فهمت قصدك 👍، أنت تقول:

> إحنا أصلاً كنا شغالين على \*\*نسب التشابه\*\*، فكيف عرف الكود رقم الـ index للسؤال والجواب في الجدول `self.data` إذا عملية التشابه كانت على أرقام؟

الجواب هو: لأن ترتيب قيم التشابه \*\*مرتبط مباشرة بترتيب الصفوف في `self.data`\*\* من البداية.

---

## 1️⃣ كيف بدأ الربط بين التشابه والـ index

\* لما درّبنا الـ TF-IDF:

```python

self.tfidf\_matrix = self.vectorizer.transform(self.data['combined\_text'])

```

كل \*\*صف\*\* في `self.tfidf\_matrix` يمثل مستند (سؤال+إجابة) من `self.data` بنفس الترتيب.

مثال:

| index في DataFrame | نص المستند | صف في tfidf\\_matrix |

| ------------------ | ------------------------ | ------------------- |

| 0 | "ما هي شروط القبول ..." | الصف 0 |

| 1 | "ما هي رسوم التسجيل ..." | الصف 1 |

| 2 | "أين تقع جامعة تبوك؟" | الصف 2 |

---

## 2️⃣ عند حساب التشابه

```python

similarities = cosine\_similarity(query\_vec, self.tfidf\_matrix).flatten()

```

\* `cosine\_similarity` يحسب التشابه بين \*\*متجه السؤال\*\* (`query\_vec`) وكل صف في `self.tfidf\_matrix`.

\* النتيجة `similarities` هي قائمة بنفس ترتيب الصفوف في `self.tfidf\_matrix`، وبالتالي نفس ترتيب `self.data`.

مثال:

```

similarities[0] → التشابه مع المستند 0 في self.data

similarities[1] → التشابه مع المستند 1 في self.data

similarities[2] → التشابه مع المستند 2 في self.data

```

---

## 3️⃣ لما نجيب أعلى التشابهات

```python

top\_indices = similarities.argsort()[-top\_n:][::-1]

```

\* `argsort()` يرجع \*\*أرقام الفهارس\*\* (indexes) مرتبة حسب القيم.

\* هذه الفهارس هي نفسها فهارس الصفوف في `self.data`.

\* لذلك لما نكتب:

```python

self.data.iloc[idx]['question']

```

فـ `idx` يطابق تمامًا صف السؤال/الإجابة اللي أعطى هذه النسبة من التشابه.

---

💡 \*\*الخلاصة:\*\*

الـ `index` ما هو ناتج عملية التشابه، لكنه \*\*يأتي من ترتيب الصفوف\*\*، لأن عملية التشابه تمت على نفس المصفوفة اللي مبنية من `self.data`، فالأماكن (الفهارس) ظلت مرتبطة.

---

إذا تحب، أقدر أعمل لك \*\*رسم جدول\*\* يربط بين:

\* الصف في `self.data`

\* الصف في `tfidf\_matrix`

\* مكان النسبة في `similarities`

عشان تشوف العلاقة أوضح.