



UNIVERSIDAD INTERNACIONAL DE
LA RIOJA

PROGRAMA DE DOCTORADO EN CIENCIAS DE LA
COMPUTACIÓN

TESIS DOCTORAL

**Metodologías de estudio de la actitud de los usuarios a
través de su huella digital en Internet**

Memoria presentada

por

Ernesto del Valle Martín

para optar al grado de Doctor

por la Universidad Internacional de La Rioja

Dirigida por las Doctoras:

Laura de Miguel Álvarez

y

M^a Begoña Yáñez Martínez

Madrid, 2022

Agradecimientos

A Laura de Miguel y Begoña Yáñez, por su paciencia y excelente guía. A mi esposa, Isabel Ruano, y a mis hijos, Laura y Ernesto, por el tiempo que se ha llevado esta tesis y que podría (o debería) haberles dedicado a ellos.

Resumen

El estudio de la actitud y la opinión de los usuarios en Internet es una materia de creciente interés en diferentes sectores académicos y profesionales, tales como el diseño de la experiencia de usuario, la educación, la innovación tecnológica, el marketing digital, la investigación de mercados, la política y la sociología. Nuevas amenazas emergentes, incompatibles con un desarrollo armónico de las sociedades, han acrecentado en tiempos recientes este interés desde perspectivas tan preocupantes como novedosas: noticias falsas, ciber acoso, discursos de odio, injerencia de potencias extranjeras en procesos electorales, y desinformación en materias de salud pública (en especial las relativas a la pandemia del Covid-19).

La preocupación sobre el estudio de la actitud de los ciudadanos en Internet, y en especial en las redes sociales, se ha trasladado a la comunidad científica, que ha intentado en los últimos años desarrollar herramientas, ya sea conceptuales, metodológicas o puramente técnicas, para responder adecuadamente a los nuevos desafíos. Sin embargo, estamos ante fenómenos inéditos y extraordinariamente recientes, donde el *background* académico y tecnológico es necesariamente limitado.

Esta tesis se plantea determinar hasta dónde se puede llegar a conocer y medir la actitud de los usuarios en Internet en el contexto actual, y si es posible utilizar ese conocimiento para influir o manipular. Para ello, hemos analizado los recursos técnicos y metodológicos que están a disposición de la comunidad académica y la población en general, con especial foco en la pandemia del COVID-19.

El análisis realizado a lo largo de la investigación abarca en profundidad los métodos y herramientas existentes, ya sean estrictamente académicos, comerciales, e incluso “informales”. Además de este análisis empírico, se han realizado diversas investigaciones de nueva planta para responder a cuestiones no resueltas y para obtener datos que permitan contestar a las preguntas de investigación sin dejar

espacios en blanco. En particular, entre los distintos métodos empleados para la redacción de estas tesis, se han realizado ocho investigaciones novedosas cuyos resultados arrojan luz a la materia de estudio.

Como resultado de la investigación, podemos concluir que es posible identificar, extraer y procesar masivamente opiniones concretas en Internet, sobre temas específicos de cualquier naturaleza, siempre que sean de acceso público. En particular, la tesis demuestra que podrían realizarse este tipo de tareas usando herramientas comerciales al alcance de cualquier persona con unos mínimos conocimientos de comunicación digital. También hemos podido determinar que no existe posibilidad de influenciar directamente a los ciudadanos con las herramientas y métodos del arte, pero sí de manera indirecta a partir de la información obtenida, ya sea al servicio de causas legítimas, como la concienciación, el desmentido de falsas informaciones y la investigación académica, o ilegítimas, como la manipulación.

Palabras clave

- Análisis automático de textos
- Medios sociales
- Opinión pública
- Inteligencia artificial

Abstract (EN)

The study of user attitudes and opinions on the Internet is a subject of growing interest in different academic and professional sectors, such as user experience design, education, technological innovation, digital marketing, market research, politics and sociology. New emerging threats, incompatible with the harmonious development of societies, have recently increased this interest from perspectives as worrying as they are novel: fake news, cyberbullying, hate speech, interference by foreign powers in electoral processes, and disinformation on public health issues (especially those related to the Covid-19 pandemic).

Concern about the study of citizens' attitudes on the Internet, and especially on social networks, has been transferred to the scientific community, which has tried in recent years to develop tools, whether conceptual, methodological or purely technical, to respond adequately to the new challenges. However, we are dealing with unprecedented and extraordinarily recent phenomena, where the academic and technological background is necessarily limited.

This thesis sets out to determine how far it is possible to get to know and measure user attitudes on the Internet in the current context, and whether it is possible to use this knowledge to influence or manipulate. To this end, we have analyzed the technical and methodological resources available to the academic community and the general public, with a special focus on the COVID-19 pandemic.

The analysis conducted throughout the research covers in depth the existing methods and tools, whether strictly academic, commercial, or even "informal". In addition to this empirical analysis, several new plant investigations have been carried out to answer unresolved questions and to obtain data to answer the research questions without leaving blanks. In particular, among the various methods used for the writing of this thesis, eight novel investigations have been carried out, the results of which shed light on the subject of study.

As a result of the research, we can conclude that it is possible to identify, extract and massively process concrete opinions on the Internet, on specific topics of any nature, provided that they are publicly accessible. In particular, the thesis demonstrates that this type of task could be performed using commercial tools within the reach of any person with a minimum knowledge of digital communication. We have also been able to determine that there is no possibility of influencing citizens directly with the tools and methods of art, but indirectly from the information obtained, either in the service of legitimate causes, such as awareness, the disproving of false information and academic research, or illegitimate, such as manipulation.

Keywords

- Automatic text analysis
- Social media
- Public opinion
- Artificial intelligence

Índice de contenido

1	Introducción a la investigación	28
1.1	Justificación de la investigación.....	28
1.1.1	Antecedentes	31
1.2	Fundamentos teóricos	33
1.2.1	Aproximación a los conceptos principales.....	33
1.2.2	Nomenclatura y debate semántico.....	34
1.2.3	Ámbitos de aplicación y áreas del conocimiento relacionadas	35
1.2.4	Desarrollo histórico y taxonómico	36
1.2.5	El concepto “huella digital”	39
1.2.6	Foco en la red social Twitter	40
1.2.7	Preferencias y limitaciones idiomáticas.....	42
1.2.8	Foco actual en <i>machine learning</i>	44
1.2.9	Interacción hombre-máquina.....	45
1.2.10	Predominio de las herramientas de código abierto	46
1.2.11	Propósito y perspectivas para la investigación en el contexto actual	46
1.3	Objetivos	48
1.3.1	Objetivo general.....	48
1.3.2	Objetivos específicos y preguntas de investigación	48
1.3.3	Enfoque de la investigación alineado con los objetivos	50
1.4	Hipótesis	51
1.5	Metodología.....	52
1.5.1	Procedimiento.....	52
1.5.2	Medios.....	53
1.5.3	Métodos.....	54
1.6	Estructura de la tesis	57
1.6.1	Contenido de la Introducción.....	57
1.6.2	Contenido del estado del arte	58
1.6.3	Contenido del desarrollo de la investigación.....	58

1.6.4	Contenido de las conclusiones	60
1.6.5	Contenido de la bibliografía.....	60
1.6.6	Contenido de los anexos.....	61
2	Contexto y estado del arte.....	62
2.1	Campo de estudio: origen, evolución y contextualización temporal.....	62
2.2	Limitaciones técnicas y desafíos en el estado del arte.....	65
2.2.1	Rendimiento y <i>benchmarks</i>	65
2.2.2	Desafío idiomático.....	68
2.2.3	Obsolescencia de los trabajos publicados.....	71
2.3	Espacios de interés académico y social.....	72
2.3.1	Sectores de aplicación.....	72
2.3.2	Transversalidad académica.....	74
2.3.3	Áreas temáticas predominantes.....	75
2.3.4	Puntos de inflexión.....	77
3	Desarrollo de la investigación.....	79
3.1	Los métodos.....	79
3.1.1	Extracción y <i>scope</i> del análisis.....	81
3.1.1.1	Nivel de granularidad.....	81
3.1.1.2	Identificación/definición de opiniones.....	83
3.1.1.3	Tareas.....	84
3.1.1.4	Spam detection.....	85
3.1.2	Preprocesamiento de textos.....	86
3.1.3	Selección de características (<i>feature selection</i>).....	88
3.1.4	Metodologías de clasificación.....	92
3.1.4.1	Machine learning.....	92
3.1.4.2	Deep Learning.....	95
3.1.4.3	Lexicon based.....	96
3.1.4.4	Métodos híbridos.....	98
3.1.4.5	Métodos “Graph based”.....	99
3.1.4.6	Otros métodos.....	100
3.1.5	Métricas de evaluación.....	101
3.1.5.1	Resultados comparativos con herramientas más habituales.....	103

3.2	Herramientas y recursos para la monitorización y análisis de la actitud de los usuarios en Internet.....	107
3.2.1	Uso y percepción de las herramientas de monitorización y análisis de la actitud de los usuarios en Internet.....	110
3.2.1.1	Nivel de conocimiento.....	111
3.2.1.2	Percepción.....	113
3.2.1.3	Uso.....	117
3.2.2	Herramientas académicas	119
3.2.3	Herramientas específicas de la comunidad investigadora hispanohablante	125
3.2.3.1	Freeling.....	125
3.2.3.2	LinguaKit.....	126
3.2.4	Herramientas comerciales empleadas por la comunidad científica.....	126
3.2.4.1	Microsoft Azure Text Analytics.....	127
3.2.4.2	IBM Watson Natural Language Understanding.....	132
3.2.4.3	Semantria	133
3.2.4.4	Repustate.....	134
3.2.5	Herramientas comerciales informales	136
3.2.5.1	Uso de herramientas comerciales informales por la comunidad investigadora.	137
3.2.5.2	Sentione	138
3.2.5.3	Brandwatch	140
3.2.5.4	Audiense.....	142
3.2.5.5	MeaningCloud.....	143
3.2.6	Evaluación del rendimiento de las herramientas académicas: <i>benchmark</i> actual	145
3.2.7	Evaluación del rendimiento de las herramientas comerciales usadas por la comunidad investigadora.....	147
3.2.7.1	Segundo estudio ampliado	149
3.2.7.2	Benchmarking de herramientas comerciales “formales”.....	152
3.3	Recursos para el desarrollo de metodologías y herramientas <i>ad hoc</i>	154
3.4	Problemas y desafíos comunes de la comunidad investigadora.....	155
3.4.1	El debate en torno a la aplicación de metodologías de <i>Machine Learning</i> ..	155
3.4.2	Polaridad y sus frentes: la gran asignatura pendiente	156

3.4.2.1	Desafío 1: la identificación de polaridad a nivel de entidad.....	157
3.4.2.2	Desafío 2: la negación y su impacto en la polaridad	159
3.4.2.3	Desafío 3: la relatividad de la polaridad por el empleo de la ironía y el sarcasmo 159	
3.4.3	Investigación multidisciplinar poco madura	160
3.4.4	Desafíos y debates locales: recursos y herramientas específicos de la comunidad investigadora de habla hispana	161
3.4.5	Problemas específicos de Twitter.....	162
3.4.6	Conjuntos de datos de referencia	163
3.4.7	Escasez de estrategias para el tratamiento de contenido multilingüe.....	163
3.4.8	Palabras de parada o <i>stop words</i>	164
3.4.9	Brevedad del texto: una anomalía con consecuencias metodológicas.....	164
3.4.10	Extracción de textos por relevancia temática y fundamentación del mapa mental de los usuarios	165
3.4.11	Tokenización.....	169
3.4.12	Geolocalización.....	169
3.5	Identificación de usuarios, profiling y manipulación	170
3.5.1	Marco legal.....	170
3.5.2	Identificación de personas concretas	172
3.5.3	Estrategias y herramientas para identificar de usuarios por patrones y potencialmente influir en su actitud.....	173
3.6	Aproximaciones metodológicas específicas sobre el impacto de la pandemia del Covid-19.....	175
3.6.1	Información al alcance de los investigadores sobre la opinión de los ciudadanos en pandemia	179
3.6.2	Aplicación práctica de la monitorización de la actitud con relación a la pandemia.....	185
3.7	Impacto de la preocupación social y académica por el discurso de odio y la polarización política	186
3.8	Accesibilidad de los recursos a personal no cualificado	189
3.8.1	Vía empírica (investigación propia mediante <i>Contextual Inquiry</i>).....	189
3.8.2	Vía estadística	191

