



Facultad de Ciencias Económicas y Empresariales

Análisis de Percepciones y Sentimientos sobre la transición energética en Redes Sociales

Autor: Andrea Soriano Fernández

Director: Jenny Alexandra Cifuentes Quintero

MADRID | Junio 2024

Resumen

Hoy en día la transición energética se ha convertido en una prioridad mundial para hacer frente al cambio climático y a todos los desafíos medioambientales a los que nuestro planeta se enfrenta. Esta transición energética se centra en reemplazar todas aquellas energías que provienen de fuentes fósiles como el carbón o la energía nuclear, y que son las responsables de la emisión de gases de efecto invernadero, por energías más limpias y renovables que contribuyan al objetivo de lograr cero emisiones netas al planeta, satisfaciendo las necesidades de las generaciones presentes sin comprometer las de las generaciones futuras. En este contexto, las redes sociales se han convertido en un medio fundamental para expresar las opiniones, preocupaciones o emociones de los usuarios especialmente de la red social X. En concreto, las publicaciones que se comparten en esta red social, son un medio vital para el debate y la discusión sobre la transición energética, y en general sobre cualquier tema relacionado con la sostenibilidad. Estas opiniones públicas ayudan a los políticos a entender el pensamiento común de sus ciudadanos y a tomar las medidas necesarias para mejorar la política energética de los países que están a su cargo.

Por ello, en este Trabajo de Fin de Grado se identifican las principales categorías de discusión y los sentimientos subyacentes asociados a la controversial temática de la transición energética en la red social X. Para ello, se emplea la técnica *Latent Dirichlet Allocation* (LDA) para el modelado de tópicos y el diccionario *Valence Aware Dictionary and sEntiment Reasoner* (VADER) para el análisis de sentimientos.

En el modelado de tópicos se identifican 5 tópicos diferentes dentro del corpus de publicaciones, cuyas temáticas son: “Políticas y desarrollo sostenible”, “Debates y desafíos en la transición energética”, “Acción climática y vista a futuro”, “Papel europeo y estrategias” y “Apoyo global a la transición energética”. Cada uno de estos tópicos aborda aspectos relevantes de la transición energética, desde los desafíos potenciales hasta la importancia de las energías renovables. Además, se destaca el papel fundamental de las políticas y la cooperación global para asegurar el éxito y la sostenibilidad de esta transición. En análisis de sentimientos realizado en los tópicos mencionados, revela un sentimiento mayoritariamente positivo o neutro, con una irrelevante puntuación negativa asociada, reflejando la importancia de esta temática, y la motivación de los usuarios hacia un mundo más limpio. Aún así, la temática del apoyo global a la transición energética presenta la mayor concentración de

sentimientos neutros y negativos, mostrando la preocupación de los usuarios al respecto, así como la importancia de que los países actúen en común, estableciendo acciones rápidas y eficaces que promuevan la transición energética.

Palabras clave: Transición energética, Energías renovables, Análisis de sentimientos, Redes sociales, Modelado de Tópicos

Abstract

Nowadays, energy transition has become a global priority to address climate change and all the environmental challenges our planet is facing. This energy transition focuses on replacing all those energies that come from fossil sources such as coal or nuclear energy, which are responsible for the emission of greenhouse gases, with cleaner and renewable energies that contribute to the goal of achieving zero net emissions to the planet, meeting the needs of present generations without compromising those of future generations. In this context, social networks have become a fundamental tool for expressing the opinions, concerns or emotions of users, especially on the X social network. Specifically, the publications shared on this social network are a key channel for debate and discussion on energy transition, and in general on any topic related to sustainability. These public opinions help politicians to understand the common thinking of their citizens and to take the necessary measures to improve the energy policy of the countries they are in charge of.

Therefore, in this Final Degree Project the main categories of discussion and the underlying sentiments associated with the controversial topic of energy transition in the social network X, are identified. For this purpose, *Latent Dirichlet Allocation* (LDA) for topic modeling and *Valence Aware Dictionary and sEntiment Reasoner* (VADER) for sentiment analysis are used.

The topic modeling identifies 5 different topics within the corpus of publications, whose themes are: “Policies and sustainable development”, “Debates and challenges in energy transition”, “Climate action and vision for the future”, “European role and strategies” and “Global support for energy transition”. Each of these topics addresses relevant aspects of energy transition, from potential challenges to the importance of renewable energies. In addition, the key role of policies and global cooperation in ensuring the success and sustainability of this transition is highlighted. The sentiment analysis conducted on the mentioned topics reveals a mostly positive or neutral sentiment, with an associated irrelevant negative score, reflecting the importance of this topic, and the motivation of users towards a cleaner world. Even so, the topic of global support for energy transition presents the highest concentration of neutral and negative sentiment, showing the concern of users in this regard, as well as the importance of countries acting together, establishing quick and effective actions to promote energy transition.

Keywords: Energy Transition, Renewable Energies, Sentiment Analysis, Social Networks, Topic Modeling

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	4
1.3. Estructura del documento	5
2. Analizando percepciones en el campo de la energía en X a través de técnicas de Machine Learning: Revisión de la literatura	6
3. Metodología de Análisis de Datos	14
3.1. Adquisición de los Datos	15
3.2. Pre-procesamiento de los Datos	16
3.3. Exploración de N-Gramas	17
3.4. Modelado de Tópicos	18
3.5. Análisis de Sentimientos	21
4. Resultados	23
4.1. Adquisición y preparación de los datos	23
4.2. Pre-procesamiento de los datos	26
4.3. Análisis de N-gramas	27
4.4. Modelado de tópicos	30
4.5. Análisis de sentimientos	35
5. Conclusiones	40
Bibliografía	45

Índice de figuras

1.1. Evolución de las emisiones globales de CO2 procedentes de combustibles fósiles. Fuente de datos: epdata. (Orús, 2023). Elaboración propia	2
1.2. Evolución de las emisiones globales de CO2 por origen. Fuente de datos: epdata. (Orús, 2023). Elaboración propia	2
1.3. Principales actividades de los usuarios de X en el tercer trimestre de 2022. Fuente de datos: Statista. (Fernández, 2023). Elaboración propia	4
3.1. Metodología del Análisis de Datos. Elaboración propia	15
3.2. Representación gráfica del modelo LDA. Fuente de datos: Journal of Machine Learning Research. (Blei, Ng, y Jordan, 2003). Elaboración propia	20
4.1. Número de publicaciones por Idioma. Elaboración propia	24
4.2. Principales Hashtags utilizados. Elaboración propia	24
4.3. Distribución de retweets por publicación. Elaboración propia	25
4.4. Distribución de <i>likes</i> por publicación. Elaboración propia	25
4.5. Wordcloud del dataset pre-procesado. Elaboración propia	27
4.6. Principales unigramas. Elaboración propia	28
4.7. Principales bigramas. Elaboración propia	29
4.8. Principales trigramas. Elaboración propia	30
4.9. Índice de coherencia para cada número de tópicos. Elaboración propia	31
4.10. Distancia intertópica para $k = 5$. Elaboración propia	32
4.11. Distribución de los tópicos seleccionados. Elaboración propia	35
4.12. Distribución de la puntuación compuesta. Elaboración propia	36
4.13. Número de publicaciones por puntuación. Elaboración propia	37
4.14. Distribución del sentimiento por tópico. Elaboración propia	39

Índice de tablas

2.1. Resumen de los estudios sobre percepción pública llevados a cabo mediante el análisis de publicaciones en X.	12
3.1. Resumen de las variables más relevantes para el análisis.	16
4.1. Tópicos, palabras clave, bigramas y trigramas más relevantes de los tópicos seleccionados	33
4.2. Estadísticas descriptivas por tópico	39

Acrónimos

<i>IPCC</i>	Panel Intergubernamental sobre Cambio Climático
<i>PLN</i>	Procesamiento de Lenguaje Natural
<i>RoBERTa</i>	Robustly optimized Bidirectional Encoder Representations from Transformers
<i>NRC</i>	National Research Center Canada
<i>VADER</i>	Valence Aware Dictionary and sEntiment Reasoner
<i>BERT</i>	Bidirectional Encoder Representations from Transformers
<i>LDA</i>	Latent Dirichlet Allocation
<i>TF-IDF</i>	Term Frequency - Inverse Document Frequency

Capítulo 1

Introducción

1.1. Motivación

En el panorama actual de desafíos ambientales y climáticos, la transición energética emerge como un proceso crucial orientado hacia la adopción de fuentes de energía más limpias y sostenibles. Este cambio, que se centra fundamentalmente en la sustitución de energías procedentes de combustibles fósiles, principales causantes de la contaminación por gases de efecto invernadero, por alternativas renovables, no es solo una cuestión técnica sino también un imperativo social y económico. La adopción de estas nuevas tecnologías energéticas representa un paso significativo hacia un futuro más sostenible y equitativo, ofreciendo un horizonte de esperanza y progreso para las generaciones venideras. Al mismo tiempo, esta transición energética conlleva desafíos complejos y multidimensionales, involucrando aspectos de justicia social, desarrollo económico, y la reconfiguración de infraestructuras y políticas globales (Energy, 2023). Su éxito no solo depende de avances tecnológicos, sino también de un cambio en la percepción y las actitudes del público, donde el diálogo y el consenso social se vuelven fundamentales para una implementación efectiva y equitativa de estas nuevas tecnologías energéticas.

En este contexto, es importante destacar el grave problema al que nuestra sociedad se enfrenta debido a las emisiones de dióxido de carbono. Las estadísticas son preocupantes: las emisiones procedentes de combustibles fósiles han seguido una tendencia ascendente año tras año, con más del 50% de estas emisiones acumuladas en los últimos 30 años, contribuyendo significativamente al incremento de los gases de efecto invernadero a nivel global (Mundo, 2021). El año 2022 marcó un récord preocupante, alcanzando una emisión de 40.000 millones de toneladas de CO₂ (Ver Figura 1.1), con la mayor parte proveniente del uso de carbón y petróleo como puede verse en la Figura 1.2. Estos combustibles no solo contribuyen al calentamiento global y, por ende, al cambio climático, sino que también tienen profundas consecuencias sociales y de salud. Un informe del Instituto Internacional de Derecho y Medio Ambiente reveló que solo en España, las emisiones de las centrales térmicas de carbón causa-

ron 1.529 muertes prematuras entre 2015 y 2016 (Orús, 2023). Este panorama ha impulsado a los países a unirse en un esfuerzo colectivo para combatir estos desafíos. Un ejemplo en esta lucha es el Acuerdo de París, firmado en 2015, donde las naciones se comprometieron a tomar medidas significativas para enfrentar el cambio climático y avanzar hacia la transición energética. Este acuerdo simboliza el reconocimiento global de la urgencia de actuar y la necesidad de colaboración internacional para mitigar los efectos devastadores del cambio climático, subrayando la importancia de transitar hacia un modelo energético más sostenible y respetuoso con el medio ambiente (Repsol, 2023).

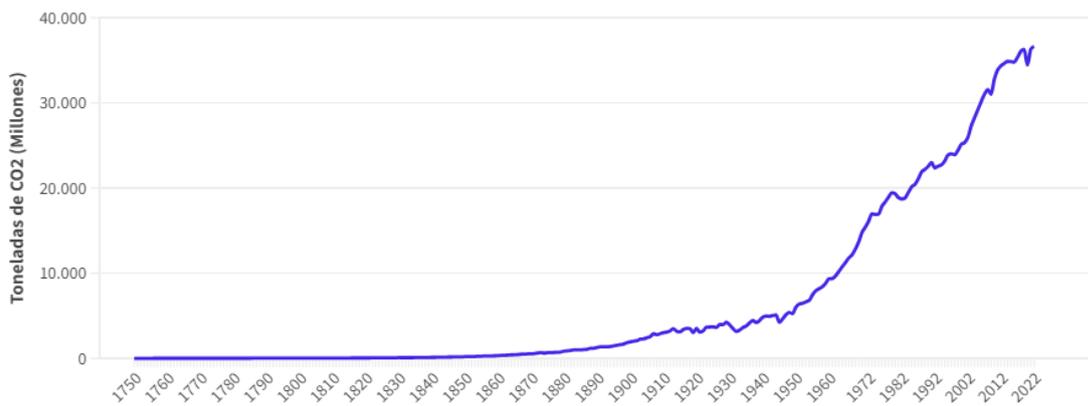


Figura 1.1: Evolución de las emisiones globales de CO2 procedentes de combustibles fósiles. Fuente de datos: epdata. (Orús, 2023). Elaboración propia

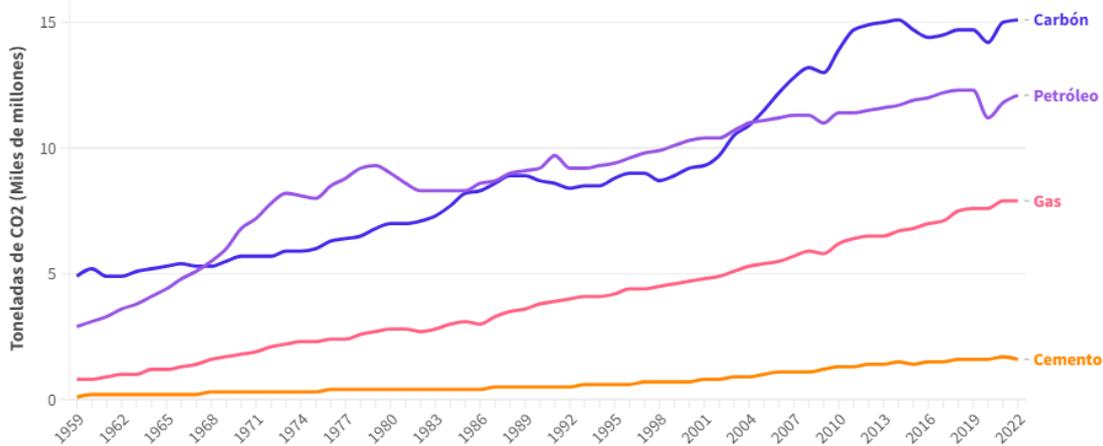


Figura 1.2: Evolución de las emisiones globales de CO2 por origen. Fuente de datos: epdata. (Orús, 2023). Elaboración propia

Es por esto que la relevancia de la transición energética se ha acentuado en el discurso académico y científico, donde diversas investigaciones han profundizado en esta temática. Un ejemplo notable es el informe de (Cantarero, 2020), que resalta las advertencias del Panel Intergubernamental sobre Cambio Climático (IPCC) sobre los efectos catastróficos del

cambio climático. La urgencia de reducir las emisiones de gases de efecto invernadero se ha vuelto ineludible frente a los evidentes riesgos del calentamiento global y el agotamiento de los recursos fósiles. La transición hacia una matriz energética más sostenible no es solo una cuestión medioambiental, sino también una necesidad social y económica, ofreciendo beneficios tangibles en múltiples ámbitos. En este escenario, entender la percepción de la sociedad y su relación con la transición energética es de suma importancia. La investigación realizada por (Gil-Pérez y Vilches, 2022) enfatiza que la aceptación y el apoyo del público son fundamentales para el éxito de cualquier política o medida orientada a esta transición. Este enfoque destaca que, más allá de las soluciones tecnológicas y políticas, la transición energética es también un proceso social que requiere la participación activa, la comprensión y el compromiso de todos los sectores de la sociedad. Así, el análisis de la percepción ciudadana sobre este tema se convierte en un elemento clave para guiar y formular estrategias efectivas que promuevan un cambio energético global hacia prácticas más sostenibles y respetuosas con el medio ambiente.

Para comprender plenamente la percepción ciudadana acerca de la transición energética, es crucial considerar el papel emergente de las redes sociales. Estos canales digitales no solo facilitan la expresión y el intercambio de opiniones y preocupaciones relacionadas con la energía y el medio ambiente, sino que también se han transformado en fuentes primordiales de Big Data para analizar tendencias y predecir comportamientos (Rodríguez y Chapis Cabrera, 2019). Según Andreu Bravo de Deloitte, las redes sociales representan una herramienta invaluable en este sentido, proporcionando una plataforma dinámica y en constante evolución para la recopilación y análisis de datos a gran escala (Bravo, 2023). Twitter, ahora X, es un ejemplo destacado de este fenómeno, albergando a más de 200 millones de usuarios activos mensualmente, una cifra en continuo crecimiento (Žiga Lesjak, 2023). La interacción diaria en esta plataforma genera un flujo constante de información, convirtiéndola en un espacio vital para el debate público. De hecho, una gran mayoría de los usuarios utiliza X principalmente para mantenerse al día con noticias y eventos (Ver Figura 1.3), lo que refleja su importancia como medio de comunicación y discusión sobre temas críticos como la transición energética. Estos debates abarcan desde la implementación de políticas más sostenibles hasta la adopción de energías renovables, atrayendo a científicos, políticos y otras figuras influyentes para intercambiar opiniones y avanzar en el discurso ambiental.

Este entorno convierte a las redes sociales en una herramienta esencial para comprender la percepción y el debate público en torno a la transición energética. La participación masiva en estas plataformas brinda una perspectiva única sobre el sentir de la sociedad, facilitando un análisis detallado de la forma en la que la ciudadanía responde a los retos implicados en el cambio hacia un sistema energético más sostenible. Esta interacción digital, que refleja una amplia gama de opiniones, es crucial para evaluar el pulso social frente a los desafíos ambientales y las oportunidades que ofrece la transición energética.

