

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE FILOLOGÍA

LA EMOTIVIDAD DE LOS «BUUS Y HURRAS». MODELO DE CLASIFICACIÓN DE COMENTARIOS DE YOUTUBE



MÁSTER EN LETRAS DIGITALES
CURSO 2020-2021

TRABAJO DE FIN DE MÁSTER

AUTORA: Mayor Dueñas, Lys

TUTOR: Martín Martín, Enrique

Departamento de Sistemas Informáticos y Computación

TUTOR: Riesco Rodríguez, Adrián

Departamento de Sistemas Informáticos y Computación

TUTOR: Robles Morales, José Manuel

Departamento de Sociología Aplicada

Madrid, septiembre 2021

Calificación: 9

*A mis tutores, Adrián, Enrique y José Manuel, y a mis compañeros del máster,
por hacer este camino transitable.*

A mi familia, por apoyarme en todas mis aventuras.

Gracias.

LA EMOTIVIDAD DE LOS «BUUS Y HURRAS». MODELO DE CLASIFICACIÓN DE COMENTARIOS DE YOUTUBE

Lys Mayor Dueñas

Resumen

Este trabajo de fin de máster presenta un modelo de aprendizaje automático para el etiquetado de comentarios de YouTube. El objetivo es medir la emotividad de los mensajes publicados en vídeos de la campaña electoral para las elecciones a la Asamblea de Madrid de 2021. El modelo, desarrollado en Python, se entrena con parámetros numéricos obtenidos con criterios lingüísticos y busca ofrecer resultados explicables, aplicando el método de árbol de decisión.

Los resultados obtenidos confirman que el Procesamiento del Lenguaje Natural es una tecnología idónea para estudiar la polarización en redes sociales; sin embargo, esta investigación pone de manifiesto las dificultades de trabajar con un conjunto de datos desequilibrado y propone posibles mejoras para futuros estudios.

El modelo está disponible en GitHub:

<https://github.com/LysMaDu/ModeloDeClasificacionDeComentarios>.

Palabras clave: polarización, redes sociales, emotivismo, PLN, aprendizaje automático, árbol de decisión

Abstract

This investigation presents a machine learning model for YouTube comments classification. The main purpose is to measure emotivity in messages posted on videos directly related to the May 2021 Community of Madrid's regional government election campaign. The model, developed in Python, is trained with numeric parameters based on linguistic criteria, and it seeks showing explainable results, by using a decision tree classifier.

The outcome shows that Natural Language Processing is a suitable technology for studies on polarization on social media. However, this research confirms the difficulty of working with imbalanced datasets and suggests possible improvements for further studies.

The model is available at GitHub:

<https://github.com/LysMaDu/ModeloDeClasificacionDeComentarios>.

Key words: polarization, social media, emotivism, NLP, Machine Learning, decision tree classifier

Índice

1. Introducción	6
2. Objetivos	8
2.1. Objetivos generales.	8
2.2. Objetivos específicos.	9
3. Estado de la cuestión	10
3.1. Las redes sociales y la comunicación política: La esfera pública, digital y política.	10
3.2. La polarización.	12
3.3. El civismo y las buenas maneras.	13
3.4. El emotivismo: en busca de «buus y hurras».	13
3.5. El Procesamiento del Lenguaje Natural y la Semántica Computacional.	14
3.6. El aprendizaje automático.	15
3.7. YouTube.	16
4. Metodología	18
4.1. Creación de corpus de comentarios: descarga y almacenamiento desde YouTube.	18
4.2. Análisis preliminar de las características de los datos descargados (corpus).	20
4.3. Definición de «buus y hurras» y parámetros lingüísticos para identificarlos.	21
4.4. Limpieza y procesamiento de los comentarios.	23
4.5. Etiquetado manual y creación de corpus de comentarios etiquetados.	26
4.6. Entrenamiento y aplicación del clasificador automático.	27
4.7. Evaluación mediante matriz de confusión.	27
5. Análisis de los resultados	29
5.1. Etiquetado manual.	29
5.2. Resultados del entrenamiento y el test del modelo.	29
5.3. Análisis mediante matriz de confusión.	31
6. Conclusiones	33
7. Bibliografía	35
7.1. Bibliografía primaria.	35
7.2. Bibliografía secundaria.	39

1. Introducción

La llegada de la tecnología supuso un cambio radical en la esfera pública. La información, mediante un proceso de desintermediación, pasó de estar en manos de los medios de comunicación y el poder institucional a estar al alcance de todo ciudadano con acceso a internet. Este nuevo panorama digital se ha leído tanto como una posibilidad de democratización de la esfera política, reforzado por la participación ciudadana, como un espacio caótico que propicia la polarización de la sociedad.

A pesar de ser relativamente reciente, este fenómeno se ha convertido en un eje central de la investigación académica relacionada con la comunicación. El comportamiento de los usuarios de internet, en concreto en redes sociales, se percibe como un buen reflejo de la situación actual. Y son las herramientas que nos proporciona la tecnología las que permiten medir las tendencias en dichos espacios.

Las tecnologías del *big data* han permitido trabajar con cantidades de datos que anteriormente era muy difícil conseguir. La informática, la sociología y la lingüística, entre otras disciplinas, se han aliado para conocer mejor la realidad del entorno virtual, en concreto, las tecnologías del Procesamiento del Lenguaje Natural han supuesto un impulso notable para dichos estudios, permitiendo analizar en detalle los mensajes publicados en internet.

Actualmente, la polarización política copa incluso los titulares de prensa. La aparición de nuevas fuerzas políticas situadas más a los extremos que las tradicionales se ha entendido como la materialización de ese distanciamiento de posturas en la sociedad. En España, en los últimos 10 años, el espectro político ha acogido a diferentes nuevos partidos, algunos de ideologías notablemente alejadas. Los estudios académicos se han interesado por esta nueva realidad y han tratado de observarla sirviéndose de las redes sociales como fuente de datos a gran escala.

Este trabajo se suma a la línea de investigación que busca entender si esa expresión pública de opiniones responde a un proceso de reflexión previo y muestra un ejercicio de argumentación por parte del individuo, o si es, más bien, una demostración emotiva de apoyo o desaprobación a una postura política o ideología, sin otro objetivo que imponer su visión, frustrando la posibilidad de debatir para llegar a acuerdos o conocer posturas diferentes a la propia. En otras palabras, esta investigación pretende estudiar la posibilidad de medir la emotividad de los comentarios en redes sociales para esclarecer si estamos ante una esfera pública digital fructífera, o nos encontramos ante un escenario de «buus y hurras».

Para ello, creamos un modelo de aprendizaje automático para la clasificación de texto en función de una serie de parámetros lingüísticos que proporcionan evidencias de la emotividad

de dicho texto. El objetivo es distinguir entre mensajes emotivos y aquellos que no muestran emotividad, para poder ahondar después en el estudio de los primeros.

En mayo de 2021, en el marco de una pandemia global desatada en 2020 por la Covid-19, envuelta en un contexto político inestable que genera alianzas de gobierno constantemente amenazadas por la crispación social, se convocan elecciones a la Asamblea de Madrid. Por un lado, el gobierno de la Comunidad de Madrid protagoniza enconados enfrentamientos con el gobierno central motivados, principalmente, por la gestión de la crisis sanitaria, además de desencuentros con sus socios de gobierno. Por otro lado, tienen lugar una sucesión de mociones de censura que ponen en peligro los acuerdos de gobierno vigentes en varias comunidades autónomas, incluida Madrid. Esta situación desemboca en la convocatoria de las mencionadas elecciones, dando lugar a una campaña electoral polarizada, lo cual podría ilustrarse con el eslogan del partido que, finalmente, obtuvo más votos, el Partido Popular: «comunismo o libertad»¹.

Este escenario presenta una oportunidad óptima para el estudio de la polarización en redes sociales y la emotividad de los comentarios de carácter político en el entorno digital. Asimismo, consideramos interesante hacerlo obteniendo los datos de YouTube, que, a pesar de ser la segunda plataforma más usada a nivel global² (por detrás de Facebook y por delante de Whatsapp, Instagram o Twitter), es la red social menos estudiada en este campo.

¹ Este artículo de *El País* muestra un estudio sobre tendencias y polarización en Twitter, realizado con técnicas de PLN y análisis del sentimiento: Andriano, B. & Pérez Colomé, J. (27 abril 2021). La polarización perfecta: así es la batalla de Madrid en Twitter. *El País*. <https://elpais.com/tecnologia/2021-04-27/la-polarizacion-perfecta-asi-es-la-batalla-de-madrid-en-twitter.html>.

² Reuters Institute for the Study of Journalism. (15 junio 2020). Global active usage penetration of leading social networks as of February 2020. *Statista*. Consultado el 29 de julio de 2021, en <https://www.statista.com/statistics/274773/global-penetration-of-selected-social-media-sites/>.

2. Objetivos

Actualmente, el término «redes sociales» comprende un amplio espectro de sitios y servicios de internet que permiten a los usuarios crear contenido e interactuar con otros usuarios. Estos entornos han supuesto un nuevo campo de estudio tanto para las ciencias sociales y, en concreto, de la comunicación, así como para el campo del Procesamiento del Lenguaje Natural (PLN). Este último se ha expandido notablemente en la última década con el objetivo de estudiar diferentes aspectos de la esfera pública, digital y política. Asimismo, aumenta el interés en los análisis de estos entornos con objetivos comerciales. En ese sentido, resulta crucial entender el papel del PLN en el campo de las redes sociales (Louis, 2016).

Cuando se habla de la polarización de la opinión pública, consciente o inconscientemente, se está haciendo referencia a estudios que conllevan, en muchas ocasiones, la implementación de PLN. Dichos estudios se han centrado, principalmente, en Twitter y Facebook puesto que ambas plataformas fueron creadas originalmente como redes sociales y son las más mayoritarias (Serrano-Contreras, 2020). Sin embargo, YouTube, que comenzó como un repositorio de video, ha ido adoptando los rasgos característicos del resto de plataformas (Ma et al., 2013) y se ha convertido en una de las más concurridas. En España es ya la tercera red social con más usuarios, según los datos de Statista³.

Como se señala en la Introducción, en mayo de 2021, en un momento de inestabilidad política y social enmarcado en la pandemia global desatada por la Covid-19 en 2020, se convocan elecciones a la Asamblea de la Comunidad de Madrid. La campaña que precede al plebiscito se describe frecuentemente en los medios de comunicación como «polarizada».

2.1. Objetivos generales.

Recogiendo esta información, y siguiendo la línea de investigación que se ocupa de estudiar la emotividad en redes sociales para profundizar en el conocimiento de la esfera pública, este trabajo trata de responder a varias preguntas:

- ¿Es posible medir la emotividad en los comentarios de YouTube?
- ¿Se pueden distinguir los comentarios emotivos de los no emotivos si observamos aquellos publicados en vídeos de la campaña de las elecciones a la Asamblea de Madrid de 2021?

³ Social networks: YouTube in Spain 2021 Brand Report. (mayo 2021). Statista Global Consumer Survey. Statista. Consultado el 29 de julio de 2021, en <https://www.statista.com/study/73553/social-networks-youtube-in-spain-brand-report/>.

2.2. Objetivos específicos.

Para poder responder a estas preguntas es necesario establecer subobjetivos que nos permitan construir un modelo para identificar dichos comentarios como emotivos o no emotivos, en base a una serie de características predefinidas:

- Crear un corpus de comentarios extraídos de vídeos de YouTube directamente relacionados con la mencionada campaña electoral.
- Establecer las características que definan los comentarios como emotivos y no emotivos.
- Seleccionar rasgos lingüísticos que identifiquen la presencia o ausencia de argumentación razonada.
- Procesar y clasificar el corpus mediante técnicas de PLN y aprendizaje automático.

