In [2]:
```
!pip install nltk
```

```
Collecting nltk
  Downloading nltk-3.9.2-py3-none-any.whl.metadata (3.2 kB)
Requirement already satisfied: click in c:\users\rahul\miniconda3\lib\site-pack
ages (from nltk) (8.3.0)
Requirement already satisfied: joblib in c:\users\rahul\miniconda3\lib\site-pac
kages (from nltk) (1.4.2)
Collecting regex>=2021.8.3 (from nltk)
  Downloading regex-2025.10.23-cp312-cp312-win_amd64.whl.metadata (41 kB)
     ---------------------------------------- 0.0/41.5 kB ? eta -:--:--
     ---------------------------------------- 0.0/41.5 kB ? eta -:--:--
     --------- ------------------------------ 10.2/41.5 kB ? eta -:--:--
     ------------------ ------------------- 20.5/41.5 kB 162.5 kB/s eta 0:00:01
     -------------------------------------  41.0/41.5 kB 279.3 kB/s eta 0:00:01
     -------------------------------------- 41.5/41.5 kB 250.3 kB/s eta 0:00:00
Requirement already satisfied: tqdm in c:\users\rahul\miniconda3\lib\site-packa
ges (from nltk) (4.66.4)
Requirement already satisfied: colorama in c:\users\rahul\miniconda3\lib\site-p
ackages (from click->nltk) (0.4.6)
Downloading nltk-3.9.2-py3-none-any.whl (1.5 MB)
   ---------------------------------------- 0.0/1.5 MB ? eta -:--:--
   -- ------------------------------------- 0.1/1.5 MB 5.1 MB/s eta 0:00:01
   ----- ---------------------------------- 0.2/1.5 MB 3.5 MB/s eta 0:00:01
   --------- ------------------------------ 0.4/1.5 MB 3.3 MB/s eta 0:00:01
   ------------ --------------------------- 0.5/1.5 MB 3.0 MB/s eta 0:00:01
   ---------------- ----------------------- 0.6/1.5 MB 3.1 MB/s eta 0:00:01
   ------------------- -------------------- 0.7/1.5 MB 2.9 MB/s eta 0:00:01
   ----------------------- ---------------- 0.9/1.5 MB 2.9 MB/s eta 0:00:01
   ----------------------- ---------------- 0.9/1.5 MB 2.7 MB/s eta 0:00:01
   --------------------------- ------------ 1.0/1.5 MB 2.6 MB/s eta 0:00:01
   ------------------------------ --------- 1.2/1.5 MB 2.6 MB/s eta 0:00:01
   --------------------------------- ------ 1.3/1.5 MB 2.6 MB/s eta 0:00:01
   ------------------------------------- -- 1.4/1.5 MB 2.7 MB/s eta 0:00:01
   ---------------------------------------  1.5/1.5 MB 2.6 MB/s eta 0:00:01
   ---------------------------------------- 1.5/1.5 MB 2.5 MB/s eta 0:00:00
Downloading regex-2025.10.23-cp312-cp312-win_amd64.whl (276 kB)
   ---------------------------------------- 0.0/276.9 kB ? eta -:--:--
   ----------------- ---------------------- 122.9/276.9 kB 3.6 MB/s eta 0:00:01
   ---------------------------------------  276.5/276.9 kB 3.4 MB/s eta 0:00:01
   ---------------------------------------- 276.9/276.9 kB 2.9 MB/s eta 0:00:00
Installing collected packages: regex, nltk
Successfully installed nltk-3.9.2 regex-2025.10.23
```

