



Revisión de artículo

TWEETS GEOETIQUETADOS COMO PREDICTORES DE RESULTADOS EN SALUD

Traducción adaptada del artículo "Geotagged US Tweets as Predictors of County-Level Health Outcomes", 2015–2016, publicado en *American Journal of Public Health* (Am J Public Health. 2017; 107:1776–1782. doi:10.2105/AJPH.2017.303993).

Jenny A. Pinzón****

Hemos realizado este resumen por considerar interesante el uso que se está dando en investigación a la información disponible en las redes sociales, particularmente en el campo de la Salud Pública.

El objetivo del estudio

Crear indicadores de modelos sociales de percepciones y comportamientos en torno a la salud usando datos obtenidos de Twitter y buscar asociaciones con las estadísticas disponibles de los condados contiguos de Estados Unidos.

La ubicación geográfica como determinante de la salud



El lugar donde vivimos, incluyendo los aspectos sociales, políticos, económicos y el entorno construido (el estatus socioeconómico, las políticas de salud pública y el acceso al cuidado de la salud), tienen gran impacto en la salud, generando, incluso, inequidades en la población. Este hecho

lo mencionan los autores del artículo refiriéndose a estudios que relacionan la calidad de vida con la enfermedad coronaria y la obesidad, citando como ejemplo la asociación entre el incremento de peso en los individuos y la densidad de restaurantes de comida rápida en la zona donde viven.

**** Coeditora Carta Comunitaria, jenny-pinzon@juanncorpas.edu.co.



Las redes sociales y la salud



http://adobe.com.mx/wp-content/uploads/2017/03/Untitled-design-4_opt.png

Además, afirman que los procesos sociales que se gestan en las redes sociales pueden influir en la salud dado que estas pueden impulsar en los usuarios nuevos intereses y hábitos de vida que varían según la ubicación de las personas; de ahí que el hacer uso de los datos públicos de dichas redes puede ayudar a entender las diferencias que se presentan y, de esa forma, realizar actividades de vigilancia en Salud Pública.

Estos procesos se facilitan por el amplio uso de Internet y las opiniones abiertas que se comparten con registros georreferenciados que permiten a los investigadores acceder a ellas y entender interacciones que, usadas en forma efectiva, pueden ayudar a descubrir patrones y problemas de salud emergentes.

Las redes sociales y las investigaciones en Salud Pública



https://i6-img-fp.akamaized.net/iconos-gratis/gorjee_318-136405.jpg?size=338&ext=.jpg

Las redes sociales han sido usadas para hacer seguimiento a enfermedades transmitidas por alimentos y para obtener información útil para la detección, seguimiento y predicción de brotes de enfermedades, así como para la localización en tiempo real de desastres y responder ante estos.

El artículo menciona que, a través de Twitter, una de las redes sociales más populares, se han obtenido percepciones sobre temas de salud. Por ejemplo, citan un estudio en el que

analizaron los *tweets* para examinar los sentires relacionados con el hábito de fumar, hallándose más opiniones positivas para el narguile y los cigarrillos electrónicos que para el tabaco tradicional. Asimismo, se ha encontrado que ser seguidor de páginas de redes sociales con *marketing* de alcohol se encuentra asociado a su consumo temprano y mayor entre los jóvenes.

Hipótesis del estudio

La hipótesis que plantean los autores es que las comunidades que son más felices, que promueven más activamente la comida saludable y la actividad física y que exhiben menos comportamientos relacionados con el alcohol en las redes

sociales, tendrán menor mortalidad, menor obesidad y menos problemas relacionados con el uso del alcohol, así como mayores niveles de actividad física.



https://t2.wp.com/t2.esmas.com/opinion/img/2011/redes_sociales_adictos.jpg

Con base en su hipótesis, los autores crearon indicadores de percepción comunitaria y de modelos sociales de dieta, actividad física y consumo de alcohol. Luego probaron esos factores del contexto cultural como predictores de resultados en salud. Aludiendo a la teoría del aprendizaje como un

proceso cognitivo que ocurre en un contexto social, concluyen que los puntos de vista y las actividades descritas en las redes sociales pueden contribuir a dar forma a costumbres, actitudes y creencias y, como consecuencia, al comportamiento de las personas.

