Preprocess text\_stemming(1)

import nltk

from nltk.corpus import stopwords

from nltk.tokenize import word\_tokenize

from nltk.stem import PorterStemmer

from nltk.stem import WordNetLemmatizer

# Download required NLTK data (you only need to do this once)

nltk.download('punkt')

nltk.download('stopwords')

nltk.download('wordnet')

# Simple text preprocessing function

def preprocess(text):

# Tokenize text (split into words)

words = word\_tokenize(text.lower()) # Convert text to lowercase

# Remove stop words

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

words = [word for word in words if word.isalnum() and word not in stop\_words]

# Stemming

ps = PorterStemmer()

stemmed\_words = [ps.stem(word) for word in words]

# Lemmatization

lemmatizer = WordNetLemmatizer()

lemmatized\_words = [lemmatizer.lemmatize(word) for word in words]

return stemmed\_words, lemmatized\_words

# Example text

text = "This is a simple example for text preprocessing, with stop word removal and stemming!"

stemmed, lemmatized = preprocess(text)

# Display results

print("Stemmed Text: ", stemmed)

print("Lemmatized Text: ", lemmatized)