

Task 3

NLP Text Generation Report

Data Cleaning and Pre-processing:

After extraction of the zipped file in the Google Collab, data pre-processing took place by changing the dashes (-) to spaces (" "), creating tokens after splitting the spaces and removing the punctuations by creating a translation table along with removing all the non-alphabetic characters and converting all the words into lower case. After this, a sequence of sentences was created from the tokens and were saved in a file.

After this, unique words were encoded in categories and sequences were made as numeric array. Now, the data was split into X and Y from which Y was converted to categorical which gave us the Y as encoded value from vocab size.

Model Training and Methodology:

An embedding layer was added to take the whole vocab as input and produce sequence length and total words, that is 50, as output. Furthermore, a LSTM layer was added to return sequences that were found to be true to add another LSTM layer. Another LSTM layer was added along with two hidden layers. Now, the model was compiled successfully using categorical cross entropy loss and Adam optimiser. Now, fitting the model with 128 batch size and epochs 200 was done.

After the model training, the model was saved and tokenizer was used to convert it into a pkl file for text generation.

To generate the result, list of results was created, encoding the words as seed text sequences and performing padding to the rest of the words along with predicting the NumPy argument for the input. The category was converted back to word from the output sequences and appending of the text that is generated by forming the result set (50 words) was done.

Input Text: I do not know whether Papa guessed my feelings on the subject, probably not, and in any case, he would not have been interested. The opinion of other people never interested him in the slightest degree. I think it was really a sign of his greatness. In the same way, he lived quite detached from the necessities of daily life. He ate what was put before him in an exemplary fashion, but seemed mildly pained when the question of paying

Output Text: the smaller well-dets go the barrel the experts a improbable was funny nonsense and he snapped not shut up a butterknife they found it over and checking asir pulled himself in the floor of an statement to the farthest stairwell the theory came in milewide swaths he was brought the