Name: Akshat Bhargava

Rollno: TEAD-22563

DS (SL-III)

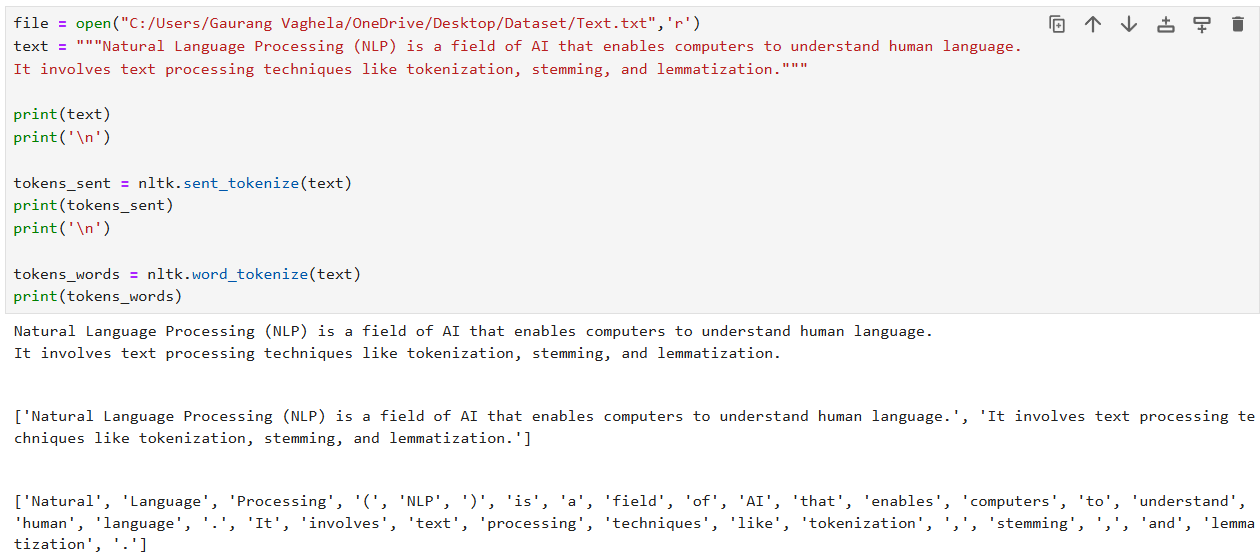
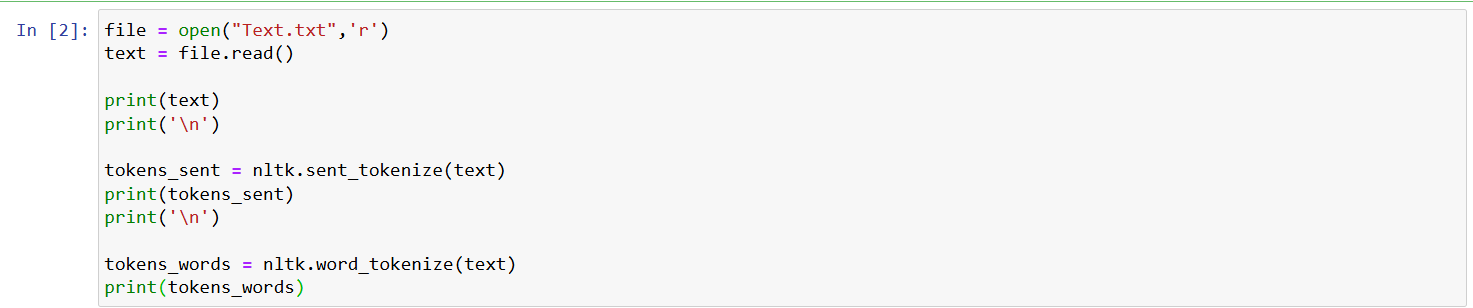
Practical 7

Code:

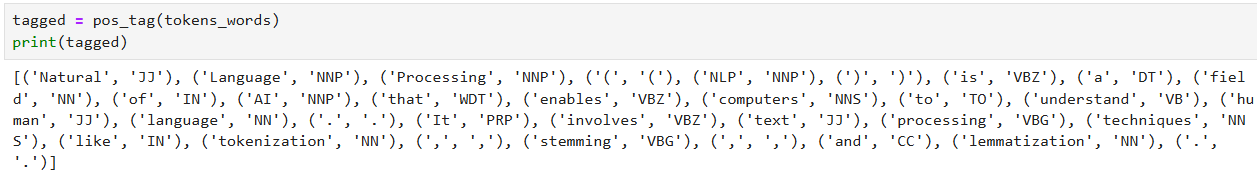
1. Extract Sample document and apply following document preprocessing methods: Tokenization, POS Tagging, stop words removal, Stemming and Lemmatization.



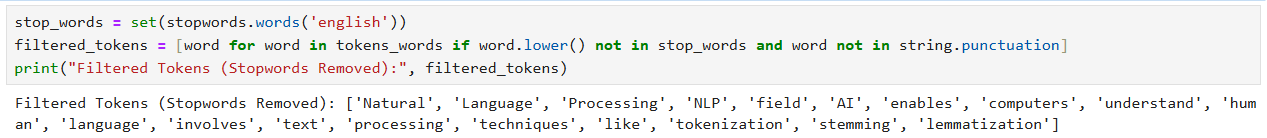
Tokenization:



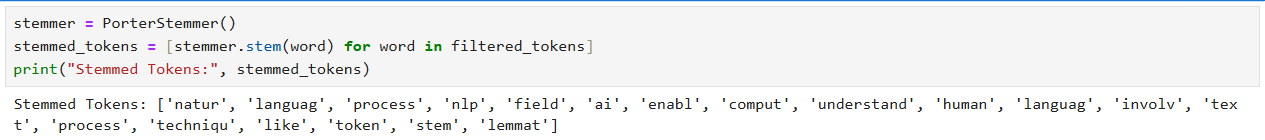
POS Tagging:



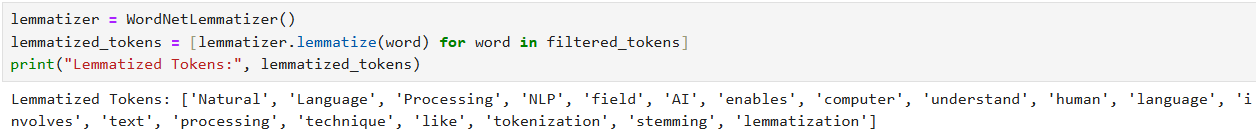
Stop Words Removing:



Stemming:



Lemmatization:



2. Create representation of documents by calculating Term Frequency and Inverse Document Frequency.