En esta línea de análisis de la percepción ciudadana a través de las redes sociales, se

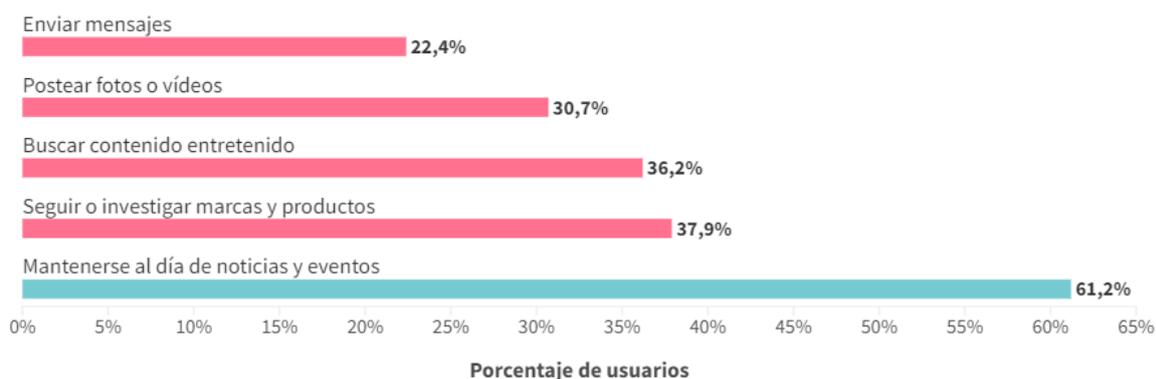


Figura 1.3: Principales actividades de los usuarios de X en el tercer trimestre de 2022.
Fuente de datos: Statista. (Fernández, 2023). Elaboración propia

enmarca el presente Trabajo de Fin de Grado. Este estudio se centra en brindar respuestas a preguntas fundamentales sobre el debate alrededor de los temas asociados a la transición energética en un contexto global. Con el objetivo de profundizar en este análisis, este trabajo propone el tratamiento de las publicaciones de X en los días previos al *COP26 Global Summit*. Este evento representa un momento clave en la agenda mundial de sostenibilidad energética, y las discusiones en la plataforma proporcionan una valiosa ventana a las opiniones y preocupaciones de la sociedad. La elección de este marco temporal y temático permite capturar una imagen instantánea de la conversación pública en un punto de inflexión en la política y la conciencia ambiental. Al analizar las interacciones, debates y opiniones publicadas en la red social X, se busca obtener información sobre la visión de los diferentes sectores de la sociedad hacia los desafíos de la transición energética. Este análisis ofrecerá información relevante no solo para facilitar la comprensión de la percepción pública, sino también para informar futuras políticas, estrategias de comunicación y acciones dirigidas a fomentar una transición energética más participativa.

1.2. Objetivos

En este contexto, el objetivo general de este Trabajo de Fin de Grado es analizar la percepción de la sociedad hacia la transición energética, utilizando el modelado de categorías de discusión y el análisis de sentimientos de publicaciones en la plataforma X (antes Twitter). Se busca comprender cómo se discuten y perciben los temas relacionados con la transición energética en esta red social, identificando tendencias, patrones de opinión y evaluando la polaridad de las opiniones expresadas. De esta manera, se busca dar soporte a la toma de decisiones más informadas en el ámbito de la energía sostenible. Con este propósito, se han establecido objetivos específicos que guiarán el desarrollo de este trabajo de fin de grado:

- Contextualizar la importancia de la transición energética en el entorno actual y destacar

cómo las redes sociales, en particular X (antes Twitter), se han convertido en un canal fundamental para la expresión de opiniones y percepciones de la sociedad en relación con temas de interés social y económico como este.

- Identificar y comprender las metodologías y herramientas disponibles para la extracción de información relevante, el modelado de tópicos y el análisis de sentimientos en publicaciones de redes sociales, a través de una revisión detallada de la literatura.
- Analizar las categorías de discusión presentes en los comentarios de X (antes Twitter) recopilados durante los cinco días previos al *COP26 Global Summit*, con el propósito de identificar las temáticas clave relacionadas con la transición energética que están atrayendo la atención en las redes sociales. Además, se llevará a cabo la evaluación de la polaridad del sentimiento asociado a cada categoría, permitiendo una comprensión detallada de si la percepción general es positiva, negativa o neutral.

1.3. Estructura del documento

Este Trabajo de Fin de Grado está estructurado en cinco capítulos. Inicialmente, se analiza la relevancia de las redes sociales como fuentes clave para la extracción de información sobre la percepción ciudadana, especialmente en temas de gran impacto socioeconómico, como la transición energética. En el capítulo 2, se llevará a cabo una revisión detallada de la literatura, dirigida a la identificación de metodologías de análisis automático que permitan el procesamiento eficaz de los enormes volúmenes de datos que se generan en las redes sociales.

Esta etapa sentará las bases para el capítulo 3 dividido en tres subapartados. La primera parte examina los datos de X centrados en los cinco días anteriores al *COP26 Global Summit*, seleccionados por su capacidad de capturar las expectativas, preocupaciones y opiniones de la sociedad en un momento crucial para la transición energética. La segunda parte se dedica a la configuración de un modelo de tópicos para identificar y clasificar las principales áreas de discusión en torno a la transición energética. La tercera parte consiste en un análisis de sentimientos, que permitirá evaluar la polaridad de las opiniones en cada categoría de discusión. Finalmente, en el capítulo 4 se presentarán los resultados obtenidos de los análisis previamente descritos, y en el capítulo 5 se presentarán las conclusiones clave obtenidas sobre la percepción pública en torno a la transición energética, y se abordarán las perspectivas y direcciones para trabajos futuros en este campo.

Capítulo 2

Analizando percepciones en el campo de la energía en X a través de técnicas de Machine Learning: Revisión de la literatura

Las redes sociales, en el panorama actual, han trascendido su rol original como plataformas de interacción social para convertirse en foros esenciales de debate público y difusión de información. Se ha observado que estas plataformas no solo facilitan el intercambio de opiniones sobre temas controvertidos, sino que también generan debates virales y contribuyen a la concienciación de la sociedad. Entre estas, X se ha distinguido como una red social líder en la comunicación de mensajes de manera instantánea y en la creación de discusiones e intercambio de diversos puntos de vista. Específicamente, X se ha destacado por ofrecer un espacio donde la libertad de expresión es accesible a todos sus usuarios, permitiendo así diálogos abiertos en una amplia gama de temas. Esta cualidad de X, permite el intercambio de información en tiempo real, enriqueciéndola con diferentes opiniones y reflexiones, aportando así un valor adicional no presente en otras redes sociales. Además, se ha identificado que X actúa como un canal informativo social, proporcionando noticias de acontecimientos mundiales algunas veces antes que los medios tradicionales (Agencia de Comunicación, 2023).

Debido a su naturaleza de *microblogging*, que promueve la difusión de mensajes cortos y concisos, y al gran volumen de datos que se genera en la plataforma, X ofrece ventajas significativas para la extracción de información sobre percepciones sociales en diversos temas. Sin embargo, este volumen masivo de datos requiere el empleo de técnicas automáticas de análisis de texto para realizar un análisis eficiente, manteniendo la fiabilidad y la integridad del contenido. En este contexto, el modelado de tópicos emerge como una de las técnicas más prominentes en el campo del Procesamiento de Lenguaje Natural (PLN). Esta metodología

se centra en la agrupación de textos en función de su temática, facilitando la identificación y categorización de las principales líneas de discusión dentro de grandes conjuntos de datos (Baviera, 2017). Al aplicar el modelado de tópicos, es posible descomponer y estructurar los comentarios recogidos de X, revelando patrones y temas subyacentes en las conversaciones, lo cual resulta invaluable para comprender las corrientes de opinión en la red social. Paralelamente, el análisis de sentimientos se posiciona como otra herramienta esencial en el estudio de datos de redes sociales. Esta técnica permite determinar las emociones, opiniones o actitudes expresadas en un conjunto de datos evaluando la polaridad del contenido: positiva, negativa o neutral (ISDI, 2023). A través del análisis de sentimientos, se puede extraer información significativa sobre la forma en la que los usuarios de X reaccionan emocionalmente a diferentes temáticas de discusión.

La utilización de técnicas avanzadas de PLN resulta sumamente útil para estudiar las percepciones sobre temas de gran interés global, como las energías renovables. Entender las opiniones públicas hacia las energías renovables es vital en el contexto de una transición energética global, ya que estas actitudes pueden tener un impacto notable en la adopción y respaldo de políticas y tecnologías sostenibles. Mediante el análisis de las conversaciones en X, estas herramientas brindan la oportunidad de identificar no solo las tendencias de discusión predominantes, sino también el tono emocional y la receptividad de la sociedad hacia los cambios en el sector energético. Este conocimiento resulta fundamental para elaborar estrategias comunicativas y políticas públicas que conecten con el público, facilitando así una transición más eficiente y aceptada hacia fuentes de energía más limpias y sostenibles.

Dentro de esta área, numerosos investigadores han abordado la temática de las percepciones sociales hacia las energías renovables en años recientes. Un estudio relevante en el campo fue el realizado por (Loureiro y Alló, 2020), el cual exploró la relación entre la percepción pública sobre el cambio climático y las preferencias respecto a políticas energéticas en Reino Unido y España. Utilizando el análisis de sentimientos basado en el léxico del National Research Center Canada (NRC), los autores analizaron comentarios generados en X, centrándose en el período comprendido entre el 1 de enero y el 30 de junio de 2019, sobre el tema del cambio climático. Los resultados del estudio indicaron diferencias notables en la connotación emocional entre los dos países. En Reino Unido, se observó una connotación más positiva hacia el cambio climático, con un sentimiento predominante de anticipación, mientras que en España, el miedo fue el sentimiento más destacado. A pesar de estas diferencias en la tonalidad emocional, ambos países mostraron preferencias similares hacia las energías sostenibles, evidenciando opiniones positivas hacia estas y negativas hacia el uso de carbón.

Siguiendo esta línea de investigación, otro trabajo destacado es el de (Kim, Ganesan, Dickens, y Panda, 2021), que se enfocó en comprender la opinión pública de los usuarios de X en Estados Unidos respecto a la energía solar y los factores determinantes del mercado energético, elementos cruciales en la transición energética. Para llevar a cabo su estudio, emplearon

una técnica de clasificación utilizando el modelo Robustly Optimized Bidirectional Encoder Representations from Transformers (RoBERTa), una variante avanzada del modelo Bidirectional Encoder Representations from Transformers (BERT). Analizaron un total de 266,686 publicaciones desde enero hasta diciembre de 2020, aplicando análisis de sentimientos para determinar la polaridad de estos mensajes. Los resultados de esta investigación resaltaron diferencias regionales significativas en las actitudes hacia la energía solar, observándose una mayor positividad en los ciudadanos del norte de Estados Unidos. Curiosamente, las opiniones eran generalmente más favorables en aquellos estados donde la mayoría había votado al partido demócrata en 2020, en regiones con mercados solares más establecidos y maduros, así como en aquellos con regulaciones de medición neta favorables al consumidor final. Estas variaciones regionales tienen implicaciones directas en las políticas y estrategias energéticas de los distintos estados de EE. UU. Por lo tanto, los autores enfatizan la importancia de que los estados implementen políticas de medición neta favorables y fomenten el apoyo a las empresas de energía solar para el desarrollo favorable de esta industria.

Continuando con el análisis de sentimientos, en un estudio relacionado realizado por (Zhang, Abbas, y Iqbal, 2022), se investigaron las percepciones y sentimientos de los ciudadanos hacia las emisiones de gases de efecto invernadero y las energías renovables, con el objetivo de comprender las posibles reacciones ante las políticas de mitigación. Este estudio analizó inicialmente 201,600 publicaciones de la red social X, recolectadas entre el 1 de septiembre y el 30 de noviembre de 2020, utilizando un algoritmo de clasificación basado en teoría de redes y similitud semántica. En lo que respecta al análisis de sentimientos, este estudio identificó la satisfacción de los ciudadanos mediante los sentimientos expresados hacia fuentes de energía específicas, mientras que la importancia se cuantificó según la frecuencia de menciones. Posteriormente, se asignaron puntuaciones a cada tema, considerando tanto la satisfacción como la importancia. Los principales resultados revelaron que los sentimientos más positivos estaban asociados con la energía de biomasa, y que tanto la energía solar como la eólica se percibían como las opciones más prometedoras para la transición energética. Este estudio subraya la importancia de que las instituciones gubernamentales establezcan políticas sostenibles que reflejen las preferencias e intereses de los ciudadanos, enfatizando la necesidad de alinear las estrategias de mitigación del cambio climático con la opinión pública.

En un estudio complementario, (Zarrabeitia-Bilbao, Morales-i Gras, Rio-Belver, y Garechana Anacabe, 2022) realizaron la recopilación y análisis de aproximadamente 200,000 publicaciones a nivel global, emitidas durante el año 2020, con el propósito de explorar las tendencias sociales en materia de energía verde. Para alcanzar este objetivo, se recopilaron los datos automáticamente mediante técnicas de *web scraping* y, a continuación, se procesaron utilizando algoritmos como el Louvain multilevel, que detecta comunidades dentro del conjunto de datos. Posteriormente, la información procesada fue sometida a diversas técnicas de aprendizaje automático, incluyendo el análisis de sentimientos mediante el modelo

Valence Aware Dictionary and sEntiment Reasoner (VADER), lo cual permitió identificar la polaridad de la información. Los resultados de este estudio identificaron nueve comunidades temáticas, reveladas por el algoritmo mencionado, que abarcaban aspectos políticos, socioeconómicos y de activismo medioambiental, con una tendencia generalmente positiva hacia la energía verde. Este resultado resalta la relevancia de analizar las percepciones en este campo, considerando su impacto significativo en la formulación de decisiones estratégicas. Asimismo, este estudio subraya la valiosa perspectiva de investigación que ofrece la red social X en el análisis de temas relacionados con la energía verde y la sostenibilidad.

Siguiendo un enfoque similar, el estudio propuesto por (Ibar-Alonso, Quiroga-García, y Arenas-Parra, 2022) se enfocó en el análisis de publicaciones generadas antes y después del inicio de la guerra en Ucrania en 2022. Específicamente, se centraron en términos relacionados con la “energía verde” desde el 16 de febrero hasta el 3 de marzo de 2022. El objetivo era discernir las percepciones y opiniones de los ciudadanos para facilitar el fomento de políticas que apoyen el desarrollo adecuado de las energías renovables. En la primera etapa, se llevó a cabo la recolección de información mediante *social listening* utilizando el programa R. Posteriormente, se procesaron los datos recopilados en la plataforma X con el mismo programa, empleando librerías como tidyverse, tidytext, stringi y topicmodel, lo que resultó en una base de datos final de 160,000 publicaciones. Finalmente, se realizó un análisis de sentimientos basado en léxico, para captar las emociones de los ciudadanos hacia la energía renovable o verde. Esto incluyó la identificación de posturas tanto pesimistas como optimistas y de expectativa, mediante la comparación del texto de las publicaciones con términos del diccionario NRC del paquete de R syuzhet. El análisis demostró cómo eventos geopolíticos significativos, como la guerra, pueden influir en la percepción pública de la transición energética, resaltando la interconexión entre sucesos globales y las opiniones ciudadanas sobre energía sostenible.

Además, resulta interesante explorar la relación entre las finanzas y la energía verde o renovable, como se investiga en el estudio realizado por (Maulida y Hakim, 2022). El propósito principal de este trabajo fue identificar las percepciones públicas sobre las finanzas verdes, examinando tanto las visiones positivas y beneficios como los aspectos negativos y amenazas, a partir del análisis de 1,779 publicaciones recogidas entre enero y diciembre de 2022. Utilizando la librería de Python VADER, se clasificó la polaridad de cada publicación para el análisis de sentimientos. Como resultado, se encontró que un 60.2 % de los sentimientos eran positivos, un 26.7 % neutros y un 13.1 % negativos, proporcionando información valiosa para que las partes interesadas puedan tomar medidas para mejorar la percepción y popularidad de las finanzas verdes.

Abordando una temática afin, el estudio de (Oji y Hameedah, s.f.) se centró en las opiniones y percepciones de los ciudadanos británicos respecto a la crisis energética, con el fin de obtener una visión más global del impacto que está teniendo en los hogares del Reino Unido. Para ello, se emplearon técnicas de análisis de sentimientos a partir de comentarios

obtenidos en la red social X, utilizando la librería de Python Snsrape con un límite de 2,000 publicaciones para evaluar la polaridad de los datos. Tras la limpieza de los datos, se aplicaron herramientas basadas en léxicos, como VADER y Textblob, para realizar el análisis de sentimientos. Los resultados obtenidos con VADER reflejaron de manera más precisa los impactos negativos de la crisis energética en el Reino Unido, así como en los comentarios de los usuarios de X. Estos estudios, al centrarse en aspectos financieros y de precios de la energía, complementan la comprensión general del contexto y las percepciones en torno a la transición energética y sostenibilidad.