3.8.3	Desk research.....	191
4	Conclusiones	193
4.1	Objetivo específico 1: acceso a la información.....	195
4.2	Objetivo específico 2: procesamiento de la información.....	196
4.3	Objetivo específico 3: posibilidad de monitorización en tiempo real.....	197
4.4	Objetivo específico 4: vinculación de actitudes a personas concretas y perfilado 197	
4.5	Objetivo específico 5: precisión, exactitud, y F-Score.....	199
4.6	Objetivo específico 6: medios requeridos.....	200
4.7	Objetivo específico 7: acceso a los medios requeridos	202
4.8	Objetivo específico 8: posibilidad de influir y manipular	202
4.9	Objetivo específico 9: estado del arte, obstáculos y desafíos de la comunidad investigadora.....	206
4.9.1	Cauces futuros previsibles.....	210
4.10	Objetivo específico 10: el impacto de la pandemia.....	210
4.11	Validación de la hipótesis.....	211
4.12	Recapitulación final.....	213
5	Referencias.....	215
6	Anexos	234

Índice de figuras

Figura 1 – Descomposición de la disciplina sentiment analysis (Yadollahi et al., 2017, p. 2)	37
Figura 2 – Tareas abarcadas por la minería de opinión y el análisis de sentimiento (Giachanou & Crestani, 2016, p. 3)	38
Figura 3 – Crecimiento de usuarios de Twitter y volumen de publicaciones sobre Sentiment Analysis. Gráfico elaborado a partir de datos de Twitter y Scopus.	42
Figura 4 – Volumen de trabajos publicados sobre sentiment analysis referidos a política y discurso de odio en España (naranja) y total mundial (azul).....	44
Figura 5 - Trabajos publicados sobre sentiment analysis y opinion mining hasta 2021. Fuente Scopus.....	63
Figura 6 - Trabajos publicados sobre sentiment analysis y opinion mining hasta 2021, para cualquier área temática, referidos específicamente a la lengua española. Fuente Scopus..	64
Figura 7 – Países de origen de los trabajos publicados con foco en la lengua árabe.....	70
Figura 8 - Países de origen de los trabajos publicados con foco en la lengua española.....	70
Figura 9 – Volumen de trabajos publicados sobre análisis de sentimiento referidos a política o a discurso de odio, total mundial.....	73
Figura 10 – Documentos publicados por especialidad académica.	75
Figura 11 – Búsquedas del término “Sentiment Analysis” – Total mundial 2004-2021. Fuente: Google Trends	77
Figura 12 - Visión simplificada de las etapas del proceso de sentiment analysis	80
Figura 13 – Proceso estándar de clasificación de textos - Mironczuk and Protasiewicz (2018)	81
Figura 14 – Niveles de granularidad: documento, frase, entidad y aspecto.....	83
Figura 15 – El marco básico teórico de trabajo para el análisis de sentimiento, convertir texto natural en datos estructurados	83
Figura 16 – Arquitectura detallada del método de Arif et al. (2018 p. 4).....	86
Figura 17 – Etapas de preprocesamiento de textos convencionales para su análisis	86

Figura 18 - Etapas de preprocesamiento con fases adicionales para análisis de textos de micro blogging (por ejemplo, Twitter).....	88
Figura 19 - Ejemplo de árbol de dependencia.....	90
Figura 20 – Arquitectura de la aproximación “híbrida” propuesta Zhang et al. (2011 p. 4)99	
Figura 21 – Gráfico que describe la lógica aplicada por Speriosu et al. Speriosu et al. (2011 p. 57).....	100
Figura 22 – Algoritmo de creación de ontologías creado por Kontopoulos (Kontopoulos et al., 2013 p. 7)	101
Figura 23 – Fórmula F-Score.	103
Figura 24 - Precisión, Recall y F1 obtenido por Ribeiro et al. (2016) con su método de evaluación Sentibench.	104
Figura 25 (CONTINUACIÓN de la figura 24) – Precisión, Recall y F1 obtenido por Ribeiro et al. (2016) con su método de evaluación Sentibench.....	105
Figura 26 - F1 score medio obtenido para cada clase (negativa, positiva, neutra) en el estudio comparativo realizado por Dolianiti et al. (2019). Se observa cómo los métodos analizados consiguen un mejor rendimiento en predicción de mensajes positivos.....	106
Figura 27 – Nivel de experiencia declarado por los encuestados que manifestaron haber usado herramientas.....	112
Figura 28 - Percepción de la relación de las herramientas de monitorización de la actitud con la actividad profesional propia.....	114
Figura 29 - Respuestas a la pregunta “¿En qué crees que pueden ser útiles estas herramientas para tus tareas profesionales o académicas? (Marca todas las que procedan)”.	114
Figura 30 - Respuesta a la pregunta: ¿Crees que la información obtenida con herramientas de monitorización/análisis automatizado de la opinión de los usuarios en internet puede usarse para manipular a las personas? Marca la respuesta que mejor describa lo que piensas.....	115
Figura 31 – Resultado de la pregunta: Marca del 1 al 5 cómo de acuerdo estarías con la siguiente afirmación "El uso de herramientas de monitorización/análisis automatizado de la opinión de los usuarios en internet compromete la privacidad de las personas".....	116

Figura 32 - Resultado de la pregunta: Marca del 1 al 5 cómo de acuerdo estarías con la siguiente afirmación "Como usuario me siento amenazado por la existencia de este tipo de herramientas".....	116
Figura 33 – Respuesta a la pregunta: ¿En qué crees que pueden ser útiles estas herramientas a la sociedad en su conjunto? (Marca todas las que procedan).....	117
Figura 34 – Respuesta a la pregunta: Si alguna vez has accedido a información procesada con herramientas de monitorización/análisis de la opinión de los usuarios en Internet, ¿cuál era el objeto de la escucha? (Marca todas las que procedan).....	117
Figura 35 - Respuesta a la pregunta: si alguna vez has trabajado directamente con herramientas de monitorización/análisis para conocer la opinión de los usuarios de Internet, ¿cuál es tu nivel general de satisfacción con los resultados?.....	118
Figura 36 - Tipología de herramientas utilizadas por los encuestados.....	119
Figura 37 – Captura de pantalla que muestra el interfaz de Sentiment Analysis de Azure Machine Learning, integrado en Excel.....	131
Figura 38 – Ejemplo de aplicación de un modelo de aspectos para identificar atributos específicos en el texto, con Repustate.....	134
Figura 39 – Análisis de sentimiento por entidades, sobre lista predeterminada de intereses sectoriales, con Repustate (en este caso, opiniones de usuarios de hotel).	135
Figura 40 - Nube de palabras asociada a un proyecto de monitorización con Repustate. Este sistema de representación visual permite obtener una visión clara de los términos que más relacionan los usuarios con el concepto (palabras clave) objeto de monitorización (en este caso un servicio público).	136
Figura 41 – Ejemplo real de creación de audiencia con diversos filtros disponibles en BrandWatch durante las investigaciones realizadas.	141
Figura 42 – Personality Dimensions en Audiense	142
Figura 43 - Identificación de partes del mensaje con potencial polaridad con MeaningCloud.	144
Figura 44 – Muestra del análisis realizado por MeaningCloud, incluyendo detección de subjetividad e ironía.....	144

Figura 45 - Procedimientos de preentrenamiento y ajuste fino con BERT (Devlin et al., 2019, p. 3) . Se utilizan las mismas arquitecturas tanto en el preentrenamiento como en el ajuste. De este modo, los mismos parámetros del modelo preentrenado se utilizan para inicializar los modelos de las diferentes tareas descendentes. Durante el ajuste fino, se ajustan todos los parámetros.	145
Figura 46 – Esquema de la metodología propuesta (investigación propia, Anexo 4).....	166
Figura 47 – Ejemplo de dendrograma con algoritmo BMM (investigación propia, Anexo 4).	168
Figura 48 – Tag Cloud de perfilado según el caso desarrollado para esta tesis (Anexo 5)	175
Figura 49 – Trabajos científicos publicados entre 2020 y abril de 2022 referidos en alguna medida al Covid-19. Téngase en cuenta que el año 2022 no está completo (por la fecha de realización de esta tesis). Fuente: base de datos de Scopus	178
Figura 50 – Trabajos científicos publicados entre 2020 y abril de 2022 referidos en alguna medida al Covid-19 – Por tipo de documento - Fuente: Scopus	178
Figura 51 – Ejemplo de tuit recogido y procesado con la herramienta SentiOne (identidad visible en el original, se ha ocultado aquí por consistencia con las normas de la API de Twitter).....	180
Figura 52 – Palabras clave más repetidas en los 92695 textos analizados	181
Figura 53 – Top hashtags usados por los usuarios que realizan publicaciones con #yonomevacuno.....	182
Figura 54 – Uso del hashtag #yonomevacuno, distribución por género	182
Figura 55 – Uso del hashtag #yonomevacuno a lo largo del tiempo en todo Internet (redes sociales, blogs, foros, medios de comunicación), por género	183
Figura 56 – Alcance de las publicaciones con el hashtag #yonomevacuno.....	183
Figura 57 – Vista de detalle sobre intereses y profesiones de profesionales de la salud en España que hablan de Covid-19, con la herramienta BrandWatch	184
Figura 58 – Vista de una query con BrandWatch sobre profesionales de la salud que hablan de Covid-19	185

Figura 59 - Vista del dashboard de análisis de un proyecto en SentiOne. En la parte inferior izquierda del tuit mostrado, puede verse la polaridad asignada por la IA del sistema (positiva).....	237
Figura 60 – Fases del procedimiento aplicado.....	274
Figura 61- Dendograma BMM.....	278
Figura 62 – Interfaz de creación del dataset en Audiense.....	282
Figura 63 – Evolución de los tuits con el hashtag “yonomevacuno” en la horquilla temporal elegida, para un total de 5316 mensajes.	283
Figura 64 – A efectos comparativos, se muestra la evolución de las menciones a energía nuclear y centrales nucleares en todo el año 2021.....	283
Figura 65 – Criterios de filtro para la definición de audiencias en Audiense.....	284
Figura 66 – Afinado de audiencias por intereses en Audiense.....	285
Figura 67 – Perfilado de audiencias por intereses comunes, extraídos de las biografías..	286
Figura 68 – Informe de seguidores de una audiencia monitorizada en Audiense.....	286
Figura 69 – Listado de cuentas de una audiencia (los nombres y fotos han sido ocultados con una franja negra por confidencialidad).....	287
Figura 70 – Seguimiento de publicaciones de las personas que forman parte de la audiencia.....	287
Figura 71 – Polaridad de las publicaciones.....	288
Figura 72 – conexión de con Twitter Ads para impactar a los miembros de la audiencia perfilada con publicidad específica.....	288
Figura 73 – Proceso en aplicación de la Metodología PRISMA.....	291
Figura 74 – Foco de los trabajos analizados (temática).....	294
Figura 75 - Foco de los trabajos analizados (tipo de investigación).....	295
Figura 76 - Vista del panel de análisis de un proyecto en SentiOne.....	302

Nota de autoría: todas las ilustraciones han sido realizadas por el autor, salvo aquellas en las que se indica la fuente (referencia bibliográfica con nº de página).

