Module 8: Recurrent Neural Networks - Assignment

Problem Statement:

Build a sentiment analysis model using Recurrent Neural Networks (RNNs) to classify movie reviews from the IMDB dataset into positive or negative sentiments.

Dataset:

The dataset comprises 25,000 movie reviews from IMDB, labeled by sentiment (positive/negative). Reviews have been preprocessed, and each review is encoded as a sequence of word indices (integers). The words in the dataset are indexed by overall frequency in the dataset, allowing for quick filtering operations such as: "only consider the top 10,000 most common words, but eliminate the top 20 most common words".

Tasks to be Performed:

Data Preprocessing:

- Load the IMDB dataset, keeping only the top 10,000 most frequently occurring words.
- Pad the sequences so that they all have the same length.

Model Building:

- Create a Sequential RNN model using TensorFlow and Keras.
- The model should consist of an Embedding layer, a SimpleRNN layer, and a Dense output layer.
- Compile the model, specifying the appropriate optimizer, loss function, and metrics.

Training:

- Train the model on the preprocessed movie reviews, using a batch size of 128 and validating on 20% of the training data.
- Run the training for 10 epochs.

Evaluation:

Evaluate the model on the test set and report the accuracy.

Expected Outcome:

A trained RNN model that can classify movie reviews into positive or negative sentiments, with an accuracy metric provided at the end of the training process.