Assignment 7

Problem Statement

Study Tag Generation model and apply the same process to generate tags of a movie through plot synopsis.

Data:

https://www.kaggle.com/datasets/cryptexcode/mpst-movie-plot-synopses-with-tags

Task:

1. Read the mpst_full_data.csv, and select title, plot_synopsis, and tags columns for further processing.
2. Preprocess the title, remove stopwords, hyperlinks, punctuations and other unwanted elements.
3. Convert tags string column into list column.
4. Apply multilabel binarizer on the tags.
5. Tokenize, and apply pad sequence on the title with maxlen as the longest title.
6. Create a sequence model using embedding with dimensions 50.
    a. Add LSTM, dropout and Batch normalization layers.
    b. Add Bidirectional LSTMs.
    c. Add Conv1D Layer.
7. Display a few title samples with actual and predicted tags.