Cómo obtuvieron los datos



https://www.rftmanmead.com/blog/content/images/2016/08/11Oracle_Stream_Analytics-7-e1469218770977.png

Usaron los datos de Twitter para conocer las actitudes, las costumbres y las conductas de la comunidad. Con el objetivo de compartir el análisis de los datos, crearon una aplicación de mapeo interactiva basada en la web mediante el uso de tecnología de código abierto que permite al público la exploración de los datos agregados por condado.

La recolección de los datos se hizo a través de la interfaz de programación de aplicaciones de transmisión de Twitter; de esa forma se obtuvieron, en forma aleatoria, aproximadamente 80 millones de tweets geoetiquetados (datos que contienen coordenadas geográficas habilitadas

por los usuarios permitiendo así ubicar el lugar donde se originan) de 603 363 usuarios en estados contiguos de Estados Unidos (excluyendo a Alaska y a Hawái) en un período de 12 meses (de abril de 2015 a marzo 2016). Estos datos estaban disponibles públicamente.

La identificación única de los tweets permitió sacar los datos duplicados; además, se eliminaron las ofertas de trabajo y las cuentas de publicidad (aproximadamente el 1 % de los tweets). Para consolidar los datos se usó un lenguaje de programación especial.

El análisis de los datos

Cómo analizaron los datos referentes a los sentimientos y las percepciones



Para el análisis de los sentimientos se usó un paquete basado en el lenguaje de programación Java creado para el procesamiento del lenguaje natural. Este paquete convierte el texto en "características" que se pueden analizar mediante la aplicación de complejos algoritmos. De acuerdo a esas características, se logró estimar la probabilidad de que un

tweet expresara sentimientos de felicidad. Para lograr lo anterior, se realizó una prueba piloto donde se evaluó el desempeño del programa comparado con la clasificación manual que un ser humano haría de los *tweets* alcanzando prevalencias similares.

Cómo analizaron los datos referentes a la alimentación y la actividad física



Los investigadores crearon una lista de 1430 alimentos populares para seguir la frecuencia de sus menciones en la red. A cada alimento le asociaron una medida de densidad calórica (calorías por cada 100 gramos de acuerdo a la base de datos del Departamento de Agricultura de EE. UU.). Se etiquetaron las frutas, los vegetales, los frutos secos y las proteínas magras, por ejemplo, el pescado, el pollo y el pavo, como "alimentos sanos", excluyendo los fritos. Las menciones al alcohol se rastrearán usando 66 términos que incluían bebidas alcohólicas populares y tipos de licor (ejemplo, vino, cerveza, etc.).

Para rastrear los *tweets* referidos a actividad física, elaboraron una lista de 376 actividades tomadas de cuestionarios, compendios y programas de entrenamiento populares. Dentro de estas actividades incluyeron ejercicios en el gimnasio y de recreación (como caminatas), y tareas del hogar, excluyendo frases comunes que no están relacionadas con la actividad física, aunque indicaran movimiento (por ejemplo, "marcharse"). Para los deportes en equipo, diferenciaron entre "jugar" y mirar un juego.







Cómo controlaron la exactitud en la interpretación de los tweets







http://images.clipartpanda.com/accuracy-clipart-arrow_target_accuracy_md_wm.jpg

Para controlar la calidad con la que el programa caracterizaba los tweets, etiquetaron manualmente 5000 tweets sobre comida y actividad física, así:

-  Tweets relacionados con alimentos: 2000.
-  Tweets no relacionados con alimentos: 500.
-  Tweets relacionados con actividad física: 2000.
-  Tweets no relacionados con actividad física: 500.