Complementando los estudios previos, la investigación realizada por (Kastrati, Imran, Daudpota, Memon, y Kastrati, 2023) se enfoca en el significativo aumento experimentado por el precio de la energía en los últimos años, lo cual ha generado malestar entre los ciudadanos expresado en redes sociales como X. A pesar de los esfuerzos gubernamentales mediante subsidios, este tema se ha convertido en un punto de discusión recurrente en la plataforma. El estudio tuvo un doble objetivo: identificar la polaridad de las opiniones ciudadanas mediante un análisis de sentimientos utilizando tecnologías como BERT, y descubrir los temas subyacentes relacionados con el incremento en los precios de la energía, empleando modelado de tópicos basado en BERTopic y Latent Dirichlet Allocation (LDA). Analizando un conjunto de 366,031 publicaciones recopiladas desde el 1 de enero de 2021 hasta el 18 de junio de 2022, se identificaron los temas más debatidos en X sobre los factores que inciden en el aumento de los precios de la energía. Asimismo, se observó la evolución de los sentimientos de los ciudadanos durante este período, siendo más negativa en 2022, en línea con los mayores incrementos en los precios de la energía. Los resultados sugieren la necesidad de implementar planes de concienciación que aborden aspectos identificados, como los vehículos eléctricos, la plantación de árboles o estrategias para reducir las facturas energéticas en los hogares, con el objetivo de mejorar el bienestar ciudadano.

En resumen, los estudios previos han demostrado que las redes sociales, en particular X, desempeñan un papel fundamental en la recopilación de percepciones y opiniones de la sociedad en torno a la energía. La influencia de las discusiones en línea sobre la formulación de políticas y estrategias se ha vuelto innegable, y estos estudios han aportado conocimientos valiosos sobre la importancia de evaluar la percepción pública en este contexto. Por esta razón, el presente Trabajo de Fin de Grado utilizará X como su fuente de datos para analizar las percepciones de la sociedad hacia la transición energética, específicamente durante los cinco días previos al *COP26 Global Summit*. A través del análisis de publicaciones en X, se busca profundizar en la comprensión de la forma en la que se discuten y perciben los temas relacionados con la energía sostenible, con el objetivo de identificar tendencias, patrones de opinión y evaluar la polaridad de las percepciones expresadas.

Además, la revisión de la literatura ha permitido identificar las diversas técnicas utilizadas por los autores en sus investigaciones. Entre ellas, el modelado de tópicos mediante el uso de *transformers* como BERTopic o enfoques probabilísticos como LDA para categori-

zar las discusiones en las publicaciones recopiladas. Específicamente, LDA se ha destacado por su capacidad para analizar grandes volúmenes de datos, siendo además fácil de utilizar y ofreciendo resultados con alta interpretabilidad; es también capaz de funcionar de forma adecuada incluso cuando las características son mayores que la muestra de entrenamiento utilizada, y sabe interpretar adecuadamente la correlación existente entre las características de los datos utilizados (Raman, 2023). Por otro lado, el análisis de sentimientos ha sido una tarea común y consistente en todos los estudios revisados, empleando diversas herramientas como VADER, BERT, NRC, RoBERTa o syuzhet. La mayoría de los trabajos reportados emplean diccionarios predefinidos para evaluar la polaridad de las emociones expresadas en las publicaciones, demostrando ser herramientas extremadamente útiles para analizar diversos temas en redes sociales. En particular, el paquete de R syuzhet, que utiliza cuatro diccionarios de sentimientos, incluyendo NRC, destaca por su disponibilidad en más de 100 idiomas y su capacidad para asignar tanto sentimientos positivos y negativos como emociones específicas como enfado, sorpresa o miedo (Isasi, 2021). Además, es importante mencionar que VADER se ha revelado como una herramienta especialmente idónea para aplicaciones en redes sociales, debido a su enfoque en el lenguaje utilizado en estas plataformas, lo que le permite captar de manera efectiva las sutilezas y matices de la comunicación en línea. Su léxico está basado en reglas y tiene en cuenta detalles como son la puntuación, las negaciones, los emoticonos, las abreviaturas o la capitalización; además, es considerado como el “léxico de oro” y contiene más de 7,500 palabras validadas por humanos consiguiendo un léxico fiable y preciso (Taylor Ray, 2021). Debido a las ventajas anteriormente descritas, este Trabajo de Fin de Grado implementará tanto VADER como LDA en su metodología.

Por otro lado, en lo que respecta a la recopilación de información para el análisis de sentimientos, la mayoría de los estudios han optado por un método de filtrado utilizando palabras clave como “green energy”, “greenhouse gas”, “renewable energy”, “solar energy”, “hydro energy” o “natural gas”. Este enfoque ha permitido acceder a temas relevantes relacionados con la energía para la elaboración de sus investigaciones. Al emplear este método de filtrado, los datos analizados se alinean más estrechamente con el tema de interés, haciendo que el proceso de filtración y limpieza de datos sea más eficaz y centrado. Este enfoque se refleja en la base de datos elegida para el análisis reportado en el presente Trabajo de Fin de Grado.

En este contexto, la Tabla 2.1 presenta un resumen de los estudios analizados, destacando sus objetivos específicos, resultados clave, así como los algoritmos y metodologías empleados en la adquisición de datos. Este resumen ilustra de manera concisa la diversidad de enfoques y metodologías utilizadas en el campo, destacando la eficacia de las diferentes técnicas de PLN y análisis de datos en la extracción de información relevante sobre las percepciones y opiniones en redes sociales, particularmente en X. Asimismo, la tabla proporciona una visión comparativa de cómo cada estudio contribuye al entendimiento general de las actitudes y percepciones públicas hacia la energía y la transición energética.

Tabla 2.1: Resumen de los estudios sobre percepción pública llevados a cabo mediante el análisis de publicaciones en X.

Referencia	Tamaño del Data-set	Método de Adquisición	Objetivo	Algoritmo	Resultados
(Loureiro y Alló, 2020)	Desconocido	Filtrado de publicaciones por fechas: del 1 de enero al 30 de junio de 2019.	Analizar la relación existente entre la percepción pública sobre el cambio climático en Reino Unido y España y las preferencias que presentan los ciudadanos respecto a políticas energéticas.	Análisis de sentimientos: NRC.	Los usuarios muestran una connotación más positiva en Reino Unido que en España siendo su sentimiento más suscitado la anticipación, mientras que en España fue el miedo.
(Kim et al., 2021)	266,686 publicaciones.	Filtrado de publicaciones por fechas: de enero a diciembre de 2020. Filtrado de publicaciones por palabras clave: "solar energy", "solar panel", "solar PV", "solar photovoltaic", "solar battery", "solar thermal", "solar power", "solar-powered", "solar generation" y "solar subsidies". Filtrado de publicaciones por geolocalización: USA.	Analizar lo opiniones y sentimientos de los estadounidenses respecto a la energía solar y los determinantes del mercado energético.	Análisis de sentimientos: RoBERTa (BERT).	Los usuarios presentan diferencias regionales notables, con actitudes más positivas respecto a la energía solar por parte de los ciudadanos del norte de Estados Unidos, aquellos que votaron al partido demócrata en ese año, y aquellos que vivían en un estado con un mercado solar más asentado y maduro.
(Zhang et al., 2022)	201,600 publicaciones aprox	Filtrado de publicaciones por fechas: del 1 de septiembre al 30 de noviembre de 2020. Filtrado de publicaciones por palabras clave: "greenhouse gas", "GHG", "renewable energy", "coal", "natural gas", "solar energy", "wind energy", "biomass", "hydro energy", "geothermal energy", "tidal energy".	Estudiar las percepciones y sentimientos de los ciudadanos hacia la emisión de gases de efecto invernadero, así como las energías renovables para entender su posible reacción ante políticas de mitigación.	Análisis de sentimientos: algoritmo basado en teoría de redes y similitud semántica	Los usuarios muestran sentimientos más positivos hacia la energía de biomasa, y tanto la energía solar como la eólica destacan como las fuentes más prometedoras para la transición energética.
(Zarrabeitia-Bilbao et al., 2022)	200,000 publicaciones	Filtrado de publicaciones por palabras clave: "green-energy" ó "green energy". Filtrado de publicaciones por fecha: del 1 de enero al 31 de diciembre de 2020.	Explorar las tendencias sociales en materia de energía verde.	Análisis de Sentimientos: VADER.	Identificación de 9 comunidades con el algoritmo empleado, que abordan temas políticos, socioeconómicos y de activismo medioambiental, con un enfoque mayoritariamente positivo centrado en la energía verde, que ayuda a la toma de decisiones.
(Ibar-Alonso et al., 2022)	160,000 publicaciones aprox	Filtrado de publicaciones por fechas: del 16 de febrero al 3 de marzo de 2022. Filtrado de publicaciones por palabras clave: "green energy".	Analizar las percepciones y opiniones de los ciudadanos antes y después de la guerra de Ucrania, para poder promover políticas que favorezcan el correcto desarrollo de las energías renovables.	Análisis de sentimientos: NRC y syuzhet.	Opiniones variadas afectadas por el impacto de los eventos geopolíticos, como la guerra, en la percepción de la transición energética por parte de la población.
(Maulida y Hakim, 2022)	1,779 publicaciones	Filtrado de publicaciones por fechas: de enero a diciembre de 2022.	Analizar las percepciones públicas sobre las finanzas verdes, estudiando las visiones positivas y beneficios de las mismas, así como los aspectos negativos y amenazas presentes.	Análisis de sentimientos: VADER.	Los usuarios muestran un 60.2 % se sentimiento positivo, un 26.7 % de sentimiento neutro, y un sentimiento negativo del 13.1 %. Además, se espera que se tomen acciones conscientes para mejorar la visión y popularidad de las finanzas verdes.

Referencia	Tamaño del Dataset	Método de Adquisición	Objetivo	Algoritmo	Resultados
(Oji y Hamedah, s.f.)	10.732 publicaciones	Filtrado de publicaciones por fechas: del 5 de diciembre de 2022 al 7 de diciembre de 2022. Filtrado por idioma: solo en lengua inglesa. Filtrado por palabras clave: "ChatGPT".	Analizar las opiniones, percepciones y comentarios de los británicos en cuanto a la crisis energética, de forma que se pueda obtener una visión más global del impacto que está teniendo.	Análisis de sentimientos: VADER y Textlob.	Se observan los impactos tan negativos de la crisis energética en Reino Unido, así como en comentarios de X.
(Kastrati et al., 2023)	2,000 publicaciones	Filtrado de publicaciones por fecha: desde el 22 de noviembre de 2022 hasta la publicación del artículo. Filtrado por palabras clave: "electricity price", "electricity bill" ó "energy bill".	Analizar la percepción ciudadana hacia la subida de los precios de la energía.	Análisis de sentimientos: BERT. Modelado de tópicos: LDA y BERTopic.	Se descubrieron los temas más comentados sobre los factores que afectan al aumento de precios de energía, así como la evolución de los sentimientos de los usuarios, siendo la polaridad más negativa en 2022 cuando los precios de la energía fueron más altos.

Capítulo 3

Metodología de Análisis de Datos

En este capítulo se detalla la metodología empleada en el presente Trabajo de Fin de Grado, cuyo objetivo es analizar la percepción ciudadana hacia la transición energética. El capítulo se estructura en cinco secciones principales: la adquisición y preparación de los datos, el preprocesamiento de los datos, el análisis descriptivo de los N-gramas, el modelado de tópicos y, finalmente, el análisis de sentimientos. Cada una de estas secciones abarca diversas actividades, tal como se ilustra en la Figura 3.1, y se explican con detalle a lo largo de este capítulo.

1. **Adquisición de los datos:** En esta primera etapa, se selecciona el conjunto de datos a analizar, utilizando datos disponibles en repositorios de acceso libre como Kaggle. La elección de los datos se ha basado en los estudios revisados en la literatura del Capítulo 2, asegurando su relevancia y adecuación para los siguientes pasos de la metodología.
2. **Pre-procesamiento de los datos:** En esta etapa, se han tratado los datos seleccionados para garantizar su máxima calidad. El proceso de limpieza y pre-procesado de datos asegura la eliminación de aquellos datos que no sean relevantes para el estudio o que puedan generar ruido, interfiriendo en la coherencia y precisión de los resultados.
3. **Exploración de N-gramas:** En esta tercera etapa se llevará a cabo un análisis descriptivo de las palabras o conjuntos de palabras más relevantes en el conjunto de datos, utilizando bigramas (conjuntos de dos palabras) y trigramas (conjuntos de tres palabras). Esta exploración se realizará empleando la técnica de Term Frequency - Inverse Document Frequency (TF-IDF).
4. **Modelado de tópicos:** En esta etapa, se describe el uso del algoritmo LDA, que permite identificar las temáticas más relevantes dentro del conjunto de datos. LDA ofrece una alta interpretabilidad y facilita el análisis de grandes volúmenes de datos.

5. **Análisis de sentimientos:** En la etapa final, se aplicará el diccionario VADER para identificar el sentimiento asociado a las publicaciones del conjunto de datos recogidos en los días previos al *COP26 Global Summit*, categorizándolos como positivos, negativos o neutros.

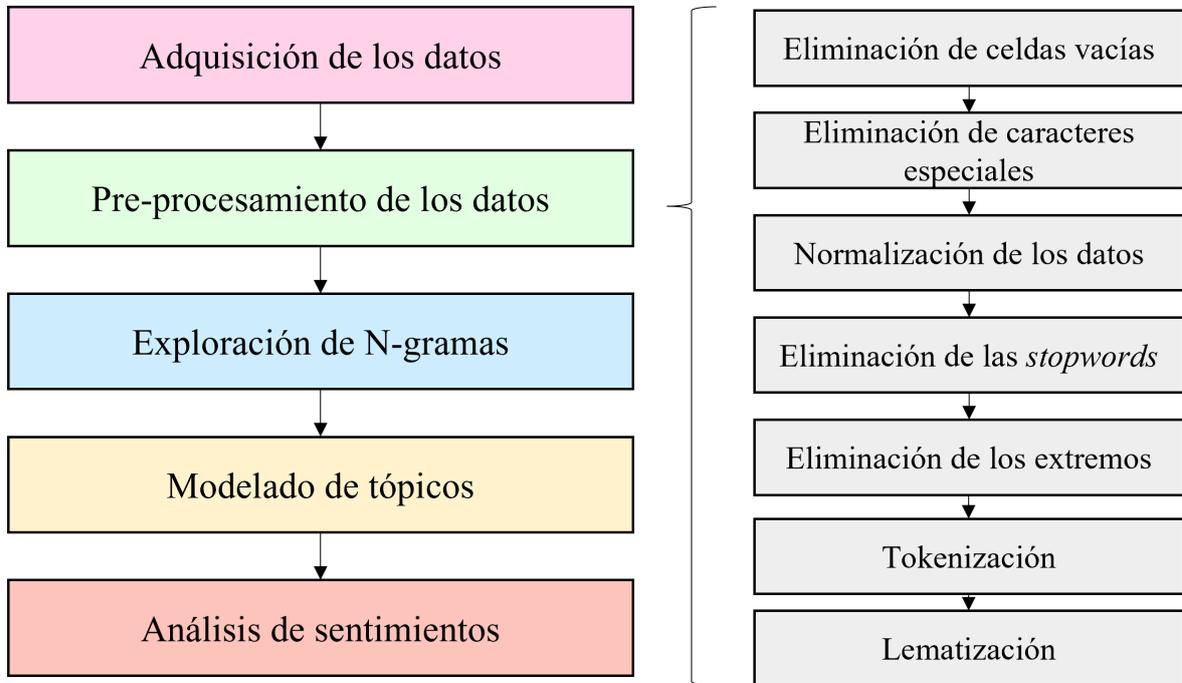


Figura 3.1: Metodología del Análisis de Datos. Elaboración propia

3.1. Adquisición de los Datos

Antes de realizar el modelado de tópicos y el análisis de sentimientos, es necesario disponer de un conjunto de datos de alta calidad que sirva como base para el trabajo. En este estudio, la recopilación de datos se ha centrado en comentarios extraídos de la red social X, específicamente relacionados con la energía y la transición energética. Estos temas se destacan en la revisión de la literatura por su relevancia en el debate público actual, su controversia y su impacto futuro en la sociedad. En este estudio, no se ha usado la API de X para la extracción de datos debido al cierre de su acceso público gratuito en 2023. Por esta razón, se realizó una búsqueda detallada en repositorios reconocidos como Kaggle, Github y Google Dataset Search para encontrar un conjunto de datos adecuado. Finalmente, se seleccionó un conjunto de publicaciones de acceso público disponible en el repositorio Kaggle, recopilados en los días previos al *COP26 Global Summit* y relacionados con la transición energética. Este dataset consta de 16,726 publicaciones y está dividido en 14 variables, lo que permite realizar un análisis descriptivo de los datos, más allá de su contenido textual, y sirve como base para

los análisis posteriores, incluyendo el modelado de tópicos y el análisis de sentimientos. Las variables que proporcionan información relevante y que serán utilizadas en este trabajo se detallan en la Tabla 3.1.

