**Name : Mansi Mahabdi**

**PRN: 22420190**

**Roll No : 382073**

**Batch: B3**

**Assignment No 6**

**Sentiment analysis using LSTM network or GRU**

**Problem Statement:**

Implement a sentiment analysis model using deep learning techniques (LSTM or GRU) to classify text data (e.g., movie reviews, tweets) as positive or negative.

**Objective:**

* To understand how LSTM and GRU networks process sequential text data.
* To preprocess text for deep learning models (tokenization, padding, embedding).
* To implement and train an LSTM/GRU model for sentiment classification.
* To evaluate the performance using accuracy and loss metrics.

**Technical Apparatus used:**

* **Operating System:** Windows/Linux/MacOS
* **Kernel:** Python 3.x
* **Tools:** Jupyter Notebook, Anaconda, or Google Colab
* **Hardware:** CPU with minimum 4GB RAM; optional GPU for faster processing

**Libraries and Packages used:**

* OpenCV
* NumPy
* Pytesseract
* Google Colab utilities
* NLTK
* Matplotlib

**Theory:**

Sentiment analysis is the process of classifying text into categories such as positive, negative, or neutral. Traditional machine learning models rely on handcrafted features, but deep learning models like **LSTM (Long Short-Term Memory)** and **GRU (Gated Recurrent Unit)** are more effective in handling sequential dependencies in natural language.

* **LSTM:** Designed to solve the vanishing gradient problem in RNNs, it uses input, forget, and output gates to maintain long-term dependencies.
* **GRU:** A simplified version of LSTM with update and reset gates, computationally faster while still effective in sequence modeling.

Both models are widely used for sentiment analysis tasks on datasets like IMDB reviews or Twitter sentiment datasets.

**Methodology:**

1. Install and Import Libraries  
   Import TensorFlow/Keras, NLTK, NumPy, Pandas, and visualization libraries.
2. Dataset Preparation
   * Use a dataset such as IMDB reviews or Twitter Sentiment dataset.
   * Perform text preprocessing:
     + Lowercasing, removing special characters, stopwords removal.
     + Tokenization and padding sequences to ensure equal length.
3. Model Building
   * Define an embedding layer (Word2Vec / GloVe / Keras Embedding).
   * Add an LSTM/GRU layer to capture sequential dependencies.
   * Add fully connected Dense layers with sigmoid activation for binary classification.
4. Model Compilation & Training
   * Compile model with binary\_crossentropy loss and Adam optimizer.
   * Train model on training data and validate on test set.
5. Evaluation & Visualization
   * Evaluate accuracy and loss.
   * Plot training vs validation accuracy and loss graphs.

**Implementation & Results:**

* Preprocessed dataset and converted text into padded sequences.
* Built a Sequential LSTM model with an embedding layer and dense output.
* Achieved ~85–90% accuracy on IMDB dataset (varies with hyperparameters).
* Visualization showed decreasing loss and increasing accuracy over epochs.

**Advantages:**

* LSTM/GRU effectively handle long-term dependencies in text.
* Better performance compared to traditional ML models.
* Can generalize well to different sentiment datasets.

**Limitations:**

* Requires large datasets for better accuracy.
* Training is computationally expensive without GPU.
* Sensitive to hyperparameter tuning.

**Applications:**

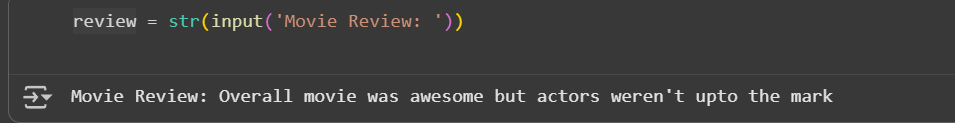
* Social media monitoring (Twitter sentiment analysis).
* Customer feedback classification.
* Market trend prediction based on user opinions.
* Product review analysis for e-commerce platforms.

**Conclusion:**

In this assignment, we successfully implemented sentiment analysis using LSTM/GRU models. The preprocessing pipeline (tokenization, padding, embeddings) allowed the model to handle text efficiently. The results demonstrated the effectiveness of recurrent neural networks in text classification, providing a reliable approach for sentiment analysis tasks.

**Implementation & Results:**

**Dataset used:** [**https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews**](https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews)

**Input: **