Al compararlos con la categorización que realizó el programa se encontró que:

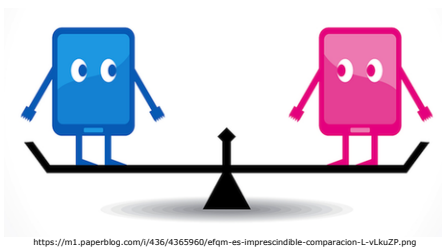
-  83 % de los tweets relacionados con alimentos se etiquetaron con precisión.
-  81 % de los tweets no relacionados con alimentos se etiquetaron con precisión.
-  82 % de los tweets relacionados con actividad física se etiquetaron con precisión.
-  97 % de los tweets no relacionados con actividad física se etiquetaron con precisión.

En general, los errores encontrados se relacionaban con el uso de lenguaje figurado como, por ejemplo, "llegar tarde" ("running late") o, "dulce como la miel" ("sweet as honey") o con referencia a ver un deporte en lugar de practicarlo.

Para evaluar la categorización de los tweets relativos al estado de ánimo, se seleccionaron al azar 500 tweets (50 % etiquetados como felices y 50 % como no felices, de acuerdo






con el algoritmo creado para el programa). Luego, a través de una plataforma de Amazon (*Amazon Mechanical Turk*), se realizaron 20 encuestas, cada una con 25 tweets ordenados en forma aleatoria, para que los participantes los clasificaran, encontrándose, en general, una precisión del 78 % en comparación con la categorización realizada por el programa.

Con qué compararon los resultados





<https://m1.paperblog.com/v/436/4365960/efqm-es-imprescindible-comparacion-l-vi.kuZP.png>


Los datos obtenidos de los tweets se agruparon por condados con el fin de compararlos con las estadísticas disponibles en fuentes externas (tasas ajustadas por edad de la población estándar de EE. UU. en el año 2000). Estas fuentes incluían los siguientes datos:

-  Mortalidad prematura, definida para el estudio como tasas X 100 000 muertes ocurridas antes de los 75 años.
-  Muertes provocadas por conducir bajo la influencia del alcohol ocurridas de 2010 a 2014.
-  Características demográficas: promedio de edad, porcentaje de blancos no hispanos e ingreso medio familiar.
-  Condiciones crónicas y comportamientos en torno a la salud (datos de 2011 a 2014).
-  Obesidad en adultos (población de 20 años en adelante), con un índice de masa corporal de 30 kg/m² o mayor.

Además, se definieron las siguientes variables:

 **Inactividad física:** porcentaje de personas de 20 años en adelante que no informaron realizar actividad física en su tiempo libre en el último mes.

 **Consumo excesivo de alcohol:** porcentaje de adultos que bebieron más de 1 bebida (mujeres) o 2 bebidas (hombres) por día en promedio.

 **Ebriedad:** porcentaje de mujeres que consumieron más de cuatro bebidas alcohólicas en una sola ocasión en los últimos 30 días o, cinco, en el caso de los hombres.

Enfoque del análisis

Los datos obtenidos de Twitter se categorizaron en terciles, correspondientes a los niveles alto, moderado y bajo (categoría de referencia). En los modelos de regresión lineal ajustados, se utilizaron los indicadores de salud derivados de Twitter para predecir los resultados en más de 3135 condados.

Para cada resultado se ejecutó un modelo diferente y el tamaño de la muestra varió debido a la falta de resultados o de variables predictoras. La mediana del número de *tweets* para las estimaciones por condado fue de 2530.


La significancia estadística se estableció en $p > 0,05$. La información se procesó con el programa *Stata*.