In [3]:
```
!pip install gensim
```

```
Collecting gensim
  Downloading gensim-4.4.0-cp312-cp312-win_amd64.whl.metadata (8.6 kB)
Requirement already satisfied: numpy>=1.18.5 in c:\users\rahul\miniconda3\lib\s
ite-packages (from gensim) (1.26.4)
Requirement already satisfied: scipy>=1.7.0 in c:\users\rahul\miniconda3\lib\si
te-packages (from gensim) (1.13.1)
Collecting smart_open>=1.8.1 (from gensim)
  Downloading smart_open-7.4.1-py3-none-any.whl.metadata (24 kB)
Requirement already satisfied: wrapt in c:\users\rahul\miniconda3\lib\site-pack
ages (from smart_open>=1.8.1->gensim) (1.17.2)
Downloading gensim-4.4.0-cp312-cp312-win_amd64.whl (24.4 MB)
   ---------------------------------------- 0.0/24.4 MB ? eta -:--:--
   ---------------------------------------- 0.0/24.4 MB ? eta -:--:--
   ---------------------------------------- 0.1/24.4 MB 1.7 MB/s eta 0:00:15
   ---------------------------------------- 0.3/24.4 MB 2.3 MB/s eta 0:00:11
    --------------------------------------- 0.4/24.4 MB 2.6 MB/s eta 0:00:10
    --------------------------------------- 0.5/24.4 MB 2.6 MB/s eta 0:00:10
   - -------------------------------------- 0.7/24.4 MB 2.6 MB/s eta 0:00:10
   - -------------------------------------- 0.8/24.4 MB 2.8 MB/s eta 0:00:09
   - -------------------------------------- 1.0/24.4 MB 2.8 MB/s eta 0:00:09
   - -------------------------------------- 1.1/24.4 MB 2.8 MB/s eta 0:00:09
   -- ------------------------------------- 1.3/24.4 MB 2.9 MB/s eta 0:00:09
   -- ------------------------------------- 1.4/24.4 MB 2.9 MB/s eta 0:00:09
   -- ------------------------------------- 1.5/24.4 MB 2.9 MB/s eta 0:00:08
   -- ------------------------------------- 1.7/24.4 MB 2.9 MB/s eta 0:00:08
   -- ------------------------------------- 1.8/24.4 MB 2.9 MB/s eta 0:00:08
   --- ------------------------------------ 1.9/24.4 MB 2.9 MB/s eta 0:00:08
   --- ------------------------------------ 2.1/24.4 MB 2.9 MB/s eta 0:00:08
   --- ------------------------------------ 2.2/24.4 MB 2.9 MB/s eta 0:00:08
   --- ------------------------------------ 2.3/24.4 MB 2.9 MB/s eta 0:00:08
   ---- ----------------------------------- 2.5/24.4 MB 2.9 MB/s eta 0:00:08
   ---- ----------------------------------- 2.6/24.4 MB 2.9 MB/s eta 0:00:08
   ---- ----------------------------------- 2.8/24.4 MB 2.9 MB/s eta 0:00:08
   ---- ----------------------------------- 2.9/24.4 MB 2.9 MB/s eta 0:00:08
   ----- ---------------------------------- 3.1/24.4 MB 2.9 MB/s eta 0:00:08
   ----- ---------------------------------- 3.2/24.4 MB 2.9 MB/s eta 0:00:08
   ----- ---------------------------------- 3.4/24.4 MB 2.9 MB/s eta 0:00:08
   ----- ---------------------------------- 3.5/24.4 MB 2.9 MB/s eta 0:00:08
   ----- ---------------------------------- 3.6/24.4 MB 2.9 MB/s eta 0:00:08
   ------ --------------------------------- 3.8/24.4 MB 2.9 MB/s eta 0:00:08
   ------ --------------------------------- 3.9/24.4 MB 2.9 MB/s eta 0:00:07
   ------ --------------------------------- 4.1/24.4 MB 2.9 MB/s eta 0:00:07
   ------ --------------------------------- 4.2/24.4 MB 2.9 MB/s eta 0:00:07
   ------- -------------------------------- 4.3/24.4 MB 2.9 MB/s eta 0:00:07
   ------- -------------------------------- 4.5/24.4 MB 2.9 MB/s eta 0:00:07
   ------- -------------------------------- 4.6/24.4 MB 3.0 MB/s eta 0:00:07
   ------- -------------------------------- 4.8/24.4 MB 3.0 MB/s eta 0:00:07
   -------- ------------------------------- 4.9/24.4 MB 3.0 MB/s eta 0:00:07
   -------- ------------------------------- 5.0/24.4 MB 3.0 MB/s eta 0:00:07
   -------- ------------------------------- 5.2/24.4 MB 2.9 MB/s eta 0:00:07
   -------- ------------------------------- 5.3/24.4 MB 3.0 MB/s eta 0:00:07
   -------- ------------------------------- 5.5/24.4 MB 3.0 MB/s eta 0:00:07
   --------- ------------------------------ 5.6/24.4 MB 3.0 MB/s eta 0:00:07
