

Visual Analytics de Emociones Geolocalizadas en Twitter: Un Análisis Espaciotemporal del Bienestar Instantáneo

Jayan Michael Caceres Cuba
Universidad Nacional de San Agustín
Arequipa, Peru
jcacerescu@unsa.edu.pe

Abstract—En la última década, Twitter se ha consolidado como una fuente clave para analizar el comportamiento emocional de las personas frente a eventos de alto impacto. Sin embargo, la mayoría de los estudios centrados en el análisis de sentimientos geolocalizados en esta red social abordan fenómenos particulares y carecen de una visión longitudinal que permita comprender cómo evolucionan las emociones a lo largo del tiempo y el espacio. En este trabajo se presenta una herramienta de visual analytics que permite explorar interactivamente emociones expresadas en Twitter, con un enfoque geoespacial y temporal. Utilizando el Twitter Sentiment Geographical Index Dataset (TSGI), se desarrolló una plataforma que integra visualización de datos espaciales, líneas de tiempo y mapas de calor para identificar patrones emocionales regionales y su evolución durante varios años. Esta solución busca facilitar el análisis del bienestar instantáneo de la población, proporcionando una base visual sólida para estudios sociales, geográficos y de salud pública. **Keywords:** bienestar instantáneo, Twitter, análisis, redes sociales, geoespacial

I. INTRODUCCIÓN

En los últimos años, las redes sociales—y en particular Twitter (X)—se han consolidado como fuentes esenciales de información en tiempo real, ya que las personas comparten noticias o reacciones sobre un evento o tema mediante breves mensajes de texto llamados «tuits». El análisis de emociones en estos tuits resulta fundamental para comprender los sentimientos de la población frente a sucesos de alto impacto. En este sentido, el Twitter Sentiment Geographical Index Dataset [4] ofrece una cobertura diaria y detallada, tanto en términos de sentimientos (positivo, negativo, neutral) como de geolocalización, lo que permite capturar el bienestar instantáneo de las personas—aquel estado emocional momentáneo que expresan en sus tuits sobre eventos específicos. A diferencia del bienestar subjetivo—que evalúa la satisfacción global con la vida a largo plazo—el bienestar instantáneo se centra en las emociones que afloran en un instante dado y está directamente vinculado a reacciones inmediatas ante sucesos como epidemias o desastres naturales.

El análisis de emociones en Twitter con enfoque geoespacial ha ganado relevancia en diversas disciplinas, permitiendo comprender cómo las emociones colectivas varían según la ubicación y el tiempo. Diversos estudios han explorado esta intersección, desarrollando herramientas y metodologías para visualizar y analizar estos datos complejos. Por ejemplo, el sistema CovidEmoVis[?] ofrece una herramienta interactiva de

análisis visual que permite explorar emociones expresadas en Twitter durante la pandemia de COVID-19. [?]En el ámbito del turismo, un estudio aplicó análisis de redes sociales y geoespaciales para visualizar las emociones de los visitantes en un parque temático, utilizando el Modelo Circunflejo del Afecto de Russell para categorizar las emociones y mapearlas geográficamente dentro del parque. Este enfoque permitió identificar áreas específicas que evocaban emociones particulares, proporcionando información valiosa para la gestión de la experiencia del visitante. Otro ejemplo es el sistema SensePlace2, [?]que integra análisis geoespacial y de contenido para apoyar la gestión de crisis mediante la visualización de tweets relevantes en mapas interactivos, permitiendo filtrar información por ubicación, tiempo y atributos específicos.

Estos estudios demuestran el potencial de combinar análisis de sentimientos, visualización de datos y componentes geoespaciales para comprender las dinámicas emocionales en diversos contextos. Sin embargo, la mayoría se centran en eventos específicos o periodos limitados, lo que resalta la necesidad de investigaciones que aborden análisis más generales y longitudinales de las emociones geolocalizadas en Twitter.

II. PROBLEMA

La falta de una solución que permita explorar de manera visual y dinámica la evolución de los sentimientos expresados en Twitter a lo largo del tiempo y el espacio, brindando una visión más profunda del bienestar instantáneo de las poblaciones en distintos contextos geográficos.

III. OBJETIVOS

Desarrollar una herramienta de visual analytics que permita explorar interactivamente las emociones geolocalizadas expresadas en Twitter a lo largo del tiempo, utilizando el dataset TSGI, con el fin de identificar patrones espaciales y temporales de bienestar instantáneo.

IV. DESCRIPCIÓN DEL DATASET

A. Contexto

El *Twitter Sentiment Geographical Index Dataset (TSGI)* [4] se basa en el *Geotweet Archive v2.0* del Harvard Center for Geographic Analysis, que agrupa miles de millones de

tuits con metadatos de geolocalización (coordenadas o *bounding box* de lugar). TSGI focaliza la recolección a partir de 2019 y clasifica cada tuit con un modelo BERT multilingüe, asignándole una probabilidad de “sentimiento positivo”. Posteriormente, se agregan diariamente esas probabilidades por las unidades administrativas definidas en GADM v3.6 (país, estado/provincia, condado/ciudad), generando índices diarios que sirven como proxy de bienestar subjetivo.

