

```
pip install nltk
```

```
import nltk  
from nltk.tokenize import sent_tokenize  
from nltk.tokenize import word_tokenize  
from nltk.corpus import stopwords  
from nltk.stem import PorterStemmer  
from nltk.stem import WordNetLemmatizer  
import re  
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
nltk.download('punkt')  
nltk.download('stopwords')  
nltk.download('wordnet')  
nltk.download('averaged_perceptron_tagger')
```

text="Tokenization replaces a sensitive data element, for example, a bank account number, with a non-sensitive substitute, known as a token. The token is a randomized data string that has no essential or exploitable value or meaning. It is a unique identifier which retains all the pertinent information about the data without compromising its security"

```
sentences = sent_tokenize(text)  
print(sentences)
```

```
words = word_tokenize(text)  
print(words)
```

```
stop_words = set(stopwords.words('english'))  
print(stop_words)
```

```
text="How to remove stop words with NLTK library in Python7"  
text= re.sub('[^a-zA-Z]', ' ', text)  
tokens= word_tokenize(text.lower())  
filtered_text= []
```

```

for w in tokens:
    if w not in stop_words:
        filtered_text.append(w)
print("Tokenized Sentence:", tokens)
print("Filtered Sentence:", filtered_text)

```

```

ps = PorterStemmer()
stemmed_words = ['wait', 'waiting', 'waited', 'waits']
for w in stemmed_words:
    print(w, ":", ps.stem(w))
print("Original Sentence:", stemmed_words)

```

```

lemmatizer = WordNetLemmatizer()
text= "studies studying cries cry"
tokenization= nltk.word_tokenize(text)
for w in tokenization:
    print("Lemma for {} is {}".format(w, lemmatizer.lemmatize(w)))

```

```

data = 'The pink sweater fits her perfectly'
words = nltk.word_tokenize(data)
for word in words:
    print(nltk.pos_tag([word]))

```

paragraph= """India is a vast country with second highest population in the world. It is a country with diverse cultures, traditions and beliefs. People in India celebrate unity in diversity. Festivals like Diwali, Holi, Navratri, Ramzan, Christmas etc. are celebrated by people across India and create a sense of brotherhood and cultural unity. Each festival has its religious and cultural importance."""

```
wn = WordNetLemmatizer()
sentences = nltk.sent_tokenize(paragraph)
corpus = []
for i in range(len(sentences)):
    review = re.sub('[^a-zA-Z]', ' ', sentences[i])
    review = review.lower()
    review = review.split()
    review = [wn.lemmatize(word) for word in review if not word in set(stopwords.words('english'))]
    review = ' '.join(review)
    corpus.append(review)
print(corpus)

tfidf = TfidfVectorizer()
X = tfidf.fit_transform(corpus).toarray()
print(X)
```