Índice de tablas

Tabla 1 - Penn Treebank Part-Of-Speech (POS) tags - extraído de (Liu, 2012)	91
Tabla 2 - Ejemplo de matriz de confusión (confusion matrix).....	102
Tabla 3 - F-Score obtenido por Dolianiti et al. (2019) en su comparativa para un mismo dataset, a nivel de frase y de documento.....	107
Tabla 4 - Clasificación de las herramientas empleadas para la monitorización y el análisis de la actitud de los usuarios en Internet.....	109
Tabla 5 - Comprobación del uso “real” de herramientas entre los encuestados.....	112
Tabla 6 - Actividad profesional de los usuarios que son capaces de mencionar herramientas correctas.....	113
Tabla 7 - Principales herramientas de uso académico	124
Tabla 8 - Funcionalidades de Azure Cognitive Service for Language.....	130
Tabla 9 - Evaluación de rendimiento GLUE de modelos BERT con diferente inicialización, (Kovaleva et al., 2019 p. 4369).....	146
Tabla 10 - F-Score comparativo de los datasets analizados (Anexo 1).....	148
Tabla 11 - Precisión comparativa de los datasets analizados.....	148
Tabla 12 - Datasets - Tamaño y distribución por polaridad	151
Tabla 13 - Comparativa medias resultados por herramienta.....	152
Tabla 14 - Resultados de la investigación sobre número de entidades en mensajes de Facebook (Anexo 7).....	158
Tabla 15- Resultados de la investigación sobre número de entidades en mensajes de Twitter (Anexo 7).....	158
Tabla 16 - Cuantificación del anonimato en Twitter según estudio de Peddinti et al. (2014)	172
Tabla 17 - Distribución de la polaridad de los textos analizados con el hashtag #yomevacuno en Facebook (Anexo 3).....	176
Tabla 18 - Datasets - Tamaño y distribución por polaridad	240
Tabla 19 - F-Score comparativo de los datasets analizados	240

Tabla 20 – Precisión comparativa de los datasets analizados.....	241
Tabla 21 – Valores obtenidos por Ribeiro et al. con su método “Sentibench” para los mejores métodos evaluados, sobre diferentes datasets.	241
Tabla 22 – Benchmarking de Zimbra et al. (2018), mostrando exclusivamente valores para la métrica accuracy.....	242
Tabla 23 – Resultados del benchmarking realizado por Abbasi et al. (2014) para 15 herramientas “stand-alone”, sobre la métrica accuracy.....	243
Tabla 24 - Datasets – Tamaño y distribución por polaridad.....	263
Tabla 25 – F-Score comparativo de los datasets analizados para SentiOne	264
Tabla 26– F-Score comparativo de los datasets analizados para MeaningCloud.....	264
Tabla 27 – F-Score comparativo de los datasets analizados para Respustate.....	264
Tabla 28 – Recall comparativo de los datasets analizados para SentiOne.....	265
Tabla 29 – Recall comparativo de los datasets analizados para MeaningCloud.....	265
Tabla 30 – Recall comparativo de los datasets analizados para Repustate.	265
Tabla 31 – Accuracy comparativo de los datasets analizados para SentiOne.....	266
Tabla 32 – Accuracy comparativo de los datasets analizados para Repustate.....	266
Tabla 33 - Accuracy comparativo de los datasets analizados para MeaningCloud.	266
Tabla 34 – Benchmarking de Dolianiti et al. (2019) para varias herramientas analizadas (F-Score – sentence level).....	267
Tabla 35 – Valores obtenidos por Ribeiro et al. con su método “Sentibench” para los mejores métodos evaluados, sobre diferentes datasets.	268
Tabla 36 – Benchmarking de Zimbra et al. (2018) para la métrica accuracy.....	269
Tabla 37 – Resultados del benchmarking realizado por Abbasi et al. (2014) para 15 herramientas “stand-alone”, sobre la métrica accuracy.....	270
Tabla 38 – Comparativa medias resultados por herramienta.....	271
Tabla 39- Trabajos seleccionados.....	294
Tabla 40 – Recursos más empleados en los trabajos analizados.....	295
Tabla 41 – Número de tuits con múltiples entidades, dataset “Antivacunas Covid-19”	298
Tabla 42 - Número de tuits con múltiples entidades, dataset “Entretenimiento TV”	298

1 Introducción a la investigación

En este primer apartado, desarrollaremos las ideas básicas que conforman la justificación de la presente investigación, el propósito de la misma y, sobre todo, las preguntas de investigación y objetivos específicos a los que responderá la tesis, además de la metodología aplicada a tal fin.

1.1 Justificación de la investigación

La popularidad de las redes sociales ha hecho que la *world wide web*¹ deje de ser un depósito estático de información para pasar a convertirse en un foro dinámico y totalmente interactivo, con información que cambia y crece continuamente. Esta información es una extraordinaria fuente potencial de oportunidades, pero también de incertidumbres y amenazas para la sociedad.

De hecho, existe una creciente preocupación social y, por extensión, política, académica y periodística, por las consecuencias de la difusión de información potencialmente dañina en Internet, así como por la propagación de “discursos de odio” y otras formas de agresión inéditas hasta tiempos recientes, como el ciberacoso.

¹ La World Wide Web (WWW) es la red informática mundial que conforma el sistema mediante el que se gestiona la información que se comparte por Internet, a través del Protocolo de Transferencia de Hipertextos (HTTP).

Durante la pandemia del Covid-19, se ha puesto de relieve un nuevo desafío: detectar y actuar sobre las corrientes de opinión y falsas informaciones incompatibles con la ciencia y la defensa de la salud pública.

Como reflejo de esta preocupación por la actitud de los ciudadanos en Internet, con especial foco en las redes sociales, en los últimos años han aparecido nuevas figuras periodísticas relacionadas con la monitorización de contenidos, como las “agencias de verificación”, así como amplios desarrollos normativos y recomendaciones desde los gobiernos y las instituciones supranacionales. También en el ámbito académico se han multiplicado las iniciativas de investigación y desarrollo orientadas a entender este fenómeno desde distintos ángulos.

El descubrimiento (marzo 2018) de que Facebook permitió a la consultora Cambridge Analytica acceso ilimitado y no autorizado a la información de identificación personal (PII por sus siglas en inglés) de más de 87 millones de usuarios, ha alimentado un creciente interés en el debate sobre el impacto social de la tecnología y los riesgos para la privacidad de los ciudadanos (Isaak & Hanna, 2018). Es inevitable preguntarse si los gobiernos e instituciones poseen la capacidad de anticipar eficazmente el impacto futuro de las nuevas tecnologías en los derechos de la ciudadanía, así como en la posible manipulación de las informaciones con fines políticos. A este respecto, la comunidad científica está jugando un papel importante en el proceso de reflexión y en el aporte de soluciones.

No obstante, no todos los focos de interés se asocian a aspectos negativos y amenazas. La difusión en la Red de opiniones e informaciones es un campo de oportunidades para las empresas que desean conocer mejor a sus clientes y entender las narrativas subyacentes que afectan a sus marcas y valores. De igual manera, las instituciones públicas, docentes, y organizaciones políticas, necesitan escuchar e interpretar lo que se dice en las redes para tomar el pulso a la sociedad a fin de poder cumplir con eficacia su función y alcanzar sus objetivos.

Este amplio interés social por el estudio la actitud de la ciudadanía en Internet se ha trasladado a la comunidad científica, que está intentando desarrollar herramientas para dar una respuesta adecuada al nuevo desafío. Sin embargo, dado que estamos ante un fenómeno muy reciente, el *background*² académico es necesariamente corto. Tengamos en cuenta que la generalización en el uso de Internet en todos los ámbitos de la sociedad es un fenómeno que comenzó a eclosionar hace apenas 20 años, y que las principales redes sociales (Twitter y Facebook) se crearon hace poco más de una década. En consecuencia, en este campo novedoso de investigación y experimentación, nos enfrentamos por un lado a numerosas incertidumbres en el área científica, y por otro a una presión social creciente para el desarrollo de herramientas eficaces de análisis.

Como consecuencia de la transversalidad del fenómeno, no solo la comunidad científica se ha puesto manos a la obra con respecto a estos desafíos. En el ámbito de la empresa y las instituciones, desde la perspectiva de la Ciencia de Datos y la filosofía subyacente al *IoT*³ y el *Big Data*⁴, se viene manifestando un profundo interés en el acceso a grandes cantidades de información con fines analíticos, a fin de entender fenómenos específicos y, sobre todo, actuar sobre ellos. Esta inclinación hacia lo que podríamos llamar “análisis accionable” plantea futuros interrogantes (y

² Background. El término se refiere a los “antecedentes”, no en el sentido de una secuencia cronológica de investigaciones previas, sino en el sentido de la base de conocimientos académicos del arte en su conjunto, sobre la que se cimentan las nuevas investigaciones.

³ IoT: Internet of Things (Internet de las cosas), es un campo de la innovación que consiste en dotar a los objetos de una conexión autónoma a Internet para intercambiar datos en tiempo real. Un ejemplo típico son las farolas del alumbrado público, que informan vía Internet sobre aspectos como autodiagnóstico, consumo eléctrico, o datos del entorno captados con sus sensores.

⁴ Big Data: término que hace referencia al procesado masivo de grandes cantidades de datos, integrados desde distintas fuentes, generalmente con la finalidad de encontrar patrones que permitan explicar y predecir los fenómenos observados (habitualmente, conducta de los consumidores) a fin de tomar decisiones fundamentadas.

quizás amenazas) en diversas áreas sociales, como la política, los negocios, la educación o la convivencia, que solo pueden ser despejados mediante la investigación científica, aplicando un enfoque totalmente académico, riguroso y profundo. Más en concreto, se hace evidente la necesidad de estudiar las herramientas y metodologías disponibles para el análisis de la actitud de los ciudadanos en Internet, identificando sus límites, desafíos, y el uso que se hace actualmente de ellas (o que potencialmente se podría llegar a realizar, de acuerdo con el estado del arte).

1.1.1 Antecedentes

A fin de conocer qué investigaciones doctorales precedentes existen sobre la materia de esta tesis, se han realizado búsquedas exhaustivas en la Base de Datos de Tesis Doctorales del Ministerio de Educación (Gobierno de España)⁵.

Aplicando un criterio de búsqueda amplio que permita encontrar el mayor número de referencias posible, incluso tangenciales, se realizaron las siguientes búsquedas por título: contiene “Internet” y “actitud”, o “Internet” y “opinión”, o “redes sociales” y “opinión”, o “redes sociales” y “actitud”. En la mayoría de los casos no hubo resultados y, cuando los hubo, se trataba de investigaciones de otras disciplinas académicas distintas a la Ciencia de Datos y sin conexión con el procesamiento del lenguaje natural. Una búsqueda aún más amplia, centrada esta vez en cualquier tipo de tesis doctoral con referencia al “procesamiento del lenguaje natural” en su título, arrojó diversos resultados, pero ninguno de ellos planteaba objetivos coincidentes con la presente tesis.

⁵ Teseo: <https://www.educacion.gob.es/teseo/>

Por otro lado, si centramos la búsqueda en el concepto “huella digital”, solo aparece un único resultado con este término en el título, sin conexión con nuestros objetivos.

A fin de aplicar una perspectiva aún más amplia, abriendo el campo a un ámbito mundial, se realizó una búsqueda avanzada en SCOPUS⁶, aplicando el siguiente criterio: (TITLE ("internet" OR "social media") AND TITLE ("attitude" OR "opinion") AND KEY ("NLP" OR "machine learning" OR "natural language processing"))⁷. Como resultado, no se obtuvo ningún libro científico publicado. El número de artículos en revistas científicas y publicaciones de menor entidad para esta misma búsqueda ascendió a 61, más del 90% de los cuales se centran en aspectos muy concretos del arte, como es previsible en un documento breve. En el capítulo relativo al estado del arte, así como en el desarrollo de la investigación, nos referiremos con frecuencia a algunos de estos trabajos, en la medida que sean relevantes para los fines de esta tesis doctoral.

⁶ SCOPUS: <https://www.elsevier.com/solutions/scopus> . Esta base de datos cuenta en la actualidad con más de 84 millones de referencias, y es una fuente de búsqueda y referencia de uso generalizado en la literatura científica.

⁷ (TITLE ("internet" OR "social media") AND TITLE ("attitude" OR "opinion") AND KEY ("NLP" OR "machine learning" OR "natural language processing")) – Esta búsqueda apunta a obras que incluyan “internet” o “social media” en el título, además de “attitude” u “opinion”. Para excluir publicaciones relativas a otras disciplinas científicas no relacionadas con esta tesis, se añadieron los filtros “NLP” y “machine learning” en las keywords.

1.2 Fundamentos teóricos

No es posible afrontar un proyecto de investigación sin un estudio previo del marco teórico preexistente. En los siguientes apartados desarrollaremos los conceptos fundamentales que articularán el estudio, desde un punto de vista histórico, técnico, semántico y taxonómico.