Con los resultados, se espera confirmar que los rasgos seleccionados sirven para la tipificación de la emotividad de los comentarios, lo cual supone que el modelo se podrá emplear en herramientas técnicamente más complejas para avanzar en el estudio e identificación de lo que se ha denominado «buus y hurras», concepto que se explica tanto en el Estado de la cuestión (§ESTADO DE LA CUESTIÓN), como en la Metodología (§METODOLOGÍA).

Este trabajo se adentra en el estudio y procesamiento de los comentarios de YouTube, la plataforma menos explorada en el campo de la investigación académica en torno a redes sociales, aportando nuevo conocimiento al respecto. Asimismo, en esta investigación se emplean por primera vez los verbos de opinión y argumentación como parámetro para medir la emotividad de un texto en un modelo de clasificación automática.

El modelo se desarrolla con lenguaje de programación y bibliotecas Python, y está disponible en el repositorio GitHub⁴ para su aprovechamiento en futuras investigaciones y para ser empleado en el procesamiento de comentarios publicados en redes sociales.

⁴ <https://github.com/LysMaDu/ModeloDeClasificacionDeComentarios>

3. Estado de la cuestión

Teniendo en cuenta el carácter multidisciplinar de esta investigación, se ha optado por dividir esta sección en diferentes apartados que permitan presentar un contexto claro de cada disciplina: el contexto sociológico, el contexto lingüístico-computacional y, finalmente, una definición de la plataforma YouTube.

3.1. Las redes sociales y la comunicación política: la esfera pública, digital y política.

3.2. La polarización.

3.3. El civismo y las buenas maneras.

3.4. El emotivismo: en busca de «buus y hurras».

3.5. El Procesamiento del Lenguaje Natural y la Semántica Computacional.

3.6. El Aprendizaje Automático.

3.7. YouTube.

3.1. Las redes sociales y la comunicación política: la esfera pública, digital y política.

El impacto que las tecnologías digitales pueden suponer sobre la esfera pública y política ha sido un tema central en el debate académico desde la aparición de las primeras. El potencial de los medios digitales para restringir o favorecer la interacción y participación pública ciudadana genera múltiples reacciones, unas veces optimistas y otras más escépticas que atienden a posibles intereses que influyen en el funcionamiento de ese entorno (Papacharissi, 2008). Habermas describió la esfera pública como un espacio de debate racional donde los actores se aproximan para intercambiar argumentos por el bien común, al margen del estado y la economía (Habermas, 1974). Por tanto, el concepto habermasiano de este lugar de encuentro implica una argumentación elaborada de las opiniones vertidas en él para que prospere.

Sin embargo, pensadores como Lyotard o Derrida disientían en categorizar la racionalidad como condición *sine qua non* para establecer una esfera pública democrática, y afirmaban que son la individualidad de opiniones y el desacuerdo lo que ha llevado y lleva a avanzar en la deliberación pública (Papacharissi, 2008). Mouffe incluso llega a reformular la esfera pública como un espacio agónico plural que engloba las múltiples percepciones de la realidad política, y afirma que la indecisión define y da forma a la situación actual (Mouffe, 2000, 2005).

Si bien es cierto que las sociedades no son comunidades en armonía, podemos afirmar que mantienen cierta estabilidad nacida del debate público y sustentada en él. Las sociedades son estructuras contradictorias que han surgido de la negociación durante conflictos y, por tanto, su estabilidad no es otra cosa que acuerdos temporales sujetos a la transformación constante

(Castells, 2009). Dichos acuerdos surgen de «la capacidad de la sociedad civil para proporcionar contenido a la acción estatal a través de la esfera pública», garantizando la democracia y el ejercicio legítimo del poder, entendiendo «el poder como representación de los valores e intereses de los ciudadanos, expresados mediante su debate en la esfera pública» (Castells, 2009, pág. 36).

Las sociedades se conforman fundamentalmente como redes y la comunicación funciona de forma análoga (Benkler, 2015; Castells, 2009). Las redes se componen de nodos interconectados, entre los cuales fluye la información. La revolución tecnológica que tuvo lugar en las décadas de 1950 y 1960 propició la transformación de las tecnologías de la información y la comunicación (Benkler, 2015), lo cual sentó las bases del nuevo paradigma tecnológico: la Era de la Información (Castells, 2009). En este nuevo escenario, la esfera pública se trasladó al entorno digital, generando un proceso de desintermediación (Benkler & Nissenbaum, 2006), es decir, la información pasó de estar en manos de los medios de comunicación y las instituciones, a estar disponible e incluso a ser generada por cualquier persona con acceso a internet (Robles et al., 2019). Al mismo tiempo, la red de redes ha expandido las fronteras, conectando a personas de cualquier parte del mundo, globalizando así la esfera pública (Castells, 2009).

Por su parte, la investigación académica se ha centrado en estudiar si esta tecnología ofrece una oportunidad sin precedentes para que los ciudadanos participen de forma directa y activa en el debate público, si este entorno acerca a los distintos actores (representantes y representados) generando organizaciones y espacios más inclusivos.

Por un lado, los defensores del poder democratizador del ciberespacio coinciden en que internet tiene el potencial inmanente de revivir la esfera pública por el mero hecho de facilitar el acceso a la misma a ciudadanos muy diversos, así como de ponerlos en contacto (Papacharissi, 2004), fomentando una sociedad red global (Castells, 2009). Los medios en línea favorecen conversaciones que trascienden límites geográficos y permiten el anonimato, lo cual puede alentar a opinar públicamente de forma desinhibida (Papacharissi, 2008, pág. 10).

Por otro lado, para los escépticos de internet, características como el anonimato o la falta de acceso a red de un porcentaje importante de la población mundial son, precisamente, las que demuestran la desigualdad en la esfera digital. Como consecuencia, no ven este espacio como la base para construir una utopía democrática (Papacharissi, 2008). Además, el acceso por sí mismo no asegura un aumento en la participación política, ni mejora la confianza en los políticos (Bimber, 2001; Kaid, 2002).

En este sentido, cabe destacar los estudios llevados a cabo por las académicas Wojcieszak, Baek y Delli Carpini (2009; Baek, et al., 2012), pues procuran arrojar luz sobre la diferencia entre entornos de debate en línea y «cara a cara». La conclusión es que, aunque la diferencia no es

muy acusada entre ambas modalidades, los espacios en línea suelen sobrerrepresentar a los usuarios con las características de hombre, blanco y joven, generan más emociones negativas y es menos probable que resulten en consenso y acción política. Asimismo, Wojcieszak y Mutz (2009) concluyen que mientras los espacios dedicados específicamente a política refuerzan la tendencia de los usuarios a la confrontación con escaso margen para el debate y la deliberación, los espacios que no están dedicados a la política, pero en los que a veces se expresan opiniones políticas surgidas al hilo de la conversación, propician el intercambio puntual de argumentos que no suele llevar a la confrontación, sino que sirven al propósito de exponer e intercambiar diferentes puntos de vista.

3.2. La polarización.

Relacionado con la confrontación, otro eje central de la investigación académica es la polarización. Los estudios se han enfocado en este fenómeno debido a que cada vez más los ciudadanos acuden a las redes sociales para informarse y debatir sobre temas políticos con el riesgo de producir lo que se ha denominado *echo-chamber* (Bail et al., 2018), espacios en los cuales los ciudadanos que participan del debate están expuestos principalmente a ideas afines a las suyas (Adamic & Glance, 2005; Colleoni et al., 2014; Conover et al., 2012). Sin embargo, cuando interactúan individuos con ideas opuestas, este intercambio se produce «mediante estrategias negativas para la comunicación política como la falta de civismo y el uso de lenguaje ofensivo» (Robles et al., 2019, pág. 194). La conjunción de estos hechos nos indica que estamos ante un proceso de polarización.

Sobre esta cuestión también hay voces que afirman que las redes sociales no solo no propician la polarización, sino que la reducen. Barberá (2014) afirma que, como demuestra su estudio, las redes sociales facilitan la exposición a ideas distintas a las del individuo, mostrándole información a la que, de otra forma o incluso en la vida fuera de la red, no tendría acceso. Puesto que, en estos entornos digitales, los individuos tienden a estar en contacto con personas con las que tienen vínculos débiles y que son políticamente heterogéneas, y por tanto distintas de su entorno personal, esa exposición reduce el extremismo político del individuo en cuestión (Barberá, 2014).

No obstante, existe una postura equidistante entre la visión positiva y la más crítica con los medios en línea y las redes sociales. Esta reconoce que hay evidencia de que se pueden producir *echo-chambers*, pero también ha observado que, dependiendo del momento, el tema e incluso del entorno (Facebook, Twitter, YouTube), el comportamiento cambia. Unas veces se muestra más polarizado y otras es permeable al debate razonado o, al menos, al intercambio de opiniones no afines desde el civismo (Serrano-Contreras et al., 2020). El equilibrio entre las

visiones utópicas y distópicas revela la verdadera situación de internet como esfera pública (Papacharissi, 2002).

3.3. El civismo y las buenas maneras.

Por tanto, ¿podría ser el civismo la clave para una comunicación política fructífera en la esfera pública y virtual? El civismo, más conocido en la literatura por el término anglosajón *civility*, se ha considerado siempre un requisito indispensable para el discurso democrático y se ha definido tradicionalmente como «buena educación» y cortesía. Papacharissi (2004) revisa este concepto y lo redefine para distinguir entre *civility* y *politeness* (podríamos traducirlo como civismo y buenas maneras). Esta distinción pretende discriminar actitudes que muestran falta de «etiqueta», de la falta de respeto por los valores democráticos, con el objetivo de, por un lado, definir más cuidadosamente el concepto de *civility* y, por otro lado, medir el potencial democrático del discurso político presente en internet.

En su investigación, Papacharissi observa los mensajes de un foro de temática política y concluye que la mayoría de esos mensajes no son predominantemente *impolite* o *uncivil*, pero sí poco concretos y difusos. También observa que aquellos mensajes o individuos de corte incívico, no lo son contra otros usuarios directamente, sino más bien hacia líderes políticos y grupos étnicos. Y, finalmente, la *impoliteness* suele mostrarse de forma encubierta, a través de rasgos como el sarcasmo, las mayúsculas (el equivalente a gritar en las «normas no escritas» de la comunicación en internet), y otras muestras de agresividad. No obstante, la mayoría de las discusiones son calmadas y mantienen un tono sosegado (Papacharissi, 2004).

3.4. El emotivismo: en busca de «buus y hurras».

Estudios más recientes apuntan a la ausencia de comunicación cara a cara como la principal causa de la falta de cortesía, lo cual propicia interacciones irracionales, emocionales e irrespetuosas (Schäfer, 2016; Serrano-Contreras et al., 2020). Esto genera polarización, alejando las posturas que se vierten por influencia de las emociones y convicciones, obviando la evidencia y los argumentos (Mason, 2014; Olsson, 2013).

Sin embargo, para poder moderar conversaciones en línea, tratando de evitar la falta de civismo y sus consecuencias, como puede ser la deshumanización de grupos sociales concretos o el aumento de la polarización, habría que idear nuevas estrategias que identifiquen dichas conductas. Ciertamente, aquello que en otras épocas se desaprobaba socialmente o se consideraba una ofensa pública, hoy día puede haber cambiado; incluso, dependiendo del entorno, una afirmación puede considerarse incívica, mientras en otro ambiente no lo es tanto (Chen et al., 2019).