Resultados


TABLE 1—Descriptive Statistics, County Level: Contiguous United States, April 2015–March 2016


County-Level Characteristics	No. of Tweets	No. of Counties	Mean \pm SD
Happiness ^a : % of tweets that are happy	79 848 992	3 135	18.54 \pm 6.29
Food culture			
Calories density of food tweets (cal/100 g)	4 041 521	3 058	238.23 \pm 65.12
% of tweets about food	4 041 521	3 058	3.85 \pm 2.40
% of tweets about healthy foods	4 041 521	2 900	0.78 \pm 0.84
% of tweets about fast food	4 041 521	2 387	0.33 \pm 0.27
Sentiment of food tweets, % happy	4 041 521	3 058	25.03 \pm 12.00
Sentiment of healthy food tweets, % happy	644 489	2 900	24.64 \pm 18.25
Sentiment of fast-food tweets, % happy	373 449	2 387	16.56 \pm 19.09
Physical activity culture			
% of tweets about physical activity	1 473 984	3 055	2.08 \pm 2.09
Sentiment of physical activity tweets, % happy	1 473 976	3 055	25.63 \pm 14.13
Substance use			
% tweets about alcohol	687 496	2 769	0.68 \pm 0.77
% tweets about drugs	687 496	1 779	0.08 \pm 0.10
% tweets about smoking	687 496	998	0.07 \pm 1.06
Sentiment of alcohol tweets, % happy	638 347	2 770	27.65 \pm 21.03
County health outcomes^b			
Premature mortality, ^c per 100 000	...	2 989	8 025.59 \pm 2 409.21
% obesity	...	3 142	30.73 \pm 4.41
% diabetes	...	3 220	9.70 \pm 2.19
% leisure-time physical inactivity	...	3 142	25.58 \pm 4.93
% binge or heavy drinking ^d	...	3 140	16.63 \pm 3.36
% driving deaths with alcohol involvement	...	3 118	31.36 \pm 15.91


La tabla anterior muestra los siguientes resultados en 3135 condados:

 La prevalencia promedio de *tweets* felices fue de alrededor del 19 % en promedio (de 80 millones de *tweets*).

 En promedio, el 4 % de los *tweets* mencionaron los alimentos (de 4 millones de *tweets*). Entre estos, la densidad calórica promedio de los alimentos mencionados fue de aproximadamente 240 calorías por cada 100 gramos.

 Los *tweets* sobre alimentos saludables contenían más expresiones de felicidad que los relacionados con la comida rápida (25 % frente a 17 %).

 En promedio, el 2 % de los *tweets* se referían a la actividad física y menos del 1 % de los *tweets* mencionaban el consumo de alcohol.

 Los *tweets* sobre el consumo de alcohol expresaron ligeramente más felicidad que los relacionados con la actividad física o los alimentos saludables.



La tabla también muestra las siguientes estadísticas descriptivas de fuentes externas:

- Tasa de mortalidad prematura promedio: aproximadamente 8000 X 100 000.
- Promedio de obesidad: 31 %.
- Promedio de adultos físicamente inactivos: un poco más del 25 %.

- Consumo excesivo de alcohol: se acercó al 17 %.
- Cerca de un tercio de las muertes estuvieron relacionadas con el alcohol.

En otros análisis, disponibles como suplementos del artículo en línea (<http://www.ajph.org>), se encontró que:

- Las menciones de alimentos saludables fueron más altas en junio, julio, agosto y noviembre, lo que refleja la posible mayor abundancia de frutas y verduras frescas disponibles durante esas épocas del año.
- La densidad calórica de los tweets de alimentos fue más alta en los meses con los principales días festivos nacionales.

- Los gastos calóricos relacionados con los tweets de actividad física fueron más bajos en los meses de primavera y verano.
- La menor prevalencia de tweets felices (15 %) se produjo en abril.
- Asimismo, analizaron las tendencias de las variables por estado.

Los resultados de la regresión lineal ajustada

TABLE 2—Twitter Characteristics as Predictors of Health Outcomes, County Level: Contiguous United States, 2011–2013

County-Level Twitter Predictors ^a	No.	Percentage Obesity, B (95% CI) ^b	Percentage Physical Inactivity, B (95% CI) ^b
Food tweets			
Third tertile (highest)	3057	-2.49 (-3.23, -1.76)	-3.62 (-4.44, -2.80)
Second tertile		-0.61 (-1.07, -0.15)	-1.46 (-2.06, -0.86)
Physical activity tweets			
Third tertile (highest)	3054	-2.40 (-3.33, -1.47)	-2.97 (-3.85, -2.08)
Second tertile		-1.01 (-1.50, -0.52)	-1.39 (-1.85, -0.93)
Happy tweets			
Third tertile (highest)	3117	-2.23 (-3.15, -1.31)	-1.97 (-3.09, -0.86)
Second tertile		-0.79 (-1.31, -0.28)	-0.68 (-1.28, -0.07)

