



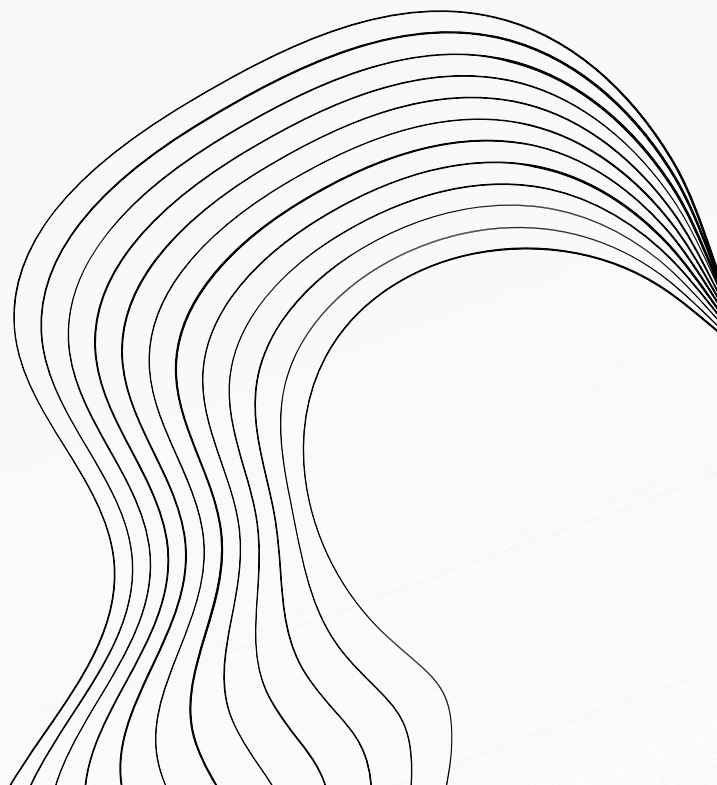
APRENDIZAJE AUTOMÁTICO PARA LA GESTIÓN DE DATOS MASIVOS

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).

Escogí poner 10 reviews de una funda para celular de iPhone, los primeros 5 reviews de personas que calificaron el producto con 5 estrellas y los últimos 5 con personas que calificaron el producto con 1 estrella. Se puede observar que los primeros 5 tienen una puntuación de polaridad arriba de 0.2. Los 5 últimos, que son los negativos, tienen una polaridad debajo de 0.2. En la subjetividad, los primeros 5 tienen arriba de 0.5 y los últimos 5, la mayoría tienen por debajo de 0.5.

```
Text: "Very happy with this phone case. Has enough of a ridge around the camera to protect the lens. Time will tell if it ends up yellowing or not. Will order again."
Polarity: 0.5, Subjectivity: 0.75

Text: "Just put this on my iPhone. It's really nice. Kind of a hard rubber feel to it. Screen protector and lens protector fit with the case. Very happy w this purchase."
Polarity: 0.4616666666666666, Subjectivity: 0.7683333333333333

Text: "Most cases make my phone heavy but this didnt. It sticks to everything and stays in place!! Very easy to put it on or take out if you need to."
Polarity: 0.2502777777777777, Subjectivity: 0.6666666666666666

Text: "I love this new case. It is light and because of the design, I don't need a screen protector. My husband ordered one as soon as he saw mine."
Polarity: 0.34545454545454546, Subjectivity: 0.5848484848484848

Text: "This cover fits perfect, looks great and I can charge it on my stand without removing the cover. Gets this for your iPhone 15 Pro Max!"
Polarity: 1.0, Subjectivity: 0.875

Text: "As in description says it's yellowing resistant but it's start getting yellow in a 4,5 days not and the quality also is not good. Please don't waste your money to t
Polarity: -0.1875, Subjectivity: 0.15000000000000002

Text: "I order the case for my iPhone 15 Pro Max, but it did not fit."
Polarity: -0.2, Subjectivity: 0.4

Text: "This product is okay. It serves its purpose adequately without any major issues. It's not exceptional, but it gets the job done. Decent value for the price."
Polarity: 0.14583333333333331, Subjectivity: 0.5999999999999999

Text: "Case is durable however although they claim it doesn't yellow it does yellow and it discolors. I wouldn't recommend it. I tried to contact them directly but they ha
Polarity: 0.03333333333333333, Subjectivity: 0.13333333333333333

Text: "This case doesn't fit my husband's iPhone 15 Plus at all. Definitely sending it back. And will just stick to Otter Box."
Polarity: 0.13333333333333333, Subjectivity: 0.3
```