Así pues, diversos estudios han tratado de medir la emotividad⁵ (Ayer, 1952; Castells, 2009) del lenguaje empleado en internet. El Emotivismo enunciado por Ayer (1952), y desarrollado por Stevenson (1960), entiende que los juicios morales no funcionan como declaraciones de hechos, sino más bien como una expresión de los sentimientos del individuo que los exterioriza, ya sea de forma escrita o hablada. Desde esta postura, se ha estudiado la emotividad del lenguaje con métodos computacionales, que han permitido reconocer información relacionada con los sentimientos en el texto (i.e.: Liu et al., 2002; Picard, 1997; Reynolds, 2006).

Concretamente, Reynolds (2006) se propone identificar comentarios que más que deliberar sobre un tema político o mostrar una serie de argumentos, alaban o condenan el mensaje original. Estos comentarios se denominan «buus y hurras» (*boo* y *hooray*). Para identificarlos, Reynolds se basa en estudios previos que han aislado una lista de *emotional words* (Cowie et al., 2001; Whissel, 1989), las cuales permiten procesar el texto y distinguir entre dos categorías: «bueno» y «malo», en función de la presencia de las palabras de la lista en el texto seleccionado y extraído de internet. A pesar de que los resultados de este estudio no son muy satisfactorios por ser un tanto difusos, sí abre una vía de investigación que se verá facilitada por el avance posterior de la programación.

3.5. El Procesamiento del Lenguaje Natural y la Semántica Computacional.

El método computacional para tratar con lenguaje se denomina Procesamiento del Lenguaje Natural (PLN). Una definición asequible es la de Chowdhury (2003), quien explica que esta área de investigación explora cómo las computadoras pueden comprender y manipular el lenguaje natural. Turing (1950) establece las bases de esta disciplina de convergencia entre lingüística e ingeniería creando el test de Turing. Uno de los objetivos de sus sucesores, desde la semántica computacional, es construir modelos que procesen lenguaje como lo hacen los humanos (Hershcovich & Donatelli, 2021; Linzen, 2020). También, la teoría de la gramática generativa de Noam Chomsky⁶ es crucial en el desarrollo de esta disciplina, pues establece un sistema de reglas de precisión matemática para la descripción de la lengua (Schaub, 2020).

La complejidad del lenguaje natural obliga a que disciplinas como la informática, la lingüística, las ciencias de la información y la psicología, entre otras, se combinen para poder comprender la comunicación humana y tratar de procesar el texto, escrito o hablado, mediante las máquinas (Verdejo Maillo, 1994).

⁵ Para entender la perspectiva emotivista es interesante familiarizarse con el trabajo de los neurocientíficos Damasio y Meyers (2008), cuya referencia se encuentra en la bibliografía secundaria de este documento.

⁶ Chomsky, 1966. Referencia incluida en la bibliografía secundaria de este documento.

Chowdhury (2003) recoge los niveles interdependientes identificados por Liddy y Feldman⁷ como necesarios para entender el lenguaje natural y que se utilizan para extraer significado de los textos:

- fonético y fonológico (pronunciación);
- morfológico (morfemas, sufijos y prefijos);
- léxico (significado léxico y sus funciones sintácticas);
- sintáctico (gramática y estructura de la oración);
- semántico (significado);
- discurso (análisis del discurso);
- pragmático (influencia del contexto externo en la interpretación del significado).

Este último es el más problemático para esta disciplina, por la dificultad para aislar las distintas capas de significado que puede tener un texto (Hershcovich & Donatelli, 2021). Tanto es así que algunos académicos, en la actualidad, llegan incluso a preguntarse si la investigación en torno al *Natural Language Understanding* es, en sus palabras, «climbing the right hill». Argumentan que un sistema entrenado sobre la forma no tiene, *a priori*, la capacidad de aprender sobre significado (Bender & Koller, 2020).

3.6. El Aprendizaje Automático.

Como pasó con la información, en la era del *big data*, hemos pasado de que solo las empresas tuvieran y almacenaran datos, a que todos nosotros, desde nuestros dispositivos tecnológicos (ordenadores, tabletas, teléfonos móviles, etc.), seamos productores y consumidores de datos.

Every time we buy a product, every time we rent a movie, visit a web page, write a blog, or post on the social media, even when we just walk or drive around, we are generating data. (Alpaydin, 2014, pág. 1)

Para gestionar esta cantidad abrumadora de información necesitamos algoritmos. Un algoritmo es una secuencia de instrucciones que deben aplicarse para transformar un *input* en un *output* (Alpaydin, 2014, pág. 2), es decir, información de entrada, incluso una «petición» o «solicitud», transformada en una respuesta o resultado. Para algunas tareas, como predecir, no existe un algoritmo predeterminado, por tanto, nos faltan las instrucciones. En ese caso, el proceso a seguir es entrenar a la máquina con los datos que sí tenemos para que aprenda qué parámetros identifican los elementos que le estamos proporcionando y, con esa información, genere un sistema de etiquetado o discriminación (modelo) para los elementos nuevos. Con este

⁷ Liddy, 1998; Feldman, 1999. Ambas referencias incluidas en la bibliografía secundaria de este documento.

sistema puede que no se consiga conocer el proceso completamente, pero se considera que se construye y se obtiene una aproximación suficientemente buena y útil para trabajar con los datos (Alpaydin, 2014). Esto, simplificando notablemente, es lo que se conoce como *machine learning* o aprendizaje automático, que es, a su vez, una rama de la inteligencia artificial.

Hoy en día, el aprendizaje automático se emplea en multitud de campos, como son la minería de datos (procesos con bases de datos), la robótica o el reconocimiento de voz (*speech recognition*). En definitiva, el *machine learning* consiste en definir un modelo de programación con una serie de parámetros que la máquina debe optimizar a través del aprendizaje mediante los datos modelo o las experiencias pasadas. Dicho modelo puede ser predictivo o descriptivo, en función de si buscamos predecir el futuro u obtener más información de los datos, o ambos (Alpaydin, 2014, pág. 3).

3.7. YouTube.

Como veíamos al inicio, la esfera pública se ha transformado, pasando del modelo jerárquico y unidireccional de los medios de comunicación de masas, a un modelo estructurado como una red de comunicación multidireccional y horizontal (López García, 2006). En los últimos años, gran parte de la investigación sobre comunicación política se ha centrado en Twitter y Facebook⁸, lo cual, según Serrano-Contreras (2020), se debe a dos motivos principales: en primer lugar, ambas son las redes sociales más populares o con mayor número de usuarios y, en segundo lugar, porque resulta más sencillo extraer datos de ellas. Sin embargo, las estadísticas apuntan a YouTube como la segunda red social más popular a nivel global⁹.

Cabe destacar que tanto Twitter como Facebook fueron diseñadas originalmente para interconectar a los usuarios. YouTube, en contraste, se concibió como un repositorio de video, pero ha ido adquiriendo características propias de las redes sociales, como pueden ser el sistema de recomendación, *likes* y *dislikes*, menciones a otros usuarios, etc., lo cual nos permite considerar a la plataforma como *social network* en la actualidad (Ma et al., 2013).

⁸ Conover et al., 2011; Grusell & Nord, 2012; Gruzd & Roy, 2014; Jaidka et al., 2019; Marichal, 2016; Oz et al., 2018. Referencias extraídas de Serrano-Contreras, 2020, e incluidas en la bibliografía secundaria en este documento.

⁹ Reuters Institute for the Study of Journalism. (15 junio 2020). Global active usage penetration of leading social networks as of February 2020. *Statista*. Consultado el 29 de julio de 2021, en <https://www.statista.com/statistics/274773/global-penetration-of-selected-social-media-sites/>.

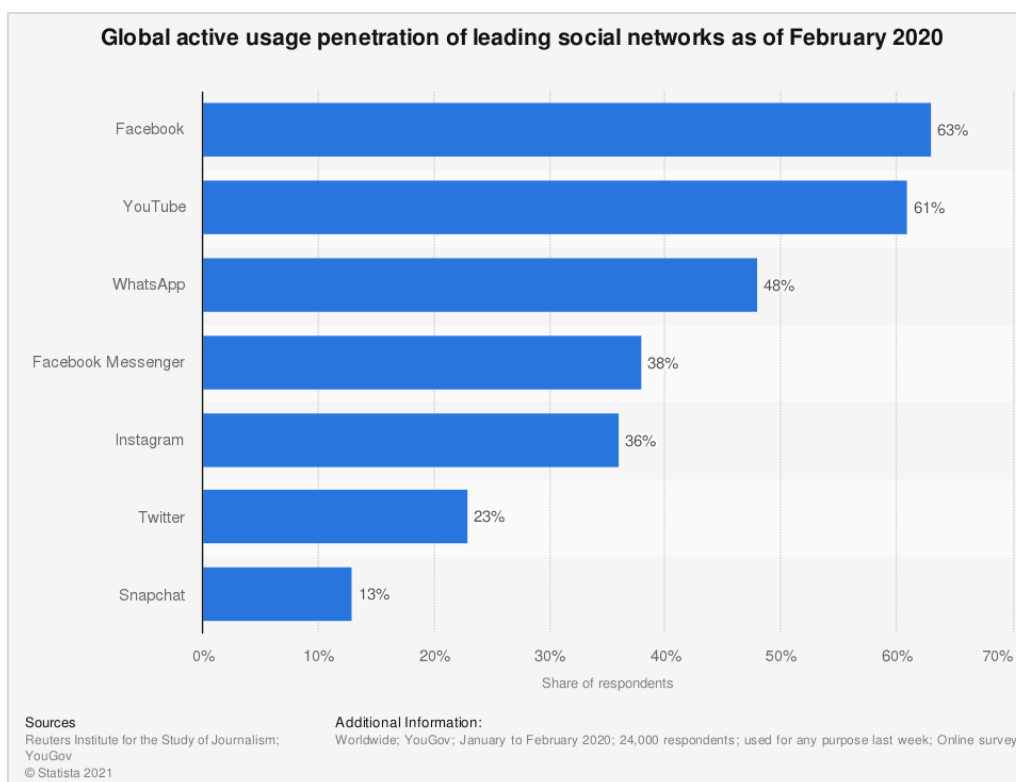


Figura 1. Gráfico de las redes sociales más populares en 2020.

YouTube fue lanzado en 2005 y adquirido por Google un año después. En la actualidad, se estima que tiene 2.100 millones de usuarios en todo el mundo; en agosto de 2020, se subieron más de 500 horas de vídeo por minuto, lo que refleja el auge en la demanda de este formato de contenido por parte de los usuarios de internet. Uno de sus mayores recursos es el contenido oficial corporativo, como son los vídeos musicales, los cuales pueden hacerse virales en cuestión de horas y alcanzar los 100 millones de visitas en menos de dos días¹⁰. En España, YouTube es la tercera red social con más usuarios, por detrás de Facebook e Instagram. Los usuarios de esta plataforma tienden a comentar las publicaciones más que el usuario de internet medio, y el 92% de los usuarios suele visitar la plataforma desde su *smartphone*¹¹.

¹⁰ YouTube - Statistics & Facts. (12 julio 2021). Statista Research Department. *Statista*. Consultado el 29 de julio de 2021, en <https://www.statista.com/topics/2019/youtube/>.

