



Fundación Universitaria Konrad Lorenz

Facultad de matemáticas e ingenierías

Taller: Creación de un Servicio de Análisis de Sentimientos para Textos Largos

Estudiantes:

Andrea Valentina Cubillos Pinto
506231711

Ivonne Nathalia Sierra Ramírez
506221064

Santiago Lozano Orrego
506221033

Profesor:

Oscar Alexander Méndez Aguirre

Asignatura:

Diseño de interfaces de usuario

Mayo, 2024

Prototipo IV

Andrea Cubillos, Ivonne Sierra y Santiago Lozano

May 30, 2024

Abstract

Este documento presenta el desarrollo de un sistema de análisis de sentimientos basado en modelos de lenguaje natural. Se utiliza la biblioteca FastAPI para crear una API web que permite a los usuarios ingresar texto y obtener el análisis de sentimiento correspondiente. Además, se implementa el procesamiento por lotes para analizar múltiples comentarios junto con su contexto. Se emplean modelos preentrenados para la clasificación de sentimientos y detección de discurso de odio en el idioma español.

Palabras clave: Análisis de sentimientos, modelos de lenguaje natural, FastAPI, API web, procesamiento por lotes, comentarios, contexto, modelos preentrenados, clasificación de sentimientos, detección de discurso de odio.

1 Introducción

El análisis de sentimientos es una tarea esencial en el procesamiento del lenguaje natural, ya que permite comprender las opiniones, actitudes y emociones expresadas en un texto. Esta tarea es crucial para conocer las percepciones del público, obtener retroalimentación de los clientes y monitorear el ambiente general en las redes sociales. En este proyecto, se aborda esta necesidad a través del desarrollo de un sistema avanzado de análisis de sentimientos en español.

Este sistema no solo analiza el sentimiento de comentarios individuales, sino que también tiene la capacidad de procesar lotes completos de comentarios de manera simultánea. Este enfoque de análisis por lotes permite obtener una visión más amplia y una comprensión más profunda de las tendencias y patrones presentes en los datos. Además, el sistema está equipado para detectar discursos de odio dentro de un contexto específico, lo que agrega una capa adicional de análisis crítico y relevante.

La implementación de este sistema se realiza utilizando FastAPI, una biblioteca de Python conocida por su rapidez y eficiencia en la creación de APIs web. Esto garantiza una interfaz de usuario fluida y receptiva, lo que facilita la interacción con el sistema, tanto para usuarios individuales como para aplicaciones que manejan grandes volúmenes de datos.

En este proyecto, se aborda esta necesidad a través del desarrollo de un sistema integral avanzado de análisis de sentimientos en español utilizando la biblioteca pysentimiento [4] y modelos preentrenados como RoBERTuito [3] en datos de redes sociales para la detección de discursos de odio en español, mediante un sistema avanzado y eficiente que aprovecha las últimas tecnologías en procesamiento del lenguaje natural y desarrollo web.

2 Resultados

El sistema desarrollado no solo ofrece una interfaz intuitiva para ingresar texto y obtener análisis de sentimiento de manera individual, sino que también proporciona una funcionalidad robusta para el análisis de lotes de comentarios junto con su contexto.

Para los comentarios individuales que, se desarrollaron siguiendo los lineamientos del tutorial de Nechu [1], el sistema muestra no solo el sentimiento asociado, sino también la puntuación correspondiente, lo que proporciona una comprensión más detallada y precisa de la evaluación del sentimiento en cada comentario.

En el caso del análisis de lotes de comentarios, el sistema procesa múltiples comentarios simultáneamente, lo que permite una visión más amplia de las tendencias y patrones en los datos. Además del análisis de sentimientos, esta función también proporciona etiquetas y probabilidades de discurso de odio para cada comentario dentro del contexto específico en el que se presentan, utilizando los conjuntos de datos y métricas de la tarea TASS de detección de emociones [2]. Esto añade una capa adicional de análisis crítico y relevante, lo que permite a los usuarios identificar áreas de preocupación o interés específicas dentro del conjunto de datos analizado.

En conjunto, estos resultados muestran el potencial del sistema desarrollado para ofrecer una comprensión profunda y completa de las opiniones expresadas en texto en español, tanto a nivel individual como a nivel de grupos de comentarios. Esta capacidad es invaluable para una amplia gama de aplicaciones, incluidas la investigación académica, la gestión de la reputación en línea, la toma de decisiones empresariales informadas, entre otras.

3 Conclusiones

El sistema de análisis de sentimientos desarrollado demuestra ser una herramienta útil para comprender las actitudes y opiniones expresadas en texto en español. La integración de modelos preentrenados facilita la clasificación precisa de sentimientos y la detección de discurso de odio. Asimismo, se identifican áreas de mejora potencial, como la optimización del procesamiento por lotes para manejar grandes volúmenes de comentarios de manera más eficiente.

En general, el sistema proporciona una base sólida para futuras investigaciones y aplicaciones en el campo del procesamiento de lenguaje natural.

References

- [1] Daniel Benzaquen. 3 - librería transformers — domina la magia de hugging face. Video, September 2023.
- [2] M. García-Vega, M. C. Díaz-Galiano, M. A. García-Cumbreras, F. M. P. Del Arco, A. Montejoráez, S. M. Jiménez-Zafra, E. Martínez Cámara, C. A. Aguilar, M. A. S. Cabezudo, L. Chiruzzo, and otros. Overview of tass 2020: Introducing emotion detection. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) Co-Located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020)*, pages 163–170, Málaga, Spain, 2020.
- [3] J. M. Pérez, D. A. Furman, L. Alonso Alemany, and F. M. Luque. Robertuito: a pre-trained language model for social media text in spanish. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 7235–7243. European Language Resources Association, 2022.
- [4] J. M. Pérez, M. Rajngewerc, J. C. Giudici, D. A. Furman, F. Luque, L. Alonso Alemany, and M. V. Martínez. pysentimiento: a python toolkit for opinion mining and social nlp tasks. *arXiv preprint*, 2021.