Nombre de la variable	Descripción de la variable
tweetID	Identificación de la publicación
date	Fecha de la publicación
tweetContent	Contenido de la publicación
tweetLanguage	Idioma de la publicación
replyCount	Número de respuestas
likeCount	Número de <i>likes</i>
retweetCount	Número de retweets

Tabla 3.1: Resumen de las variables más relevantes para el análisis.

Para el procesamiento posterior de los datos, se analizarán exclusivamente las publicaciones del conjunto de datos que están en inglés, ya que el 96 % de ellos están realizadas en este idioma. Esta decisión facilita el uso de herramientas y bibliotecas de análisis de texto automático en inglés, que son de alta calidad y ampliamente validadas. Además, trabajar con un idioma predominante en el dataset garantiza una mayor coherencia en el análisis y minimiza las complicaciones asociadas a la traducción y el procesamiento multilingüe.

3.2. Pre-procesamiento de los Datos

En esta segunda etapa, se emplearán herramientas específicas para la limpieza y el tratamiento de los datos, garantizando así su máxima calidad. Primero, se llevará a cabo un proceso de limpieza del dataset, eliminando todas las observaciones vacías y aquellos caracteres especiales, URLs, emoticonos, hashtags y signos de puntuación que puedan interferir en los resultados finales. Posteriormente, se normalizarán todas las publicaciones a minúsculas, de manera que no exista ninguna diferencia entre términos como “Climate” y “climate”. Este proceso de normalización es necesario para asegurar que el análisis sea más eficaz y de mayor calidad, evitando inconsistencias y mejorando la precisión en la identificación de patrones y tendencias en los datos.

En segundo lugar, se eliminarán las *stopwords*, también conocidas como palabras vacías, que son muy comunes en el idioma y no aportan información significativa. Al eliminarlas, el análisis puede enfocarse en palabras que realmente aporten valor y relevancia, mejorando así la precisión de los resultados. Algunos ejemplos de estas palabras vacías en inglés son “a”, “of”, “so”, “me”, “and” o “you”. Además, se realizará la eliminación de extremos, es decir, se suprimirán todas aquellas palabras que se repiten en la mayoría de las publicaciones (más del 95 % del corpus) o que son muy infrecuentes (menos del 2 % del corpus). Este proceso

es necesario, ya que permite identificar más fácilmente las categorías de discusión, al filtrar el ruido generado por términos excesivamente comunes o extremadamente raros.

En tercer lugar, se procede a la tokenización, un proceso que divide cada publicación en unidades más pequeñas denominadas *tokens*. Por ejemplo, la publicación “Está en nuestras manos mejorar el futuro global” se dividiría en “está”, “en”, “nuestras”, “manos”, “mejorar”, “el”, “futuro”, “global”. Posteriormente, se lleva a cabo la lematización, un proceso mediante el cual las palabras o *tokens* identificados se reducen a su forma base o lema. Esto implica que todas las palabras que comparten una misma raíz se unifican bajo una única forma, facilitando el análisis. Por ejemplo, las palabras “cleaning”, “cleanest” y “cleaner” se convierten en “clean”, lo que ayuda a agrupar las palabras bajo un mismo término. Este proceso mejora la interpretabilidad de los resultados y genera consistencia en el conjunto de datos, siendo de gran utilidad para etapas posteriores como el análisis de N-gramas y el modelado de tópicos, al permitir una identificación más sencilla de las categorías de discusión más relevantes. Sin embargo, el proceso descrito anteriormente no se aplica a la etapa de análisis de sentimientos ya que tanto las mayúsculas y minúsculas, *stopwords*, emoticonos o signos de puntuación, son muy relevantes en el análisis, al proporcionar matices necesarios para una evaluación precisa de las emociones expresadas en el conjunto de datos estudiado.

Esta etapa de preprocesamiento de los datos se centra en la limpieza de los datos, garantizando la máxima depuración y preparación de los mismos para las etapas siguientes. La cuidadosa depuración de la información es indispensable para asegurar que el análisis subsiguiente, tanto en el modelado de tópicos como en el análisis de sentimientos, se realice de manera efectiva.

3.3. Exploración de N-Gramas

Tras el procesamiento y limpieza de los datos, se procede a realizar un análisis descriptivo de los N-gramas para identificar las palabras o conjuntos de palabras más relevantes en el texto. Los N-gramas en el conjunto de datos pueden clasificarse como unigramas (una sola palabra), bigramas (dos palabras) o trigramas (tres palabras). Este análisis permite cuantificar la relevancia de estos N-gramas en el conjunto de datos utilizando la técnica de TF-IDF. Esta métrica se compone de dos partes: la Frecuencia de Término (TF) y la Frecuencia Inversa de Documento (IDF). La TF calcula las veces que un término específico aparece en un documento, proporcionando una medida de su relevancia dentro del mismo (Yang, 2017). Este elemento se calcula mediante la Ecuación 3.1.

$$TF(x) = \frac{\text{Número de veces que aparece la palabra } x \text{ en nuestro conjunto de datos}}{\text{Número de palabras que tiene nuestro conjunto de datos}} \quad (3.1)$$

Por otra parte, según describe (Yang, 2017), la segunda parte de la métrica TF-IDF corres-

ponde a la IDF. Esta componente asigna un peso diferente a cada palabra en el documento, disminuyendo el peso de aquellos términos que son muy frecuentes y aumentando el peso de aquellos que no lo son. Este ajuste permite resaltar las palabras más informativas y distintivas del texto, mejorando la capacidad del análisis para identificar términos clave del corpus. La IDF se calcula mediante la Ecuación 3.2.

$$IDF(x) = \ln \left(\frac{\text{Número total de documentos}}{\text{Número de documentos que tienen la palabra } x} \right) \quad (3.2)$$

La fórmula compuesta de TF-IDF se obtiene multiplicando la TF por la IDF, como se muestra en la Ecuación 3.3.

$$TF-IDF(x) = TF(x) \times IDF(x) \quad (3.3)$$

Esta métrica permite identificar los términos más relevantes dentro del corpus específico de publicaciones de X, analizado en el presente Trabajo de Fin de Grado. La aplicación de la TF-IDF facilita la detección de palabras que son informativas y distintivas, eliminando el ruido de los términos comunes que no aportan valor significativo al análisis. Al destacar las palabras más importantes, se mejora la comprensión de los temas discutidos en las publicaciones, lo que brinda un análisis preliminar, de gran utilidad para las etapas posteriores.

3.4. Modelado de Tópicos

Para realizar el modelado de tópicos, se ha empleado la técnica de LDA. Esta técnica permite identificar los temas más relevantes presentes en el conjunto de datos, es decir, en el corpus de publicaciones de X. LDA es un modelo de aprendizaje no supervisado que facilita la extracción de información significativa a partir de grandes volúmenes de datos, garantizando que los análisis sean precisos e interpretables. Como se ha observado en el Capítulo 2, una gran diversidad de las investigaciones revisadas han optado por utilizar LDA como su herramienta principal para el modelado de tópicos, destacando su precisión y eficiencia. LDA permite descubrir patrones ocultos en los datos textuales, agrupando palabras que frecuentemente aparecen juntas y asignándolas a diferentes temas, lo cual permite comprender la estructura subyacente de las discusiones y percepciones expresadas en las publicaciones.

En este caso, para poder realizar el análisis de sentimientos a posteriori, LDA nos permite identificar las palabras, o conjuntos de palabras, que tienden a aparecer con mayor frecuencia en un texto, en función de una estimación de sus probabilidades. Por ejemplo, en este trabajo que analiza la transición energética a partir de un conjunto de publicaciones de X, las palabras que podrían aparecer con mayor frecuencia son “energía”, “transición” o “renovable”. Tras el cálculo de estas probabilidades, LDA permite comprender la relación existente entre estas palabras y los temas en torno a los cuales se agrupan, detectando así las temáticas más

relevantes del conjunto de publicaciones. Por ejemplo, podrían identificarse temáticas como la energía renovable o el gas natural. Finalmente, este modelo asigna dos o más temáticas a cada uno de los documentos, permitiendo un análisis más detallado sobre el contenido de cada documento. En resumen, gracias a que utiliza inferencias bayesianas, LDA en el modelado de tópicos puede estimar las probabilidades de que estas palabras vuelvan a aparecer en el documento, generando tópicos a partir de las mismas y clasificando estos tópicos dentro de los documentos. Este proceso proporciona una categorización adicional que permitirá abordar el análisis de sentimientos con mayor profundidad, además de mejorar la comprensión de los principales temas de discusión relacionados con la transición energética (Seth, 2021).

Para entender adecuadamente el funcionamiento del modelo LDA, es fundamental considerar dos grandes asunciones clave:

1. Los documentos son una mezcla de diferentes tópicos.
2. Los tópicos son una mezcla de diferentes palabras.

Básicamente, el modelo asume que los documentos se generan mediante un proceso generativo estadístico, donde cada documento está compuesto por un conjunto de tópicos y cada tópico está compuesto por un conjunto de palabras. Otra consideración importante al aplicar este modelo es la determinación de los tópicos que se generarán en el conjunto de datos, representada por la variable “ k ”, la cual debe estimarse a priori. Si la “ k ” asignada es demasiado grande, se corre el riesgo de una sobre-división de los datos, resultando en categorías innecesarias y cálculos imprecisos. Por el contrario, si la “ k ” asignada es muy pequeña, se pueden agrupar incorrectamente ciertas palabras en categorías inapropiadas, lo que también conduce a resultados inexactos. Por lo tanto, es necesario encontrar un equilibrio adecuado para asegurar un cálculo efectivo y preciso. Con el fin de estimar el parámetro asociado al número de tópicos, es decir, “ k ”, se requiere calcular primero el valor de coherencia, que representa la similitud semántica entre las palabras más relevantes dentro de un mismo tópico. Un valor de coherencia alto indica que las palabras dentro del mismo tópico son muy similares entre sí, mientras que un valor de coherencia bajo sugiere que las palabras dentro del mismo tópico no son similares. Para determinar el número óptimo de tópicos, se generan varios modelos con diferentes valores de “ k ” y se selecciona aquel que logre el mejor equilibrio entre un mayor índice de coherencia y un menor número de tópicos. Este proceso se realiza mediante la técnica conocida como el método del codo, en el cual se grafica la coherencia de los tópicos en función del número de tópicos y se elige el valor de “ k ” donde se observa un cambio notable en la pendiente de la curva, indicando el punto de compromiso óptimo entre complejidad y coherencia. Es importante tener en cuenta que el modelo LDA no considera el orden de las palabras dentro de un documento; en su lugar, las analiza de manera individual, como si estuvieran sacadas de una bolsa, un enfoque conocido como *bag of words*. La Figura 3.2 ilustra el funcionamiento del modelo LDA, destacando los parámetros

requeridos en el proceso de ajuste e identificación de los tópicos. A continuación, se listan las consideraciones más relevantes para comprender el proceso de modelado con LDA:

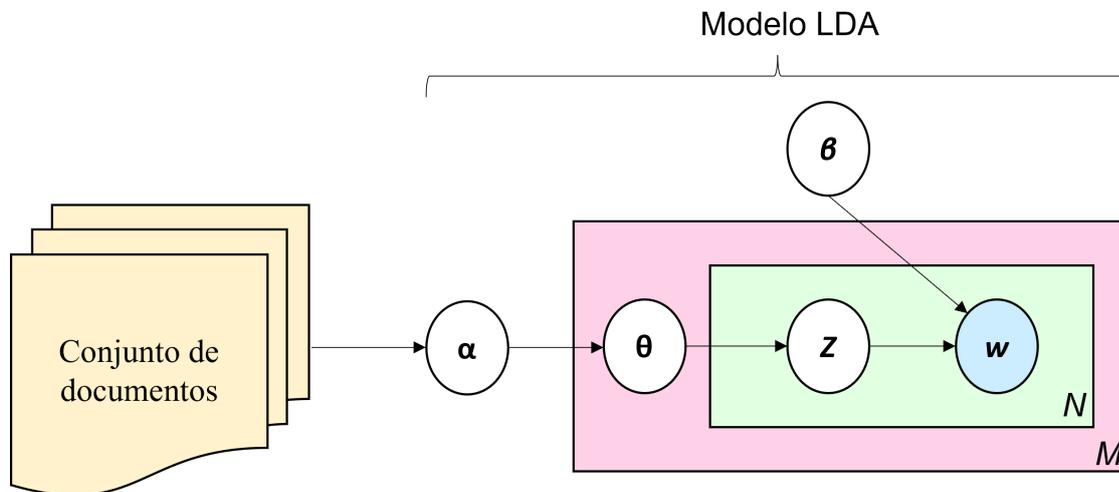


Figura 3.2: Representación gráfica del modelo LDA. Fuente de datos: Journal of Machine Learning Research. (Blei et al., 2003). Elaboración propia

- En primer lugar, el recuadro rosa hace referencia a todos los documentos que están presentes en el corpus M .
- Por otro lado, el recuadro verde engloba el número de palabras que hay en un documento, representado por la letra N .
- Además, dentro de este recuadro rosa existe una gran diversidad de palabras. Para referirse a una palabra específica en el documento, se utiliza la notación w , representada en color azul.
- Según este modelo, cada palabra está asociada a un tópico, que se representa con la letra Z , lo que da lugar a la distribución de tópicos, representada por theta (θ).
- Finalmente, alpha (α) hace referencia al parámetro que controla la distribución de los tópicos por documento, mientras que beta (β) controla la distribución de palabras por tópicos.

Finalmente, se visualizará la distancia intertópica de los tópicos identificados mediante LDA utilizando la librería *pyLDAvis*. Esta herramienta permite representar de forma bidimensional los distintos tópicos mediante círculos. De esta manera, es posible observar si existe algún tipo de solapamiento entre los diferentes tópicos y minimizarlo al máximo, ya que el solapamiento puede dificultar la interpretación clara de las temáticas modeladas (Van der Maaten y Hinton, 2008).

3.5. Análisis de Sentimientos

El análisis de sentimientos en redes sociales permite identificar y evaluar las emociones o actitudes expresadas en las publicaciones de los usuarios. Este proceso es fundamental para capturar las opiniones y percepciones sobre temas de alto impacto social, como la transición energética, que es el foco de este Trabajo de Fin de Grado. Mediante este análisis, se cuantifica el sentimiento subyacente en el texto, clasificándolo como positivo, negativo o neutro. Las conclusiones obtenidas de este análisis brindan información sobre la opinión de diferentes grupos de interés, ya sean relacionados con una marca, un país o una red social específica. Los datos obtenidos son valiosos para diversas aplicaciones, como la implementación de medidas gubernamentales adecuadas, la optimización de estrategias de marketing para aumentar las ventas de un producto, y la mejora de las estrategias de comunicación de una empresa o producto. Además, el análisis de sentimientos puede ayudar a identificar tendencias emergentes, gestionar la reputación en línea y desarrollar políticas que reflejen mejor las preocupaciones y necesidades del público.

Para este estudio, se empleará un diccionario basado en léxico conocido como VADER. Se trata de un modelo preentrenado específicamente diseñado para identificar el sentimiento de las publicaciones en X, clasificándolas como positivas, negativas o neutras. Este diccionario es especialmente eficaz para analizar documentos provenientes de redes sociales, debido a su sensibilidad para detectar sentimientos a partir de lenguaje coloquial, expresiones comunes, mayúsculas, puntuación y emoticonos. Una de las ventajas de utilizar VADER es que no requiere una extensa limpieza o preprocesamiento de los datos, como la eliminación de *stopwords*, mayúsculas o puntuaciones, procesos que son necesarios en el modelado de tópicos. En cambio, estos elementos se consideran valiosos, ya que proporcionan información relevante para captar los matices de los sentimientos en las publicaciones. La capacidad de VADER para manejar estos aspectos del lenguaje informal lo convierte en una herramienta ideal para el análisis de sentimientos en el contexto de las redes sociales, asegurando resultados representativos del tono emocional expresado por los usuarios en sus publicaciones. Además, este diccionario ofrece otras ventajas, como el hecho de que no requiere ningún tipo de entrenamiento previo, ya que se basa en un léxico generalizable. Esto lo hace rápido y eficaz, permitiendo su uso *online*.

VADER es entonces capaz de cuantificar el sentimiento positivo o negativo de una publicación, teniendo en cuenta todos los aspectos mencionados anteriormente (Bonta, Kumaresh, y Janardhan, 2019). Para ello, VADER analiza si las palabras del texto están presentes en su léxico con el fin de determinar su sentimiento y asignarles una polaridad. Esta polaridad se representa como un valor numérico normalizado entre [-1] y [+1], donde [-1] indica el sentimiento más negativo y [+1] el sentimiento más positivo. Para asignar esta puntuación a cada publicación, se calculan y suman las puntuaciones individuales de cada una de las palabras que los componen, y luego se normalizan dentro del rango de [-1] a [+1]. Esto proporciona

una puntuación numérica compuesta que permite comparar la emoción o sentimiento de cada publicación de manera rápida y efectiva (Bonta et al., 2019).