¹¹ Social networks: YouTube in Spain 2021 Brand Report. (mayo 2021). Statista Global Consumer Survey. *Statista*. Consultado el 29 de julio de 2021, en <https://www.statista.com/study/73553/social-networks-youtube-in-spain-brand-report/>.

4. Metodología

En esta sección se explica el proceso de trabajo seguido en esta investigación, basado en los siguientes pasos:

- 4.1.** Creación de corpus de comentarios: descarga y almacenamiento desde YouTube.
- 4.2.** Análisis preliminar de las características de los datos descargados (corpus).
- 4.3.** Definición de «buus y hurras» y parámetros lingüísticos para identificarlos.
- 4.4.** Limpieza y procesamiento de los comentarios.
- 4.5.** Etiquetado manual y creación de corpus de comentarios etiquetados.
- 4.6.** Entrenamiento y aplicación del clasificador automático.
- 4.7.** Evaluación mediante matriz de confusión.

4.1. Creación de corpus de comentarios: descarga y almacenamiento desde YouTube.

El lenguaje de programación utilizado ha sido Python y el entorno escogido para el desarrollo ha sido Jupyter Notebook, por facilitar el control sobre las partes de código que se ejecutan, agilizando así las múltiples pruebas que se llevan a cabo durante la construcción de un programa.

Primeramente, es necesario definir qué vídeos interesan para la investigación. Como se expone en los objetivos (§OBJETIVOS), el marco escogido para este estudio es la convocatoria de elecciones a la Asamblea de Madrid en 2021. Este escenario ilustra la situación de crispación y polarización actuales, lo cual, cabe asumir, se refleja en los comentarios publicados por los ciudadanos en los vídeos de campaña de los principales partidos. Puesto que esta investigación comienza antes de la celebración del plebiscito, se tiene en cuenta la representación parlamentaria en la cámara en la legislatura XI. Los partidos con representación en ese momento son: PSOE, PP, Ciudadanos, Más Madrid, Vox y Unidas Podemos¹².

Para ser elegibles, los vídeos deben ser para la campaña electoral específicamente y estar colgados en el canal oficial de YouTube de los partidos. Además, la selección se acota a los dos vídeos más vistos y comentados de cada partido. Tras una primera búsqueda, observamos que hay una diferencia considerable entre las visualizaciones y el número de comentarios publicados en los vídeos de, por un lado, Unidas Podemos y Vox y, por otro, el resto de partidos.

¹² Datos extraídos de la página web oficial de la Asamblea de Madrid (5 agosto 2021)
<https://www.asambleamadrid.es/la-asamblea/historia/legislatura-xi/grupos-parlamentarios>.

Canal	Playlist/Título del video	Playlist/video ID	Visualizaciones	Comentarios
Video más visto/más comentado de cada partido (27/04/21)				
PP	Pablo Casado e Isabel Díaz Ayuso interviene	TihucvwpOio	46.192	7
PP	Isabel Díaz Ayuso participa en un Acto de Ca	_UGG4mF-d2c	21.480	19
Vox España	Rocío Monasterio desmonta a la activista pr	yuO06O4L1tg	345.961	7952
Vox	Discurso completo de Santiago Abascal en e	yNq1qN8aCPM	100.081	846
Mas Madrid	Arrancamos la campaña en Vallecas! #ElMac	vFuGZlZV8TY	4.980	49
Mas Madrid	Cierre de campaña 🗳️ #VotaMásMadrid4M	hm6ecgMmh4w	4763	127
Podemos	Tú decides Un corto de Daniel Guzmán.	AhgmHGdx7ZM	238.082	6498
Podemos	Entrevista a Pablo Iglesias en Telemadrid	Csqchc_EMwQ	66.127	1468
PSOE-M	CAMPAÑA 2021 PSOE-M HAZLOXMADRID 02	rY5GRbjxVew	114.882	31
PSOE-M	CAMPAÑA 2021 PSOE-M #HAZLOXMADRID 0	Gt-vpfjeaE8	105.529	3
PSOE-M	Ángel Gabilondo, candidato a la presidencia	CszwJJMWNWc	23.082	71
Cs Madrid	Elige Centro	g0tReXDhb7E	50	0
Cs	Bal, Edmundo Bal	n3ndkApOHTs	3.346	172
Cs	Acto de Cierre de Campaña con Inés Arrima	9XZ7yfd9dL0	4.245	222

Figura 2. Datos de los vídeos de los principales partidos de la Asamblea de Madrid.

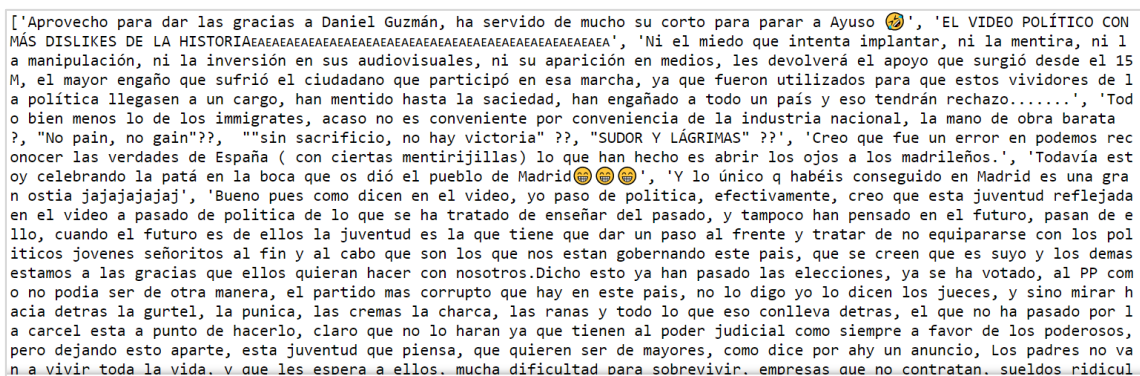
Ante estos datos, decidimos estudiar el vídeo más visto y comentado de Vox (7.952 comentarios) y el de Podemos (6.498 comentarios), por mostrar números comparables entre ellos. Por tanto, el corpus se compone de 14.450 muestras. Es importante tener en cuenta que estas cifras corresponden a la fecha en la que se recogen los datos y pueden sufrir leves modificaciones si se emplean datos recogidos en una fecha diferente, pues el número de comentarios puede variar con el paso del tiempo, ya que el vídeo sigue disponible en la plataforma.

A modo de contexto, es interesante conocer el contenido de los vídeos para, después, trabajar con los comentarios. El vídeo de Vox muestra parte de una entrevista a la candidata del partido, Rocío Monasterio, conducida por la ahora presentadora de TVE Mónica López, quien ha sido presentadora del tiempo en las noticias de esta cadena durante años. El vídeo de Podemos es un corto dirigido por Daniel Guzmán, que presenta a dos jóvenes intercambiando opiniones en un entorno de clase obrera.

Para interactuar con YouTube hacemos uso de la YouTube Data API v3, en concreto empleamos la biblioteca `youtube-data-api`, que facilita la comunicación con dicha plataforma, a través de Python. El primer paso es darse de alta como desarrollador y obtener una clave (`api_key`) que otorga Google Cloud Console. Con esta credencial, la plataforma te permite realizar peticiones y descargar información pública de la misma. Una vez asignada dicha clave, esta se utiliza para hacer una petición a través de la API para descargar los comentarios del vídeo mediante su código de identificación, `video_id`, y el método `get_video_comments`. Después, se almacenan los comentarios creando una función que los archiva en formato JSON.

4.2. Análisis preliminar de las características de los datos descargados (corpus).

Para poder establecer las necesidades de procesamiento del texto, es imprescindible realizar un análisis previo del corpus recopilado. En este caso, obtenemos una lista de cadenas de texto con todos los comentarios archivados, como muestra la figura 3, y realizamos búsquedas teniendo en cuenta los parámetros habituales que se aplican en la limpieza de un texto, como pueden ser errores habituales de escritura o abreviaturas típicas de la escritura informal, por ejemplo: «xq» o «pq» en lugar de «porque/por qué».



['Aprovecho para dar las gracias a Daniel Guzmán, ha servido de mucho su corto para parar a Ayuso 🙄', 'EL VIDEO POLÍTICO CON MÁS DISLIKES DE LA HISTORIA', 'Ni el miedo que intenta implantar, ni la mentira, ni la manipulación, ni la inversión en sus audiovisuales, ni su aparición en medios, les devolverá el apoyo que surgió desde el 15 M, el mayor engaño que sufrió el ciudadano que participó en esa marcha, ya que fueron utilizados para que estos vividores de la política llegasen a un cargo, han mentido hasta la saciedad, han engañado a todo un país y eso tendrán rechazo.....', 'Todo bien menos lo de los inmigrantes, acaso no es conveniente por conveniencia de la industria nacional, la mano de obra barata?', 'No pain, no gain?', '"sin sacrificio, no hay victoria" ??', 'SUDOR Y LÁGRIMAS" ??', 'Creo que fue un error en podemos reconocer las verdades de España (con ciertas mentirijillas) lo que han hecho es abrir los ojos a los madrileños.', 'Todavía estoy celebrando la patá en la boca que os dió el pueblo de Madrid 🤔🤔', 'Y lo único q habéis conseguido en Madrid es una gran ostia jajajajajaj', 'Bueno pues como dicen en el video, yo paso de política, efectivamente, creo que esta juventud reflejada en el video a pasado de política de lo que se ha tratado de enseñar del pasado, y tampoco han pensado en el futuro, pasan de ello, cuando el futuro es de ellos la juventud es la que tiene que dar un paso al frente y tratar de no equipararse con los políticos jóvenes señoritos al fin y al cabo que son los que nos estan gobernando este país, que se creen que es suyo y los demas estamos a las gracias que ellos quieran hacer con nosotros.Dicho esto ya han pasado las elecciones, ya se ha votado, al PP como no podia ser de otra manera, el partido mas corrupto que hay en este país, no lo digo yo lo dicen los jueces, y sino mirar hacia detras la gurtel, la punica, las cremas la charca, las ranas y todo lo que eso conlleva detras, el que no ha pasado por la carcel esta a punto de hacerlo, claro que no lo haran ya que tienen al poder judicial como siempre a favor de los poderosos, pero dejando esto aparte, esta juventud que piensa, que quieren ser de mayores, como dice por ahy un anuncio, Los padres no van a vivir toda la vida, v que les espera a ellos. mucha dificultad para sobrevivir. empresas que no contratan. sueldos ridiculos

Figura 3. Muestra de parte de la lista de comentarios extraídos de los vídeos estudiados.

Tras un trabajo exhaustivo de observación de los comentarios, se identifican ciertas características habituales en la escritura:

- Uso de mayúsculas para expresar énfasis. Ejemplo: 'EL vídeo POLÍTICO CON MÁS DISLIKES DE LA HISTORIA'.
- Uso excesivo de signos de interrogación y admiración para expresar énfasis. Ej.: '31.155 dislikes!!!! Que hable la mayoría!!!!. Podemos=Estafa'.
- Abreviación de las palabras como «porque» o «que», entre otras. Ej.: 'q asco d periodismo'.
- Uso errático de la puntuación. Ej.: 'q cara tienes .tu no defiendes la libertad .nos metería a la carcel .venga mentirosa .eso hay una doble bara de medir . [...]'.
- Presencia moderada de emoticonos (emoji UNICODE). Ej.: 'Chepasssssssss!!!!😂😂😂😂😂'
- Presencia escasa de etiquetas o *hashtags*. Ej.: 'la mayoría esta con vox #mueraelchavismo #mueralarevolucionbolivariana #mueraelcomunismochavista #mueraelcomunismo viva españa es'.
- Menciones escasas a otros usuarios. Ej.: '@Ruben Ibañez te van a borrar el comentario, las verdades le duelen mucho a los fanáticos.'

