# Assignment No:6

Practical Exercise: Sentiment Analysis using LSTM Network or GRU

## 1) Problem Statement

Implement sentiment analysis using Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU) for text classification. The goal is to predict the sentiment of movie reviews (positive or negative) using the IMDB dataset.

## 2) Libraries Used

Python:  
1. **Pandas**: For loading and manipulating the IMDB dataset.  
2. **NumPy**: For mathematical operations and array handling.  
3. **NLTK (Natural Language Toolkit)**: For handling stop words in text preprocessing.  
4. **TensorFlow/Keras**: For building and training the deep learning model (LSTM/GRU).  
5. **Scikit-learn**: For splitting the dataset into training and test sets.

## 3) Theory

Sentiment analysis is a natural language processing task that involves classifying text into positive or negative sentiment. Recurrent Neural Networks (RNNs) are a class of neural networks specifically designed to handle sequential data like text. LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Units) are advanced RNN architectures that can capture long-term dependencies in the input sequence, making them effective for sentiment analysis tasks.  
  
In this assignment, an LSTM/GRU-based network will be used to classify movie reviews as positive or negative based on their textual content.

## 4) Methods

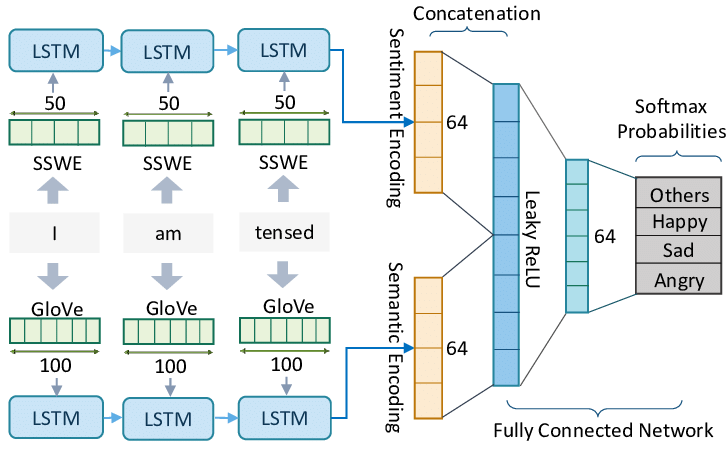
1. **Data Preprocessing**:  
 - The IMDB dataset is loaded, and the reviews are cleaned by removing HTML tags, non-alphabetic characters, and stop words.  
 - The sentiment labels (positive/negative) are encoded as binary values (1 for positive, 0 for negative).  
   
2. **Model Architecture**:  
 - The model is built using the Keras Sequential API with the following layers:  
 - An Embedding layer to convert words into dense vector representations.  
 - An LSTM (or GRU) layer to capture sequential dependencies in the text.  
 - A Dense layer with a sigmoid activation function for binary classification.  
   
3. **Training**:  
 - The model is trained using the Adam optimizer and binary crossentropy loss function.  
 - The dataset is split into training and test sets, and the model is trained on the training set and evaluated on the test set.  
   
4. **Evaluation**:  
 - The performance of the model is evaluated based on accuracy, and sample predictions are generated for unseen reviews.

## 5) Advantages and Disadvantages

- **Advantages**:  
 - LSTM and GRU networks are effective at capturing long-term dependencies in sequential data, making them well-suited for sentiment analysis tasks.  
 - TensorFlow/Keras provides easy-to-use APIs for building, training, and evaluating LSTM/GRU models.  
   
- **Disadvantages**:  
 - Training LSTM/GRU models can be time-consuming, especially for large datasets.  
 - LSTM/GRU models require careful tuning of hyperparameters (e.g., number of units, learning rate) for optimal performance.

## 6) Diagram

## Building a Twitter Sentiment Analysis System with Recurrent Neural Networks



## 7) Conclusion

LSTM and GRU models are powerful tools for sentiment analysis, as they can effectively capture the dependencies in sequential data like movie reviews. By leveraging these architectures, the model is able to predict the sentiment of reviews with high accuracy. These models are widely used in various natural language processing (NLP) tasks, including text classification, translation, and summarization.