**SHARON P. A**

PRACTICAL TECHNICAL ASSESMENT

Activity 1

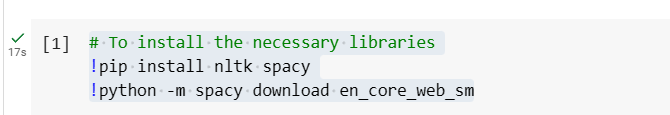
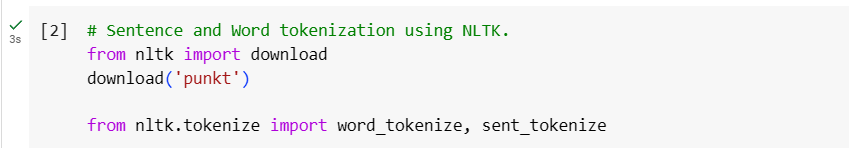
**NLP Preprocessing Techniques**

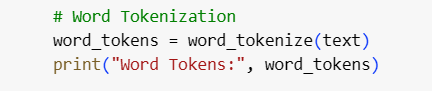
comprehensive guide to NLP preprocessing techniques using NLTK and Spacy, including tokenization, stemming, and lemmatization.

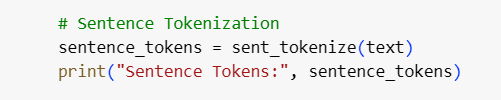
Requirements

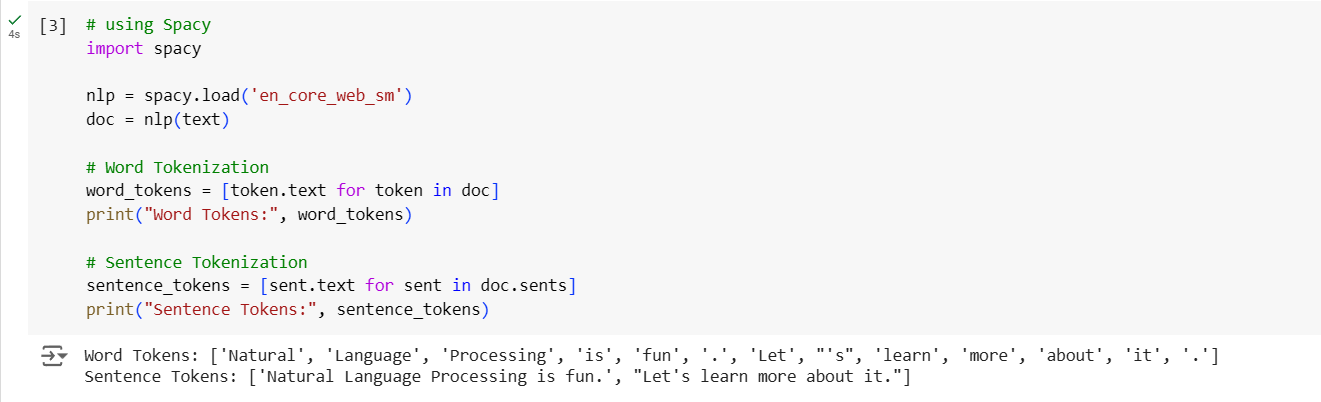
* Personal computer/laptop
* Google Collab

Procedure

1. Install Necessary Libraries
2. Import Libraries and Download NLTK Resources
3. Define the Text
4. Tokenization Using NLTK





1. Tokenization Using Spacy
2. Stemming Using NLTK
3. Lemmatization Using NLTK

Activity 2

**Python implementation of BoW**

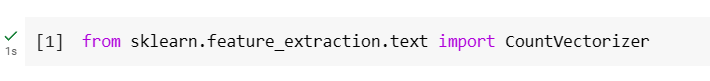
Here's a step-by-step procedure documentation for converting a collection of text documents into a Bag of Words (BoW) representation using CountVectorizer from scikit-learn

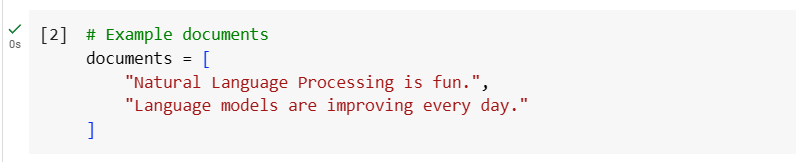
Requirements

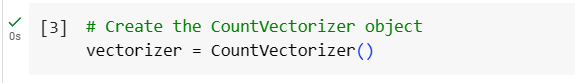
* Personal computer/laptop
* Google Collab

