1.Write a Python program for the following preprocessing of text in NLP:

* Tokenization
* Filtration
* Script Validation
* Stop Word Removal
* Stemming

import nltk

import re

from nltk.tokenize import word\_tokenize

from nltk.corpus import stopwords

from nltk.stem import PorterStemmer

# Download necessary datasets (Run this once)

#nltk.download('punkt\_tab')

#nltk.download('punkt')

#nltk.download('stopwords')

def preprocess\_text(text):

# Tokenization

tokens = word\_tokenize(text)

print("\nTokens:", tokens)

# Filtration: Removing special characters, digits, and extra spaces

tokens = [re.sub(r'[^A-Za-z]', '', token) for token in tokens]

tokens = [token for token in tokens if token] # Remove empty strings

print("\nFiltered Tokens:", tokens)

# Script Validation: Removing non-English words

tokens = [token for token in tokens if token.isalpha()]

print("\nScript Validated Tokens:", tokens)

# Stop Word Removal

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

tokens = [token.lower() for token in tokens if token.lower() not in stop\_words]

print("\nTokens after Stop Word Removal:", tokens)

# Stemming

stemmer = PorterStemmer()

tokens = [stemmer.stem(token) for token in tokens]

print("\nStemmed Tokens:", tokens)

return tokens

# Example text

text = "Natural Language Processing (NLP) is a fascinating field of Artificial Intelligence! NLP helps computers understand human language."

processed\_tokens = preprocess\_text(text)

print("Final Processed Tokens:", processed\_tokens)