

**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS  
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA**



**TALLER DE APLICACIONES SOCIALES**

**Resumen Artículo: Wisdom of crowds detects COVID-19 severity ahead of  
officially available data**

**PRESENTADO POR:**

**GRUPO 5**

- Lizana Yauri Alexander
- Ginocchio Salcedo Gino
- Muñoz Silva Brayan

**DOCENTE:**

Yury Limachi Cartolin

**LIMA– PERÚ**

**2023**

## **Introducción**

En la introducción del artículo, se nos menciona que el conocimiento colectivo de una multitud logró ser utilizado con éxito en escenarios de pronóstico desafiantes en ciencias sociales y de datos. Asimismo, se estableció que las opiniones colectivas formadas por un grupo de individuos a veces pueden ser más precisas que las opiniones de expertos individuales, a este fenómeno se le denominó "la sabiduría de las multitudes".

Se explica también que en lugar de utilizar consultas de motores de búsqueda web que solo indican patrones de búsqueda de conocimiento, los investigadores se enfocaron en evaluar la sabiduría de las multitudes representada en la plataforma de redes sociales Twitter. Para ello, se recopiló la reacción en las redes sociales, expresada por la intensidad de los tweets geolocalizados relacionados con COVID-19, en las regiones geopolíticas de Italia, España y Estados Unidos al comienzo de la pandemia. Luego, investigaron las relaciones con los datos de mortalidad regional después de un mes.

El objetivo principal del estudio con esto fue determinar si la intensidad de la atención inicial en Twitter relacionada con COVID-19 se correlaciona con la gravedad de la pandemia en esas regiones después de un mes. Los resultados revelaron una relación significativa entre la intensidad de los tweets relacionados con COVID-19 y el impacto de la pandemia en esas regiones después de un mes.

Estos hallazgos podrían respaldar la creación de sistemas de detección de novedades en tiempo real, con el objetivo de informar tempranamente sobre la gravedad de las crisis que afectan a un territorio y activar medidas de control en una etapa en la que los datos disponibles son extremadamente limitados.

## **Métodos y Materiales**

### **Recopilación de la Data**

Los investigadores utilizaron fuentes oficiales y confiables para obtener datos sobre la propagación y las víctimas de COVID-19 en cada región, así como datos de población por región, por ejemplo: Para Italia, se obtuvieron datos de

población por región del Istituto Nazionale di Statistica (ISTAT), actualizados hasta 2019. Para España, se obtuvieron datos de población por región del Instituto Nacional de Estadística (INE), también actualizados hasta 2019. Para Estados Unidos, se utilizaron datos de población por región del United States Census Bureau. Esto les permitió realizar un análisis comparativo entre la intensidad de los tweets relacionados con COVID-19 y la gravedad de la pandemia en cada región.

### **Preprocesamiento de la Data**

Los investigadores utilizaron un repositorio de datos de Twitter para obtener información sobre los tweets relacionados con COVID-19. Utilizaron técnicas de rastreo para recopilar los datos y luego los procesaron para su posterior análisis. Esto les permitió evaluar la intensidad de los tweets relacionados con COVID-19 en cada región y su relación con la gravedad de la pandemia.

### **Identificación de los picos de intensidad**

Los investigadores utilizaron el puntaje Z y un umbral establecido de un Z mayor a 3 para identificar los picos de intensidad de los tweets relacionados con COVID-19. Esto les permitió determinar el momento en que comenzó la reacción en las redes sociales a la epidemia en cada país.

### **Análisis de Regresión**

Los investigadores utilizaron tres modelos específico de regresión para analizar los datos: el modelo 1, que comparaba la intensidad ajustada de los tweets con los casos de muerte registrados; el modelo 2, que comparaba el logaritmo de la población con los casos de muerte registrados; y el modelo 3, que combinaba la intensidad ajustada de los tweets y el logaritmo de la población como variables predictoras de los casos de muerte. Asimismo también se realizaron análisis ponderados y no ponderados, y se presentaron los coeficientes de regresión y los valores de R-cuadrado para evaluar la significancia y la calidad de ajuste de los modelos.

## **Análisis de Correlación**

Los investigadores utilizaron el coeficiente de correlación de rango de Spearman para evaluar la relación entre la intensidad de los tweets y los casos de muerte. Utilizaron valores de correlación de hipótesis nula para evaluar la significancia de la correlación encontrada y determinar si había una correlación significativa entre las variables.

## **Resultados**

Los investigadores encontraron una correlación significativa entre la intensidad de los tweets relacionados con COVID-19 y los casos de muerte en las regiones de Italia, España y Estados Unidos. Sugieren que la intensidad de los tweets puede ser un indicador útil de la percepción de la gravedad de la pandemia por parte de la población y que la información de la sabiduría de la multitud puede ser más efectiva en las primeras etapas de una pandemia emergente. También reconocen que su metodología tiene limitaciones y que se necesitan más datos y modelos más sofisticados para desarrollar herramientas de monitoreo y pronóstico más precisas en el futuro. Además, sugieren que se requiere recopilar datos regionales adicionales para otros países con el fin de ajustar de manera más precisa la intensidad de los tweets.

## **Discusión**

De acuerdo a los resultados mostrados, se plantea que los trabajos futuros deben centrarse en refinar los modelos e incorporar fuentes de datos adicionales para mejorar la precisión y confiabilidad de los indicadores de percepción de la multitud. Los datos regionales de otros países también serían valiosos para validar y ajustar aún más la medida de intensidad de los tweets.

## **Conclusión**

El estudio resalta el potencial de la actividad en Twitter como una fuente de información valiosa para comprender la percepción y respuesta del público ante una pandemia. Al aprovechar la sabiduría de la multitud, se puede obtener información que complementa los datos oficiales y mejorar la capacidad para monitorear y responder ante una crisis de salud pública.