Project Title: Paraphrase Detection Using NLP

Abstract:

Paraphrase detection is an essential task in Natural Language Processing (NLP) that determines whether two sentences have same meaning. This project aims to develop a paraphrase detection model using NLP, deep learning techniques. The system takes two input sentences and classifies them as paraphrase or not by analysing their semantic similarity and Machine Learning, Deep Learning techniques for classification. This paraphrase detection has applications in plagiarism detection, chatbots, text summarization, content validation and intelligent search engines.

Team:

1. **Name:** Goli Srikar
   Reg No: 22BCE9946
2. **Name:** Raya Veera Venkata Naga Rahul Reddy
   Reg No: 22BCE8855

Introduction:

With the rapid growth of AI Automation, digital communication, the need for understanding sentence similarity by an AI system for automating the tasks has increased. In this project, we will preprocess text data, extract meaningful representations, and train a deep learning model to identify paraphrased sentences. By leveraging Sentence-BERT (SBERT) embeddings and cosine similarity, we aim to improve classification accuracy.

Libraries & Technologies Used:

- Python (Programming Language, Pandas, NumPy)
- NLTK, spacy, gensim (Preprocessing , word embeddings , Named Entity Recognition (NER) and Cosine Similarity etc.)
- Scikit learn, TensorFlow, keras (for DL models)
- Sentence-Transformers (SBERT) (Deep Learning-based NLP for embeddings)

Expected Input:

- **Two sentences for comparison**
- **Example:** sentence1: AI is transforming the world,
  sentence2: Artificial intelligence is changing the world.

Expected Output:

- **Classification result, Similarity score(optional):** "Paraphrase" or "Not Paraphrase", Similarity Score (0-1)
- **Example Output:**
  paraphrase: true,
  similarity_score: 0.87