- Presencia marginal de enlaces externos (*links*). Ej.: 'Jajajjaaj este me gusta más <https://youtu.be/HBxC88kINEI>'.

4.3. Definición de «buus y hurras» y parámetros lingüísticos para identificarlos.

Puesto que es el objetivo principal de la investigación, en este punto del proceso es imprescindible definir qué entendemos por mensaje emotivo para poder identificarlo, o lo que es lo mismo, «buus y hurras».

Reynolds, referente pionero en acuñar este término, lo explica así:

In light of this recent work on computational processing of information dealing with feelings, a sort of “Computational Emotivism” is possible in which statements are analyzed for their content to determine if writers referring to the statement are “praising” or “condemning.” Or, more colloquially, if a group of documents is shouting “boo” or cheering “hooray” with regard to a statement. (Reynolds, 2006, pág. 2).

En otras palabras, son expresiones que emplea el usuario/escritor para manifestarse a favor o en contra de una postura o mensaje. Habitualmente tienen forma imperativa y son meras expresiones de un sentimiento que carecen de una reflexión razonada previa ('Pues la mayoría ha hablado. No se pudo. Hala, a pastar todos.', 'Asco de sectaria la Monica López.pero rocio tiene muchos ovarios por mucho que la ataquen , 🤔🤔🤔🤔 Rocio ❤️❤️').

A la luz de esta definición, establecemos dos rasgos lingüísticos principales a la hora de identificar la emotividad en un comentario: conectores textuales y marcadores discursivos de relación lógica y argumentativa (López, 2006) y verbos que indican opinión o argumentación. Estos elementos dan cuenta de la presencia de argumentación en el texto, por lo que, en cierta medida, actúan como evidencias lingüísticas de que el autor del comentario ha pasado por un proceso de reflexión previo al acto de escribir dicho comentario.

Así pues, los conectores y marcadores seleccionados son los siguientes (Martín Zorraquino & Portóles Lázaro, 1999)¹³: a propósito, ahora bien, de hecho, debido a, en cambio, en cualquier caso, en realidad, en todo caso, es decir, no obstante, o sea, pero, por el contrario, por eso, por lo tanto, por qué, por tanto, porque, sin embargo.

¹³ También se consultaron las páginas web del Centro Virtual Cervantes (30 junio 2021) https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/conectordiscursivo.htm y https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/marcadoresdiscurso.htm.

En el proceso de estudio de este parámetro, consideramos que deben quedar fuera conectores consecutivos o ilativos (pues, entonces) y aditivos (además, incluso, encima), puesto que no necesariamente implican argumentación. Asimismo, suelen formar parte de comentarios sencillos, por tanto, de incluirlos en el listado, podrían generar errores en la identificación de mensajes argumentados.

En el caso de los verbos de opinión y argumentación, decidimos introducir este parámetro como evidencia de razonamiento para fortalecer el proceso de clasificación automática, el cual se explica más adelante. Seleccionamos una serie no muy extensa de verbos que se emplean para argumentar o expresar una opinión mínimamente razonada. Puesto que el objetivo principal es medir la emotividad y, si es posible, distinguir entre los mensajes emotivos de los que no lo son, este parámetro, al igual que el anterior, actúan como evidencia de un proceso de reflexión mínimo por parte del autor del comentario.

El espectro de emotividad es, como cabe esperar, muy amplio. Desde mensajes formados exclusivamente por emoticonos, hasta mensajes argumentados de 1007 palabras (ver figura 4), como hemos encontrado en nuestro corpus. Partiendo de esta realidad y sin olvidar la subjetividad inmanente al lenguaje, se trata de identificar rasgos que pueden constituir evidencia, o al menos indicio, de un comportamiento específico en la expresión escrita. En este caso, aquellos elementos del lenguaje asociados a la argumentación nos sirven para identificar mensajes con emotividad probablemente baja, frente a otros que son puramente emotivos. La lista de verbos seleccionados es la siguiente: considerar, creer, cuestionar, discutir, explicar, imaginar, opinar, pensar, proponer, sospechar, sostener, suponer.

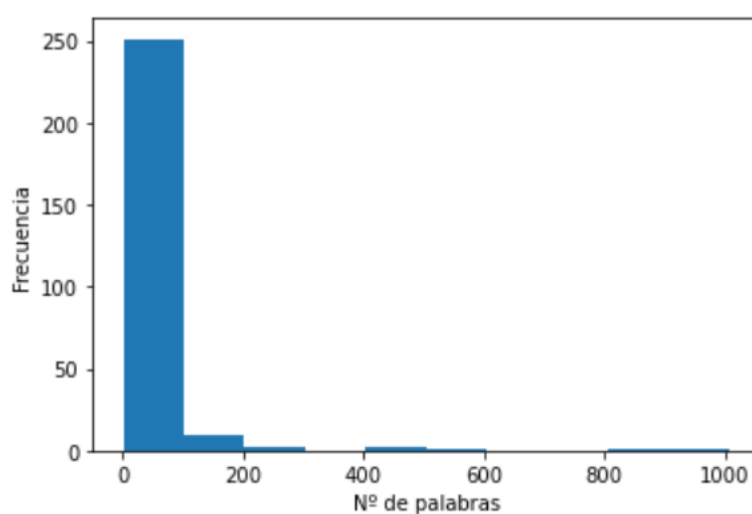


Figura 4. Histograma de frecuencia y número de palabras por comentario.

Para ilustrar esta aparente dicotomía, tomemos dos ejemplos. Por un lado, un mensaje claramente emotivo, *a priori*, no contaría ni con conectores, ni con verbos de argumentación, es decir, un comentario que no muestra un razonamiento, sino más bien, que expresa una opinión subjetiva con el objetivo de imponer dicha opinión sin aportar argumentos. Estos comentarios cumplen con todo lo anteriormente descrito: 'Isobaras, dedícate a lo tuyo, aquí, no das la talla!'; 'Toma paraguazos, doña borrascas.'; 'Enorme Rocio Monasterio'. Por otro lado, el siguiente comentario es, inequívocamente, razonado: '@News24HoursONLINE Totalmente de acuerdo, es tremendo. Es que no se comprende cómo se puede llegar a ser así ante unas cámaras; tampoco estaría bien en una conversación personas (maltrato) pero ahí en TV1 pagado por los españoles y tratándola así que ni un inspector de policía con un delincuente se comporta así. Como dices, raya lo delictivo'. Objetivamente, cuenta con dos elementos de nuestras listas: el verbo «comprender» y el conector «pero», además de mostrar una clara intención de diálogo, búsqueda de acuerdo y exposición de argumentos.

Además de estos dos rasgos lingüísticos principales, identificamos el uso de interrogaciones y exclamaciones, mayúsculas y emoticonos como elementos a tener en cuenta para medir la emotividad, puesto que se usan para intensificar la opinión que se está mostrando.

En conclusión, los denominados «buus y hurras» son comentarios de alta emotividad que muestran rechazo o apoyan un mensaje o postura expresada por otro (en este caso, el mensaje original es el propio vídeo o un comentario de otro usuario) sin aportar argumentos. La presencia excesiva de emoticonos, signos de interrogación y exclamación y mayúsculas se consideran evidencias de emotividad, mientras que la presencia de verbos y conectores de opinión y argumentación se identifican como muestras de razonamiento y, por tanto, de menor emotividad.

4.4. Limpieza y procesamiento de los comentarios.

Una vez definidos el concepto y el objetivo último de la investigación, identificar comentarios emotivos, el siguiente paso es decidir qué herramientas emplear para conseguirlo. Teniendo en cuenta que tenemos un corpus extenso de comentarios y cinco parámetros a identificar en el texto, la herramienta más eficiente para este estudio es el aprendizaje automático. Esta tecnología nos permite, primero, entrenar un modelo para que aprenda a identificar en el texto los parámetros que le indicamos, para, después y basándose en ellos, clasificar automáticamente los comentarios, acorde a las necesidades de la investigación.

Es el paradigma de «aprendizaje automático» que se ha impuesto en las últimas décadas en Inteligencia Artificial: los algoritmos infieren las posibles respuestas a partir de los datos observados anteriormente en el corpus. (Moreno, 2021)

Puesto que los modelos de aprendizaje automático funcionan con parámetros numéricos, el modo de expresar en una cifra nuestros parámetros lingüísticos es hallando la proporción de estos elementos en cada comentario. Para conseguirlo, sometemos al texto a los procesos de limpieza y análisis que se explican a continuación.

El PLN ofrece las siguientes acciones para acondicionar el texto con el objetivo de extraer información de él (Velez et al., 2016; Robles et al., 2020):

- Normalización: identificar y corregir errores tipográficos, duplicidad de caracteres, espacios, puntuación, etc.
- *Tokenización*: identificación de cada uno de los términos/elementos (*tokens*) que forman la frase o cadena de texto.
- Análisis morfosintáctico (*POS*): asignación de categoría lingüística a cada *token* en función de su rol en la frase.
- Limpieza de *stopwords*: términos vacíos de significado para la clasificación, como pueden ser preposiciones, conjunciones, etc.
- Lematización o *stemming*: procesos similares para extraer o bien el lema, o bien la raíz (lexema) de los términos.

Dentro de la normalización, identificamos la necesidad de eliminar la puntuación, los enlaces externos, las tildes y lo que denominamos gazapos, que son diferentes abreviaturas de las palabras: aunque, con, de, por, porque, por qué, que y también. Esta limpieza se realiza mediante expresiones regulares, utilizando la biblioteca *re* de Python¹⁴. Se descarta tanto eliminar *hashtags*, por su escasa presencia, como menciones a otros usuarios, ya que en YouTube es complicado puesto que, si el nombre del usuario mencionado está formado por más de una palabra y estas aparecen separadas, la tarea se vuelve ineficaz. Es decir, si tenemos un usuario llamado «@Lys mayor», podríamos eliminar la palabra que aparece inmediatamente después de la @, sin miedo a eliminar otra palabra que forme parte del mensaje; sin embargo, nos resulta imposible identificar la palabra «mayor» como parte del nombre de usuario. El siguiente ejemplo es un comentario real del corpus que ilustra la dificultad descrita: '@Ruben Ibañez te van a borrar el comentario, las verdades le duelen mucho a los fanáticos.'

Utilizamos el texto limpio para encontrar algunos de nuestros parámetros lingüísticos, como son los conectores, los verbos y los emoticonos. Para la búsqueda y contabilización de estos elementos en el texto es necesario *tokenizarlo*, por esta razón necesitamos que el texto se haya limpiado previamente. Esta contabilización se realiza mediante una función que,

¹⁴ El código se puede consultar en el repositorio GitHub:
<https://github.com/LysMaDu/ModeloDeClasificacionDeComentarios>

primero, encuentra todas las apariciones de los elementos que incluimos en una lista (conectores, verbos, emojis UNICODE) mediante el método `re.findall`, después, *tokeniza* el texto para obtener el número de *tokens* que tiene, mediante el método `word_tokenize` de la biblioteca NLTK, y, finalmente, divide el total de conectores/verbos/emoticonos entre el número total de *tokens*, obteniendo la proporción de dicho elemento en el comentario.

El proceso para obtener la proporción de verbos argumentativos requiere pasos adicionales. En esta tarea, empleamos las bibliotecas NLTK, spaCy y re. Existen otras herramientas de PLN como IBM Watson, pero en este caso se descartó por no tratarse de *software* libre.