   --------- ------------------------------ 5.7/24.4 MB 3.0 MB/s eta 0:00:07
```

```
--------- ------------------------------ 5.9/24.4 MB 3.0 MB/s eta 0:00:07
--------- ------------------------------ 6.1/24.4 MB 3.0 MB/s eta 0:00:07
---------- ----------------------------- 6.2/24.4 MB 3.0 MB/s eta 0:00:07
---------- ----------------------------- 6.3/24.4 MB 3.0 MB/s eta 0:00:07
---------- ----------------------------- 6.5/24.4 MB 3.0 MB/s eta 0:00:06
---------- ----------------------------- 6.7/24.4 MB 3.0 MB/s eta 0:00:06
----------- ---------------------------- 6.8/24.4 MB 3.0 MB/s eta 0:00:06
----------- ---------------------------- 6.9/24.4 MB 3.0 MB/s eta 0:00:06
----------- ---------------------------- 7.1/24.4 MB 3.0 MB/s eta 0:00:06
----------- ---------------------------- 7.2/24.4 MB 3.0 MB/s eta 0:00:06
------------ --------------------------- 7.3/24.4 MB 3.0 MB/s eta 0:00:06
------------ --------------------------- 7.5/24.4 MB 3.0 MB/s eta 0:00:06
------------ --------------------------- 7.6/24.4 MB 3.0 MB/s eta 0:00:06
------------ --------------------------- 7.8/24.4 MB 3.0 MB/s eta 0:00:06
------------ --------------------------- 7.9/24.4 MB 3.0 MB/s eta 0:00:06
------------- -------------------------- 8.0/24.4 MB 3.0 MB/s eta 0:00:06
------------- -------------------------- 8.1/24.4 MB 3.0 MB/s eta 0:00:06
------------- -------------------------- 8.2/24.4 MB 3.0 MB/s eta 0:00:06
------------- -------------------------- 8.4/24.4 MB 3.0 MB/s eta 0:00:06
------------- -------------------------- 8.5/24.4 MB 3.0 MB/s eta 0:00:06
-------------- ------------------------- 8.6/24.4 MB 3.0 MB/s eta 0:00:06
-------------- ------------------------- 8.8/24.4 MB 3.0 MB/s eta 0:00:06
-------------- ------------------------- 8.9/24.4 MB 2.9 MB/s eta 0:00:06
-------------- ------------------------- 9.0/24.4 MB 2.9 MB/s eta 0:00:06
-------------- ------------------------- 9.1/24.4 MB 2.9 MB/s eta 0:00:06
--------------- ------------------------ 9.3/24.4 MB 2.9 MB/s eta 0:00:06
--------------- ------------------------ 9.4/24.4 MB 2.9 MB/s eta 0:00:06
--------------- ------------------------ 9.5/24.4 MB 2.9 MB/s eta 0:00:06
--------------- ------------------------ 9.6/24.4 MB 2.9 MB/s eta 0:00:06
--------------- ------------------------ 9.8/24.4 MB 2.9 MB/s eta 0:00:06
---------------- ----------------------- 9.9/24.4 MB 2.9 MB/s eta 0:00:05
---------------- ----------------------- 10.0/24.4 MB 2.9 MB/s eta 0:00:05
---------------- ----------------------- 10.1/24.4 MB 2.9 MB/s eta 0:00:05
---------------- ----------------------- 10.2/24.4 MB 2.9 MB/s eta 0:00:05
---------------- ----------------------- 10.3/24.4 MB 2.9 MB/s eta 0:00:05
----------------- ---------------------- 10.4/24.4 MB 2.9 MB/s eta 0:00:05
----------------- ---------------------- 10.5/24.4 MB 2.9 MB/s eta 0:00:05
----------------- ---------------------- 10.7/24.4 MB 2.9 MB/s eta 0:00:05
----------------- ---------------------- 10.8/24.4 MB 2.9 MB/s eta 0:00:05
----------------- ---------------------- 11.0/24.4 MB 2.9 MB/s eta 0:00:05
------------------ --------------------- 11.1/24.4 MB 2.9 MB/s eta 0:00:05
------------------ --------------------- 11.2/24.4 MB 2.9 MB/s eta 0:00:05
------------------ --------------------- 11.4/24.4 MB 2.9 MB/s eta 0:00:05
------------------ --------------------- 11.5/24.4 MB 2.9 MB/s eta 0:00:05
------------------- -------------------- 11.6/24.4 MB 2.9 MB/s eta 0:00:05
------------------- -------------------- 11.7/24.4 MB 2.9 MB/s eta 0:00:05
------------------- -------------------- 11.9/24.4 MB 2.9 MB/s eta 0:00:05
------------------- -------------------- 12.0/24.4 MB 2.9 MB/s eta 0:00:05
------------------- -------------------- 12.1/24.4 MB 2.9 MB/s eta 0:00:05
-------------------- ------------------- 12.3/24.4 MB 2.9 MB/s eta 0:00:05
