Assignment 3 [6 Marks + 2 Bonus]
Amazon product reviews using RNN and Word Embeddings

Dataset: Amazon reviews dataset; This dataset was created from the scraped reviews from products in Amazon, it contains 17000+ records attached with the assignment "amazon_reviews.csv". The classes are three in number namely; Negative Reviews, Neutral Reviews, Positive Reviews.

Requirements:

Data Pre-processing (if needed): to clean your data and provide a valid dataset for the models to be trained, like removing stopwords using NLTK.

Data Splitting: apply data splitting for your; 80% as training set and 20% as validation set.

Word Embedding: build your vocabulary by extracting and indexing unique words, convert each review to a sequence of indices, then apply sequence padding to have all sequences of the same length in preparation for input to the embedding layer.

Model Training: You will train two models simpleRNN and LSTM and print the accuracy for each model on testing data.

Bonus:

1. Allow the user to input a new review and predict the result.
2. Provide a report that shows model summary of each model and the best hyperparameters for each model (splitting ratio, sequence padding length …) with a table showing the accuracy against each parameter (i.e. 80% 20% ratio, 70% 30% ratio, and same for sequence padding length).

Submission Rules:

1. Max number of members per team is 5 from any group.
2. Deadline is 17 May 2024 @ 11:59 PM.
3. Submission will be on google classroom.
4. No late submission is allowed.
5. No submission through e-mails.
6. Cheating cases; you will get a negative grade whether you give the code to someone or take the code from someone/internet.

7. You have to write clean code and follow a good coding style including choosing meaningful variable names, Also your code must be modular (All processes are encapsulated into methods & you may use classes If you like) ? Code Modularity included in the grading criteria.