



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Escola Tècnica Superior  
d'Enginyeria Agronòmica i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica  
y del Medio Natural

Aplicación de técnicas de Procesamiento de Lenguaje  
Natural (NLP) en Twitter para la evaluación de Políticas  
Agrarias y del Medio Rural.

Trabajo Fin de Máster

Máster Universitario en Ingeniería Agronómica

AUTOR/A: Gutiérrez Domínguez, Alba

Tutor/a: Roig Tierno, Honorat

Cotutor/a: García Alvarez-Coque, José María

CURSO ACADÉMICO: 2021/2022

## RESUMEN

– **Título del TFG/M:** “Aplicación de técnicas de Procesamiento de Lenguaje Natural (NLP) en Twitter para la evaluación de Políticas Agrarias y del Medio Rural”.

– **Resumen del TFG/M:** El contenido publicado en las redes sociales (RRSS) se ha convertido en una valiosa fuente de información. Poniendo el foco en el sector agroalimentario, el contenido generado por los usuarios en medios como Twitter puede ser de utilidad a pequeños y grandes agricultores para conocer la opinión de los consumidores sobre los productos que adquieren a través de los diferentes canales, la forma en la que se han producido o bien su origen. La aplicación de técnicas de procesamiento de lenguaje natural (NLP) aplicados a grandes volúmenes de información puede ser de utilidad al conjunto del sector (incluyendo policymakers y agentes sociales) para conocer los intereses y necesidades de la población y mejorar la forma de comunicación con la sociedad. En este trabajo, se ha empleado el lenguaje de programación R para estudiar la aplicación de un tipo de modelos NLP de análisis del sentimiento sobre publicaciones en RRSS de temas agronómicos de actualidad, como son la agricultura ecológica y la nueva PAC. Asimismo, se ha valorado el grado de precisión de distintas técnicas de *sentiment analysis*, así como la viabilidad de utilizar estos modelos como una fuente fiable de información para desarrollar futuras acciones y medidas políticas que vayan en línea con la opinión pública. Concretamente, se ha demostrado la utilidad de emplear el *sentiment analysis* para recabar una gran cantidad de datos de usuarios anónimos en cuestión de minutos, identificar el sentimiento de estos respecto a un tema (sea este una política, un modelo de producción o una práctica agronómica concreta) y también el motivo por el que el individuo se muestra a favor o en contra.

– **English:** The content published on social media (RRSS) has become a valuable source of information. Putting the focus on the agri-food sector, the content generated by users in media such as Twitter can be useful to small and large farmers to know what consumers think about the products they acquire, the way in which they have been produced or their origin. The application of natural language processing (NLP) techniques applied to large volumes of information can be useful to the entire sector (including policymakers and social agents) to discover the interests and needs of the population and improve the way they communicate with society. In this final thesis, the programming language R has been used to study the application of NLP models of sentiment analysis on publications in social media of current agronomic topics, such as organic farming and the new CAP. Likewise, precision of different sentiment analysis techniques has been assessed, as well as the feasibility of using these models as a reliable source of information to develop future actions and political measures that are aligned with public opinion. Specifically, the usefulness of sentiment analysis has been demonstrated in order to collect large amounts of data from anonymous users in a matter of minutes, to identify their opinion and find the reason behind it.

– **Palabras claves (key words):** sentimentanalysis; NLP; redessociales (socialmedia); RRSS; Twitter; RStudio; opinionmining; política agraria (agricultural policy)

– **Autor del TFM:** Dña. Alba Gutiérrez Domínguez

– **Localidad y fecha:** (Valencia, Julio de 2022)

– **Tutor Académico:** D. Honorat Roig Tierno

– **Cotutor:** D. José María García Álvarez-Coque

# ÍNDICE

RESUMEN .....	i
<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
OBJETIVOS DEL TFM .....	2
<b>2. MARCO TEÓRICO .....</b>	<b>3</b>
2.1. Política agroalimentaria .....	3
2.2. Las RRSS como fuente de datos para el análisis de políticas agrarias .....	9
2.3. Técnicas de Procesamiento del Lenguaje Natural .....	15
2.4. Contribución del NLP a la evaluación y configuración de políticas agrarias .....	21
<b>3. INTRODUCCIÓN A LA METODOLOGÍA.....</b>	<b>24</b>
<b>4. RESULTADOS .....</b>	<b>25</b>
4.1. Caso 1: agricultura ecológica .....	26
4.2. Caso 2: la PAC .....	26
<b>5. LIMITACIONES DEL ANÁLISIS DE OPINIÓN EN RRSS .....</b>	<b>38</b>
<b>6. CONCLUSIONES .....</b>	<b>39</b>
<b>BIBLIOGRAFÍA .....</b>	<b>41</b>
<b>ANEXOS .....</b>	
<b>I. Tablas de abreviaturas y definiciones</b>	
<b>II. Metodología</b>	
<b>III. Registro y acceso a la API de Twitter</b>	
<b>IV. Desarrollo del caso 1 (caso piloto)</b>	
<b>V. Código R del Caso 1</b>	
<b>VI. <i>Dataframe</i> del Caso 1</b>	
<b>VII. Código R del Caso 2</b>	
<b>VIII. <i>Dataframe</i> del Caso 2</b>	
<b>IX. Limitaciones del análisis de opinión en RRSS</b>	

## Índice de imágenes, tablas y gráficos de la memoria

*Imagen 1. Línea de tiempo con los cambios introducidos por cada reforma de la PAC. 7*

*Imagen 2. Esquema del proceso SMA. 11*

*Imagen 3. Análisis de sentimiento empleando el método de aprendizaje supervisado. 18*

*Imagen 4. Arquitectura del método de aprendizaje no supervisado para análisis de sentimientos. 19*

*Imagen 5. Metodología de trabajo SMA aplicada a la analítica de sentimientos u opinion mining. 24*

**Imagen 6.** Ejemplo de varios tuits antes y después de su “limpieza”. 29

**Imagen 7.** Resultados en R utilizando la metodología de aprendizaje no supervisado. 29

**Imagen 8.** Captura del tuit de Greta Thunberg sobre la PAC (25 de octubre de 2020). 29

**Imagen 9.** Resultados en R utilizando la metodología de aprendizaje no supervisado. 32

**Imagen 10.** Nube de palabras del conjunto de tuits de sentimiento positivo sobre la PAC. 33

**Imagen 11.** Nube de palabras del conjunto de tuits de sentimiento negativo sobre la PAC. 33

**Tabla 1.** Relación de los eco-regímenes con las prácticas asociadas y los cultivos que pueden acoogerse a cada 8

**Tabla 2.** Comparativa de las ventajas y limitaciones de las encuestas tradicionales vs UGC. 14

**Tabla 3.** Ventajas e inconvenientes de cada método de análisis de opiniones. 20

**Tabla 4.** Métodos híbridos empleados en diversos estudios para el análisis de sentimientos en Twitter. 21

**Tabla 5.** Resumen de las principales características de cada caso de estudio. 26

**Tabla 6.** Algunos de los tuits clasificados como “Neutral” dentro del caso 2. 30

**Gráfico 1.** Distribución de la muestra global de tuits desde 2017 hasta la actualidad. 27

**Gráfico 2.** Evolución temporal del número de tuits publicados desglosado por categoría de sentimiento. 30

**Tabla 7.** Las publicaciones que han conseguido más RT, clasificadas por sentimiento y temática. 36

## 1. INTRODUCCIÓN

En el sector de la agricultura cada vez es más frecuente escuchar conceptos como *Big Data*, *Data Analytics*, *Machine Learning* (ML) o Inteligencia Artificial (IA). La agricultura 4.0. ha permitido a muchos agricultores mejorar su productividad y reducir la incertidumbre gracias a la utilización de herramientas de ayuda a la toma de decisiones (Navarro-Hellín et al, 2016), imágenes por satélite y drones (Yang et al, 2012) o bien modelos de predicción del rendimiento de cosecha (Salman et al, 2021). Sin embargo, estas innovaciones se han introducido con mucha fuerza en los ámbitos privados de la producción vegetal y animal, la industria agroalimentaria y las empresas de servicios, mientras que el sector público ha quedado relegado a un segundo plano. Las administraciones públicas, encargadas de gestionar las políticas agraria y de desarrollo rural en torno a las cuales se estructura el conjunto del sector agropecuario de un país, todavía no han incorporado este tipo de tecnologías dentro de su estructura de trabajo (Criado & Rojas, 2013).

La agricultura europea se enfrenta en la actualidad a múltiples desafíos globales. La actividad humana responsable del cambio climático provocará un aumento en la temperatura de la Tierra de al menos 1,5 °C (IPCC, Febrero de 2022). Ello dará lugar a un mayor número de fenómenos meteorológicos y climáticos extremos en regiones como el sur de Europa, donde los períodos de sequía y las olas de calor alcanzarán umbrales de tolerancia críticos para la biodiversidad, la salud humana y la agricultura. Asimismo, el conflicto interterritorial entre Rusia y Ucrania ha desembocado en una crisis energética y una subida general de los precios que pone en riesgo la seguridad alimentaria en Europa y compromete las rentas de los agricultores, agudizando las tensiones sociales en el mundo rural. En este contexto de incertidumbre, las nuevas políticas agrarias europeas, en su mayoría englobadas dentro de la futura Política Agraria Común (PAC) post-2020, están obligadas a adaptarse a esta nueva realidad (Moraleda, 2022). Para ello, entre otras cosas, es necesario que tomen en cuenta las necesidades de productores y consumidores por igual. Escucharles es una buena manera de hacerlo. Pero, ¿cómo pueden los organismos públicos conocer las opiniones y las demandas de la población de forma masiva, rápida y que no suponga un gran desembolso económico? En este aspecto, las redes sociales (RRSS) se han convertido en una eficaz fuente de información sobre la ciudadanía, una ventana a las opiniones de los ciudadanos, expresadas de forma libre y gratuita (Batrinca & Treleaven, 2015).

Las RRSS se han incorporado a nuestra vida cotidiana de una manera rápida y progresiva desde hace más de una década, constituyendo un *“fenómeno social, político, económico y tecnológico que está modificando la forma en la que nos relacionamos”* (Criado & Rojas, 2013). En diferentes ámbitos, estas plataformas han alcanzado a una masa crítica de millones de usuarios que comparten su día a día, sus opiniones, quejas y reclamaciones con su núcleo de seguidores continuamente y desde cualquier parte. Sólo se requiere de un *smartphone* y conexión a Internet.

Las administraciones públicas han sufrido una profunda transformación en materia de digitalización y de comunicación en RRSS en las últimas dos décadas (Criado & Rojas, 2013), apostando de una manera generalizada por el uso de las RRSS para difundir su mensaje mediante la apertura de perfiles en redes y de otros canales de comunicación. Si bien, no es esta la única manera para el sector público de aprovechar el potencial de las redes como fuente de información. La tecnología también permite que las entidades

públicas recaben datos de valor procedentes de las RRSS para que, a través de la analítica de datos, contribuyan a mejorar los mecanismos de evaluación y toma de decisiones en materia de políticas públicas, como la PAC.

La tecnología de procesamiento del lenguaje natural (NLP) es una de estas herramientas de análisis de datos, un campo poco explorado hasta ahora dentro del ámbito agroalimentario. Haciendo uso de estas técnicas, el contenido generado por los usuarios en redes como Twitter puede ser de gran utilidad a pequeños y grandes agricultores, para conocer qué opinan los consumidores sobre los productos que adquieren en los supermercados (o en mercados locales), la forma en la que se han producido o su origen. Yendo un paso más allá, aplicar técnicas NLP puede ser de utilidad al conjunto del sector (incluyendo *policymakers* y agentes sociales) para analizar de qué forma podría mejorar la relación actual con el consumidor así como la información transmitida al público general desde las empresas de la industria agroalimentaria y las instituciones académicas.

En concreto, los modelos informáticos de analítica del sentimiento son capaces de extraer de un texto opiniones, actitudes u emociones sobre cualquier tema a partir del lenguaje empleado y según el contexto en el que se formulan (Patel & Patel, 2020). El estudio del sentimiento es una herramienta fundamental en las ciencias sociales y políticas, empleada para evaluar desde la polarización en la comunicación política hasta el análisis de reseñas de productos y comentarios en RRSS en diversos subcampos (Van Atteveldt et al, 2021). Durante la elaboración de este trabajo, sin embargo, no se han encontrado investigaciones que aborden el análisis del sentimiento (también llamado *opinion mining*) y cómo este puede ser útil a la hora de configurar políticas públicas agrarias alineadas con los intereses y necesidades que los ciudadanos expresan a través de sus perfiles de RRSS. En el futuro, esta metodología podría servir como complemento a otras herramientas empleadas durante las etapas de desarrollo de las políticas agrarias y rurales, abriendo la puerta a una transformación de calado en el sector público agrario, que tenga mucho más en cuenta las opiniones de gran parte de la población.

## **OBJETIVOS DEL TFM**

En este Trabajo de Fin de Máster, se evaluará el método NLP de aprendizaje no supervisado para el análisis de opiniones de los usuarios de una de las RRSS más populares, Twitter, acerca de algunos de los temas del sector agroalimentario que más debate han generado en los últimos años, con particular interés en conocer las reacciones de los ciudadanos frente a la PAC europea. Se valorará el grado de fiabilidad de este método frente a la comprobación humana y se estudiará la viabilidad de utilizar este tipo de modelos del lenguaje (LMs por sus siglas en inglés) como una fuente de información rápida y fiable para contribuir al desarrollo de futuras medidas o iniciativas sociopolíticas del ámbito agroalimentario que vayan en línea con el sentimiento mayoritario que expresa la población.

Entre las acciones que se han llevado a cabo, se encuentran:

- Valoración de distintas metodologías NLP y *opinion mining* para su aplicación al contenido generado en RRSS por los usuarios (*User-Generated Content*).
- Aplicación en lenguaje de programación R de la metodología de aprendizaje no supervisado para el análisis de sentimientos en publicaciones de Twitter que tienen que ver con la agricultura ecológica y la PAC.

- Estudio de la evolución de la opinión de los usuarios sobre la PAC a lo largo de los últimos 5 años y detección de aquellos sucesos que han disparado las reacciones de *tuits* positivos y/o negativos.
- Comparativa de distintos *lexicon* empleados habitualmente en los métodos de aprendizaje no supervisado para mejorar la precisión del modelo de análisis de opinión ejecutado.
- Análisis de los temas más comentados en las publicaciones (los *tuits*) de sentimiento positivo o negativo y de las causas por las que una política o práctica genera mayor aceptación o rechazo entre la población.
- Análisis de las opiniones positivas y negativas sobre la PAC desglosadas por países.
- Detección de las limitaciones de tipo lingüístico, técnico, legal, etc. que encierran este tipo de metodologías.

En este trabajo, se presenta una breve revisión del marco teórico de la PAC y su última reforma 2023-2027, se introducen los conceptos necesarios de conocer y que se abordarán más adelante como RRSS, NLP o *sentiment analysis* y se presentan las técnicas de procesamiento del lenguaje más habituales para el análisis de sentimientos (Apartado 2). Después de destacar el papel del NLP en la configuración de las políticas agrarias, en el Apartado 3 se explica paso a paso el método elegido para el análisis de opiniones (método de aprendizaje no supervisado o *lexicon-based*) en base a la metodología de *Social Media Analytics* de Stieglitz et al (2018). Posteriormente, se introducen en el Apartado 4 los dos casos de uso sobre los que se probará el modelo de lenguaje y se analizan los resultados obtenidos. Por último, el Apartado 5 recoge las limitaciones a esta tecnología y la discusión de los resultados obtenidos en el Apartado 4.

## 2. MARCO TEÓRICO

### 2.1. Política agroalimentaria

- *Impulso de nuevas políticas, leyes e iniciativas públicas dentro del ámbito agroalimentario*

La agricultura es una de las actividades con mayor impacto sobre la ciudadanía. Concretamente, el sector agroalimentario en España supone un 10% del Producto Interior Bruto y genera más de 2 millones de puestos de trabajo directos (Moraleta, 2022). En un momento en el que la población crece aceleradamente y el cambio climático y la actividad humana ponen en riesgo la disponibilidad de recursos del planeta, avances en el terreno agronómico son indispensables para garantizar la producción de alimentos a nivel mundial y de forma sostenible. Innovaciones como la carne cultivada (o sintética), los organismos transgénicos o la vuelta a sistemas productivos más sostenibles y a las cadenas de distribución local pueden resultar irrenunciables para sostener la industria y la sociedad del futuro.

Según Ramos-Sandoval & Beltran (2021), en los últimos años se están produciendo cambios importantes en la dinámica de los sistemas alimentarios en todo el mundo, cambios alejados del paradigma actual centrado en unos pocos actores del sistema agroalimentario (agroindustria y distribución), cambios que buscan incluir a la sociedad civil y a los

consumidores como actores clave en esta transición. De igual modo, Renting et al (2003) defendieron *“la creación, operación y evolución de nuevas cadenas alternativas de suministro de alimentos como una de las dimensiones clave de los nuevos patrones emergentes de desarrollo rural”*, subrayando por tanto la importancia de explorar el rol del consumidor y del agricultor en la dinámica de los patrones de producción y consumo emergentes.

Dado que estas innovaciones son clave para el progreso económico y social, es fundamental comprender la posición de los consumidores al respecto (Siegrist & Hartmann, 2020), así como de los productores. Su apoyo es necesario si se pretenden implementar con éxito las innovaciones tecnológicas que se necesitan y la industria agroalimentaria debe examinar si una nueva tecnología será aceptada o rechazada por la ciudadanía antes de implementarla y comercializarla (Cifci et al., 2020). Lo mismo puede extrapolarse al papel de las instituciones públicas, a la hora de lanzar nuevos planes estratégicos que configuren un nuevo sistema de producción agraria. Por ejemplo, algunas tecnologías, como los organismos genéticamente modificados (GMOs, por sus siglas en inglés) o la irradiación de alimentos, son propensas a recibir críticas negativas, a menudo fruto del miedo o bien del desconocimiento (Vàzquez-Salat & Houdebine, 2013).

La industria agroalimentaria debe aprovechar todas las fuentes de información y análisis a su alcance para comprender mejor el pensamiento del consumidor y, en consecuencia, desarrollar tecnologías e implementar estrategias socialmente aceptadas y que tengan en cuenta las necesidades y valores de la sociedad actual, logrando que la población se involucre en el proceso de cambio (Kemp, 2013). De igual modo, conocer las opiniones de la población puede ayudar a esclarecer la forma en la que las empresas, las administraciones públicas y las universidades brindan información a la ciudadanía, pudiendo contribuir al desarrollo de campañas de información y estrategias de comunicación y márketing que muestren a los consumidores todos los beneficios que esas innovaciones pueden traer a la sociedad y cómo estas contribuyen a solucionar desafíos actuales como crisis de suministro, problemas de seguridad alimentaria e incluso el cambio climático (Tao, Yang & Feng, 2020).

Para todo ello, y como se verá más adelante, el análisis de los contenidos generados por los usuarios (UGC por sus siglas en inglés) en Internet (y más concretamente, en RRSS) puede llegar a ser determinante en el desarrollo de las nuevas tecnologías e innovaciones del sector agroalimentario (Recuero-Virto & Valilla-Arróspide, 2022), así como para el impulso de nuevas estrategias y políticas agroalimentarias que cuenten con el respaldo de gran parte de la ciudadanía.

Por tanto, uno de los casos de estudio que se plantean en el presente trabajo busca responder a la siguiente pregunta:

**“¿Es posible utilizar el análisis de las opiniones que los consumidores (y los productores) vierten en RRSS respecto a temas agronómicos de interés, tales como la agricultura ecológica, los alimentos genéticamente modificados o las macrogranjas, como herramienta de ayuda a la configuración de nuevas iniciativas políticas?”**



– *Evaluación de políticas agrarias*

La agricultura europea de las últimas décadas ha estado irremediablemente ligada al nacimiento y la evolución de la Política Agraria Común. Tras la Segunda Guerra Mundial, la principal preocupación de los mandatarios políticos europeos era encontrar la forma de aumentar significativamente la producción agrícola en el continente (De Castro, Miglietta & Vecchio, 2020). Teniendo en cuenta el hecho de que la Comunidad producía solo el 80% de sus necesidades totales de productos agroalimentarios en el momento de la firma del Tratado de Roma (1957), los objetivos iniciales de la PAC, según los autores, fueron:

1. elevar la productividad apostando por la tecnificación y la optimización de todos los factores de producción;
2. asegurar la estabilización del mercado y garantizar la seguridad en el suministro;
3. garantizar unos precios aceptables para el consumidor.

Para ello, la PAC se estructuró en torno a cuatro principios: eliminación de las barreras aduaneras a los productos agroalimentarios entre los países miembros, armonización de los instrumentos de cada país en materia de política agraria (algo que se está revirtiendo a partir de las últimas reformas de la PAC), protección frente a la importación para garantizar que se atiende la demanda de productos alimentarios dentro de la Comunidad y por último, la solidaridad financiera, es decir, la financiación común de todas las medidas y mecanismos de la PAC por parte de los países miembros. Marković & Pejanović (2012) explican la historia de la PAC y sus sucesivas reformas a lo largo de estos años:

La PAC comenzó a implementarse en el año 1962 mediante un sistema de precios garantizados, la protección de las importaciones y las intervenciones en el mercado. Consiguió numerosos efectos positivos ya en la primera década de su implementación, facilitando la autosuficiencia en la producción, la estabilidad del mercado y el crecimiento de los ingresos de los productores. Sin embargo, al mismo tiempo, comenzaron a manifestarse las primeras consecuencias negativas de los mecanismos intervencionistas de la PAC. Principalmente, los precios garantizados y los subsidios a la exportación desencadenaron grandes excedentes de ciertos productos agrícolas y un rápido crecimiento de los costes presupuestarios. La política proteccionista de precios tensó las relaciones que la Comunidad mantenía con sus viejos socios comerciales. Además, la intensificación de la producción y la implementación descontrolada de los insumos químicos provocaron una creciente degradación de los recursos naturales y numerosos problemas medioambientales.

Las primeras reformas (1968, 1972, 1985) buscaron paliar los problemas anteriormente mencionados con poco éxito. No fue hasta 1992 cuando la nueva reforma, conocida como reforma MacSharry, empezó a dar pasos hacia el sistema actual de ayudas directas a productores, mediante la reducción de los precios de los cereales y la carne de vacuno, el pago de compensaciones a los agricultores por la exención de la producción de tierras, así como la compensación por la extensificación de la producción ganadera. Asimismo, por primera vez la PAC contempló el desarrollo rural como pieza clave dentro de sus medidas de acompañamiento.

El gran giro que supuso la implementación de la Agenda 2000 (Berlín, 1999) fue la introducción de una política de desarrollo rural como el “segundo pilar” de la PAC. Por primera vez aparecieron como criterios para la concesión de ayudas los llamados principios ambientales y de producción agrícola sostenible. Esta tendencia se mantuvo con la reforma Fischler (2003), reforzando la política rural como parte integral del paquete de reformas. Asimismo, con la reforma Fischler se continuó en la línea de reducir las medidas con impacto distorsionador en el mercado mundial. Se introdujo el esquema de pago único por finca. El régimen de pago único (2005) implica que los agricultores reciben subvenciones cuyo importe no depende del ámbito de producción. En esencia, esto elimina la relación entre la producción y las ayudas, y proporciona a los agricultores europeos una mayor libertad para comportarse como productores orientados al mercado.

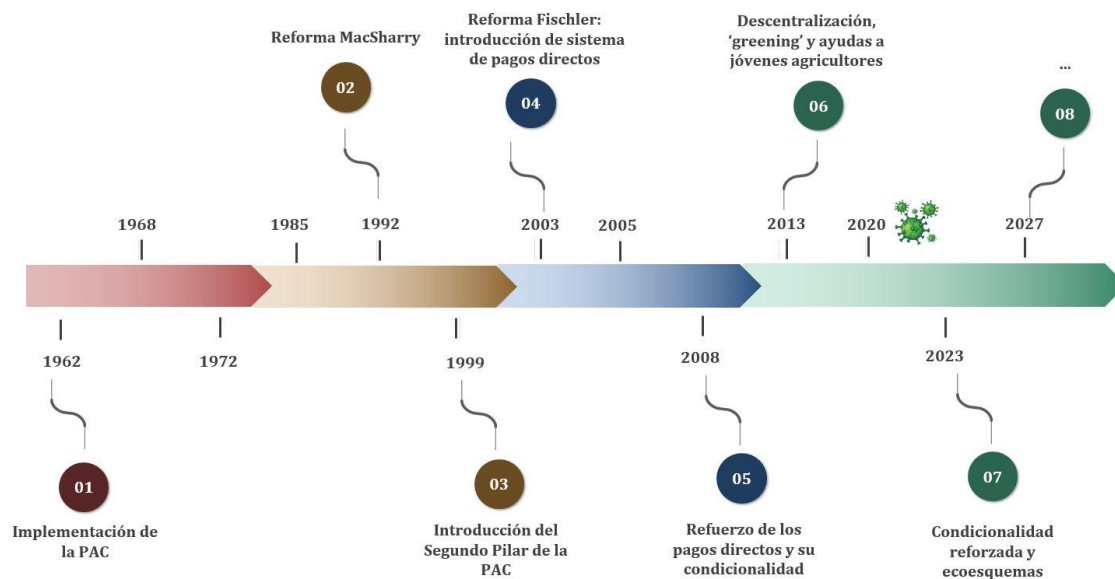
Tras la reforma de 2003, la PAC se estructuró en dos pilares principales que contenían los siguientes instrumentos:

1<sup>er</sup> Pilar: sistema de pagos directos e intervención en ciertos mercados. En este pilar hay un concepto fundamental, el de “condicionalidad”. La condicionalidad establece que el alcance del pago directo, es decir, el incentivo para los productores, está sujeto a que estos respeten distintas normas relacionadas con la protección del medio ambiente, la seguridad de los alimentos y el bienestar animal.

2<sup>o</sup> Pilar: política de desarrollo rural. Los fondos destinados a este pilar de la PAC dan soporte a la mejora estructural y la implantación de nuevas tecnologías para que las condiciones de vida y trabajo en las zonas rurales de la Unión Europea se ajusten a las necesidades de la población.

Las modificaciones introducidas en el marco de la reforma de la PAC de 2008 siguieron la dirección establecida en 2003 al desacoplar aún más los pagos directos de la producción y reforzar la condicionalidad de los pagos. La última reforma de la PAC se aprobó en el año 2013 y continúa en vigor. Por primera vez, la reforma de 2014-2020 dotó a los Estados Miembros (EM) de la posibilidad de adaptar los pagos directos de la PAC para poder abordar mejor sus necesidades nacionales (Henke et al, 2018). Los EM debían decidir: qué pagos adoptar (de aquellos no obligatorios), la cantidad de recursos que asignar a cada pago y los criterios de elegibilidad asociados, la distribución de fondos entre los dos pilares de la PAC, el perfil de los beneficiarios así como los criterios para garantizar una distribución justa de los pagos entre agricultores, sectores y regiones dentro de cada Estado miembro. Asimismo, se introdujeron novedades como las ayudas a jóvenes agricultores y el 30% del apoyo condicionado al cumplimiento de las medidas “verdes” o *greening* (Dupraz & Guyomard, 2019).

Estos sucesivos cambios a lo largo de su corta historia, han ido siempre de la mano de las necesidades socioeconómicas emergentes y los cambios dentro del escenario de la UE y el dominio global (De Castro, Miglietta & Vecchio, 2020). También, de las exigencias de los consumidores europeos, especialmente en materia de protección medioambiental, seguridad alimentaria y bienestar animal.



**Imagen 1.** Línea de tiempo con los cambios introducidos por cada reforma de la PAC.  
Fuente: Elaboración propia.

En el año 2023, está previsto que entre en vigor la nueva PAC post-2020, que introducirá una serie de modificaciones al plan anterior. La nueva política pretende apostar por un enfoque más ecológico debido a la presión social que existe para adoptar una posición más responsable con el medio ambiente (Horák, 2022). Dichas modificaciones irán principalmente enfocadas a reforzar el papel ambiental y de lucha contra el cambio climático de la agricultura, así como en lograr una adecuada alineación de los resultados esperados con los estándares impuestos por los Objetivos de Desarrollo Sostenible (ODS) recogidos en la Agenda 2030 de Naciones Unidas (Matthews, 2020).

La nueva PAC llega en un momento de convulsión, retrasada a consecuencia de la pandemia del virus SARS-CoV-2 y forzada a lidiar con la crisis energética y de precios que sacude a los agricultores y consumidores europeos tras el estallido del conflicto bélico entre Rusia y Ucrania en febrero de 2022 (Consejo Europeo, 2021; Martínez Arroyo, 2021; Maté, 2022). **Se trata, por tanto, de un momento propicio para evaluar los resultados alcanzados gracias a la última PAC (período 2013-2020, extendida hasta 2023) y determinar cuáles son los retos que el nuevo plan debe abordar a lo largo de los próximos cuatro años (2023-2027).**

La última reforma pretende ir en la misma línea de estos últimos años, apostando por los pagos directos vinculados al logro de objetivos medioambientales y climáticos pero bajo una nueva figura denominada “condicionalidad reforzada” (Petsakos et al, 2022). Las medidas de *greening* introducidas con la reforma de 2013 se reemplazan por dicho principio, que engloba una serie de normas y buenas prácticas agrarias y medioambientales que todo beneficiario de las ayudas directas ha de cumplir en relación al medio ambiente, el cambio climático, la salud pública, la salud animal, la sanidad vegetal y el bienestar animal (MAPA, 2021). Bajo esta nueva figura, se incluyen aspectos de la condicionalidad actual, se incorporan normas anteriormente incluidas en el pago verde o *greening* y se añaden también algunas novedades.

Por lo tanto, se pretende que la condicionalidad reforzada se convierta en la base de las obligaciones ambientales y climáticas obligatorias sobre las cuales los EM diseñen

medidas aún más ambiciosas (aunque voluntarias) a través de sus planes nacionales estratégicos (Azcárate & Folkesson, 2020). Estas medidas, que cada EM adaptará a las particularidades nacionales propias, se denominan eco-esquemas o eco-regímenes, y los agricultores que se adhieran a alguna de ellas recibirán un pago por hectárea adicional. En España, por ejemplo, se han aprobado 7 prácticas susceptibles de remuneración, aplicables a dos eco-regímenes: agricultura baja en carbono y agroecología (Villanueva, 2022). Estas prácticas se recogen en el Plan Estratégico de la PAC de España (PEPAC). Al menos un 25% del presupuesto nacional de los EM a las ayudas directas deberán ir dirigidas al pago de estos eco-esquemas de adopción voluntaria [Art. 97.1 del Reglamento (UE) 2021/2115, de 2 de diciembre de 2021, CE, 2021].

Eco-régimen	Objetivo principal	Prácticas	Cultivos que pueden acogerse
Agricultura baja en carbono	Mejorar la estructura de los suelos, reducir la erosión y la desertificación, aumentar el contenido en carbono de los mismos y reducir las emisiones.	Pastoreo extensivo (P1).	Pastos permanentes y pastizales
		Agricultura de conservación: siembra directa (con gestión sostenible de insumos de regadío) (P4).	Cultivos herbáceos
		Cubiertas vegetales espontáneas o sembradas en cultivos leñosos (P6)	Cultivos permanentes
		Práctica de cubiertas vegetales inertes en cultivos leñosos (P7).	Cultivos permanentes
Agroecología	Favorecer la biodiversidad asociada a espacios agrarios, los paisajes y la conservación y la calidad de los recursos naturales, agua y suelo	Siega sostenible y el establecimiento de islas de biodiversidad en las superficies de pastos. (P2).	Pastos permanentes y pastizales
		Rotaciones en tierras de cultivo con especies mejorantes (con gestión sostenible de insumos en regadíos) (P3).	Cultivos herbáceos
		Espacios de biodiversidad en tierras de cultivo y cultivos permanentes (con gestión sostenible de insumos en regadío): Superficies no productivas y elementos del paisaje (P5).	Cultivos herbáceos y permanentes. Condiciones específicas para cultivo bajo agua

**Tabla 1.** Relación de los eco-regímenes con las prácticas asociadas y los cultivos que pueden acogerse a cada eco-régimen.

Fuente: PEPAC (MAPA, 2021).

## Limitaciones en la evaluación de la PAC

En la última década, se ha multiplicado la literatura existente acerca del impacto de las medidas y reformas de la PAC sobre la actividad agraria europea. La apuesta por medidas de fuerte carácter ambiental tiene efectos importantes sobre otros elementos clave, como la competitividad del sector primario europeo (De Castro, Miglietta & Vecchio, 2020). Existe por tanto la necesidad de abrir una discusión sobre las implicaciones que podría tener la implementación de nuevos enfoques.

La aplicación de los denominados Métodos de Evaluación de Programas (PEM por sus siglas en inglés) para la evaluación de la PAC y de otras políticas públicas goza cada vez de mayor aceptación entre la comunidad sociocientífica (Esposti, 2022), ya sea *ex ante*, para analizar los posibles impactos de la implementación de nuevas políticas, o *ex post*, para un análisis de impactos después del período de implementación (Horák, 2022). El uso de conceptos estadísticos como la Inferencia Causal (IC) se ha convertido en un campo en sí mismo a la hora de evaluar políticas (Matthews, 2021). Sin embargo, la PAC presenta ciertas características que dificultan el estudio y la medida de los resultados obtenidos siguiendo este tipo de técnicas.

Principalmente, se trata de una política y no de un programa, es decir, está compuesta de un conjunto de medidas heterogéneas e interdependientes (Lassance, 2020). Si bien cada una de estas medidas podría evaluarse por separado, lo habitual es que actúen de forma conjunta y que los beneficiarios de la PAC se adhieran a varias de ellas a la vez (Matthews, 2020), por lo que los resultados son el fruto de la ejecución simultánea de todas ellas y es difícil atribuir dichos resultados a una u otra medida de forma independiente, sino que es la suma de todas ellas la responsable de los ya mencionados resultados.

A esto se suma el carácter voluntario de algunas de ellas (Dupraz & Guyomard, 2019), que acrecienta el reto de evaluar la PAC puesto que las condiciones particulares de cada beneficiario juegan un papel importante a la hora de que los agricultores decidan adoptar estas medidas voluntarias, así como el hecho de que el conjunto de políticas no solo está relacionado con la agricultura, sino que se ve afectado por muchos otros sectores de la economía europea, tales como el sector energético y el sector industrial (Bureau & Swinnen, 2018), y esto podría distorsionar los posibles resultados de la investigación. Por último, la reforma de 2014-2020 introdujo un sistema de pagos más selectivo, dentro del cual cada EM podía elegir y adaptar las diversas medidas a sus necesidades y características particulares (Henke et al, 2018), complicando todavía más la evaluación de la PAC a nivel europeo.

**Se trata, por tanto, de una labor compleja y que requiere de un gran volumen de datos. Una vez que el objetivo de la evaluación ha quedado definido, lo más importante es acceder a los antecedentes empíricos que requiere dicha evaluación, es decir, disponer de los conjuntos de datos necesarios para la labor de investigación** (Esposti, 2022). En este sentido, la Unión Europea ha destacado en numerosas ocasiones las ventajas de involucrar a los ciudadanos en la evaluación de la PAC, así como en las discusiones para las futuras reformas (Marković & Pejanović, 2012). Como consumidores de productos agroalimentarios, se convierten en una fuente de datos clave para el análisis de los mecanismos de ayuda europeos. De hecho, ha sido gracias a las exigencias del consumidor europeo que a lo largo de estos años se han aprobado leyes clave dentro del sistema comunitario, tales como la directiva de bienestar animal en explotaciones ganaderas (Directiva 98/58/CE) o la ley sobre producción ecológica y etiquetado de productos ecológicos (Reglamento 2018/848).

En consecuencia, la siguiente pregunta que se plantea en este trabajo es:

**“¿Puede resultar útil analizar las opiniones de la población sobre la política agraria europea de cara a evaluar los resultados conseguidos y plantear nuevas reformas, empleando para ello el enorme conjunto de datos que los usuarios generan en RRSS?”**

## **2.2. Las RRSS como fuente de datos para el análisis de políticas agrarias**

Las RRSS se han convertido en la mejor plataforma para que las personas compartan ideas, pensamientos e información de manera virtual (Susmitha&Pranitha, 2022). Se trata del instrumento de comunicación que más ha crecido en los últimos años (Criado & Rojas, 2013). Según Kaplan & Haenlein (2010), el término “red social” hace referencia a toda plataforma (en Internet) *“en la que el contenido y las aplicaciones ya no son creados y*

*publicados por individuos, sino que todos los usuarios los modifican continuamente de manera participativa y colaborativa*". En otras palabras, cualquier aplicación basada en Internet que permita "la creación y el intercambio de contenido generado por el usuario", el conocido como *User-Generated Content (UGC)*.

De acuerdo con Antypas, Preece & Collados (2022), el 48% de los europeos usan las RRSS todos los días o casi todos los días. Estas herramientas han favorecido la creación en los últimos años de redes temáticas de usuarios que generan conocimiento de valor disponible para que cualquier organización (incluida la propia RRSS) conozca los intereses de estos usuarios y puedan tomarlas en consideración para sus actividades (Criado & Rojas, 2013). Debido a la cantidad de contenido producido diariamente y la cantidad de usuarios activos en las plataformas de RRSS, cualquier empresa u organización tiene la posibilidad de emplearlas para comprender cómo evolucionan las tendencias en su sector (así como los problemas), identificar riesgos y oportunidades en la comunicación a través de este medio y extraer conclusiones útiles que les ayuden en su toma de decisiones (Stieglitz et al, 2018). Esto mismo puede ser de utilidad a las instituciones públicas, como se verá más adelante.

La acumulación de esta enorme cantidad de datos, ofrecida en múltiples formatos (incluidos datos de texto, imágenes, vídeos, sonidos y geolocalizaciones), se conoce como *Social Media Big Data* (Stieglitz et al, 2018) y de su análisis puede extraerse información de valor sobre las últimas tendencias, problemas, necesidades o actores influyentes que interesan a los usuarios de un lugar y en un momento determinado.

El campo de *Social Media Analytics (SMA)* engloba todos aquellos procesos necesarios para extraer conocimiento de las RRSS, esto es, recopilar un gran conjunto de datos UGC durante un período de tiempo específico sobre un tema concreto y analizarlo de forma cuantitativa (Stieglitz et al, 2018). El proceso de SMA viene marcado por una serie de pasos que definen Weiguo & Gordon (2014):

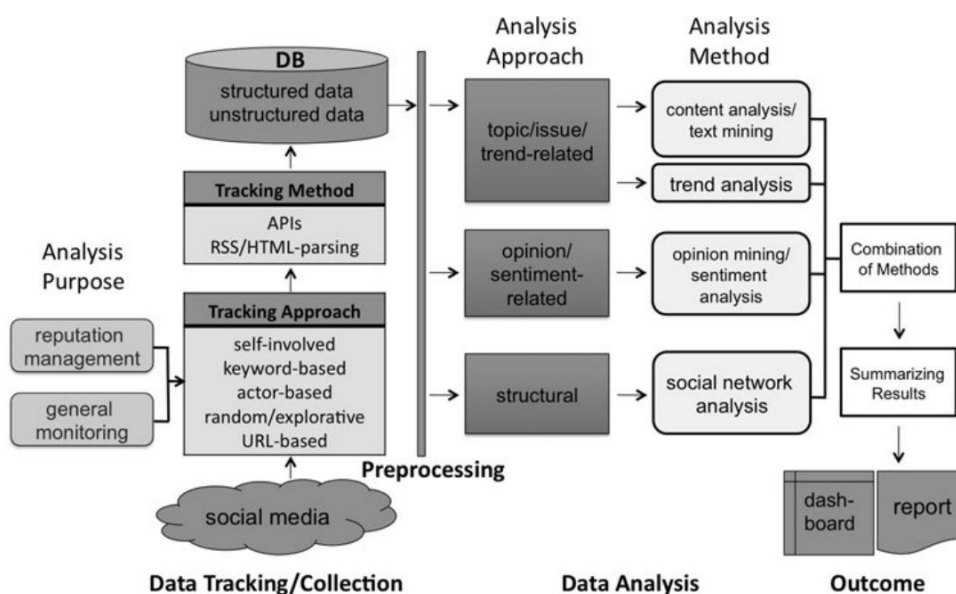
**1. Captura de datos (*Data tracking and collection*).** La etapa de descubrimiento y captura de datos permite identificar conversaciones en redes relacionadas con la actividad o el tema de interés y archivarlas. Esto se hace mediante la recopilación de enormes cantidades de datos relevantes de cientos o miles de fuentes a través del uso de APIs (*Application Programming Interface*) o mediante el rastreo. El origen de datos pueden ser tanto las plataformas más populares (Facebook, LinkedIn, Twitter, YouTube, etc.) como fuentes más pequeñas y especializadas (foros, blogs, sitios de noticias, podcasts, etc.).

**2. Preparación o "limpieza" de datos (*Preprocessing*).** Dado que la etapa de captura recopila datos de muchos usuarios y fuentes, una parte considerable puede tener "ruido" y, por lo tanto, debe eliminarse antes de realizar un análisis significativo. Para ello pueden llevarse a cabo varias fases de preprocesamiento, incluida la vinculación de datos y registros de diferentes fuentes, el etiquetado de partes del discurso, la extracción de características, el modelado de datos y otras operaciones sintácticas y semánticas que respaldan el análisis. Posteriormente a la limpieza, se utilizan clasificadores de texto de complejidad variable (consultar subapartado 2.3. Técnicas NLP) que permiten analizar el conjunto de datos según una serie de parámetros predefinidos.

**3. Análisis de datos (*Data analysis*).** La etapa de análisis es el núcleo de todo el proceso SMA. Evaluar el significado de los datos una vez procesados abarca métodos estadísticos y técnicas derivadas de la extracción de texto, NLP, la traducción automática y el análisis de

redes. El análisis de datos proporciona información sobre el sentimiento de los usuarios: opiniones de clientes acerca de una empresa y sus productos, hábitos y comportamientos populares, reacciones frente a sucesos, respuestas a campañas publicitarias, etc. En esta etapa se pueden producir muchas métricas y tendencias útiles sobre los usuarios, cubriendo sus antecedentes, intereses, preocupaciones y redes de relaciones.

**4. Visualización (Outcome).** Los resultados del análisis se resumen y muestran a los usuarios en un formato fácil de entender (paneles de visualización, informes, etc.). Pueden utilizarse técnicas de visualización que agreguen y muestren información de múltiples fuentes y la presenten de forma útil e intuitiva. El análisis visual ayuda a dar sentido a grandes cantidades de información, incluidos patrones que son más evidentes para las personas que para las máquinas.



**Imagen 2.** Esquema del proceso SMA.

Fuente: Stieglitz & Dang-Xuan, 2013.

Estas técnicas de SMA (entre las que se encuentra el análisis de sentimientos u *opinion mining* aplicado al UGC, sobre el que se hablará en el subapartado 2.3. Técnicas NLP) han ganado una aceptación considerable dentro de los grupos de investigación académica (Stieglitz et al, 2018), sin embargo, se trata de un campo emergente sobre el que todavía no existe un criterio de actuación estandarizado que sirva de guía a los investigadores para futuras aplicaciones del SMA. Lo que sí está claro es que **el SMA es aplicable a cualquier ámbito de investigación que pueda extraer información relevante procedente de las RRSS (política, sociología, marketing, publicidad, psicología, economía, etc.),** y asimismo, según Stieglitz (2018), se trata de un proceso relativamente similar en cualquiera de los casos de aplicación (descubrimiento de datos, recopilación, preparación y análisis). **Por tanto, merece la pena profundizar en su estudio y debatir sobre posibles modelos y enfoques generales del SMA que puedan aplicarse también a la evaluación de temas relacionados con la política agraria y el sector agroalimentario en su conjunto.**

Tal y como defienden Rojas-Martín & Criado (2015), las RRSS han permitido a las instituciones públicas contar con un canal de comunicación multidireccional con la ciudadanía, sin intermediarios y obteniendo una respuesta casi instantánea. Esta presencia efectiva de la Administración en redes le acerca más a las demandas de los ciudadanos y contribuye a involucrarlos más en las decisiones públicas (Agostino, 2013). A estas facilidades comunicativas, se añade la posibilidad para los organismos públicos de aplicar técnicas de SMA que les faciliten gestionar el diseño de futuras políticas y servicios públicos, en línea con las indicaciones directas que les transmite la ciudadanía “a golpe de *tuit*”.

Criado & Rojas-Martín (2013) exponen de manera concisa las razones por las que las Administraciones Públicas (y por extensión, cualquier organismo político) deben hacer uso de las RRSS para extraer información de valor de sus usuarios:

- Facilidad para compartir la información y generar debate.
- Transparencia en la comunicación.
- Libertad de expresión de sus usuarios.
- Enorme repercusión de las opiniones y recomendaciones de los usuarios. La información se difunde con enorme rapidez, llegando de forma eficaz a un gran número de interlocutores.
- Incremento constante del número de usuarios que utilizan las redes sociales en los últimos meses: en la UE, el 57 % de las personas de entre 16 y 74 años participaron en RRSS en 2020 (Eurostat, 2020), un aumento de 3 puntos porcentuales en comparación con 2019. En España, esta cifra aumenta hasta un 65%.
- Facilidad en la segmentación de la información.
- Comodidad para el usuario, este está familiarizado con la herramienta y la utiliza con frecuencia, lo que le aporta confianza y credibilidad. Además, le permite interactuar cómodamente desde su dispositivo móvil, esté donde esté.
- El ciudadano puede compartir sus propias propuestas directamente.
- Utilización de diferentes soportes en la comunicación (texto, vídeo, imagen).
- Creación de grupos o foros de debate o de trabajo, para dialogar sobre temas concretos o relevantes.
- Permiten la geolocalización de los participantes y contenidos.
- Uso en múltiples dispositivos.
- Coste reducido para la organización.

Antypas, Preece & Collados (2022) profundizan en la idea de que las RRSS se han vuelto extremadamente influyentes en lo que respecta a la formulación de políticas en el mundo occidental. Los usuarios debaten en redes sobre temas de actualidad y futuras acciones de los políticos, discuten las consecuencias, expresan su escepticismo, furia, satisfacción general o proponen cambios (Stieglitz et al, 2018). El SMA permite, por ejemplo, estudiar estos datos e identificar los factores que influyen en la participación política en las elecciones (Johannessen & Følstad, 2014). Mediante la aplicación de SMA, los gobiernos y los partidos políticos buscan obtener información para desarrollar estrategias políticas útiles de cara a las citas electorales (Vaccari et al., 2013). Las posibilidades que ofrece el SMA han acelerado la participación de figuras públicas y partidos políticos en redes como Twitter. Utilizando este nuevo altavoz, los políticos expresan sus opiniones, debaten temas



de actualidad y promueven su agenda política con el fin de influir en el comportamiento de los votantes (Antypas, Preece & Collados, 2022). Por otro lado, estas plataformas permiten a sus usuarios seguir a los políticos e interactuar con ellos, incrementando el interés de la ciudadanía por la discusión política.

Stieglitz et al (2018) comparten una serie de estudios recientes que se centran en la clasificación de *tuits* en función del sentimiento que denotan (positivo, negativo o neutral) para investigar la popularidad de cierto candidato político o bien para extraer la intención de voto de los usuarios. Este tipo de técnicas, conocidas por el nombre de *sentiment analysis* u *opinion mining* (ver subapartado), también se han empleado en varios países para comprobar si existe alguna correlación entre el sentimiento extraído de un conjunto de publicaciones en redes y los resultados electorales obtenidos. Estos mismos autores defienden que el sentimiento afecta a la difusión de mensajes en RRSS, siendo aquellos mensajes con una mayor carga emocional los que tienden a compartirse con mayor rapidez comparado con otros mensajes de tono neutral. Antypas, Preece & Collados (2022) profundizan en esta idea de las emociones en su estudio acerca de la relación entre el sentimiento de un *tuit* político y la viralidad que alcanza, para lo cual recogieron un enorme conjunto de *tuits* recientes de miembros de los parlamentos de Grecia, España y Reino Unido, realizando un análisis multifacético que tenía en cuenta aspectos contextuales tan influyentes como el partido político al que pertenecía el parlamentario, el contexto temporal y la ubicación, entre otros. Según los autores, existe una fuerte relación entre el sentimiento que denota un *tuit* político y la popularidad que alcanza, con *tuits* de carácter negativo que muestran una mayor difusión en redes frente a aquellos que transmiten un sentimiento positivo, y esta tendencia parece haberse acentuado en los últimos años, lo que podría indicar que la sociedad cada vez sufra una mayor polarización.

Pero las RRSS no sólo permiten conocer la opinión de la población acerca de ciertas figuras políticas influyentes. Recuero-Virto & Valilla-Arróspide (2022) revelaron que nuevas innovaciones de la industria agroalimentaria (biotecnología, *sustainable farming*, etc.) eran más susceptibles de rechazo por parte de los ciudadanos en base a las opiniones recogidas de usuarios de Twitter. **Conocer estos datos podría ayudar a empresas y administraciones públicas a diseñar futuros planes para hacer frente a las reticencias de los consumidores, como campañas educativas y de divulgación (también por RRSS), así como a formular nuevas políticas públicas que tengan en cuenta la postura mayoritaria de la población.** Asimismo, el estudio reveló aquellos temas sobre los que la población adoptaba una postura mayoritariamente neutra (por ejemplo, el hambre o el agotamiento de los recursos del planeta). Esta detección temprana de la indiferencia de los usuarios sobre ciertos temas cruciales podría ayudar a desarrollar una respuesta rápida que haga que los usuarios modifiquen su percepción y acepten las estrategias o innovaciones propuestas.

Concretamente, Twitter parece haberse establecido como la principal plataforma *online* donde políticos y ciudadanos interactúan para comentar temas sociopolíticos (Antypas, Preece & Collados, 2022). Se trata de una de las redes sociales con un mayor número de usuarios, con alrededor de 500 millones de *tuits* publicados por día (Araujo Britto Sass et al., 2020), lo que la convierte en una excelente plataforma de comunicación pública. La variedad de temas tratados en Twitter hace de él un canal de información prácticamente

ilimitado, del cual poder extraer una enorme cantidad de datos relacionados con cualquier temática para posteriormente analizarlos.

Twitter ofrece noticias en tiempo real y un espacio a sus usuarios para tratar una amplia variedad de temas y publicar información acerca de sus rutinas diarias y hábitos de consumo (Ramos-Sandoval & Beltran, 2021). Dado que comer y beber son algunas de las actividades humanas fundamentales, así como actividades de disfrute y de ocasión para socializar, los *tuits* relacionados con esta temática pueden ser una fuente de datos muy útil para conocer el comportamiento del consumidor respecto a los alimentos y la forma actual de producirlos (Moreno-Sandoval et al, 2018), así como sobre sus elecciones dietéticas y los paradigmas de consumo actuales. Las ventajas que presenta Twitter frente a otras plataformas, según Ramos-Sandoval & Beltran (2021), residen en su capacidad para recopilar una enorme cantidad y variedad de datos de los usuarios (*tuits*, perfiles de usuarios, mensajes directos, listas, etc.) en un corto período de tiempo y a un bajo coste. Asimismo, el gran número de usuarios que tiene Twitter (211 millones de usuarios activos en 2021 según datos de la compañía) supone una cantidad enorme de sujetos potenciales de estudio, que supera con creces el alcance que pueden tener otras metodologías de recolección de datos, como son las encuestas especializadas. (Moreno-Sandoval et al, 2018) La ventaja de contar con una recolección de información continua permite, además, un aumento significativo del volumen de usuarios que pueden ser estudiados a lo largo del tiempo, pudiendo identificar patrones o cambios de comportamiento, lo cual constituye un aspecto muy relevante.

De todo ello, cabe concluir que **el uso del UGC en RRSS, específicamente en Twitter, al ser la plataforma que se ha erigido como referente de la comunicación política en redes, puede ser una herramienta de captación de datos complementaria a otras formas tradicionales de recogida de información para la evaluación y configuración de políticas públicas, tales como la elaboración de encuestas.**

ENCUESTAS	UGC
Tamaño de la muestra condicionado por aspectos operativos y económicos	Permite tamaños de muestra de un enorme volumen
Complicado para definir y acotar la muestra	Complicado para definir y acotar la muestra
Coste alto	Coste bajo
El contexto en el que el usuario responde a la encuesta es forzado	Los usuarios expresan sus opiniones de forma libre y espontánea
Riesgo de sesgo alto: sesgo de muestreo, ausencia de respuesta, sensibilidad de los encuestados al cuestionario, sesgo del entrevistador, errores en la codificación	Sesgo de muestro
Exige mucho tiempo	Rápido y eficiente
Respuestas concretas	Más próximo al pensamiento del consumidor
Es necesario formar a los entrevistadores	Es necesario el preprocesamiento de los datos
Versatilidad de formato	Permite estudiar a los usuarios a lo largo del tiempo de forma muy fácil

**Tabla 2.** Comparativa de las ventajas y limitaciones de las encuestas tradicionales vs UGC a la hora de recopilar información para la evaluación de políticas.

Fuente: Elaboración propia a partir de (Treasury Board of Canada Secretariat, 1998).

### 2.3. Técnicas de Procesamiento del Lenguaje Natural

#### – ¿Qué son las técnicas NLP?

El auge del UGC en RRSS y el interés en explotar estos datos plantea nuevas dificultades para capturar, examinar y traducir este contenido escrito, ya que la información publicada está dispersa, es confusa y no estructurados (Alsaeedi & Khan, 2019).

El Procesamiento del Lenguaje Natural (NLP, por sus siglas en inglés) es un campo de las ciencias computacionales, la IA y la lingüística enfocado a enseñar a los ordenadores a comprender declaraciones o palabras del lenguaje humano (Srivastava, R. et al, 2022). Fue diseñado para facilitar las comunicaciones entre los usuarios y los ordenadores por medio del lenguaje natural. A continuación, se hace un resumen de las tareas más importantes que el NLP es capaz de ejecutar, siguiendo el esquema de SAS Institute (2021):

- Clasificación y resumen de contenido. Incluyendo búsqueda e indexación, alertas de contenido y detección de duplicidades. Generación automática de sinopsis de grandes cuerpos de texto.
- Extracción contextual automática de información estructurada de fuentes basadas en texto.
- Análisis de sentimiento. Identificación del estado de ánimo u opiniones subjetivas en texto.
- Conversión de habla a texto y de texto a habla. Transformación de comandos de voz en texto escrito y viceversa.
- Traducción automática de texto o habla de un idioma a otro.

Por todo ello, en los últimos años el NLP ha demostrado ser muy útil para la extracción de información de valor comercial (Alsaeedi & Khan, 2019). Cada año, se revisan y se proponen nuevas tareas, que van desde clásicos como el análisis de sentimientos hasta la detección de ironía o predicción de emojis, y se comparten y analizan grandes conjuntos de datos.

#### – El análisis de opiniones o sentiment analysis

De acuerdo con Patel y Patel (2020), el análisis de opiniones, más conocido por los términos en inglés *sentiment analysis* (SA) u *opinion mining*, es una rama del NLP enfocada en extraer de un texto opiniones, sentimientos, actitudes, emociones o puntos de vista sobre un tema de interés y clasificarlos en distintas categorías (“positivo”, “negativo” o “neutral”), según el contexto en el que se formulan. En ocasiones, el sistema de clases puede ser sustituido por una escala numérica subjetiva (Feldman, 2013). Por ejemplo, desde “-1” para las impresiones más negativas hasta “+1” para las impresiones más positivas, siendo “0” el indicativo de neutralidad. Esta tarea analítica puede probarse en cualquier tipo de texto, incluyendo publicaciones de las RRSS.

El SA recurre al NLP, la minería de textos y la lingüística computacional para reconocer, extraer, evaluar y examinar metódicamente los estados emocionales y la información subjetiva que refleja un texto para tratar de determinar la postura del autor frente a un tema particular a través de su respuesta emocional y su forma de comunicación (Alsaeedi & Khan, 2019). Las técnicas de análisis de opinión existentes a menudo tratan de identificar palabras, frases o patrones que revelen un punto de vista (Valsamidis et al, 2013). Sin

embargo, se trata de un proceso complejo, ya que no es solo la presencia de una palabra clave lo que importa, sino el contexto en el que se formula.

Las empresas han utilizado ampliamente el análisis de opiniones para detectar las preferencias de los clientes y su actitud respecto a sus marcas, productos y servicios (Antypas, Preece & Collados, 2022). Sin embargo, el valor del análisis de sentimientos no solo se aprecia en entornos comerciales. Se ha utilizado con éxito en numerosos temas sociopolíticos, tales como la perspectiva del público acerca de la pandemia del coronavirus (Barkur & Vibha, 2020), los estereotipos de género (Bhaskaran & Bhallamudi, 2019) o bien para conocer cómo los votantes valoran a sus representantes políticos (Antypas, Preece & Collados, 2022).

El proceso de *sentiment analysis* lo explica Feldman (2013) de la siguiente forma: el documento objeto de estudio (en formato PDF, HTML, XML, Word, entre otros), llamado corpus, se convierte en texto y se procesa previamente, es decir, se “limpia”, utilizando una variedad de herramientas lingüísticas, como derivación, tokenización, etiquetado, extracción de entidades y extracción de relaciones (consultar Apartado 3. Metodología). Posteriormente, el sistema utiliza una serie de léxicos y recursos lingüísticos a su alcance para clasificar el documento pre-procesado en base a las anotaciones de sentimientos. Estas anotaciones se pueden hacer a nivel del documento completo, a nivel de oraciones individuales o a nivel de sintagmas, como se explica más adelante. Estas configuran el resultado del sistema (el *output*) y se pueden presentar al usuario mediante una serie de herramientas de visualización.

Las técnicas de *opinion mining* pueden probarse a distintos niveles del texto,

- Sobre el conjunto del documento: se hace la prueba sobre el conjunto del texto analizado, obteniendo un único resultado (“positivo”, “negativo”, “neutral”) para todo el documento (Patel&Patel, 2020), pudiendo este ser un extracto de una novela, la transcripción de un diálogo o bien un conjunto de *tuits*.
- Sobre una oración: un único documento puede contener múltiples opiniones incluso sobre un mismo tema (Feldman, 2013). A veces, es necesario tener una visión más detallada de las diferentes opiniones. En el análisis a nivel de oración, cada oración completa de un texto es analizada y clasificada. Cada oración presenta, por tanto, su propia polaridad. Posteriormente, puede calcularse el sentimiento global del documento a partir de los resultados obtenidos en cada oración (sentimiento agregado).
- Sobre un sintagma: la lengua inglesa distingue entre el término “sentence”, entendiendo el mismo como una oración completa que cuenta con sujeto y predicado, y el término “phrase”, cuya traducción al castellano, “sintagma”, hace referencia, según la Real Academia Española, a un grupo de palabras en las que se puede descomponer una oración, con una estructura interna lógica articulada en torno a un sustantivo (nombre), un adjetivo, un verbo, un adverbio, etc. Implica por tanto una mayor profundidad a la hora de analizar cada oración por separado así como el texto global en su conjunto (Patel&Patel, 2020). Este enfoque es útil cuando el autor habla de distintos atributos de la entidad o temática en cuestión y presenta una opinión diferente para cada uno de estos atributos (Feldman, 2013).

A continuación, se explican en detalle los diversos métodos existentes para el análisis de opiniones. Estos pueden clasificarse fundamentalmente bajo tres categorías: el método de anotación manual, el método de aprendizaje supervisado o *Machine Learning* (ML) y el método de aprendizaje no supervisado o *lexicon-based*. El acceso de miles de usuarios a este tipo de técnicas y su aplicación a múltiples dominios han contribuido a que continuamente aparezcan nuevas metodologías, modelos de ML y *Deep Learning* (DL), *lexicon* adaptados a nuevos dominios, etc., y por tanto, a que cada vez existan más enfoques aplicables al análisis de opiniones y con una mayor precisión.

#### *Método de anotación manual*

La anotación manual no es una técnica computacional, ya que requiere de la intervención humana para el análisis y la clasificación del texto. Sin embargo, es muy frecuente recurrir a la anotación manual como forma de hacer la comparativa con los métodos de *opinion mining* basados en el aprendizaje supervisado y no supervisado, al tratarse de la metodología que refleja los resultados más fiables al haber sido elaborada por personas físicas.

Un ejemplo de una tarea de anotación manual se llevó a cabo en Antypas, Preece & Collados (2022), donde se designaron tres hablantes nativos como anotadores, a quienes se les confió un conjunto de *tuits* cuya polaridad debían evaluar en base a una escala numérica del 1 al 5. Para consolidar las anotaciones, debe haber un nivel aceptable de consenso entre todos los anotadores.

#### *Método de aprendizaje supervisado (o enfoque ML)*

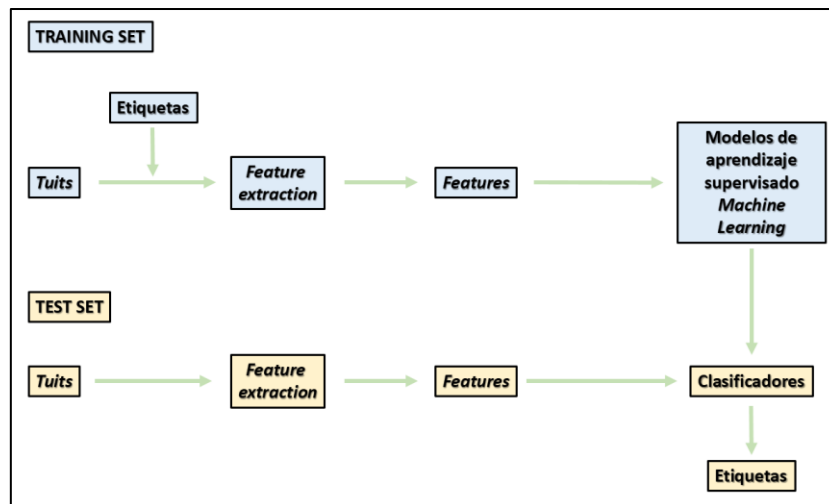
En este método, empleando un conjunto de datos históricos de los que conocemos el sentimiento, se puede determinar el sentimiento de un nuevo texto (Melcher, 2018). Para ello, se trabaja con dos conjuntos de datos: el *training set* y el *test set*.

-*Training set*: se trata de un texto clasificado manualmente, que se utiliza para entrenar el algoritmo (Bermeo-Almeida et al, 2019). El proceso de etiquetado (*labelling*) es costoso y lleva mucho tiempo.

-*Test set*: una vez que el modelo ha sido entrenado, este se prueba en el *test set* (no clasificado) para ver qué etiquetas devuelve.

Finalmente, existe un tercer conjunto de datos anotados que se reservan y se utilizan para probar el sistema periódicamente después de que se hayan producido varias iteraciones de prueba y mejora (Bermeo-Almeida et al, 2019).

El rendimiento de los clasificadores de sentimientos depende en gran medida de la cantidad de datos que constituyen el *training set* (Alsaeedi & Khan, 2019), así como del proceso de *feature extraction* (transformación de un grupo de datos a un conjunto reducido de características para facilitar su procesamiento y evitar redundancias).



**Imagen 3.** Análisis de sentimiento empleando el método de aprendizaje supervisado.  
Fuente: Elaboración propia a partir de Alsaeedi & Khan, (2019).

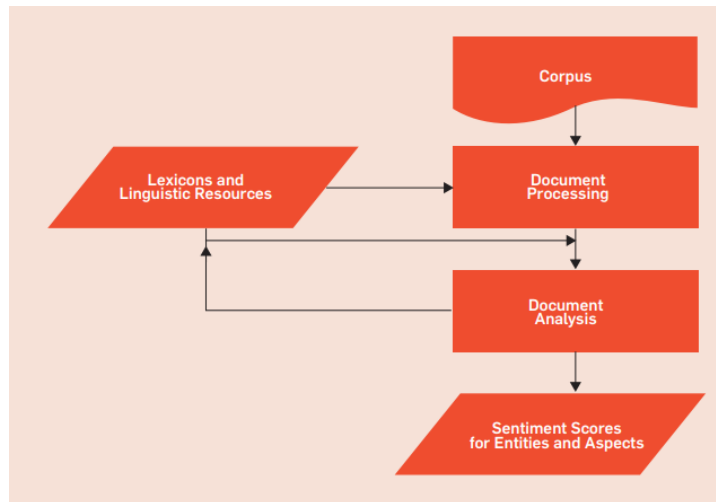
Existen cientos de modelos de ML que pueden realizar tareas de *opinion mining*, entre ellos:

- Modelos ML clásicos: Algunos ejemplos son los algoritmos para clasificación de sentimientos Naïve Bayes, Support Vector Machine (SVM) o RoBERTa (Bermeo-Almeida et al, 2019).
- Modelos DL o *transformer-based models*: de aparición reciente, los modelos de lenguaje basados en transformadores, como GPT, BERT o XLNET, han transformado el campo de la NLP, superando a los modelos lineales y redes neuronales anteriores (Barbieri et al, 2020). Su funcionamiento es ligeramente distinto a los modelos ML tradicionales. Primero, se entrenan en un gran corpus sin etiquetar y posteriormente, se ajustan a la tarea en la que existe un *training set* apropiado.

#### *Método de aprendizaje no supervisado (o lexicon-based)*

El aprendizaje no supervisado entrena al algoritmo utilizando información que no está clasificada ni etiquetada, agrupando información no ordenada basándose en similitudes, patrones y diferencias que encuentre entre los datos no etiquetados y el *lexicon* (diccionario o corpus con el que se hace la comparativa), sin requerir de un entrenamiento previo de datos (Patel&Patel, 2020). El proceso consiste en buscar cada palabra dentro del *lexicon* y asignarle un valor positivo, negativo o neutro, obteniendo un valor promedio que represente la polaridad global del texto (Bermeo-Almeida et al, 2019).

Las técnicas de aprendizaje no supervisado pueden limitarse a un enfoque simplificado, basado en un recuento de la frecuencia de aparición de ciertas palabras clave dentro de un texto (Young&Soroka, 2012), o bien pueden ser más complejas y estudiar el sentimiento agregado de una oración o incluso de un texto completo. Esto último se consigue por medio del análisis a nivel de sintagma. Dividiendo el texto en una serie de frases de pequeña escala, se determina el sentimiento de un texto completo agregando la polaridad de cada microfrase o sintagma (Alsaeedi & Khan, 2019). El análisis a nivel de sintagma permite, entre otros, invertir la polaridad cuando se encuentra una negación en el texto.



**Imagen 4.** Arquitectura del método de aprendizaje no supervisado para análisis de sentimientos.  
Fuente: (Feldman, 2013).

Los *lexicon* contienen una lista de miles de palabras positivas y negativas que se utilizan para determinar la polaridad del sentimiento de un texto en función de si el conjunto de datos en origen contiene un mayor número de palabras positivas o negativas (Mitra, 2020). Por tanto, la clasificación del texto dependerá enormemente del *lexicon* empleado, su tamaño y su adecuación al dominio o contexto en cuestión. Algunos de los *lexicons* más utilizados para detectar el sentimiento de un conjunto de *tuits* son: SentiWordNet, iSOL, eSOL, ML-Senticon, MPQA, Vader y Afinn (Srivastava, R. et al, 2022).

#### *Comparativa de los distintos métodos para el análisis de sentimientos*

Los enfoques computacionales (*lexicon-based* y sobre todo ML) para el análisis de sentimientos tienen el potencial de remediar los problemas de escalabilidad y replicabilidad inherentes a la codificación manual (Van Atteveldt et al, 2021).

Las metodologías *lexicon-based* dependen de la presunción de que la polaridad de un texto puede medirse sobre la base de la polaridad de las palabras que la componen (Alsaeedi & Khan, 2019). Sin embargo, debido a la complejidad del lenguaje, tal enfoque presenta una limitación básica ya que no se toman en cuenta numerosos aspectos del lenguaje, como la negación y otros aspectos como la ironía o el sarcasmo.

Asimismo, los métodos de aprendizaje no supervisado, también conocidos como “*off-the-shelf*”, generalmente son ejecutados para una única tarea y validados en un dominio específico. Al estar basados en diccionarios ya existentes extraídos directamente de un repositorio, es frecuente que no se adapten bien a realizar nuevas tareas. En un estudio conducido por Young y Soroka (2012), distintos diccionarios empleados para evaluar el sentimiento de diversos textos de índole política, mostraron escasa similitud en los resultados, y una escasa correlación con las anotaciones manuales hechas para los mismos.

Del mismo modo, los modelos ML, especialmente los modelos DL, si bien muestran resultados más precisos que los métodos no supervisados, aún no alcanzan el nivel de validez generalmente requerido para los métodos de procesamiento del lenguaje (Van Atteveldt et al, 2021). A este hecho se suma la dificultad para clasificar grandes cantidades de datos para la elaboración de los datos de entrenamiento (Alsaeedi & Khan, 2019), algo que no ocurre en los enfoques de análisis de sentimiento no supervisados. Los métodos de

aprendizaje no supervisados se consideran cada vez más vitales a medida que aumenta el volumen de información no etiquetada en las RRSS. Por ejemplo, para detectar opiniones en Twitter en tiempo real (Feldman, 2013).

Un estudio concluyó que, al comparar los enfoques supervisado y no supervisado, el método de ML (que empleó el modelo SVM) ofrecía un rendimiento superior con una precisión TF-IDF del 96,3% respecto al método no supervisado, que empleó el *lexicon* VADER alcanzando una precisión del 88,7% (Srivastava, R. et al, 2022). No obstante, ambas técnicas obtuvieron una precisión muy alta, condicionando entonces la utilización de uno u otro método a las condiciones particulares del de cada caso de estudio.

Anotación manual	<i>Lexicon-based</i>	<i>Machine Learning</i>
Precisión	Precisión	Precisión
Rentabilidad	Rentabilidad	Rentabilidad
Tiempo	Tiempo	Tiempo
Adaptabilidad	Adaptabilidad	Adaptabilidad
Limitaciones lingüísticas	Limitaciones lingüísticas	Limitaciones lingüísticas
Detección de polaridad	Detección de polaridad	Detección de polaridad

**Tabla 3.** Ventajas e inconvenientes de cada método de análisis de opiniones.

Fuente: Elaboración propia a partir de múltiples fuentes.

Observando las ventajas y desventajas que presenta cada enfoque, cabe concluir que todavía **es necesaria una supervisión manual de los modelos automáticos y semi-automáticos actuales, para garantizar la validez del análisis de sentimiento aplicado** (Van Atteveldt et al, 2021). Un posible enfoque híbrido puede consistir en una primera validación manual que permita ajustar el *lexicon* o el modelo a un nuevo contexto, y posteriormente continuar con el método de aprendizaje automático elegido, reduciendo así el tiempo de análisis y los costes del mismo. De hecho, múltiples estudios han hecho la prueba de evaluar el sentimiento con ambas técnicas para un mismo conjunto de *tuits*, obteniendo mejores resultados que con cualquiera de las dos técnicas por separado. Alsaeedi & Khan (2019) hacen una clasificación de distintos métodos híbridos para el análisis del sentimiento en Twitter, mostrando cómo en todos los casos, los estudios llevados a cabo demostraron la mejoría obtenida gracias a la combinación de ambas técnicas.

Los algoritmos de aprendizaje automático, como Naive Bayes, Maximum Entropy y SVM, lograron una precisión de aproximadamente el 80%. Los algoritmos híbridos de análisis de sentimientos en Twitter tendieron a funcionar mejor que las técnicas de aprendizaje automático supervisado, ya que pudieron lograr una precisión de clasificación de aproximadamente el 85% (Alsaeedi & Khan, 2019).

ESTUDIO	MÉTODO	MODELOS	DATASETS	RESULTADOS
Balage Filho and Pardo	<b>Híbrido</b>	SVM (como modelo ML) y SentiStrength	SemEval-2013 (de	F-Score de 0.563 frente al obtenido por SVM solo (0.499)



		(como modelo <i>lexicon-based</i> )	libre acceso)	
Ghiassi et al.	<b>Híbrido</b>	DAN2 (ML) y Twitter-specific ( <i>lexicon</i> )	Propios	Para el grupo negativo, la precisión ( <i>accuracy</i> ) fue de 92.5 frente a 91.45 del SVM solo.  Para el grupo positivo, la precisión ( <i>accuracy</i> ) fue de 68.2 frente a 67.6 del SVM solo.
Khan et al.	<b>Híbrido</b>	EEC (ML), IPC (ML) y SentiWordNet ( <i>lexicon</i> )	Propios	<i>Accuracy</i> de 85.7
Zainuddin et al.	<b>Híbrido</b>	Análisis de Componentes Principales (ACP) y SVM (ML)	STS, HCTS, STC (de libre acceso)	<i>Accuracy</i> de 76.55 (para STS), 71.62 (HCTS) y 74.24 (STC)
Asghar et al.	<b>Híbrido</b>	SC, EC, IDSC (todos modelos ML) y SentiWordNet ( <i>lexicon</i> )	Propios	F-Score de 0.88 frente al obtenido por SentiWordNet solo (0.81)

**Tabla 4.** Métodos híbridos empleados en diversos estudios para el análisis de sentimientos en Twitter.  
Fuente: Elaboración propia a partir de Alsaeedi & Khan (2019).

#### 2.4. Contribución del NLP a la evaluación y configuración de políticas agrarias

Técnicas NLP como el análisis de sentimientos se utilizan en una gran diversidad de dominios, como el turismo, las finanzas, la política y la salud, entre otros (Bermeo-Almeida et al, 2019). El UGC comenzó a minarse alrededor del año 2015, sin embargo, en el campo de la agricultura, existen hasta la fecha muy pocos estudios publicados que utilicen el NLP para analizar diferentes áreas de la industria alimentaria (Recuero-Virto & Valilla-Arróspide, 2022), tales como la experiencia del consumidor, los servicios de entrega, la seguridad alimentaria, etc. En 2019, Bermeo-Almeida et al afirmaron haber encontrado un único estudio relacionado con la aplicación del *opinion mining* en redes para prevenir la expansión de plagas en cultivos. Tampoco otras temáticas de interés en este ámbito como el apoyo a las SFSC (Moreno-Sandoval et al, 2018) o bien la competitividad del sector agropecuario en su conjunto (Borrero & Zabalo, 2021) han sido exploradas adecuadamente.

El contenido publicado en RRSS, páginas webs, blogs y videoblogs se encuentra tremendamente fragmentado. No obstante, tiene mucho potencial en términos de investigación teórica, matemática y empírica en el dominio agroalimentario. Pese a encontrarse en una etapa preliminar, el sector agroalimentario ha comenzado a explorar estas técnicas y a obtener resultados prometedores. Pese a que no hay datos publicados sobre las investigaciones internas que las empresas de la industria agroalimentaria llevan a cabo en este ámbito, a tenor de las últimas publicaciones científicas, las cuales revelan un creciente interés en este área, todo indica que el sector privado ya se está movilizandoy está llevando a cabo sus propios proyectos de SMA, enfocándolo como una herramienta más de marketing y/o análisis de mercado.

Como muestra de ello, en el año 2018 Moreno-Sandoval et al analizaron un conjunto de *tuits* (incluidos emojis) para extraer información sobre el comportamiento de los consumidores en relación con ciertos alimentos en función de su perfil demográfico (edad, sexo, nivel socioeconómico). Concluyeron que, incluir el análisis de datos de Twitter como herramienta de marketing en empresas agroalimentarias, podría ser muy útil para avanzar en estrategias centradas en el cliente, puesto que permite conocer las preferencias de los consumidores por ciertos alimentos, productos o marcas, a partir de los comentarios que vierten en la red. Anteriormente, Vidal et al. (2015) ya recogieron miles de *tuits* y los clasificaron para averiguar cuáles son los temas más comentados en redes relacionados con la comida, de cara a evaluar los hábitos de compra de los consumidores, sus preferencias, tiempos, planes, etc.

En esta misma línea, Borrero & Zabalo (2021) emplearon la técnica NLP conocida como *Content Analysis* para determinar la utilidad de Twitter para el diseño de estrategias de marketing en la industria de frutos rojos. Como resultado de este análisis, encontraron diferencias significativas en los patrones de consumo de fresas de los consumidores de países de habla inglesa y los de habla hispana, así como en aquellos asuntos que más preocupaban a estos consumidores en relación a los sistemas actuales de producción de frutos rojos: los consumidores angloparlantes parecen estar más preocupados por la presencia de plaguicidas en las fresas, mientras que los hispanoparlantes *tuitean* más sobre el contenido nutricional de estas. También hallaron diferencias a la hora de consumirlos: el primer grupo prefiere consumir frutos rojos en diversos formatos como acompañando a yogures o batidos, mientras que el segundo grupo *tuitea* más sobre consumirlos en situaciones espontáneas.

Sin embargo, no solo las empresas pueden beneficiarse de las técnicas NLP. Dentro del ámbito científico-académico, los blogs han demostrado ser una fuente de información muy útil para informar a los agricultores sobre mejores prácticas de gestión agronómica (Valsamidis et al, 2013). En dicho estudio, se aplicaron técnicas de *opinion mining* a textos extraídos de blogs de agricultura para investigar la opinión de los autores y los usuarios que dejaron comentarios sobre distintos temas agrícolas. Bermeo-Almeida et al (2019) propusieron un marco de aplicación del análisis de sentimientos sobre blogs de agricultura. Aplicaron técnicas de minería de datos y SA para analizar *posts* sobre agricultura y así conocer las opiniones de los agricultores y otros usuarios sobre las plagas que afectan a cultivos como el arroz, la soja y el cacao, sustancias químicas utilizadas en el control integrado de plagas, síntomas que presentan los cultivos cuando son afectados por una plaga, etc. Conociendo toda la información disponible al respecto en este tipo de blogs, pueden tomarse acciones para reducir el uso de químicos que causen daño a la salud, prevenir la aparición y la expansión de plagas hacia nuevas zonas o conocer tratamientos alternativos contra plagas (previamente empleados en otros cultivos).

Otro estudio reciente comparó los discursos de los consumidores en redes sobre el movimiento *Zero Waste* antes y después del estallido de la pandemia de covid-19, y sus resultados revelaron que el interés por la corriente *Zero Waste* ha ido en aumento en estos últimos años (Park et al, 2022) por parte de la ciudadanía, si bien la participación activa de las empresas agroalimentarias en dicho movimiento continúa siendo escasa. Un caso de estudio diferente investigó, en Perú (Ramos-Sandoval & Beltran, 2021), la polaridad de las opiniones en Twitter acerca de las cadenas cortas de suministro de alimentos (SFSC por sus

siglas en inglés), en el marco del programa de apoyo a pequeños agricultores del gobierno peruano (#DeLaChacraALaOlla). Fueron incapaces de extraer resultados definitivos, concluyendo que los usuarios de Twitter publicaban influidos en gran medida por el contexto, siendo muy sensibles a las circunstancias del momento (la pandemia de covid-19, principalmente). Este aspecto del contexto no se debe pasar por alto a la hora de llevar a cabo este tipo de análisis, en especial en un sector como el agroalimentario, profundamente influido por factores externos volátiles como el clima, los costes energéticos, los precios de mercado o las preferencias del consumidor.

Otros investigadores han profundizado en otros temas interesantes como la seguridad alimentaria y la calidad de los alimentos. Goldberg et al. (2020) utilizaron la extracción de datos y el análisis de sentimientos para clasificar los riesgos de seguridad atribuidos al consumo de alimentos, basándose en reseñas de páginas web. Singh et al. (2018) analizaron las opiniones de miles de clientes en Twitter sobre la calidad de ciertos productos alimenticios y su proceso de distribución y venta para identificar situaciones que afectaban a la satisfacción de los consumidores.

Como se ha visto, la tecnología NLP puede aplicarse para resolver multitud de problemas y contribuir a la generación de conocimiento. En el subapartado 2.1. Políticas agroalimentarias, se ha mencionado la importancia de contar con el apoyo de productores y consumidores de cara a implementar con éxito los desarrollos sociales y tecnológicos que el sector agroalimentario necesita. Recuero-Virto & Valilla-Arróspide (2022) demostraron cómo las técnicas de análisis de sentimiento en redes pueden ayudar a conocer las opiniones de los ciudadanos: recopilaron y clasificaron más de 7.000 *tuits* en 14 temáticas diferentes relacionadas con el mundo de la alimentación. De todas ellas, un total de 6 fueron mayoritariamente percibidas como “positivas” por los usuarios (temas relacionados con *food technology* y agricultura sostenible), 5 de ellas como “negativas” (dimensiones relacionadas con la industria y escenarios de crisis) y “neutrales” las restantes. Los resultados conducen a pensar que las RRSS tienen un potencial de explotación enorme para llevar a cabo análisis cuantitativos básicos, que podrían usarse como base para análisis prospectivos en el campo de las políticas públicas agrarias (Ramos-Sandoval & Beltran, 2021), contribuyendo al diseño de nuevas políticas agroalimentarias más ajustadas a las necesidades y valores de la sociedad actual.

Las posibles aplicaciones son muy numerosas, pudiendo clasificarse de la siguiente manera:

**A. Ayuda a la toma de decisiones:** la toma de decisiones es un aspecto fundamental dentro de cualquier entidad, ya sea de carácter público o privado. Las opiniones extraídas del UGC pueden ayudar a tomar decisiones en múltiples entornos (Patel&Patel, 2020), incluyendo la formulación y la evaluación de políticas e iniciativas públicas que afecten al sector agroalimentario.

**B. Previsiones y análisis de tendencias:** las RRSS congregan a millones de usuarios a diario, y es en ellas donde surgen y se difunden a mayor velocidad las noticias de última hora, las tendencias del momento y las opiniones individuales. Estar al tanto de todo ello puede ayudar a los organismos públicos a anticipar cambios significativos en las

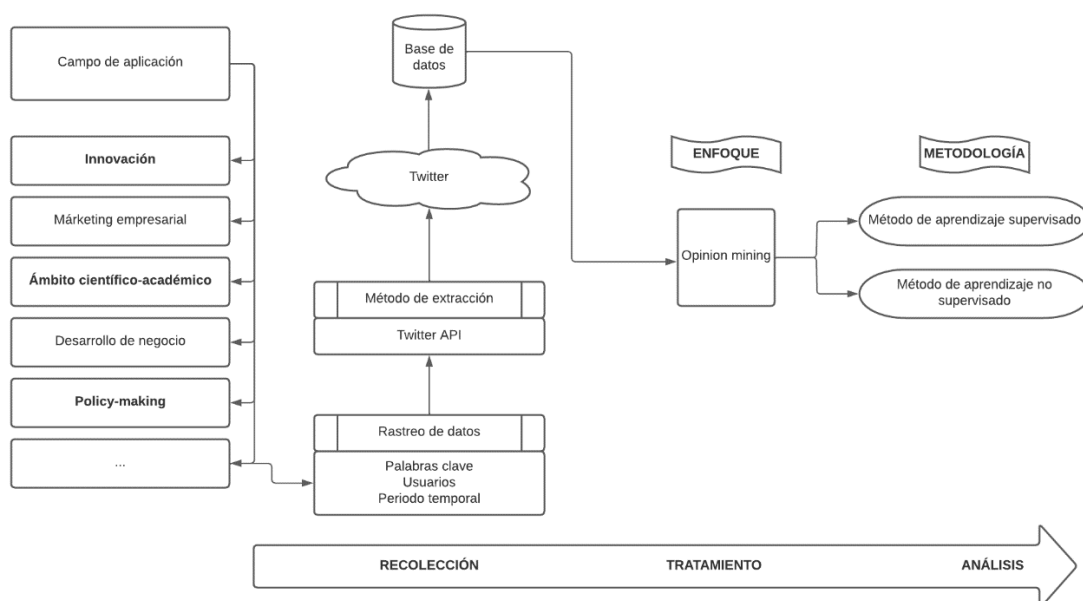
necesidades y demandas sociales de la ciudadanía y a adaptar los planes estratégicos a tiempo.

**C. Cocreación de políticas:** de manera indirecta, el estudio del UGC para la evaluación de políticas públicas contribuye al fortalecimiento de un modelo de política colaborativa, que fomente la participación ciudadana y disminuya las intermediaciones entre la organización y el individuo particular.

Por último, es importante resaltar la importancia de estas herramientas más allá de la utilidad que presentan para las instituciones públicas. Asociaciones de agricultores y cooperativas, pymes, universidades, centros de investigación, organizaciones sin ánimo de lucro, etc. pueden beneficiarse del NLP en RRSS para averiguar si el mensaje que buscan enviar a los ciudadanos se transmite correctamente y cuáles son los fallos que restan por solucionar para que, conociendo esta información, pueda generarse un impacto positivo sobre el conjunto del sector agropecuario y la industria agroalimentaria y en su forma de comunicación con la ciudadanía.

### 3. INTRODUCCIÓN A LA METODOLOGÍA

La metodología a seguir para el análisis de sentimiento de *tuits* en el presente trabajo se basa en el marco SMA que plantean Stieglitz et al en su trabajo (2018).



**Imagen 5.** Metodología de trabajo SMA aplicada a la analítica de sentimientos u opinion mining.  
Fuente: Elaboración propia a partir de Stieglitz et al (2018) y Stieglitz & Dang-Xuan (2013).

Para que el UGC resulte útil como fuente de datos para el análisis de sentimientos, es imprescindible, al igual que en cualquier otro método de recogida de datos, como son las encuestas especializadas, tener en cuenta tres aspectos básicos antes de extraer la información (Treasury Board of Canada Secretariat, 1998):

**1. Definición de los requerimientos.** Previamente y con exactitud, se define el tema a analizar para determinar qué tipo de datos ofrecerán suficientes evidencias. Es

necesario hacerse preguntas como: ¿qué se hará con los datos después de su recogida? ¿Qué conclusiones se buscan extraer del análisis de estos datos? ¿Los datos necesarios están disponibles en algún otro sitio? ¿Es el método de recogida de datos el más adecuado y eficiente en términos económicos y de tiempo? Tal y como detalla el **Anexo II: Metodología**, para este caso se ha elegido R como el lenguaje de programación más adecuado para el análisis de datos y Twitter como la red social utilizada como fuente de datos. Al mismo tiempo, ha sido necesario definir el método de extracción de los datos (Twitter API) y hacer una caracterización pormenorizada del conjunto de datos disponible.

**2. Determinación de la muestra.** La metodología de muestreo a la hora de utilizar las redes sociales como fuente de datos es necesariamente distinta a los planteamientos tradicionales de recolección de datos (Salmons, 2017). Dicha metodología deberá adaptarse a la naturaleza del estudio, la red social empleada y el público objetivo. Para este trabajo se ha configurado una muestra de *tuits* teniendo en cuenta **el volumen, la zona geográfica, el idioma, el periodo temporal, la frecuencia de extracción y las palabras clave** utilizadas.

**3. Procesamiento de datos y testeo.** Por último, el contenido que los usuarios publican en RRSS genera un enorme volumen de texto que debe filtrarse adecuadamente, simplificando la muestra (eliminación de caracteres especiales, corrección de errores ortográficos, etc.) y reduciendo su tamaño para facilitar el manejo de estos datos posteriormente. En última instancia, se debe elegir la metodología más acertada para el procesamiento y análisis de la información. En este caso, se ha seleccionado el método de aprendizaje supervisado de *opinion mining* explicado en el apartado 2.3.

Cada uno de estos puntos se detalla **en el Anexo II**.

## 4. RESULTADOS

A continuación, se han llevado a cabo dos casos de estudio para evaluar la metodología de aprendizaje no supervisado, aplicándola sobre dos temáticas diferentes dentro del ámbito de la política agroalimentaria. El caso 1 (Anexo IV) se centra en determinar e incrementar la precisión de la metodología de *sentiment analysis* seleccionada (la técnica *lexicon-based*), mientras el caso 2 aplica directamente el método sobre un conjunto real de datos que no han sufrido ningún filtrado manual previo, imitando así las condiciones que se seguirían en un proceso de analítica de sentimientos real.

CARACTERÍSTICAS	CASO 1	CASO 2
<b>Metodología</b>	Híbrida (Anotación manual + <i>Lexicon-based</i> )	<i>Lexicon-based</i>
<b>Objetivo</b>	Elección del <i>lexicon</i> + Comprobación de la precisión del modelo bajo condiciones óptimas	Validación de la metodología en un caso real
<b>Temática</b>	Agricultura ecológica	PAC
<b>Idioma</b>	Inglés	Inglés
<b>Número de <i>tuits</i></b>	275 (limitados por la anotación manual)	73.850
<b>Período temporal</b>	Mayo de 2022	2017-2022
<b>Geolocalización</b>	Sí (Europa)	No

<b>Resultados del caso</b>	<ul style="list-style-type: none"> <li>-Selección del <i>lexicon</i> más preciso y comprobación de la metodología</li> <li>-Distribución del sentimiento (positivo, negativo, neutral)</li> <li>-Temáticas más mencionadas en cada categoría de sentimiento</li> </ul>	<ul style="list-style-type: none"> <li>-Distribución del sentimiento (positivo, negativo, neutral)</li> <li>-Evolución del sentimiento a lo largo del tiempo</li> <li>-Eventos responsables de cambios bruscos en la distribución del sentimiento</li> <li>-Distribución del sentimiento en el grupo de <i>tuits</i> originales</li> <li>-Temáticas más mencionadas en cada categoría de sentimiento</li> <li>-<i>Clusterización</i> de temáticas según categoría de sentimiento</li> <li>-<i>Tuits</i> con mayor difusión según categoría de sentimiento</li> </ul>
----------------------------	--	--

**Tabla 5.** Resumen de las principales características de cada caso de estudio.  
Fuente: Elaboración propia.

#### 4.1. Caso 1: agricultura ecológica

El caso 1 está documentado en el **Anexo IV: Desarrollo del caso 1**. Para esta prueba piloto, se ha planteado el objetivo de comprobar la precisión de la metodología de aprendizaje no supervisado y los *lexicon* frente al proceso de anotación manual, que requiere de comprobación humana. De esta forma, se ha seleccionado un tema de gran interés mediático, la agricultura ecológica, para analizar si los métodos automatizados NLP son confiables a la hora de descifrar el sentimiento de un *tuit*.

Al tratarse de un proceso lento, que requiere de un anotador (o varios) que clasifiquen el sentimiento denotado por cada *tuit*, la muestra es considerablemente menor a la que se empleará posteriormente para analizar el caso 2.

#### 4.2. Caso 2: la PAC

Para el segundo caso, el objetivo principal es analizar un gran volumen de datos de Twitter para extraer el sentimiento, tratando de imitar al máximo el procedimiento que se seguiría para un caso real. En esta ocasión, se utilizará el *lexicon* y la función validados en el caso anterior para extraer información de una muestra grande de *tuits*, sin depender para ello de la anotación humana. De esta forma, se pretende comprobar la utilidad de este sistema a la hora de procesar conjuntos de datos de gran tamaño sin contar con un “filtrado” previo manual.

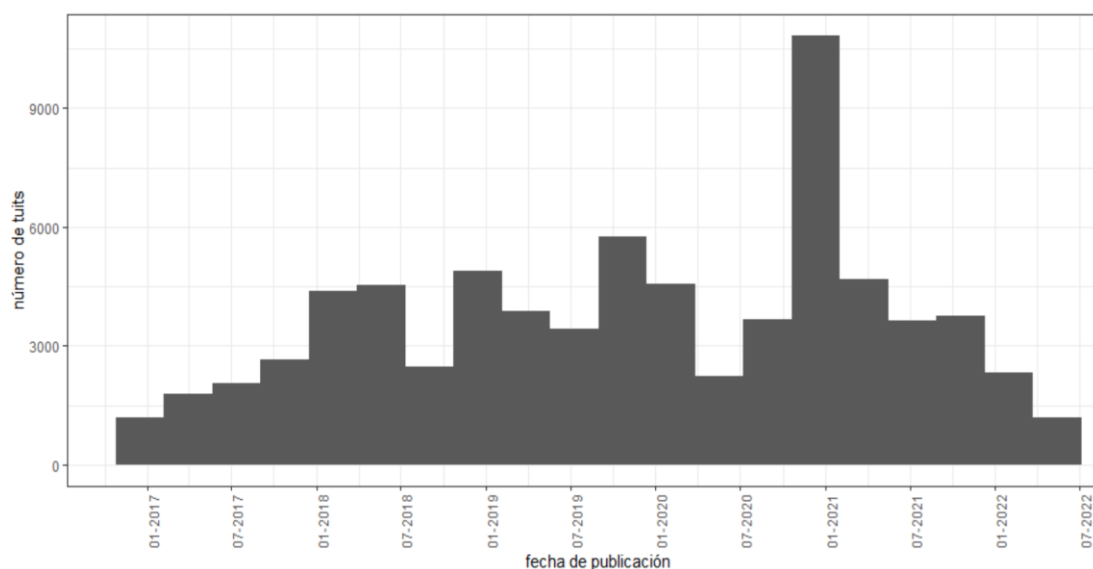
El tema elegido para analizar la opinión de los usuarios en Twitter es la PAC, en un momento en el que está a punto de entrar en vigor su nueva reforma (ver Subapartado 2.1.), lo cual ha podido favorecer la creación de conversaciones en redes que evalúan el trabajo de la anterior y sobre todo, que lanzan propuestas de mejora para la futura PAC 2023-2027. Como novedad, en esta ocasión se estudiarán *tuits* publicados desde el año 2017 hasta ahora, lo que permite asimismo observar la evolución del sentimiento de los usuarios a favor y en contra de la PAC a lo largo del tiempo. Especialmente interesante es comprobar el impacto sobre las opiniones de aquellos eventos desestabilizadores ocurridos en Europa en los últimos años, como la pandemia de covid-19, la reciente guerra en Ucrania o el Brexit.

### Conjunto de datos:

Las palabras clave escogidas fueron “*Common Agricultural Policy*”. Si bien emplear su abreviatura (CAP) hubiera devuelto un mayor número de *tuits*, dada la tendencia de los usuarios en Twitter a acortar las palabras para no exceder el límite de caracteres, en este caso, la palabra “cap” en inglés tiene distintos significados (tope, límite, gorra, tapa...) y por tanto incluirla como palabra clave devolvía miles de *tuits* no relacionados con la temática. La abreviatura en español (PAC) devolvía la mayor parte de *tuits* en español invalidando la muestra, ya que esta es contrastada posteriormente con un *lexicon* de palabras en inglés (Hu and Liu, 2004). En consecuencia, la muestra también ha de ser en inglés. Otra ventaja de seleccionar estas palabras clave fue su especificidad, reduciendo al límite el número de *tuits* extraídos que no están relacionados con la temática del caso.

El tamaño de la muestra ha sido de 73.850 *tuits*, repartidos a lo largo de casi cinco años y medio. El período temporal de la muestra se extiende desde el día 01/01/2017 hasta el día 31/05/2022. La distribución de *tuits* a lo largo del tiempo es la siguiente:

- 2017: 8.521 *tuits*.
- 2018: 14.018 *tuits*.
- 2019: 15.826 *tuits*.
- 2020: 19.451 *tuits*.
- 2021: 12.851 *tuits*.
- 2022: 3.183 *tuits* (hasta el 31/05/2022).



**Gráfico 1.** Distribución de la muestra global de tuits desde 2017 hasta la actualidad.

Fuente: Fuente: Elaboración propia en RStudio.

En esta ocasión, no se ha filtrado según la geolocalización de los usuarios para poder contar con un tamaño de muestra mayor. Pese a las ventajas que ofrece este filtro explicadas en el caso anterior, se prioriza alcanzar una muestra de gran tamaño. Asimismo, no se debe olvidar tampoco el efecto que la PAC tiene en otras regiones fuera de Europa. Reacciones desde otras partes del mundo han de tenerse en cuenta también si lo que se busca es hallar una valoración global de esta política.

Otro aspecto importante a considerar es la inclusión de los RT dentro del análisis. Incluirlos supone introducir en la muestra un mismo *tuit* (el original),  $x$  número de veces (tantas veces como RT haya conseguido), lo cual implica que un error en la clasificación de un *tuit* se reproducirá tantas veces como haya RTs, distorsionando los resultados (ver Subapartado Errores detectados). No obstante, ¿recibe el mismo apoyo un *tuit* negativo que consigue más de 1.500 RT, que otro positivo que tan sólo recibe 2 RT? Un *retuit* demuestra que un usuario está de acuerdo con lo que otro ha publicado y por ello decide difundirlo con sus seguidores. Por tanto, cabe pensar que aquel que *retuitea* el *tuit* de otra persona, comparte la misma postura sobre el tema que se trata. Incluir los RT permitirá obtener una visión más ajustada a la realidad.

Al tratarse de una gran cantidad de datos, a la hora de recopilarlos se ha seguido un flujo de trabajo riguroso, consistente en: 1) almacenar las credenciales de acceso a Twitter API (ver Anexo VII); 2) almacenar eficientemente los *tuits* devueltos y 3) vincular los datos en un *dataframe* que contenga únicamente las variables imprescindibles (fecha y texto).

*Pre-procesamiento de datos:*

La función empleada para la limpieza de *tuits* en R (ver Anexo VII) es más compleja que la aplicada en el caso 1, puesto que ahora no contamos con la comprobación adicional de la anotación manual. Existen múltiples paquetes en R que ayudan a automatizar la limpieza y *tokenización* de texto, por ejemplo, ***quanteda***. Sin embargo, esta vez se ha implementado una función propia que, si bien puede estar menos optimizada, permite seleccionar cada uno de los pasos de limpieza para así poder adaptarla al tipo de texto que se está analizando, en este caso, un conjunto de *tuits*.

En este ejercicio, dado que el principal objetivo es estudiar el sentimiento que transmiten los usuarios en un *tuit*, se procede a eliminar todos aquellos elementos del texto que no contribuyan a dicha función:

- Mayúsculas
- Patrones no informativos (URLs de páginas web)
- Signos de puntuación
- Caracteres sueltos
- Caracteres de control
- Etiquetas HTML
- Números
- Gráficos
- Espacios en blanco múltiples

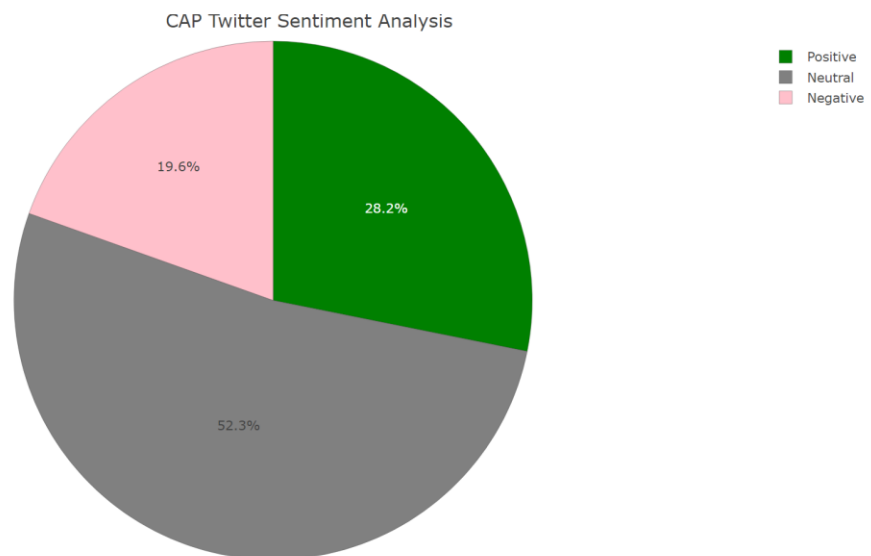


texto	texto_tokenizado
RT @skisidjames: Leaving the EU: <U+2705> Out of the cus...	rt skisidjames leaving the eu <u+ > out of the customs unio..
The Common Agricultural Policy was designed by a contine...	the common agricultural policy was designed by a continen..
@GrantWhiteTZ Margret Thatcher was a consequence of La...	grantwhitetz margret thatcher was a consequence of labou..
RT @skisidjames: Leaving the EU: <U+2705> Out of the cus...	rt skisidjames leaving the eu <u+ > out of the customs unio..
@AgriSkippy @Pat_Ronche @actufr pour la PAC on est jeun...	agriskippy pat ronche actufr pour la pac on est jeune agricu..
RT @EUAgri: Grants for information measures on the Comm...	rt euagri grants for information measures on the common a..
RT @keepcalm_change: The Common Agricultural Policy: Pr...	rt keepcalm change the common agricultural policy promoti..
@Jamken22 The EU common agricultural policy keeps price...	jamken the eu common agricultural policy keeps prices arti..
RT @EUAgri: Grants for information measures on the Comm...	rt euagri grants for information measures on the common a..
The common agricultural policy <U+0001F447> https://t.co...	the common agricultural policy <u+ f >
RT @EUAgri: Grants for information measures on the Comm...	rt euagri grants for information measures on the common a..
@Jamken22 @hodder44 The Common Agricultural Policy th...	jamken hodder the common agricultural policy that keeps f..
RT @EUAgri: Grants for information measures on the Comm...	rt euagri grants for information measures on the common a..
@TomTugendhat It is, a shame it contains misleading infor...	tomtugendhat it is a shame it contains misleading informati..

**Imagen 6.** Ejemplo de varios tuits antes y después de su “limpieza”.  
Fuente: Rstudio.

### Testeo:

En este caso, para el análisis de sentimiento se ha seguido el mismo procedimiento del caso 1, empleando la función **score.sentiment** desarrollada en el caso anterior con el *lexicon* Hu and Liu. Del conjunto de *tuits* analizados, cataloga a un 52% como neutrales, un 28,2% como positivos y solo un 19,6% como negativos. En total, se han obtenido 20.789 *tuits* positivos y 14.441 negativos. El elevado número de *tuits* de categoría neutral puede deberse a la presencia de publicaciones de medios de comunicación y otras organizaciones que informan sobre la PAC desde la objetividad, o bien que se hacen eco de los anuncios de la UE acerca de este tema o sobre la organización de webinarios o eventos que tratan esta temática, los cuales tienden a difundirse con rapidez dentro de las redes (elevado número de RT).



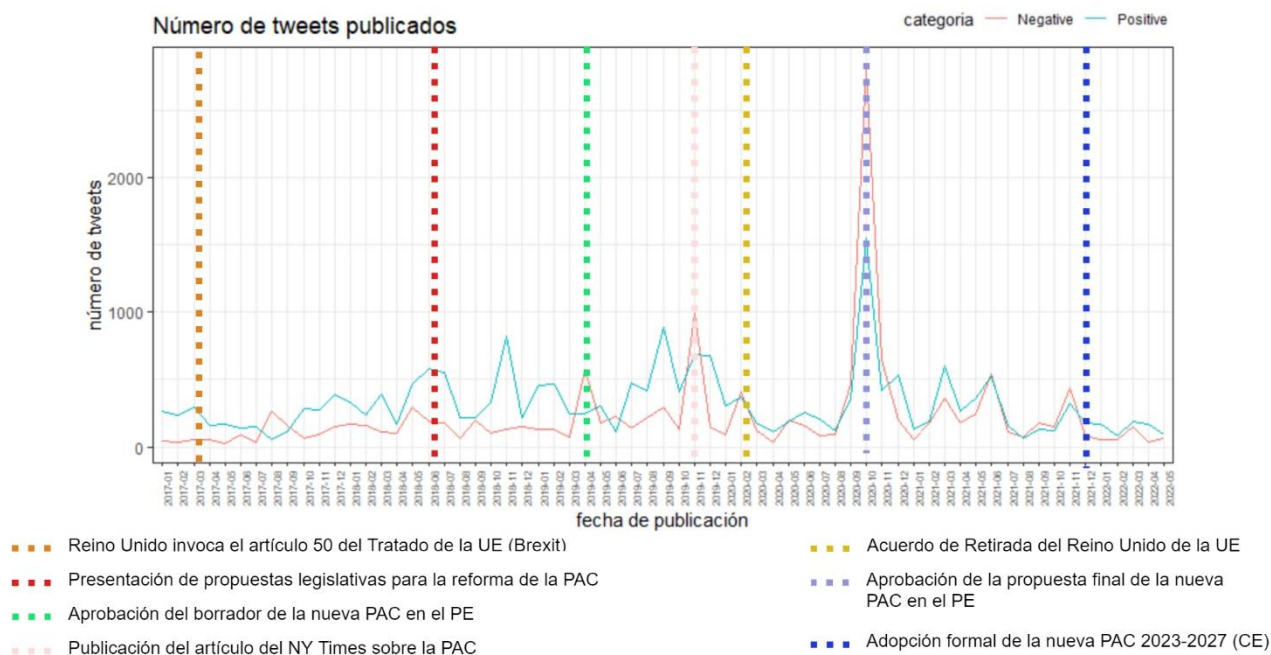
**Imagen 7.** Resultados en R utilizando la metodología de aprendizaje no supervisado.  
Fuente: Elaboración propia en RStudio.

Fecha	Texto	Puntos	Categoría
14/01/2017	RT @Macra_dduggan: Register now to have your say on any future Common Agricultural Policy. 1st of 5 sessions kick off in Navan on 18th #Mac...	0	Neutral
14/01/2017	Nature describes rethinking Common Agricultural Policy This time lets look at the evidence <a href="https://t.co/AMCoYvrUBw">https://t.co/AMCoYvrUBw</a> !	0	Neutral
16/01/2017	Agriculture Committee to discuss review of Common Agricultural Policy <a href="https://t.co/3yl6Pqm2pY">https://t.co/3yl6Pqm2pY</a>	0	Neutral
16/01/2017	RT @MacranaFeirme: Should farmers receive a pension and a Common Agricultural Policy Direct Payment? Have Your Say! #MacraCap2020 <a href="https://t.co/...">https://t.co/...</a>	0	Neutral
17/01/2017	What's @greenpeace view on the Common Agricultural Policy? Why #TheresaMay pledging to retain similar subsidies? #biodiversity #rewilding	0	Neutral
17/01/2017	RT @COPACOGECA: .@EP_ThinkTank in depth analysis on Common Agricultural Policy 2014-2020: Direct payments - a reference note. #CAP <a href="https://t.co/...">https://t.co/...</a>	0	Neutral
18/01/2017	Can the 'greening' be greener? Ecological Focus Areas under the EU's Common Agricultural Policy <a href="https://t.co/Jp9ewmLm19">https://t.co/Jp9ewmLm19</a>	0	Neutral
19/01/2017	At #igw2017 @PhilHoganEU announces start of Public Consultation on 2 Feb on new Common Agricultural Policy. "It has to be simp..."	0	Neutral
23/01/2017	Open letter from organic farmers calls for a new approach to the Common Agricultural Policy. <a href="https://t.co/AvP438Rk6r">https://t.co/AvP438Rk6r</a> <a href="https://t.co/aKNu5j09Hk">https://t.co/aKNu5j09Hk</a>	0	Neutral
23/01/2017	Copa and Cogeca outline key elements of future Common Agricultural Policy #CAP to EU Farm Ministers <a href="https://t.co/elHUhSIFFb">https://t.co/elHUhSIFFb</a>	0	Neutral

**Tabla 6.** Algunos de los tuits clasificados como "Neutral" dentro del caso 2.  
Fuente: Excel.

### Evolución temporal

Los sentimientos de las personas son volátiles y pueden cambiar con el tiempo (Patel & Patel, July 2020). Estas transformaciones pueden monitorizarse para estudiar qué eventos han podido ser motivo de cambio en la actitud de la población respecto a un tema y cómo evoluciona la tendencia general de pensamiento a lo largo del tiempo. Por ello, se va a analizar la distribución temporal de los *tuits* a lo largo de más de cinco años (2017-hasta la actualidad) según su categoría de sentimiento (positivo o negativo).



**Gráfico 2.** Evolución temporal del número de tuits publicados desglosado por categoría de sentimiento.  
Fuente: Elaboración propia en RStudio.

Puede observarse un perfil de actividad similar entre ambas categorías: cuando aumentan las publicaciones positivas, aumentan también las publicaciones negativas. A lo largo de toda la línea temporal, los *tuits* positivos son más numerosos que los negativos. Sin embargo, esta respuesta no se da en algunos momentos en los que se dispara el número de *tuits* de categoría negativa. Aquí se enumeran algunos ejemplos y sus posibles causas:

-Julio y agosto de 2017: gran parte de los *tuits* que se muestran críticos con la PAC están relacionados con la salida de Gran Bretaña de la Unión Europea, el Brexit. Tan sólo un año después del referéndum del Brexit (23 de junio de 2016), la incertidumbre acerca de los próximos pasos a seguir por parte de Gran Bretaña en materia de política agraria se refleja entre las publicaciones de los usuarios, divididos entre quienes muestran su preocupación por las consecuencias de esta ruptura para los agricultores ingleses y quienes defienden su marcha para dejar de financiar unas medidas que consideran ineficaces.

-Marzo y abril de 2019: a principios de abril, la comisión de agricultura del Parlamento Europeo aprobó una serie de propuestas para la reforma de la PAC post-2020 (Evans, 2 de abril de 2019). Coincidiendo con la aprobación del borrador, en Twitter muchos usuarios se mostraron reticentes a algunas de las medidas de la nueva PAC.

-Noviembre de 2019: La PAC fue ampliamente cuestionada después de la publicación de un artículo del NY Times, el 3 de noviembre de 2019, donde se identificaron prácticas que favorecían la concentración de tierras y la percepción de ayudas por parte de oligarcas en Hungría y otros países de Europa Central y del Este (Gebrekidan, Apuzzo & Novak, 2019), así como por parte de la mafia en Italia (Roberts, 2018). La publicación del artículo disparó los comentarios en redes cuestionando los beneficios de esta política y denunciando la necesidad de una reforma estructural.

-Octubre de 2020: se dispararon el número de *tuits* publicados, tanto positivos como negativos, coincidiendo con la semana del 19 al 23 de octubre, cuando el Parlamento Europeo debatía sobre la PAC y votaba su aprobación definitiva el 23 de octubre (Sottile, 23 de octubre de 2020). La mayoría de críticas mencionan la falta de ambición en materia de clima y protección del medioambiente. Concretamente, un *tuit* de Greta Thunberg del 25 de octubre 2020 criticando la votación por considerarla “desastrosa” para el clima, la biodiversidad y la agricultura sostenible, superó los 1.800 RT.



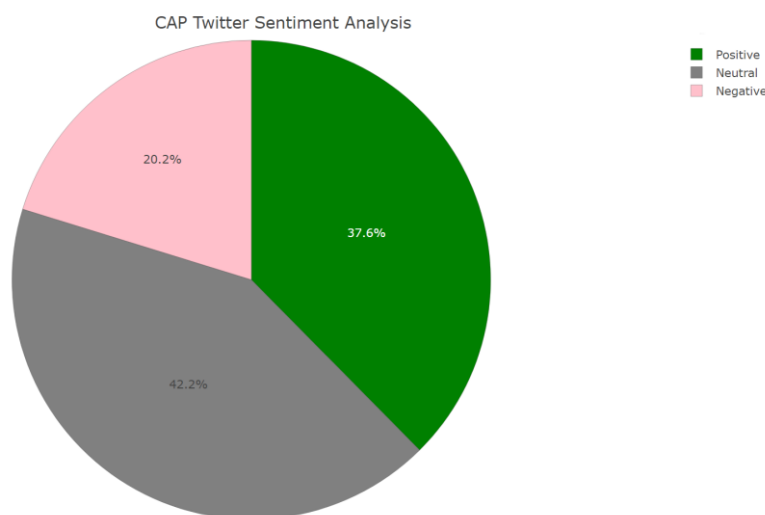
**Imagen 8.** Captura del tuit de Greta Thunberg sobre la PAC (25 de octubre de 2020), que obtuvo una gran difusión.

Fuente: Twitter.

### *Análisis de opinión de los tuits originales*

Otro aspecto que se ha estudiado es el cambio en la distribución de las distintas categorías de sentimiento cuando se tiene en cuenta únicamente el conjunto de *tuits*

originales, excluyendo los RT. Como se mencionó al principio de este caso, si bien incluir estos últimos nos da un resultado más ajustado a la realidad, excluirlos minimiza posibles errores atribuibles al modelo de cálculo. Como puede observarse (ver Imagen 9), se reduce el porcentaje de *tuits* neutrales en más de un 10%, previsiblemente porque se tratan en su mayoría de publicaciones anunciando webinaros, eventos informativos, artículos científicos y periodísticos que obtienen una amplia difusión.



**Imagen 9.** Resultados en R utilizando la metodología de aprendizaje no supervisado.  
Fuente: Elaboración propia en RStudio.

La proporción de publicaciones negativas apenas sube ligeramente, mientras que las publicaciones de categoría positiva crecen casi un 10%, lo cual indica que, si la proporción de *tuits* positivos aumenta habiendo considerado solo los *tuits* originales, **los *tuits* de categoría positiva tienden a ser menos *retuiteados* que los de categoría negativa**, en línea con lo que se verá en el siguiente subapartado (véase tabla 7).

#### *Frecuencia de palabras*

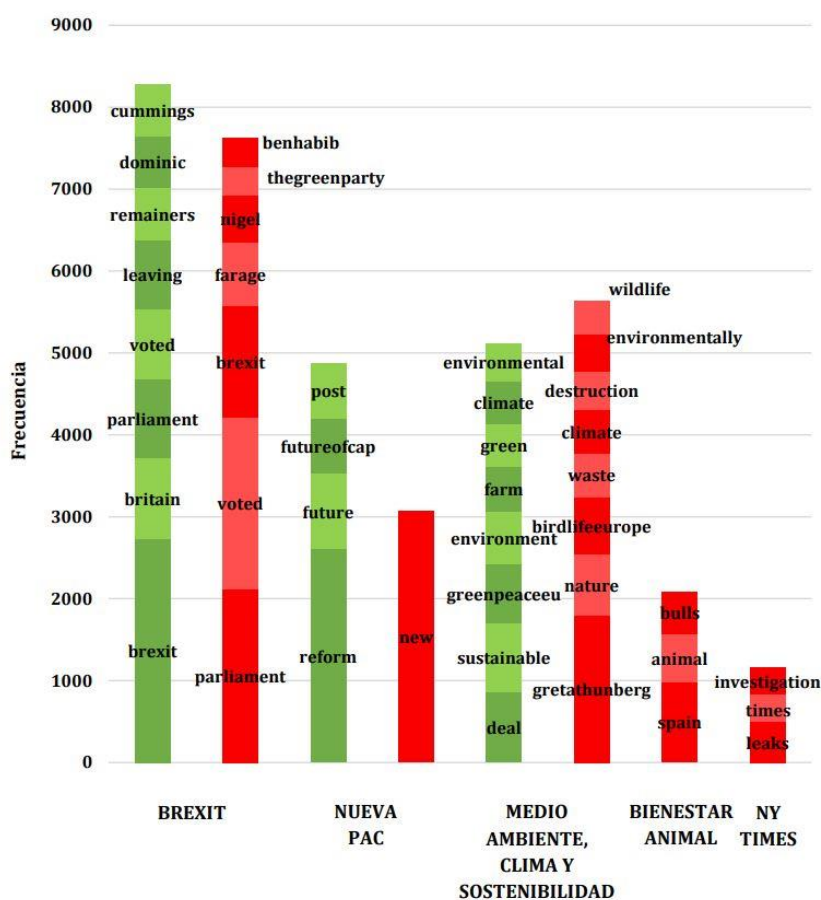
A continuación, se extrajeron las palabras más mencionadas dentro de los grupos de *tuits* categorizados como “*positive*” y “*negative*” para evaluar cuáles son los temas que la ciudadanía asocia con la PAC, qué beneficios detectan en esta política y asimismo cuáles son los problemas que le atribuyen. Para poder generar la nube de palabras, se hizo uso de la función *rquery.wordcloud* (del paquete *wordcloud*), configurándola de manera que sólo aparecieran las palabras mencionadas un mínimo de 4 veces en el conjunto de datos y asimismo eliminando del análisis las palabras clave empleadas durante la recogida de datos (“*Common Agricultural Policy*”), así como las stop words y otras palabras que no aportan valor.

Cabe destacar que en ninguna de las dos categorías aparece el covid-19 como un tema frecuentemente mencionado por los usuarios de Twitter, pese a haber provocado un retraso en la aprobación de la nueva reforma de la PAC de más de 2 años.

Los resultados que devolvió la nube de palabras para los *tuits* de categoría positiva fueron poco concluyentes (ver Imagen 10). Se aprecia cómo el principal tema de conversación en relación a la PAC es el Brexit. Se trata del tema más mencionado a lo largo de estos últimos 5 años. El siguiente término es “*reform*” y dentro de este grupo se enmarcan



En ambos casos, muchas de las palabras de la nube frecuentemente mencionadas no ofrecen detalle sobre lo que la población valora positiva o negativamente de la PAC. Incluso habiendo hecho un filtrado de las *stop words*, las nubes de palabras no son lo suficientemente aclaratorias para extraer conclusiones. Por esta razón, resulta de interés analizar las palabras por *clusters* o agrupaciones de palabras, en función de la temática de la que hablen. De esta forma, se pueden identificar con mayor claridad los beneficios y los defectos que los usuarios asocian con la PAC. Se han seleccionado las 30 palabras más mencionadas en los *datasets* positivo y negativo y se han agrupado según la temática a la que hacen referencia (ver Gráfico 3). Algunas palabras son difíciles de agrupar en un único grupo (parliament, new, animal), pero aparecen frecuentemente en publicaciones relacionadas con la temática señalada.



**Gráfico 3.** Agrupación de los temas más comentados entre el dataset de tuits positivos y negativos.  
Fuente: Elaboración propia.

Se aprecia cómo el Brexit ha sido el tema que más ha acaparado la atención de los usuarios en relación a la PAC. Si bien esto está relacionado con el hecho de que la mayoría de *tuits* son en inglés y por tanto, la cantidad de *tuits* procedentes de Reino Unido es mayor en proporción con el resto de países, es importante no olvidar la trascendencia de la salida de Gran Bretaña para la futura reforma de la PAC, al tratarse del tercer Estado Miembro que más contribuía a los presupuestos de la Unión, solo por detrás de Alemania y Francia (EAE Business School, 2019). No hay duda de que es un aspecto que preocupa a los europeos y que debe ser abordado por los países miembros, tanto en el aspecto que concierne al presupuesto global de la PAC como en la inclusión de planes de contingencia para las empresas del sector.

El otro tema más comentado está relacionado con el medio ambiente y la crisis climática. En este punto, sorprende que los *tuits* negativos superan a los positivos en cantidad de menciones, lo cual refleja que los usuarios en Twitter podrían percibir las medidas incluidas hasta ahora dentro de la PAC en materia de medio ambiente y sostenibilidad como poco ambiciosas. Si bien uno de los conceptos más mencionados dentro del grupo de *tuits* positivos es el *Green Deal*, un conjunto de iniciativas políticas aprobadas por la Comisión Europea cuyo objetivo es que Europa sea el primer continente en alcanzar la neutralidad climática para el año 2050. Este plan podría marcar la hoja de ruta de las futuras medidas de la reforma post-2020 de la PAC en materia climática y de sostenibilidad ambiental.

La configuración de una “nueva PAC” ha suscitado mucho interés en Twitter. Los usuarios comentaban los buenos resultados alcanzados hasta el momento, o bien incidían en aquellos puntos que son necesarios abordar en la nueva reforma. Prueba del interés por este asunto está también en la cantidad de *hashtags* generados sobre la nueva PAC 2023-2027, tales como #FutureOfCAP, #CAPreform, #CAPStrategicPlan o los más críticos como #VotheThisCAPDown o #WithdrawTheCAP.

En cuanto a la temática del “Bienestar animal”, se ha criticado ampliamente la destinación de fondos de la PAC a la cría de ganado de lidia. Este tema aún está pendiente del voto del Consejo Europeo después de que el Parlamento Europeo aprobara en 2020 una enmienda en el marco de la reforma de la PAC que veta destinar los pagos acoplados a ganadería de toro de lidia (La Vanguardia, 2020). Por último, la publicación del artículo del NY Times sobre la asignación de fondos de la PAC a beneficiarios bajo sospecha de llevar a cabo operaciones opacas de cesión y acaparamiento tierras ha entrado dentro de los temas más comentados (consultar Subapartado de Evolución temporal).

#### ¿De qué hablan los *tuits* con mayor difusión?

En relación con el subapartado anterior, interesa conocer cuáles son los *tuits* que han conseguido un mayor número de RT dentro de ambas categorías. ¿Coinciden estas publicaciones con las temáticas de los *clusters*? ¿Qué *tuits* tienden a ser más *retuiteados*, los de sentimiento positivo o negativo? La siguiente tabla agrupa los 7 *tuits* con mayor número de RT dentro de la categoría “Positiva” del análisis de opiniones y los 7 de la categoría “Negativa”.

Fecha	Texto	RT	Temática
25/10/2020	This week the European Parliament voted for a new 7yr Common Agricultural Policy (CAP) that will be disastrous for climate, biodiversity and sustainable farming. But it's not over yet. Sign the open letter demanding the @EU_Commission to #WithdrawTheCAP <a href="http://WithdrawTheCAP.org">http://WithdrawTheCAP.org</a>	1.769	MEDIO AMBIENTE, CLIMA Y SOSTENIBILIDAD
20/10/2020	The European Parliament has voted against any meaningful reform of the EU's common agricultural policy #CAP They have signed a death sentence for small farms, nature and the climate #FutureofCAP #CAPreform	712*	NUEVA PAC
10/09/2019	Dominic Cummings had a go at Remainers saying we were all rich elite. This is where he lives. His father-in-law lives in a castle & receives EU funding from the Common Agricultural Policy. <a href="https://theredroar.com/2019/08/revealed-anti-elite-dominic-cummings-lives-in-1-6-million-islington-townhouse/">https://theredroar.com/2019/08/revealed-anti-elite-dominic-cummings-lives-in-1-6-million-islington-townhouse/</a>	625*	BREXIT
20/09/2020	In Spain they enjoy setting bulls on fire & to torture them in barbaric arenas. Spain is a EU member since 1986. 34 years the	478	TAUROMAQUIA

	EU has failed to put an end to this torture. Even worse EU budgets have been going to support this under Common Agricultural Policy.		
20/12/2018	"The biggest waste of EU resources is Nigel Farage's salary"	471	BREXIT
25/06/2021	This pro-extinction policy will hit Europe's wildlife with extreme violence. Read our press release for more #WithdrawTheCAP #EUGreenDeal #CAPreform <a href="https://birdlife.org/europe-and-central-asia/news/press-release-bad-CAP-Deal-kill-nature-25June2021">https://birdlife.org/europe-and-central-asia/news/press-release-bad-CAP-Deal-kill-nature-25June2021</a>	360	MEDIO AMBIENTE, CLIMA Y SOSTENIBILIDAD
07/11/2019	Why is @TheGreenParty pro EU when the single most environmentally damaging policy in Europe is the Common Agricultural Policy? It costs a fortune, enables farmers to grow things we could buy more cost effectively elsewhere and uses land which should be used for reforestation.	349	MEDIO AMBIENTE, CLIMA Y SOSTENIBILIDAD
18/10/2020	1) for the last 60 years, we've been funding a campaign to exterminate Europe's wildlife. through the EU's Common Agricultural Policy (CAP), hundreds of billions of euros of public money have been funnelled into destroying wildlife-rich meadows, woods, wetlands and rivers.	333	MEDIO AMBIENTE, CLIMA Y SOSTENIBILIDAD
04/11/2019	New York Times investigation finds the EU's common agricultural policy "is deliberately opaque, grossly undermines the EU's environmental goals and is warped by corruption and self-dealing", fuelling "Mafia-style land grabs". <a href="https://nytimes.com/2019/11/03/world/europe/eu-farm-subsidy-hungary.html">https://nytimes.com/2019/11/03/world/europe/eu-farm-subsidy-hungary.html</a>	307	INVESTIGACIÓN DEL NY TIMES
10/11/2019	7. The EU's Common Agricultural Policy supports farmers, ensures food security and delivers reasonable prices to consumers. <a href="https://instituteforgovernment.org.uk/explainers/common-agricultural-policy">https://instituteforgovernment.org.uk/explainers/common-agricultural-policy</a>	161	SEGURIDAD ALIMENTARIA, PRECIOS JUSTOS
01/06/2018	Presenting a modernised & simplified Common Agricultural Policy for the future of food & farming in Europe: <a href="http://europa.eu/!wn47yD">http://europa.eu/!wn47yD</a> 🎯 Targeted, Flexible, Effective #FutureofCAP @PhilHoganEU	111	NUEVA PAC
29/11/2017	The Future of Food & Farming – a flexible, fair & sustainable Common Agricultural Policy: <a href="http://europa.eu/!Nh37NV">http://europa.eu/!Nh37NV</a> #FutureofCAP	86	NUEVA CAP
09/04/2021	Week 32 of my #ClimateStrike for #FridaysForFuture My sign says (in Finnish): Better Agricultural Policy! I am asking for the CAP (Common Agricultural Policy) to be good for the small farmers, the climate and biodiversity.	80	MEDIO AMBIENTE, CLIMA Y SOSTENIBILIDAD
08/04/2019	We are prepared to support European farmers in a possible "no-deal" Brexit scenario. Our Common Agricultural Policy has existing tools that can be activated in the event of market disturbances to #Prepare4Brexit. Learn more <a href="https://europa.eu/!Uu97Vj">https://europa.eu/!Uu97Vj</a>	73	BREXIT

**Tabla 7.** Las publicaciones que han conseguido más RT, clasificadas por sentimiento y temática.

\*Ver Subapartado Errores detectados.

Fuente: RStudio.

Lo más destacable es que los *tuits* negativos consiguen más RT que los positivos, hasta 10 veces más. Pese a que el número total de *tuits* de categoría positiva es mayor, estos son menos *retuiteados*. A esta misma conclusión llegaron Antypas, Preece & Collados (2022), quienes comprobaron que existe una fuerte relación entre el sentimiento y la popularidad de los *tuits* de tinte político. En su estudio, aquellos *tuits* clasificados negativamente mostraban una mayor penetración en redes que aquellos que transmitían un sentimiento positivo.



## Conclusiones del caso

La mayor parte de los temas con más RT coinciden con los temas señalados en los *clusters*. Teniendo en cuenta, por tanto, qué publicaciones han despertado mayor interés así como qué *clusters* han sido más mencionados, podrían extraerse varias conclusiones interesantes de cara a la constitución de la nueva PAC:

1) A lo largo de su historia, la PAC se ha ido adaptando a las necesidades socioeconómicas del momento y a los cambios dentro del escenario europeo y mundial. El Brexit es una de las grandes preocupaciones de los europeos y uno de los aspectos más delicados con los que la nueva PAC deberá tratar para garantizar que las rentas de los agricultores no se vean comprometidas y que los precios de los productos no suban (algo que ahora también se ve afectado por la guerra en Ucrania). El Brexit afectará a la definición del futuro presupuesto y a la solidaridad financiera de cada EM. Asimismo, obligará a tener contemplados planes de contención de esta crisis y de protección de las exportaciones europeas frente a los nuevos trámites aduaneros.

2) La PAC es tachada por muchos de ser poco ambiciosa en sus compromisos medioambientales y con el Pacto Verde Europeo (*Green Deal*). Por poner un ejemplo, a pesar de las aspiraciones que presentan los nuevos eco-esquemas, su aplicación por parte de los agricultores es voluntaria. Además, cada EM puede decidir qué pagos voluntarios adoptar y cuáles no y cuáles son los criterios de elegibilidad para obtenerlos (Petsakos et al, 2022). Movimientos como Fridays 4 Future, organizaciones como GreenPeace Europe o Birdlife e incluso partidos políticos instan a la Comisión a aprobar propuestas más “verdes”, al considerar que con la nueva PAC que se plantea, no se alcanzarán a tiempo los objetivos del *Green Deal* y de la Agenda 2030.

3) Uno de los puntos positivos de la PAC que los ciudadanos tienen más en cuenta es la seguridad alimentaria, tanto en garantizar la calidad y la seguridad del producto consumido como en asegurar unos precios asequibles para el consumidor. También se ha mencionado como punto positivo el apoyo y la protección de las rentas de los agricultores.

4) Hay preocupación en los entornos animalistas sobre el apoyo de la PAC al sector ganadero de lidia. Muchos defensores del bienestar animal cuestionan que estos fondos europeos se destinen a financiar la tauromaquia, frente a quienes defienden que se trata de una tradición española con derecho a ser conservada. En la muestra actualizada, se ha detectado un apoyo mayoritario a la abolición de estos fondos.

5) Se deben eliminar las sospechas de corrupción y de prácticas ilegales a la hora de distribuir los fondos PAC. Esto solo es posible a través de la mejora de los métodos de solicitud y asignación de fondos. En muchos países (como España), el cuaderno de campo de los agricultores, instrumento clave durante el proceso de solicitud de ayudas de la PAC, no está digitalizado. La digitalización y automatización de estos procesos es clave para reducir el número de infracciones.

6) Hay escasas menciones a los agricultores, sus condiciones laborales, los precios que perciben, etc. Esto probablemente indica que los agricultores no están presentes en RRSS y no cuentan con núcleos de apoyo dentro de estas plataformas, organizaciones visibles que generen debate en torno a los problemas que les conciernen. Mientras que los

consumidores están ampliamente representados en Twitter, la opinión de los productores es casi desconocida y este método no sería el más adecuado para cubrir este último punto.

Es posible, por tanto, contestar afirmativamente a la pregunta formulada al inicio de este trabajo: ¿puede resultar de utilidad analizar las opiniones en RRSS de los usuarios acerca de la PAC de cara a evaluar los resultados conseguidos y plantear nuevas medidas?. Este caso demuestra que las RRSS contribuyen a obtener una visión más completa de la realidad de la política agraria en Europa tal y como lo expresa la ciudadanía.

#### *Errores detectados*

En el subapartado ¿De qué hablan los tuits con mayor difusión? se han detectado algunos *tuits* de sentimiento positivo cuya categorización es errónea. La aparición de errores en dicha clasificación, habiendo revisado algunos de los *tuits* mal clasificados, se ha debido principalmente a las referencias (positivas o negativas) dentro del *tuit* a otros temas que no son la PAC. Concretamente, muchos *tuits* que apoyaban el Brexit hablaban de las ventajas para el Reino Unido de salir de la Política Agraria Común. Al introducir palabras de tinte positivo, el modelo ha confundido el resultado como un *tuit* en apoyo a la PAC cuando en realidad puede considerarse de categoría negativa al respaldar que Gran Bretaña deje de aplicar la PAC.

Esta limitación del modelo se discutirá más adelante en el Apartado 5. Limitaciones del análisis de opinión en RRSS.

## **5. LIMITACIONES DEL ANÁLISIS DE OPINIÓN EN RRSS**

Si bien todas las funcionalidades descritas en los Casos 1 y 2 suponen un instrumento de gran valor para académicos, *policymakers* e incluso empresas, es importante destacar que estas técnicas presentan una serie de limitaciones de carácter técnico, lingüístico, legal e incluso ético, que se ven acentuadas en ámbitos de aplicación como el agronómico, donde el empleo de estas herramientas está muy poco extendido y todavía existe mucho margen de mejora. En líneas generales, las barreras detectadas pueden dividirse en seis grupos:

A) Sesgos en la muestra: los usuarios de Twitter son más jóvenes y presentan un grado de educación superior a la media poblacional (Borrero & Zabalo, 2021). Por ello, cualquier muestra que se tome de esta plataforma podría no ser una muestra representativa del conjunto de la población.

B) Calidad de la muestra: la calidad de la muestra con la que se trabaja al procesar datos de RRSS se puede ver comprometida por múltiples factores, entre ellos, el gran volumen de datos, la presencia de *bots* y publicaciones *spam* o la “suciedad” de los datos.

C) Limitaciones lingüísticas: asociadas al idioma particular del texto y al uso que hace el usuario del lenguaje en un contexto lingüístico determinado (sarcasmo, ambigüedad, faltas de ortografía, etc.).

D) Precisión del método de aprendizaje no supervisado: se han identificado barreras inherentes a la metodología utilizada en este trabajo para la clasificación de sentimientos,

relacionadas con el tamaño del *lexicon*, el límite de caracteres aceptado por Twitter o el uso de la negación en un tuit.

E) Temas legales y de privacidad de los usuarios: conseguir extraer grandes volúmenes de *tuits* sigue estando restringido únicamente a usuarios de perfil académico.

F) Limitaciones éticas: en línea con el punto anterior, cabe por último recordar que en trabajos de SMA se está tratando con datos personales de usuarios de RRSS, de los que muchas veces también se conoce su nombre de usuario, sexo y hasta su geolocalización.

El detalle de cada grupo puede consultarse en el **Anexo IX: Limitaciones del análisis de opinión en RRSS.**

## 6. CONCLUSIONES

A lo largo del presente trabajo, se ha podido comprobar la utilidad de los métodos de *opinion mining* en Twitter para recabar una gran cantidad de información en poco tiempo sobre la opinión de los usuarios en RRSS frente a temáticas del ámbito agronómico, véase la Política Agraria Común y el modelo de agricultura ecológica. **Se ha demostrado cómo, aplicadas de forma combinada con la anotación manual, las técnicas NLP pueden ayudar en la valoración de antiguas políticas e incluso en el impulso y configuración de nuevas iniciativas políticas.** Se ha visto también que estas técnicas pueden ayudar a identificar el motivo o los motivos por los que la ciudadanía se posiciona a favor o en contra de un tema, sea este una política, un modelo de producción o una práctica agronómica. El *sentiment analysis* contribuye incluso a detectar qué sucesos desencadenan un inusual número de reacciones negativas o positivas que alteran la percepción que el usuario tiene acerca de un tema en un momento determinado.

Como muestran los resultados, la recopilación, procesamiento y análisis de datos generados en RRSS no es una tarea fácil. En cada uno de los pasos del proceso de *opinion mining* se han encontrado desafíos de distinta naturaleza que son necesarios abordar. Los investigadores señalan la sobreabundancia de datos que están disponibles en redes sociales y la naturaleza dinámica de las publicaciones en redes hace que recopilarlos sea un proceso largo y caótico. Prueba de ello está en la muestra final del Caso 2, donde se ha detectado una fracción de *tuits* que no tratan sobre la PAC o que la mencionan sólo de pasada y expresan su opinión sobre otro asunto, pese a haber hecho una selección cuidadosa de las palabras clave. Aprender a filtrar estos datos y a descartar aquellos que son irrelevantes es necesario para poder aprovechar la oportunidad de trabajar con infinidad de datos útiles que cientos de miles de usuarios de todo el planeta han compartido libremente y de forma gratuita. Los algoritmos avanzados de detección de temáticas, más conocidos como TDT (*Topic Detection and Tracking*), como es el caso del modelo DL TopicBERT (Asgari-Chenaghlu et al., 2021), irán aumentando su precisión gracias a su entrenamiento hasta aproximarse a la alcanzada mediante el proceso de anotación manual (empleado en el Caso 1).

En etapas posteriores del análisis de opinión, el método de aprendizaje a aplicar es otro punto a tener en cuenta. Aspectos del lenguaje natural como el uso del sarcasmo, la ironía o la ambigüedad, tan frecuentes de encontrar en contextos comunicativos espontáneos como lo son las RRSS, afectan a la capacidad interpretativa de los LMs que no

tienen una etapa de entrenamiento previo enfocada en la detección de este tipo de figuras lingüísticas. La riqueza del *lexicon* condicionará en buena medida el acierto en los resultados. Configurar un *lexicon* adaptado al contexto o dominio de la temática en cuestión es una buena manera de reducir errores.

Otro obstáculo son los problemas en la detección de los *valence shifters*, los cuales no pueden solventarse con los métodos de aprendizaje no supervisado. Se trata de una de las limitaciones de esta técnica que puede dar lugar a importantes distorsiones en los resultados. Los modelos ML pueden solucionar este problema siendo entrenados para detectar estructuras de negación en el texto. Sin embargo, el entrenamiento de estos LM precisa de un volumen de datos sensiblemente mayor que el método *lexicon-based*. Esto supone emplear mucho tiempo en el entrenamiento y la posterior validación de los modelos, además de exigir un esfuerzo computacional mayor. **Antes de elegir un método de aprendizaje automático u otro, se debe valorar el nivel de acierto que es necesario alcanzar en cada caso y tener en cuenta los esfuerzos que hacen falta para llevar a cabo dicho análisis.**

El análisis de redes sociales sigue siendo un área de investigación relativamente nueva. Dado el carácter interdisciplinario del SMA, los datos de estas plataformas pueden resultar útiles para investigadores que trabajan en campos muy diferentes (Stieglitz et al., 2018). Este trabajo resalta la necesidad de continuar investigando en la aplicación de estas disciplinas en sectores distintos al ámbito de las TIC, donde los investigadores no están familiarizados con los métodos informáticos que descubren, recopilan y preparan grandes volúmenes de datos de RRSS de forma automatizada. Al haber demostrado la utilidad del método de aprendizaje no supervisado para la clasificación del sentimiento en las temáticas de agricultura ecológica y la PAC, se abren ahora oportunidades dentro del sector público agrario para introducir la ciencia y la analítica de datos en RRSS dentro de los métodos de valoración e impacto de las medidas que implementan.

El empleo de R como lenguaje de programación ha permitido llevar a cabo análisis estadísticos en profundidad para una interpretación más exacta de los resultados. R es conocido por ser una de las mejores herramientas en estadística computacional y por ello fue el programa elegido para elaborar los dos casos de estudio. Poseer un conocimiento medio del lenguaje es suficiente para que un usuario, de cualquier tipo de perfil investigador, pueda hacer uso de esta herramienta de forma gratuita para extraer conclusiones y visualizar los resultados de manera gráfica gracias a los miles de paquetes de trabajo disponibles para la creación de gráficos de calidad.

**Como futuras líneas de investigación**, sería interesante aplicar otros modelos de lenguaje de aprendizaje automático, basados en ML o redes neuronales, y comparar el rendimiento de cada uno de ellos respecto a los métodos *lexicon-based*. Analizar periodos temporales más extensos o comparar las opiniones según el género de los usuarios (actualmente esta información solo está disponible en la biografía de los usuarios que deciden hacerlo público, por lo que no es fácil recopilar esta información) o su país de residencia (el número de *tuits* geolocalizados es escaso y por tanto la muestra es muy pequeña para poder estudiarla) pueden ser otras acciones interesantes de cara a obtener una imagen más exacta del perfil de usuario que se muestra a favor o en contra de estos temas. Además, la extracción automática de datos podría ampliarse a nuevos temas, palabras clave y *hashtags*, otros idiomas e incluso a otras plataformas de RRSS.