-------------------- ------------------- 12.4/24.4 MB 2.9 MB/s eta 0:00:05
-------------------- ------------------- 12.5/24.4 MB 2.9 MB/s eta 0:00:05
-------------------- ------------------- 12.6/24.4 MB 2.9 MB/s eta 0:00:05
--------------------- ------------------ 12.8/24.4 MB 2.9 MB/s eta 0:00:05
```

```
--------------------- ------------------ 12.9/24.4 MB 2.8 MB/s eta 0:00:05
--------------------- ------------------ 13.0/24.4 MB 2.9 MB/s eta 0:00:04
--------------------- ------------------ 13.1/24.4 MB 2.8 MB/s eta 0:00:04
--------------------- ------------------ 13.3/24.4 MB 2.8 MB/s eta 0:00:04
---------------------- ----------------- 13.4/24.4 MB 2.8 MB/s eta 0:00:04
---------------------- ----------------- 13.6/24.4 MB 2.8 MB/s eta 0:00:04
---------------------- ----------------- 13.7/24.4 MB 2.8 MB/s eta 0:00:04
---------------------- ----------------- 13.8/24.4 MB 2.8 MB/s eta 0:00:04
---------------------- ----------------- 14.0/24.4 MB 2.8 MB/s eta 0:00:04
----------------------- ---------------- 14.1/24.4 MB 2.8 MB/s eta 0:00:04
----------------------- ---------------- 14.2/24.4 MB 2.8 MB/s eta 0:00:04
----------------------- ---------------- 14.3/24.4 MB 2.8 MB/s eta 0:00:04
----------------------- ---------------- 14.4/24.4 MB 2.8 MB/s eta 0:00:04
----------------------- ---------------- 14.6/24.4 MB 2.8 MB/s eta 0:00:04
------------------------ --------------- 14.7/24.4 MB 2.8 MB/s eta 0:00:04
------------------------ --------------- 14.8/24.4 MB 2.8 MB/s eta 0:00:04
------------------------ --------------- 15.0/24.4 MB 2.8 MB/s eta 0:00:04
------------------------ --------------- 15.1/24.4 MB 2.8 MB/s eta 0:00:04
------------------------ --------------- 15.2/24.4 MB 2.8 MB/s eta 0:00:04
------------------------- -------------- 15.4/24.4 MB 2.8 MB/s eta 0:00:04
------------------------- -------------- 15.5/24.4 MB 2.8 MB/s eta 0:00:04
------------------------- -------------- 15.6/24.4 MB 2.8 MB/s eta 0:00:04
------------------------- -------------- 15.8/24.4 MB 2.8 MB/s eta 0:00:04
-------------------------- ------------- 15.9/24.4 MB 2.8 MB/s eta 0:00:04
-------------------------- ------------- 16.0/24.4 MB 2.8 MB/s eta 0:00:04
-------------------------- ------------- 16.2/24.4 MB 2.8 MB/s eta 0:00:03
-------------------------- ------------- 16.3/24.4 MB 2.8 MB/s eta 0:00:03
-------------------------- ------------- 16.4/24.4 MB 2.8 MB/s eta 0:00:03
--------------------------- ------------ 16.6/24.4 MB 2.8 MB/s eta 0:00:03
--------------------------- ------------ 16.7/24.4 MB 2.8 MB/s eta 0:00:03
--------------------------- ------------ 16.8/24.4 MB 2.8 MB/s eta 0:00:03
--------------------------- ------------ 17.0/24.4 MB 2.8 MB/s eta 0:00:03
---------------------------- ----------- 17.1/24.4 MB 2.8 MB/s eta 0:00:03
---------------------------- ----------- 17.2/24.4 MB 2.8 MB/s eta 0:00:03
---------------------------- ----------- 17.3/24.4 MB 2.8 MB/s eta 0:00:03
---------------------------- ----------- 17.5/24.4 MB 2.7 MB/s eta 0:00:03
---------------------------- ----------- 17.6/24.4 MB 2.8 MB/s eta 0:00:03
----------------------------- ---------- 17.7/24.4 MB 2.8 MB/s eta 0:00:03
----------------------------- ---------- 17.9/24.4 MB 2.8 MB/s eta 0:00:03
----------------------------- ---------- 18.1/24.4 MB 2.8 MB/s eta 0:00:03
----------------------------- ---------- 18.2/24.4 MB 2.8 MB/s eta 0:00:03
------------------------------ --------- 18.3/24.4 MB 2.8 MB/s eta 0:00:03
------------------------------ --------- 18.4/24.4 MB 2.8 MB/s eta 0:00:03
------------------------------ --------- 18.6/24.4 MB 2.8 MB/s eta 0:00:03
------------------------------ --------- 18.7/24.4 MB 2.8 MB/s eta 0:00:03
------------------------------ --------- 18.9/24.4 MB 2.8 MB/s eta 0:00:02
------------------------------- -------- 19.0/24.4 MB 2.8 MB/s eta 0:00:02
