrytart custard pineapple roast turkey toffee and hot buttered toast she very soon finished it off a curious said alice must be shutting up like a and so it was indeed she was now only ten inches high
Generated Text: to the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the and and and the the the the

2. Seed Text: the mock turtle replied counting off the subjects on his flappers ancient and modern with seaography then drawlingmaster was an old congereel that used to come once a week he taught us drawling stretching and fainting in was that said alice i show it you the mock tu
Generated Text: the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the and and the and the the the the the

3. Seed Text: began picking them up again as quickly as she could for the accident of the goldfish kept running in her head and she had a vague sort of idea that they must be collected at once and put back into the jurybox or they would die trial cannot said the king
Generated Text: the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the and and the and the the the the the

4. Seed Text: might find another key on it or at any rate a book of rules for shutting people up like telescopes this time she found a little bottle on it certainly was not here said alice and round the neck of the bottle was a paper label with the words beautifully printed
Generated Text: to the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the and and and the the the the

5. Seed Text: alice and all her wonderful adventures till she too began dreaming after a fashion and this was her first she dreamed of little alice herself and once again the tiny hands were clasped upon her knee and the bright eager eyes were looking up into could hear the very
Generated Text: the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the and and the the the the and and the

6. Seed Text: tells us a afraid i know said alice rather alarmed at the proposal the dormouse they both cried up and they pinched it on both sides at once the dormouse slowly opened his eyes he said in a hoarse feeble voice heard every word you fellows were us a said the
Generated Text: the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the and the the the the the the the the

7. Seed Text: you do i to go bill the master says to go down the so got to come down the chimney has said alice to herself they seem to put everything upon bill i be in place for a good deal this fireplace is narrow to be sure but i think i
Generated Text: and the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the and and the the and the the the the

8. Seed Text: taller and the other side will make you grow side of what the other side of thought alice to herself the said the caterpillar just as if she had asked it aloud and in another moment it was out of sight alice remained looking thoughtfully at the mushroom for a minute
Generated Text: the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the and and the the the the the the

9. Seed Text: cried alice she was so much surprised that for the moment she quite forgot how to speak good english opening out like the largest telescope that ever was goodbye for when she looked down at her feet they seemed to be almost out of sight they were getting so far off
Generated Text: the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the and and the the the the the the

10. Seed Text: outoftheway things to happen that it seemed quite dull and stupid for life to go on in the common way so she set to work and very soon finished off the cake chapter ii the pool of tears and cried alice she was so much surprised that for the moment she
Generated Text: the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the the and and and and the and the the the the the
```

The outputted sequences for the simpleRNN model was:

```
1. Seed Text: so awfully have answered three questions and that is said his father give yourself airs do you think i can listen all day to such stuff be off or kick you down is not said said the caterpillar quite right said alice timidly of the words have got is wrong from
Generated Text: and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and

2. Seed Text: silent the king laid his hand upon her arm and timidly said my dear she is only a the queen turned angrily away from him and said to the knave them the knave did so very carefully with one foot said the queen in a shrill loud voice and the three
Generated Text: and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and

3. Seed Text: flapper across his eyes he looked at alice and tried to speak but for a minute or two sobs choked his voice as if he had a bone in his said the gryphon and it set to work shaking him and punching him in the back at last the mock turtle
Generated Text: and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and

4. Seed Text: the way of expecting nothing but outoftheway things to happen that it seemed quite dull and stupid for life to go on in the common way so she set to work and very soon finished off the cake chapter ii the pool of tears and cried alice she was so much
Generated Text: and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and

5. Seed Text: had been for some time busily writing in his notebook cackled out and read out from his book fortytwo all persons more than a mile high to leave the everybody looked at alice not a mile said alice said the king two miles added the queen i go at any said
Generated Text: and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and

6. Seed Text: of them with her arms folded frowning like a thunderstorm fine day your the duchess began in a low weak voice i give you fair shouted the queen stamping on the ground as she spoke you or your head must be off and that in about half no time take your
Generated Text: and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and

7. Seed Text: in a large ring with the mouse in the middle alice kept her eyes anxiously fixed on it for she felt sure she would catch a bad cold if she did not get dry very soon said the mouse with an important air you all ready this is the driest thing
Generated Text: and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and

8. Seed Text: she had quite forgotten the duchess by this time and was a little startled when she heard her voice close to her ear thinking about something my dear and that makes you forget to talk i tell you just now what the moral of that is but i shall remember it
Generated Text: and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and

9. Seed Text: see some meaning in them after all i could not you swim can he added turning to the knave the knave shook his head sadly i look like he said which he certainly did not being made entirely of cardboard right so said the king and he went on muttering over
Generated Text: and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and

10. Seed Text: to open them again and all would change to dull grass would be only rustling in the wind and the pool rippling to the waving of the rattling teacups would change to tinkling sheepbells and the shrill cries to the voice of the shepherd the sneeze of the baby the sh
Generated Text: and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and and
```

Reflection

I learned a lot from this assignment, especially about the hundreds of ways to optimize a model. I spent over 40 hours trying hundreds of different combinations of hyperparameters, models, layers, regularization techniques, tokenizations, loss functions, activation functions, batch sizes, epochs, callbacks, and books. So I was really able to learn how everything works with each other and what scenarios to use which combination and hyperparameter values. In the end, I gave up trying to get higher validation accuracy and just decided to let it ruin my grade, as I now have to turn it in 5 days late with no results to show for it. However, I gained a lot of knowledge from it which is what really matters.