Assignment 6

CS 6379, NLP Lab

1. Write a Python program to implement a **text classification model using RNN and LSTM** on the Twitter Sentiment140 dataset.

    - Use pre-trained **Word2Vec embeddings**.
    - Train both RNN and LSTM models.
    - Compare their accuracy and training time.

2. Write a Python program using a pre-trained **Transformer model** (like BART or T5) from Hugging Face to perform **text summarization**.

    - Accept a long article as input.
    - Generate a concise summary.
    - Evaluate using the **different evaluation matrix scores**.

1. Which model best captures **long-term dependencies** in a sequence?
2. What is the primary disadvantage of using Bag-of-Words?
3. Name the two-word embedding techniques.
4. What is the major innovation in **Transformer architecture** over RNNs?
5. What does the **tokenizer** in NLP do?
6. Write down different prepossessing steps (at least 3) in NLP.
7. Write the formula for TF-IDF.
8. Why do we remove stopwords in text processing?
9. Name two types of Word2Vec architectures.
10. Which Transformer-based model is designed for text generation?