

MIRA

Monitor de redes y medios

Antecedentes y metodología
para el monitoreo de temas
de salud pública en Twitter



DOCUMENTO
METODOLÓGICO

Índice

Acerca de Twitter	3

Antecedentes de análisis de Twitter como insumo para investigación de salud pública	3

El discurso sobre vacunación en Twitter	4

MIRA: Monitor de temas de Inmunización en Redes sociales de Argentina	5

Clasificación automatizada de tweets	7

Hallazgos	10

Acerca de Twitter

1 about.twitter.com

2 investor.twitterinc.com/financial-information/quarterly-results/default.aspx

En su página institucional¹, Twitter se define como como “lo que está pasando en el mundo, y aquello de lo que la gente habla, ahora mismo.” En la práctica, se trata de un servicio de mensajería instantánea y pública, que permite a sus usuarios publicar mensajes breves -o “tuits”- que pueden ser leídos, reenviados y comentados por cualquier otro usuario de la plataforma.

A principios de 2010, la empresa que provee el servicio, Twitter Inc., reportó 330 millones de usuarios activos². Si bien Twitter no reporta cantidad de usuarios por país, a mediados de 2018 una consultora de marketing estimó que 7,6 millones de Argentinos utilizan el servicio.

Los usuarios de la plataforma generan un enorme volumen de contenido; a nivel global, unos 500 millones de mensajes diarios, cubriendo toda la gama de temas imaginable, incluyendo opiniones personales. Es por ello que los tuits se han convertido en una fuente recurrente en investigaciones sobre fenómenos sociales, incluyendo aplicaciones de monitoreo de salud pública.

Antecedentes de análisis de Twitter como insumo para investigación de salud pública

Existe una extensa literatura documentando la captura y análisis de tuits como insumo para la investigación en el contexto de la salud pública.

Sinnenberg et al. (2016) realizaron un exhaustiva inspección de los catálogos de diversos jornales y archivos de publicación científica. A 2015, identificaron 137 artículos en los cuales se describía el uso de Twitter como fuente para estudios de salud pública. Entre ellos, algo más de una cuarta parte (36 artículos) correspondían a casos donde la fuente se utilizó como base de sistemas de monitoreo. En general, los procesos de monitoreo basados en actividad de usuarios se realiza mediante la recopilación a gran escala de mensajes de usuarios (en el orden de las decenas o cientos de miles de tuits) que luego son analizados en forma automatizada, en ocasiones con la ayuda de técnicas de machine learning, para inferir actitudes de la población general respecto a actividades con impacto en la salud (como fumar) o a la presencia de síntomas asociados a enfermedades (por ejemplo, usuarios quejándose de experimentar fiebre o dolor abdominal).

Entre quienes utilizaron Twitter como fuente para sistemas de monitoreo de salud pública Rodríguez-Martínez y Garzón-Alfonso (2018)³ describen un sistema de “Vigilancia de la Salud en Twitter” (THS por sus siglas en inglés). THS se presenta como una herramienta diseñada para asistir a funcionarios en el

³ Rodríguez-Martínez, Manuel, and Cristian C. Garzón-Alfonso. “Twitter Health Surveillance (THS) System.” Proceedings:... IEEE International Conference on Big Data. IEEE International Conference on Big Data. Vol. 2018. NIH Public Access, 2018.

4 Jason B., et al. "Toward real-time infoveillance of twitter health messages." *American journal of public health* 108.8 (2018): 1009-1014.

campo de la salud pública permitiendo recopilar tuits, determinar si se relacionan con alguna condición médica, extraer datos asociados (como ubicación del usuario) y crear un repositorio de alto volumen de datos que puede ser utilizado para posterior análisis en profundidad.

Por su parte, Jason et al. (2018)⁴ presentan el resultado de un proyecto multidisciplinario que aunó esfuerzos de investigadores médicos, computadores científicos y metodólogos para implementar un sistema de monitoreo en tiempo real, basado en Twitter y utilizando herramientas de código abierto. El sistema, diseñado para permitir alerta temprana en caso de la irrupción de epidemias u otras amenazas para la salud pública, combina captura y clasificación automatizada de tuits con intervención humana especializada, en la forma de técnicos encargados de analizar "en vivo" el contexto de los tuits clasificados como signo de alerta para decidir su importancia real.