Identificar formas verbales requiere un proceso de derivación regresiva, para lo cual utilizamos el stemmer de NLTK. Si bien es cierto que el lemmatizer de esta misma biblioteca es muy potente y realiza tanto la lematización como el análisis *POS* en una sola acción, este es ineficaz en español. Sin embargo, el stemmer es la herramienta finalmente seleccionada debido a que genera mejores resultados en este idioma, en lo que se refiere a identificar la raíz. Procesamos nuestra lista de verbos con el stemmer para fijar el grupo de raíces que buscaremos en el texto.

El siguiente paso es identificar los verbos de cada comentario. Para ello realizamos el análisis *POS* de todo el comentario con el método `spacy.load('es_core_news_sm')`, que muestra mejores resultados que la herramienta lemmatizer de NLTK, ya que spaCy funciona mejor con español. Una vez etiquetadas las partes del texto, podemos identificar los verbos, aplicar stemmer y comprobar si las raíces de los verbos del comentario coinciden con las de nuestra lista. Contamos aquellas que coinciden y dividimos el número de apariciones entre el número de *tokens* total del comentario, obteniendo la proporción de verbos argumentativos de dicho comentario.

Finalmente, para hallar la proporción de exclamaciones, interrogaciones y mayúsculas, el proceso es similar, pero dicha proporción se calcula respecto al número de caracteres del texto, en lugar de calcularlo conforme al número de *tokens*. Asimismo, estos elementos se observan en los comentarios originales, es decir, que no han pasado por el proceso de limpieza. Primero, se aplica la expresión regular `re.findall` para identificar las apariciones en el comentario del elemento que buscamos y, después, se cuentan los caracteres del comentario mediante el método `len`. Para terminar, se halla la proporción dividiendo la primera cifra entre la segunda.

Una vez diseñados todos los procesos para depurar el texto y para extraer la proporción de los parámetros elegidos para la investigación, podemos proceder a la fase de etiquetado de los comentarios.

4.5. Etiquetado manual y creación de corpus de comentarios etiquetados.

Para poder entrenar el modelo de clasificación, necesitamos extraer una pequeña parte, pero significativa, del corpus y etiquetar los comentarios manualmente, ya que debemos entrenar al modelo alimentándolo con ejemplos del resultado que esperamos. Establecemos cuatro individuos etiquetadores y acordamos unos criterios comunes para llevar a cabo la tarea. El grupo de etiquetadores se caracteriza, en este caso, por su bagaje dispar, pues está formado por un sociólogo, dos informáticos y una filóloga. Ante esta circunstancia decidimos medir la fiabilidad del etiquetado empleando el método Kappa de Fleiss.

Para evaluar el grado de coincidencia entre etiquetadores existen varias herramientas, entre las cuales Kappa de Fleiss se emplea en los casos en que hay más de dos individuos desarrollando el trabajo. Esta función toma como argumento una lista que expresa de forma numérica las etiquetas aplicadas a cada mensaje por los etiquetadores, y evalúa el grado de coincidencia entre ellos (Fleiss et al., 2003). Los resultados de esta evaluación, así como las decisiones derivadas de ellos, se explican en el apartado de Análisis de los resultados (§ANÁLISIS DE LOS RESULTADOS, sección 5.1).

A priori, se establecen dos etiquetas: «emotivo» o «no emotivo». Sin embargo, rápidamente identificamos la necesidad de una tercera etiqueta neutra, denominada «no sabe», para clasificar aquellos comentarios que no se ajustan a ninguna de las dos opciones previas.

Los criterios son los siguientes:

- No basarse en la idea general del mensaje, sino en palabras o grupos de palabras concretos.
- Emotivos:
 - a. Mensajes pensados para influir/convencer a alguien mediante exclamaciones, llamamientos, espoleando, etc., pero nunca con razones.
 - b. Mensajes para imponer o expresar nuestra posición, sin intención de informar.
- No emotivos:
 - a. Un mensaje que anuncia.
 - b. Un mensaje que expresa razones que respaldan una posición.

Extraemos 650 comentarios aleatoriamente, los cuales dividimos en subcorpora para llevar a cabo el etiquetado manual. Clasificamos 50 comunes, es decir, cada uno los etiqueta y después contrastamos el resultado para evaluar la fiabilidad del etiquetado mediante Kappa de Fleiss, como se explica anteriormente, y 150 individualmente. Trabajamos en formato Excel usando las bibliotecas `xlrd` y `xlswriter`, las cuales se utilizan para interactuar con hojas de cálculo de Office desde Python (Caballero Roldán et al., 2018). Una vez etiquetados, los comentarios se

procesan aplicando todas las funciones diseñadas para obtener las proporciones de los distintos parámetros seleccionados previamente. Todo ello se almacena en una lista de diccionarios que utilizaremos para entrenar al modelo.

4.6. Entrenamiento y aplicación del clasificador automático.

Para trabajar con la información almacenada en `lista_etiquetada`, el resultado del proceso descrito anteriormente, la convertimos en un `DataFrame` con la biblioteca `pandas`. Este formato es similar a Excel en cuanto a la disposición de la información en columnas y filas. Como indicábamos antes, el modelo funciona con valores numéricos, por tanto, agrupamos, por un lado, los atributos expresados en números, que son la proporción de verbos, conectores, exclamaciones e interrogaciones, mayúsculas y emoticonos en cada comentario, y por otro, la clase que hemos asignado a cada comentario, es decir, la etiqueta, también expresada con un número: 0, 1 o 2.

Seguidamente, dividimos la muestra de comentarios en 80% conjunto de entrenamiento y 20% conjunto de test. Esta división se lleva a cabo para reservar el conjunto de test con el cual se valida la calidad del modelo, empleando así comentarios no incluidos en el entrenamiento. Creamos un objeto de la clase `DecisionTreeClassifier` de `Scikit-learn` y ajustamos el modelo con el método `fit`, lo que generará un modelo de clasificación como árbol de decisión que permitirá visualizar el conocimiento extraído en un formato comprensible a simple vista. Otros métodos más opacos, como las redes neuronales, hacen más difícil explicar los resultados obtenidos, por lo tanto, pueden ser más potentes en la clasificación, pero dificultan la comprensión del proceso y los resultados. Esto es especialmente importante en esta investigación porque necesitamos saber cómo ha decidido el modelo para poder evaluar su capacidad de predicción.

Finalmente, una vez hemos entrenado al modelo con ese primer conjunto formado por el 80% de la muestra, evaluamos el clasificador para comprobar su tasa de acierto (*accuracy*). Para ello, clasificamos las instancias del conjunto de test usando el árbol y el método `predict`. Después, medimos la precisión con la función `accuracy_score` de `Scikit-learn`, que compara las etiquetas reales del conjunto de test con la predicción realizada, obteniendo la tasa de acierto del modelo.

4.7 Evaluación mediante matriz de confusión.

Por último, consideramos interesante incluir una matriz de confusión para entender mejor el rendimiento del modelo y poder analizar los resultados con más información. Una matriz de confusión es una herramienta que permite visualizar el desempeño de un modelo de

clasificación¹⁵ ya que muestra los aciertos en cada categoría, esto es, «emotivo», «no emotivo» y «no sabe» en nuestro caso. Para visualizarla se necesita importar `matplotlib.pyplot`, así como el propio método `plot_confusion_matrix` de `sklearn.metrics`, y aplicar este al conjunto de test y al árbol de clasificación. El detalle de los resultados se puede ver en el siguiente apartado (§ANÁLISIS DE LOS RESULTADOS, sección 5.3).

¹⁵ Documentación de Scikit-learn (11 agosto 2021) https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html

5. Análisis de los resultados

Esta sección analiza los resultados obtenidos en la investigación. Se divide en tres partes para revisar la información que podemos extraer de cada fase del procesamiento de los comentarios objeto de estudio:

5.1. Etiquetado manual.

5.2. Resultados del entrenamiento y el test del modelo.

5.3. Análisis mediante matriz de confusión.

5.1. Etiquetado manual.

El etiquetado manual puede ser evaluado para conocer su grado de fiabilidad (§METODOLOGÍA, sección 4.5). Kappa de Fleiss es un parámetro estadístico empleado en los casos en que hay más de dos etiquetadores o *raters* (Fleiss et al., 2003). Como se explica anteriormente, el equipo etiquetador de esta investigación es diverso, por lo que decidimos llevar a cabo la comprobación.

El rango de valores de resultado puede oscilar entre 0 y 1, por tanto, el resultado advierte de una fiabilidad aceptable pero baja: 0.348¹⁶. Por ello, decidimos repetir todo el proceso con un solo etiquetador, para observar el impacto sobre los resultados y evitar esta problemática. Estos mejoran ligeramente, pero entendemos que la incidencia sobre el modelo no es significativa.

5.2. Resultados del entrenamiento y el test del modelo.

Como se explica en la sección 4.6 del apartado de Metodología (§METODOLOGÍA), entrenamos y probamos el modelo con el método `DecisionTreeClassifier` de Scikit-learn sobre 650 comentarios del corpus, y evaluamos la tasa de acierto con `accuracy_score`, obteniendo un 0.5 de exactitud.

Las herramientas *scoring*¹⁷ de Scikit-learn sirven para evaluar diferentes parámetros, entre ellos, la tasa de acierto o exactitud de un modelo. La cifra que obtenemos al aplicar esta técnica es, lógicamente, mejor cuanto más se acerca a uno, ya que uno representa un 100% de acierto. Una exactitud en torno al 0.5 no es lo suficientemente alta como para considerar al

¹⁶ de Bruijn, L. (30 agosto 2021) Inter-Annotator Agreement (IAA). *towards data science*. <https://towardsdatascience.com/inter-annotator-agreement-2f46c6d37bf3>

En la página se puede consultar la tabla de valores Kappa y su interpretación, donde 0.21-0.40 se considera «*Fair agreement*».

¹⁷ Documentación oficial de la biblioteca Scikit-learn sobre *scoring* (26 de agosto de 2021) https://scikit-learn.org/stable/modules/model_evaluation.html#scoring-parameter.

modelo eficaz para discriminar por sí solo los mensajes con las tres etiquetas que hemos empleado.

Para entender mejor los datos obtenidos, generamos un informe mediante ProfileReport de pandas_profiling que muestra el detalle estadístico de nuestro corpus etiquetado. El informe nos confirma que, de los 650 comentarios, 359 son «emotivos», 234 son «no sabe» y 57 «no emotivos». La infrarrepresentación de comentarios «no emotivos» hace que al modelo le sea difícil aprender a identificarlos. Asimismo, el resto de las variables presentan datos similarmente dificultosos. El 90,6% de los comentarios no contiene verbos argumentativos, el 84,3% no contiene conectores de nuestra lista, el 72,2% no contiene exclamaciones o interrogaciones y el 84,3% no contiene emoticonos. Con el parámetro mayúsculas ocurre, previsiblemente, lo contrario, el 96,3% las contiene.

La tasa de «zeros», es decir, el porcentaje de comentarios que no contienen ninguno de los parámetros, es importante para determinar la dificultad a la que se ha enfrentado el modelo para aprender. Cuando dicha tasa es equilibrada, el aprendizaje es más profundo y los resultados más fiables, puesto que los datos disponibles han sido suficientes para trazar conclusiones informadas.

