In [1]:
```
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer, WordNetLemmatizer
from nltk import pos_tag
from sklearn.feature_extraction.text import TfidfVectorizer
```

In [2]:
```
# Tokenization
document = "Data science is a field of study that uses modern tools and techniques
tokens = word_tokenize(document.lower())
print(tokens)
```

```
['data', 'science', 'is', 'a', 'field', 'of', 'study', 'that', 'uses', 'modern',
'tools', 'and', 'techniques', 'such', 'as', 'machine', 'learning', 'algorithms',
'to', 'unify', 'statistics', ',', 'data', 'analysis', ',', 'informatics', 'and',
'their', 'related', 'methods', 'in', 'order', 'to', 'understand', 'and', 'analyz
e', 'actual', 'phenomena', 'with', 'data', '.']
```

In [3]:
```
# POS Tagging
tagged_tokens = pos_tag(tokens)
print(tagged_tokens)
```

```
[('data', 'NNS'), ('science', 'NN'), ('is', 'VBZ'), ('a', 'DT'), ('field', 'NN'),
('of', 'IN'), ('study', 'NN'), ('that', 'WDT'), ('uses', 'VBZ'), ('modern', 'JJ'),
('tools', 'NNS'), ('and', 'CC'), ('techniques', 'NNS'), ('such', 'JJ'), ('as', 'I
N'), ('machine', 'NN'), ('learning', 'VBG'), ('algorithms', 'NNS'), ('to', 'TO'),
('unify', 'VB'), ('statistics', 'NNS'), (',', ','), ('data', 'NNS'), ('analysis',
'NN'), (',', ','), ('informatics', 'NNS'), ('and', 'CC'), ('their', 'PRP$'), ('rel
ated', 'JJ'), ('methods', 'NNS'), ('in', 'IN'), ('order', 'NN'), ('to', 'TO'), ('u
nderstand', 'VB'), ('and', 'CC'), ('analyze', 'VB'), ('actual', 'JJ'), ('phenomen
a', 'NN'), ('with', 'IN'), ('data', 'NNS'), ('.', '.')]
```

In [4]:
```
# Stop Words Removal
stop_words = set(stopwords.words('english'))
filtered_tokens = [token for token in tokens if token not in stop_words]
print(filtered_tokens)
```

```
['data', 'science', 'field', 'study', 'uses', 'modern', 'tools', 'techniques', 'ma
chine', 'learning', 'algorithms', 'unify', 'statistics', ',', 'data', 'analysis',
',', 'informatics', 'related', 'methods', 'order', 'understand', 'analyze', 'actua
l', 'phenomena', 'data', '.']
```

In [5]:
```
# Stemming
stemmer = PorterStemmer()
stemmed_tokens = [stemmer.stem(token) for token in filtered_tokens]
print(stemmed_tokens)
```

```
['data', 'scienc', 'field', 'studi', 'use', 'modern', 'tool', 'techniqu', 'machi
n', 'learn', 'algorithm', 'unifi', 'statist', ',', 'data', 'analysi', ',', 'inform
at', 'relat', 'method', 'order', 'understand', 'analyz', 'actual', 'phenomena', 'd
ata', '.']
```

In [6]:
```
# Lemmatization
lemmatizer = WordNetLemmatizer()
lemmatized_tokens = [lemmatizer.lemmatize(token) for token in filtered_tokens]
print(lemmatized_tokens)
```

```
['data', 'science', 'field', 'study', 'us', 'modern', 'tool', 'technique', 'machin
e', 'learning', 'algorithm', 'unify', 'statistic', ',', 'data', 'analysis', ',',
'informatics', 'related', 'method', 'order', 'understand', 'analyze', 'actual', 'p
henomenon', 'data', '.']
```

In [7]:
```
# TF-IDF
tfidf = TfidfVectorizer()
tfidf_matrix = tfidf.fit_transform([document])
print(tfidf_matrix)
```

```
  (0, 31)       0.14002800840280097
  (0, 17)       0.14002800840280097
  (0, 0)        0.14002800840280097
  (0, 3)        0.14002800840280097
  (0, 28)       0.14002800840280097
  (0, 16)       0.14002800840280097
  (0, 8)        0.14002800840280097
  (0, 13)       0.14002800840280097
  (0, 18)       0.14002800840280097
  (0, 25)       0.14002800840280097
  (0, 9)        0.14002800840280097
  (0, 2)        0.14002800840280097
  (0, 20)       0.14002800840280097
  (0, 29)       0.14002800840280097
  (0, 26)       0.28005601680560194
  (0, 1)        0.14002800840280097
  (0, 11)       0.14002800840280097
  (0, 12)       0.14002800840280097
  (0, 5)        0.14002800840280097
  (0, 22)       0.14002800840280097
  (0, 23)       0.14002800840280097
  (0, 4)        0.42008402520840293
  (0, 27)       0.14002800840280097
  (0, 14)       0.14002800840280097
  (0, 30)       0.14002800840280097
  (0, 24)       0.14002800840280097
  (0, 21)       0.14002800840280097
  (0, 15)       0.14002800840280097
  (0, 7)        0.14002800840280097
  (0, 10)       0.14002800840280097
  (0, 19)       0.14002800840280097
  (0, 6)        0.42008402520840293
```