Capítulo 4

Resultados

En este capítulo se presentan los resultados obtenidos a partir de la metodología descrita en el Capítulo 3 de este Trabajo de Fin de Grado. Se abordan los principales resultados sobre el conjunto de datos relacionado con la transición energética en cada una de las etapas de la metodología, comenzando por el preprocesamiento de los datos y la exploración de los N-gramas, y finalizando con el modelado de tópicos y el análisis de sentimientos. El código utilizado para la realización de este análisis está disponible en el siguiente repositorio: (Soriano, 2024).

4.1. Adquisición y preparación de los datos

El conjunto de datos utilizado proviene de un repositorio en Kaggle y consiste en publicaciones recopiladas de la red social X en los días previos al *COP26 Global Summit*. Este conjunto de datos, relacionado con la transición energética, originalmente contenía 16,726 observaciones distribuidas en 14 variables. Durante el proceso de limpieza de datos, se eliminaron 7 variables que no aportaban información relevante para el análisis, conservando únicamente las variables relevantes mencionadas en la sección 3.1. Estas variables incluyen el contenido de cada publicación, la fecha de publicación, el idioma, y el número de *likes*, respuestas y retweets.

Asimismo, antes del proceso de limpieza, se realizaron varios análisis preliminares para estudiar la distribución de las publicaciones a lo largo del tiempo, el idioma principal utilizado en este conjunto de datos y el número de *likes* y retweets por publicación. Como se muestra en la Figura 4.1, el inglés es el idioma predominante con 16,014 publicaciones, lo que supone un 95,74 % del conjunto de datos. Los demás idiomas representan una fracción muy pequeña y son prácticamente irrelevantes para el análisis.

Además, como se muestra en la Figura 4.2, se analizaron los 20 hashtags más utilizados por los usuarios en sus publicaciones. Este análisis resalta la temática principal del trabajo sobre la transición energética y sus temas derivados, como las energías renovables y el

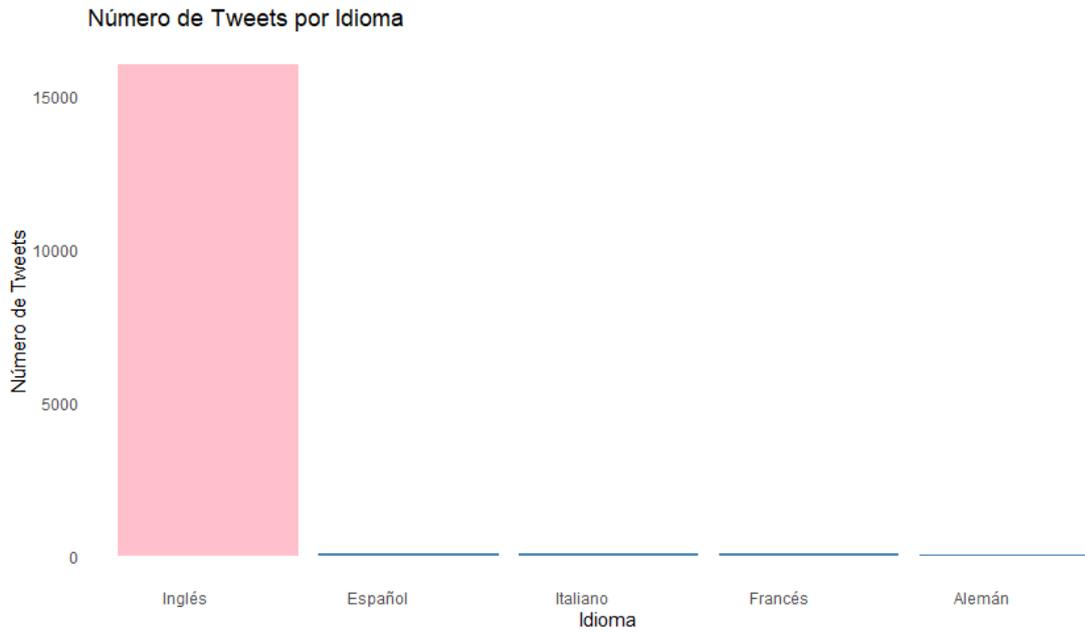


Figura 4.1: Número de publicaciones por Idioma. Elaboración propia

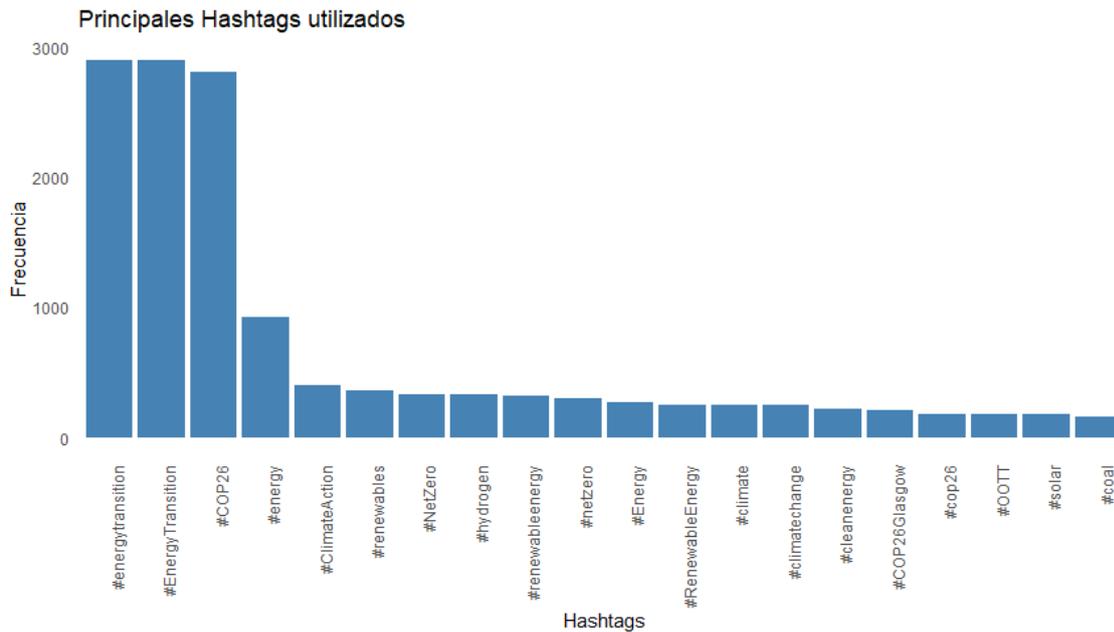


Figura 4.2: Principales Hashtags utilizados. Elaboración propia

cambio climático (#ClimateAction, #renewable, #RenewableEnergy, #climatechange). Estos hashtags reflejan un enfoque hacia sistemas energéticos sostenibles y el acceso a energía limpia, destacando también la estrecha relación entre la transición energética y la mitigación del cambio climático, con un énfasis particular en las energías renovables.

Por otro lado, se han estudiado también los retweets y *likes* por publicación a través de dos gráficos de cajas para observar su distribución en el conjunto de datos, como se puede obser-

var en las Figuras 4.3 y 4.4. En estos gráficos, se aprecia que la mayoría de las publicaciones presentan valores cercanos a cero en términos de *likes* y retweets. El rango intercuartílico de ambas figuras muestra que los *likes* tienen un rango ligeramente más amplio, lo que indica una mayor dispersión de los datos en comparación con los retweets. Esto sugiere que los usuarios son más propensos a dar *likes* a las publicaciones relacionadas con la transición energética que a republicarlas. En general, las publicaciones no están captando la atención suficiente de los usuarios como para volverse virales, ya que no logran una cantidad significativa de retweets ni de *likes*. Este comportamiento puede reflejar que, aunque hay un interés en el dialogo relacionado con la transición energética, este interés no es lo suficientemente fuerte como para generar una amplia difusión o interacción en términos de retweets y *likes*.

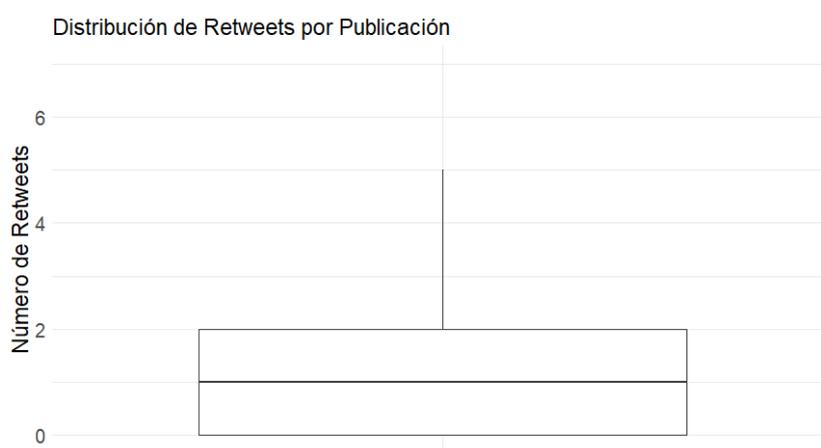


Figura 4.3: Distribución de retweets por publicación. Elaboración propia

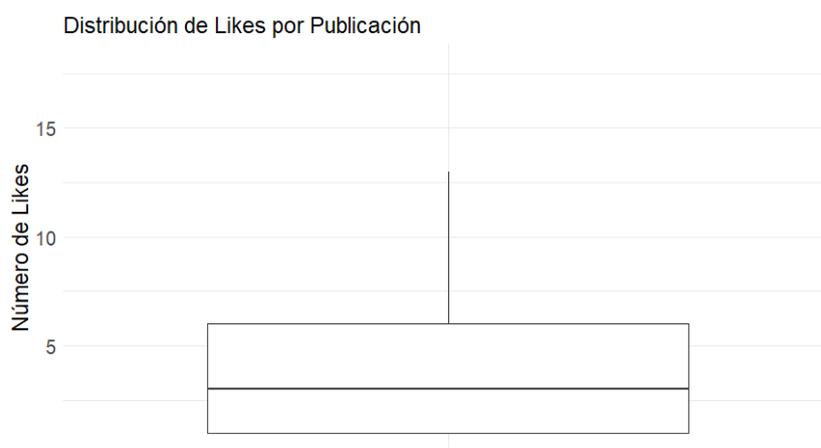


Figura 4.4: Distribución de *likes* por publicación. Elaboración propia

4.2. Pre-procesamiento de los datos

Como se ha expuesto en la sección 3.2, el preprocesamiento de los datos se centra en las tareas de limpieza y normalización, garantizando la máxima calidad y la eliminación de cualquier ruido o incoherencia que pueda interferir en los resultados finales.

En primer lugar, se eliminaron elementos no deseados del conjunto de datos, como URLs, hashtags, menciones, espacios en blanco adicionales, números y puntuación. Además, se normalizó el texto convirtiendo todos los caracteres a minúsculas para evitar distinciones entre palabras como “Climate” y “climate”. Esto es especialmente relevante, como se observó en la Figura 4.2, donde los principales hashtags empleados eran “energytransition” y “Energy-Transition”.

Posteriormente, se procedió a la eliminación de las *stopwords*, o palabras que no aportan información significativa, así como de las palabras extremadamente frecuentes o infrecuentes que no influyen en el análisis. Para ello, se descargaron diccionarios específicos para el procesamiento de datos en inglés, permitiendo detectar y eliminar todas aquellas palabras comunes que no aportan valor relevante al análisis. Ejemplos de estas *stopwords* incluyen “me”, “you”, “some”, “so” y “all”.

Después de la eliminación de las *stopwords*, se continuó con el proceso de lematización, mediante el cual las palabras se redujeron a su lema o raíz. Por ejemplo, palabras como “join”, “joined” y “joining” se redujeron a “join”, ya que comparten el mismo significado y no afectan el resultado del análisis. Además, se eliminaron todas las publicaciones que pudiesen contener contenido duplicado. Es importante destacar que, a pesar del uso de las últimas actualizaciones de las librerías para la limpieza del conjunto de datos, fue necesario añadir manualmente algunas *stopwords* y lemas que no habían sido eliminados o modificados adecuadamente. Un ejemplo de estos lemas añadidos manualmente incluye “discuss” en lugar de “discussion”, “achieve” en lugar de “achievement” y “develop” en lugar de “development”. En este punto del estudio, es importante destacar que este flujo de preprocesamiento de datos varía según la etapa del análisis. Para el análisis de sentimientos, el tratamiento de puntuaciones y emoticonos será diferente, ya que estos elementos son relevantes en este contexto. Estas diferencias se explicarán detalladamente en los siguientes subapartados.

Como resultado de esta etapa, el conjunto de datos inicial, que contenía 16,726 publicaciones, se redujo a 13,856 publicaciones después del preprocesamiento, lo que representa la eliminación de aproximadamente el 17% de los datos iniciales. Posteriormente, se generó una nube de palabras (*wordcloud*) para visualizar las 100 palabras más frecuentes en el conjunto de datos limpios, como se muestra en la Figura 4.5.

El análisis de la nube de palabras muestra que las palabras más frecuentes en el conjunto de datos están altamente relacionadas con el contexto de la transición energética. Términos como “clean”, “climate”, “new”, “need”, “global”, “renewable” y “green” destacan la conversación sobre la transición hacia energías sostenibles. La prominencia de estas palabras indica

las palabras, tal como se explicó en la sección 3.3. En este sentido, como se puede observar en la Figura 4.6, los unigramas o palabras individuales más relevantes según la métrica TF-IDF, son “clean”, “climate”, “need”, “global”, “new”, “green”, “accelerate”, “renewable”, “join” y “support”. Estas palabras también se destacan con un tamaño mayor en la nube de palabras de la Figura 4.5. Además, el predominio de unigramas como “need” y “support” resalta la urgencia y la demanda de acciones relacionadas con la transición energética. Valores de relevancia altos para términos como “clean” y “green” destacan la aspiración hacia energías más limpias y sostenibles. Además, palabras como “accelerate” y “renewable” indican un fuerte impulso hacia la adopción rápida de energías renovables. El término “global”, por su parte, refleja la escala y la importancia internacional de la transición energética, mientras que “join” y “support” sugieren un llamado a la acción y colaboración entre distintos actores.

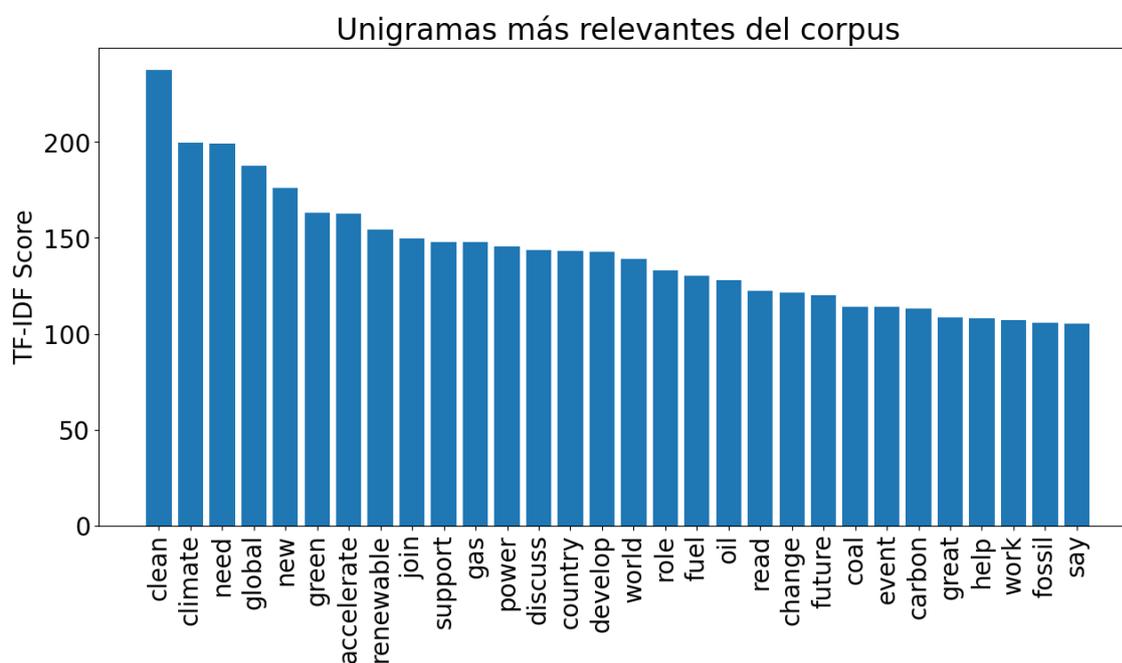


Figura 4.6: Principales unigramas. Elaboración propia

Posteriormente, se ha calculado el TF-IDF de los bigramas que conforman el corpus de la base de datos limpia, como se puede observar en la Figura 4.7. Las parejas de palabras que obtienen un mayor valor de relevancia son “fossil fuel”, “net zero”, “climate change”, “oil gas”, “low carbon”, “south africa”, “develop country”, “natural gas”, “green hydrogen” y “long term”. Algunos de estos bigramas, como “net zero”, “climate change” y “low carbon”, indican que los usuarios de la red social X están particularmente interesados en fuentes de energía limpia y renovable que minimicen el impacto ambiental. Estos términos reflejan un fuerte enfoque en la reducción de emisiones y en la adopción de políticas y tecnologías sostenibles. La frecuencia de bigramas como “fossil fuel” y “oil gas” sugiere que, además de promover energías limpias, hay una preocupación continua por el uso y el impacto de los combustibles fósiles. Bigramas como “south africa” y “develop country” pueden reflejar

discusiones sobre la equidad en la transición energética, destacando la importancia de considerar las necesidades y desafíos específicos de diferentes regiones y países en desarrollo. La presencia de términos como “green hydrogen” y “natural gas” sugiere un interés en tecnologías emergentes y alternativas dentro del ámbito de las energías sostenibles. Finalmente, el bigrama “long term” subraya la perspectiva de sostenibilidad y planificación a largo plazo en las discusiones sobre la transición energética.

