Name: Nishant Dhotre
Roll No: 23CS60R48

1] Word2vec

**Hyperparameters:**

- **Batch Size: 32 (used in train_model function during batching)**
- **Learning Rate: 0.001 (specified in train_model function)**
- **Number of Epochs: 20 for training the model (train_model function call uses epochs=20)**
- **Max Sequence Length: Not explicitly defined in this code, but it influences the embedding averaging in average_embeddings. The actual sequence length handled would depend on the Word2Vec model's behavior and the vocabulary processing.**
- **Vocabulary Size: Determined dynamically based on the CountVectorizer settings (min_df=5, max_df=0.8) in the create_vocabulary function. The exact size depends on the dataset.**

2]RNN

**Hyperparameters:**

- **Batch Size:** 32, used for training and evaluating the models, influencing how many samples are processed before the model is updated.
- **Learning Rate:** 0.001, determining the step size at each iteration while moving toward a minimum of a loss function.
- **Number of Epochs:** 10, indicating how many times the entire dataset is passed forward and backward through the neural network.
- **Max Sequence Length:** 100, setting a limit to the length of the tokenized inputs. This is crucial for processing text data where input sizes can vary.
- **Vocabulary Size:** Dynamically determined based on the **CountVectorizer** settings in **create_vocabulary**, with **min_df=5** and **max_df=0.8**. The vocabulary size directly affects the size of the embedding layer.
- **Embedding Dimension:** 300, corresponding to the size of the Word2Vec embeddings used.
- **Hidden Dimension:** 256, specifying the size of the RNN/LSTM's hidden layers.
- **Output Dimension:** 4, representing the number of target classes for the model to predict.
- **Bidirectionality:** Enabled, indicating the RNN/LSTM processes the input data in both forward and backward directions, potentially capturing more contextual information.

Name: Nishant Dhotre
Roll No: 23CS60R48

3] LSTM

**Hyperparameters:**

- **Batch Size:** The batch size, which determines the number of samples to work through before updating the internal model parameters, is not explicitly mentioned in the **train_model** function call. However, it's determined by how **train_data_loader** is initialized elsewhere in the code. Common practice would suggest values like 32, 64, or 128.
- **Learning Rate:** 0.001, a critical hyperparameter that influences the speed and quality of the training process by determining the step size at each iteration when minimizing the loss function.
- **Number of Epochs:** 10, specifying how many complete passes the training dataset will go through.
- **Max Sequence Length:** Determined by the **max_len** parameter in the **DataFrameDataset** class and set to 100. This affects how text sequences are truncated or padded.
- **Vocabulary Size:** Dynamically calculated as **len(word_index) + 1**, accommodating for zero padding. The actual size depends on the processed dataset and the **CountVectorizer** settings.
- **Embedding Dimension:** 300, chosen to match the dimensions of the Word2Vec embeddings.
- **Hidden Dimension:** 256, indicating the size of the LSTM's hidden layers.
- **Output Dimension:** 4, corresponding to the number of classes in the classification task.
- **Bidirectionality:** Enabled (**True**), suggesting the LSTM processes the input data in both forward and backward directions to capture better contextual relationships.