El discurso sobre vacunación en Twitter

5 McNeill, Andrew, Peter R. Harris, and Pam Briggs. "Twitter influence on UK vaccination and antiviral uptake during the 2009 H1N1 pandemic." *Frontiers in public health* 4 (2016): 26.

Como plataforma de comunicación en tiempo real, Twitter aloja interacción entre usuarios que opinan sobre eventos de actualidad, funcionando como muestra del tono y sentimiento del debate pública que se desarrolla en torno a temas de coyuntura. En el caso de la vacunación pública, un estudio en el Reino Unido monitoreó la conversación en Twitter en torno a la respuesta del gobierno frente a la pandemia del virus H1N1 en 2009-2010, hallando corrientes de escepticismo en torno a los beneficios del gasto público en vacunas⁵. También se ha estudiado si el tono de los mensajes compartidos en la red afectan el grado de participación (respuestas y reenvíos) de otros usuarios, encontrando que las posiciones extremas, tanto a favor como en contra de la vacunación, generan mayor repercusión sin importar su contenido⁶.

6 Blankenship, Elizabeth B., et al. "Sentiment, Contents, and Retweets: A Study of Two Vaccine-Related Twitter Datasets." *The Permanente journal* 22 (2018).

7 blog.twitter.com/en_us/topics/company/2019/helping-you-find-reliable-public-health-information-on-twitter.html

La comprobada difusión de noticias y mensajes anti-vacunación en las redes sociales llevó a Twitter a reaccionar en 2019. La empresa anunció su compromiso con "la salud de la conversación pública" e implementando un sistema automático para combatir la desinformación y las llamadas fake news: a partir de entonces, los usuarios que buscan términos relacionados con la vacunación son dirigidos a un sitio con información oficial⁷. A pesar de esfuerzos de esta clase, la exposición automática a información de fuentes oficiales podría ser ineficaz ante posturas extremistas. Tras el análisis realizado sobre la conversación en Twitter de diversas comunidades de usuarios, se encontró que las personas reconocibles como *anti-establishment* reciben mayor exposición a mensajes negativos respecto a la vacunación, y a la vez desconfían de las instituciones tradicionales y descreen los argumentos científicos⁸. Tal evidencia sugiere que la simple publicación y circulación de argumentos científicos podría ser insuficiente para recuperar la confianza de los grupos más refractarios, requiriendo estrategias holísticas que incluyan una mejor comprensión de sus opiniones y temores.

8 Lutkenhaus, Roel O., Jeroen Jansz, and Martine PA Bouman. "Mapping the Dutch vaccination debate on Twitter: Identifying communities, narratives, and interactions." *Vaccine: X* 1 (2019): 100019.

MIRA: Monitor de temas de Inmunización en Redes sociales de Argentina

Con el objetivo de cuantificar a través del tiempo el volumen y tono del discurso público relacionado con la vacunación, la Fundación Bunge y Born ha comenzado el desarrollo e implementación de un sistema de Vigilancia Inmunológica en Redes Argentinas (MIRA). Al momento, MIRA utiliza a Twitter como su fuente de datos para el análisis. Se espera que el desarrollo de ésta aplicación resulte en una metodología, conjunto de mejores prácticas, y documentación de técnicas que permitan extender el sistema a una gran variedad de otros temas cruciales para la salud pública.

⁹ developer.twitter.com/en/docs/tweets/filter-realtime/guides/connecting.html

MIRA accede en tiempo real al “*stream*”⁹ de Twitter, es decir la secuencia de mensajes originados en todo el mundo a medida que ocurren. MIRA utiliza el acceso oficial a Twitter, siguiendo los parámetros y reglas de uso que la plataforma exige.

En cada segundo, la comunidad global de usuarios genera miles de tuits. El acceso público de Twitter, utilizado por las entidades que recopilan esta información, limita la cantidad de tuits que se pueden recuperar en tiempo real. En la práctica, quienes monitorean la fuente pueden recuperar una fracción pequeña del total de mensajes, que algunas fuentes estiman en 1%. Esta limitación se ve paliada al utilizar filtros de búsqueda, que permiten obtener sólo los tuits que cumplen determinadas condiciones. MIRA realiza una búsqueda por palabras clave, asegurando que todos los tuits de la muestra recibida segundo a segundo contengan términos de interés para el monitoreo de discurso sobre vacunas. En la práctica, dado que el discurso sobre cualquier tema en particular representa una pequeña fracción de todo lo que se dice en la plataforma, es posible asumir que todos o una gran parte de los tweets de interés son capturados con esta modalidad.

La lista de palabras clave es la siguiente:

‘vacuna’, ‘vacunas’, ‘vacunar’, ‘vacunación’, ‘vacunacion’, ‘vacunados’, ‘vacunadas’, ‘vacunado’, ‘vacunada’, ‘vacuné’, ‘vacunen’, ‘vacunemos’, ‘vacunando’, ‘vacunaron’

MIRA consulta en forma continua los sistemas de Twitter, recopilando cada tuit encontrado con alguna de las palabras clave en su contenido. Los tuits son almacenados en una base de datos para su posterior análisis. Cada registro contiene los siguientes campos:

TABLA 1: ESTRUCTURA DE DATOS DE UN TWEET

CAMPO	ATRIBUTO REPRESENTADO	EJEMPLO
id	número único de identificación del tuit	40904081
in_reply_to_status_id	si el tuit es una respuesta a otro previo, identificador único del tuit al que se está respondiendo	40860874
in_reply_to_user_id	si el tuit es una respuesta a otro previo, identificador único del usuario al que se le está respondiendo	127387408
text	contenido del tuit (el mensaje)	"RT @Pajaropolitico: La antigua administración de la @SSalud_mx no alertó sobre el desabasto de vacunas contra el sarampión"
created	fecha y hora de emisión del tuit	2019-01-30 00:37:25
lang	lenguaje del tuit (clasificado en forma automática por Twitter mediante machine learning)	es
source	dispositivo/aplicación desde donde se envió el tuit (cliente web, móvil android, móvil iOS, Instagram, Foursquare, etc)	Twitter for Android
lat	si el tuit incluye coordenadas que lo georeferencian, la latitud correspondiente	19.40599
lon	si el tuit incluye coordenadas que lo georeferencian, la longitud correspondiente	-99.13929
user_name	nombre de usuario del autor del tuit	jere_1985
user_id	identificador único del usuario/a que publicó el tuit	12141728
user_created	fecha de alta del usuario/a en Twitter	2014-12-16 04:39:04
user_description	descripción provista por el usuario/a sobre sí mismo/a	egresado de la FI UNAM Coyoacán
user_location	lugar de residencia provista por el usuario/a	Ciudad de México
user_followers	cantidad de usuarios que "siguen" al autor/a del tuit	882
user_followed	cantidad de usuarios "seguidos" por el autor/a del tuit	95112
user_statuses_count	cantidad de tuits publicados por el usuario/a hasta el momento	105154
user_verified	variable binaria (SI/NO) que indica si el usuario/a ha verificado su identidad ante Twitter	F
retweeted	variable binaria (SI/NO) que indica si el tuit es una republicación ("un retuit") de un mensaje original de otro usuario/a	T
place_name	en caso de que Twitter haya podido identificar el lugar de emisión del tuit, el nombre de sitio	DF, Mexico
place_country	en caso de que Twitter haya podido identificar el lugar de emisión del tuit, el nombre del país al que corresponde	MX
place_type	la clase del lugar identificado	city