1.2.1 Aproximación a los conceptos principales

El proceso que abarca la minería de opiniones (*opinion mining*, frecuentemente abreviado como OM) y el análisis de sentimiento (*sentiment analysis*, frecuentemente abreviado como SA) se define como la tarea de detectar, extraer y clasificar opiniones sobre una materia, involucrando el procesamiento de lenguaje natural (NLP por sus siglas en inglés) para hacer un seguimiento del estado de ánimo del público en relación un tema concreto (Saad & Saberi, 2017).

Algunos autores plantean una visión más amplia de la materia, pero sin apartarse en lo esencial del consenso, definiendo este campo como: “...todas las áreas de detección, análisis y evaluación del estado mental del ser humano con relación a diferentes eventos, problemas, servicios o cualquier otro interés. Más precisamente, este campo apunta a la extracción de opiniones, sentimientos y emociones basándose en la observación de las acciones de la gente que pueden ser capturadas usando sus escritos, expresiones faciales, verbales, música, movimientos...” (Yadollahi et al., 2017, p1).

Otros investigadores, como Pereira et al., prefieren definiciones más amplias limitándose a describir el análisis de sentimiento como una técnica usada para discernir opiniones y sentimientos en los textos (Pereira-Kohatsu et al., 2019).

Para la lingüista Maite Taboada, “The main goal of sentiment analysis is to determine whether a text, or a part of it, is subjective or not and, if subjective, whether it

expresses a positive or negative view. The direction of the opinion (i.e., whether positive or negative) is sometimes referred to as semantic orientation.⁸ (Taboada, 2016, p.326).

Por último, También podemos definir el SA, de una manera más enfocada, como el estudio de los sentimientos de las personas a nivel de entidad⁹ (Fang & Zhan, 2015).

1.2.2 Nomenclatura y debate semántico

Existe un consenso muy claro en la comunidad científica en torno al uso del término *opinion mining*, pero en el caso de *sentiment analysis* no es infrecuente encontrar la variante *sentimental analysis*. Una búsqueda de *sentiment analysis* y *sentimental Analysis* en Google Scholar arroja los siguientes resultados:

- *Sentiment analysis*: 149.000
- *Sentimental analysis*: 31.700

Por tanto, la variante más utilizada es *sentiment analysis* con una prevalencia del 82,5%.

En cuanto a las equivalencias para *opinion mining*, no observamos referencias relevantes a términos sinónimos, pero sí a dos términos afines, *opinion analysis* y *emotion mining* con los siguientes resultados en Google Scholar:

- *Opinion mining*: 1.810.000

⁸ “El objetivo principal del análisis de sentimientos es determinar si un texto, o una parte de él, es subjetivo o no y, si es subjetivo, si expresa una opinión positiva o negativa. La dirección de la opinión (es decir, si es positiva o negativa) a veces se denomina orientación semántica.”

⁹ En SA entendemos por “entidad” una parte del texto que hace referencia a un aspecto concreto de una opinión, sobre el que el autor podría tener un sentimiento distinto a otros aspectos de la misma frase. Ejemplo: “ADIF ha hecho un excelente trabajo, pero el servicio de RENFE es un desastre”.

- *Opinion analysis*: 1.730.000
- *Emotion mining*: 303.000

Si nos centramos en los 100 estudios más relevantes según los criterios definidos en el presente estudio, observamos que ninguno de ellos contiene el término *opinion analysis* en el título.

Por otro lado, es frecuente que algunos autores consideren *opinion mining* y *sentiment analysis* no como dos aspectos de una *task* (tarea) sino directamente como sinónimos (Del et al., 2016) (Yadollahi et al., 2017).

Cuando se aporta una visión de *opinion mining* y *sentiment analysis* como aspectos no intercambiables, hay consenso a la hora de asociar el primero al proceso de encontrar/extraer opiniones y el segundo al proceso de establecer su polaridad o valencia (positiva o negativa).

En definitiva, mientras para algunos autores *sentiment analysis* y *opinion mining* son sinónimos, otros establecen claras diferencias. En el segundo caso, esas diferencias se refieren a distintos aspectos de un mismo proceso, es decir, mantienen una estrecha relación entre ambos conceptos. Esta falta de homogeneidad puede confundir a la comunidad académica y refleja una evidente necesidad de estandarización en los protocolos del arte.

1.2.3 Ámbitos de aplicación y áreas del conocimiento relacionadas

La especialidad científica del *sentiment analysis*, generalmente basada en técnicas de inteligencia artificial (IA), es un conjunto de metodologías aplicadas en diversos campos de investigación, desde diferentes disciplinas académicas, y con distintos propósitos. Como veremos con sumo detalle más adelante, podemos identificar varios focos de interés para la comunidad, como la política, las marcas y productos, la salud, la educación, y cualquier tipo de fenómeno sociológico.

Se ha realizado un análisis de términos relacionados con ayuda de la herramienta Google Scholar¹⁰, para determinar la frecuencia de correlación entre *sentiment analysis* y otras áreas de conocimiento relevantes en la ciencia de datos. El algoritmo de búsqueda asocia *sentiment analysis* y cada una de las siguientes áreas en el número indicado de referencias (búsqueda estricta):

- Machine learning¹¹: 68.600
- Computational linguistics¹²: 41.800
- NLP¹³: 30.700
- Big data: 28.600
- Neural network: 28.200
- Deep learning: 21.500

Es evidente, por tanto, que estamos ante un campo de estudio que forma parte del ecosistema del procesamiento del lenguaje natural y la IA.

1.2.4 Desarrollo histórico y taxonómico

Aunque las raíces del análisis de sentimiento se remontan a los años 90, el 99% de la producción bibliográfica sobre la materia es posterior a 2004 (Mäntylä et al.,

¹⁰ <https://scholar.google.es/>

¹¹ Machine Learning, o Aprendizaje Automático, es un subcampo de las ciencias de la computación, englobado en la disciplina de la inteligencia artificial, que tiene por objetivo desarrollar técnicas que permitan que las máquinas aprendan a resolver problemas, mediante la construcción de modelos basado en datos.

¹² Lingüística computacional.

¹³ NLP: acrónimo de Natural Language Processing. En español, PNL o Procesamiento de Lenguaje Natural.

2018). Por tanto, estamos ante un marco de referencia temporal muy estrecho y, además, muy reciente.

Desde un punto de vista taxonómico, el contexto de la materia tratada se puede estructurar según se muestra en el siguiente esquema (Yadollahi et al., 2017, p. 2).

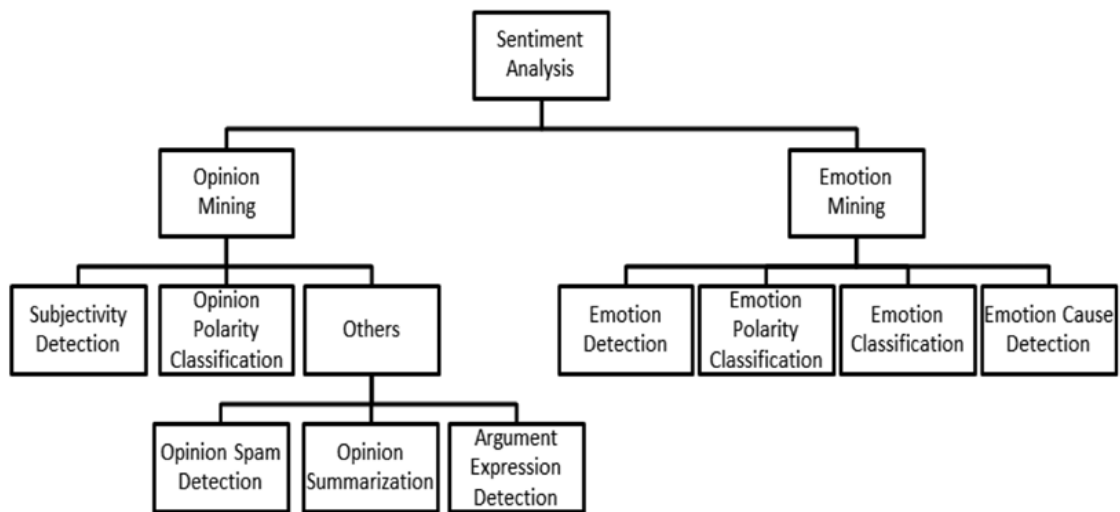


Figura 1 – Descomposición de la disciplina sentiment analysis (Yadollahi et al., 2017, p. 2)

Con independencia de la posición relativa (jerárquica o no) de los conceptos “análisis de sentimiento” y “minería de opinión”, toda la literatura analizada, así como las herramientas de análisis correspondientes, suponen una clara diferenciación entre dos esferas:

- Significado de la opinión: actitud del usuario
- Polaridad de la opinión: signo de las emociones asociadas a las actitudes

El siguiente cuadro resume el procedimiento en el que se contextualizan las tareas abarcadas por la minería de opinión y el análisis de sentimiento, tomando como ejemplo el caso de Twitter (Giachanou & Crestani, 2016b, p. 3):

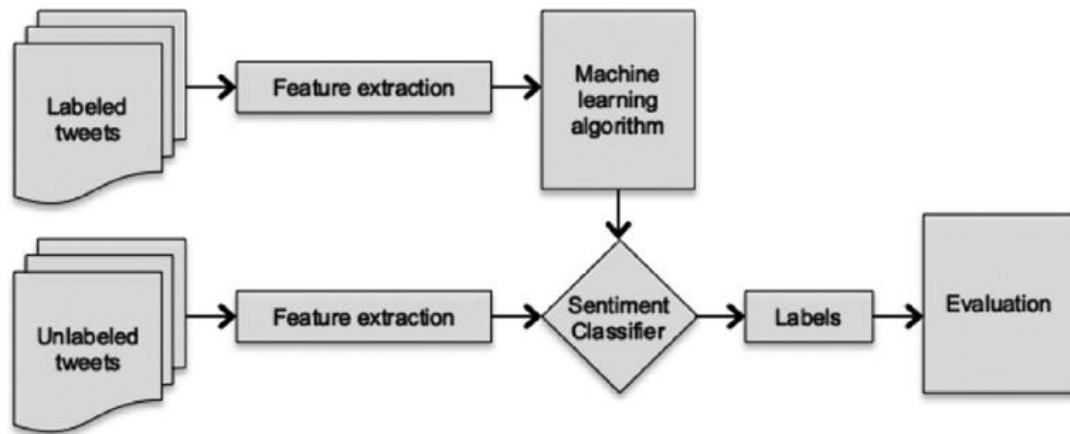


Figura 2 – Tareas abarcadas por la minería de opinión y el análisis de sentimiento (Giachanou & Crestani, 2016, p. 3)

Pero, ¿cómo podemos definir técnicamente una opinión? Según Giachanou & Crestani (2016) “an opinion is a quintuple $(e_i, a_{ij}, s_{ijkl}, h_k, t_i)$ where e_i is the name of an entity, a_{ij} is an aspect of e_i , s_{ijkl} is the sentiment on aspect a_{ij} of entity e_i , h_k is the opinion holder, and t_i is the time when the opinion is expressed by h_k .”¹⁴

A la hora de proyectar y ejecutar los estudios de minería de opinión y análisis de sentimiento, distintos autores defienden diferentes aproximaciones metodológicas y herramientas, como veremos en detalle más adelante, pero existe un consenso total, en la literatura analizada, sobre los tipos de tareas a realizar:

Extracción → Clasificación → Evaluación

¹⁴ “una opinión es un quintuple $(e_i, a_{ij}, s_{ijkl}, h_k, t_i)$ donde e_i es el nombre de una entidad, a_{ij} es un aspect de e_i , s_{ijkl} es el sentimiento del aspecto a_{ij} de la entidad e_i , h_k es el sujeto que opina, y t_i es el tiempo cuando la opinión es expresada por h_k .”

En la fase de **extracción**, se obtiene la data desde la fuente (aunque es frecuente utilizar sets de datos previamente extraídos y puestos a disposición de la comunidad académica). Los datos son a continuación procesados para su **clasificación**, normalmente mediante tecnologías de *machine learning*, y finalmente los resultados son **evaluados**.

1.2.5 El concepto “huella digital”

Al interactuar en Internet, todas las acciones que los ciudadanos realizan se traducen en datos, algunos de los cuales son potencialmente rastreables, ya sea de forma pública o por los prestadores de servicios. Esta “huella” puede ser activa o pasiva (Madden et al., 2007), dependiendo de si es o no deliberada por parte del usuario¹⁵.