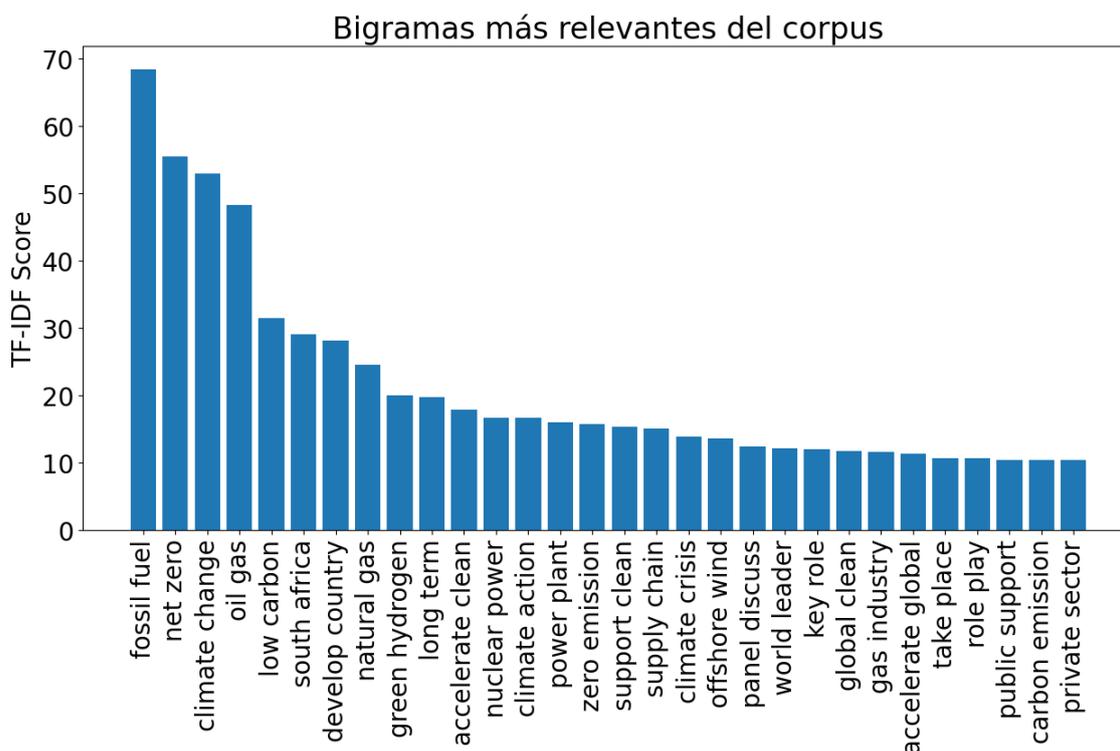


Figura 4.7: Principales bigramas. Elaboración propia

Finalmente, se ha calculado el TF-IDF de los 30 trigramas más relevantes del corpus de publicaciones, como se puede observar en la Figura 4.8. Entre los trigramas más destacados se encuentran “oil gas industry”, “net zero emission”, “play key role”, “oil gas company”, “achieve net zero” y “partnership south africa”. Estos conjuntos de palabras refuerzan la idea de que la sociedad es cada vez más consciente de la necesidad de una energía más limpia con el objetivo de lograr las cero emisiones. Trigramas como “achieve net zero” y “net zero emission” destacan el énfasis en la meta de alcanzar emisiones netas cero, reflejando un compromiso con la sostenibilidad ambiental. La relevancia de trigramas como “partnership south africa”, cuyos términos también aparecieron en el análisis de bigramas (Figura 4.7), sugiere la relevancia particular de este país en el contexto del *COP26 Global Summit*. Sudáfrica, con su economía altamente dependiente del carbón, anunció importantes iniciativas y proyectos basados en energía renovable durante el evento, lo que explica su prominencia en las discusiones. Finalmente, trigramas como “oil gas industry” y “oil gas company” indican una continua atención a las industrias tradicionales de combustibles fósiles, lo que puede reflejar

tanto la preocupación por su impacto ambiental como el reconocimiento de la necesidad de una transición justa y equilibrada hacia energías más limpias.

En resumen, los unigramas, bigramas y trigramas estudiados giran en torno a la misma temática, enfatizando la urgente necesidad de abordar el cambio climático y de implementar acciones que promuevan energías renovables, limpias y verdes que no perjudiquen el medioambiente. Este análisis destaca el creciente compromiso de la sociedad hacia la sostenibilidad y la importancia de una transición energética efectiva y ética.

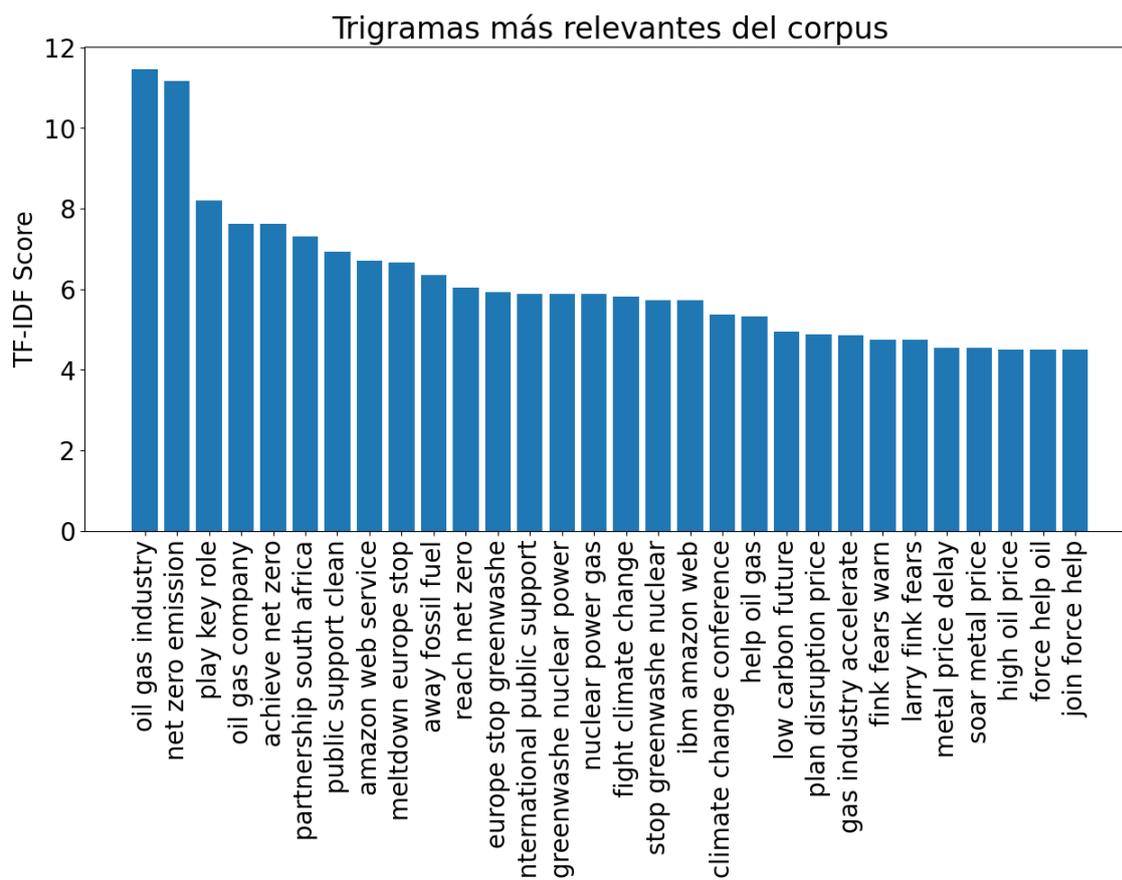


Figura 4.8: Principales trigramas. Elaboración propia

4.4. Modelado de tópicos

El modelado de tópicos es una técnica ampliamente utilizada para identificar y categorizar las discusiones presentes dentro de un conjunto de datos. En este estudio, se ha utilizado la técnica de LDA, descrita en la sección 3.4. Esta técnica permite descubrir patrones ocultos en el conjunto de datos y agrupar palabras, asignándolas a diferentes temas, lo que facilita una comprensión más estructurada del contenido.

Antes de implementar LDA, se llevó a cabo una etapa de preprocesamiento de los datos. Este proceso incluyó la normalización de las palabras a minúsculas, la lematización, la

filtración de extremos y la eliminación de caracteres especiales innecesarios para esta etapa del análisis, entre otras técnicas. Es importante destacar que, aunque estos elementos (como mayúsculas y signos de puntuación) son eliminados en el preprocesamiento para el modelado de tópicos, serán cruciales en la siguiente etapa de análisis de sentimientos. Por lo tanto, estas operaciones de limpieza no se llevarán a cabo en esa fase.

Tras el preprocesamiento de los datos, se procede a la implementación del modelado de tópicos. En primer lugar, es necesario determinar a priori el número de tópicos, “k”, adecuado para el conjunto de datos. Este número no debe ser ni demasiado grande ni demasiado pequeño, para evitar tanto la sobredivisión de los datos como la asignación a categorías inapropiadas. Para seleccionar el valor óptimo de “k”, se ha calculado el índice de coherencia o similitud semántica para diferentes valores de “k”, permitiendo evaluar la calidad de los tópicos generados. El objetivo es encontrar un equilibrio entre un alto índice de coherencia y un número razonable de tópicos, asegurando así que los temas identificados sean tanto relevantes como interpretables.

Como se puede observar en la Figura 4.9, se evaluaron diferentes valores para el número de tópicos, desde 2 hasta 14. En esta figura, se muestra que los valores de coherencia más altos son superiores a 0.39, y que a partir de 6 tópicos, el índice de coherencia disminuye drásticamente. Finalmente, se escogió un valor de “k” igual a 5, ya que es el número más bajo de tópicos que posee un mayor índice de coherencia, alcanzando 0.3999. Además, este es el punto donde se observa un cambio notable en la pendiente de la curva, indicando el compromiso óptimo entre complejidad y coherencia. Como se mencionó anteriormente, el objetivo es encontrar un equilibrio entre un mayor índice de coherencia y un menor número de tópicos, asegurando que cada uno de estos 5 tópicos contenga palabras relacionadas y se diferencien claramente entre sí.

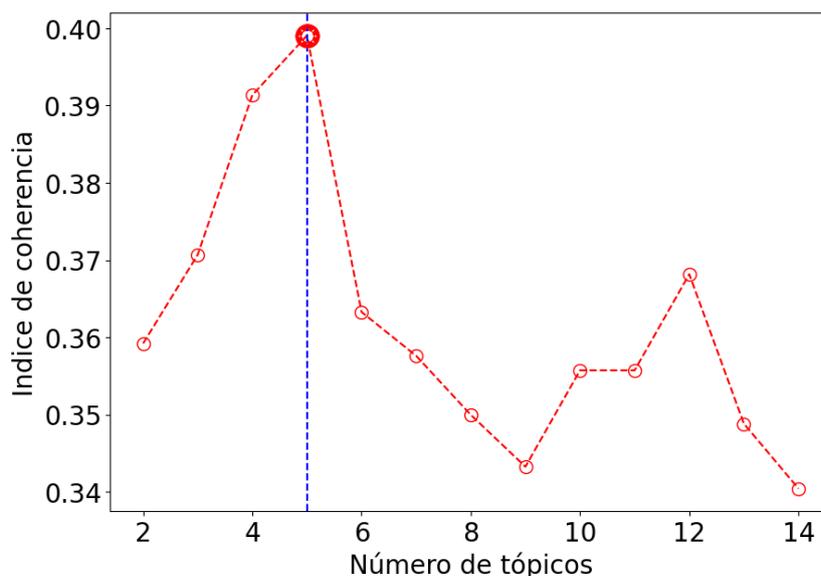


Figura 4.9: Índice de coherencia para cada número de tópicos. Elaboración propia

Durante este análisis, se utilizó también la herramienta *pyLDAvis* para visualizar la distancia intertópica de los 5 tópicos identificados. Esta visualización permite observar si existe algún tipo de solapamiento que pueda interferir en el análisis. Como se muestra en la Figura 4.10, existe una distancia intertópica considerable entre los tópicos, lo que indica que son distintos entre sí y no hacen referencia al mismo tema o contienen las mismas palabras. Sin embargo, se observa un ligero solapamiento entre los tópicos 1 y 4. A pesar de este solapamiento parcial, no es lo suficientemente significativo como para afectar negativamente el análisis, ya que las diferencias entre estos tópicos siguen siendo claras y relevantes. Esto asegura que cada tópico identificado aporta información única para entender mejor las discusiones sobre la transición energética en el conjunto de datos seleccionado.



Figura 4.10: Distancia intertópica para $k = 5$. Elaboración propia

Finalmente, para diferenciar claramente los 5 tópicos seleccionados, se les ha asignado una categoría basada en sus palabras clave, bigramas y trigramas. Esta categorización permite identificar la temática principal de cada tópico. Como resultado, la Tabla 4.1 muestra la etiqueta asignada a cada tópico, caracterizando su contenido principal y facilitando una interpretación más precisa de los temas presentes en el conjunto de datos.

Tabla 4.1: Tópicos, palabras clave, bigramas y trigramas más relevantes de los tópicos seleccionados

Tópico	Categoría	Palabras Clave	Bigramas	Trigramas
1	Políticas y desarrollo sostenible	“gas”, “need”, “carbon”, “fuel”, “oil”, “renewable”, “price”, “fossil”, “low” y “develop”	“net zero”, “south africa”, “accelerate clean”, “low carbon” y “long term”	“partnership south africa”, “world economic forum”, “net zero emission”, “post rise quarter” y “rise quarter profit”
2	Debates y desafíos en la transición energética	“climate”, “change”, “day”, “clean”, “talk”, “event”, “work”, “great”, “challenge” y “discuss”	“fossil fuel”, “oil gas”, “net zero”, “low carbon” y “natural gas”	“net zero emission”, “away fossil fuel”, “fink fears warn”, “larry fink fears” y “metal price delay”
3	Acción climática y vista a futuro	“clean”, “new”, “support”, “future”, “accelerate”, “global”, “partnership”, “launch”, “opportunity” y “green”	“climate change”, “offshore wind”, “climate action”, “accelerate inclusive” y “net zero”	“fight climate change”, “climate change conference”, “day opening event”, “coal day number” y “day number determined”
4	Papel europeo y estrategias	“coal”, “need”, “country”, “clean”, “power”, “plan”, “nuclear”, “climate”, “develop” y “say”	“oil gas”, “panel discuss”, “meltdown europe”, “nuclear power” y “take place”	“meltdown europe stop”, “amazon web service”, “oil gas industry”, “europe stop greenwashe” y “greenwashe nuclear power”
5	Apoyo global a la transición energética	“join”, “global”, “register”, “discuss”, “role”, “event”, “accelerate”, “november”, “week” y “company”	“fossil fuel”, “south africa”, “develop country”, “net zero” y “power plant”	“public support clean”, “plan disruption price”, “international public support”, “coal fire power” y “fossil fuel project”

En este contexto, el tópico 1 se centra en las publicaciones relacionadas con temas políticos y el desarrollo sostenible, como indica su categoría asignada, **políticas y desarrollo sostenible**. Este tópico aborda el controvertido tema de Sudáfrica durante el *COP26 Global Summit*, reflejado en los bigramas y trigramas “south africa” y “partnership south africa”. Durante el evento, Sudáfrica, cuya economía era altamente dependiente del carbón, anunció importantes iniciativas, políticas y proyectos basados en energía renovable. Además, se destacan palabras clave relacionadas con los combustibles fósiles, como “gas”, “carbon”, “fossil” y “oil”. Estas palabras destacan la necesidad de reducir el uso de combustibles fósiles en favor de energías más limpias y sostenibles, lo cual es fundamental para garantizar la salud del planeta a largo plazo. Este enfoque se refleja también en los bigramas “accelerate clean” y “long term”, que enfatizan la urgencia de acelerar la transición hacia energías limpias y la visión de sostenibilidad a largo plazo.

Por su parte, el tópico 2 se ha etiquetado en la categoría de **debates y desafíos de la transición energética**, abarcando temáticas relacionadas con las conversaciones enfocadas en la transición hacia una energía más limpia. Palabras clave como “work”, “challenge” y “discuss” reflejan los diálogos necesarios para cambiar el rumbo de la sociedad hacia una más limpia. El trígama más relevante, “net zero emission”, destaca el objetivo de alcanzar

una sociedad con cero emisiones. Estas palabras también evidencian los desafíos reales y potenciales que enfrenta la sociedad para alcanzar estas metas y alejarse de los combustibles fósiles. Bigramas relevantes como “fossil fuel” y “oil gas” hacen referencia a los obstáculos relacionados con la dependencia actual de estos recursos.