## BIBLIOGRAFÍA

- Agostino, D. (2013). *Using social media to engage citizens: A study of Italian municipalities*. Public relations review, 39(3), 232-234.
- Alsaeedi, A., & Khan, M. Z. (2019). *A study on sentiment analysis techniques of Twitter data*. International Journal of Advanced Computer Science and Applications, 10(2), 361-374.
- Antypas, D., Preece, A., & Collados, J. C. (2022). *Politics and Virality in the Time of Twitter: A Large-Scale Cross-Party Sentiment Analysis in Greece, Spain and United Kingdom*. arXiv preprint arXiv:2202.00396.
- Araujo Britto Sass, C., Colombo Pimentel, T., Gonzalez Barandela Aleixo, M., Mendes Dantas, T., Cyrino Oliveira, F.L., Queiroz de Freitas, M., Gomes da Cruz, A. and Almeida Esmerino, E. (2020). *Exploring social media data to understand consumers' perception of eggs: a multilingual study using Twitter*. Journal of Sensory Studies, Vol. 35 No. 6
- Art. 97.1 del Reglamento (UE) 2021/2115 del Parlamento Europeo y del Consejo, de 2 de diciembre de 2021, por el que se establecen normas en relación con la ayuda a los planes estratégicos que deben elaborar los Estados Miembros en el marco de la política agrícola común (...), y por el que se derogan los Reglamentos (UE) n.o 1305/2013 y (UE) n.o 1307/2013.
- Asgari-Chenaghlu, M., Feizi-Derakhshi, M. R., Farzinvash, L., Balafar, M. A., & Motamed, C. (2021). *Topic detection and tracking techniques on Twitter: a systematic review*. Complexity, 2021.
- Azcárate, T. G., & Folkesson, C. (2020). *The new delivery model of the CAP: Some relevant issues*. Economía Agraria y Recursos Naturales-Agricultural and Resource Economics, 20(1), 149-167.
- Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. (2020). *Tweeteval: Unified benchmark and comparative evaluation for tweet classification*. arXiv preprint arXiv:2010.12421.
- Barkur, G. & Vibha, G. B. K. (2020) *Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India*. Asian Journal of Psychiatry.
- Batrinca, B., & Treleaven, P. C. (2015). *Social media analytics: a survey of techniques, tools and platforms*. Ai & Society, 30(1), 89-116.
- Bermeo-Almeida, O., del Cioppo-Morstadt, J., Cardenas-Rodriguez, M., Cabezas-Cabezas, R., & Bazán-Vera, W. (2019, January). *Sentiment analysis in social networks for agricultural pests*. In 2nd International Conference on ICTs in Agronomy and Environment (pp. 122-129). Springer, Cham.
- Bhaskaran, J. & Bhallamudi, I. (2019) *Good Secretaries, Bad Truck Drivers? Occupational Gender Stereotypes in Sentiment Analysis*. In Proceedings of the First Workshop on Gender Bias in Natural Language Processing.
- Borrero, J. D., & Zabalo, A. (2021). *Identification and analysis of strawberries' consumer opinions on twitter for marketing purposes*. Agronomy, 11(4), 809.
- Bureau, J. C., & Swinnen, J. (2018). *EU policies and global food security*. Global food security, 16, 106-115.
- Cifci, I., Demirkol, S., Altunel, G.K. and Cifci, H. (2020). *Overcoming the food neophobia towards science-based cooked food: the supplier perspective*. International Journal of Gastronomy and Food Science, Vol. 22, p. 100280.
- Consejo Europeo (2 de diciembre de 2021) *El Consejo adopta para el período 2023-2027 una política agrícola más justa, más ecológica y más basada en los resultados*. Comunicado de prensa. Disponible

en: <https://www.consilium.europa.eu/es/press/press-releases/2021/12/02/council-adopts-fairer-greener-and-more-performance-based-farming-policy-for-2023-2027> [Visto el 27/05/2022].

Criado, J. I., & Rojas Martín, F. (2013). *Las redes sociales digitales en la gestión y las políticas públicas. Avances y desafíos para un gobierno abierto*. Barcelona, Escola d'Administració Pública de Catalunya.

Dupraz, P., & Guyomard, H. (2019). *Environment and climate in the Common Agricultural Policy*. EuroChoices, 18(1), 18-25.

EAE Business School (17 de Septiembre de 2019) *España es el quinto país que más contribuye al presupuesto de la UE en 2019*. [Comunicado de prensa] Disponible en: <https://www.eae.es/actualidad/notas-de-prensa/espana-es-el-quinto-pais-que-mas-contribuye-al-presupuesto-de-la-ue-en-2019#:~:text=Por%20delante%20de%20ella%2C%20se.con%2017.060%20millones%20de%20euros> [Visto el 27/06/2022]

European Commission (s.f.) Principales objetivos estratégicos de la nueva PAC. Disponible en: [https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/new-cap-2023-27/key-policy-objectives-new-cap\\_es#briefs](https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/new-cap-2023-27/key-policy-objectives-new-cap_es#briefs) Visto[20/04/2022].

Eurostat (2020) *Individuals and Internet activities*. Disponible en: [https://ec.europa.eu/eurostat/databrowser/view/ISOC\\_CI\\_AC\\_I\\_custom\\_1058781/bookmark/table?lang=en&bookmarkId=0dd5535e-a7e5-4ff8-a2a4-134d5b56718d&page=time:2020](https://ec.europa.eu/eurostat/databrowser/view/ISOC_CI_AC_I_custom_1058781/bookmark/table?lang=en&bookmarkId=0dd5535e-a7e5-4ff8-a2a4-134d5b56718d&page=time:2020) [Visto el 06/05/2022]

Esposti, R. (2022). *The Coevolution of Policy Support and Farmers' Behaviour. An investigation on Italian agriculture over the 2008-2019 period* (No. 464).

Evans, O. (2 de abril de 2019) *MEPs vote in favour of Hogan's more flexible CAP*. Irish Farmers Journal. Disponible en: <https://www.farmersjournal.ie/meps-vote-in-favour-of-hogan-s-more-flexible-cap-453041> [Visto el 29/05/2022].

Feldman, R. (2013) Techniques and Applications for Sentiment Analysis. Communications of the ACM. Vol. 56, No 4., 82-89.

Gabriel, D., Sait, S. M., Hodgson, J. A., Schmutz, U., Kunin, W. E., & Benton, T. G. (2010). *Scale matters: the impact of organic farming on biodiversity at different spatial scales*. Ecology letters, 13(7), 858-869.

Gattinger, A., Muller, A., Haeni, M., Skinner, C., Fliessbach, A., Buchmann, N., ... & Niggli, U. (2012). *Enhanced top soil carbon stocks under organic farming*. Proceedings of the National Academy of Sciences, 109(44), 18226-18231.

Gebrekidan, S., Apuzzo, M. & Novak, B. (3 de noviembre de 2019) *The Money Farmers: How Oligarchs and Populists Milk the E.U. for Millions*. New York Times. Disponible en: <https://www.nytimes.com/2019/11/03/world/europe/eu-farm-subsidy-hungary.html> [Visto el 24/06/2022].

Goldberg, D.M., Khan, S., Zaman, N., Gruss, R.J. and Abrahams, A.S. (2020) *Text mining approaches for postmarket food safety surveillance using online media*. Risk Analysis, in press, doi: 10.1111/risa.13651.

Henke, R., Benos, T., De Filippis, F., Giua, M., Pierangeli, F., & Pupo D'Andrea, M. R. (2018). *The new common agricultural policy: How do member states respond to flexibility?*. JCMS: Journal of Common Market Studies, 56(2), 403-419.

Horák, I. (2022). *The Competitiveness Outlook of the European Agriculture with the new Green Deal Policy* (No. 2022-78). Mendel University in Brno, Faculty of Business and Economics.

Huber, M., Rembiałkowska, E., Średnicka, D., Bügel, S., & Van De Vijver, L. P. L. (2011). *Organic food and impact on human health: Assessing the status quo and prospects of research*. NJAS: Wageningen Journal of Life Sciences, 58(3-4), 103-109.

Hu, M., & Liu, B. (2004). *Mining and summarizing customer reviews*. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004). Seattle, Washington

Intergovernmental Panel on Climate Change, IPCC (Febrero de 2022) *Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Pörtner, H. O., Roberts, DC, Tignor, M., Poloczanska, ES, Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Lösschke, S., Möller, V., et al., Eds, 2022.

Jayasinghe, U. & Ghoshal, D. (3 de marzo de 2022) *Fertiliser ban decimates Sri Lankan crops as government popularity ebbs*. Reuters. Disponible en: <https://www.reuters.com/markets/commodities/fertiliser-ban-decimates-sri-lankan-crops-government-popularity-ebbs-2022-03-03/> [Visto el 29/05/2022].

Jockers, M. L. (2017). *Syuzhet: Extract sentiment and plot arcs from Text*. Retrieved from <https://github.com/mjockers/syuzhet>

Kaplan, A. M., & Haenlein, M. (2010). *Users of the world, unite! The challenges and opportunities of Social Media*. Business horizons, 53(1), 59-68.

Kemp, S. E. (2013). *Consumers as part of food and beverage industry innovation*. In Open innovation in the food and beverage industry (pp. 109-138). Woodhead Publishing.

La Vanguardia (22 de octubre de 2020) *Ecologistas celebran que UE no financie la cría de toros de lidia en la PAC*. La Vanguardia. Disponible en: <https://www.lavanguardia.com/vida/20201022/484227132701/ecologistas-celebran-que-ue-no-financie-la-cria-de-toros-de-lidia-en-la-pac.html> [Visto el 29/06/2022]

Lassance, A. (2020). *What Is a Policy and What Is a Government Program? A Simple Question With No Clear Answer, Until Now*. Social-Science Research Network (SSRN). Disponible en: <https://ssrn.com/abstract=3727996> Visto [20/04/2022]

Marković, K., Njegovan, Z., & Pejanović, R. (2012). *Former and future reforms of Common Agricultural Policy of the European Union*. Економика пољопривреде, 59(3), 483-498.

Martínez Arroyo, F. (22 de enero de 2021) *Una PAC para la reconstrucción del país tras la COVID*. El Diario. Disponible en: [https://www.eldiario.es/castilla-la-mancha/palabras-clave/pac-reconstruccion-pais-covid\\_132\\_7039332.html](https://www.eldiario.es/castilla-la-mancha/palabras-clave/pac-reconstruccion-pais-covid_132_7039332.html) [Visto el 27/05/2022]

Maté, V. (17 de marzo de 2022) *La guerra de Ucrania pone contra las cuerdas la política verde de la PAC*. El País. Disponible en: <https://elpais.com/economia/2022-03-17/la-guerra-de-ucrania-pone-contra-las-cuerdas-la-politica-verde-de-la-pac.html> [Visto el 27/05/2022].

Matthews, A. (2021). *The contribution of research to agricultural policy in Europe*. Bio-based and Applied Economics, 10(3), 185-205.

Matthews, A. (2020). *The new CAP must be linked more closely to the UN Sustainable Development Goals*. Agricultural and Food Economics, 8(1), 1-4.

Melcher, K. (2018). *Sentiment Analysis: Deep Learning, Machine Learning, Lexicon Based?* [Webinar]. Knime. Disponible en: <https://www.youtube.com/watch?v=By4IZelzxIw> [Visto el 28/04/2022]

Ministerio de Agricultura, Pesca y Alimentación (Diciembre de 2021) *El Plan Estratégico de la PAC de España (2023-2027). Resumen de la propuesta*. Subdirección General de Planificación de Políticas Agrarias. Centro de Publicaciones del MAPA.

Mitra, A. (2020) *Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)*. Journal of Ubiquitous Computing and Communication Technologies (UCCT) Vol.02/ No.03, 145-152. Disponible en: <https://www.ijournals.com/jucct/> [Visto el 29/04/2022].

Moraleda, F. (2022) *Sector agroalimentario Next Generation: incertidumbres y certezas*. Agricultura y ganadería familiar en España: Anuario 2022, pp. 90-96. Fundación de Estudios Rurales.

Moreno-Sandoval, L. G., Sánchez-Barriga, C., Buitrago, K. E., Pomares-Quimbaya, A., & Garcia, J. C. (2018, August). *Spanish Twitter data used as a source of information about consumer food choice*. In International Cross-Domain Conference for Machine Learning and Knowledge Extraction (pp. 134-146). Springer, Cham.

Navarro-Hellín, H., Martínez-del-Rincon, J., Domingo-Miguel, R., Soto-Vallés, F., & Torres-Sánchez, R. (2016). *A decision support system for managing irrigation in agriculture*. Computers and Electronics in Agriculture, 124, 121-131.

Olthof, A. W., van Ooijen, P. M., & Cornelissen, L. J. (2021). *Deep learning-based natural language processing in radiology: The impact of report complexity, disease prevalence, dataset size, and algorithm type on model performance*. Journal of medical systems, 45(10), 1-16.

Park, I. H., Park, Y. M., Lee, C., Sun, J. E., Hu, W., & Chung, J. E. (2022). *A Comparative Study of Dietary Related Zero-waste Patterns and Consumer Responses Before and After COVID-19*. Family and Environment Research, 60(1), 21-38.

Patel, M. A. K., & Patel, B. R. (July 2020) *Sentiment Analysis for Social Media: An Overview and Open Challenges*. Int.J.Computer Technology & Applications, Vol 11 (3), 35-39.

Petsakos, A., Ciaian, P., Espinosa, M., Perni, A., & Kremmydas, D. (2022). *Farm-level impacts of the CAP post-2020 reform: A scenario-based analysis*. Applied Economic Perspectives and Policy.

Prusa, J., Khoshgoftaar, T. M., & Seliya, N. (2015, December). *The effect of dataset size on training tweet sentiment classifiers*. In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA) (pp. 96-102). IEEE.

Ramos-Sandoval, R., & Beltran, C. S. (2021, November). *Using Twitter data for consumer research: The Peruvian's consumer perspective on short food supply chains*. In 2021 IEEE Sciences and Humanities International Research Conference (SHIRCON) (pp. 1-4). IEEE.

Recuero-Virto, N., & Valilla-Arróspide, C. (2022). *Forecasting the next revolution: food technology's impact on consumers' acceptance and satisfaction*. British Food Journal

Rekha, Naik, S. N., & Prasad, R. (2006). *Pesticide residue in organic and conventional food-risk analysis*. Journal of Chemical Health & Safety, 13(6), 12-19.

Renting, H., Marsden, T. K., & Banks, J. (2003). *Understanding alternative food networks: exploring the role of short food supply chains in rural development*. Environment and planning A, 35(3), 393-411.

Rinker, T. (October 2021) *Package 'sentimentr'*. CRAN repository.

Roberts, H. (8 de noviembre de 2018) *How the mafia got to our food*. Financial Times. Disponible en: <https://www.ft.com/content/73de228c-e098-11e8-8e70-5e22a430c1ad> [Visto el 01/07/2022]

Rodríguez-Ibáñez, M. et al (2021). *Sentiment Analysis of Political Tweets from the 2019 Spanish Elections*. IEEE Access, 9, 101847-101862.



- Rohde Johannessen, M., & Følstad, A. (2014). *Political social media sites as public sphere: A case study of the norwegian labour party*. Communications of the Association for Information Systems, 34(1), 56.
- Rojas-Martín, F., & Criado, J. I. (2015). *Avanzando en la medición de las redes sociales digitales: una propuesta para su análisis desde una perspectiva institucionalista*. Revista de Gestión Pública, 4(1), 43-64.
- SAS Institute (2021) *What is Natural Language Processing (NLP)*. Disponible en: [https://www.sas.com/es\\_ar/insights/analytics/what-is-natural-language-processing-nlp.html#nlphistory](https://www.sas.com/es_ar/insights/analytics/what-is-natural-language-processing-nlp.html#nlphistory) [Visto el 29/04/2022].
- Salman, M., García Vila, M., Fereres Castiel, E., Raes, D., & Steduto, P. (2021). *The AquaCrop model—Enhancing crop water productivity*. CSIC.
- Salmons, J. (2017). *Using social media in data collection: designing studies with the qualitative e-research framework*. The SAGE handbook of social media research methods, 177-197.
- Siegrist, M., & Hartmann, C. (2020). *Consumer acceptance of novel food technologies*. Nature Food, 1(6), 343-350.
- Singh, A., Shukla, N. and Mishra, N. (2018) *Social media data analytics to improve supply chain management in food industries*. Transportation Research, Vol. E 114, pp. 398-415, doi: 10.1016/j.tre.2017.05.008.
- Sottile, F. (23 de octubre de 2020) *Common Agricultural Policy: a decisive vote in Parliament*. Slow Food. Disponible en: <https://www.slowfood.com/common-agricultural-policy-a-decisive-vote-in-parliament/> [Visto el 24/06/2022].
- Srivastava, R. et al (2022) *Comparative Analysis of Lexicon and Machine Learning Approach for Sentiment Analysis*. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 13, No. 3.
- Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). *Social media analytics—Challenges in topic discovery, data collection, and data preparation*. International journal of information management, 39, 156-168.
- Stieglitz, S., & Dang-Xuan, L. (2013). *Social media and political communication: a social media analytics framework*. Social network analysis and mining, 3(4), 1277-1291.
- Susmitha, M., & Pranitha, R. L. (2022). *Performance Assessment Using Supervised Machine Learning Algorithms of Opinion Mining on Social Media Dataset*. In Proceedings of Second International Conference on Advances in Computer Engineering and Communication Systems (pp. 419-427). Springer, Singapore.
- Tao, D., Yang, P. and Feng, H. (2020). *Utilization of text mining as a big data analysis tool for food science and nutrition*. Comprehensive Reviews in Food Science and Food Safety, Vol. 19, pp. 875-894. 10.1111/1541-4337.12540.
- Treasury Board of Canada, Secretariat (March 1998) *Program Evaluation Methods: Measurement and Attribution of Program Results*. © Minister of Public Works and Government Services. Public Affairs Branch 3rd Edition.
- Twitter (Octubre 2021) *Q3 2021: Letter to Shareholders*. Disponible en: [https://s22.q4cdn.com/826641620/files/doc\\_financials/2021/q3/Final-Q3'21-Shareholder-letter.pdf](https://s22.q4cdn.com/826641620/files/doc_financials/2021/q3/Final-Q3'21-Shareholder-letter.pdf) [Visto el 06/05/2022]

- Turgut, C., Ornek, H., & Cutright, T. J. (2011). *Determination of pesticide residues in Turkey's table grapes: the effect of integrated pest management, organic farming, and conventional farming*. *Environmental monitoring and assessment*, 173(1), 315-323.
- Vaccari, C., Valeriani, A., Barber, P., Bonneau, R., Jost, J. T., Nagler, J., & Tucker, J. (2013). *Social media and political communication. A survey of Twitter users during the 2013 Italian general election*. *Rivista italiana di scienza politica*, 381-410. Disponible en: <https://doi.org/10.1426/75245> Visto [21/04/2022].
- Valkanas, G., Katakis, I., Gunopulos, D., & Stefanidis, A. (2014). *Mining Twitter Data with Resource Constraints*. In 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) pp. 157-164. <https://doi.org/10.1109/WI-IAT.2014.29>.
- Valsamidis, S., Theodosiou, T., Kazanidis, I., & Nikolaidis, M. (2013). *A framework for opinion mining in blogs for agriculture*. *Procedia Technology*, 8, 264-274.
- Van Atteveldt, W., van der Velden, M. A., & Boukes, M. (2021). *The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms*. *Communication Methods and Measures*, 15(2), 121-140.
- Vázquez-Salat, N., & Houdebine, L. M. (2013). *Society and GMOs. The lessons from the GM crops debate in Europe over risk management and communication provide valuable pointers for the upcoming debate on GM animals*. *EMBO reports*, 14(8), 671-674.
- Vidal, L., Ares, G., Machin, L. and Jaeger, S.R. (2015) *Using Twitter data for food-related consumer research: a case study on "what people say when tweeting about different eating situations"*. *Food Quality and Preference*, Vol. 45, pp. 58-69.
- Villanueva, A. J. (2022) *Los ecosquemas de la nueva PAC*. *Agricultura y ganadería familiar en España: Anuario 2022*, pp. 90-96. Fundación de Estudios Rurales.
- Weiguo, F. & Gordon, M. D. (2014) *The power of social media analytics*. *Communications of the ACM* Volume 57, Number 6, Pages 74-81.
- Yang, C., Everitt, J. H., Du, Q., Luo, B., & Chanussot, J. (2012). *Using high-resolution airborne and satellite imagery to assess crop growth and yield variability for precision agriculture*. *Proceedings of the IEEE*, 101(3), 582-592.
- Young, L., & Soroka, S. (2012). *Affective news: The automated coding of sentiment in political texts*. *Political Communication*, 29(2), 205-231.