Tradicionalmente, los organismos públicos e instituciones han venido mostrando especial interés en el aspecto pasivo de la huella digital, muy especialmente en la llamada “huella de dispositivo”, ya que es la que en mayor medida afecta a la protección de datos de carácter personal. El Comité Europeo de Protección de Datos, en su Dictamen 9/2014 sobre la aplicación de la Directiva 2002/58/CE, recoge la definición de la RFC¹⁶, que define la huella digital de dispositivo como un conjunto de elementos de información que identifican a un dispositivo o una instancia de aplicación. Por su parte, la Agencia Española de Protección de Datos (AEPD) amplía la definición poniendo el foco en el propósito del acceso a los datos: “La huella digital del dispositivo es una recopilación sistemática de información sobre un determinado dispositivo remoto con el objetivo de identificarlo, singularizarlo y, de

¹⁵ Un ejemplo de huella “deliberada” sería la publicación de opiniones o informaciones personales en una red social. En usuario es consciente de que está dando información públicamente accesible.

¹⁶ RFC 6973 Privacy Considerations for Internet Protocols: <https://www.rfc-editor.org/rfc/rfc6973>

esa forma, poder hacer un seguimiento de la actividad del usuario del mismo con el propósito de perfilarlo.”¹⁷

Sin embargo, el tipo de huella digital que afecta a nuestra investigación y donde se concentra gran parte de la actividad académica pertenece a una categoría distinta, la relativa a la huella digital “activa” o “active digital footprint” (Madden et al., 2007), que tiene que ver con las opiniones expresadas por los ciudadanos de en las redes. El acceso masivo y procesamiento de estas opiniones es lo que permite a la comunidad investigadora (y otros agentes) entender la actitud de los ciudadanos frente a los asuntos de trascendencia social y académica pasados y presentes.

1.2.6 Foco en la red social Twitter

Con más de 300 millones de usuarios activos en Twitter y 500 millones de tweets diarios¹⁸, esta red es una fuente prioritaria para la comunidad científica dedicada a la minería de opiniones. El hecho que la totalidad de los textos (exceptuando cuentas protegidas por sus usuarios¹⁹) sean accesibles a través de una *API*²⁰, facilita la recolección automatizada de literales, ya sea en tiempo real o retrospectivamente.

¹⁷ Estudio presentado en 2019 por la Agencia Española de Protección Datos “Fingerprinting o huella digital de dispositivo”, disponible en: <https://www.aepd.es/sites/default/files/2019-09/estudio-fingerprinting-huella-digital.pdf>

¹⁸ <https://about.twitter.com/company>

¹⁹ <https://help.twitter.com/es/safety-and-security/how-to-make-twitter-private-and-public>

²⁰ API: Application Programming Interface (interfaz de programación de aplicaciones). Se define como un conjunto de funciones y procedimientos que ofrece cierta biblioteca para que pueda ser accedida y utilizada de forma remota por otro software. En este caso, la API de Twitter permite el acceso programático a Twitter de forma avanzada, accediendo a los elementos principales de esta red, tales como: tuits, mensajes directos, usuarios y listas, entre otros.

Enlace: <https://developer.twitter.com/en/docs/twitter-api>

La mayor parte de los trabajos disponibles relacionados con la minería de emociones en Twitter se refieren al idioma inglés (Yadollahi et al., 2017).

En la revisión bibliográfica realizada para la preparación de esta tesis, se pudo comprobar que existen más de 2.000 trabajos publicados sobre *sentiment analysis* que llevan a Twitter referenciado en el título²¹. La misma prospección, realizada sobre otras redes, arrojó los siguientes resultados comparativos:

- Twitter: 2.152
- Facebook: 1.064
- Instagram: 241
- Reddit: 111
- LinkedIn: 0

Como revela la gráfica a continuación (Figura 3), hay una correlación inequívoca entre el interés social por el análisis de sentimiento y el crecimiento de la red social Twitter.

²¹ Prospección realizada con la base de datos de Scopus en julio de 2022

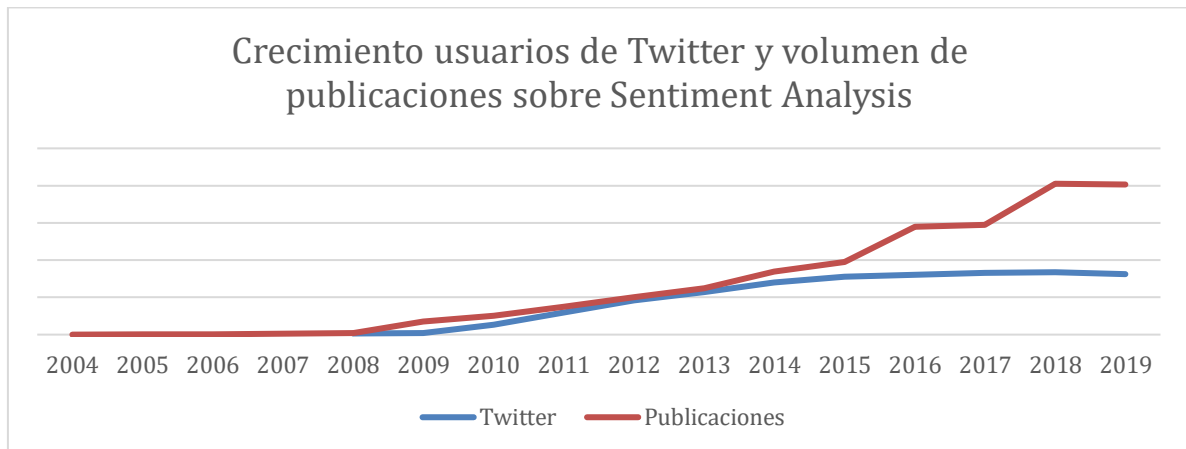


Figura 3 – Crecimiento de usuarios de Twitter y volumen de publicaciones sobre Sentiment Analysis. Gráfico elaborado a partir de datos de Twitter y Scopus.

En el caso de los trabajos focalizados en política y discurso de odio, el protagonismo de Twitter es aún mayor. Un 87% de los trabajos publicados sobre estos dominios para textos en lengua española citan como única fuente de literales a esta red social. En este apartado, destaca especialmente el trabajo pionero de Cuesta et al. (2014) sobre extracción masiva y análisis de textos de Twitter. Algo similar sucede con el dominio “covid-19”, otro de los focos de atención de la comunidad investigadora, como veremos.

1.2.7 Preferencias y limitaciones idiomáticas

La mayoría de los recursos más desarrollados han sido creados originalmente para el idioma inglés, lo cual es especialmente problemático en el caso de los *corpora* ²².

²² Corpora: plural de corpus (latín). Un corpus es un conjunto estructurado de textos en un formato legible para las máquinas, que se han producido en un entorno de comunicación humano natural, como por ejemplo una red social.

Los *corpora* anotados son cruciales para mejorar el rendimiento de los sistemas de minería de opinión (Sun et al., 2017).

Una circunstancia especialmente reseñable en la definición del marco teórico es el peso de los textos en español como objeto de estudio, ya que el idioma, como veremos, tiene un impacto considerable en los modelos metodológicos. Menos del 2% de los trabajos publicados sobre *sentiment analysis* y *opinion mining* hasta la fecha están referidos a la lengua española de forma específica, lo cual no correlaciona proporcionalmente con el peso del español como lengua en el contexto global. Según el Instituto Cervantes, un 6,3% de la población mundial habla español como lengua nativa (492.990.519 personas)²³. Desde esta perspectiva, podríamos decir que el español está infra referenciado en una proporción de 1 a 3 en los trabajos sobre *sentiment analysis* y *opinion mining*.

Si observamos la evolución en el ritmo de publicación de trabajos, centrándonos en los dominios más directamente relacionados con esta tesis, comprobamos que las publicaciones referidas al idioma español crecen rápidamente, pero a un ritmo más bajo (Figura 4).

²³ La lengua española en el mundo, en datos y gráficos - Datos actualizados el 9 de diciembre de 2021 - <https://www.epdata.es/datos/lengua-espanola-mundo-datos-graficos/513>

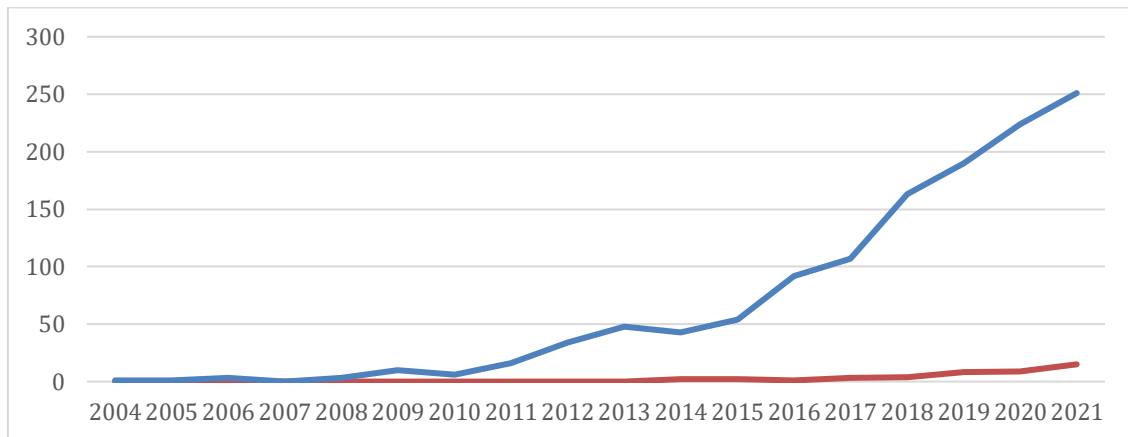


Figura 4 – Volumen de trabajos publicados sobre sentiment analysis referidos a política y discurso de odio en España (naranja) y total mundial (azul).

Además del español, podemos comprobar que otros idiomas también concentran la atención de los investigadores de forma relevante. A partir de 2016, el árabe (Ghallab et al., 2020) (Boudad et al., 2018), y en menor medida el chino mandarín (Peng et al., 2017) (Teo et al., 2020) y el ruso (Loukachevitch & Levchik, 2016), han cobrado especial fuerza. En una revisión bibliográfica sobre las técnicas en procesamiento de lenguaje natural para la minería de opinión, Sun et al. (2017) identifican ya en 2017 tres “kits” de herramientas para investigadores específicamente desarrollados para la lengua china.

1.2.8 Foco actual en *machine learning*

Las metodologías de clasificación de los textos predominantes en los orígenes del arte se basaban en *Lexicons*, listas de palabras previamente anotadas según su polaridad, con las que se comparan los documentos analizados mediante una serie de procesos. Los trabajos basados en este tipo de aproximación se han venido aplicando comúnmente en textos convencionales como blogs y reseñas de productos, pero están menos explorados en *micro blogging* en comparación con los métodos basados en aprendizaje automático (Giachanou & Crestani, 2016). Existe cierto consenso acerca de las limitaciones implícitas al hecho de trabajar con

Lexicons. De hecho, ya en fases muy tempranas, se señalaba este tipo de recurso como “necesario pero no suficiente” (Liu, 2012).

Las aproximaciones basadas en *lexicons*, o que utilizan este tipo de recurso en algún momento del proceso (hay metodologías “híbridas”) siguen siendo parte del marco teórico de los investigadores, pero han perdido mucho terreno frente a las metodologías basadas en *machine learning*. La principales diferencias entre ambas aproximaciones son básicamente dos: las metodologías de *machine learning* requieren uno o varios procesos de entrenamiento de los modelos, y generalmente obtienen mejores resultados en la predicción.

1.2.9 Interacción hombre-máquina

Los métodos de detección de opinión y sentimiento, cuando se trata de analizar la interacción hombre-máquina son muy escasos (Clavel & Callejas, 2016). El foco de la comunidad científica está puesto en el análisis de los textos, fundamentalmente en las redes sociales, pero desde disciplinas como el diseño de la Experiencia de Usuario (UX), que requieren un enfoque muy orientado a la interacción entre los usuarios y los sistemas, se necesitan soluciones metodológicas que están por llegar²⁴.

²⁴ En el Anexo 8 se propone un enfoque metodológico inédito basado contextual inquirí para resolver uno de los desafíos de esta tesis.