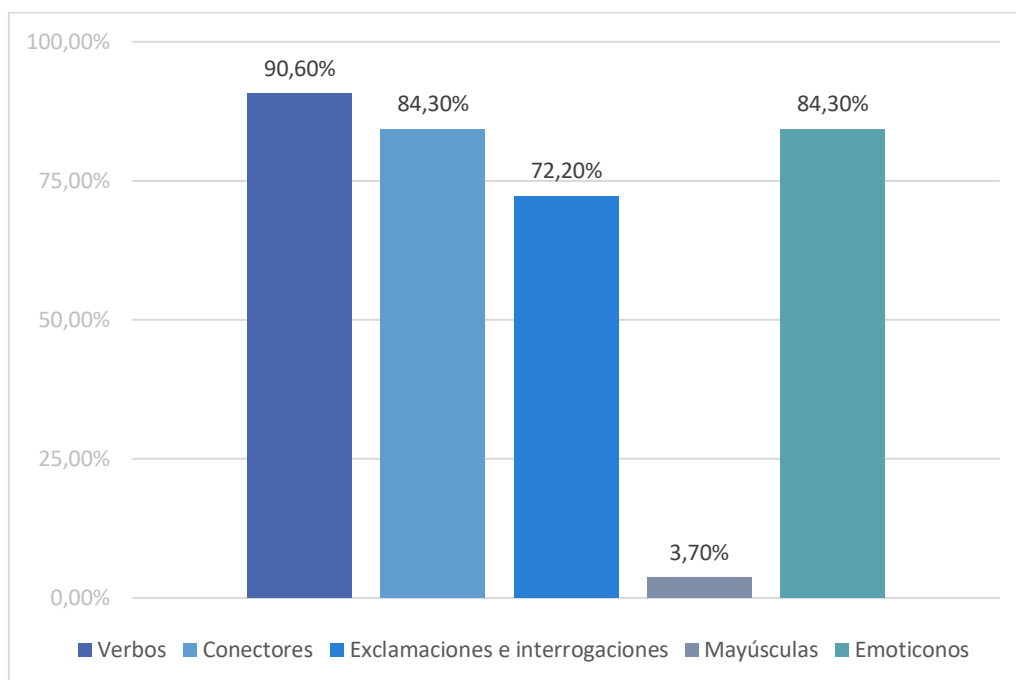


Figura 5. Gráfico del porcentaje de comentarios que no contienen los parámetros investigados.

Es importante recordar que la presencia de verbos y conectores argumentativos en los comentarios implica menor emotividad, mientras que el resto de los parámetros implica mayor

emotividad. Sin embargo, esta distinción no se proporciona como instrucción al algoritmo. Un modelo de aprendizaje automático funciona tomando los parámetros e interpretándolos en función de los datos de entrenamiento que le hemos dado (*train set*). Por ello, la infrarrepresentación de una de las clases, en este caso los comentarios «no emotivos», o la baja incidencia de los parámetros en los comentarios, dificulta dicho aprendizaje, pues el modelo dispone de muy pocos ejemplos para aprender a identificar las características que los definen.

Asimismo, habitualmente se parte de una clasificación binaria, es decir, se asumen dos clases. En nuestro caso existen tres etiquetas: «emotivo», «no emotivo» y «no sabe». Puesto que no es posible establecer la dicotomía «emotivo» vs. «no emotivo» conforme a las definiciones establecidas previamente (§METODOLOGÍA, secciones 4.3 y 4.5), ya que la naturaleza del texto no lo permite (necesitamos una etiqueta neutra o intermedia para aquellos comentarios que no cumplen con las condiciones de las otras dos), para establecer una solución binaria satisfactoria podríamos asumir las clases «emotivo» y «sin garantías». Esto es, asegurar que aquello que se etiqueta como emotivo lo es, cumpliendo así con el primer objetivo de esta investigación, que es responder a la pregunta «¿Es posible medir la emotividad en los comentarios de YouTube?» (§OBJETIVOS). Esto sería un paso adelante en el estudio de «buus y hurras», pues se podrían aplicar técnicas de análisis de sentimiento, por ejemplo, para distinguir «buus» de «hurras», sabiendo que estamos haciéndolo sobre comentarios que han sido etiquetados como emotivos con una alta fiabilidad.

5.3. Análisis mediante matriz de confusión.

Para poder confirmar si el modelo sirve al propósito de identificar comentarios emotivos con seguridad, debemos observar cómo ha acertado y fallado en el proceso de clasificación. Esto se consigue mediante una matriz de confusión (§METODOLOGÍA, sección 4.7). Esta herramienta muestra el detalle de la predicción realizada por el modelo comparando dicha predicción (eje horizontal *Predicted label*) con las etiquetas asignadas originalmente de forma manual (eje vertical *True label*).

En la figura 6, vemos que la columna de la izquierda presenta los 95 comentarios etiquetados por el modelo como emotivos, de ellos, 57 son emotivos en su etiqueta original, es decir, ha acertado en un 54,15% de los casos. Como cabe esperar, donde más falla es en la identificación de comentarios «no emotivos» (columna central). De los 8 etiquetados así, ha coincidido solo en uno, es decir, un 12,5% de acierto. Similarmente, con respecto a los comentarios etiquetados como «no sabe» (columna derecha), ha coincidido con el original en un 33,3%.

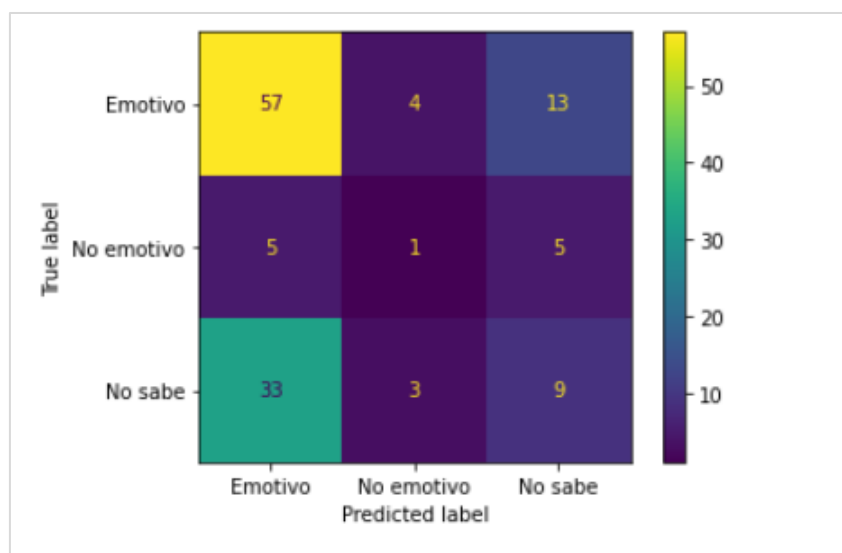


Figura 6. Matriz de confusión.

A la luz de este análisis, podríamos concluir que la respuesta a la segunda pregunta de investigación «¿Se pueden distinguir los comentarios emotivos de los no emotivos si observamos aquellos publicados en vídeos de la campaña de las elecciones a la Asamblea de Madrid de 2021?» (§OBJETIVOS) es que, con el diseño actual del modelo, no es posible hacer dicha distinción con garantías suficientes.

6. Conclusiones

En conclusión, esta investigación nos ha permitido ahondar en el estudio de la polarización y la emotividad en redes sociales a través del análisis del lenguaje de los comentarios publicados en vídeos de YouTube de temática política. Hemos podido constatar que es posible medir la emotividad de dichos mensajes mediante herramientas de programación e inteligencia artificial. El aprendizaje automático ha confirmado su potencial para el análisis del lenguaje con objetivos sociolingüísticos, ya que hemos podido clasificar un corpus para este fin.

Sin embargo, los resultados obtenidos también han apuntado a la necesidad de afinar dicho análisis para conseguir resultados más fiables. La naturaleza subjetiva y las múltiples capas de significado del lenguaje hacen del análisis pragmático automático una tarea complicada. Sin embargo, un estudio lingüístico *ad hoc* más profundo del corpus podría, posiblemente, ayudar a definir características de los comentarios más estables que proporcionen más información a la hora de entrenar al algoritmo.

A pesar de que, con las variables propuestas y estudiadas en esta investigación, no ha sido posible realizar una adecuada clasificación de los comentarios, pues dichas variables han demostrado estar infrarrepresentadas, generando resultados de fiabilidad baja, sí ha sido posible crear un modelo que supone un primer paso en la tipificación de mensajes emotivos. Estos resultados y el estudio de los mensajes etiquetados como «emotivos» pueden mejorar aplicando posteriormente técnicas más potentes, como el análisis de sentimiento o un clasificador probabilístico de Naive Bayes. Sin embargo, aunque otras tecnologías son más potentes en la clasificación, se perdería la claridad que ofrece el árbol de decisión y los resultados dejarían de ser explicables.

No obstante, teniendo un conjunto de datos desequilibrado, aplicar *random resampling* podría mejorar los resultados (Branco et al., 2015). Esto consiste en modificar el conjunto de entrenamiento del modelo, ya sea mediante *oversampling* o *undersampling*, para ajustar los niveles de clases infrarrepresentadas hasta conseguir equilibrar dicho conjunto. Aunque estas estrategias implican el riesgo de perder información valiosa sobre la muestra, pues consiste en añadir o eliminar instancias del conjunto de entrenamiento, habitualmente suponen una solución sencilla y eficaz para mejorar el aprendizaje y, en consecuencia, los resultados.

Si bien es cierto que la subjetividad inmanente al lenguaje complica el análisis automático del mismo, generando amplias dudas sobre su viabilidad y posibilidades de éxito en algunas voces (Bender & Koller, 2020), las tecnologías informáticas de procesamiento del lenguaje natural han demostrado una evidente capacidad evolutiva. Por eso, es importante llevar a cabo estudios de esta naturaleza y más necesario, si cabe, en español, donde es más difícil encontrar

recursos ya probados que sirvan para acelerar en nuevas investigaciones. Por ejemplo, sería interesante y muy útil que instituciones oficiales como las universidades públicas, la RAE o el Instituto Cervantes ofrecieran o acogieran listados de palabras de diferente naturaleza, orientados al PLN, como podrían ser listas de *stopwords*, listados genéricos orientados al análisis de sentimiento (palabras con connotación negativa o positiva), etc. Dichos listados tendrían que ser revisados para cada estudio para asegurar la validez de los mismos en ese determinado contexto, pero supondrían una ayuda notable en la realización y planteamiento de nuevas investigaciones de carácter lingüístico.

7. Bibliografía

7.1. Bibliografía primaria.

- Adamic, L. A. & Glance, N. (2005). The political blogosphere and the 2004 US election: divided they blog. *Proceedings of the 3rd international workshop on Link discovery. ACM*, 36–43.
<https://doi.org/10.1145/1134271.1134277>
- Apaydin, E. (2014). Introduction to machine learning. *ProQuest Ebook Central*.
<https://ebookcentral.proquest.com>
- Ayer, A. J. (1952). *Language, Truth, and Logic*. New York: Dover Publications.
- Baek, Y. M., Wojcieszak, M., & Delli Carpini, M. X. (2012). Online versus face-to-face deliberation: Who? Why? What? With what effects? *New Media and Society*, 14(3), 363–383.
<https://doi.org/10.1177/1461444811413191>
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., ... Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221.
<https://doi.org/10.1073/pnas.1804840115>
- Barberá, P. (2014). How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the US. *2015 American Political Science Association Conference*.
- Bender, E. M. & Koller, A. (2020). Climbing towards NLU: on meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 5185–5198.
<https://doi.org/10.18653/v1/2020.acl-main.463>
- Benkler, Y., & Nissenbaum, H. (2006). Commons-based Peer Production and Virtue. *Journal of Political Philosophy*, 14(4), 394–419.
- Benkler, Y. (2015). *Las riquezas de las redes*. Barcelona: Ikaria.
- Bimber, B. (2001). Information and Political Engagement in America: The Search for Effects of Information Technology at the Individual Level. *Political Research Quarterly*, 54(1), 53–67.
<https://doi.org/10.1177/106591290105400103>
- Branco, P., Torgo, L., & Ribeiro, R. (2015). A Survey of Predictive Modelling under Imbalanced Distributions. *ArXiv:1505.01658 [Cs]*. <http://arxiv.org/abs/1505.01658>
- Caballero Roldán, R., Martín Martín, E., & Riesco Rodríguez, A. (2018). *Big data con Python: recolección, almacenamiento y proceso*. Madrid: RC Libros.
- Castells, M. (2009). *Comunicación y poder*. Madrid: Alianza Editorial.

