

TEXT GENERATOR

To create a text generator, the following have been implemented in this notebook:

The text file location : `"/content/drive/MyDrive/61262-0.txt"`

Preprocessing:

1. After import, the text is preprocessed using the following methods:
 - a. Any special characters except period and apostrophe have been removed
 - b. The text is converted to lower case
 - c. The text is then tokenized to sentences using `sent_tokenize`
 - d. The sequences have been padded to make sure all inputs are the same size

Currently implemented Methodolgy:

1. Embedding layer has been used to convert words into vectors
2. 3 layers of LSTM (Long-Short Term Memory), a type of RNN has been used and it captures long-term memory
3. A layer of Normalization has also been added to prevent overfitting
4. Early stopping has been used to ensure that the model stops if it senses overfitting. i.e., if the validation loss starts increasing for more than 3 continuous epochs (with patience = 3)

The case of the repeating words:

Ideas to fix it:

1. Train for more epochs
2. Adjust temperature to increase diversity in predicted text
3. Try adding more LSTM layers and dropout layers
4. Try Bidirectional LSTM - It predicts based on both future and past data
5. Reducing batch size - will increase training time but can provide better results
6. Increasing the training data by adding more Agatha Christie books
7. Trying Transfer learning from a pretrained text generator model

Method to execute has to be added to each task, as well

ChatGPT vs Model Result

ChatGPT hey chat can you predict the outcome for this like Agatha Christie. What is your text generation for this: (50 words) There was a murder in the building

There was a murder in the building. The body lay motionless in the dimly lit hallway, a crimson pool spreading across the floor. The residents whispered in hushed tones, eyes darting with suspicion. Inspector Graves arrived, his gaze sharp. "Locked doors. No forced entry. The killer is among us."

Comparison with ChatGPT:

hey chat can you predict the outcome for this like Agatha Christie.
What is your text generation for this: (50 words)
There was a murder in the building

There was a murder in the building. The body lay motionless in the dimly lit hallway, a crimson pool spreading across the floor. The residents whispered in hushed tones, eyes darting with suspicion. Inspector Graves arrived, his gaze sharp. "Locked doors. No forced entry. The killer is among us."



Model Result

Define a seed text

```
text_prompt = "There was a murder in the building"
```

There was a murder in the building point about your not too that s was temporary sold heeds of poirot i packet our look was drawer i or the which fianc influence concerning street the my table house match number from of hastings out very arm do to that seen the sad but fibre