1.2.10 Predominio de las herramientas de código abierto

La tendencia general en la comunidad apunta al desarrollo de herramientas *open source*²⁵ (Ribeiro et al., 2016). Esto sin duda facilita el acceso de la comunidad a la investigación y a futuros desarrollos. En particular, hay herramientas y librerías especialmente utilizadas, como Freeling²⁶ para procesos de normalización léxica en textos en español y otros idiomas, SentiStrength²⁷ para determinar la polaridad del sentimiento (positiva o negativa) y Stanford NLP²⁸ para procesamiento del lenguaje natural mediante inteligencia artificial, entre muchas otras. No obstante, a medida que el desarrollo de aplicaciones se va normalizando y escapando al ámbito académico, no han dejado de aparecer recursos “de pago”, en algunos casos desde las áreas de producto basadas en IA de los grandes gigantes tecnológicos, como IBM y Microsoft. También han proliferado las herramientas informales, cuya eficacia estudiaremos a fondo, al alcance de todo tipo de usuarios por su bajo coste.

1.2.11 Propósito y perspectivas para la investigación en el contexto actual

Los propósitos y perspectivas que dan forma al marco teórico actual vienen conducidos por dos grandes ejes, uno puramente técnico y otro de utilidad social. Por un lado, se hace evidente un interés en los métodos por sí mismos (por ejemplo, rendimiento comparativo de modelos), y por otro, una preocupación intensa por su

²⁵ Open source: código abierto. Es un modelo de desarrollo de software que se basa en la colaboración abierta (Levine & Prietula, 2014), de modo que la comunidad es libre de editarlo y adaptarlo. Open source no es sinónimo de “gratis”, pero este tipo de software generalmente lo es.

²⁶ <https://nlp.lsi.upc.edu/freeling/>

²⁷ <http://sentistrength.wlv.ac.uk/>

²⁸ <https://nlp.stanford.edu/>

adecuación a objetivos concretos de aplicación práctica (por ejemplo, monitorización de la actitud con relación a la pandemia del Covid-19).

Otra característica especialmente crítica para definir el marco teórico, desde el punto de vista del propósito de los investigadores, es la extrema rapidez de los cambios. Las herramientas a disposición de la comunidad para conocer las actitudes de los usuarios sobre temas de interés social, mediante técnicas de minería de opinión y procesamiento de lenguaje natural, no han parado de evolucionar a gran velocidad. Este cambio continuo conlleva una rápida obsolescencia de los trabajos publicados, difumina el marco teórico en comparación con otras disciplinas más consolidadas, y alude a un necesario propósito de estandarización metodológica como perspectiva de futuro.

En cuanto a los propósitos concretos de utilidad social sobre los que se aplican las metodologías de análisis, ya hemos visto en este capítulo su gran diversidad: abarcan todos los ámbitos de la sociedad, desde la educación a la salud, pasando por la política y la empresa. No obstante, en los últimos años han surgido nuevas problemáticas especialmente conflictivas para la opinión pública, con el consiguiente reflejo en las redes sociales, que preocupan seriamente a los investigadores del comportamiento y la actitud digital, como la pandemia del COVID-19 y los “discursos de odio”. En torno a estas problemáticas encontramos dos aspectos que están siendo objeto de atención especial, como la polarización política y, sobre todo, las *fake news*²⁹.

Mirando al futuro, en un contexto de riesgo de pandemia, las noticias falsas y la polarización extrema pueden marcar un antes y un después también desde el punto

²⁹ Fake News: noticias falsas. “Las noticias falsas, también denominadas fake news, son aquellas informaciones carentes de veracidad que tienen como objetivo llegar a un gran número de personas y generar desinformación.” (Peiró, 2020)

de vista de las metodologías de investigación y las estrategias de aproximación a los problemas por parte de la comunidad investigadora.

1.3 Objetivos

1.3.1 Objetivo general

El objetivo principal de este proyecto de investigación es aportar información **nueva y objetiva** a la comunidad profesional y académica, basada en la **investigación científica**, sobre el estudio de la actitud de los usuarios a través de la **huella que deja su interacción en Internet**, con especial foco en las **redes sociales**, teniendo en cuenta el contexto actual. Se pretende así construir y fundamentar un cuerpo de conocimiento libre, novedoso y transversal que pueda servir de base para otros investigadores de diversos ámbitos, tales como el diseño de la experiencia de usuario, la innovación tecnológica y el marketing digital.

1.3.2 Objetivos específicos y preguntas de investigación

Partiendo de este objetivo general, se plantean los siguientes objetivos específicos, alineados con las preguntas de investigación, que ayudarán a formular la hipótesis posterior:

- **Objetivo específico 1: determinar la posibilidad de acceso a la opinión e informaciones publicadas por los usuarios en Internet.** ¿Es accesible la información? ¿con qué limitaciones? ¿se puede recopilar masivamente? ¿es posible identificar asuntos de interés específicos?
- **Objetivo específico 2: conocer las posibilidades de procesamiento de la información.** ¿Con qué nivel de granularidad se puede realizar el

procesamiento y análisis? ¿qué se puede llegar a conocer con el análisis de los datos?

- **Objetivo específico 3: establecer la posibilidad de monitorización.** ¿es posible realizar seguimientos continuados y en tiempo real sobre temas concretos?
- **Objetivo específico 4: determinar la posibilidad de perfilado y vinculación de actitudes a personas concretas.** ¿Se puede estudiar a usuarios concretos? ¿qué se puede saber sobre ellos? ¿se pueden crear arquetipos por atributos o intereses comunes de un segmento de usuarios?
- **Objetivo específico 5: medir la precisión y exactitud en el análisis.** ¿Cuál es el rendimiento real que podemos esperar con los métodos y herramientas actuales? ¿Cómo de eficaces son los recursos que están “a disposición de cualquiera” en comparación con los métodos más avanzados de la comunidad científica?
- **Objetivo específico 6: analizar el volumen y coste de los medios requeridos.** ¿Qué medios técnicos son necesarios para acceder a la información y procesarla? ¿qué herramientas del conocimiento se precisan? ¿cuáles son los costes involucrados?
- **Objetivo específico 7: delimitar el acceso a los medios requeridos.** ¿Quién puede acceder a la información y a su procesamiento para conocer efectivamente la actitud de los ciudadanos en Internet?
- **Objetivo específico 8: determinar la capacidad de influir.** ¿Pueden las herramientas de monitorización y análisis de la actitud de los ciudadanos en Internet ser utilizadas para influir en la toma de decisiones de las personas, directa o indirectamente?
- **Objetivo específico 9: establecer el Estado del Arte – obstáculos y desafíos de la comunidad investigadora.** ¿En qué punto se encuentra el estado del arte? ¿cuáles son los principales desafíos a los que se enfrentan los

investigadores? ¿Qué desafíos específicos tenemos por delante para el análisis en lengua española?

- **Objetivo específico 10: analizar el impacto de la pandemia del Covid-19 en el tema de estudio.** ¿Ha influido la pandemia en el estado del arte? ¿Cuál ha sido el impacto, si lo ha habido, en la comunidad investigadora?

1.3.3 Enfoque de la investigación alineado con los objetivos

Para cumplir con los objetivos indicados, será preciso desarrollar **una investigación exhaustiva sobre los métodos y herramientas actualmente disponibles**, ya sean **académicos o comerciales**. No será suficiente un enfoque descriptivo. Esta investigación deberá tener un componente empírico y todo el apoyo experimental necesario para arrojar luz sobre las áreas donde no se disponga de datos de estudios previos.

A fin de mantener la consistencia científica y una arquitectura de la información clara, **el apartado de conclusiones se articulará sobre los objetivos descritos**, dando respuesta a cada uno de ellos de forma sistemática.

1.4 Hipótesis

Este proyecto de investigación se basa en la siguiente premisa-hipótesis fundamental, a partir de las preguntas de investigación:

Cualquier persona u organización con conocimientos técnicos avanzados de marketing digital está potencialmente capacitada para generar y procesar información exhaustiva sobre la actitud de los usuarios en Internet, ya sea con fines puramente analíticos o con intención de influenciar procesos de formación de opinión y toma de decisiones. Para este fin, es preciso utilizar herramientas tecnológicas y metodologías de distinta naturaleza, precisión y alcance, que están en continua transformación.

Hablamos de “cualquier persona u organización”, y a este respecto es importante considerar varios aspectos clave en los que profundizar en la validación de la hipótesis:

- Los conocimientos y herramientas necesarios para acceder a los datos y procesarlos.
- El tipo de información accesible.
- Los límites técnicos y metodológicos existentes.

Otro aspecto esencial apunta a la finalidad última de este acceso y procesamiento de datos, ya que ésta puede llevar asociados distintos escenarios técnicos y normativos. Por este motivo distinguimos en la premisa-hipótesis dos espacios: el mero análisis, habitual en disciplinas como la investigación de mercados, UX o sociología, y la influencia sobre procesos de formación de opinión o toma de decisiones, más propios de la política y el marketing.

1.5 Metodología

En este apartado se describe con detalle el enfoque metodológico aplicado en el desarrollo de la presente investigación, de acuerdo con el siguiente esquema estandarizado:

- Procedimiento.
- Medios.
- Métodos.

Como veremos a continuación, se han utilizado en distinta medida todas las categorías de recursos habituales en la investigación científica de la especialidad, ya sea bibliográficos, técnico/experimentales, participativos, o formativos, con el fin de alcanzar los objetivos planteados y responder a las preguntas de investigación sin lagunas metodológicas. En los apartados siguientes se detalla de forma precisa cada uno de estos aspectos.

1.5.1 Procedimiento

De acuerdo con los procedimientos del método científico habituales en la especialidad, partimos del siguiente esquema para el desarrollo de los trabajos, en orden cronológico:

1. Revisión bibliográfica.
2. Contextualización sectorial mediante trabajo etnográfico previo.
3. Determinación del estado del arte y del marco teórico.
4. Definición de objetivos y preguntas de investigación necesarios para validar la hipótesis.
5. Planificación de los trabajos de investigación.
6. Desarrollo de la investigación para cada aspecto que debe ser objeto de estudio.

7. Documentación sistemática de los resultados.
8. Desarrollo de las conclusiones.
9. Validación o rechazo de la hipótesis.

1.5.2 Medios

Siguiendo el procedimiento descrito en el punto anterior, el desarrollo de la investigación ha requerido el acceso a los medios siguientes:

- Recursos bibliográficos: literatura científica incluida en el apartado “Referencias”.
- Documentación técnica: manuales, guías y pliegos de especificación técnica de las herramientas analizadas.
- Recursos formativos: formación específica en materias concretas, de carácter académico (cursos y seminarios de doctorado) y extraacadémico (estudio y entrenamiento en el uso de aplicaciones técnicas analizadas).
- Acceso a herramientas académicas de *sentiment analysis* y procesamiento de lenguaje natural.
- Uso experimental de aplicaciones comerciales de *social listening*³⁰.
- Herramientas auxiliares para investigadores: test con usuarios, análisis estadístico y presentación visual.
- Acceso a recursos de Twitter para desarrolladores.
- *Scholar Peer Counseling*: participación como miembro del Grupo de Investigación de Ciencia de Datos de UNIR, bajo la dirección del Dr. Luis de la Fuente.

³⁰ Social Listening: “Escucha Social”. En marketing y comunicación, se usa este término para referirse a las labores de monitorización de los comentarios en las redes sociales sobre marcas, productos, sectores y tendencias.

- *Private Peer Counseling*: participación en proyectos de investigación privados de Psyma Ibérica³¹.
- Bases de datos de publicaciones académicas y sus herramientas estadísticas asociadas.

1.5.3 Métodos

Como veremos a continuación, ha sido preciso aplicar una amplia variedad de métodos para el desarrollo de la investigación, más de los inicialmente previstos. Esto se debe a que muchos de los aspectos a cubrir, por su novedad, han requerido una arquitectura compleja de fuentes primarias que supliera la falta de información previamente disponible. A continuación, se enumeran todos los métodos empleados:

- **DESK RESEARCH**
Estudio de informaciones, estadísticas y *White Papers*, con objeto de documentar objetivamente el contexto social y justificar la necesidad y utilidad de este proyecto de investigación, así como fijar un enfoque de partida sólido.
- **BIBLIOGRAPHIC RESEARCH**
Recopilación, acceso y estudio de la bibliografía existente para determinar el estado del arte y adquirir un conocimiento profundo del dominio, desde el punto de vista teórico.
- **SYSTEMATIC REVIEW**
Con fines de publicación en revista especializada, y como parte de la investigación para esta tesis, se realizó una revisión sistemática de aspectos

³¹ Psyma Ibérica forma parte de PSYMA GROUP, un grupo internacional de investigación de mercados nacido en Alemania en 1957. <https://www.psyma.com/es/>

especialmente relevantes para la tesis, cuyo estado del arte aún no había sido definido en la literatura disponible.