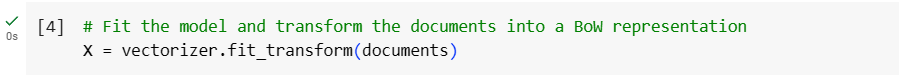
Procedure

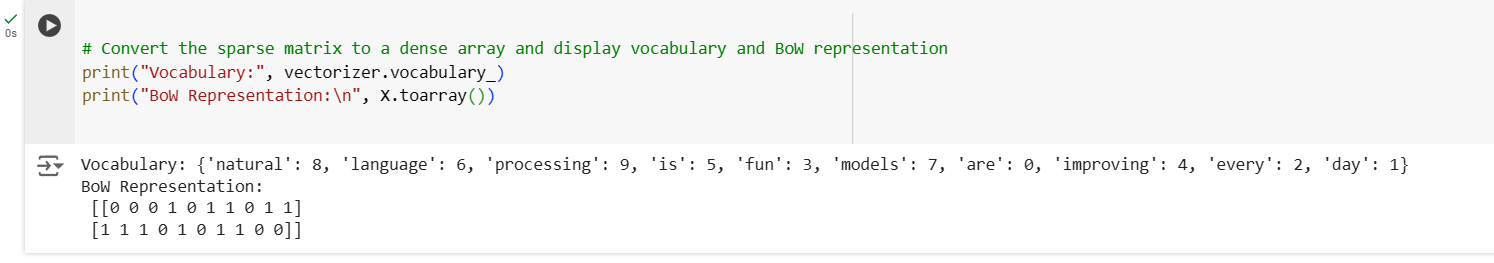
1. Import the CountVectorizer class from the sklearn.feature\_extraction.text module.



1. Prepare a list of example documents. Each document is a string of text.
2. Create an instance of CountVectorizer. This object will be used to convert the text documents into a matrix of token counts.



1. Fit the model and transform the documents into a BoW representation
2. Convert the resulting sparse matrix into a dense array for easier inspection.

Print the vocabulary, which is a dictionary mapping terms to their indices in the matrix, and print the BoW representation, which shows the token counts for each document.

Activity 3

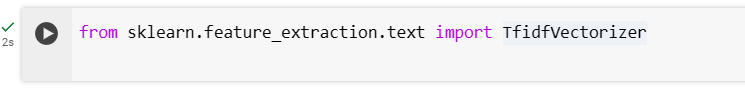
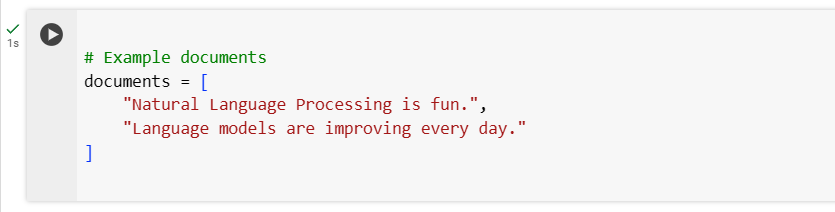
**Python implementation of BoW**

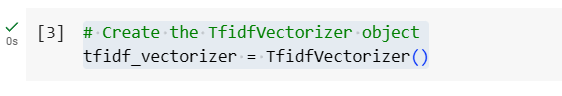
Here's a step-by-step procedure documentation for converting a collection of text documents into a Bag of Words (BoW) representation using CountVectorizer from scikit-learn

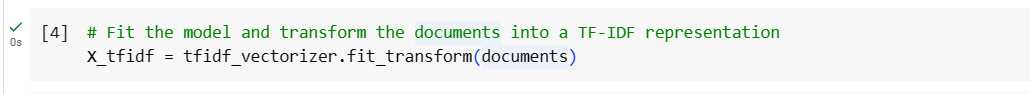
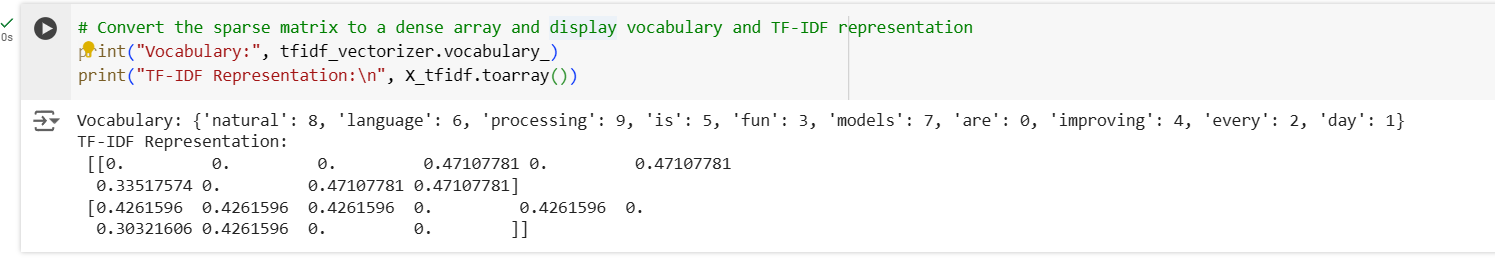
Requirements

* Personal computer/laptop
* Google Collab

Procedure

1. Start by importing the TfidfVectorizer from the sklearn.feature\_extraction.text module.
2. Create a list of example documents to be transformed using TF-IDF.
3. Initialize the TfidfVectorizer object.



1. Fit the TfidfVectorizer to the documents and transform them into a TF-IDF representation.
2. Display the Vocabulary and TF-IDF Representation

This shows the mapping of each word to a unique index in the TF-IDF matrix.