

# T11. Actividad Sumativa 3

## 1. ¿Qué es Opinion Spam Detection?

*Opinion Spam Detection* se refiere a la tarea de identificar y filtrar reseñas falsas o engañosas que buscan manipular la percepción de productos, servicios u organizaciones.

El spam de opinión se ha convertido en un problema importante teniendo incidencia incluso sobre cuestiones sociales y políticas, que según Bing Lui puede incluso llegar a ser aterrador, "ya que puede deformar las opiniones y movilizar a las masas hacia posiciones contrarias a las costumbres legales o éticas".

Estas reseñas pueden ser positivas, para promover injustamente un producto, o negativas, para desacreditar a la competencia. El autor clasifica el spam de opiniones en tres tipos principales:

- Tipo 1 (Reseñas falsas): Reseñas que deliberadamente engañan a los lectores al ofrecer opiniones positivas o negativas no merecidas sobre un producto o servicio.
- Tipo 2 (Reseñas sobre marcas únicamente): Comentarios que no se centran en un producto específico, sino en la marca o el fabricante en general, a menudo con sesgos.
- Tipo 3 (No reseñas): Textos que aparentan ser reseñas pero no contienen opiniones reales, como anuncios o contenido irrelevante.

La detección de este tipo de spam es crucial para mantener la integridad y confiabilidad de las plataformas de reseñas en línea.

---

## 2. Según Bing Liu, ¿cuáles son los componentes de una tupla de opinión?

Bing Liu define formalmente una opinión como una quintupla:

$(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$

Donde:

- $e_i$ : nombre de la entidad sobre la cual se expresa la opinión (por ejemplo, una película).
  - $a_{ij}$ : aspecto o componente específico de la entidad (por ejemplo, "cinematografía" o "diseño de producción").
  - $s_{ijkl}$ : sentimiento expresado hacia el aspecto  $a_{ij}$  de la entidad  $e_i$ . Puede ser positivo, negativo, neutro, o representarse con distintos niveles de intensidad (como 1 a 5 estrellas).
  - $h_k$ : la persona o autor de la opinión.
  - $t_l$ : el momento en el que se emite la opinión.
- 

## 3. ¿Todas las oraciones subjetivas producen al menos una opinión?

No necesariamente. Si bien las oraciones subjetivas expresan sentimientos, creencias o juicios, no todas sirven para constituir una opinión según la definición de Bing Liu.

Recordemos que para que podamos extraer una opinión de una oración, debemos poder extraer la información sobre a una entidad específica a la que se refiere y expresar un sentimiento claro hacia un aspecto de esa entidad. Por ejemplo, una oración como "Me encantó!" es subjetiva, pero no proporciona suficiente información sobre qué causa o a qué se refiere tal expresión para ser clasificada como una opinión estructurada.

---

## 4. Menciona la clasificación de emociones que hace el autor.

Recordemos que una frase objetiva presenta información objetiva sobre el mundo, mientras que una frase subjetiva expresa sentimientos, opiniones o creencias personales. Las emociones son nuestros sentimientos y pensamientos subjetivos.

Liu diferencia entre sentimientos y emociones. Mientras que el análisis de sentimientos se centra en determinar la orientación (positiva, negativa o neutral) de una opinión, el análisis de emociones busca identificar emociones específicas expresadas en el texto, como alegría, tristeza, ira, miedo, sorpresa o disgusto. Esta clasificación permite una comprensión más profunda y matizada de las opiniones expresadas por los usuarios.

---

## 5. ¿Qué es Document Sentiment Classification?

La clasificación de sentimientos a nivel de documento es una tarea que busca determinar la orientación general (positiva, negativa o neutral) de un texto completo, como una reseña o un artículo. Este enfoque considera el documento en su totalidad, sin analizar entidades o aspectos específicos dentro del texto. Es una de las tareas más estudiadas en el análisis de sentimientos y se basa en técnicas de clasificación de texto tradicionales, utilizando algoritmos de aprendizaje supervisado y características como la frecuencia de palabras.

De acuerdo a la definición formal propuesta por Liu:

Dado un documento de opinión  $d$  que evalúa una entidad determinar el sentimiento general  $s$  del opinante sobre la entidad, es decir, determinar  $s$  expresado sobre el aspecto GENERAL en la quintupla  $(\text{GENERAL}, s, e, h, t)$  donde la entidad  $e$ , el titular de la opinión  $h$  y el momento de la opinión  $t$  se suponen conocidos o irrelevantes (no importa).

---

## 6. Describe la clasificación de opiniones con base en aprendizaje no supervisado.

El aprendizaje no supervisado en la clasificación de opiniones implica identificar patrones y estructuras en los datos sin utilizar etiquetas predefinidas.

Para esta tarea se pueden aplicar técnicas como agrupamiento (clustering) para agrupar opiniones similares o el modelado de temas (topic modeling) para descubrir aspectos comunes en las opiniones. Técnicas que permiten explorar grandes volúmenes de datos y extraer insights sin la necesidad de un conjunto de datos etiquetado. Aquí es donde los embeddings de oraciones tienen mucho sentido y ayudan a estas tareas.