

# Evaluación de efectividad de monitores de actividad física mediante el análisis de reseñas en tiendas en línea

## PROBLEMA

Los monitores de actividad física (MAF) son dispositivos cuya popularidad está en aumento como instrumento para mejorar la salud y el estilo de vida de una persona. Los usuarios a menudo usan las descripciones provistas por los fabricantes respecto a las características funcionales del dispositivo al comprarlo, que prometen efectos positivos; sin embargo, no hay resultados significativos que comprueben la veracidad de estas afirmaciones.

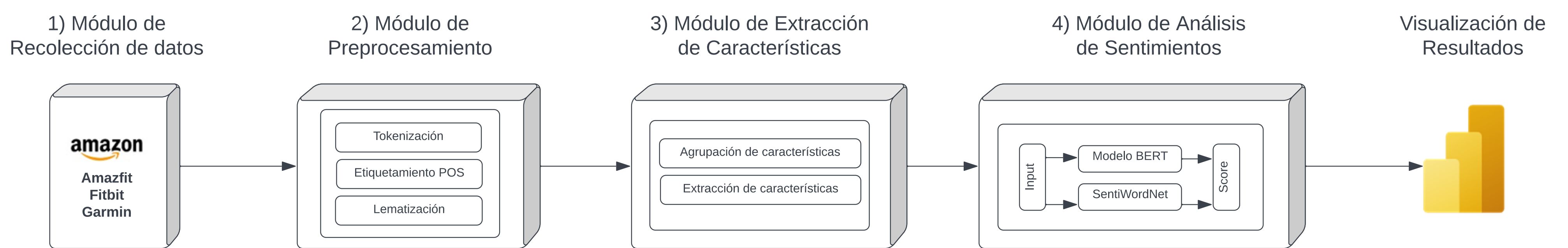
## OBJETIVO GENERAL

Crear un *framework* de evaluación de monitores de actividad física, mediante el análisis de reseñas publicadas en tiendas en línea, para medir la efectividad de estos productos.



## PROPUESTA

- Se utilizó un *dataset* de reseñas extraídas de Amazon de tres dispositivos, Fitbit, Amazfit y Garmin, con un total de 9716 reseñas. Estas reseñas corresponden a tres países de habla inglesa: Estados Unidos, Canadá y Reino Unido.
- Se diseñó un *framework* que consistió en cuatro módulos: 1) recolección de datos, 2) preprocesamiento, 3) extracción de características y 4) análisis de sentimientos. Para el módulo de extracción de características se realizó un análisis a nivel de cláusulas.
- Para el módulo de análisis de sentimientos se experimentó con SentiWordNet y el modelo BERT para obtener un *score* que represente la polaridad de la opinión de un usuario sobre una característica de un monitor de actividad física. Posteriormente, se clasificó este *score* en tres categorías para etiquetar el sentimiento: Positivo, Negativo y Neutral.
- Tanto el módulo de extracción de características como el módulo de análisis de sentimientos se dividieron en dos fases: una para los fabricantes y otra para las reseñas de los usuarios.



## RESULTADOS

- El modelo que presenta mejores resultados en 3 de las 4 métricas de evaluación utilizadas para el análisis de sentimientos en las reseñas en tiendas en línea fue BERT.
- Se identificaron a aproximadamente 16 mil cláusulas con una característica de un dispositivo.
- El 75% de las cláusulas obtenidas tenían una longitud de aproximadamente 20 tokens

Métricas	BERT	SentiWordNet
<b>Accuracy</b>	0.7154	0.6466
<b>Recall</b>	0.6021	0.6731
<b>Precision</b>	0.8092	0.5559
<b>F1 Score</b>	0.6905	0.6089

User and Manufacturer Score per Aspect

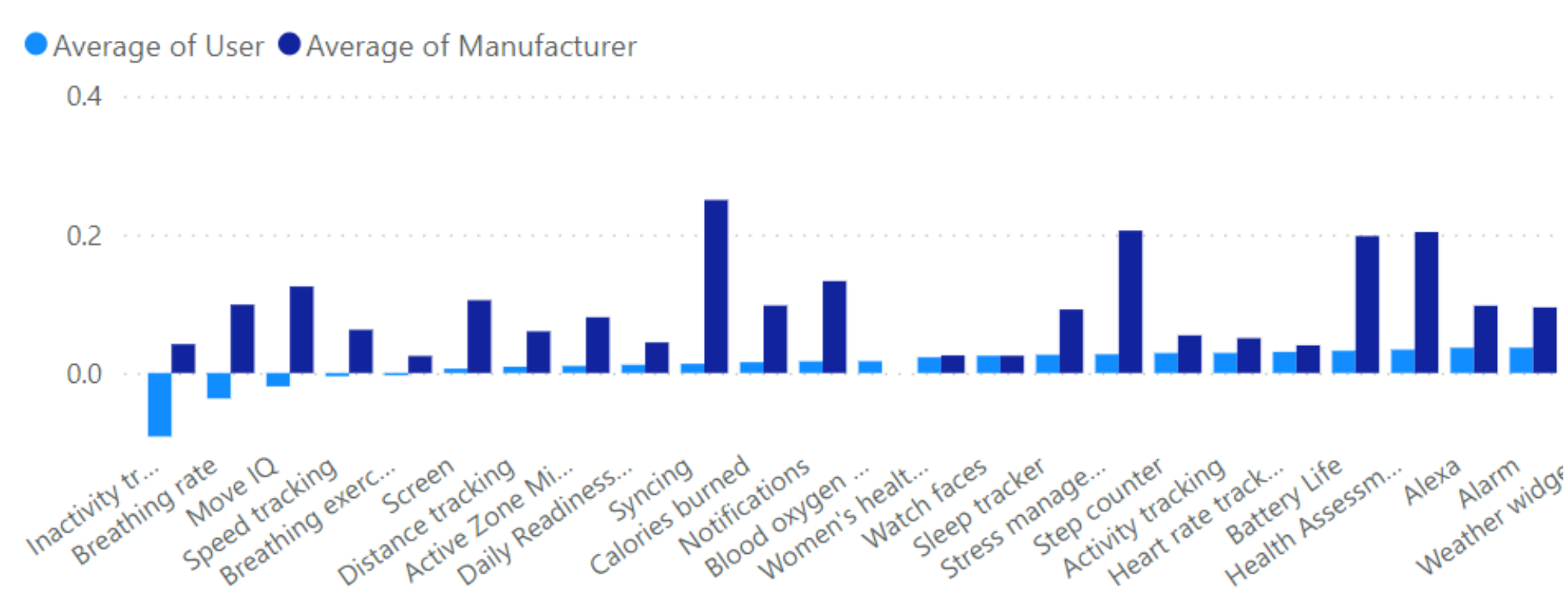


Figure 2: Average Score Difference between User and Manufacturer per Aspect

## CONCLUSIONES

- El uso del modelo BERT permite obtener un *score* más ajustado a los valores reales para clasificar los sentimientos de las cláusulas, con 81% de precisión.
- La efectividad de los monitores de actividad física es sobrestimada por parte de los fabricantes. Los usuarios perciben que el rendimiento de las características ofrecidas en los dispositivos es menor al publicitado.
- La metodología aplicada en el presente proyecto puede ser extrapolada para la evaluación de otras clases de dispositivos en tiendas en línea.
- El presente proyecto provee a los consumidores de MAF otra herramienta para escoger su próxima compra y a los fabricantes para identificar rápidamente la opinión de los usuarios de las características de sus dispositivos.

