cosinesimilarity

October 12, 2025

[1]:
```
import numpy as np
import pandas as pd

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

[2]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity

def read_file(filename):
    with open(filename, 'r', encoding='utf-8') as file:
        return file.read()

def compute_cosine_similarity(doc1, doc2):
    vectorizer = TfidfVectorizer()
    tfidf_matrix = vectorizer.fit_transform([doc1, doc2])
    return cosine_similarity(tfidf_matrix[0:1], tfidf_matrix[1:2])[0][0]

file1 = 'document1.txt'
file2 = 'document2.txt'

doc1 = read_file(file1)
doc2 = read_file(file2)

similarity = compute_cosine_similarity(doc1, doc2)

print(f"Cosine Similarity: {similarity:.4f}")
```

```
Cosine Similarity: 0.4119
```

[3]:
```
def read_file(filename):
    with open(filename, 'r', encoding='utf-8') as file:
        return file.read()


doc1 = read_file(file1)
```

```
doc2 = read_file(file2)

print("Content of Document 1:\n")
print(doc1)
print("\nContent of Document 2:\n")
print(doc2)
```

```
Content of Document 1:

Artificial intelligence and machine learning are transforming industries. AI is
enabling computers to perform tasks that were once thought to be exclusively
human capabilities, such as understanding language, recognizing patterns, and
making decisions. The future of AI is vast, with possibilities ranging from
healthcare to autonomous vehicles.


Content of Document 2:

Machine learning and artificial intelligence are changing the way industries
operate. AI allows machines to understand patterns, make decisions, and even
process language, which were once human-only abilities. The applications of AI
are extensive, from self-driving cars to advancements in healthcare.
```

[ ]: