

Identificación de usuarios ansiosos *para detectar condiciones de salud mental en la población tuitera de México*

Identification of Anxious Users *to Detect Mental Health Conditions in the Twitter Population of Mexico*

Ana Miriam Romo Anaya, Víctor Silva Cuevas e Irving Gibran Cabrera Zamora*

Basándonos en múltiples estudios que muestran que las palabras que usa una persona en sus escritos transmiten información sobre su personalidad, motivaciones, estado de ánimo y emocional e incluso su nivel económico, podemos entonces pretender que la información subyacente de los usuarios de *Twitter* permite generar indicadores del estado de ánimo e indicadores asociados con la prevalencia de condiciones de salud mental. Como primera aproximación, este trabajo se enfoca en identificar a quienes más expresan sentimientos negativos en sus textos acompañados de palabras relacionadas con ansiedad y que los hacen diferenciables de los demás. Para esto, usamos toda la muestra de tuits recolectados por el INEGI desde el 2017 al 2020 que se utilizó para la construcción del Índice del Estado de Ánimo de los Tuiteros en México junto con el recurso léxico *LIWC*

Based on multiple studies that show that the words a person uses in their writings convey information about their personality, motivations, moods, emotional states, and even economic level, we can claim that the underlying information of Twitter users allows us to generate mood indicators as well as indicators associated with the prevalence of mental health conditions. As a first approximation, this work focuses on identifying those who express more negative feelings in their texts accompanied by anxiety-related words and who differentiate them from others. For this we use the entire sample of tweets collected by INEGI from 2017 to 2020 that was used for the construction of the Mood Index of Twitter users in Mexico along with the lexical resource *LIWC* (Linguistic Inquiry Word Count) designed to analyze psychological implications of word use. This resource

* Instituto Nacional de Estadística y Geografía (INEGI), miriam.romo@inegi.org.mx, victor.silvac@inegi.org.mx e irving.cabrera@inegi.org.mx, respectivamente.

Nota: colaboradores en el área de investigación del INEGI; los juicios contenidos en este artículo son responsabilidad exclusiva de los autores y no reflejan una posición oficial del Instituto al respecto.

(*Linguistic Inquiry Word Count*) diseñado para analizar implicaciones psicológicas del uso de las palabras muy empleado por psicólogos, sociólogos y comunicólogos.

Los resultados nos informan que, a pesar de que la captación de usuarios y sus textos en estos cuatro años está descendiendo, existe una incidencia creciente de personas con textos ansiosos cargados de valencia negativa a través del tiempo. Como un monitoreo de la prevalencia de ansiedad, se construyó un índice para cada tuitero recolectado en el periodo de referencia en las 32 entidades federativas del país. Este nos reveló el impacto que han tenido eventos acontecidos en estos últimos años, como el sismo del 2017 en ciertos estados y el inicio del confinamiento por la pandemia por COVID-19 en el 2020.

Palabras clave: *Twitter*; Índice de Ansiedad; aprendizaje estadístico; ánimo tuitero; *datificación* de la salud; INEGI.

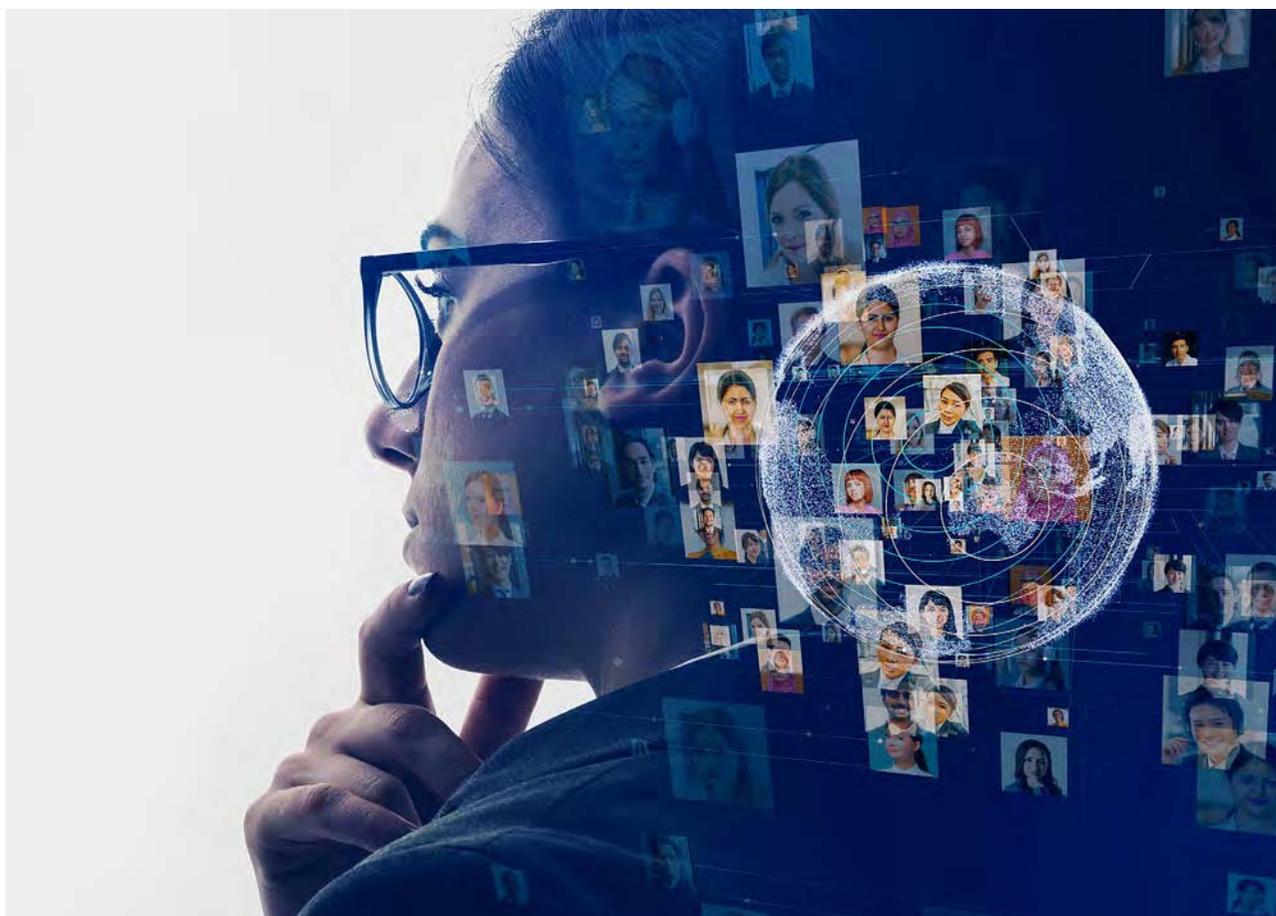
Recibido: 21 de febrero de 2022.

Aceptado: 29 de junio de 2022.

has often been used by psychologists, sociologists and communication scholars.

The results inform us that, although the number of users and texts has decreased in these four years, there is an increasing incidence of people with texts reflecting anxiety, loaded with negative valence over time. As a monitoring of the prevalence of anxiety, an index was constructed for each Twitter user collected in the reference period in the 32 states of the country. This revealed the impact of events that have occurred in recent years, such as the 2017 earthquake in certain states, and the beginning of the COVID-19 pandemic confinement in 2020.

Key words: *Twitter*; Anxiety Index; statistical learning; Twitter mood; datafication of health; INEGI.



AI (Inteligencia Artificial). Red de comunicación.

Introducción

La extracción de conocimiento de manera automática con distintos tipos de datos es un área de la inteligencia artificial que está cobrando cada vez más interés en distintos sectores. Con mucho éxito, se ha realizado una gran variedad de estudios que usan la enorme cantidad de datos disponibles en el mundo de la tecnología y que, junto con las técnicas del aprendizaje estadístico, han logrado diagnosticar y predecir más información de interés con alta precisión.

La tendencia tecnológica que logra convertir o mapear los datos cualitativos obtenidos de las acciones de las personas a través del uso de dispositivos y/u ordenadores a datos cuantificables para ser medidos y descubrir nuevo conocimiento o *insight*¹ se conoce como *datificación*. Gracias a ella, el procesamiento y análisis de datos no estructurados obtenidos de estos medios (texto, imágenes y audio) están ayudando a medir el impacto de eventos relevantes, opiniones políticas, preferencias, comportamientos y sentimientos de los individuos. Es por ello que iniciativas gubernamentales, instituciones y empresas regionales y globales reconocen el valor de estos datos, por lo que están invirtiendo en capital humano e infraestructura tecnológica para analizar y predecir el comportamiento de la población a la hora de comprar un producto o de opinar sobre una decisión política, entre otros aspectos.

En particular, en el mundo de las redes sociales, los datos que se pueden obtener y procesar son inagotables, pues muchos de sus usuarios comparan sus opiniones y emociones día a día. Sus palabras pueden reflejar la personalidad, el estado de ánimo e incluso su nivel económico. Por ello, estas plataformas se consideran una fuente potencial de información para el estudio de una gran variedad de temas.

El Instituto Nacional de Estadística y Geografía (INEGI), por ejemplo, propuso un indicador que mide el estado de ánimo de los tuiteros en México (aplicación web²) por medio de técnicas de análisis de sentimiento, el cual se obtiene mediante una muestra recolectada diariamente de tuits georreferenciados clasificándolos de manera automática como un mensaje con carga positiva o negativa. Esta información llega a ser útil cuando se relaciona con eventos políticos, sociales, deportivos o económicos que se dan en el país para conocer el sentimiento de cierta parte de la población ante estos sucesos.

Dentro de este mismo enfoque del análisis de sentimiento, se puede determinar no solo la valencia del texto, sino también la carga emocional negativa integrada por miedo, ira, tristeza o asco, entre otras que pueden existir detrás de la información que el individuo expone en sus redes sociales; es un tópico que está contribuyendo de manera importante al campo de la medicina y la psicología, pues su alcance es tal que, si llegan a ser muy intensas, estas pueden afectar la calidad de vida del individuo, mostrando un gran riesgo de padecer un trastorno mental, como depresión y ansiedad, con la posibilidad de llegar al suicidio (Piqueras Rodríguez *et al.*, 2009).

Según Richard Layard (2017), el efecto de un problema de salud mental como la depresión va más allá del bienestar social: afecta la producción en las empresas debido a que este padecimiento prevalece más en las personas en edad laboral. Se hace notable la reducción de la concentración y memoria del trabajador, aumentan sus ausencias y se generan más gastos médicos que, a su vez, merman los presupuestos de las agencias públicas de salud. Toda esta información muestra la importancia no solo de atender los casos ya diagnosticados de padecimientos mentales, sino también detectar las posibles situaciones en riesgo de padecerlos.

Este tipo de estudios se relacionan directamente con la *datificación* de la atención médica, también

¹ Es un término utilizado en psicología proveniente del inglés, que se puede traducir al español como *visión interna* o, más genéricamente, *percepción* o *entendimiento*.

² <https://www.inegi.org.mx/app/animotuitero/#/app/multiline>

llamada *datificación de la salud* (Ruckenstein et al., 2017), que se desarrolla con diferentes registros, como la investigación médica basada en datos (de gobierno, biobancos), la infraestructura de la salud pública, la monitorización continua de pacientes, biosensores implantables, el uso de internet para la interacción médico paciente, prácticas de autocuidado, dispositivos de salud, así como *fitness* portátiles y aplicaciones para teléfonos inteligentes. Estos datos recolectados son de gran interés para el gobierno, los especialistas o las instituciones públicas de la salud, pues reconocen su potencial para prevenir y mitigar las cargas físicas y financieras de las *enfermedades del estilo de vida actual*, como obesidad, diabetes y enfermedades cardiovasculares. Además, proporciona evidencia para la generación de programas o políticas de salud pública, y a los profesionales de la salud les ayuda a un temprano diagnóstico para enfocarse en dar pronto seguimiento a quienes necesiten más atención.

El estudio presentado en este artículo pretende generar información relevante acerca de la emoción negativa *ansiedad* en los mensajes de los tuiteros en México, identificando *el grado de ansiedad* y *en qué momentos se incrementa* como indicadores en el tema de salud pública y bienestar social utilizando datos no tradicionales (los cuales no se obtienen por medio de censos o encuestas).

Consideramos que los tuits georreferenciados a los que accede el INEGI desde el 2016 son una base de datos suficiente para lograr estimaciones sobre los patrones de escritura de los tuiteros, que indicarán la vulnerabilidad de expresar emociones negativas que los hace diferenciable de los demás. Además, nos permite realizar comparaciones de tipos de tuiteros por unidad geográfica a través de los años.

Como punto de partida, decidimos empezar a identificar la *ansiedad* que se puede ver reflejada en la escritura de los tuiteros tomando como hipótesis de que esta forma parte de las emociones negativas presentes comúnmente en padecimientos mentales. Algunos psicólogos (Piqueras

Rodríguez et al., 2009) indican que la ansiedad es el resultado de una expresión patológica del *miedo* y mencionan que "... La expresión conductual es la confrontación, la reacción agresiva y lleva al individuo, para mitigar su malestar, a conductas no saludables, como beber, fumar y no alimentarse sanamente...". Además, la Organización Mundial de la Salud (OMS) ha manifestado varias veces que tanto la depresión como la ansiedad son consideradas como enfermedades mentales y que estos padecimientos se están convirtiendo en un problema mundial, pues sucede en países ricos y pobres (OMS, 2021).

Realizar esta medición complementará la información que se obtiene del Índice del Estado de Ánimo de los Tuiteros en México pues, aparte de cuantificar el sentimiento negativo de los textos de cada tuitero, identifica el grado de ansiedad que los hace diferenciables de otros.

La estrategia consistió en asociar un valor a cada tuitero, el cual cuantifica qué tanta ansiedad negativa o positiva expresa en sus textos. Hacemos un comparativo por entidad geográfica y temporalidad al comparar este índice de todos los tuiteros recolectados para el 2017, 2018, 2019 y 2020. Realizar esta comparación ayudó a identificar el impacto que tienen ciertos eventos políticos, económicos y sociales en la población.

Además, perfilamos aquellos tuiteros que suelen expresar más palabras de ansiedad con polaridad negativa, ya que estos pueden ser candidatos idóneos para ser evaluados por psicólogos para asignar de manera cualitativa una propensión a padecer o no alguna condición de riesgo en cuanto a su salud mental. Para descubrir la incidencia de este tipo de usuarios, su tipificación se llevó a cabo por año.

Conscientes estamos de que se deben considerar factores que no siempre son detectados de forma automática para decidir si un usuario tiende a tener un problema real de salud mental; por ejemplo, la detección de ciertas emociones no es altamente precisa debido a que el tono de la expresión

textual puede ser interpretada de formas distintas según la zona geográfica de donde provenga el tuitero. Hay regiones en este país en las que sus habitantes suelen ser más sarcásticos o bromistas que en otras. Es entonces muy importante ratificar de manera cualitativa que el grupo de tuiteros que detectemos con métodos cuantitativos en realidad esté padeciendo ansiedad, por lo que expertos en psicología y/o comunicación deberán validar la detección.

En los dos siguientes apartados se relata cómo se están aplicando las técnicas estadísticas y recursos léxicos con los datos de las redes sociales para pronosticar el riesgo de que un usuario pueda estar padeciendo un problema de salud mental con base en lo que expresan en dichas plataformas.

Sobre los datos que usamos y su procesamiento se detalla en la tercera parte del este documento, continuando con un análisis exploratorio de estos donde se reflejan tendencias a expresar más ansiedad en ciertos meses relacionados con eventos políticos y sociales.

En las dos penúltimas secciones hablamos de la construcción del Índice de Ansiedad, así como la aplicación de métodos de aprendizaje no supervisado y supervisado para la detección de los usuarios con mayor ansiedad; por último, presentamos algunas conclusiones y consideraciones.

Técnicas de aprendizaje estadístico para la obtención de información

Las empleadas en el campo de la inteligencia artificial son algoritmos estadísticos y matemáticos que ayudan a identificar de manera automática patrones en datos y a resolver problemas complejos, por ejemplo, de *clasificación* y *agrupamiento* en distintos sectores de la ciencia.

Uno de los pioneros en aprovechar los modelos estadísticos para descubrir patrones de salud mental de los usuarios a través de redes sociales fue De Choudhury (2013), quien recopiló datos mediante

una base de usuarios diagnosticados clínicamente con depresión y que, a su vez, estuvieran inscritos en *Twitter*. A través de sus publicaciones un año antes de ser diagnosticados, se midieron los estilos lingüísticos, los atributos de sus propios grupos sociales (por medio de gráficos de redes), la carga emocional de sus textos y los tópicos a los que hacen referencia. Con un método de aprendizaje supervisado, alcanzó 72 % de precisión en el diagnóstico de depresión.

Actualmente, el autor David Brooks, columnista de *The New York Times*, señala en su artículo “How Artificial Intelligence Can Save Your Life...” (2019) que estos métodos son: “... [técnicas] aterradoras, difíciles de entender, y bastante asombrosas pues ayudan a salvar la vida de personas, ya que con estas técnicas se puede detectar y diagnosticar a tiempo a personas que intentarán suicidarse...”. Hace referencia a varios estudios que han usado estas técnicas para el diagnóstico temprano de padecimientos mentales.

El artículo escrito por Eric Topol en su libro llamado *Deep Medicine...* (2019) muestra un estudio en el que se introduce la información de los registros médicos de los pacientes a un algoritmo de aprendizaje estadístico para predecir quién es probable de cometer suicidio, y lo hizo correctamente en 93 % de las veces.

Otro trabajo muy innovador es el que llevó a cabo una empresa llamada Mindstrong³ fundada por un equipo de especialistas en neurociencia, medicina e inteligencia artificial. En ella se está desarrollando una aplicación que controla la forma en la que la persona escribe, con qué frecuencia borra caracteres y cuánto se desplaza mientras usa otras aplicaciones, tratando de descubrir cómo cambia su cognición mental a lo largo del tiempo. Los datos que recopila se cifran y analizan en forma remota mediante aprendizaje automático, y se comparten con el paciente y su médico. A diferencia de otras aplicaciones que tienen la misma finalidad, Mindstrong asegura que —considerando las

³ <https://mindstronghealth.com/>

interacciones físicas de los usuarios con sus teléfonos identificando no lo que hacen, sino cómo lo hacen— puede ofrecer pistas más precisas de rastrear estos problemas con el tiempo.

Además, existen otras investigaciones (Reece y Danforth, 2017) donde la unidad de análisis no es solo el texto, ahora se vuelve la atención hacia las imágenes, de manera que analizaron 43 950 fotos de *Instagram* de 166 personas. Sus algoritmos lograron 70 % de precisión al reconocer quién estaba deprimido.

En México también hay estudios con el mismo objetivo: identificar de manera temprana indicios de riesgo de personas propensas a depresiones, tristeza y/o suicidios en función de los textos que escriben en sus plataformas digitales (Rodríguez-Esparza *et al.*, 2016).

Recursos usados en trabajos relacionados

El tener un conjunto de usuarios etiquetados con ansiedad o depresión (conjunto de entrenamiento) diagnosticados por especialistas psicólogos o psiquiatras es la clave para generar un conjunto ideal de entrenamiento para que, junto con un buen método de aprendizaje estadístico supervisado, se obtengan predicciones confiables con relación a que un usuario exhiba un riesgo de ansiedad o depresión. Sin embargo, no siempre se tiene el tiempo, el dinero y la disponibilidad de información para conformar uno con estas características. Es necesario acudir a otros recursos más económicos para identificar a estos usuarios; por ejemplo, según lo demostraron Tsugawa *et al.* (2015), el uso de un *bag of words*⁴ relacionado con depresión o tristeza es útil para identificar emociones negativas.

Otra forma para detectar de manera automática una colección de usuarios en *Twitter* con algún pro-

blema de salud mental es el de localizar a los usuarios que hayan expresado públicamente tener una afección, ya sea creyendo que pueden tener apoyo de otros que padecen lo mismo o para combatir el tabú de las enfermedades mentales o también como explicación de sus comportamientos (Coppersmith *et al.*, 2014; y Nadeem *et al.*, 2014). Por lo regular usan un filtrado de tuits buscando expresiones como “me diagnosticaron X”. Esta X puede ser algún trastorno, según lo que se necesita detectar, como: depresión, ansiedad, trastorno bipolar o de estrés postraumático, entre otros. En estos estudios se está suponiendo que, debido a que existe un estigma alrededor de los problemas de salud mental, parecería poco probable que se declare padecerlos cuando en realidad no es este el caso. Aun así, se sabe que en este tipo de redes se suelen dar declaraciones a modo de bromas y/o sarcasmos, por lo que es muy importante que se verifique manualmente que el significado del texto sea genuino.

Otro enfoque es el de realizar una identificación de categorías de emociones en un texto usando recursos léxicos, es decir, diccionarios que contienen un conjunto de palabras etiquetadas por emociones. La mayoría de estos recursos están en el idioma inglés; algunos de ellos tienen traducción al español, pero no todos son evaluados manualmente para asegurarse de que la connotación de la palabra sea en verdad etiquetada con la emoción correcta. En el cuadro 1 se describen algunos diccionarios desarrollados en español disponibles para los interesados en estos temas.

Es importante mencionar que *LIWC* se consigue como un programa computacional ejecutable comercial que hace un conteo de las palabras para 74 categorías agrupadas en cinco procesos psicológicos: social, afectivo, cognitivo, biológico y personal. Además, proporciona el número de palabras relacionadas con el lenguaje estándar de escritura, por ejemplo: total de pronombres, total de pronombres personales (1.ª persona, 1.ª persona plural, 2.ª persona, 3.ª persona singular, pronombres, verbos —pasado, presente, futuro—, adverbios, negaciones y otros).

⁴ Bolsa de palabras relacionadas con un tema que sirven como un filtro para detectar aquellos textos de tuits que mencionen dichas palabras.

Cuadro 1

Algunos recursos léxicos marcados por emociones en el idioma español

Diccionario	Clases	Creado por	Idioma	Verificado manual	Referencia
<i>SEL Spanish Emotion Lexico</i>	Tristeza, alegría, ira, enojo, sorpresa, desagrado.	Camacho, V., G. Sidorov, & N. Galicia (2016).	Español	Sí	https://www.cic.ipn.mx/~sidorov/
<i>NFC National Research Council Canada</i>	Alegría, sorpresa, confianza, miedo, ira, tristeza, aversión, anticipación.	Saif Mohammad and Peter Turney.	Inglés, traducido a 40 idiomas.	No	https://saifmohammad.com/WebPages/lexicons.html
<i>Linguistic Inquiry Word Count (LIWC)</i>	Ansiedad, tristeza, enfado, emoción negativa, emoción positiva.	Pennebaker, J. W., R. L. Boyd, K. Jordan & K. Blackburn (2015).	Inglés, español, alemán, holandés noruego, italiano y portugués.	Sí	https://liwc.wpengine.com/

Un ejemplo del uso de *LIWC* es una herramienta expuesta en la web *Analyze Words*,⁵ la cual arroja un indicador que mide el aspecto emocional, social y racional del tuitero con base en sus últimas 25 publicaciones teniendo solo el nombre de la cuenta pública del usuario.

Datos y su procesamiento

Este trabajo se realizó usando los tuits públicos y georreferenciados dentro del territorio nacional recolectados por el INEGI desde el 2017 al 2020. Se cuenta con alrededor de 2 millones de tuits por mes. A partir de estos datos, detectamos a todos los usuarios que hayan publicado más de 25 do-

cumentos por año, identificando así a aproximadamente 600 mil. La gráfica 1 muestra la distribución por año. Notamos que la captación de usuarios va disminuyendo a través del tiempo.

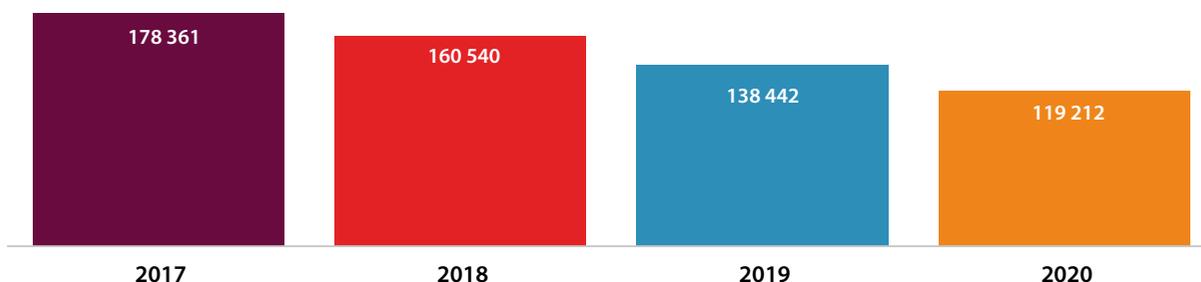
Se adquirió el *LIWC*, programa ejecutable diseñado para hacer los conteos de palabras pertenecientes a cada clase cuando se imputa una serie de documentos, sin embargo, el formato y la dimensión de los tuits a estudiar diferían de las del programa, por lo que fue necesario crear nuestro propio código para la lectura y conteo. Decidimos realizarlo con dos tipos *software*, *Python* y *R*, con la finalidad de llevar a cabo una comparación de resultados y garantizar la calidad de estos.

Además, se implementó un ambiente específico para el almacenamiento y las consultas bajo *Spark*

5 <https://www.analyzewords.com/science.php>

Gráfica 1

Número de tuiteros que publicaron más de 25 tuits por año



Fuente: INEGI. *Estado de ánimo de los tuiteros en México* (DE) <https://www.inegi.org.mx/app/animotuitero/#/app/multiline>

y *Apache Zeppeling*, pasando de un formato muy general, como *json*, a archivos *Parquet*, logrando disminuir significativamente la velocidad de consultas.

Como sabemos, la limpieza y el preproceso de datos es una etapa importante en cualquier estudio donde se analiza información no estructurada: quitar caracteres y símbolos especiales ayuda a eliminar ruido cuando se requiere interpretar resultados y reducir el texto a palabras básicas, como *tokens* y lemas, y facilita su conteo. Para nuestro estudio, eliminamos enlaces http, signos de puntuación, espacios innecesarios, emojis, etiquetas (#), menciones (@), y *tokenizamos*.⁶ Ni los números ni *stopwords*⁷ fueron omitidos, pues estos los contabiliza *LIWC* para diagnosticar la forma de escritura del usuario.

Se procesaron 124 627 640 tuits de 600 mil usuarios desde enero del 2017 a noviembre del 2020. Con estos datos se construyó una matriz de infor-

mación donde cada registro es un usuario y las columnas están dadas por características que representan sus textos (ver cuadro 2).

Para la identificación del municipio correspondiente a cada tuit, se utilizó el elemento *bounding_box*, que contiene un mínimo de cuatro pares de puntos geográficos (latitud, longitud) que hace referencia a la zona geográfica donde el texto fue creado; con estos se genera un polígono que será identificado por su valor central (centroide). Se realizó un cruce de estos valores con los polígonos que identifican a cada municipio del país para asociarlo con el tuit de acuerdo con el polígono donde cae el centroide. Con este método logramos 99 % de clasificación correcta de municipios a todos los tuits.

Aunque tenemos el conteo de palabras de todas las clases de *LIWC* por cada tuitero, este trabajo solo se concentra en la clase *ansiedad*; algunas palabras con esta etiqueta se muestran en el cuadro 3. Sin duda, esta matriz de información será de gran utilidad para investigaciones más robustas futuras al presente estudio.

6 Proceso de segmentar el texto en unidades lingüísticas: palabras, signos, números, alfanuméricos, etcétera.

7 Palabras muy comunes y poco informativas desde el punto de vista léxico, por ejemplo, conjunciones (y, o, ni, que), artículos (la, los, el), preposiciones (a, en, por, etcétera).

Cuadro 2

Información obtenida para cada usuario

Concepto	Descripción
id usuario	Identificador único del usuario.
N_word_clase_i	Número de palabras para cada clase de <i>LIWC</i> .
Tweet_positivos	Número de tuits positivos según ánimo tuitero (AT).
Tweet_negativos	Número de tuits negativos según AT.
Tweet_neg_ansi	Número de tuits negativos y que tenga al menos una palabra de cada categoría de <i>LIWC</i> .
Tweet_pos_ansi	Número de tuits positivos y que tenga al menos una palabra de cada clase <i>i</i> .
Clase_i	Conteo de palabras para la clase <i>i</i> de <i>LIWC</i> .
Clase_Ansiedad	Conteo de palabras relacionadas con <i>ansiedad</i> .
N_tweet	Número de tuits.
N_word	Número de palabras totales.
N_word_liwc	Número de palabras captadas por <i>LIWC</i> .
Entidad	Entidad federativa donde el tuitero publica la mayoría de sus tuits.
Municipio	Municipio donde el tuitero publica la mayoría de sus tuits.
Mes	Mes en el que el tuitero realiza sus documentos.

Palabras relacionadas a la clase *ansiedad* en *LIWC*

abruma*	apremia*	chocad*	desconcert*	enoj*	falte*	inadecuad*	miserable	preocupad*	sobresalta	sobresalté	temerari*	trauma
	aprensi*	compulsivo	desdichad*	escandal*	faltó	inciért*	nervios*	preocupé	sobresaltáis	sobresaltó	temid*	turba
abyect*	apuro*	confundi*	deseos*	espanto*	<i>fear</i>	incómod*	obsesión	preocupe*	sobre-saltamos	suda*	temiendo	turbad*
aflicci*	<i>ashamed</i>	confus*	desequilib*	esperpento*	histéric*	incomodo	obsesion*	preocupo	sobresaltan	sudando	temor*	turbar
afligi*	asust*	<i>confused</i>	desespera*	estres	horrend*	inquieta*	pánic*	preocupó	sobresaltando	sudar*	tensa*	<i>upset</i>
agita*	aterrad*	crisis	desgana*	estrés	horrible*	inquieta*	pavor	presión*	sobresaltar	sudó	tensión	urgencia
alarm*	aterroriza*	critic	deshonra*	estresante	horror	insegur*	peligr*	presiona*	sobresaltará*	suscep-tible*	tension*	vacilante*
angusti*	aturdid*	<i>critic</i>	desorde*	estruja*	horroriz*	intimid*	pena*	problema	sobresaltare*	susto*	tenso*	vengue*
ansiedad*	aturdi-miento	depresión	<i>disturb</i> *	estupefact*	humilla*	<i>late</i>	petrifica*	reacia*	sobresaltare*	temblar*	terco*	vergonzos*
anticipa*	aturdir	depresiones	embarazos*	excitable	impacien*	mello	prejuicio*	reacio*	sobresaltaron	temblor*	terror*	vergüenza*
aplstar	aversi*	deprimid*	emoción	faltaron	impotencia	miedo*	preocupa	riesgo	sobresaltas	teme	tímid*	vulnerable*
aprehensión	<i>breakdown</i>						preocupaci*	sacud*	sobresaltast*	temer	torcedura*	<i>worried</i>

Nota: las palabras con * son raíces mejor conocidas como *steam* (en inglés), i. e., *angusti* es la raíz de varias palabras: angustiar, angustiado, angustia, etcétera.

Análisis exploratorio de datos

Una vista general del comportamiento de la información a través del tiempo por medio de gráficas puede ayudar a obtener conocimiento a partir de patrones naturales de los propios datos, o bien, buscar elementos que confirmen si las hipótesis o técnicas establecidas al inicio de cualquier estudio son las adecuadas.

La gráfica 2 nos muestra la evolución por año de los tipos de tuits de todos los tuiteros. Como se puede notar, existe una disminución de tuits recolectados en estos cuatro años. Tanto los positivos como los negativos (de acuerdo con la etiqueta que proporciona el clasificador del ánimo tuitero —AT—) decaen, aunque la velocidad de disminución de los últimos es menor que la de positivos. Se aprecia un ligero aumento de *ansioso-negativos*, los cuales son definidos por aquellos que tienen al menos una palabra relacionada con ansiedad y que son clasificados como negativos por el AT. Una disminución de tuits *ansioso-positivos* y el número de palabras relacionadas con ansiedad por año parece constante con un ligero aumento en el 2018. La información que nos da esta gráfica nos lleva a pensar que, aun cuando los tuiteros y los tuits van en disminución, pueden existir subpoblaciones de

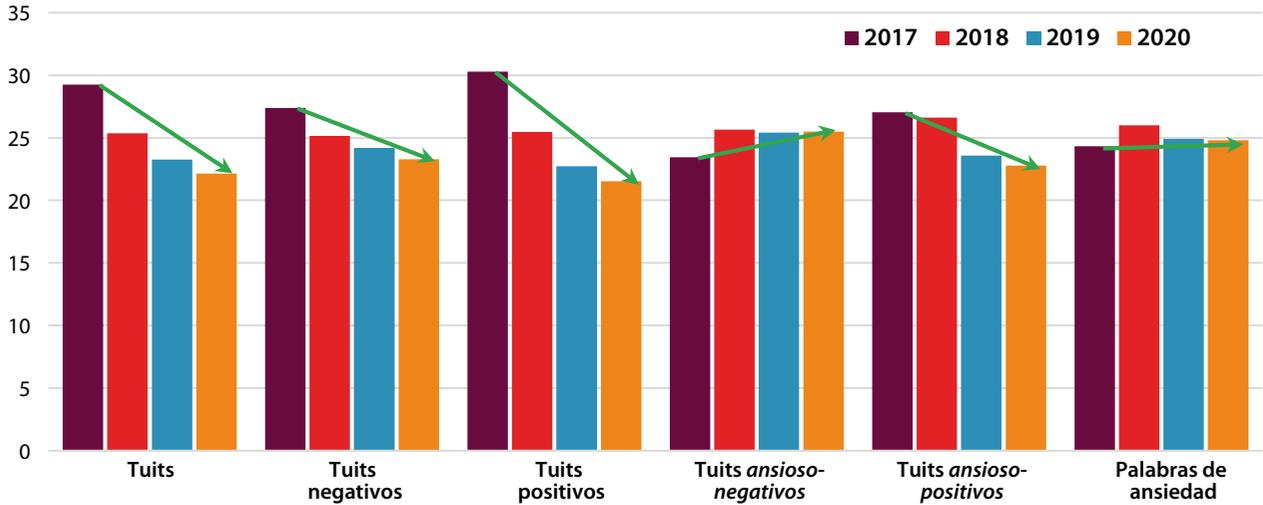
usuarios que expresan cierta temática que no necesariamente sigue la misma tendencia de disminución de la población total, por ejemplo, el caso de los que escriben textos *ansioso-negativos*, hecho que se demuestra más adelante al usar la técnica de aprendizaje supervisado.

Un análisis de los textos, tuiteros y palabras de ansiedad a través de los cuatro años, pero ahora por mes, nos revela ciertas fluctuaciones interesantes. Observamos en la gráfica 3 que la cantidad de tuits positivos siempre rebasa a los negativos y, efectivamente, tienden a disminuir a través del tiempo. Además, de manera general, se aprecia una relación entre estos dos tipos de tuits, si aumentan los positivos, se incrementan los negativos y viceversa. Note que estas fluctuaciones de aumento se dan en determinados meses de cada año.

En la gráfica 4, donde se muestra el conteo de palabras relacionadas con ansiedad por mes y año, se destacan barras más altas en ciertos meses. Resulta muy tentador empezar a relacionar estos aumentos con eventos importantes que sucedieron en cada año, por ejemplo, septiembre del 2017, mes en el que ocurrió un sismo; en mayo-julio del 2018 tuvieron lugar las elecciones presidenciales; enero del 2019, desabasto por gasolina; y marzo

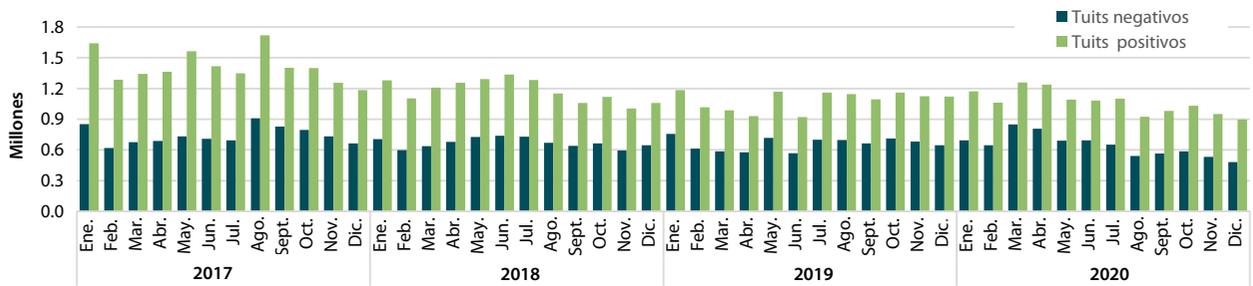
Gráfica 2

Evolución en el tiempo del tipo de tuits (porcentaje)



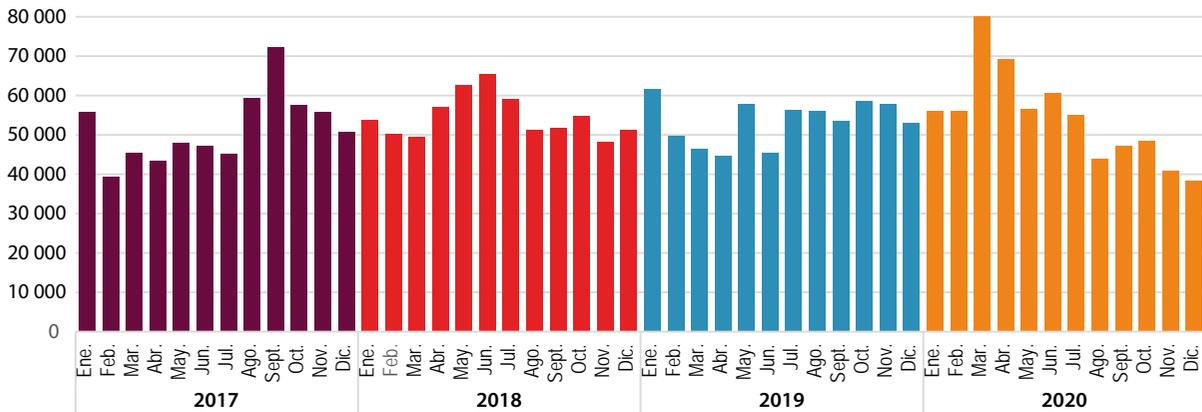
Gráfica 3

Distribución de los tuits negativos y positivos por mes a través de los años



Gráfica 4

Número de palabras relacionadas con ansiedad por mes a través del tiempo



del 2020 empezó el confinamiento en México por la pandemia de COVID-19.

Una pregunta natural que surge después de analizar la gráfica anterior es: ¿hacia qué valencia se inclinan las palabras relacionadas con ansiedad, a la negativa o a la positiva? Note que en la gráfica 5 solo se examinan a los tuiteros que expresaron palabras de ansiedad. Las barras de color verde oscuro representan a quienes publicaron textos clasificados como negativos; las de color verde claro, los positivos. Vemos que existe una incidencia mayor de tuiteros con textos *ansioso-negativos* que *ansioso-positivos*, y la presencia de los primeros se hace más notable en los mismos meses en los que tuvimos eventos que ya mencionamos.

Confirmando lo antes dicho, la correlación estadística asociada de las palabras de ansiedad vs. los textos negativos y positivos de todos los tuiteros por año se presentan en el cuadro 4, aunque existe una mayor correlación hacia lo negativo; la que hay con positivos no es tan baja, pues presenta un valor mayor a 0.6 en todos los años, incluso, en el

2018, estos coeficientes se igualan; por lo tanto, también deben contemplarse en el estudio y es por ello que en la siguiente sección se construye un índice que informa qué tan positivos o negativos son los textos que muestran ansiedad.

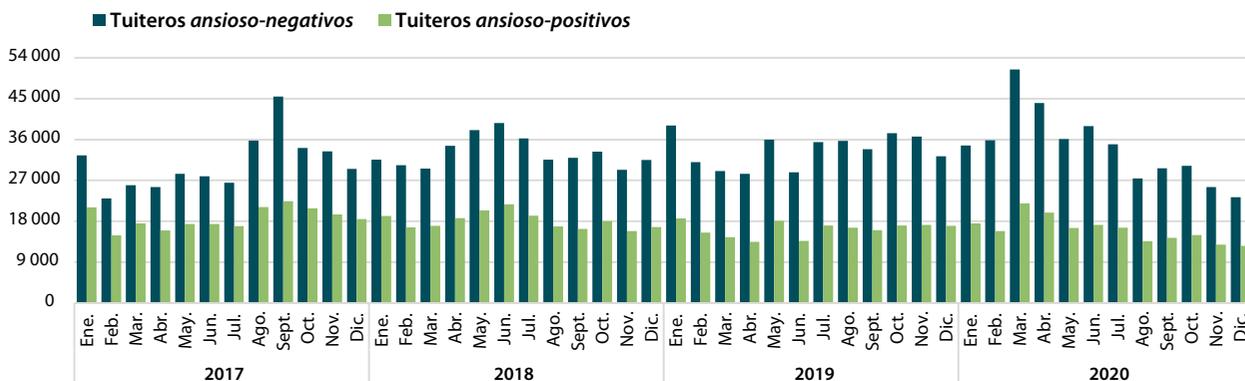
Como resumen, la información de las gráficas anteriores nos muestra que existen meses relevantes donde la población tuitera expresa más negatividad y ansiedad, que existe una tendencia más pronunciada a expresar ansiedad con contexto negativo que positivo y que los tuiteros *ansioso-negativos* van aumentando en cada año a pesar de que la población usuaria es cada vez menor.

Generación del Índice de Ansiedad por tuitero

Para realizar comparaciones entre usuarios y poder diferenciar si sus expresiones sobre ansiedad se encuentran en un contexto positivo o negativo, se pensó en generar un indicador que refleje esas diferencias.

Gráfica 5

Número de tuiteros ansiosos por mes a través del tiempo



Cuadro 4

Correlación de tuits *ansioso-negativos* y *ansioso-positivos* por año

Año	Tuits <i>ansioso-negativos</i>	Tuits <i>ansioso-positivos</i>
2017	0.61	0.56
2018	0.72	0.71
2019	0.76	0.63
2020	0.85	0.63

Asignación de clave en textos para el cálculo del Índice de Ansiedad

Polaridad de textos (AT)	Tuits con al menos una palabra relacionada con ansiedad	Tuits que no tienen palabras relacionadas con ansiedad
Positivos	<i>B</i>	<i>C</i>
Negativos	<i>A</i>	<i>D</i>

Cada usuario identificado en nuestra base de datos tendrá un valor entre -1 y 1, que indica el grado de ansiedad en función de sus textos: si este se acerca a -1, entonces podemos decir que tiende a expresar ansiedad en un contexto negativo; si se acerca a 0, el usuario es neutro en cuanto a su ansiedad; y si se aproxima a 1, entonces la mayoría de sus textos contienen palabras de ansiedad con valencia positiva.

Hicimos una partición excluyente de las clases de AT para cada tuitero, como se muestra en el cuadro 5, de tal manera que *N* representa la suma de todos sus textos.

El Índice de Ansiedad está determinado solo por las cantidades *A* y *B*:

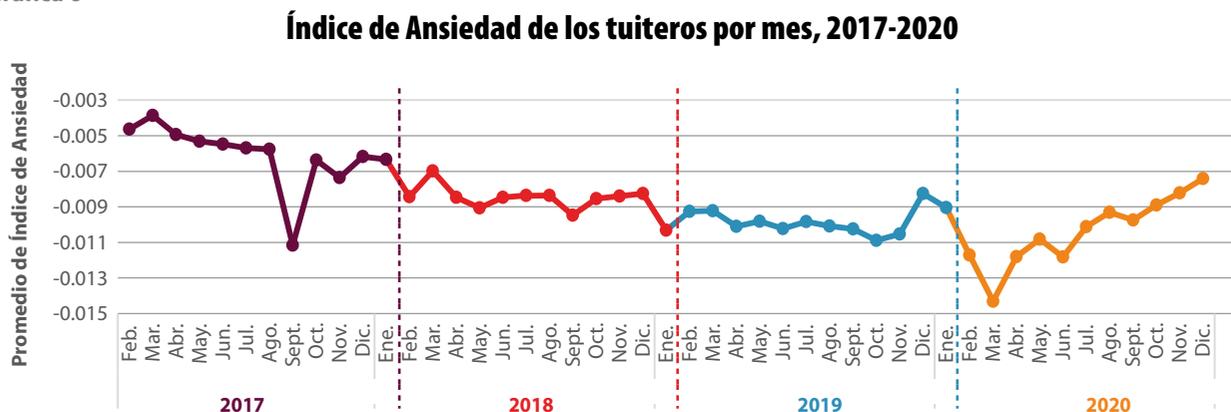
$$I_{ansiedad} = \frac{B}{N} - \frac{A}{N} = \frac{B-A}{N} \quad (1)$$

Note que si $A = 0$ y $B > 0$, entonces todos los textos con palabras de ansiedad del tuitero se clasificaron como positivos; por lo tanto, el Índice tiene un valor de $I_{ans} = 1$. Caso contrario, si $B = 0$ y $A > 0$, entonces no hubo textos *ansioso-positivos* y el Índice es $I_{ans} = -1$.

Un índice igual a 0 indica que el usuario es neutro en ansiedad; sin embargo, esto se puede presentar en dos casos: 1) si *A* es igual a *B*, el usuario expresa ansiedad, pero sus textos positivos igualan a sus negativos; por lo tanto, no tiene un sesgo de ansiedad que muestre una valencia de sus textos y 2) no usa palabras de ansiedad, por lo que *A* y *B* son iguales a 0; este usuario puede expresar gran cantidad de negativos (positivos), pero esa negatividad (positividad) no tiene nada que ver con la ansiedad.

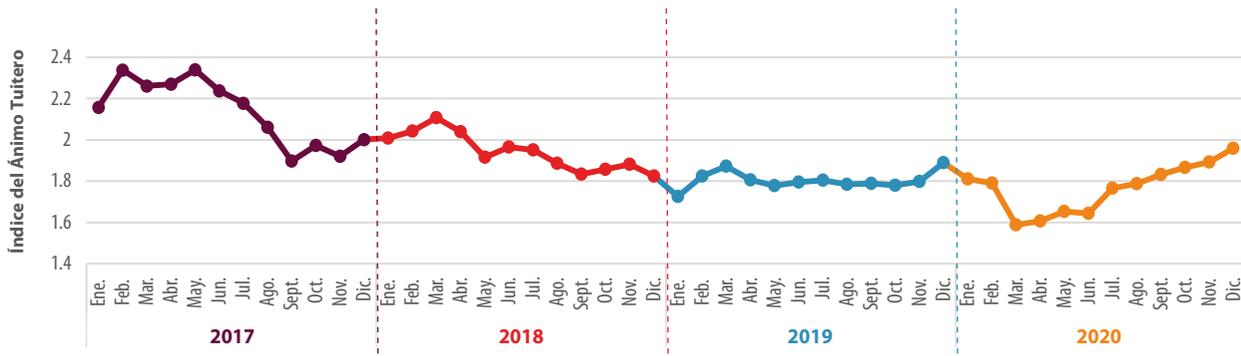
La gráfica 6 representa el promedio del Índice de Ansiedad de los tuiteros por mes desde el 2017 al 2020. Los picos más bajos en toda la serie corresponden a septiembre del 2017 y marzo del 2020. Para fines de comparación, en la gráfica 7 se visualiza el Índice del Ánimo Tuitero para el mismo periodo. Ambas gráficas muestran una tendencia similar, coinciden en ciertos aumentos o disminuciones a pesar de que las unidades de medición son diferentes; por un lado, el Índice del Ánimo Tuitero corresponde a la clasificación del propio texto de los tuits y el de Ansiedad, a la clasificación del tuitero. Ambos oscilan en distintas escalas. Podemos decir que no son excluyentes, pero tampoco

Gráfica 6



Gráfica 7

Índice del Ánimo Tuitero por mes, 2017-2020



complementarios. Los valores más bajos del Índice del Ánimo Tuitero en toda la serie se dan en enero del 2019 y marzo del 2020.

El Índice de Ansiedad ahora se representa por entidad federativa en las gráficas de 8, 10, 11 y 12. Podemos notar diferentes comportamientos para los distintos años; por ejemplo, en el 2017, las entidades con valores más bajos durante septiembre son: Chiapas, Ciudad de México, Guerrero, Morelos, Puebla y Tlaxcala; estas fueron las más afectadas por los sismos sucedidos en ese mes. Para confirmar esta sospecha, nos dimos a la tarea de encontrar las palabras más frecuentes asociadas con ansiedad en dichos estados para septiembre,

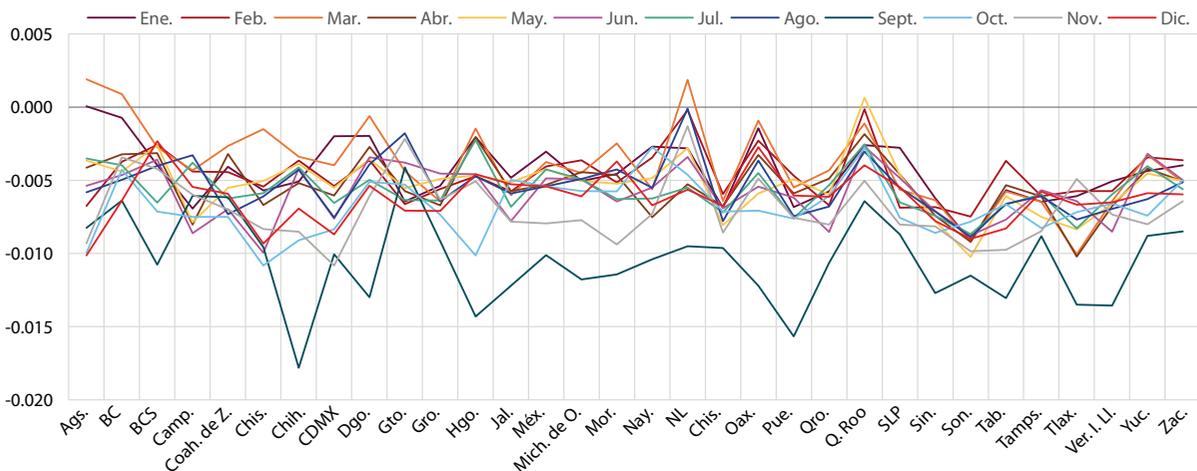
las cuales se muestran en la gráfica 9 a través de nubes de palabras de algunas entidades afectadas.

Se observa que los valores del Índice cambian en los siguientes años; el lector puede identificar los extremos y relacionarlos con algún suceso acontecido en el país. De estas graficas se desprende una interpretación meramente tendencial; para corroborar las asociaciones con eventos, hay que verificarlo profundizando el análisis.

Para el 2020, según la gráfica 11, destaca marzo con los índices más bajos para casi todas las entidades, hecho que claramente podemos relacionar con el inicio de la cuarentena por la pandemia de COVID-19.

Gráfica 8

Índice de Ansiedad por mes según entidad federativa, 2017



Gráfica 9

Top ten de palabras relacionadas con ansiedad en septiembre del 2017 para entidades seleccionadas

Guerrero



Chiapas



Puebla

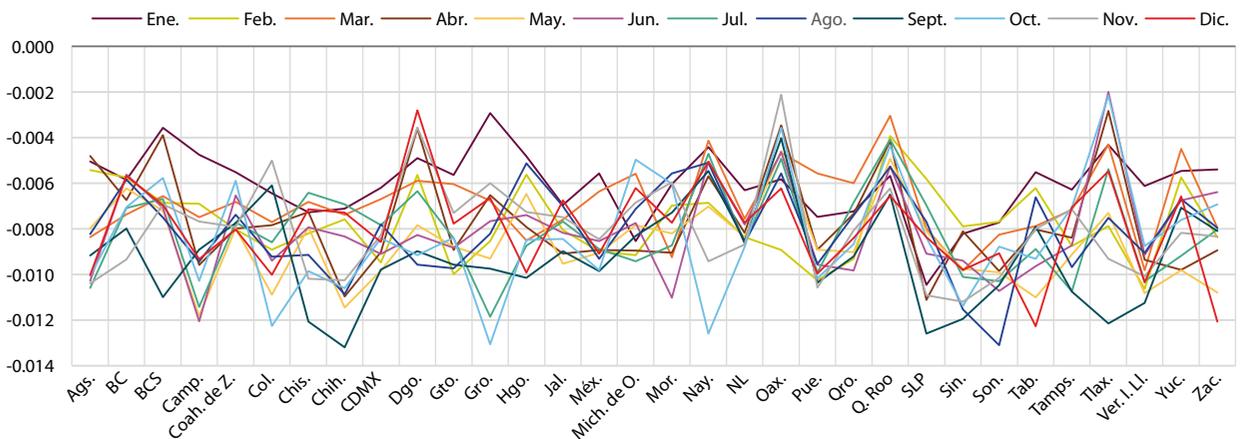


Tlaxcala



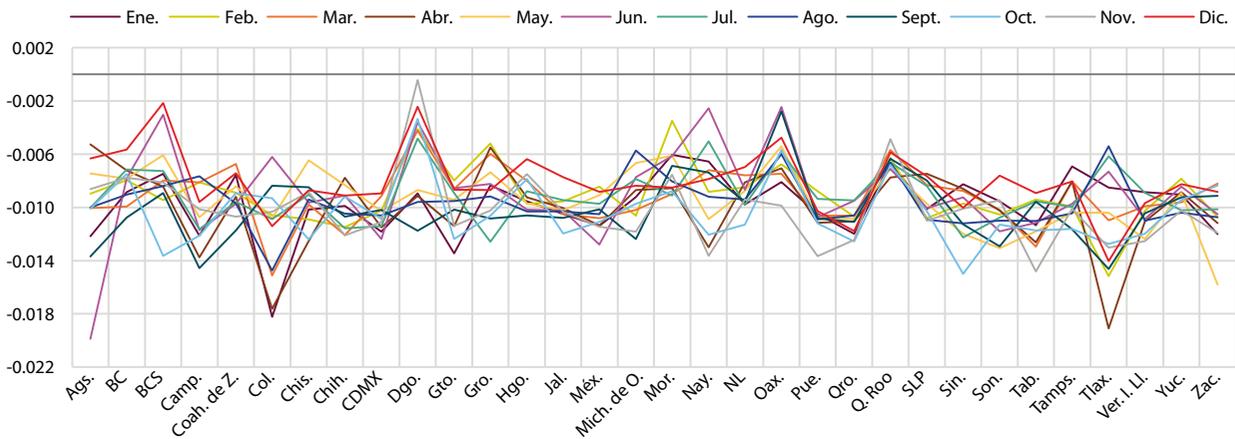
Gráfica 10

Índice de Ansiedad por mes según entidad federativa, 2018



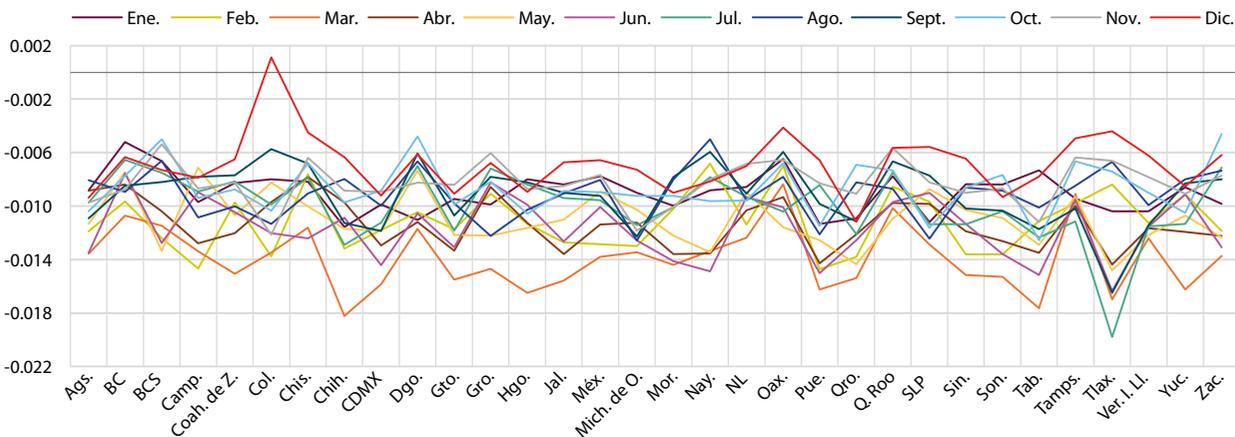
Gráfica 11

Índice de Ansiedad por mes según entidad federativa, 2019



Gráfica 12

Índice de Ansiedad por mes según entidad federativa, 2020



Para diagnosticar la presencia de otras emociones presentes en los tuitos, se tiene la posibilidad de construir índices similares para casi todas las clases de LIWC, de tal manera que puedan ser comparados y correlacionados. En pruebas previas a este análisis, notamos que las clases relacionadas con afecciones positivas más expresadas por los usuarios son *alegría* y *sorpresa*. Sobre las afecciones negativas, se puede decir que la ansiedad está muy relacionada con la emoción *miedo* (tal como lo señalan Piqueras Rodríguez *et al.*, 2009); la *tristeza*, con el *enojo*; y las palabras que expresan *repulsión* suelen ser menos utilizadas por los tuitos.

Si quisiéramos detectar a los usuarios que expresan más ansiedad con más carga negativa en sus textos que los distingue de los demás, usar este índice como referencia no sería lo más adecuado. Imagine a un tuitero que escribió solo un documento *ansioso-negativo*, por lo que su valor es -1; ahora, si existiera otro que escribió 10 documentos *ansioso-negativos* y ninguno *ansioso-positivo*, su índice también es igual a -1. El segundo ha usado más palabras de ansiedad que el primero y, sin embargo, el índice los iguala. La siguiente sección explica la construcción de grupos para diferenciar a este tipo de usuarios.

Construcción de tipos de ansiosos de acuerdo con la formación de grupos

Bajo el supuesto de que cuantas más palabras de ansiedad y más textos negativos escriba un tuitero, este tendrá mayor probabilidad de padecer un riesgo de salud mental, y si esta situación se corrobora con la evaluación cualitativa por parte de los expertos psicólogos, podemos entonces conformar una base de datos etiquetada para futuros estudios. Este es uno de los objetivos que buscamos al perfilar a los usuarios por medio de la técnica estadística *clustering*; el otro es el de medir la prevalencia de este tipo de tuiteros a través de los años, tal como lo sospechamos desde la gráfica 2.

Primeramente, la agrupación fue realizada tomando a todos los usuarios captados en el 2017 con más de 25 tuits. Se usaron tres variables para el agrupamiento: número de tuits *ansioso-negativos*, número de tuits *ansioso-positivos* y número de palabras relacionadas con ansiedad. Probamos diferentes técnicas de agrupamiento como *k-medias* y las basadas en mezcla de funciones gaussianas. Considerando el criterio *Bayesian Information Criterion (BIC)* decidimos quedarnos con seis grupos y usar un modelo con distribución elipsoidal con iguales tamaños, igual forma y diferentes orientaciones (EEV) (Scrucca *et al.*, 2016).

Como resultado de la agrupación, obtenemos las etiquetas que identifican al grupo de pertenencia de cada usuario en ese año; esta categorización se convierte en la base para generar un clasificador que determine las etiquetas de los tuiteros para los años subsecuentes, es decir, el proceso determinará el grupo al que pertenece el tuitero según la semejanza que tiene con los del año base. Por lo tanto, los grupos formados en el 2017 sirven como un conjunto de patrones fijos en los que el modelo se basa para asignar a nuevos usuarios al patrón más similar, de tal manera que sea más sencillo notar el crecimiento o disminución de tuiteros a través de los años bajo un año base. En otras palabras, se usa un método

de aprendizaje supervisado, suponiendo que las etiquetas de los usuarios del 2017 son las que determinan la diferencia de escritura de los usuarios.

Con la información que nos presenta el cuadro 5 es posible caracterizar a cada grupo formado en términos de los promedios de las variables usadas para la agrupación y otras que no fueron usadas. Podemos notar sus similitudes y diferencias.

El grupo 1 está caracterizado por muy pocos tuiteros, tan solo 67 que escriben mucho más texto que los demás; en promedio, cada tuitero escribe más de 3 mil tuits al año categorizados en su mayoría como positivos; sin embargo, tienden a expresar más palabras de ansiedad con una media de 167 por tuitero. Los textos *ansioso-negativos* son casi el doble que sus *ansioso-positivos*.

El grupo 2 está compuesto por tuiteros que, aunque no están en el extremo como en el grupo 1, sí reflejan una negatividad ansiosa en sus textos, con un promedio de 38 palabras de ansiedad por usuario y 24 tuits *ansioso-negativos* vs. 9.4 *ansioso-positivos*.

El grupo 3 es más neutro en las expresiones de ansiedad negativa y positiva.

Por otro lado, el 4 está compuesto por tuiteros que expresan palabras de ansiedad, pero con tendencia positiva, sus tuits *ansioso-positivos* son poco más de seis veces de sus *ansioso-negativos*; podemos llamar a este grupo como los tuiteros *ansioso-positivos*. Aunque pareciera que los grupos 4 y 5 tienen características similares, el algoritmo los diferencia al obtener promedios menores de este último.

Por su parte, en el 6 quedan clasificados la mayoría de los tuiteros, quienes no muestran ansiedad.

Pondremos atención en los grupos 1 y 2, pues creemos que pueden ser más vulnerables a seguir expresando textos negativos con palabras de ansiedad.

El método de clasificación (aprendizaje supervisado) que usamos fue el de análisis discriminante

Cuadro 6

Promedios de las variables por grupo resultado de la agrupación

Grupo	1	2	3	4	5	6
Número de tuitos	67	2 092	10 830	2 194	894	162 284
Tuits	3 449.51	892.61	376.2	728.8	285.58	155.99
Tuits negativos	1 373.01	412.34	120.16	122.05	55.18	51.73
Tuits positivos	2 076.49	480.27	256.04	606.75	230.39	104.26
Tuits <i>ansioso-negativos</i>	93.19	24.27	5.3	2.75	1.68	2.06
Tuits <i>ansioso-positivos</i>	54.78	9.48	4.43	13.58	8.06	1.1
Palabras de ansiedad por tuitero	167.04	38.34	11.1	16.57	12.01	3.21
Índice de Ansiedad	-0.024	-0.02	-0.0097	0.031	0.013	-0.024

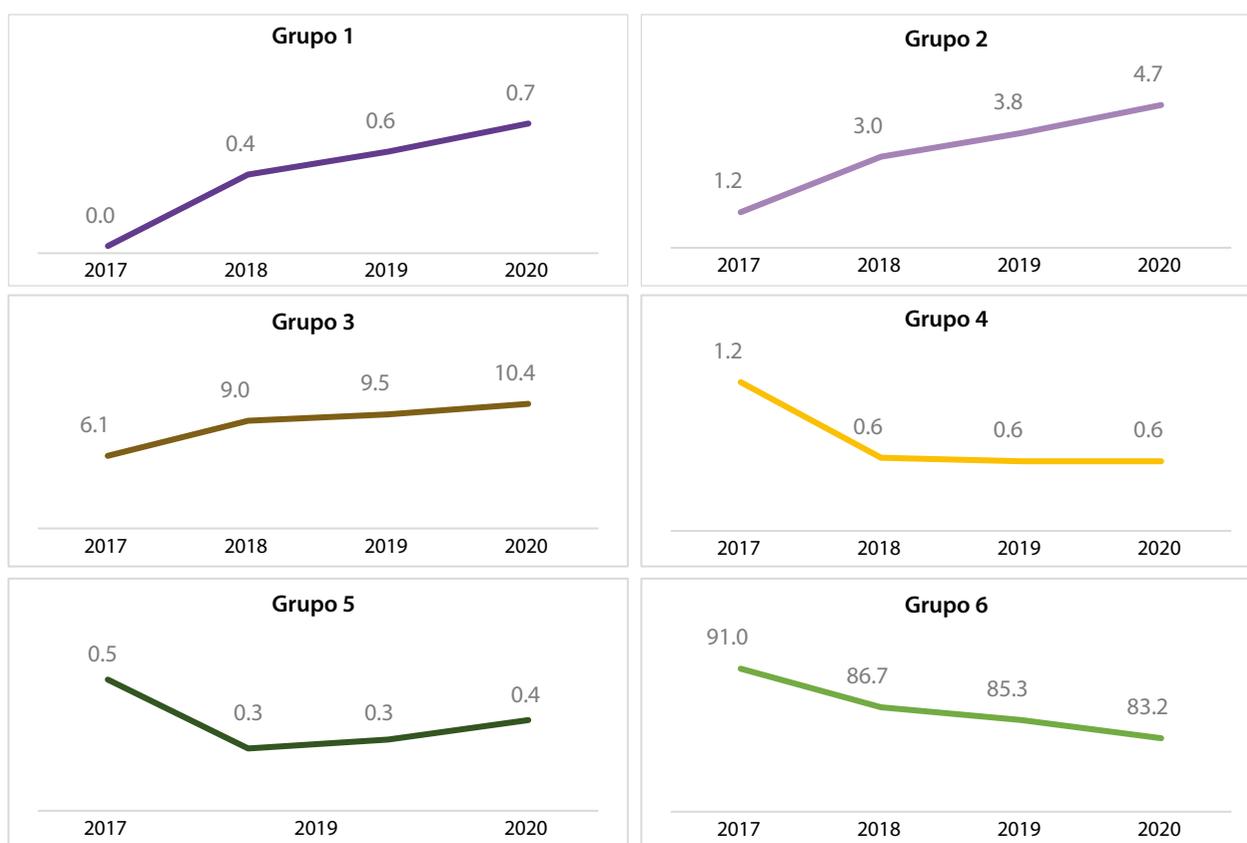
basado en mezclas, donde se asume que la densidad de cada clase sigue una distribución gaussiana. El corpus de usuarios del 2017, que está dado por un total de 178 361, se dividió en 70 % como base de entrenamiento y 30 % de prueba. Para una validación cruzada con $k = 10$ interacciones, ob-

tuvimos un error de predicción del clasificador de 0.030 (0.0012 desviación estándar).

La evolución de los usuarios en cada grupo a través de los años se muestra en las gráficas 13. Para los grupos más ansiosos (1-3), notamos una tendencia

Gráficas 13

Distribución porcentual de los tuitos en cada grupo a través de los años



creciente de tuiteros a través de los años; mientras que para los menos ansiosos (4-6) ocurre lo contrario.

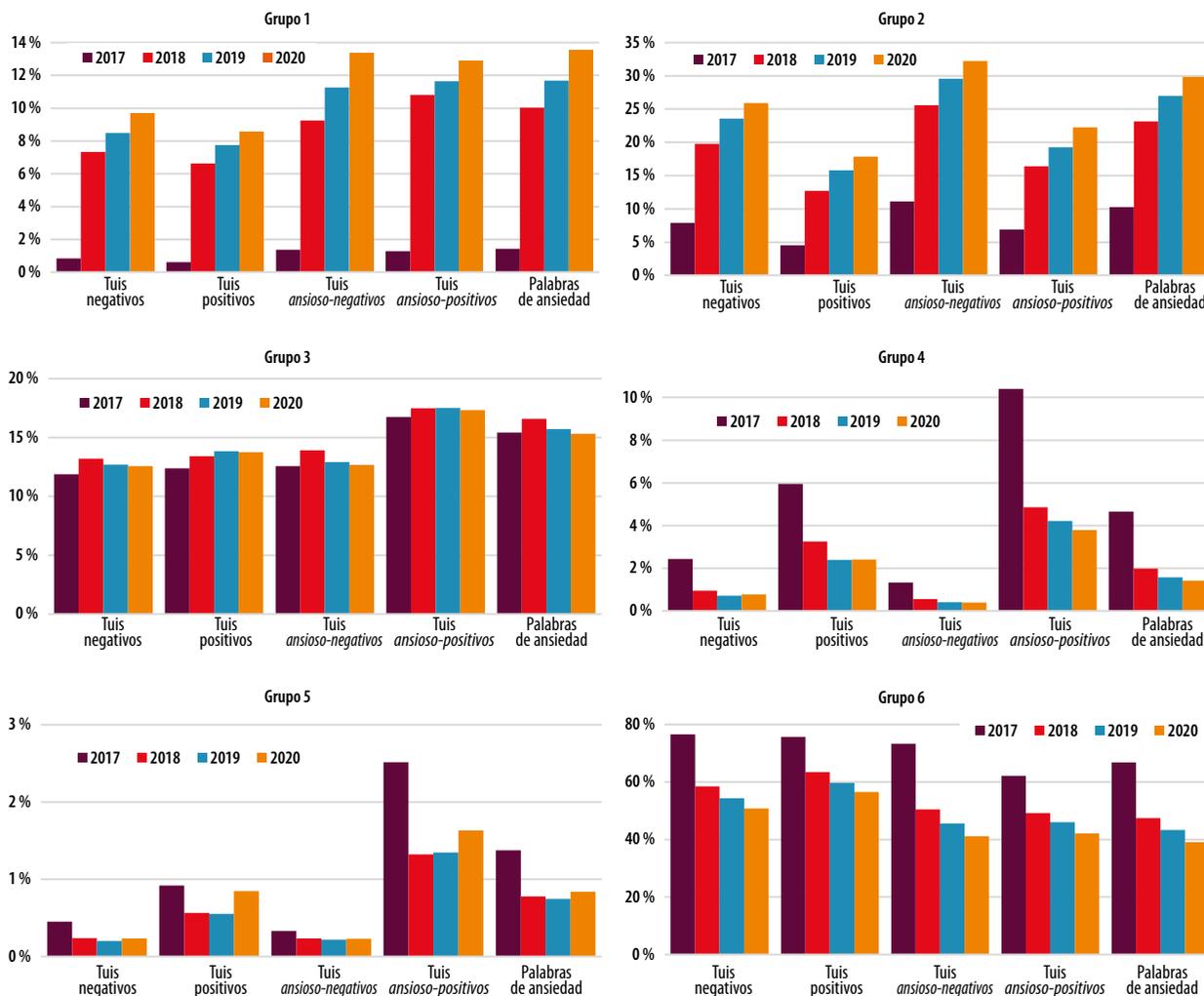
El crecimiento de tuiteros tipo 1 y 2 a través de los cuatro años implica que los tuits *ansioso-negativos* también van en aumento, tal como se refleja en las gráficas 14. Estos dos grupos de usuarios tienen un comportamiento diferente al de todos los tuiteros de nuestra muestra contemplados en la gráfica 2.

Aunque existe un aumento de usuarios clasificados en el grupo 3 en los cuatro años, el por-

centaje de sus tipos de textos se mantiene casi constante; es decir, no hay cambios identificables en la tendencia de escritura de los usuarios de este grupo de un año a otro. Por otra parte, los categorizados en el grupo 4 en el 2017 son más *ansioso-positivos*, sin embargo, quienes se clasificaron en este mismo grupo durante el 2020 se caracterizan por escribir con menos ansiedad positiva. Algo similar sucede para el 5. Es importante resaltar que los tuiteros del 6 son cada vez menos, por lo tanto, hay una disminución de sus tipos de textos.

Gráficas 14

Porcentaje de los tuiteros en cada grupo a través de los años



Conclusiones y consideraciones

Como ya lo comentamos en la introducción, la disposición de datos de redes sociales atrae a los investigadores en distintas áreas precisamente por la gran variedad de temas que abarcan para tener conocimiento sobre gustos, eventos, opiniones y sentimiento de los usuarios prácticamente a un costo cero, a diferencia de los obtenidos de fuentes oficiales, como encuestas y censos. Sin embargo, emplear los de las redes no proporciona un cierto grado de confiabilidad en las estimaciones o predicciones que se realicen, situación que sí se tiene con los datos oficiales; es decir, existen ventajas notables al usar datos abiertos, pero también hay riesgos metodológicos y éticos. Lo ideal es implementar estrategias metodológicas para combinar estos dos tipos de fuentes, como lo están haciendo ya los organismos nacionales e internacionales.

Aun así, a pesar de todo lo que se está haciendo al respecto y que se planea hacer para combinar datos tradicionales con no tradicionales, en este ejercicio, donde examinamos la ansiedad como una emoción intrínseca a un problema de salud mental y cómo es el comportamiento de una muestra relativamente grande de usuarios de *Twitter* referente al uso de palabras relacionadas con ansiedad y la negatividad de sus textos, establecemos una estadística a través de datos no tradicionales que muestra indicadores en el tema de salud pública y bienestar social que puede interesar a las autoridades gubernamentales, a profesionales o planificadores de la salud y organizaciones no gubernamentales implicadas o interesadas en proporcionar servicios de salud mental como evidencia para la generación de programas o políticas de salud pública.

Bajo una perspectiva de estudio basada en el análisis cuantitativo y automático de datos, observamos en nuestros resultados:

- Tendencias de ansiedad a nivel poblacional en ciertos periodos coincidentes con eventos desafortunados en el tiempo, como el sismo del 2017 en algunas entidades del país y la ansie-

dad mostrada en el inicio de la cuarentena por COVID-19.

- El Índice de Ansiedad proporciona un indicador de alerta del bienestar social.
- La prevalencia creciente de usuarios *ansioso-negativos* a través del tiempo.
- Diagnóstico inicial de posibles usuarios afectados que serán etiquetados por expertos para la conjunción de una base de datos que sirva para realizar predicciones del grado de ansiedad de otros usuarios.
- Los resultados complementan aún más la información que muestra el ánimo tuitero sobre la salud mental de la población tuitera en México.

Este estudio abre la puerta para el análisis de lenguaje asociado a otros trastornos mentales considerando las estrategias que han empleado otros investigadores al usar *LIWC* como diagnóstico de personalidades de una población de usuarios o de personas en lo individual.

Contamos con una estructura de información bastante amplia para robustecer esta investigación. Algunos trabajos futuros que podríamos explorar: identificar palabras usadas con frecuencia que acompañen a otra afección emocional negativa y observar la diferencia de uso de esas palabras con las emociones antagónicas; comprobar si existe un patrón similar en nuestro país con respecto a otras naciones donde los estudios indican que las personas deprimidas (Argamon, 2005) tienden a usar más pronombres en primera persona, mientras que las que reflejan miedo en sus textos son propensas a expresar más palabras en tercera persona; identificar las cinco grandes (*big-five*) diferentes personalidades de los usuarios: neuroticismo, extroversión, apertura a la experiencia, conciencia y afabilidad (Costa y McCrae, 1992); así como generar más indicadores emocionales según las clases que detecta *LIWC* y correlacionarlos para una detección más robusta de salud mental.

Fuentes

Brooks, David. "How Artificial Intelligence Can Save Your Life. The machines know you better than you know yourself", en: *The New York Times*. 24 de junio de 2019.

- Coppersmith, Glen, Mark Dredze y Craig Harman. "Quantifying Mental Health Signals in Twitter", en: *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Baltimore, Maryland, USA. Association for Computational Linguistics, 2014, pp. 51-60.
- Costa, P.T. Jr., & R. R. McCrae. Four ways five factors are basic", en: *Personality and Individual Differences*. 13, 1992, pp. 653-665 (DE) Google Scholar.
- De Choudhury, Munmun, Michael Gamon, Scott Counts y Eric Horvitz. "Predicting depression via social media", en: *Seventh Int. AAAI Conf. Weblogs Soc. Media. Proceedings of the International AAAI Conference on Web and Social Media*. 7(1), 2013, pp. 128-137 (DE) <https://ojs.aaai.org/index.php/ICWSM/article/view/14432>
- Díaz Rangel, Ismael, Grigori Sidorov y Sergio Suárez-Guerra. "Creation and evaluation of a dictionary tagged with emotions and weighted for Spanish", en: *Onomázein*. Núm. 29, junio. Santiago de Chile, Pontificia Universidad Católica de Chile, 2014, pp. 34-46.
- Fontaine, Johnny R. J., Klaus R. Scherer, Etienne Roesch y Phoebe C. Ellsworth. "The World of Emotions Is Not Two-Dimensional", en: *Psychological Science*. 18, núm. 12, 2017, pp. 1050-1057.
- Layard, Richard. *The economics of mental health" With modern psychological therapy, mentally ill people can become more productive and more satisfied with life*. LSE, UK and IZA, German, IZA World of Labor, 2017.
- Mohammad, Saif M. y Peter Turney. *Emotions Evoked by Common Words*. Ottawa, Ontario, Canada, K1A, Institute for Information Technology, National Research Council Canada, 2010.
- Nadeem, Moin, Mike Horn, Glen Coppersmith, Johns Hopkins University y Sandip Sen. *Identifying depression on Twitter*. PhD, University of Tulsa, 2014.
- OMS. *Depresión*. OMS, 2021 (DE) https://www.who.int/health-topics/depression#tab=tab_1, consultado el 28 de febrero de 2021.
- Piqueras Rodríguez, José Antonio, Victoriano Ramos Linares, Agustín Ernesto Martínez González y Luis Armando Oblitas Guadalupe. "Emociones negativas y su impacto en la salud mental y física", en: *Suma Psicológica (Fundación Universitaria Konrad Lorenz)*. 16, núm. 2, diciembre de 2009, pp. 85-112.
- Reece, Andrew G. y Chris M. Danforth. "Instagram photos reveal predictive markers of depression", en: *EPJ Data Science*. 6, núm. 15, 2017.
- Rodríguez-Esparza, Luz Judith, Diana Barraza-Barraza, Jesús Salazar-Ibarra y Rafael Gerardo Vargas-Pasaye. "Índice de riesgo al suicidio en México utilizando Twitter", en: *Revista de Ciencias de la Salud*. 5, núm. 15, 2016, pp. 1-13.
- Ruckenstein, Minna Susanna, and Natasha Dow Schull. "The datafication of health", en: *Annual Review of Anthropology*. 46, 2017, pp. 261-278.
- Scrucca, Luca, Michael Fop, T. Brendan Murphy y Adrian E. Raftery. "Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models", en: *The R Journal*. 8, núm. 20, 2016, pp. 5-33.
- Stirman, Shannon Wiltsey y James W. Pennebaker. "Word use in the poetry of suicidal and nonsuicidal poets", en: *Psychosomatic Medicine*. 63, núm. 4, 2001, pp. 517-522.
- Topol, Eric. *Deep Medicine, How Artificial Intelligence Can Make Healthcare*. New York, Basic Books, 2019.
- Tsugawa, Sho, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh y Hiroyuki Ohsaki. "Recognizing Depression from Twitter Activity", en: *33rd Annual ACM Conference on Human Factors in Computing Systems*. Editado por ACM Digital Library, 2015, pp. 3187-3196.