NLP ASSIGNMENT

Q1:- Take a custom paragraph, perform the entire pipeline and Print results at each step.

Tokenization → Stopword Removal → Stemming → Lemmatization.

SOL:-

1) Import Libraries

```
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer, WordNetLemmatizer
```

✓ 5.3s

2) Download required resources

```
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('omw-1.4')
```

✓ 19.3s

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\ESHA\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping tokenizers\punkt.zip.
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\ESHA\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\stopwords.zip.
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\ESHA\AppData\Roaming\nltk_data...
[nltk_data] Downloading package omw-1.4 to
[nltk_data]     C:\Users\ESHA\AppData\Roaming\nltk_data...

True
```

3) Custom Paragraph

```
paragraph = """Natural Language Processing is an exciting field of Artificial Intelligence.
It allows computers to understand, interpret, and generate human language effectively."""
```

✓ 0.0s

```
print("Original Paragraph:")
print(paragraph)
print("--------------------------------------------------")
```

✓ 0.0s

```
Original Paragraph:
Natural Language Processing is an exciting field of Artificial Intelligence.
It allows computers to understand, interpret, and generate human language effectively.
--------------------------------------------------
```

4) Tokenization

```
# Ensure punkt_tab is downloaded
import nltk
nltk.download('punkt_tab')


tokens = word_tokenize(paragraph)
print("Step 1 - Tokenization:")
print(tokens)
print("--------------------------------------------------")
```

✓ 1.5s Python

```
[nltk_data] Downloading package punkt_tab to
[nltk_data]     C:\Users\ESHA\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping tokenizers\punkt_tab.zip.
Step 1 - Tokenization:
['Natural', 'Language', 'Processing', 'is', 'an', 'exciting', 'field', 'of', 'Artificial', 'Intelligence', '.', 'It', 'allows', 'computers', 'to', 'understand', ',', 'interpret', ',',
--------------------------------------------------
```

5) StopWord Removal

```
stop_words = set(stopwords.words('english'))
filtered_tokens = [word for word in tokens if word.lower() not in stop_words and word.isalpha()]
print("Step 2 - Stopword Removal:")
print(filtered_tokens)
print("--------------------------------------------------")
```

✓ 0.0s Python

```
Step 2 - Stopword Removal:
['Natural', 'Language', 'Processing', 'exciting', 'field', 'Artificial', 'Intelligence', 'allows', 'computers', 'understand', 'interpret', 'generate', 'human', 'language', 'effectively
--------------------------------------------------
```

6) Stemming

```
stemmer = PorterStemmer()
stemmed_words = [stemmer.stem(word) for word in filtered_tokens]
print("Step 3 - Stemming:")
print(stemmed_words)
print("--------------------------------------------------")
```

✓ 0.0s

```
Step 3 - Stemming:
['natur', 'languag', 'process', 'excit', 'field', 'artifici', 'intellig', 'allow', 'comput', 'understand', 'interpret', 'gener', 'human', 'languag', 'effect']
--------------------------------------------------
```

7) Lemmatization

Q2:- Define NLP and its real time application in a specific domain base.

SOL:- Natural Language Processing (NLP) is a subfield of Artificial Intelligence (AI) and Linguistics that focuses on enabling computers to understand, interpret, and generate human languages in a meaningful way. It combines computational linguistics, machine learning, and deep learning to process text and speech data.

Domain: Healthcare

In the healthcare domain, NLP plays a **critical role** in analyzing unstructured medical data and improving patient care.

Applications:

1. **Clinical Document Processing** – NLP extracts key information (like symptoms, diagnosis, medications) from doctors' notes, prescriptions, and electronic health records (EHR).
2. **Medical Chatbots & Virtual Assistants** – Used for answering patient queries, scheduling appointments, or providing preliminary health advice.
3. **Disease Prediction & Risk Analysis** – NLP Analyzes patient history and lab reports to detect early signs of diseases like diabetes, cancer, or heart conditions.
4. **Voice Recognition in Healthcare** – Doctors can dictate notes, and NLP systems automatically convert them into structured medical records.

Example: IBM Watson Health uses NLP to read and interpret clinical research papers and patient data, helping doctors make faster and more accurate treatment decisions.

Q3:- What is NLU and NLG?

SOL:- NLU (Natural Language Understanding)

- Definition:
  NLU is a subfield of NLP that focuses on enabling machines to understand the meaning, intent, and context behind human language.
- What it does:
  - Extracts entities (names, dates, locations).
  - Identifies intent (what the user wants).
  - Handles ambiguity, synonyms, and context.
- Example:
  - If a user types: *"Book me a flight to Delhi tomorrow morning"* → NLU extracts:
    - Intent: Book flight
    - Entities: Destination = Delhi, Date = Tomorrow, Time = Morning

NLG (Natural Language Generation)

- **Definition:**
  NLG is the opposite of NLU. It focuses on enabling machines to **generate human-like text or speech** from structured data or internal representations.
- **What it does:**
  - Converts data into meaningful sentences.
  - Creates reports, summaries, or conversational replies.
- **Example:**
  - Input data: {Destination: Delhi, Date: Tomorrow, Time: Morning}
  - NLG generates: *"Your flight to Delhi is booked for tomorrow morning."*