En los análisis de regresión lineal ajustada se examinó la asociación entre las características derivadas de Twitter y las estadísticas disponibles. Los principales resultados mostraron que:

- La edad mediana a nivel de condado se asoció positivamente con una mortalidad prematura más baja; mientras que se asoció negativamente con la mediana de los ingresos de los hogares de blancos no hispanos; solo el ingreso medio de los hogares fue estadísticamente significativo para dicha variable.
- Los condados con más tweets sobre actividad física experimentaron 714 menos muertes prematuras por cada 100 000 habitantes en comparación con aquellos que presentaron menor número de dichos tweets.
- Los condados con los porcentajes más altos de tweets con expresiones sobre felicidad, comida y actividad física

- tuvieron prevalencias de obesidad que fueron menores entre un 2,23 % y un 2,49 %.
- Los condados con mayor número de tweets sobre felicidad, comida y actividad física tuvieron prevalencias más bajas de inactividad física (entre 1,97 % y 3,62 %).
- Mayores porcentajes de tweets sobre alimentos saludables y tweets de alimentos bajos en calorías también se asociaron estadísticamente de forma significativa con mejores resultados de salud.
- La percepción positiva hacia los alimentos saludables y la actividad física se asoció significativamente con una menor obesidad e inactividad física.



Otros resultados

Los condados con el tercil más alto de *tweets* relacionados con el alcohol tuvieron un 3,6 % más de muertes por conducir bajo los efectos del alcohol y un 2,3 % más de consumo excesivo de alcohol en comparación con los condados con menos *tweets* sobre este tema.

DISCUSIÓN

Los tres hallazgos principales en el estudio fueron:

1. El aprendizaje social de las conductas expresadas en Twitter en torno a los alimentos y la actividad física se asoció con una menor mortalidad, obesidad e inactividad física a nivel de condado.
2. La felicidad y un sentimiento positivo en torno a comportamientos saludables están vinculados a mejores resultados de salud.
3. Una mayor promoción a una cultura de consumo de alcohol se relacionó con tasas más altas de abuso de esta sustancia y de mortalidad relacionada con esta conducta.

En conclusión, las características de Twitter fueron predictivas de los resultados de salud (según la composición demográfica y económica estudiada).



Los hallazgos de acuerdo al contexto

En las redes sociales las personas comparten noticias, opiniones e información sobre sus actividades; por ejemplo, lo que están a punto de comer (esto refleja sus elecciones dietarias y si estas están relacionadas con la obesidad y el riesgo de diabetes).

Es interesante que los indicadores de Twitter para felicidad y mejores comportamientos en torno a la salud, mostraron peores patrones para la zona sur, la cual se ve afectada con más pobreza, menos acceso a los recursos, incluida la atención médica y peores resultados en salud.

La influencia de los procesos sociales

Los resultados de este estudio destacan que las costumbres, valores, suposiciones y creencias en torno a la salud se enriquecen en el entorno social y pueden tener un impacto en el desarrollo y mantenimiento de comportamientos. El hallazgo de que las menciones del alcohol en las redes sociales se asocian con comportamientos de consumo y con la mortalidad, está en consonancia con otras investigaciones

sobre la influencia de las redes sociales. En este estudio se encontró que los condados con los *tweets* de alcohol más bajos tuvieron 3,6 % menos muertes relacionadas con el alcohol. En 2014, 9967 personas murieron en accidentes por conducir en estado de embriaguez. Una reducción de 3,6 % menos de muertes relacionadas con el alcohol se traduciría en alrededor de 113 muertes menos.