- **ETNOGRAFÍA / ENTREVISTA (CONTEXTUAL INQUIRY)**
Entrevistas y observación en el contexto con profesionales, a fin de evaluar aspectos específicos relativos al uso de las herramientas sin recurrir a fuentes secundarias.
- **ENCUESTA**
Aproximación estadística al conocimiento de determinados aspectos no disponibles en la literatura científica, sobre la relación que se establece sobre las aplicaciones de análisis de la actitud digital de los usuarios y los usuarios de las mismas (en cualquier ámbito).
- **LAB RESEARCH**
Análisis en profundidad de las herramientas, métodos e información disponibles, con foco en cada pregunta de investigación y objetivos.
- **PRODUCT TRIAL & WALKTHROUGH**
Recorrido práctico por los sistemas, herramientas y métodos, con especial atención a las tecnologías más utilizadas e innovadoras, a fin de obtener la información necesaria para alcanzar y documentar resultados.
- **EXPERIMENTAL RESEARCH**
Se han desarrollado varios trabajos de investigación novedosos (incluidos en los anexos de la tesis) de componente experimental, aplicando metodologías nuevas o enfoques de utilidad, teniendo en cuenta el estado del arte y las necesidades de la comunidad investigadora. En concreto, se han realizado estudios comparativos y experimentos de alcance práctico relativos a la funcionalidad de las herramientas, además de un trabajo para prospectar una metodología complementaria al estado del arte que ilustra uno de los puntos débiles de la investigación de la opinión de los usuarios en Internet (presentado en congreso).

- PARTICIPATION RESEARCH

Realización de diversas pruebas técnicas y colaboración en proyectos de iniciativa privada, a fin de ampliar la base de información.

1.6 Estructura de la tesis

Para la redacción de la presente tesis doctoral se ha seguido un modelo estructural canónico, acorde a los criterios oficiales de UNIR, organizado en cuatro partes: Introducción, Contexto y Estado del Arte, Desarrollo de la investigación, y Conclusiones. Cada una de ellas se articula en una serie de apartados, cuyo propósito se expone con detalle a continuación.

1.6.1 Contenido de la Introducción

La introducción consta de cuatro capítulos que constituyen el punto de partida de la investigación, con especial foco en la justificación y en la definición de la hipótesis.

- **Justificación de la investigación:** aquí se describen las razones objetivas por las que se considera esta investigación como necesaria y útil para la comunidad académica y la sociedad en su conjunto.
- **Fundamentos Teóricos:** este apartado desarrolla los conceptos fundamentales que articularán el estudio, desde un punto de vista histórico, técnico, semántico y taxonómico, siempre con una justificación objetiva basada en la investigación previa.
- **Objetivos:** en este apartado se establece un objetivo general para la investigación, seguido de un desarrollo sistemático de objetivos específicos, además de establecer las preguntas de investigación que se plantean como punto de partida inicial para el trabajo.
- **Hipótesis:** aquí definimos, en sus términos exactos, la hipótesis que deberá ser validada o desechada en las conclusiones finales.
- **Metodología:** en este apartado se describe la metodología empleada en el desarrollo de la investigación, teniendo en cuenta procedimientos, medios y métodos. El objetivo de este apartado es facilitar a otros investigadores una

comprensión profunda del camino recorrido por el investigador, que debería posibilitar una reproducción de la investigación con los mismos resultados.

1.6.2 Contenido del estado del arte

En este segundo apartado, se establece el contexto que enmarcará toda la investigación, partiendo de un análisis profundo de la literatura científica disponible, teniendo muy en cuenta las inquietudes más relevantes de la comunidad investigadora.

- Campo de estudio: origen, evolución y contextualización temporal
- Limitaciones y desafíos del estado del arte
- Espacios de interés académico

1.6.3 Contenido del desarrollo de la investigación

En este apartado se aplican de forma sistemática la metodologías elegidas sobre el campo de estudio, con los objetivos específicos que han sido establecidos previamente, a fin de obtener la información necesaria que nos permitirá responder después responder a las preguntas de investigación y validar (en su caso) la hipótesis, dentro de los límites del marco teórico previamente fijado. Los contenidos de este apartado son los siguientes:

- **Investigación sobre los métodos:** este capítulo tiene por objeto responder a los objetivos específicos relacionados con las metodologías más actuales a disposición de la comunidad investigadora, incluyendo aspectos como su diversa base tecnológica y su precisión.
- **Investigación sobre las herramientas:** tanto los investigadores como el público (usuarios no académicos) necesitan herramientas para acceder a la

información y procesarla, ya sea con un método o de una manera informal. En este apartado se obtiene la información necesaria para responder a los objetivos planteados al respecto. Adicionalmente, se estudiará la cuestión también desde una perspectiva lingüística, centrándonos en los textos de habla hispana.

- **Investigación sobre los recursos disponibles para el desarrollo de nuevos métodos:** los dos capítulos anteriores estarían incompletos si no prestamos también atención a otro tipo de recursos que usan habitualmente los investigadores para crear sus propias herramientas o diseñar nuevas metodologías. Este capítulo resuelve ese importante aspecto.
- **Investigación sobre los límites y desafíos actuales en el dominio:** en este último capítulo del apartado, partiendo de la información reflejada en los anteriores, se desarrollan las cuestiones relativas a la capacidad real que tenemos en este momento para analizar la actitud de los ciudadanos en la Red, con sus límites, sus obstáculos y su proyección futura.
- **Investigación sobre posibilidades de identificación, *profiling* y manipulación:** un aspecto esencial de esta tesis es alcanzar un conocimiento cierto no solo del alcance de los recursos disponibles para la comunidad, sino también del potencial uso práctico de estos medios para influir en la población, ya sea con una intencionalidad legítima o ilegítima.

Adicionalmente, se desarrollará un análisis del impacto de dos aspectos esenciales del contexto social actual que afectan a la comunidad investigadora:

- **Impacto de la pandemia del Covid-19:** A fin de lograr una contextualización plena, se ha considerado esencial incluir un apartado con especial atención al impacto de la pandemia del Covid-19 en la literatura científica.
- **Impacto de la inquietud social sobre el discurso de odio y otros aspectos de actualidad política:** además del Covid-19, hay aspectos especialmente relevantes para esta tesis que tienen un gran impacto en el estado del arte y

que involucran cuestiones de gran utilidad social y académica, como el *hate speech* y la manipulación de la opinión política. Todos ellos son cubiertos en este tercer capítulo del estado del arte.

- **Accesibilidad de los métodos a personas fuera del ámbito académico:** No sería posible responder la pregunta de investigación sin comprobar empíricamente, y también por otros medios, la accesibilidad de usuarios “no cualificados” al estudio de la actitud de los ciudadanos en Internet.

1.6.4 Contenido de las conclusiones

El paso siguiente en el proceso de la tesis es la reflexión sobre los resultados obtenidos, así como su análisis con relación a la hipótesis planteada, a fin de tener los elementos de juicio objetivos que son necesarios para llegar a una conclusión. Con esta finalidad, el apartado se articula dando respuesta, una a una, a todas las preguntas de investigación.

Validación de la hipótesis: en este subapartado final, se decide y justifica la validación o rechazo de la hipótesis. La reflexión se articula en torno a los objetivos específicos, a fin de asegurar una cobertura íntegra de los distintos aspectos incluidos en el *scope* del estudio.

1.6.5 Contenido de la bibliografía

La bibliografía referencia los 132 trabajos académicos que han sido estudiados para la realización de esta tesis doctoral.

1.6.6 Contenido de los anexos

Como información complementaria, pero esencial para la obra, se incluye documentación de los estudios realizados y otros datos y materiales de referencia que pueden ser necesarios para una plena comprensión de los trabajos realizados en esta tesis. En total, este apartado consta de 8 anexos.

2 Contexto y estado del arte

El estudio de las metodologías profesionales para el análisis del comportamiento y la actitud de los usuarios en Internet, y muy en especial en las redes sociales, no puede concebirse sin una contextualización previa desde un punto de vista tecnológico, metodológico, social y académico. Para ello, como parte de la investigación realizada para esta tesis, se ha llevado a cabo un estudio en profundidad del estado del arte y una prospección de la tendencia y la evolución del dominio a lo largo del tiempo.

A continuación, se expondrán con detalle estos aspectos, dotando a las preguntas de investigación de esta tesis de una contextualización precisa.

2.1 Campo de estudio: origen, evolución y contextualización temporal

Aunque las raíces del análisis de sentimiento se remontan a los años 90 del siglo pasado, el 99% de la producción bibliográfica sobre el tema es posterior a 2004 (Mäntylä et al., 2018). Si ponemos el foco en los últimos 5 años, encontramos un total de 15.929 trabajos publicados³². La misma prospección, referida a los cinco años anteriores a ese período, arroja la cifra de 5.575, lo que supone un ritmo de producción en torno a un 300% menor que el actual. Esto nos da una idea precisa del creciente interés científico por la especialidad.

³² TITLE-ABS-KEY ("sentiment analysis" OR "opinion mining") AND PUBYEAR > 2016 AND PUBYEAR < 2022

El siguiente gráfico refleja claramente la evolución desde los orígenes, con un brusco incremento en 2018³³ :

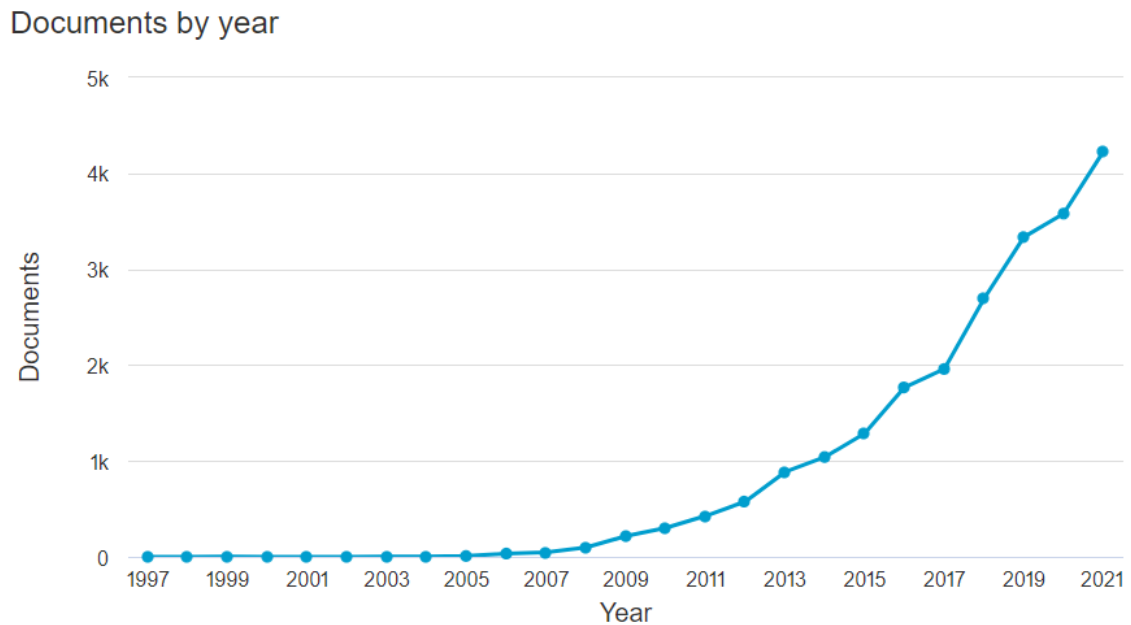


Figura 5 - Trabajos publicados sobre sentiment analysis y opinion mining hasta 2021. Fuente Scopus.

A la vista de este gráfico, resulta evidente la extraordinaria “juventud” de este campo de investigación. Hay una correlación muy interesante, por otra parte, entre el “despegue” de la materia y la fecha de creación de las principales redes sociales (Facebook 2004 y Twitter 2006).