- Chen, G. M., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We Should Not Get Rid of Incivility Online. *Social Media + Society*, 5(3), 1-5.
<https://doi.org/10.1177/2056305119862641>
- Chowdhury, G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, 37(1), 51-89. <https://doi.org/10.1002/aris.1440370103>
- Colleoni, E., Rozza, A. & Arvidsson, A. (2014). Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *Journal of Communication*, 64(2), 317–332. <https://doi.org/10.1111/jcom.12084>
- Conover, M. D., Gonçalves, B., Flammini, A., & Menczer, F. (2012). Partisan Asymmetries in Online Political Activity. *EPJ Data Science*, 1(6), 1–19. <https://doi.org/10.1140/epjds6>
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion Recognition in Human-Computer Interaction. *IEEE Signal Processing Magazine*, 18(1), 32-80. <https://doi.org/10.1109/79.911197>
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). The Measurement of Interrater Agreement. En *Statistical Methods for Rates and Proportions*, 2(212-236), 22-23. John Wiley & Sons, Inc. <https://doi.org/10.1002/0471445428>
- Habermas, J. (1974). *Theory and practice*. Londres: Heinemann.
- Herscovich, D., & Donatelli, L. (2021). It's the Meaning That Counts: The State of the Art in NLP and Semantics. *KI - Künstliche Intelligenz*. <https://doi.org/10.1007/s13218-021-00726-6>
- Kaid, L. (2002). Political Advertising and Information Seeking: Comparing Exposure via Traditional and Internet Channels. *Journal of Advertising*, 31(1), 27-35.
<https://doi.org/10.1080/00913367.2002.10673658>
- Linzen, T. (2020). How Can We Accelerate Progress Towards Human-like Linguistic Generalization? *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 5210–5217.
<https://doi.org/10.18653/v1/2020.acl-main.465>
- Liu, H., Lieberman, H., & Selker, T. (2002). Automatic Affective Feedback in an Email Browser. *MIT Media Lab Software Agents Group Technical Report*.
- López, C. (2006). Marcas de subjetividad y argumentación en tres géneros especializados del español. *Revista signos*, 39(61), 205-229.
<https://dx.doi.org/10.4067/S0718-09342006000200004>
- López García, G. (2011). Comunicación en red y mutaciones de la esfera pública. *Servicio Editorial de la Universidad del País Vasco/Euskal Herriko Unibertsitatearen Argitalpen Zerbitzua*, 11(20), 231-249. <http://hdl.handle.net/10810/40907>

- Louis, A. (2016). Natural Language Processing for Social Media. (Review). *Computational Linguistics*, 42(4), 833–836. https://doi.org/10.1162/COLI_r_00270
- Ma, X., Wang, H., Li, H., Liu, J., & Jiang, H. (2013). Exploring Sharing Patterns for Video Recommendation on YouTube-like Social Media. *Multimedia Systems*, 20(6), 675–691. <https://doi.org/10.1007/s00530-013-0309-1>
- Martín Zorraquino, M. A. & Portóles Lázaro, J. (1999). Los marcadores del discurso. En Bosque, I. y Demonte, V., (dirs.). *Gramática descriptiva de la lengua española*, 3(63), 4051-4213. Madrid: Espasa.
- Mason, L. (2014). “I disrespectfully agree”: The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science*, 59(1), 128–145. <https://doi.org/10.1111/ajps.12089>
- Moreno, A. (5 agosto 2021) Procesamiento del lenguaje natural ¿qué es? *Blog del Instituto de Ingeniería del Conocimiento*. <https://www.iic.uam.es/inteligencia/que-es-procesamiento-del-lenguaje-natural/>
- Mouffe, C. (2000). *The Democratic Paradox*. Londres: Verso.
- Mouffe, C. (2005). *On the Political*. Londres: Routledge.
- Olsson, E. J. (2013). A Bayesian Simulation Model of Group Deliberation and Polarization. En Zenker, F. (Ed.), *Bayesian Argumentation. Synthese Library (Studies in Epistemology, Logic, Methodology, and Philosophy of Science)*, 362, 113–133. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-5357-0_6
- Papacharissi, Z. (2002). The virtual sphere: The Internet as a Public Sphere. *New Media & Society*, 4(1), 9–27. <https://doi.org/10.1177/14614440222226244>
- Papacharissi, Z. (2004). Democracy Online: Civility, Politeness, and the Democratic Potential of Online Political Discussion Groups. *New Media & Society*, 6(2), 259-283. <https://doi.org/10.1177/1461444804041444>
- Papacharissi, Z. (2008). The Virtual Sphere 2.0: The Internet, the Public Sphere and Beyond. En Chadwick, A. & Howard, P. N. (Eds.), *The Routledge Handbook of Internet Politics*, 246-261. Londres: Routledge.
- Picard, R. W. (1997) *Affective Computing*. Cambridge, MA: MIT Media Laboratory.
- Reynolds, C. (2006). Boo-Hooray and Effective Approaches to Ethical Textual Analysis. *Computers & Philosophy, an International Conference*, 233-246.
- Robles, J. M., Atienza, J., Gómez, D. & Guevara Gil, J. (2019). La polarización de “La Manada”. El debate público en España y los riesgos de la comunicación política digital. *Tempo Social*, 31(3), 193-216. <https://doi.org/10.11606/0103-2070.TS.2019.159680>

- Robles, J. M., Velez, D., De Marco, S., Rodríguez, J. T., & Gomez, D. (2020). Affective Homogeneity in the Spanish General Election Debate. A Comparative Analysis of Social Networks Political Agents. *Information, Communication & Society*, 23(2), 216-233.
<https://doi.org/10.1080/1369118X.2018.1499792>
- Schäfer, M. S. (2016). Digital Public Sphere. *The International Encyclopedia of Political Communication*, 1–7. <https://doi.org/10.1002/9781118541555.WBIEPC087>
- Schaub, L-P. (2020). La industria del lenguaje en la era del dato. *Revista Abaco*, hal-02912828.
- Serrano-Contreras, I. J., García Marín, J. & Luengo, O. (2020). Measuring Online Political Dialogue: Does Polarization Trigger More Deliberation? *Media and Communication*, 8(4), 63-72. <https://doi.org/10.17645/mac.v8i4.3149>
- Stevenson, C. L. (1960). *Ethics and Language* (Ser. A yale paperbound, 19). New Haven: Yale University Press.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460.
<https://doi.org/10.1093/mind/LIX.236.433>
- Velez, D., Sueiras, J., Ortega, A., & Velez, J. F. (2016). A Method for K-means Seeds Generation Applied to Text Mining. *Statistical Methods & Applications*, 25(3), 477–499.
<https://doi.org/10.1007/s10260-015-0345-4>
- Verdejo Maillo, M. F. (1994). Procesamiento del Lenguaje Natural: fundamentos y aplicaciones. *UNED, Curso de Verano*.
- Whissel, C. M. (1989). The Dictionary of Affect in Language. En Plutchik, R. y Kellerman, H. (Eds.), *The Measurement of Emotions*, 4, 113-131. New York: Academic Press.
<https://doi.org/10.1016/B978-0-12-558704-4.50011-6>
- Wojcieszak, M. E., Baek, Y. M. & Delli Carpini, M. X. (2009). What is really going on? Structure underlying face-to-face and online deliberation. *Information, Communication & Society* 12(7), 1080–1102. <https://doi.org/10.1080/13691180902725768>
- Wojcieszak, M. E., & Mutz, D. C. (2009). Online Groups and Political Discourse: Do Online Discussion Spaces Facilitate Exposure to Political Disagreement? *Journal of Communication*, 59(1), 40–56. <https://doi.org/10.1111/j.1460-2466.2008.01403.x>

7.2. Bibliografía secundaria.

- Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on Twitter. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 5(1), 89–96.
- Chomsky, N. (1966). *Syntactic structures* (6th printing, Ser. Janua linguarum, 4). La Haya: Mouton.
- Damasio, A., & Meyer, K. (2008). Behind the looking-glass. *Nature*, 454(7201), 167-8. <http://dx.doi.org/10.1038/454167a>
- Eger, S., Şahin, G. G., Rücklé, A., Lee, J.U., Schulz, C., Mesgar, M., Swarnkar, K., Simpson, E., & Gurevych, I. (2019). Text Processing Like Humans Do: Visually Attacking and Shielding NLP Systems. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 1634–1647. Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1165>
- Eichmann, D. (1995). Ethical Web Agents. *Computer Networks and ISDN Systems*, 28(1-2), 127-136. [https://doi.org/10.1016/0169-7552\(95\)00107-3](https://doi.org/10.1016/0169-7552(95)00107-3)
- Feldman, S. (1999). NLP meets the Jabberwocky: Natural Language Processing in Information Retrieval. *Information Today*, 23(3), 62-72.
- Finlayson, J. G. (2005). *Habermas: A very short introduction*. ProQuest Ebook Central <https://ebookcentral.proquest.com>
- Grusell, M., & Nord, L. (2012). Three Attitudes to 140 Characters: The Use and Views of Twitter in Political Party Communications in Sweden. *Public Communication Review*, 2(2), 48–61. <https://doi.org/10.5130/pcr.v2i2.2833>
- Gruzd, A., & Roy, J. (2014). Investigating Political Polarization on Twitter: A Canadian Perspective. *Policy & Internet*, 6(1), 28–45. <https://doi.org/10.1002/1944-2866.POI354>
- Jaidka, K., Zhou, A., & Lelkes, Y. (2019). Brevity is the Soul of Twitter: The Constraint Affordance and Political Discussion. *Journal of Communication*, 69(4), 345–372. <https://doi.org/10.1093/joc/jqz023>
- Liddy, E. (1998). Enhanced Text Retrieval Using Natural Language Processing. *Bulletin of the American Society for Information Science*, 24, 14-16.
- Lleó, J. (2020). Modelo para la minería de textos en el sector periodístico. *Universidad Complutense de Madrid*. <https://eprints.ucm.es/id/eprint/62598/>
- Marichal, J. (2016). *Facebook Democracy: The Architecture of Disclosure and the Threat to Public Life*. Taylor & Francis. <http://library.oapen.org/handle/20.500.12657/30601>

- Oz, M., Zheng, P., & Chen, G. M. (2018). Twitter versus Facebook: Comparing Incivility, Impoliteness, and Deliberative Attributes. *New Media & Society*, 20(9), 3400–3419.
<https://doi.org/10.1177/1461444817749516>
- Pariser, E. (2011). *The Filter Bubble: What the Internet is Hiding from You*. London: Viking/Penguin Press.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, 12, 2825–2830.