------------------------------- -------- 19.1/24.4 MB 2.8 MB/s eta 0:00:02
------------------------------- -------- 19.3/24.4 MB 2.8 MB/s eta 0:00:02
------------------------------- -------- 19.4/24.4 MB 2.8 MB/s eta 0:00:02
-------------------------------- ------- 19.5/24.4 MB 2.8 MB/s eta 0:00:02
-------------------------------- ------- 19.7/24.4 MB 2.8 MB/s eta 0:00:02
-------------------------------- ------- 19.8/24.4 MB 2.8 MB/s eta 0:00:02
-------------------------------- ------- 19.9/24.4 MB 2.8 MB/s eta 0:00:02
```

```
   -------------------------------- ------- 20.1/24.4 MB 2.8 MB/s eta 0:00:02
   --------------------------------- ------ 20.2/24.4 MB 2.8 MB/s eta 0:00:02
   --------------------------------- ------ 20.3/24.4 MB 2.8 MB/s eta 0:00:02
   --------------------------------- ------ 20.5/24.4 MB 2.8 MB/s eta 0:00:02
   --------------------------------- ------ 20.6/24.4 MB 2.8 MB/s eta 0:00:02
   --------------------------------- ------ 20.7/24.4 MB 2.8 MB/s eta 0:00:02
   ---------------------------------- ----- 20.9/24.4 MB 2.8 MB/s eta 0:00:02
   ---------------------------------- ----- 21.0/24.4 MB 2.8 MB/s eta 0:00:02
   ---------------------------------- ----- 21.1/24.4 MB 2.8 MB/s eta 0:00:02
   ---------------------------------- ----- 21.3/24.4 MB 2.8 MB/s eta 0:00:02
   ----------------------------------- ---- 21.4/24.4 MB 2.8 MB/s eta 0:00:02
   ----------------------------------- ---- 21.4/24.4 MB 2.8 MB/s eta 0:00:02
   ----------------------------------- ---- 21.5/24.4 MB 2.8 MB/s eta 0:00:02
   ----------------------------------- ---- 21.6/24.4 MB 2.8 MB/s eta 0:00:02
   ----------------------------------- ---- 21.7/24.4 MB 2.8 MB/s eta 0:00:01
   ----------------------------------- ---- 21.8/24.4 MB 2.8 MB/s eta 0:00:01
   ----------------------------------- ---- 21.8/24.4 MB 2.8 MB/s eta 0:00:01
   ------------------------------------ --- 22.0/24.4 MB 2.7 MB/s eta 0:00:01
   ------------------------------------ --- 22.1/24.4 MB 2.7 MB/s eta 0:00:01
   ------------------------------------ --- 22.2/24.4 MB 2.7 MB/s eta 0:00:01
   ------------------------------------ --- 22.4/24.4 MB 2.7 MB/s eta 0:00:01
   ------------------------------------ --- 22.5/24.4 MB 2.7 MB/s eta 0:00:01
   ------------------------------------- -- 22.7/24.4 MB 2.7 MB/s eta 0:00:01
   ------------------------------------- -- 22.8/24.4 MB 2.7 MB/s eta 0:00:01
   ------------------------------------- -- 22.9/24.4 MB 2.8 MB/s eta 0:00:01
   ------------------------------------- -- 23.1/24.4 MB 2.8 MB/s eta 0:00:01
   -------------------------------------- - 23.2/24.4 MB 2.8 MB/s eta 0:00:01
   -------------------------------------- - 23.3/24.4 MB 2.8 MB/s eta 0:00:01
   -------------------------------------- - 23.5/24.4 MB 2.8 MB/s eta 0:00:01
   -------------------------------------- - 23.6/24.4 MB 2.8 MB/s eta 0:00:01
   -------------------------------------- - 23.7/24.4 MB 2.8 MB/s eta 0:00:01
   ---------------------------------------  23.8/24.4 MB 2.8 MB/s eta 0:00:01
   ---------------------------------------  24.0/24.4 MB 2.8 MB/s eta 0:00:01
   ---------------------------------------  24.1/24.4 MB 2.8 MB/s eta 0:00:01
   ---------------------------------------  24.2/24.4 MB 2.7 MB/s eta 0:00:01
   ---------------------------------------  24.4/24.4 MB 2.8 MB/s eta 0:00:01
   ---------------------------------------  24.4/24.4 MB 2.8 MB/s eta 0:00:01
   ---------------------------------------  24.4/24.4 MB 2.8 MB/s eta 0:00:01
   ---------------------------------------- 24.4/24.4 MB 2.7 MB/s eta 0:00:00
Downloading smart_open-7.4.1-py3-none-any.whl (63 kB)
   ---------------------------------------- 0.0/63.2 kB ? eta -:--:--
   ---------------------------------------- 63.2/63.2 kB 3.3 MB/s eta 0:00:00
Installing collected packages: smart_open, gensim
Successfully installed gensim-4.4.0 smart_open-7.4.1
```

