Name : Devkumar Biswas

Class: BE(AI&DS)

Div: B

Subject : IR(CL-II Lab)

Roll no. : BEAD21267

**Assignment No. - 1**

**Problem Statement** : Write a program for pre-processing of a text document such as stop word removal, stemming.

Code:-

pip install nltk

import nltk

from nltk.corpus import stopwords

from nltk.tokenize import word\_tokenize

from nltk.stem import PorterStemmer

# Download required NLTK resources

nltk.download('punkt')

nltk.download('stopwords')

# Example text

text = "Natural Language Processing is an exciting field of artificial intelligence."

# Tokenize the text into words

tokens = word\_tokenize(text)

print("Tokens:", tokens)

# Define English stopwords

stop\_words = set(stopwords.words('english'))

# Remove stopwords from tokens

filtered\_tokens = [word for word in tokens if word.lower() not in stop\_words]

print("Filtered Tokens (no stop words):", filtered\_tokens)

# Initialize the Porter Stemmer

ps = PorterStemmer()

# Apply stemming to each word

stemmed\_tokens = [ps.stem(word) for word in filtered\_tokens]

print("Stemmed Tokens:", stemmed\_tokens)

OUTPUT: -



