

Datos *online* no-solicitados en Twitter y toma de decisiones en políticas públicas sobre el uso de sustancias psicoactivas en Colombia

Peña, E. y Fontes, C. (2021). Datos *online* no-solicitados en Twitter y toma de decisiones en políticas públicas sobre el uso de sustancias psicoactivas en Colombia. *Revista Cultura y Droga*, 26(32), 213-233. <https://doi.org/10.17151/culdr.2021.26.32.10>

Enrique Peña*
Carlos Fontes**

Recibido: 16 de diciembre de 2020
Aprobado: 6 de mayo de 2021

Resumen

Hoy en día observamos cómo muchas de las actividades humanas cotidianas se han trasladado desde el mundo material al mundo digital y con ellas los datos que generan. Entre estas nuevas relaciones digitales se destacan las interacciones comerciales, nuevas relaciones laborales derivadas del incremento del trabajo virtual remoto y novedosas formas de relacionamiento y convivencia social facilitadas por una amplia gama de tecnologías, entre ellas las redes sociales como Facebook, Twitter e Instagram. Cada vez que se tiene acceso a más y mejor tecnología al servicio de la sociedad, los Estados y organismos privados pueden verse beneficiados por este nuevo paradigma a la hora de cumplir sus objetivos sociales. Estamos en una era en donde el énfasis se trasladó de la generación de datos hacia el procesamiento de estos en grandes volúmenes. Entre las distintas herramientas que generan datos *online* no solicitados y para el análisis del presente escrito se destaca la plataforma Twitter,

* Politólogo UNAL; Especialización en Política de Drogas, Regulación y Control, Universidad de la República, Uruguay. Bogotá, Colombia. E-mail: enriquep36@hotmail.com.  orcid.org/0000-0001-6646-2580.

Google Scholar

** Economista Universidad de Brasilia; Especialización en Política de Drogas, Regulación y Control, Universidad de la República, Uruguay; Especialización en Políticas de Drogas, Salud y Derechos Humanos, CIDE Región Centro – México. Montevideo, Uruguay. E-mail: kdufontes@gmail.com.  orcid.org/0000-0002-4729-9105.

Google Scholar



red social de uso gratuito de comunicación bidireccional que ha sido el centro de distintos ejercicios políticos a lo largo del mundo, usada por muchos políticos de Occidente y que se caracteriza por tener una interfaz de consulta masiva de datos, apta para centros que pretenden análisis social.

En el siguiente documento se pretende evidenciar a partir de dos ejemplos prácticos (análisis de Word Cloud y análisis de sentimientos), como el uso del Application Programming Interface (API) de Twitter puede arrojar información útil para mejorar la toma de decisión en políticas públicas, aplicable a temas específicos de difícil enfoque como por ejemplo el uso de sustancias psicoactivas en la sociedad. Los datos obtenidos, debidamente procesados, permitirán un mejor contexto de partida para la toma de decisiones, contexto que suele estar ensombrecido por las dinámicas propias del prohibicionismo en las políticas de drogas.

Palabras clave: *big data*, política de drogas, Twitter, datos *online*, toma de decisiones.

Unsolicited online data on Twitter and decision-making in public policies on the use of psychoactive substances in Colombia

Abstract

Today it can be observed how many of the daily human activities have been transferred from the material world to the digital world and with them the data they generate. Among these new digital relationships are commercial interactions, new labor relationships derived from the increase in remote virtual work and new forms of social and coexistence relationships facilitated by a wide range of technologies, including social networks such as Facebook, Twitter and Instagram. Every time there is access to more and better technology at the service of society, States and private organizations can benefit from this new paradigm when it comes to meeting their social objectives. This is an era in which the emphasis shifted from data generation to data processing in large volumes. Among the different tools that generate unsolicited online data and for the analysis of this document, the Twitter platform stands out, a social network for free use of bidirectional communication that has been the center of different political exercises around the world, used by many politicians from the West and characterized by having a massive data query interface suitable for centers that seek social analysis.

This document aims at demonstrating, from two practical examples (Word Cloud analysis and sentiment analysis), how the use of the Application Programming Interface (API) of Twitter can provide useful information to improve decision-making in public policies, applicable to specific issues that are difficult to approach, such as the use of psychoactive substances in society. The data obtained, duly processed, will allow a better starting context for decision-making that is usually overshadowed by the dynamics of prohibitionism in drug policies.

Key words: *big data*, drug policy, Twitter, online data, decision making.

Justificación y marco conceptual

La toma de decisiones en políticas públicas

Según Subirats y Dente (2014), los actores en políticas públicas están definidos como aquellos que llevan a cabo las acciones relevantes dentro de un proceso de toma de decisiones. Conceptualmente hay que diferenciarlos de los *stakeholders*, definidos por los autores como los sujetos interesados o implicados en algún tema de política pública, pero que no tienen actuación directa sobre el proceso de toma de decisiones. En este sentido, los autores sintetizan que “no son actores aquéllos que están interesados en actuar, ni tampoco aquéllos que deberían actuar, sino sólo aquellos que actúan”. Aún cabe destacar que los actores no son solo aquellos que deberían actuar en función de su estatus legal dentro del proceso de toma de decisiones, sino todos aquellos que actúan con el fin de obtener un resultado, sean ellos sujetos individuales o colectivos.

En relación con los actores colectivos, Subirats y Dente (2014) definen que para que un actor colectivo pueda ser considerado como tal, es necesario que este pueda tener preferencias propias, distintas a las de los individuos que lo constituyen. En este sentido, es fundamental que los intereses del grupo estén bien definidos y que exista un conjunto de reglas, formales o informales, que tiene que cumplir quien actúe en nombre del actor colectivo. Como señalan los autores, rescatando la distinción conceptual elaborada por Bennett-Segerberg (2012) entre acción conectiva y acción colectiva, algunas acciones conjuntas, como la intervención

en forma de acción viral en las redes sociales, difícilmente cumplen con las condiciones necesarias para definir las conceptualmente como acción conjunta de un actor colectivo. Sin embargo, las redes sociales se han convertido en un medio donde actores relevantes pueden difundir sus intereses y opiniones de forma más rápida y con mayor alcance, lo que convierte a las redes sociales en un importante medio de actuación de los actores en políticas públicas.

Definido el concepto de actores en políticas públicas, pasamos a analizar brevemente cómo se da el proceso de toma de decisiones. Hammond *et al.* (2000) sugieren el método PROACT de toma de decisiones como el más efectivo, y que se constituye por ocho elementos principales: problema, objetivos, alternativas, consecuencias, transacciones, incertidumbre, tolerancia del riesgo y decisiones vinculadas. En resumen, hay que definir un problema con precisión, especificar los objetivos, crear alternativas posibles para la resolución del problema planteado, entender las consecuencias de cada alternativa, estudiar las contradicciones y los *trade-off* inherentes a la toma de decisiones, aclarar las incertidumbres, considerar el grado de tolerancia al riesgo, y considerar todas las decisiones vinculadas en diferentes períodos del tiempo. Dado el grado de complejidad de la toma de decisiones en políticas públicas, algunos de estos elementos del método PROACT toman mayor relevancia, como la incertidumbre. Es fundamental contar con información abundante y relevante para mitigar los riesgos asociados a la incertidumbre y maximizar la racionalidad del proceso de toma de decisiones en políticas públicas.

Bracho (2010) define la política basada en evidencia como el uso explícito e intencional de la mejor evidencia de investigación disponible para la toma de decisiones y la elección entre opciones de política pública. Esta corriente teórica surge como respuesta a la demanda creciente, según la autora, por sustentar las acciones y decisiones públicas en un marco racional y con respaldo en evidencia científica. Esto exige la creación de sistemas de información y evaluación que generen evidencia relevante continua, lo que aumenta el costo y el nivel de complejidad de esta relación entre investigación y políticas públicas. Bracho (2010) identifica, de este modo, cinco categorías de políticas públicas en función de su relación con la evidencia científica: a) políticas que no toman en cuenta la evidencia; b) políticas que están al tanto de la existencia de la evidencia pero no la aplican; c) políticas que consideran la evidencia pero no están totalmente moldeadas por ella; d) políticas que se perciben con influencia de la evidencia; e) la política que está moldeada y

basada en la evidencia intencionalmente buscada, relativa a relaciones causales entre propósitos buscados, opciones de acción y resultados.

La importancia de los datos para las políticas públicas queda más evidente en la etapa de evaluación y monitoreo de los resultados, aunque la política basada en evidencia recomienda la incorporación de la evidencia científica, ya en la etapa de diseño e implementación de las políticas públicas. En el caso concreto de las políticas de drogas, el Consorcio Internacional sobre Políticas de Drogas (IDPC en su sigla en inglés), en su *Guía sobre política de drogas* (2012), sugiere que para garantizar que las políticas y estrategias en materia de drogas sean adecuadas, deben basarse en una evaluación objetiva de prioridades y pruebas empíricas. La guía concluye en este sentido que dada la complejidad de los factores que influyen en los niveles y patrones de producción y consumo de sustancias psicoactivas, los responsables de adoptar políticas deben buscar constantemente mejores respuestas, basándose en pruebas empíricas y experiencias, y no por la influencia de ideologías e intereses políticos.

Sin embargo, la recolección de evidencia científica en el campo de las drogas no es una tarea simple debido a las restricciones impuestas por la normativa tradicionalmente prohibicionista. Los datos relacionados al mercado de drogas provienen muchas veces de muestras no contrastadas, como es el caso de los datos vinculados exclusivamente a incautaciones realizadas por la policía, datos relacionados con usuarios problemáticos que recurren a los sistemas de asistencia, datos sobre procesamientos judiciales y datos de carácter auto declaratorio provenientes de las encuestas oficiales que en ocasiones privilegian la estigmatización sobre el nivel de prevalencia del consumo de sustancias psicoactivas en un sector determinado de la sociedad.

A pesar de eso, los investigadores sociales han desarrollado varias estrategias metodológicas para estudiar el fenómeno de las drogas a lo largo del tiempo. Son estrategias metodológicas de distinta naturaleza, cada una con sus limitaciones, pero que combinadas pueden ser útiles para la comprensión global del fenómeno. Pérez y Quintana (2003) resumen las principales estrategias metodológicas, diseños de estudio y técnicas de investigación relacionadas con el estudio del consumo de sustancias psicoactivas. Desde un enfoque cualitativo se destacan los relatos y narraciones, los estudios especiales con técnicas cualitativas y estrategias de información rápida. Desde un enfoque cuantitativo se destacan el conteo, registro

de casos y la metodología de encuesta (población general y poblaciones especiales). Desde un enfoque combinado se destacan los sistemas de información sobre consumo de drogas y estudios combinados con el empleo de nuevos métodos. Dentro de los nuevos métodos, los autores resaltan el análisis de redes sociales, el muestreo por bola de nieve y los métodos de entrevistadores con acceso privilegiado.

Los datos en la era digital

La recolección de datos en internet para la investigación en políticas de drogas no es una novedad. Miller y Sonderlund (2010) hicieron un levantamiento de las investigaciones en políticas de drogas que han utilizado datos online y encontraron tres tipos principales de metodología de recolección: análisis de contenido, entrevistas y cuestionarios online. Los autores encontraron que el método más utilizado fue el cuestionario online: 44 de los 46 estudios analizados utilizaron este método de recolección de datos. Según los autores, la utilización de cuestionarios online aumenta la probabilidad de alcanzar poblaciones estigmatizadas dado su carácter anónimo y disminuye los costos asociados a la investigación. Sin embargo, esta forma de recolección de datos *online* no elimina los desafíos metodológicos asociados a la construcción de una muestra confiable con base en cuestionarios de autodeclaración, como ya mencionamos.

Hay otros tipos de datos online que no han sido muy utilizados en la investigación en política de drogas pero que pueden aportar *insights* significativos sobre los patrones y comportamientos de consumo y venta de sustancias psicoactivas. Enghoff y Aldridge (2019) los definen como datos *online* no solicitados. La creciente utilización de los medios digitales para la comunicación diaria en varios ámbitos de la vida ha generado un nuevo ambiente social digital que, según Décary-Héту y Aldridge (2015) se caracteriza por la generación de ‘rastros digitales’ de estas actividades, que se convierten en datos que pueden ser de mucha utilidad para la investigación social por su naturaleza espontánea y por su cantidad.

El desarrollo de la tecnología de almacenamiento ha permitido la recolección y procesamiento de una cantidad creciente de datos no solicitados. Mayer-Schonberger y Cukier (2013) estiman que en el año 2000, solo 25% de todos los datos almacenados en el mundo eran digitales y que, ya en el 2013, la cantidad total de información almacenada en el mundo alcanzaba 1.200 exabytes, de los cuales 98% eran digitales. Esa revolución digital, a su vez, ha permitido el desarrollo de nuevas tecnologías

como la Inteligencia Artificial y la *Internet of Things*. Esto abre muchas posibilidades en el campo de la ciencia social, principalmente en temas donde normalmente la recolección de datos se presenta como un desafío importante, como es el caso de las políticas de drogas, que históricamente han encontrado desafíos significativos en la recolección de datos debido a la prohibición y al estigma social asociado a esas prácticas, lo que dificulta la expansión del conocimiento científico en el área.

Enghoff y Aldridge (2019) definen la existencia de dos grupos principales de personas que generan datos no solicitados vinculados al consumo de sustancias psicoactivas en el mundo digital.

Personas que consumen sustancias psicoactivas y participan en comunidades online

En estas comunidades *online* los usuarios pueden interactuar entre sí, intercambiar información sobre patrones y prácticas de consumo, creencias y conocimiento sobre las sustancias, casi siempre de forma anónima. Esto amplía el acceso de los investigadores del campo de drogas a la información sobre las dinámicas de estas comunidades, pues elimina algunas de las barreras de acceso típicas de las comunidades *offline*. Entre estas barreras, podemos nombrar el tamaño y la dispersión geográfica de las comunidades *offline*, y las medidas adoptadas por sus miembros para mantener en secreto las actividades desarrolladas en el ámbito comunitario con el objetivo de mitigar los riesgos de exposición derivados de la ilegalidad y del estigma social. En comparación, las comunidades *online*, dado su carácter anónimo, muchas veces no exigen el establecimiento de vínculos de confianza entre investigador y sujeto investigado, y los miembros de la comunidad se sienten más seguros a compartir información sobre sus prácticas.

Personas que venden sustancias psicoactivas

Enghoff y Aldridge (2019) afirman que el tráfico ilícito de sustancias que se desarrolla en la red convencional ('clearnet') normalmente está asociado a sustancias legales o de distintos estatus legales, como las drogas de prescripción. No fue hasta el 2011 que se escuchó hablar de un mercado *online* de ventas de sustancias ilícitas, el *Silkroad*, que operaba en lo que es conocido como 'darknet'. El éxito de estos mercados, también conocidos como criptomercados, está asociado a la tecnología de encriptación que permite esconder la identidad física de los agentes y sus

transacciones. En este sentido, el anonimato también opera como un factor clave para la existencia de estos mercados, como en el caso de las comunidades *online* de usuarios. Entre los datos que se pueden extraer de los criptomercados, los autores destacan: el listado de productos que incluye información sobre el tipo y la calidad, la cantidad y el precio, el país de origen de los vendedores, los destinos habilitados para el envío de la sustancia, y métricas de reputación de los vendedores obtenidas a través del *feedback* de los compradores.

Dadas las restricciones técnicas asociadas al acceso a los cripto mercados, nos interesa en este trabajo analizar los datos no solicitados que se pueden encontrar en plataformas de uso corriente, como la red social Twitter. Según Enghoff y Aldridge (2019), la utilización de datos provenientes de esta plataforma requiere aplicar estrategias de selección de datos para identificar contenido relevante y excluir la mayoría de los resultados irrelevantes. A continuación, realizaremos dos breves ejercicios para demostrar el potencial de utilizar los datos no solicitados en Twitter para la toma de decisiones en política de drogas.

Uso de datos no solicitados de Twitter en la toma de decisiones de políticas públicas

El uso de datos no solicitados para la toma de decisiones en políticas públicas de sustancias psicoactivas depende enteramente de su contexto particular de uso. Sin embargo, no es descabellado afirmar que presentan información que difícilmente se encontraría en otras metodologías de recolección más “tradicionales” como entrevistas o encuestas longitudinales de prevalencia. Y que permiten ampliar el foco y obtener mayor contexto de una realidad específica. Lo anterior, en ningún modo debe entenderse como fuentes de datos excluyentes a las usadas normalmente, sino más bien como complementarias. Por ejemplo, la discusión pública de nuevas sustancias psicoactivas en Twitter podría actuar como señal (*input*) para los sistemas de alerta temprana (SAT) nacionales.

Twitter, por su parte, se ha consolidado como el escenario de difusión por defecto de muchos políticos y personalidades influyentes de Occidente, donde la censura es relativamente baja a comparación de otras redes sociales con cantidades similares de usuarios, el contenido sexual es relativamente tolerado y se usa inclusive como medio de difusión de comunicados por parte de organizaciones al margen de la ley de muchos países. De esta manera, sea cual sea el caso de uso de los datos no solicitados, es prudente decir que brindan una óptica alternativa y complementaria,

ya sea por sus actores o por el tipo de información que permite circular. Aunado a lo anterior se encuentra la arquitectura que la plataforma dispone para la consulta de datos, que a pesar de requerir un mínimo conocimiento técnico para su consumo, es de fácil acceso. No obstante, vale la pena mencionar que existen planes de pago en la medida en que se requieran hacer mayor cantidad de consultas, costos que un centro de investigación podría asumir.

Aspectos metodológicos relacionados a la recolección de Tweets que mencionan ‘cocaína’ en población colombiana entre el 12 de junio y el 12 de julio de 2019

El API de twitter como recolector de datos no solicitados para la toma de decisiones de políticas públicas de sustancias psicoactivas

Según Mozilla Foundation (2019), las interfaces de programación de aplicaciones (API, por sus siglas en inglés) son “construcciones disponibles en los lenguajes de programación que permiten a los desarrolladores crear funcionalidades complejas de una manera simple”. Abstraen el código complejo hacia conceptos más sencillos de usar. En otras palabras, brindan puntos de salida que los usuarios pueden usar sin tener que preocuparse de la complejidad que los subyace. Por ejemplo, al manejar un carro, el conductor interactúa directamente con el volante para girar en las curvas, en vez de preocuparse por el complejo mecanismo que se desencadena por debajo. Existen dos categorías de API, las que están integradas localmente en el navegador web del cliente y aquellas que se encuentran del lado del servidor. El API de Twitter es de segunda categoría, pues se encuentra alojada y gestionada del lado de los servidores de Twitter.

Para poder interactuar con el API de Twitter es necesario realizar peticiones mediante un protocolo de transferencia de hipertexto o HTTP (por sus siglas en inglés) y en el cual se basa la transferencia de información en la World Wide Web (más conocida como www). Entre otros usos, el API de Twitter nos permite enviar peticiones HTTP y recibir contenido de sus bases de datos. Por ejemplo, digamos que queremos saber hace cuánto la palabra “muro” hace parte de la narrativa de Twitter de Donald Trump, para ello podemos solicitarle a Twitter mediante su API que nos muestre todos los *tweets* en los que Trump haya mencionado “muro” e identificar cuándo fue el primero. Son muchas las posibilidades de este tipo de herramientas, pues nos permiten interactuar y manipular los datos *online* no solicitados que se generan en esta plataforma. Para mayor información sobre el API, puede recurrirse directamente a su documentación (Twitter Inc., 2019).

El API de Twitter tiene muchas funcionalidades, van desde poder automatizar una respuesta hasta permitir la recolección de datos para análisis estadísticos. Igualmente, tienen distintos planes para poder acceder a su API, algunos gratuitos y otros de pago, para efectos de este ejercicio utilizaremos el API endpoint “Search Tweets: 30 days” bajo el plan “sandbox”, que si bien tiene limitaciones en la medida que no da acceso al total de sus herramientas, es suficiente para el propósito de este documento. Con el ánimo de demostrar uno de los tantos usos que el API de Twitter puede tener en la toma de decisiones en políticas públicas de sustancias psicoactivas, procedimos a realizar un ejemplo de cómo usar esta herramienta y demostrar de esta manera su capacidad de aportar datos.

La recolección de datos se realizó entre los usuarios de Twitter ubicados en Colombia que hayan escrito la palabra ‘cocaína’ en algún tweet entre el 12 de junio y el 12 de julio de 2019.

Especificaciones técnicas del entorno de desarrollo

Para poder interactuar con el API “Search Tweets: 30 days” es necesario configurar un ambiente de desarrollo que permita realizar peticiones al API y manipular los datos obtenidos. Las peticiones se obtienen a través de peticiones HTTP, usando Javascript como lenguaje de programación (Node.js). En este caso se usó como entorno de desarrollo:

- Sistema operativo: Windows 10 Pro.
- Editor de Texto: Atom (1.38.2)¹
- Entorno de Ejecución: Node.js (8.9.3)²
- Manejador de paquetes: NPM (6.9.2)³
- Módulos NPM: Chalk (2.4.2)⁴ y Twit (2.2.11)⁵

Para realizar peticiones de manera exitosa, son necesarios algunos accesos especiales de desarrollador que Twitter otorga, estos precisan una fase de aprobación por parte de la plataforma. Por otro lado, se utilizó GitHub como repositorio de código basado

¹ Más información en: <https://atom.io/>

² Más información en: <https://nodejs.org/es/>

³ Más información en: <https://www.npmjs.com/>

⁴ Más información en: <https://www.npmjs.com/package/chalk>

⁵ Más información en: <https://www.npmjs.com/package/twit>

en el sistema de control de versiones Git. El repositorio de este proyecto, así como sus resultados puede consultarse en su totalidad en el siguiente enlace:
<https://github.com/enpepolicy/DrugPolicyTwitterMonitor>

Peticiones al API “Search Tweets: 30 days”

La recolección de datos se realizó entre los usuarios de Twitter ubicados en Colombia que hayan escrito la palabra ‘cocaina’ en algún tweet entre el 12 de junio y el 12 de julio de 2019. Se utilizaron peticiones GET habilitadas por el API, que pueden ser consultadas en su documentación oficial⁶, apoyándonos en un módulo NPM llamado ‘Twit’ que otorga una configuración base para usar el API con mayor facilidad. De acuerdo con las especificaciones del módulo, entre las varias opciones, fueron pasados como parámetros de petición:

- query : ‘cocaina place_country:CO’
- Permite pasar hasta 256 caracteres en entorno Sandbox, es el encargado de definir las reglas/filtros sobre las que se hacen las peticiones. Fue usada la palabra ‘cocaina’ en conjunto con el operador ‘place_country:CO’, es decir todos aquellos *tweets* que tienen Colombia como lugar o locación asociado.
- maxResults : ‘100’
- Permite definir la cantidad de tweets que se espera recibir, en nuestro caso se solicitaron 100 por petición.
- Next : ‘*’
- Debido a que hay un número máximo de *tweets que pueden recibirse por cada petición, para seguir recolectando los faltantes es necesario adjuntar un token que la petición anterior nos devuelve. Realizamos este proceso hasta que el API deje de devolver el token.*

El comportamiento de este proceso puede consultarse en:

<https://github.com/enpepolicy/DrugPolicyTwitterMonitor/blob/master/peticion.js>

Guardado de respuestas

Una vez se efectúa la petición es necesario guardar la información recibida, para ello utilizamos el módulo ‘fs’ que viene integrado de forma nativa con nuestra versión

⁶ Más información en: <https://developer.twitter.com/en/docs>

de Node.js. Cada petición fue guardada en un archivo ‘cocaina-x.json’ y pueden ser consultados en:

<https://github.com/enpepolicy/DrugPolicyTwitterMonitor/tree/master/resultados>

Procesamiento de datos

Los archivos recibidos directamente por el API se encuentran en formato .JSON, tienen información adicional que no necesitamos y deben procesarse a un formato de texto más fácil de manipular como .TXT. Para ello, diseñamos una serie de condiciones que separan los tweets dependiendo de su tipo:

- Retweet: publica en el “muro” el *tweet* de otro usuario sin agregar texto adicional.
- Quote Tweet: publica en el “muro” el *tweet* de otro usuario agregando texto adicional.
- Main Tweet: publica en el “muro” un *tweet* de autoría propia.
- Reply Tweet: replica sobre otro *tweet* algún comentario.

En esta fase se pasaron los archivos ‘cocaina-x.json’, marcados con el módulo NPM ‘Chalk’ y *logueados* en consola cada tipo de *tweet*. Posteriormente, se agruparon manualmente en archivos según tipo. Para realizar este proceso fue ejecutado el siguiente código:

<https://github.com/enpepolicy/DrugPolicyTwitterMonitor/blob/master/procesarATexto.js>

Y los resultados de ese proceso pueden consultarse acá:

<https://github.com/enpepolicy/DrugPolicyTwitterMonitor/tree/master/resultados/procesado>

Resultados

A partir de la búsqueda descrita, se obtuvieron 370 *tweets* distribuidos en categorías de la siguiente manera (ver Tabla 1):

Tabla 1. Tweets distribuidos por categorías

| | |
|-------------|-----|
| Retweet | 31 |
| Quote Tweet | 54 |
| nativoReply | 164 |
| nativoMain | 121 |
| Total | 370 |

Fuente: Peña & Fontes (2019).

Debido a que el propósito de este documento es demostrar con ejemplos sencillos que el uso del API de Twitter puede arrojar datos útiles para mejorar la toma de decisiones en políticas públicas de sustancias psicoactivas. Procederemos a realizar dos tipos de análisis que den cuenta de la versatilidad de este tipo de fuente de información. Inicialmente, realizaremos un análisis de nube de palabras o ‘Word Cloud’ a los tweets en los cuales el usuario directamente escribe una opinión: Quote Tweet, nativoReply y nativoMain. Después, realizaremos un análisis de sentimiento sobre los Quote tweet.

Análisis de WordCloud

Según Atenstaedt (2012), un Word Cloud es una representación gráfica de frecuencia de palabras en un texto. La mayor cantidad de veces que una palabra aparece, más grande se verá en la imagen generada. Suelen usarse en entornos políticos para analizar discursos y tienen valor en la medida que hacen evidentes conceptos que subyacen intenciones políticas.

Para nuestro ejemplo, antes que nada, debemos consolidar un archivo que contenga el tipo de tweets que pretendemos analizar: Quote Tweet, nativoReply y nativoMain. Este consolidado cuenta con 11096 palabras, 67450 caracteres y puede consultarse directamente acá:

<https://github.com/enpepolicy/DrugPolicyTwitterMonitor/blob/master/resultados/procesado/consolidadoNativoYQuote.txt>

Tabla 2. Palabras con más frecuencia sin conectores

| | |
|-------------------|-----|
| cocaína | 341 |
| azúcar | 78 |
| más | 75 |
| Colombia | 58 |
| coca | 35 |
| mundo | 33 |
| Santos | 30 |
| producción | 23 |
| toneladas | 23 |
| negocio | 20 |
| glifosato | 19 |

Fuente: Peña & Fontes, (2019).

A continuación, se presentan dos casos en los que los resultados coinciden con discusiones nacionales sobre políticas públicas de sustancias psicoactivas y dan cuenta de su relevancia:

¿El azúcar es más malo para la salud que la cocaína?

El excandidato presidencial Gustavo Petro fue el centro de la polémica nacional al asegurar en un tweet que el azúcar era una droga más peligrosa que la marihuana o la cocaína (Canal Red Más, 2019), lo que llevó el debate a los medios y al país en general. Para una parte del imaginario social el azúcar no tiene efectos nocivos, y si los tiene, son evidentemente menores que los de la marihuana o la cocaína. En el debate público fueron escasos los soportes investigativos de afecciones relacionadas con el azúcar o las sustancias, y más las opiniones emotivas que normalmente pululan en las redes. Esto se da en un contexto bastante particular, en donde meses atrás, el congreso con un discutido *lobby*, no permitiera subir los impuestos a las bebidas azucaradas que pretendía desincentivar su consumo por razones de salud pública. Por otro lado, el tweet de Petro se dio como respuesta a un tweet iniciado por la senadora del partido Centro Democrático, María Fernanda Cabal, en el que

criticaba a la Corte Constitucional de Colombia, por haber declarado inexecutable unos artículos del Código de Policía que prohibía consumir alcohol y sustancias psicoactivas en espacios públicos (Noticias Canal RCN, 2019).

Santos y su relación con las 180.000 hectáreas sembradas de coca

Una de las principales polémicas que dejó el expresidente Juan Manuel Santos justo antes de salir del cargo, fue la cantidad de hectáreas de coca sembradas: 180.000. Lo anterior se une con la agenda política de los contradictores del Acuerdo de paz con las FARC, que ven una intención de Santos de promover el cultivo a través del Plan Nacional de Sustitución de Cultivos Ilícitos, uno de los pilares del acuerdo. Por otro lado, su nuevo rol le ha permitido al expresidente Santos ser más abierto frente a sus posturas en política de drogas. Recientemente se unió a la Comisión Global de Política de Drogas, organización que promueve la regulación de sustancias psicoactivas a lo largo del mundo y de la que hacen parte varios exmandatarios. La frecuencia de su nombre nos indica el peso que Santos juega en el país como actor en política de drogas y de la importancia de ser tomado en cuenta a la hora de promover políticas públicas al respecto.

Análisis de sentimientos

Según Sande (2018), el análisis de sentimientos es un área de investigación enmarcada dentro del campo del procesamiento del lenguaje natural y cuyo objetivo fundamental es el tratamiento computacional de opiniones, sentimientos y subjetividad en textos. En este sentido, una opinión es una valoración positiva o negativa acerca de un producto, servicio, organización, persona o cualquier objeto sobre el cual se expresa un texto determinado. El proceso de análisis de sentimientos aplicado a mensajes publicados en Twitter puede arrojar información similar a la que se obtiene mediante encuestas de opinión, pero de manera inmediata, menos costosa y con actualizaciones en tiempo real.

Las técnicas computacionales de análisis de sentimientos normalmente se aplican a datos masivos, dada la dificultad del procesamiento manual de esta cantidad de información. Sin embargo, para este ejercicio seleccionamos una muestra de 54 ‘quote tweets’ y los clasificamos de forma manual. Primeramente, los ‘quote tweets’ se clasificaron según dos criterios: sentimiento hacia la sustancia (positivo, neutro

y negativo) y sentimiento general del mensaje (positivo, neutro, negativo), con los siguientes resultados (ver Tabla 3).

Tabla 3. Análisis de sentimientos, Sustancia y Mensaje vs. Positivo, Neutro, Negativo y No relevante

| | Sentimiento (Sustancia) | Sentimiento (Mensaje) |
|--------------|-------------------------|-----------------------|
| Positivo | 3 | 1 |
| Neutro | 44 | 26 |
| Negativo | 5 | 25 |
| No relevante | 2 | 2 |

Fuente: Peña & Fontes (2019).

Dos *tweets* se descartaron de la muestra pues el término de búsqueda ‘cocaína’ no estaba relacionado al asunto de interés de este trabajo. De los 52 *tweets* relevantes, encontramos que 85% expresaban una opinión neutral hacia la sustancia, 10% expresaban una opinión negativa y 5% expresaban una opinión positiva. En relación al sentimiento del mensaje en sí, este porcentaje cambia sustancialmente, a pesar de que aún predomina el sentimiento neutral. En este caso, 50% de los mensajes expresaban una opinión neutral, 48% una opinión negativa, y 2% una opinión positiva. Para profundizar el análisis y explicar la diferencia de opinión encontrada en las dos categorías anteriores, decidimos clasificar los 52 ‘quote tweets’ relevantes por tema, separándolos en tres categorías distintas: consumo/adicción, narcotráfico/corrupción y otros (ver Tabla 4).

Tabla 4. Quote tweets agrupados

| | |
|-------------------------|----|
| Consumo/Adicción | 14 |
| Narcotráfico/Corrupción | 30 |
| Otros | 8 |

Fuente: Peña & Fontes (2019).

58% de los tweets relevantes estaban relacionados al tema narcotráfico/corrupción, 27% estaban relacionados a consumo/adicción, y 15% no estaban relacionados a ninguno de los anteriores. De los 30 ‘quote tweets’ relacionados a narcotráfico/corrupción, 67% expresaban una opinión negativa en general, pero solo 13%

expresaban una opinión negativa hacia la sustancia, lo que explica la diferencia encontrada en la tabla 2. Eso nos lleva a concluir que la mayoría de los ‘quote tweets’ que expresaban una opinión negativa lo hacían en relación con el vínculo entre narcotráfico y corrupción estatal, y no en relación con la sustancia en sí. Por otro lado, de los 14 ‘quote tweets’ relacionados a consumo/adicción, 36% expresaban una opinión negativa y 64% expresaban una opinión neutral, y no había diferencias significativas entre las dos categorías de sentimiento.

Los resultados encontrados en este breve ejercicio de análisis de sentimiento coinciden y complementan los resultados del análisis de Word Cloud. Las opiniones reflejan los sentimientos de una muestra de la población con relación a las discusiones nacionales sobre políticas públicas de sustancias psicoactivas descritas anteriormente, y sirven como un termómetro de la opinión pública que puede influenciar el proceso de toma de decisiones de los actores más relevantes en el tema.

Conclusiones

- **Viabilidad, versatilidad y bajo costo de uso de herramientas**
La era de la información y el reciente acceso a tecnologías de recopilación y procesamiento de grandes volúmenes de datos abre inmensas posibilidades para optimizar la toma de decisiones e implementación de políticas públicas relacionadas con la regulación del uso de sustancias psicoactivas. El costo de efectuar investigaciones decrece exponencialmente en la medida que se tenga acceso a plataformas generadoras de datos como Twitter, permitiendo recolectar de manera automatizada grandes cantidades de datos en periodos muy cortos de tiempo y procesarlos oportunamente, haciendo viables investigaciones que de otro modo no podrían realizarse.
- **Uso de datos no solicitados de Twitter en la toma de decisiones de políticas públicas**
El uso de datos no solicitados para la toma de decisiones en políticas públicas de sustancias psicoactivas depende enteramente de su contexto particular de uso. Lo anterior en ningún modo debe entenderse como fuentes de datos excluyentes a las usadas normalmente, sino más bien como complementarias. Twitter por su parte se ha consolidado como el escenario de difusión por defecto de muchos políticos y personalidades influyentes de Occidente.

Es prudente decir que brindan una óptica alternativa y complementaria, ya sea por sus actores o por el tipo de información que permite circular. Aunado a lo anterior se encuentra la arquitectura que la plataforma dispone para la consulta de datos, que a pesar de requerir un mínimo conocimiento técnico para su consumo, es de fácil acceso.

- **Análisis de Word Cloud como ‘identificador’ de tensiones en el campo de las políticas de sustancias psicoactivas**

El análisis de Word Cloud es una técnica sencilla y de fácil implementación que solo necesita un buen volumen de palabras para llevarse a cabo. Hace evidentes conceptos que subyacen intenciones políticas, que, dentro del contexto de la plataforma de Twitter, puede dar cuenta de discusiones nacionales y sus intensidades. Esta herramienta es lo suficientemente versátil como para permitir utilizar cualquier texto o grupo de textos, y permite su visualización de forma efectiva y agradable.

- **Análisis de sentimientos como termómetros de opinión pública**

Así como el análisis de Word Cloud, el análisis de sentimientos es una técnica simple que, aplicada a grandes cantidades de datos, genera información importante sobre los sentimientos de la opinión pública hacia determinado tema. La opinión pública juega un papel importante en el proceso de toma de decisiones en una sociedad democrática, y tener acceso a una herramienta que la pueda medir en tiempo real, sin necesidad de construcción de muestras, y a un costo reducido, es de mucha utilidad para los actores en políticas públicas.

- **Cuestiones éticas relacionadas al uso de datos no solicitados**

Finalmente, la recolección de datos *online* no solicitados levanta una serie de cuestiones éticas relacionadas a la privacidad de los usuarios y de sus mensajes. Por un lado, se cuestiona si los comentarios publicados en plataformas digitales como Twitter pueden entenderse como datos públicos o privados. Por otro lado, se discute la necesidad de consentimiento previo de cualquier sujeto que participe en alguna investigación. Aunque las posibilidades abiertas por estos nuevos métodos de recopilación y procesamiento de datos sean muchas, igualmente muchos serán los desafíos éticos que tendrán que enfrentar los investigadores, ya que no existen reglas claras, legales o sociales, que definan los límites de la privacidad en el nuevo mundo digital.

Referencias

- Atenstaedt, R. (2012). Word cloud analysis of the BJGP. *British Journal of General Practice*, 62(596), 148
- Bennett, W. L., & Segerberg, A. (2012). The Logic of Connective Action. *Information, Communication & Society*, 15(5), 739-768. <http://doi.org/10.1080/1369118X.2012.670661>
- Bracho, T. (2010). Políticas basadas en evidencia. La política pública como acción informada y objeto de investigación. En *CIDE Problemas, decisiones y soluciones. Enfoques de política pública*. CIDE.
- Canal Red Más. (2019). *Polémica de Petro, tras asegurar que el azúcar es más peligrosa que la cocaína*. <http://www.redmas.com.co/colombia/polemica-de-petro-tras-asegurar-que-la-azucar-es-mas-peligrosa-que-la-cocaina/>
- Consortio Internacional sobre Políticas de Drogas. (2012). *Guía sobre políticas de drogas*. http://www.cicad.oas.org/fortalecimiento_institucional/planesnacionales/docs/guia_politicas_drogas_spa.pdf
- Décary-Héту, D., & Aldridge, J. (2015). *Sifting through the net: Monitoring of online offenders by researchers*. https://www.researchgate.net/publication/312446591_Sifting_through_the_net_Monitoring_of_online_offenders_by_researchers
- Enghoff, O. & Aldridge J. (2019) The value of unsolicited online data in drug policy research. *International Journal of Drug Policy*. <https://doi.org/10.1016/j.drugpo.2019.01.023>
- Hammond, J., Keeney, R. & Raiffa, H. (2000). *Decisiones inteligentes*. Ediciones Gestión 2000.
- Mayer-Schonberger, V. & Cukier, K. (2013). *Big Data. La revolución de los datos masivos*. Turner Publicaciones.
- Miller, P. G., & Sonderlund, A. L. (2010). Using the internet to research hidden populations of illicit drug users: A review. *Addiction*, 105(9), 1557-1567. <https://doi.org/10.1111/j.1360-0443.2010.02992.x>.
- Mozilla Foundation. (2019). *¿Qué son las APIs?* https://developer.mozilla.org/es/docs/Learn/JavaScript/Client-side_web_APIs/Introducci%C3%B3n
- Noticias Canal RCN. (2019). *Corte Constitucional tumba prohibición de consumo de droga y alcohol en el espacio público*. <https://noticias.canalrcn.com/nacional-pais/corte-constitucional-tumba-prohibicion-consumo-droga-y-alcohol-el-espacio-publico>
- Pérez, N. & Quintana, P. (2003). Principales enfoques y estrategias metodológicas empleados en la investigación del consumo de drogas: la experiencia en México. *Revista Cubana de Medicina General Integral*. 19(2). http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21252003000200012&lng=es&nrm=iso
- Peña, E & Fontes, C. (2019). Piscina de textos de tweets. *Repositorio Github*. <https://github.com/enpepolicy/DrugPolicyTwitterMonitor/blob/master/resultados/procesado/consolidadoNativoYQuote.txt>

- Sande, J. C. (2018). *Análisis de Sentimientos en Twitter*. Universidad Oberta de Catalunya. <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/81435/6/jsobrinosTFM0618memoria.pdf>
- Subirats, J. & Dente, B. (2014). *Decisiones públicas. Análisis y estudio de los procesos de decisión en políticas públicas*. Ariel.
- Twitter Inc. (2019). *Docs*. <https://developer.twitter.com/en/docs>