In [4]:

```python
import nltk
from nltk.tokenize import word_tokenize, sent_tokenize
from nltk.corpus import stopwords
import re
import gensim
from gensim.models import Word2Vec
```

In [5]:

```python
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('punkt_tab')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\rahul\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping tokenizers\punkt.zip.
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\rahul\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\stopwords.zip.
[nltk_data] Downloading package punkt_tab to
[nltk_data]     C:\Users\rahul\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping tokenizers\punkt_tab.zip.
```

Out[5]:

```
True
```

In [6]:

```
sample_text = """
Machine learning is a field of computer science that gives computers the abili
It focuses on the development of algorithms that can analyze and interpret pat
Machine learning is used in a variety of applications such as email filtering,
The algorithms improve their performance as they are exposed to more data over
Learning from data enables machines to make predictions and decisions based or
"""
```

In [7]:

```
sentences = re.sub('[^A-Za-z]+', ' ', sample_text)
sentences = re.sub(r'(?:^| )\w(?:$| )', ' ', sentences).strip()
sentences = sentences.lower()
sentences
```

Out[7]:

```
'machine learning is field of computer science that gives computers the abili
ty to learn without being explicitly programmed it focuses on the development
of algorithms that can analyze and interpret patterns in data machine learnin
g is used in variety of applications such as email filtering speech recogniti
on and computer vision the algorithms improve their performance as they are e
xposed to more data over time learning from data enables machines to make pre
dictions and decisions based on past experiences'
```

In [8]:

```
all_sentences = sent_tokenize(sentences)
all_words = [word_tokenize(word) for word in all_sentences]
print(all_words)
```

```
[['machine', 'learning', 'is', 'field', 'of', 'computer', 'science', 'that', 'g
ives', 'computers', 'the', 'ability', 'to', 'learn', 'without', 'being', 'expli
citly', 'programmed', 'it', 'focuses', 'on', 'the', 'development', 'of', 'algor
ithms', 'that', 'can', 'analyze', 'and', 'interpret', 'patterns', 'in', 'data',
'machine', 'learning', 'is', 'used', 'in', 'variety', 'of', 'applications', 'su
ch', 'as', 'email', 'filtering', 'speech', 'recognition', 'and', 'computer', 'v
ision', 'the', 'algorithms', 'improve', 'their', 'performance', 'as', 'they',
'are', 'exposed', 'to', 'more', 'data', 'over', 'time', 'learning', 'from', 'da
ta', 'enables', 'machines', 'to', 'make', 'predictions', 'and', 'decisions', 'b
ased', 'on', 'past', 'experiences']]
```