B. Objeto o Entidad de Estudio

El objeto de estudio son los *índices diarios de sentimiento* de Twitter en distintas divisiones administrativas georreferenciadas. Cada registro representa la combinación de:

- Una **fecha** específica (variable temporal).
- Una **unidad administrativa** en tres niveles (país, provincia/departamento, ciudad/condado).
- Dos métricas agregadas calculadas a partir de todos los tuits georreferenciados que caen dentro de esa unidad el día en cuestión:
 - SCORE: Valor promedio de probabilidad de sentimiento positivo ($\text{float} \in [0, 1]$).
 - N: Número de tuits georreferenciados usados para el cálculo (entero ≥ 0).

C. Atributos del Dataset

A continuación se describen los principales atributos del archivo CSV agregado por año (por ejemplo, `num_posts_and_sentiment_summary_2023.csv`):

- **DATE**
 - Descripción: Fecha del índice.
 - Tipo: YYYY-MM-DD (cadena de caracteres, *string*).
 - Rango de valores: Desde 2019-01-01 hasta la fecha más reciente disponible.
- **NAME_0**
 - Descripción: Nombre del país (nivel ADMIN 0).
 - Tipo: Cadena de caracteres (*string*).
 - Ejemplos: “Peru”, “United States”, “Spain”.
- **NAME_1**
 - Descripción: Nombre de la subdivisión de nivel 1 (estado/provincia/región según GADM).
 - Tipo: Cadena de caracteres (*string*).
 - Ejemplos (Perú): “Lima”, “Cusco”, “Arequipa”.
- **NAME_2**
 - Descripción: Nombre de la subdivisión de nivel 2 (ciudad/condado/provincia secundaria).
 - Tipo: Cadena de caracteres (*string*).
 - Ejemplos (Perú): “Lima Province”, “Cusco Province”, “Arequipa Province”.
 - Notas: Puede estar vacía si no existen datos georreferenciados a nivel 2 en esa fecha/ubicación.
- **SCORE**
 - Descripción: Valor promedio diario de probabilidad de “sentimiento positivo” para todos los tuits georreferenciados dentro de la unidad administrativa y fecha especificadas.

- Tipo: Número de punto flotante (*float*).
- Rango de valores: $[0.0, 1.0]$ (valores más cercanos a 1 indican mayor proporción de tuits con sentimiento positivo).

• N

- Descripción: Número total de tuits georreferenciados (originales) utilizados para calcular el SCORE en esa fecha y unidad administrativa.
- Tipo: Entero (*int*).
- Rango de valores: $[0, +\infty)$ (un valor de 0 indica que no hubo tuits, en cuyo caso puede omitirse la fila en algunos niveles).

D. Cuadro Resumen de Atributos

TABLE I
RESUMEN DE ATRIBUTOS DEL TSGI (POR REGISTRO DIARIO).

Atributo	Descripción	Tipo	Rango / Ejemplos
DATE	Fecha del índice	string	“2023-02-15”, “202
NAME_0	Nombre del país (ADMIN 0)	string	“Peru”, “United Sta
NAME_1	Nombre del estado/provincia (ADMIN 1)	string	“Lima”, “Cusco”, “
NAME_2	Nombre de la ciudad/condado (ADMIN 2)	string	“Lima Province”, “
SCORE	Promedio diario de sentimiento positivo	float	$[0.0, 1.0]$
N	Número de tuits utilizados	int	$[0, +\infty)$

AGRADECIMIENTOS

Se agradece al Harvard Center for Geographic Analysis por proveer el *Geotweet Archive* y al equipo de Panacea Lab por publicar el *Twitter Sentiment Geographical Index Dataset* de acceso libre.

REFERENCES

- [1] M. Hammad, M. K. H. Rashed, F. Al-Turjman, *CovidEmoVis: An interactive visual analytics system for tracking emotions in COVID-19 tweets*, in: *Advances in Intelligent Systems and Computing*, vol. 1261, pp. 123–133, 2021. DOI:10.1007/978-3-030-66919-5_10.
- [2] L. Zhao, Y. Sun, H. Wang, J. Sun, *Understanding emotions in location-based social networks: A study of tourist emotions in a theme park*, *Tourism Management*, vol. 81, 2020, Art. no. 104162. DOI:10.1016/j.tourman.2020.104162.
- [3] A. M. MacEachren, A. C. Robinson, A. Jaiswal, S. Pezanowski, J. Savelyev, P. Mitra, X. Zhang, M. Blanford, *SensePlace2: Geotwitter analytics support for situational awareness*, in: *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 181–190, 2011. DOI:10.1109/VAST.2011.6102456.
- [4] Y. Chai, D. Kakkar, J. Palacios, S. Zheng, *Twitter Sentiment Geographical Index Dataset*, *Scientific Data*, vol. 10, Art. no. 572, 2023. DOI:10.1038/s41597-023-02572-7.