Las coordenadas de geolocalización permiten identificar tanto país como ciudad de procedencia de un tuit. Sin embargo, dado que menos del 1% de los tweets capturados presentan coordenadas, el principal atributo que permite determinar la procedencia del usuario es el campo *user_location*, que responde al lugar de residencia que cada usuario de la plataforma elige declarar. Respondiendo a que los términos de búsqueda son palabras del lenguaje español, los tuits acumulados provienen de hispanoparlantes en todo el mundo. Analizando el campo *user_location* se ha logrado determinar como Argentina la procedencia de algo más de un 12% del total de los tuits. Dado que muchos usuarios optan por especificar su provincia de residencia, es posible analizar los datos a ese nivel, e incluso a escala ciudad para ciertas urbes suficientemente representadas: Ciudad de Buenos Aires, Bahía Blanca, La Plata entre otras.

Clasificación automatizada de tweets

10 Zhang, Yin, Rong Jin, and Zhi-Hua Zhou. "Understanding bag-of-words model: a statistical framework." *International Journal of Machine Learning and Cybernetics* 1.1-4 (2010): 43-52.

11 Wang, Min, et al. "Microblog sentiment analysis based on cross-media bag-of-words model." *Proceedings of international conference on internet multimedia computing and service*. 2014.

12 Joulin, Armand, et al. "Bag of tricks for efficient text classification." *arXiv preprint arXiv:1607.01759* (2016).

Debido al caudal de tweets recolectados (alrededor de 6.5 millones al año) resulta de interés la implementación de sistemas automatizados de clasificación de contenidos. La clasificación por origen basa en un sistema simple, que identifica a la Argentina o a alguna de sus ciudades o provincias en el campo donde los usuarios auto-reportan su procedencia. En contraste, la separación del contenido de los mensajes en distintas categorías resulta más complejo. Para identificar mensajes relacionados a la conversación en torno a la importancia de las vacunas -por ejemplo, sobre antivacunismo-, versus aquellos que tratan de otros temas: *carne* vacuna, vacunas para mascotas, etc. Para esa tarea se implementó un sistema de clasificación de texto basado en *fastText*, un algoritmo desarrollado para clasificar texto y representar texto siguiendo el modelo de procesamiento de lenguaje natural conocido como "bag of words". El modelo, uno de los más utilizados para sistemas automatizados de clasificación¹⁰, se basa en convertir a cada palabra una serie de atributos numéricos (un vector). Al analizar un texto, sólo se extrae el identificador numérico de cada palabra, sin que sea necesario considerar el orden en que aparecen los términos ni su semántica -de allí el nombre bag of words, una simple "bolsa de palabras". Esta simplificación facilita en gran medida la aplicación de técnicas de aprendizaje automático, ya que la comparación entre textos se reduce a la comparación de conjuntos de vectores numéricos¹¹. El método *fastText* incorpora una mejora respecto al modelo *bag of words* tradicional, al considerar n-gramas (secuencias de palabras) y no sólo vocablos sueltos al realizar la conversión en vectores¹². Esto es, si se eligiera trabajar con n-gramas de extensión dos, *fastText* asignará identificadores distintos a las palabras "campo", "partido" y "verde", así como también a secuencias de hasta dos palabras, incluyendo "campo verde" y "partido verde".

fastText es parte de la familia de algoritmos de aprendizaje automático de tipo "supervisado", ya que requiere de un conjunto de ejemplos previamente clasi-

ficados -mediante intervención humana- de los cuales “aprender” una serie de atributos que diferencian una categoría de otra. Durante ésta etapa, llamada *de entrenamiento*, el algoritmo produce un modelo estadístico basado en la información conocida. Una vez realizado el aprendizaje, el modelo puede aplicarse a ejemplos de texto nuevos (nunca “vistos” por el algoritmo de clasificación) sobre los cuales estima la probabilidad de pertenecer a alguna de las categorías con las que fue entrenado; por ejemplo “antivacunismo”, “faltante de vacunas”, o “tema no relacionado”.

