import nltk

from nltk.tokenize import TreebankWordTokenizer

from nltk.corpus import stopwords

from nltk.stem import PorterStemmer

# Force download necessary resources

nltk.download('stopwords', force=True)

def nlp\_preprocessing\_pipeline(sentence):

tokenizer = TreebankWordTokenizer()

# Step 1: Tokenization (no 'punkt' needed)

tokens = tokenizer.tokenize(sentence)

print("Original Tokens:")

print(tokens)

# Step 2: Stopword Removal

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

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

print("\nTokens Without Stopwords:")

print(filtered\_tokens)

# Step 3: Stemming

stemmer = PorterStemmer()

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

print("\nStemmed Words:")

print(stemmed\_words)

# Example sentence

sentence = "NLP techniques are used in virtual assistants like Alexa and Siri."

nlp\_preprocessing\_pipeline(sentence)