La felicidad y la mortalidad



El artículo reconoce que la investigación sobre los efectos de la felicidad en la salud es aún incipiente; sin embargo, afirma que existen vínculos entre la felicidad y la mortalidad, con algunos indicios de que tales vínculos pueden estar parcialmente mediados por la influencia de la felicidad en los comportamientos adoptados en torno a la salud. Por ejemplo, menciona las teorías psicosomáticas que postulan que las personas consumen alimentos que consideran reconfortantes para combatir los síntomas de angustia psicológica; también

que se ha encontrado que las emociones negativas están asociadas con la inactividad física, posiblemente por la disminución de la motivación. En el estudio, se encontró que los *tweets* relacionados con felicidad y con actividad física se asociaron con una reducción del 2 % al 3 % de inactividad. Además, señalan los autores, las investigaciones han encontrado que involucrarse incluso en niveles bajos de actividad física en comparación con ser inactivo se relaciona con disminuciones sustanciales (20 %) en la mortalidad.

Las limitaciones y las fortalezas del estudio

Sin duda, como lo indican los autores, los métodos innovadores utilizados en este estudio ofrecen a los investigadores una forma de medir el "pulso" de una comunidad mediante el análisis de las expresiones y opiniones compartidas en línea gracias al intercambio de información que se da en las redes sociales. Además, los datos geotiquetados facilitan examinar asociaciones según la ubicación de la fuente de los datos. Aun así, mencionaron estas limitaciones:

- La falta de datos estadísticos disponibles de áreas pequeñas para su correspondiente análisis.
- La construcción de indicadores por área se restringió a la recolección de información de *tweets* geotiquetados (usuarios que habilitaron la ubicación en sus teléfonos móviles). De acuerdo con algunos estudios, solo del 1 % al 2 % de *tweets* contienen información de ubicación GPS.
- Los usuarios que habilitan el geotiquetado de sus *tweets* difieren demográficamente de los que no lo hacen; por ejemplo, son un poco mayores y es más probable que sean hombres, aunque estas diferencias son pequeñas.
- La falta de representación de los usuarios de Twitter en este estudio, ya que solo el 23 % de todos los usuarios de Internet y el 20 % de la población adulta de EE. UU. usan Twitter. Además, esta red social es más popular entre las personas que viven en áreas urbanas que en las rurales (el 30 % frente al 15 %) y, entre los adultos menores de 50 años frente a los mayores de 50 años (el 30 % frente al 11 %).
- Respecto al análisis de las percepciones, el modelo solo pudo procesar *tweets* en inglés, limitando, posiblemente, las conclusiones a los hablantes de este idioma.
- La diferencia de culturas hace difícil evaluar la percepción de la felicidad ya que, mientras en algunas de ellas se fomenta la expresión abierta de sus emociones, otras que enfatizan el individualismo manifiestan de forma diferente sus emociones.
- El análisis de sentimientos en el estudio se enfocó en la clasificación de sentimientos como "feliz" frente a "no feliz" (que abarca emociones tanto neutrales como tristes), así que no se evaluó la prevalencia de los *tweets* que expresaron tristeza.

Lo que implica para la salud pública



<http://blogs.sld.cu/patricia/files/2017/03/MEDICORRSS.jpg>

A pesar de la falta de representatividad y las demás limitaciones descritas en el estudio, los autores afirman que los datos obtenidos de las redes sociales ofrecen nuevas oportunidades en la investigación por su potencial para evaluar las necesidades en salud de diferentes comunidades. Lo anterior se demuestra por las asociaciones encontradas entre los *tweets* e importantes indicadores de salud, como la mortalidad prematura, la obesidad y los comportamientos relacionados con la salud.

Concluyen entonces que estas nuevas fuentes de datos, obtenidos en tiempo real, permiten a los funcionarios de salud pública examinar las tendencias de las costumbres, las percepciones y los comportamientos que pueden augurar problemas de salud o brotes emergentes de enfermedades, proporcionándoles así una forma de intervenir para prevenir eventos adversos de salud y también para medir el impacto de las intervenciones en salud.