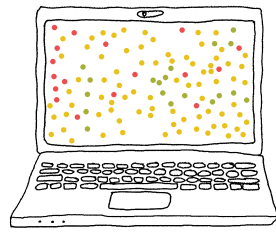
Para la clasificación de tweets según tema, el algoritmo fue entrenado contra un corpus de tweets previamente caracterizados mediante intervención humana (n ~ 10.000). A cada ejemplo de entrenamiento se le asignó una etiqueta entre tres posibles: “antivacunismo” (mensajes que se refieren a la polémica en torno al valor de las vacunas), “problemas de acceso” (mensajes que se refieren a barreras para el acceso a las vacunas, como faltantes de stock), y “otros” para mensajes fuera del tema de interés. Para el entrenamiento, se alimentó al algoritmo un 80% de los tweets preclasificados para la extracción de un modelo. Concluido el entrenamiento, la calidad del modelo fue evaluada contra el 20% restante de los tweets preclasificados, extrayendo una “predicción” de su categoría utilizando la información aprendida en el paso previo. El modelo predijo correctamente la categoría asignada por un humano con precisión y exhaustividad superiores al 95% para todos los casos (Tabla 2)

TABLA 2: MÉTRICAS DE PERFORMANCE DEL MODELO DE CLASIFICACIÓN AUTOMÁTICA DE TEMAS

ETIQUETA	PRECISIÓN	EXHAUSTIVIDAD (RECALL)	MEDIDA-F (F-SCORE)
Antivacunismo	97,7%	99,2%	98,5%
Problemas de acceso	96,5%	98%	97,2%
Tema no relacionado	98%	95,7%	96,8%

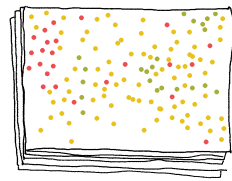
La alta eficiencia de los algoritmos de fastText permite realizar la clasificación de decenas de miles de tweets en pocos segundos sin necesidad de hardware especializado, por lo cual es posible identificar los temas de los tweets en tiempo real, a medida que son recolectados y almacenados en una base de datos. El “dataducto” (data pipeline) que recorren los tweets procesados por MIRA está completamente basado en sistemas operando en la nube (Figura 2).

FIGURA 1: "DATA PIPELINE" DEL SISTEMA DE RECOLECCIÓN, ALMACENAMIENTO Y CLASIFICACIÓN DE TWEETS DE MIRA



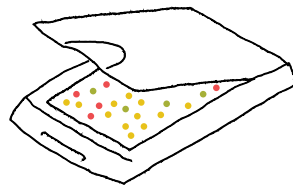
Accede

a la secuencia de mensajes originados en Twitter en tiempo real en todo el mundo y al conjunto de noticias online en más de 100 idiomas.



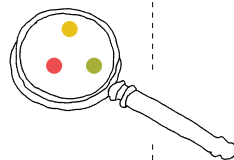
Recopila

los tweets que contienen palabras clave vinculadas al tema de vacunación y los textos completos de noticias pertenecientes a las categorías de Salud Pública y Vacunas.



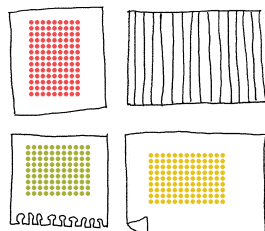
Procesa

8.000 noticias y 10.000.000 de tuits por año vinculados a temas de vacunación.



Analiza

una serie de atributos mediante un sistema automático de interpretación de lenguaje que permite diferenciar una categoría de otra.



