

---

# Resumen

La detección de opiniones falsas o verdaderas acerca de un producto o servicio, se ha convertido en un problema muy relevante de nuestra época. Según estudios recientes hasta el 80 % de las personas han cambiado su decisión final basados en las opiniones revisadas en la web. Algunas de estas opiniones pueden ser falsas positivas, con la finalidad de promover un producto, o falsas negativas para desacreditarlo.

Para ayudar a resolver este problema se propone en esta tesis un nuevo método para la detección de opiniones falsas, llamado PU-Learning modificado. Este método aumenta la precisión mediante un algoritmo iterativo y resuelve el problema de la falta de opiniones etiquetadas.

Para el funcionamiento del método propuesto se utilizan un conjunto pequeño de opiniones etiquetadas como falsas y otro conjunto grande de opiniones no etiquetadas, del cual se extraen las opiniones faltantes y así lograr una clasificación de dos clases. Este tipo de escenario se ha convertido en una situación muy común en los corpus de opiniones disponibles.

Como una segunda contribución se propone una representación basada en n-gramas de caracteres. Esta representación tiene la ventaja de capturar tanto elementos de contenido como del estilo de escritura, permitiendo con ello mejorar la efectividad del método propuesto en la detección de opiniones falsas.

La evaluación experimental del método se llevó a cabo mediante tres experimentos de clasificación de opiniones utilizando dos colecciones diferentes. Los resultados obtenidos en cada experimento permiten ver la efectividad del método propuesto así como también las diferencias entre la utilización de varios tipos de atributos.

Dado que la falsedad o veracidad de las opiniones vertidas por los usuarios, se convierte en un parámetro muy importante en la toma de decisiones, el método

que aquí se presenta, puede ser utilizado en cualquier corpus donde se tengan las características mencionadas antes.