APRENDIZAJE AUTOMÁTICO PARA LA GESTIÓN DE DATOS MASIVOS

BY DR. PAULO LOPEZ MEYER

30/07/24

TAREA 10

ISAAC MENCHACA

ANÁLISIS DE TEXTO

1. Implementar el código de Embeddings.

El resultado muestra una lista de documentos similares a 'doc1', junto con sus puntuaciones de similitud. Las puntuaciones indican cuán similares son los documentos a 'doc1'. Una puntuación más alta significa mayor similitud. En este caso:

- 'doc2' tiene una similitud de 0.1625 con 'doc1'.
- 'doc3' tiene una similitud de 0.0014 con 'doc1'.
- 'doc4' tiene una similitud de -0.0161 con 'doc1'.

```
Documentos similares a 'doc1':
 [('doc2', 0.16251666843891144), ('doc3', 0.0014203854370862246), ('doc4', -0.016057293862104416)]
```

2. Implementar el código de TF-IDF.

el resultado representa la importancia de cada palabra en cada documento del conjunto de datos. En el primer documento, las palabras an, field, e interdisciplinary tienen los valores más altos, indicando que son más relevantes en este documento. El segundo documento son learning, machine, of, y part. Y finalmente el tercer documento lo son involves y statistics.

	an	data	field	interdisciplinary	involves	is	\
0	0.483591	0.285617	0.483591	0.483591	0.000000	0.367784	
1	0.000000	0.257129	0.000000	0.000000	0.000000	0.331100	
2	0.000000	0.359594	0.000000	0.000000	0.608845	0.000000	

	learning	machine	of	part	science	statistics
0	0.000000	0.000000	0.000000	0.000000	0.285617	0.000000
1	0.435357	0.435357	0.435357	0.435357	0.257129	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.359594	0.608845

3. Implementar el código de análisis de sentimiento para 10 reviews diferentes.

get_sentiment proporciona dos métricas para cada texto: polaridad y subjetividad.

1. Polaridad: Indica el grado de positividad o negatividad del texto. Los valores van de -1 (muy negativo) a 1 (muy positivo).
2. Subjetividad: Indica cuánto de subjetivo u objetivo es el texto. Los valores van de 0 (muy objetivo) a 1 (muy subjetivo).

En los resultados podemos concluir que todas las opiniones son bastante subjetivas, es decir, basado en opiniones personales. Y que la mayoría de los resultados son positivos 3/5.

```
Text: "I love this product! It's absolutely amazing."
Polarity: 0.6125, Subjectivity: 0.75

Text: "This is the worst experience I've ever had."
Polarity: -1.0, Subjectivity: 1.0

Text: "The movie was okay, not great but not terrible either."
Polarity: 0.19999999999999998, Subjectivity: 0.75

Text: "I'm extremely happy with the service."
Polarity: 0.8, Subjectivity: 1.0

Text: "I hate waiting in long lines."
Polarity: -0.42500000000000004, Subjectivity: 0.65
```