Clasifica y almacena

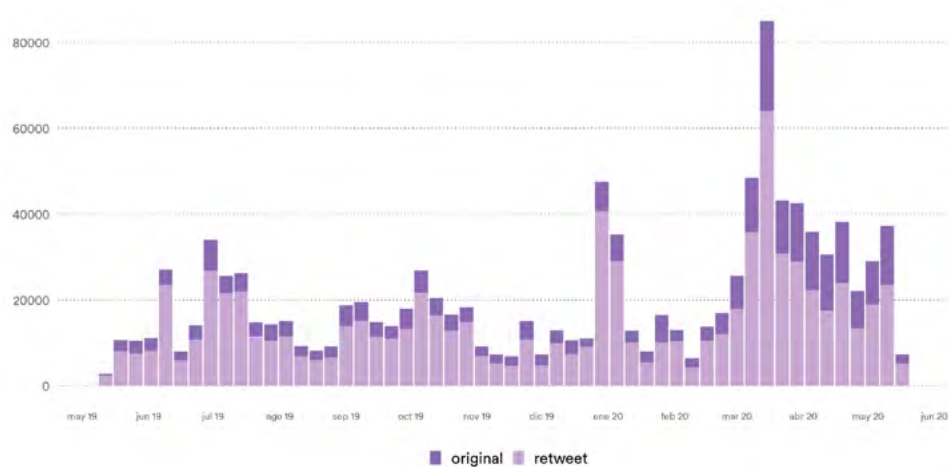
la información en diferentes temas.

Hallazgos

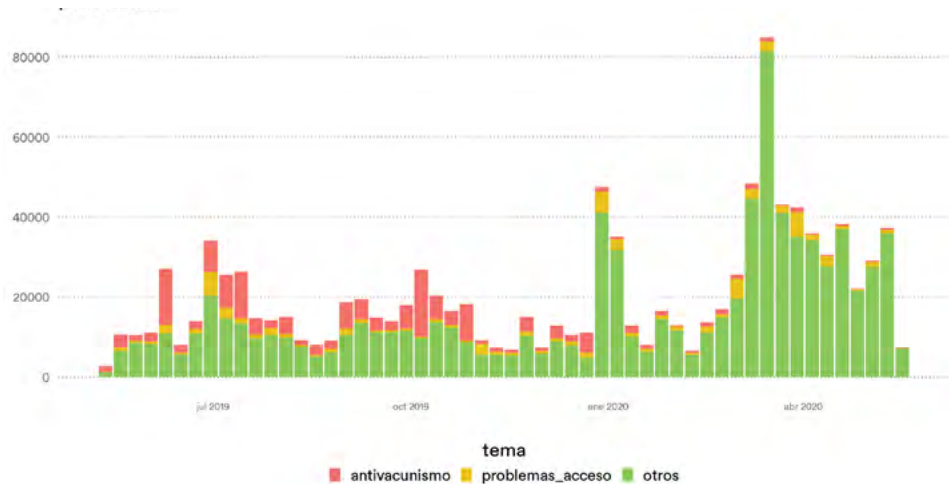
13 Encontrar fuente

El discurso en Twitter sobre vacunas en la Argentina, representado por aquellos tweets que mencionan las palabras clave antes mencionadas, exhibe una distribución entre contenido original versus “retuiteos” similar a la encontrada en otros campos de estudio¹³. En relación al volumen total de tuits, es una pequeña cantidad de usuarios la que genera los mensajes originales, que luego son replicados (reenviados o comentados) por una cantidad mucho mayor de participantes en la conversación:

TWEETS SOBRE VACUNAS DE USUARIOS EN ARGENTINA POR SEMANA

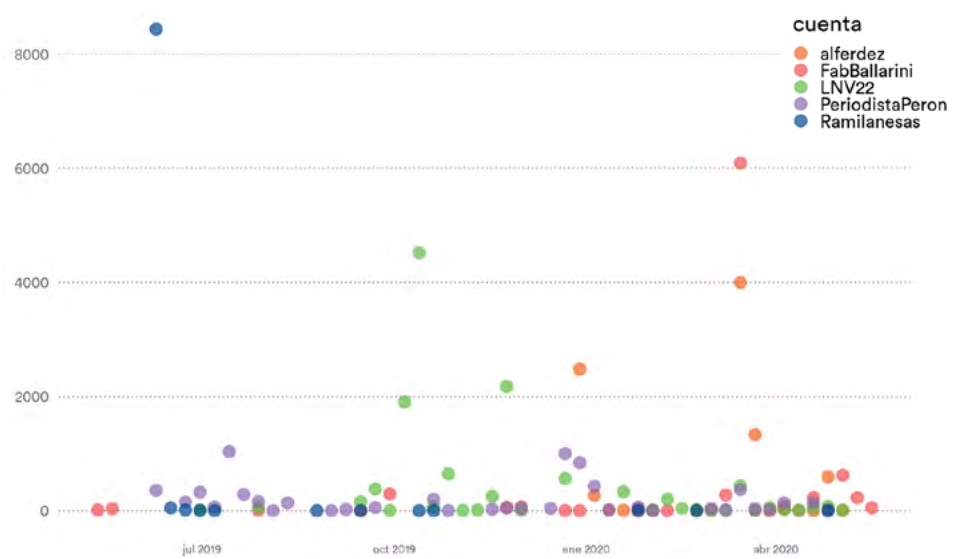


TWEETS SOBRE VACUNAS DE USUARIOS EN ARGENTINA POR SEMANA



Ulterior análisis indica que los picos de actividad se deben a la respuesta ante mensajes específicos que alcanzan la viralidad, es decir que son re-enviados y respondidos masivamente.

USUARIOS DE LA RED CON MÁS REPERCUSIÓN INMEDIATA (RETWEETS) POR SEMANA



Al identificar el contenido de los tuits viralizados, así como sus autores, es posible determinar los temas debatidos. Para los picos de actividad hallados, se verifica que los mensajes con mayor repercusión podrían clasificarse como *pro-vacunas*. Por ejemplo, una muestra de los mensajes del @MelissaMazonR, cuenta que originó varios tuits virales:



“Fallece el día de hoy en urgencias pediátricas paciente masculino de 2 años de edad recién cumplidos el día domingo 14 de Julio. Antecedentes de importancia: **NO TIENE VACUNAS. Diagnóstico: Tétanos.**”

CLASIFICACIÓN RECIBIDA: ANTIVACUNISMO



“Ese niño debió el domingo celebrar su cumpleaños, no estar hospitalizado conectado a un ventilador luchando por su vida. Por favor, vacunen a sus hijos, ellos no tienen la culpa.”

CLASIFICACIÓN RECIBIDA: ANTIVACUNISMO



“**VACUNEN A SUS HIJOS, ES UNA ORDEN**”

CLASIFICACIÓN RECIBIDA: ANTIVACUNISMO

Los tuits viralizados alcanzan gran repercusión pero no necesariamente representan a personas o entidades que emiten en forma activa y sostenida discurso sobre la vacunación. Para identificar a los actores que publican con frecuencia mensajes al respecto de las vacunas, se recurre a contabilizar la cantidad de días diferentes en los que cada usuario representado en la base ha emitido tuits.

Al determinar las cuentas con mayor actividad sostenida, se observa que ninguna de las que producen interacción en forma frecuente pertenece a fuentes oficiales.

CUENTA	DIAS DISTINTOS	TOTAL RETWEETS ACUMULADOS
LibertadDoctora	70	1015
LNV22	50	11100
lucagarciaa05	38	1892
PeriodistaPeron	35	2680
Ramilanesas	29	8514
unamadresoltera	29	1062
TereGarciaOK	28	700
nawilo_0	25	2440
Candee_Millan	24	981
Manci_L	23	1958

Se trata de cuentas personales, cuyo discurso es decididamente pro-vacunación. De hecho, la mayoría de los mensajes en la red -un 65%- expresa rechazo, con frecuencia en términos drásticos, al movimiento antivacunas. La principal subcategoría dentro de la conversación en torno al valor de las vacunas son expresiones de rechazo, incredulidad o ataque a las posturas antivacunistas (Tabla X).

La segunda categoría más común (12%) la ocupan las noticias con desinformación, por ejemplo aludiendo a supuestos estudios científicos que probarían que las vacunas causan mayor mortalidad entre infantes. Estos tweets se originan en muy pocas cuentas, que los repiten una y otra vez. Los mensajes que simplemente expresan desconfianza personal hacia las vacunas son relativamente escasos: apenas un 4,4% de los mensajes encontrados expresa rechazo a la vacunación propia o de los menores a cargo.

IMPACTO DE LA PANDEMIA EN EL DISCURSO SOBRE VACUNAS