El tópico 3 se ha asociado a la categoría de **acción climática y perspectiva a futuro**, ya que tanto sus palabras clave como los bigramas y trigramas indican un enfoque orientado hacia el futuro y las acciones potenciales para combatir el cambio climático. Palabras relevantes como “future”, “opportunity”, “partnership” y “launch” refuerzan esta temática, mostrando la importancia de un futuro sostenible y las posibles alianzas necesarias para lograrlo. Además, trigramas como “fight climate change” y “climate change conference” también destacan la acción climática requerida para la transición energética, haciendo referencia a diversas iniciativas y conferencias dedicadas a abordar este problema que concierne a toda la sociedad.

El tópico 4, por otro lado, está asociado a la temática del **papel europeo y estrategias**. Este tópico abarca las discusiones y estrategias necesarias, como “develop” y “plan”, para realizar la transición hacia energías más limpias, destacando términos como “clean”. El papel fundamental de Europa se resalta con la mención del continente en bigramas y trigramas relevantes como “meltdown europe” y “europe stop greenwashing”. Estos términos también hacen referencia al fenómeno conocido como *greenwashing*, que implica engañar al consumidor sobre la sostenibilidad de una empresa, país o, en este caso, continente. Además, el trigramas “amazon web service” se destaca, lo que podría estar vinculado a la implementación de tecnologías innovadoras por parte de grandes corporaciones para apoyar la transición energética. Este tópico se enfoca en la necesidad de estrategias bien desarrolladas y la adopción de tecnologías avanzadas, así como la importancia de la transparencia y la autenticidad en las acciones ambientales de Europa para lograr una transición energética sostenible.

Por último, el tópico 5 se asocia con la temática de **apoyo global a la transición energética**. Este tópico está reforzado por palabras clave como “join”, “global”, “role” y “discuss”, que muestran la importancia del apoyo y la colaboración internacional para lograr una transición energética efectiva y poner fin a la dependencia de los combustibles fósiles. La idea de que la unión hace la fuerza es fundamental, especialmente cuando se persigue un objetivo común. En este contexto, se necesita una cooperación global para alcanzar rápidamente los objetivos de cero emisiones, como se refleja en los términos “accelerate” y “net zero”.

En la Figura 4.11 se muestra la frecuencia absoluta de cada uno de los tópicos. Se observa que el tópico 1, relacionado con las políticas y el desarrollo sostenible, cuenta con el mayor número de publicaciones, alcanzando un total de 3,568. Esto sugiere que las discusiones sobre políticas y desarrollo sostenible son las más prevalentes dentro del conjunto de datos analizado, reflejando la importancia de estos temas en el discurso público sobre la transición energética. Por otro lado, el tópico 5, que se refiere al apoyo global a la transición energética, es el menos frecuente, con 2,293 publicaciones. Esto podría indicar que, aunque el apoyo internacional es importante para la transición energética, no es tan discutido como

las políticas y estrategias específicas. La menor frecuencia de publicaciones en este tópico puede reflejar la necesidad de un mayor énfasis en la cooperación global y la colaboración internacional en el diálogo sobre energía sostenible.

En resumen, el modelado de tópicos ha permitido identificar una variedad de temas relacionados con la transición energética, que abarcan desde la acción global hasta estrategias, políticas y desafíos específicos. Esta técnica ha facilitado la comprensión de las opiniones y perspectivas de los usuarios de la red social X en relación con este tema de gran relevancia social.

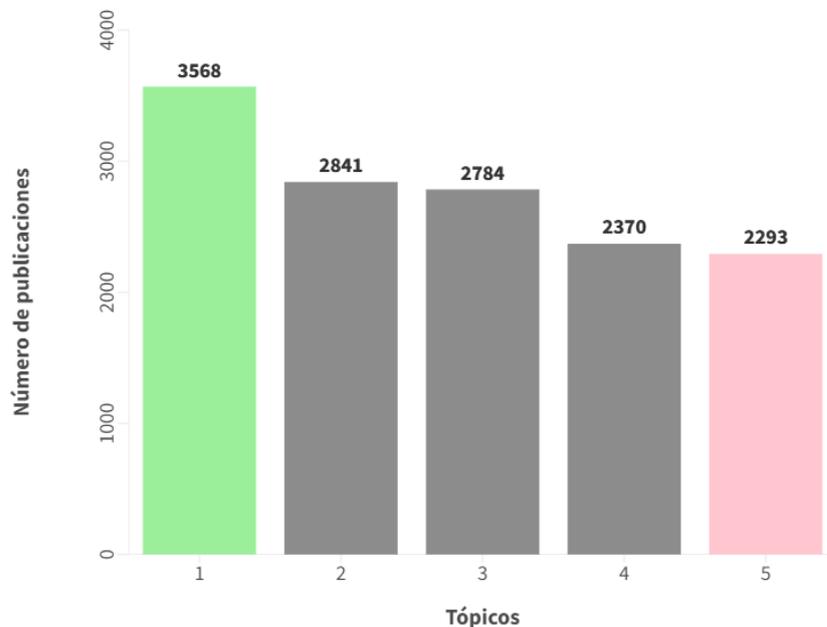


Figura 4.11: Distribución de los tópicos seleccionados. Elaboración propia

4.5. Análisis de sentimientos

Durante esta etapa, a través del análisis de sentimientos se pretende identificar y evaluar las emociones expresadas por los usuarios de la red social X en sus publicaciones sobre la transición energética. Para captar estas emociones, se ha utilizado el modelo basado en diccionarios llamado VADER, que permite identificar los sentimientos asociados a las publicaciones, clasificándolos como positivos, negativos o neutros, como se menciona en la Sección 3.5. Este modelo asigna a cada palabra una polaridad entre [-1] y [+1]. Posteriormente, se suman estas polaridades y se normalizan para obtener un valor final dentro de ese mismo rango.

Con el fin de preservar los elementos de análisis necesarios para el algoritmo VADER, se ha llevado a cabo un proceso de limpieza de datos diferente, en el cual se han mantenido

las mayúsculas, puntuaciones y emoticonos, ya que estos aportan información relevante sobre los sentimientos asociados a las publicaciones. Tras este procesamiento y la aplicación del modelo VADER, se elaboró un gráfico de cajas, como se muestra en la Figura 4.12, que representa la puntuación compuesta del conjunto de publicaciones, permitiendo observar la distribución del sentimiento asociado. A través de este diagrama de cajas, se puede observar la prevalencia de sentimientos positivos y neutros asociados a las publicaciones del conjunto de datos. Los sentimientos negativos, en comparación, no tienen una presencia significativa. La mediana de las puntuaciones de sentimiento se encuentra en 0.450909, lo que indica una tendencia general hacia sentimientos positivos. El valor mínimo de la puntuación es -0.969300, sugiere que hay una minoría de publicaciones con sentimientos negativos, pero estos no son lo suficientemente prevalentes como para afectar significativamente la distribución general.

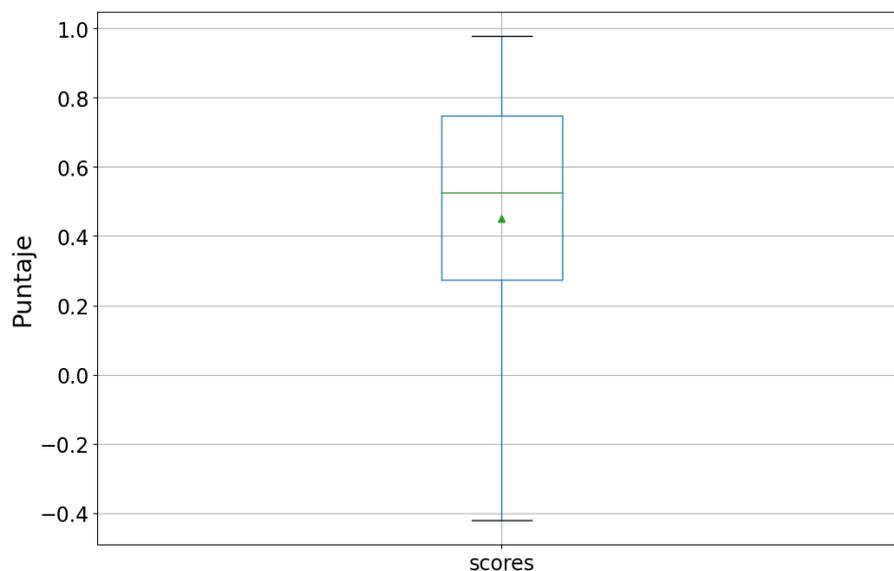


Figura 4.12: Distribución de la puntuación compuesta. Elaboración propia

Adicionalmente, para complementar el análisis de la puntuación compuesta, se ha realizado un histograma que representa las puntuaciones de las publicaciones estudiadas (ver Figura 4.13). El histograma muestra un comportamiento asimétrico hacia la derecha, indicando que la mayoría de las publicaciones están inclinadas hacia un sentimiento positivo. Esto es consistente con el gráfico de cajas de la Figura 4.12, donde se observa una mayor concentración de publicaciones positivas, predominantemente en el rango de 0.5 a 1. Esta predominancia de sentimientos positivos en este rango destaca el apoyo general y el optimismo de los usuarios hacia temas relacionados con energías limpias y sostenibles. Asimismo, se observa que uno de los picos más altos del histograma se encuentra en 0.0, lo que indica una significativa presencia de valores neutros. En contraste, los sentimientos negativos son poco frecuentes en comparación con los sentimientos positivos y neutros.

A continuación, tras analizar la densidad y el número de publicaciones por rango de puntuación, se ha examinado la distribución del sentimiento en cada uno de los cinco tópicos para

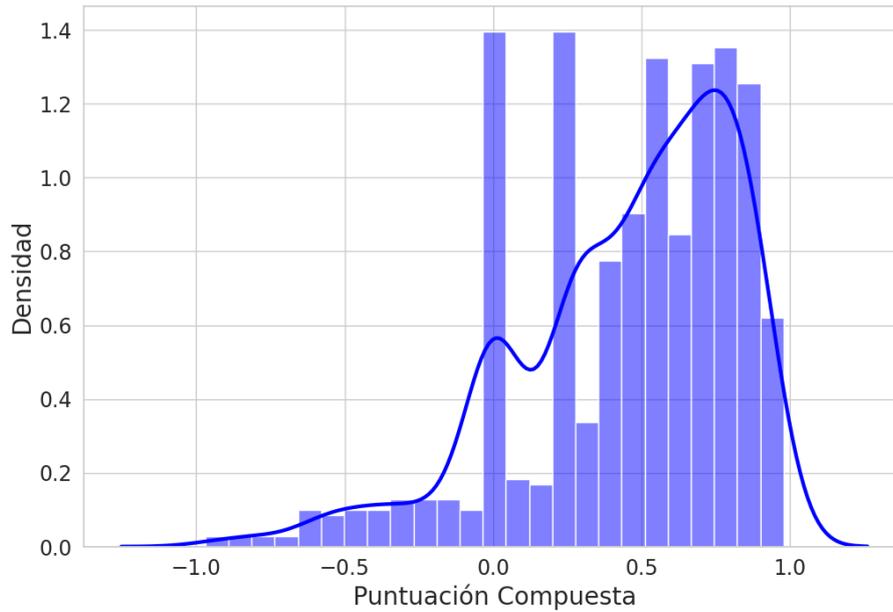


Figura 4.13: Número de publicaciones por puntuación. Elaboración propia

identificar aquellos con sentimientos más y menos positivos, como se puede observar en la Figura 4.14. Además, se han realizado análisis estadísticos descriptivos para cada tópico, incluyendo la media, la desviación estándar y el coeficiente de asimetría, como se muestra en la Tabla 4.2.

En relación con el tópico 1, **políticas y desarrollo sostenible**, se observa que presenta una mayor concentración de publicaciones con un sentimiento positivo asociado, con una media superior a la de los demás tópicos, alcanzando un valor de 0.554573. Además, de todos los tópicos, este es el que muestra menos opiniones negativas o neutras, con una puntuación mínima de -0.5875. Esto podría deberse a que los usuarios apoyan firmemente la implementación de nuevas políticas que promuevan la transición energética, reduciendo las emisiones causadas por los combustibles fósiles y fomentando una visión hacia el desarrollo sostenible a largo plazo.

En cuanto al tópico 2, **debates y desafíos de la transición energética**, muestra una distribución que se inclina hacia el lado positivo; sin embargo, es uno de los tópicos con mayor presencia de sentimientos negativos y neutros, con una media de 0.403564 y una puntuación mínima de -0.8555. El coeficiente de asimetría de -0.815814 indica una distribución sesgada hacia la izquierda, lo que significa que hay una mayor concentración de puntuaciones por encima de la media, pero también una cola larga hacia los valores más bajos o negativos. Esto sugiere que aunque muchos participantes tienen una visión positiva o neutral sobre la transición energética y sus desafíos, existe una minoría significativa con sentimientos muy negativos. Estos sentimientos negativos pueden estar relacionados con preocupaciones específicas, como el impacto económico de la transición, la pérdida de empleos en industrias tradicionales, o el costo de implementar nuevas tecnologías.

En el tópico 3, **acción climática y perspectiva a futuro**, se observa una media de 0.462357, marcada por un predominio de sentimientos positivos hacia un futuro sostenible. Este tópico fomenta las acciones y conferencias que abordan la transición energética, con el objetivo de lograr un planeta más limpio y sostenible. Sin embargo, también se percibe una cierta tendencia hacia sentimientos neutros e incluso negativos, posiblemente causados por preocupaciones sobre la efectividad de estas medidas y la rapidez de su implementación para alcanzar un mejor futuro para la sociedad. Esta diversidad de opiniones refleja tanto el apoyo entusiasta como las dudas y preocupaciones existentes sobre la viabilidad y el impacto de las acciones climáticas propuestas.

El tópico 4, **papel europeo y estrategias**, presenta una distribución similar a la del tópico 3, con estadísticas comparables que reflejan una prevalencia de sentimientos positivos y una menor cantidad de sentimientos negativos y neutros. La puntuación mínima de este tópico es -0.5994. Este patrón sugiere un sólido apoyo a las nuevas estrategias que fomentan la transición energética, similar al apoyo observado en el tópico 3. Sin embargo, también puede reflejar cierta preocupación o sentimiento de indiferencia hacia el *greenwashing*, un tema destacado en este tópico, así como sobre el papel de Europa en esta transición. La presencia de sentimientos negativos y neutros, aunque menor, indica que hay inquietudes sobre la sinceridad y efectividad de las estrategias europeas en la promoción de una transición energética genuina y sostenible.

El tópico 5, **apoyo global a la transición energética**, es el último estudiado en este trabajo y también el menos frecuente, presentando únicamente 2,293 publicaciones. Este tópico muestra una media inferior a la de los demás, con un valor de 0.340892, y alcanza una puntuación negativa mayor que el resto, con -0.9693. Además, presenta una mayor desviación estándar, lo que indica una mayor variabilidad en las opiniones de los usuarios y la presencia de opiniones extremas. Las puntuaciones compuestas de sentimiento están más dispersas respecto de la media en comparación con los demás tópicos. Una mayor presencia de sentimientos negativos y neutros podría reflejar la preocupación y el pensamiento crítico respecto al apoyo global actual hacia la transición energética. Estas críticas podrían estar fundamentadas en percepciones sobre la falta de unión y coordinación por parte de los diferentes países del mundo para alcanzar un objetivo común. Este análisis destaca la necesidad de una mayor cooperación internacional y de esfuerzos coordinados para lograr una transición energética efectiva y sostenible.