In [9]:

```
for i in range(len(all_words)):
    all_words[i] = [w for w in all_words[i] if w not in stopwords.words('engli
data = all_words
data1 = sum(data, [])
print(data)
print(data1)
```

```
[['machine', 'learning', 'field', 'computer', 'science', 'gives', 'computers',
'ability', 'learn', 'without', 'explicitly', 'programmed', 'focuses', 'developm
ent', 'algorithms', 'analyze', 'interpret', 'patterns', 'data', 'machine', 'lea
rning', 'used', 'variety', 'applications', 'email', 'filtering', 'speech', 'rec
ognition', 'computer', 'vision', 'algorithms', 'improve', 'performance', 'expos
ed', 'data', 'time', 'learning', 'data', 'enables', 'machines', 'make', 'predic
tions', 'decisions', 'based', 'past', 'experiences']]
['machine', 'learning', 'field', 'computer', 'science', 'gives', 'computers',
'ability', 'learn', 'without', 'explicitly', 'programmed', 'focuses', 'developm
ent', 'algorithms', 'analyze', 'interpret', 'patterns', 'data', 'machine', 'lea
rning', 'used', 'variety', 'applications', 'email', 'filtering', 'speech', 'rec
ognition', 'computer', 'vision', 'algorithms', 'improve', 'performance', 'expos
ed', 'data', 'time', 'learning', 'data', 'enables', 'machines', 'make', 'predic
tions', 'decisions', 'based', 'past', 'experiences']
```

In [10]:

```
context_target_pairs = []
window_size = 2
```

In [11]:

```
for i in range(window_size, len(data1) - window_size):
    context = [data1[i - 2], data1[i - 1], data1[i + 1], data1[i + 2]]
    target = data1[i]
    context_target_pairs.append((context, target))
```

In [12]:

```
print("First 5 context-target pairs:")
print(context_target_pairs[:5])
```

```
First 5 context-target pairs:
[(['machine', 'learning', 'computer', 'science'], 'field'), (['learning', 'fiel
d', 'science', 'gives'], 'computer'), (['field', 'computer', 'gives', 'computer
s'], 'science'), (['computer', 'science', 'computers', 'ability'], 'gives'),
(['science', 'gives', 'ability', 'learn'], 'computers')]
```

In [13]:

```
model = Word2Vec(sentences=data, vector_size=50, window=window_size, min_count
```

In [14]:

```
word = input("Enter a word: ")
```

In [15]:

```
print(f"\nMost similar words to '{word}':")
similar_words = model.wv.most_similar(word)
for sim_word, score in similar_words:
    print(f"{sim_word}: {score:.4f}")
```

```
Most similar words to 'science':
performance: 0.2402
based: 0.2400
analyze: 0.2369
patterns: 0.2242
decisions: 0.2032
variety: 0.1825
gives: 0.1683
enables: 0.1576
experiences: 0.1413
vision: 0.1351
```

In [16]:

```
print(list(model.wv.key_to_index.keys())[:50])
```

```
['data', 'learning', 'algorithms', 'computer', 'machine', 'experiences', 'pas
t', 'based', 'decisions', 'predictions', 'make', 'machines', 'enables', 'time',
'exposed', 'performance', 'improve', 'vision', 'recognition', 'speech', 'filter
ing', 'email', 'applications', 'variety', 'used', 'patterns', 'interpret', 'ana
lyze', 'development', 'focuses', 'programmed', 'explicitly', 'without', 'lear
n', 'ability', 'computers', 'gives', 'science', 'field']
```

In [17]:

```
#sg
modelsg = Word2Vec(sentences=data, vector_size=50, window=window_size, min_cou
```

