TF-IDF

{'جامعة': 4, 'تبوك': 3, 'شروط': 5, 'القبول': 2, 'في': 6, 'الرسوم': 1, 'الدراسية': 0, 'لجامعة': 7}

from sklearn.feature\_extraction.text import TfidfVectorizer

import numpy as np

# مجموعة النصوص للتدريب

documents = [

"جامعة تبوك",

"شروط القبول في جامعة تبوك",

"الرسوم الدراسية لجامعة تبوك"

]

# إنشاء الـ TF-IDF Vectorizer

vectorizer = TfidfVectorizer()

# ===== 1) fit =====

vectorizer.fit(documents)

# طباعة القاموس

print("📌 القاموس (Vocabulary):")

for word, idx in vectorizer.vocabulary\_.items():

print(f"{word} → index {idx}")

# طباعة IDF لكل كلمة

print("\n📌 قيم IDF لكل كلمة:")

for word, idx in vectorizer.vocabulary\_.items():

print(f"{word}: {vectorizer.idf\_[idx]:.4f}")

# ===== 2) حساب TF يدوي =====

def term\_frequency(doc, word):

words = doc.split()

return words.count(word) / len(words) # عدد مرات الكلمة / عدد كلمات النص

print("\n📌 قيم TF لكل كلمة في كل مستند:")

for i, doc in enumerate(documents):

print(f"\nالمستند {i+1}: \"{doc}\"")

for word in vectorizer.vocabulary\_:

print(f"{word}: TF = {term\_frequency(doc, word):.4f}")

# ===== 3) transform لحساب TF-IDF =====

tfidf\_matrix = vectorizer.transform(documents)

print("\n📌 مصفوفة TF-IDF (القيم النهائية):")

print(tfidf\_matrix.toarray())

### ****القاموس (Vocabulary****

|  |  |
| --- | --- |
| الكلمة | Index |
| جامعة | 4 |
| تبوك | 3 |
| شروط | 5 |
| القبول | 2 |
| في | 6 |
| الرسوم | 1 |
| الدراسية | 0 |
| لجامعة | 7 |
|  |  |

كل كلمة لها رقم (Index) في القاموس.

قيم IDF لكل كلمة

|  |  |
| --- | --- |
| الكلمة | IDF |
| جامعة | 1.2877 |
| تبوك | 1 |
| شروط | 1.6931 |
| القبول | 1.6931 |
| في | 1.6931 |
| الرسوم | 1.6931 |
| الدراسية | 1.6931 |
| لجامعة | 1.6931 |

الكلمات الأكثر تكرارًا عبر المستندات تأخذ IDF أصغر، والنادرة تأخذ IDF أكبر

قيم TF لكل كلمة في كل مستند

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| المستند | النص | الكلمة | TF |  |
| 1 | جامعة تبوك | جامعة | 0.5 |  |
| 1 | جامعة تبوك | تبوك | 0.5 |  |
| 1 | جامعة تبوك | شروط | 0 |  |
| 1 | جامعة تبوك | القبول | 0 |  |
| 1 | جامعة تبوك | في | 0 |  |
| 1 | جامعة تبوك | الرسوم | 0 |  |
| 1 | جامعة تبوك | الدراسية | 0 |  |
| 1 | جامعة تبوك | لجامعة | 0 |  |
| 2 | شروط القبول في جامعة تبوك | جامعة | 0.2 |  |
| 2 | شروط القبول في جامعة تبوك | تبوك | 0.2 |  |
| 2 | شروط القبول في جامعة تبوك | شروط | 0.2 |  |
| 2 | شروط القبول في جامعة تبوك | القبول | 0.2 |  |
| 2 | شروط القبول في جامعة تبوك | في | 0.2 |  |
| 2 | شروط القبول في جامعة تبوك | الرسوم | 0 |  |
| 2 | شروط القبول في جامعة تبوك | الدراسية | 0 |  |
| 2 | شروط القبول في جامعة تبوك | لجامعة | 0 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | جامعة | 0 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | تبوك | 0.25 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | شروط | 0 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | القبول | 0 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | في | 0 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | الرسوم | 0.25 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | الدراسية | 0.25 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | لجامعة | 0.25 |  |
|  |  |  |  |  |

### ****مصفوفة TF-IDF النهائية****

كل صف = مستند، كل عمود = كلمة من القاموس.[

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| المستند | النص | الكلمة | TF-IDF |  |
| 1 | جامعة تبوك | جامعة | 0.7898 |  |
| 1 | جامعة تبوك | تبوك | 0.6134 |  |
| 1 | جامعة تبوك | شروط | 0 |  |
| 1 | جامعة تبوك | القبول | 0 |  |
| 1 | جامعة تبوك | في | 0 |  |
| 1 | جامعة تبوك | الرسوم | 0 |  |
| 1 | جامعة تبوك | الدراسية | 0 |  |
| 1 | جامعة تبوك | لجامعة | 0 |  |
| 2 | شروط القبول في جامعة تبوك | جامعة | 0.3838 |  |
| 2 | شروط القبول في جامعة تبوك | تبوك | 0.298 |  |
| 2 | شروط القبول في جامعة تبوك | شروط | 0.5046 |  |
| 2 | شروط القبول في جامعة تبوك | القبول | 0.5046 |  |
| 2 | شروط القبول في جامعة تبوك | في | 0.5046 |  |
| 2 | شروط القبول في جامعة تبوك | الرسوم | 0 |  |
| 2 | شروط القبول في جامعة تبوك | الدراسية | 0 |  |
| 2 | شروط القبول في جامعة تبوك | لجامعة | 0 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | جامعة | 0 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | تبوك | 0.3227 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | شروط | 0 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | القبول | 0 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | في | 0 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | الرسوم | 0.5465 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | الدراسية | 0.5465 |  |
| 3 | الرسوم الدراسية لجامعة تبوك | لجامعة | 0.5465 |  |

[0.0, 0.0, 0.0, 0.6134, 0.7898, 0.0, 0.0, 0.0],

[0.0, 0.0, 0.5046, 0.2980, 0.3838, 0.5046, 0.5046, 0.0],

[0.5465, 0.5465, 0.0, 0.3227, 0.0, 0.0, 0.0, 0.5465]

]

القيمة = **TF × IDF** لكل كلمة في كل مستند.