Como conclusión de este análisis de sentimientos, los resultados muestran que la temática de la transición energética está asociada mayoritariamente a sentimientos positivos entre los usuarios de la red social X. Esto indica un amplio apoyo público hacia las iniciativas de energías limpias y sostenibles. Este sentimiento positivo sugiere que hay una conciencia creciente y un respaldo significativo para las políticas y acciones destinadas a mitigar el cambio climático y promover un futuro energético sostenible. Sin embargo, la presencia de sentimientos neutros y negativos en algunos tópicos, como en el de debates y desafíos de la

Tabla 4.2: Estadísticas descriptivas por tópico

Tópico	Punt. Máxima	Punt. Mínima	Media	Desv. Típica	Coef. Asimetría
1	0.9678	-0.5875	0.554573	0.332239	-1.111790
2	0.9531	-0.8555	0.403564	0.376210	-0.815814
3	0.9776	-0.7906	0.462357	0.375258	-1.109842
4	0.9118	-0.5994	0.467031	0.325084	-0.819790
5	0.9548	-0.9693	0.340892	0.415152	-0.931972

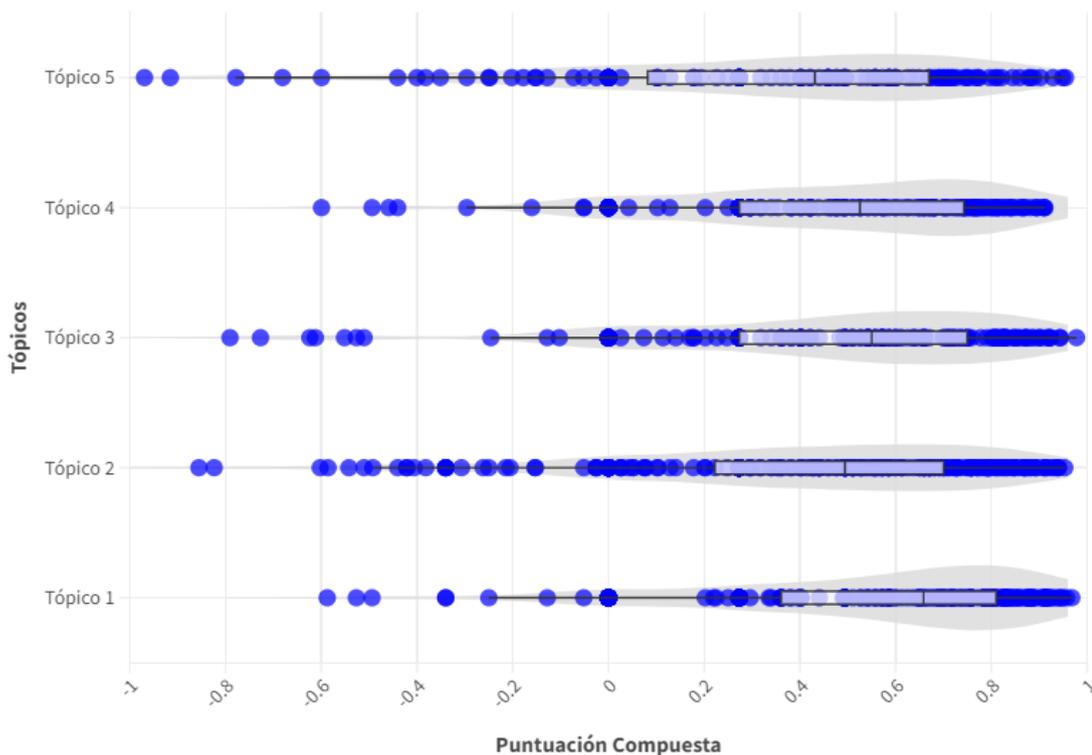


Figura 4.14: Distribución del sentimiento por tópico. Elaboración propia

transición energética, y en el de apoyo global a la transición energética, refuerza la necesidad de abordar las preocupaciones y críticas existentes. Estos sentimientos reflejan inquietudes sobre la efectividad de las medidas propuestas, la rapidez de su implementación, y la falta de coordinación internacional.

Capítulo 5

Conclusiones

La transición energética, campo que implica el cambio hacia fuentes de energía más limpias y sostenibles, es de gran relevancia para mitigar los efectos del cambio climático y asegurar un futuro más ecológico para las próximas generaciones. Este proceso no solo busca reducir las emisiones de gases de efecto invernadero, sino también promover la eficiencia energética y la innovación tecnológica en el sector energético. En este contexto, entender las percepciones de la sociedad sobre este tema es fundamental, ya que se requiere la aceptación pública y el apoyo para la implementación exitosa de políticas y medidas relacionadas con la transición energética. Las redes sociales, en particular, se han convertido en una herramienta útil y efectiva para extraer esta información, ya que proporcionan una plataforma donde los individuos pueden expresar sus opiniones y preocupaciones de manera libre y abierta. Al analizar las discusiones y sentimientos expresados en estas plataformas, es posible extraer información relevante, en tiempo real, de las actitudes de la sociedad hacia la transición energética, lo que puede guiar la toma de decisiones y la formulación de políticas más efectivas y alineadas con las necesidades de la sociedad.

En este contexto, este Trabajo de Fin de Grado se ha centrado en brindar respuestas en este campo. Se ha logrado comprender la percepción ciudadana acerca de la transición energética a través de la red social X, mediante la realización de un modelado de tópicos y un análisis de sentimientos. Este análisis se basó en un conjunto de datos recopilados en los días previos al *COP26 Global Summit*, un evento que representa un momento clave en la agenda mundial de sostenibilidad energética.

Con el fin de alcanzar este objetivo, se realizó una revisión de la literatura que abarca diversos estudios realizados a lo largo de los últimos años, relacionados con la percepción social hacia las energías renovables a través de la red social X, anteriormente conocida como Twitter. Gracias a este análisis de la literatura, se concluyó que la técnica LDA ha sido ampliamente aplicada para el modelado de tópicos previo al análisis de sentimientos, debido a su capacidad para analizar grandes volúmenes de datos, su facilidad de uso y los resultados altamente interpretables que ofrece. Además, se identificó que la técnica más adecuada para

el análisis de sentimientos en este estudio es la herramienta VADER. Este modelo se destaca por su léxico detallado, que considera aspectos como la puntuación y las negaciones, y contiene más de 7,500 palabras validadas por humanos, garantizando así un léxico fiable y preciso.

Posteriormente, se siguió un enfoque metodológico para identificar los temas de discusión y las percepciones y sentimientos de los usuarios en relación con la transición energética. Para ello, se recopilaron 16,726 publicaciones de la red social X en los días previos al *COP26 Global Summit*. Seguidamente, se realizó el preprocesamiento de los datos para limpiar y garantizar la calidad del conjunto de datos. Tras esta fase de limpieza, se procedió al análisis de N-gramas para identificar los términos más relevantes del conjunto de datos, analizando unigramas, bigramas y trigramas relacionados con la temática de la transición energética. Posteriormente, se llevó a cabo el modelado de tópicos utilizando la técnica LDA. Finalmente, se realizó un análisis de sentimientos empleando VADER, cuantificando las emociones asociadas a las publicaciones del conjunto de datos analizado.

En relación con el modelado de tópicos, se han identificado cinco tópicos mediante el análisis del índice de coherencia y la distancia intertópica. El primer tópico, denominado Políticas y desarrollo sostenible, se centra en la necesidad de dejar de utilizar combustibles fósiles en favor de energías más limpias, apoyando así el desarrollo sostenible. Este tópico también hace referencia a Sudáfrica, un tema destacado durante el *COP26 Global Summit*, donde el país anunció iniciativas para promover la transición energética. El segundo tópico, clasificado bajo la categoría de Debates y desafíos de la transición energética, aborda los desafíos, diálogos y conversaciones necesarias para alcanzar el objetivo de cero emisiones, alejándose de las energías basadas en combustibles fósiles. Este tópico resalta la importancia de la discusión y la colaboración para superar los obstáculos y avanzar hacia un futuro energético más sostenible. El tópico 3 se refiere a la acción climática y la perspectiva futura, enfocándose en la búsqueda de oportunidades y alianzas que puedan facilitar la transición energética en nuestra sociedad. Este tópico destaca la importancia de iniciativas proactivas y colaborativas para enfrentar los desafíos climáticos y avanzar hacia un futuro sostenible. El tópico 4, denominado “El papel europeo y estrategias”, resalta el papel relevante que tiene Europa en el contexto de la transición energética. Este tópico también aborda el problema del *greenwashing*, revelando palabras clave relacionadas con esta práctica y su impacto en las percepciones públicas sobre la sostenibilidad. Finalmente, el tópico 5 identificado como apoyo global a la transición energética, enfatiza la importancia de la cooperación internacional para lograr la transición energética y alcanzar las cero emisiones de manera rápida y efectiva.

En cuanto al análisis de sentimientos, tras la evaluación de los diferentes tópicos y el tratamiento específico de los datos para su limpieza, se observó una gran cantidad de sentimientos positivos asociados a las publicaciones del dataset, así como una notable presencia de sentimientos neutros, mientras que los sentimientos negativos eran poco prevalentes. Es-

te resultado podría revelar un amplio apoyo y optimismo hacia la transición energética por parte de los usuarios de la red social X, destacando un consenso general sobre la importancia y la necesidad de avanzar hacia energías más limpias y sostenibles. Asimismo, se realizó un análisis de la distribución del sentimiento en cada uno de los cinco tópicos identificados. A pesar de la predominancia de sentimientos positivos en todos ellos, se encontró que el tópico 2, **debates y desafíos de la transición energética**, y el tópico 5, **apoyo global a la transición energética**, presentaban una mayor carga de sentimientos neutros y negativos. El tópico 2, que aborda los desafíos y discusiones necesarios para lograr la transición energética, refleja la preocupación de los usuarios sobre la efectividad de las políticas actuales y las dificultades inherentes al proceso de cambio. Este sentimiento mixto sugiere que, aunque hay un reconocimiento de la necesidad de debate y colaboración, podrían existir dudas significativas sobre la viabilidad y el impacto real de las medidas propuestas. Por otro lado, el tópico 5, centrado en el apoyo global a la transición energética, muestra una mayor cantidad de sentimientos neutros y negativos en comparación con los demás tópicos, lo que puede indicar la percepción de una falta de coordinación y cooperación internacional. Los usuarios parecen expresar incertidumbre respecto a la capacidad de los diferentes países para unirse o trabajar juntos hacia un objetivo común de sostenibilidad energética.

Por su parte, el tópico 1, denominado **políticas y desarrollo sostenible**, no solo presentaba el mayor número de publicaciones, sino también el mayor sentimiento positivo asociado, con una media superior respecto al resto. Esto refleja un fuerte apoyo a las políticas y estrategias que promueven el desarrollo sostenible y la transición hacia energías más limpias. Los usuarios parecen estar a favor de las iniciativas que buscan reducir la dependencia de los combustibles fósiles y fomentar un desarrollo a largo plazo basado en la sostenibilidad.

Finalmente, como posibles líneas de investigación futura, se propone realizar un análisis de series temporales para evaluar la evolución del sentimiento a lo largo de los años. Dado que este dataset fue recopilado días previos al *COP26 Global Summit*, un análisis temporal con estos datos no es suficiente para evaluar la variabilidad del sentimiento a lo largo del tiempo. Resulta relevante observar la evolución de las preocupaciones, sentimientos y necesidades de los usuarios respecto a la transición energética, y determinar si han aumentado en los últimos años debido a la creciente incertidumbre sobre el cambio climático. Asimismo, la incorporación de variables demográficas podría ofrecer un estudio más detallado de las variaciones en cuanto a sentimiento y opinión sobre la transición energética en diferentes países. Esto permitiría identificar qué países tienen una actitud más positiva y cuáles presentan una actitud más negativa o neutra, proporcionando una visión comparativa del papel que desempeñan la cultura y el desarrollo industrial en las opiniones de los ciudadanos.

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que *ChatGPT* u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Andrea Soriano Fernández, estudiante de E2-Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado “Análisis de percepciones y sentimientos sobre la transición energética en redes sociales”, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa *ChatGPT* u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.
2. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
3. Interpretador de código: Para realizar análisis de datos preliminares.
4. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
5. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
6. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
7. Traductor: Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han

dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado *ChatGPT* u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: Junio 2024

Firma: Andrea Soriano Fernández

Referencias

- Agencia de Comunicación. (2023). *La influencia de twitter como medio de comunicación*. (acceso Enero 20, 2023) <https://www.agenciasdecomunicacion.org/actualidad/la-influencia-de-twitter-como-medio-de-comunicacion.html>.
- Baviera, T. (2017). Técnicas para el análisis del sentimiento en twitter: Aprendizaje automático supervisado y sentistrength. *Dígitos*, 1(3), 33–50.
- Blei, D. M., Ng, A. Y., y Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993–1022.
- Bonta, V., Kumares, N., y Janardhan, N. (2019). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, 8(S2), 1–6.
- Bravo, A. (2023). *La importancia de las redes sociales en las empresas y su gestión*. (acceso Noviembre 15, 2023) <https://www2.deloitte.com/es/es/pages/governance-risk-and-compliance/articles/importancia-redes-sociales-empresas-gestion.html>.
- Cantarero, M. M. V. (2020). Of renewable energy, energy democracy, and sustainable development: A roadmap to accelerate the energy transition in developing countries. *Energy Research & Social Science*, 70, 101716.
- Eneryou. (2023). *La transición energética: ¿qué es y por qué es importante?* (acceso Noviembre 14, 2023) <https://www.eneryou.org/post/la-transici%C3%B3n-energ%C3%A9tica-qu%C3%A9-es-y-por-qu%C3%A9-es-importante>.
- Fernández, R. (2023). *X (twitter) - datos estadísticos*. (acceso Noviembre 15, 2023) <https://es.statista.com/temas/10170/twitter-datos-mundiales/#topicOverview>.
- Gil-Pérez, D., y Vilches, A. (2022). Necesidad de un nuevo impulso a la transición a la sostenibilidad. *Documento de Trabajo*.
- Ibar-Alonso, R., Quiroga-García, R., y Arenas-Parra, M. (2022). Opinion mining of green energy sentiment: A russia-ukraine conflict analysis. *Mathematics*, 10(14), 2532.
- Isasi, J. (2021). *Análisis de sentimientos en r con 'syuzhet'*. (acceso Enero 28, 2023) <https://programminghistorian.org/es/lecciones/analisis-de>

-sentimientos-r.

- ISDI. (2023). *Sentiment analysis: ¿cómo funciona?* (acceso Enero 20, 2023) <https://www.isdi.education/es/blog/sentiment-analysis-como-funciona>.
- Kastrati, Z., Imran, A. S., Daudpota, S. M., Memon, M. A., y Kastrati, M. (2023). Soaring energy prices: Understanding public engagement on twitter using sentiment analysis and topic modeling with transformers. *IEEE Access*, *11*, 26541–26553.
- Kim, S. Y., Ganesan, K., Dickens, P., y Panda, S. (2021). Public sentiment toward solar energy—opinion mining of twitter using a transformer-based language model. *Sustainability*, *13*(5), 2673.
- Loureiro, M. L., y Alló, M. (2020). Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the uk and spain. *Energy Policy*, *143*, 111490.
- Maulida, S., y Hakim, B. A. (2022). Twitter sentiment analysis on green finance. *Accounting and Sustainability*, *1*(1).
- Mundo, B. N. (2021). *Los gráficos que muestran que más del 50% de las emisiones de co2 ocurrieron en los últimos 30 años.* (acceso Noviembre 15, 2023) <https://www.bbc.com/mundo/noticias-59013521>.
- Oji, O. P., y Hameedah, A. A. (s.f.). Lexicon-based sentiment analysis approach on the energy crisis in the united kingdom.
- Orús, A. (2023). *El cambio climático, en datos y gráficos.* (acceso Noviembre 15, 2023) <https://www.epdata.es/datos/cambio-climatico-datos-graficos/447?accion=2>.
- Raman. (2023). *ML | linear discriminant analysis.* (acceso Enero 31, 2023) <https://www.geeksforgeeks.org/ml-linear-discriminant-analysis/>.
- Repsol. (2023). *Un mundo con menos emisiones y más energía.* (acceso Noviembre 15, 2023) <https://www.repsol.com/es/energia-futuro/futuro-planeta/que-es-la-transicion-energetica/index.cshtml>.
- Rodríguez, V. A., y Chapis Cabrera, E. (2019). Importancia de las tecnologías de la información y las comunicaciones, el internet y las redes sociales en el mejoramiento y desarrollo de las empresas. *contribuciones a la Economía*(marzo).
- Seth, N. (2021). *Topic modeling and latent dirichlet allocation (lda) using gensim and sklearn.* (acceso Mayo 6, 2023) <https://www.analyticsvidhya.com/blog/2021/06/part-2-topic-modeling-and-latent-dirichlet-allocation-lda-using-gensim-and-sklearn/#:~:text=A%20tool%20and%20technique%20for,a%20mixture%20of%20topics%2C%20and>.
- Soriano, A. (2024). *Tfg energía.* (acceso Mayo 31, 2023) <https://github.com/andreasorianof/Andrea-Soriano/tree/main>.
- Taylor Ray, S. A. (2021). *Using vadersentiment to intuitively predict the sentiment of social media posts.* (acceso Enero 31, 2023) <https://ds3.ssrc.msstate.edu/2021/01/14/using-vadersentiment-to-intuitively-predict-the-sentiment-of>

-social-media-posts/.

- Van der Maaten, L., y Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(11).
- Yang, Y. (2017). Research and realization of internet public opinion analysis based on improved tf-idf algorithm. En *2017 16th international symposium on distributed computing and applications to business, engineering and science (dcabes)* (pp. 80–83).
- Zarrabeitia-Bilbao, E., Morales-i Gras, J., Rio-Belver, R. M., y Garechana Anacabe, G. (2022). Green energy: identifying development trends in society using twitter data mining to make strategic decisions. *Profesional de la Información*, 31(1).
- Zhang, Y., Abbas, M., y Iqbal, W. (2022). Perceptions of ghg emissions and renewable energy sources in europe, australia and the usa. *Environmental Science and Pollution Research*, 29(4), 5971–5987.
- Žiga Lesjak. (2023). *Twitter ahora se llama x - los usuarios de twitter se convirtieron en usuarios de x.* (acceso Noviembre 15, 2023) <https://tridenstechnology.com/es/twitter-usuarios-estadisticas/>.