In [ ]:

```

```

In [18]:

```
import numpy as np
import matplotlib.pyplot as plt
```

In [19]:

```
from sklearn.decomposition import PCA
```

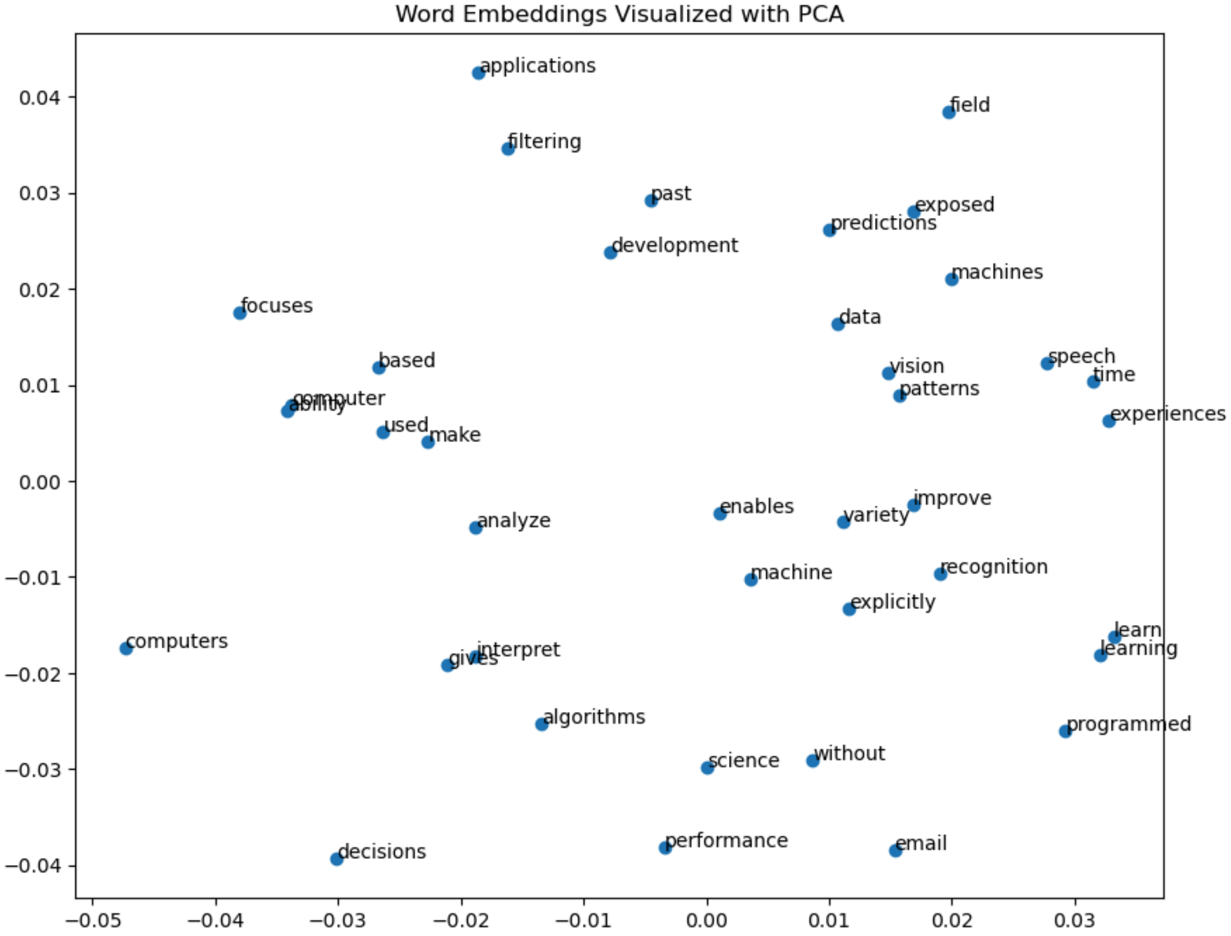
In [20]:

```
X = model.wv[model.wv.index_to_key]
pca = PCA(n_components=2)
result = pca.fit_transform(X)
```

In [21]:

```
plt.figure(figsize=(10, 8))
plt.scatter(result[:, 0], result[:, 1])
words = model.wv.index_to_key

for i, word in enumerate(words):
    plt.annotate(word, xy=(result[i, 0], result[i, 1]))
plt.title("Word Embeddings Visualized with PCA")
plt.show()
```


Word Embeddings Visualized with PCA
applications
field
filtering
past
exposed
predictions
development
machines
focuses
data
based
vision
speech
time
patterns
computer
ability
used
make
experiences
enables
improve
variety
analyze
machine
recognition
explicitly
learn
learning
computers
interpret
gives
algorithms
programmed
science
without
performance
email
decisions
0.04
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
−0.05
−0.04
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03

```
In [ ]:
```