³³ Formato de consulta: (TITLE-ABS-KEY ("sentiment analysis") OR TITLE-ABS-KEY ("opinion mining")) AND PUBYEAR < 2022)

En el caso de investigaciones dirigidas específicamente a contenidos españoles, la producción científica comienza más tarde. Por tanto, nos encontramos ante un marco temporal muy estrecho desde el punto de vista mundial, y aún más estrecho desde el punto de vista del idioma español. Si realizamos la misma consulta de la figura 5 limitándonos a trabajos orientados a la lengua española, obtenemos el siguiente resultado:

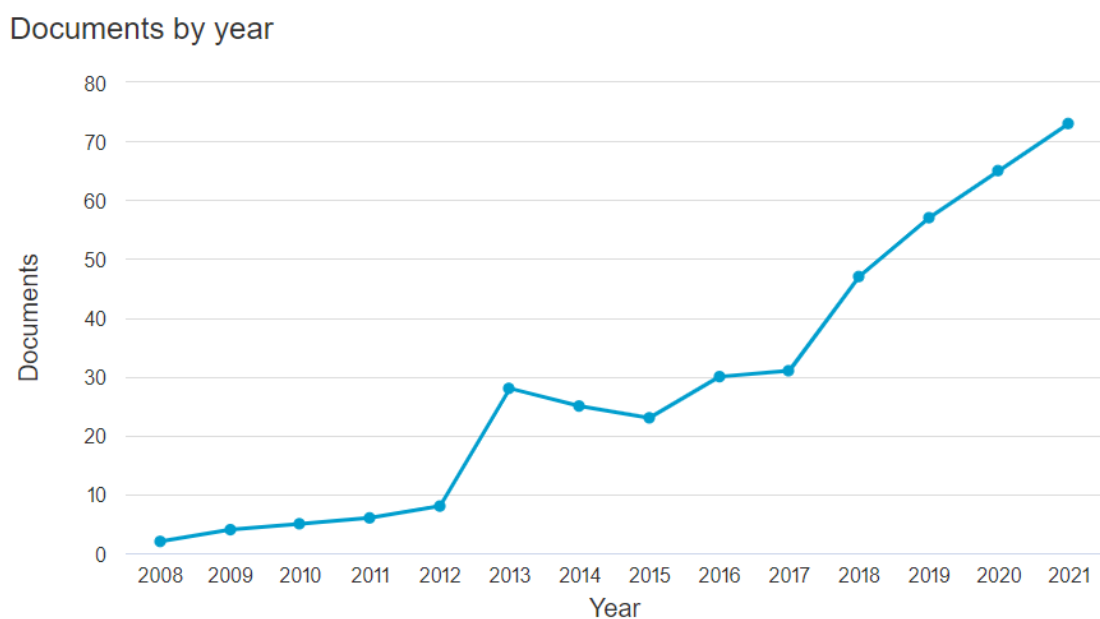


Figura 6 - Trabajos publicados sobre sentiment analysis y opinion mining hasta 2021, para cualquier área temática, referidos específicamente a la lengua española. Fuente Scopus.

2.2 Limitaciones técnicas y desafíos en el estado del arte

Como hemos visto, el estudio del comportamiento y la actitud de los usuarios en Internet, con especial foco en el ecosistema *social media*, es una materia de creciente interés en diferentes sectores académicos y profesionales, tales como el diseño de la experiencia de usuario, la educación, la innovación tecnológica, el marketing digital, la investigación de mercados, la política, la sociología e incluso las ciencias de la salud. Sin embargo, esta disciplina está lejos de superar severas dificultades de muy distinta naturaleza. De acuerdo con la investigación bibliográfica realizada, podemos resumir los obstáculos y retos de la comunidad en cuatro grandes ejes:

- Limitaciones técnicas de las herramientas de recopilación y análisis de los datos.
- Metodologías y literatura científica de muy reciente desarrollo, y por tanto poco consolidadas/desarrolladas.
- Riesgo de obsolescencia a muy corto plazo de los trabajos publicados.
- Progresos limitados en técnicas auxiliares de la ciencia de datos.

Por otro lado, aparte de los aspectos puramente técnicos, la prospección de la actitud y el comportamiento de la población en Internet, especialmente cuando afecta a materias políticas o socialmente “sensibles”, está fuertemente condicionada por “líneas rojas” de carácter legal, además de importantes sesgos ideológicos y culturales, producto de un creciente interés social.

2.2.1 Rendimiento y *benchmarks*

Una de las áreas científicas más correlacionadas con el *sentiment analysis* en la literatura reciente es el *machine learning* (y especializaciones subsidiarias, como *deep learning*). Sin embargo, los autores más citados insisten en que de momento los resultados obtenidos son limitados (Sun et al., 2017), lo que deja espacio a la investigación y optimización de las metodologías/tecnologías existentes. Si unimos

esto al estrecho marco temporal analizado en el apartado anterior, podemos considerar que el estado del arte se encuentra aún en una fase muy temprana, con mucho terreno por delante. Especialmente revelador es el dato aportado por Zimbra et al. (2018), que apunta a una precisión a menudo por debajo del 70%³⁴. Sin embargo, para ilustrar la rápida evolución de la disciplina, no podemos olvidar que, tan solo un año después del trabajo de estos autores, hacían irrupción en el panorama académico tecnologías que elevaban el *benchmark* por encima del 80% para ciertas tareas (Devlin et al., 2019).

Dos aspectos que preocupan a los investigadores en los trabajos más recientes son la interpretación del contexto y el análisis de la polaridad a nivel de entidad³⁵. Según Pereira-Kohatsu et al. (2019), las *surface forms* (enfoques basados en el uso de *lexicons* y *machine learning*) han funcionado notablemente bien para muchos problemas de procesamiento del lenguaje natural, pero no son capaces de explicar la semántica de las palabras. Los métodos basados en incrustaciones de palabras mediante el entrenamiento de una red neuronal permiten acercarse a este objetivo. Criado y Villodre (2018) concluyen que los sistemas actuales basados en el aprendizaje automático aún son incapaces de capturar todas las vicisitudes de los contextos. Esto es un problema general no resuelto en el ámbito de la IA: la interpretación del contexto es el mayor abismo actual entre el cerebro humano y las máquinas, y el terreno donde se avanza más despacio. Pla y Hurtado (2014) sostienen que obtener la polaridad a nivel de entidad introduce una complejidad adicional al análisis de sentimiento, porque primero se debe determinar la parte del

³⁴ "State-of-the-art TSA approaches continue to perform poorly, with reported sentiment classification accuracies typically below 70%" (Zimbra et al., 2018, p. 24)

³⁵ Ejemplo: frases con dos entidades distintas, cada una con una polaridad diferente, como "el CEO es un desastre y lo está haciendo todo mal, pero hay un nuevo partido en la oposición con grandes ideas para solucionar los problemas de la gente")

texto que hace referencia a cada una de las entidades. Este es un problema (dentro de otro problema) aún no resuelto en procesamiento del lenguaje natural.

Por otra parte, algunos autores también identifican desafíos importantes de otra naturaleza, incluso antes de aplicar metodologías de análisis a los textos. Estos retos tienen que ver con el descubrimiento de temas y con la recopilación de datos (Stieglitz et al., 2018).

Si ponemos el foco en el análisis desde la perspectiva de la interacción hombre-máquina, observamos que los pocos recursos existentes apenas difieren de las metodologías convencionales de minería de opinión y no han sido diseñados para el análisis de interacciones socio-afectivas (Yadollahi et al., 2017). El centro de atención de la comunidad científica está puesto en el análisis de los textos, fundamentalmente en las redes sociales, pero desde la perspectiva de disciplinas que requieren un enfoque muy orientado a la interacción hombre-máquina, como el diseño de la Experiencia de Usuario (UX), se demandan soluciones metodológicas que aún están por llegar (Clavel & Callejas, 2016).

En cuanto la diversidad de aproximaciones a la solución de los problemas, se han documentado a lo largo de los últimos 10 años diversas tecnologías y estrategias que permiten conocer con precisión los intereses de los usuarios mediante el estudio de su actitud digital, incluyendo gustos, hábitos de consumo, aficiones, estilo de vida, ideología política, o incluso orientación religiosa y sexual. Por ejemplo, en un estudio realizado por Ribeiro et al. (2016) se referencian hasta 24 *sentence-level methods*³⁶ disponibles en la literatura científica, siete de ellos de pago y el resto *open source*. En estudios más recientes, se identifican hasta ocho técnicas que incorporan *machine learning* (Qazi et al., 2017). Por otro lado, proliferan las herramientas, no

³⁶ Métodos de análisis de textos con un nivel de granularidad a escala de “frase”.

necesariamente científicas, centradas en analítica de las redes sociales (Batrinca & Treleaven, 2014), dado el especial interés que este ámbito tiene para el marketing y el mundo de la política.

Todo apunta, por tanto, a una falta de estandarización y de consenso, con un aluvión bibliográfico de soluciones a los mismos problemas, síntoma de lo que podríamos interpretar como una saludable insatisfacción de la comunidad investigadora por estado actual del arte.

2.2.2 Desafío idiomático

Otro importante frente para los investigadores (especialmente para los de habla no inglesa), es la adecuación de los métodos a los distintos idiomas. El estudio de la literatura académica refleja claramente que existen serias dificultades para el análisis de textos en idiomas distintos del inglés (y en menor medida, chino). Se han publicado notables trabajos recientes para la lengua española (Del et al., 2016), pero la mayoría de los recursos más desarrollados han sido creados originalmente para el ámbito anglosajón, lo cual es especialmente problemático en el caso de los *corpora anotados*, cruciales para mejorar el rendimiento de los sistemas de minería de opinión (Sun et al., 2017). En el caso concreto de la minería de emociones en Twitter, uno de los espacios de investigación más recurrentes entre la comunidad científica, la mayor parte de los trabajos disponibles se refieren al idioma inglés (Yadollahi et al., 2017).

Una investigación bibliométrica³⁷ detallada sobre el número de trabajos publicados en función del idioma objeto de estudio (distinto del inglés), arroja los siguientes resultados:

- Arabic: 290
- Russian: 54
- Indonesian: 52
- Spanish: 49
- Chinese: 42
- Hindi: 38
- Portuguese: 36
- Bengali: 33
- German: 16
- French: 13

La extraordinaria pujanza de los trabajos orientados al árabe se explica especialmente por el especial interés de la comunidad científica de Arabia Saudí, origen de casi la cuarta parte de los trabajos publicados, sobre un total de 10 países involucrados.

³⁷ Prospección realizada en la base de datos de Scopus (julio 2022), sobre los 10 idiomas con mayor número de hablantes.

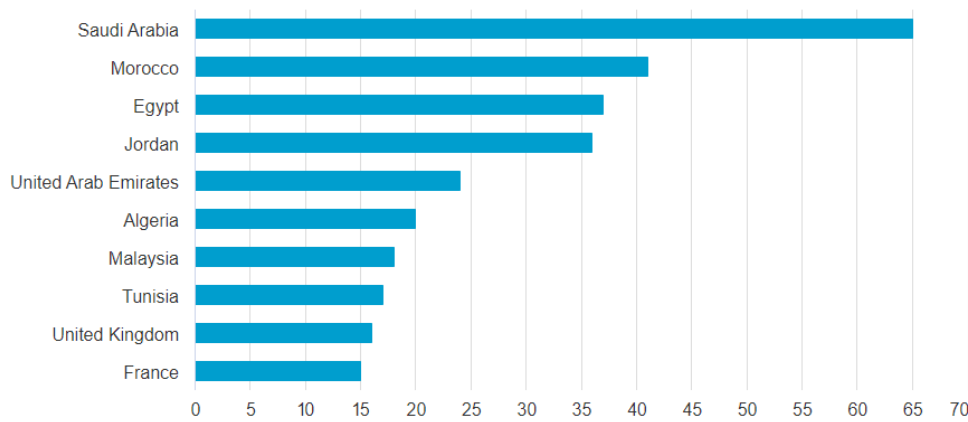


Figura 7 – Países de origen de los trabajos publicados con foco en la lengua árabe

En el caso de los trabajos orientados al idioma español, obtenemos la distribución por territorios que refleja la Figura 8. El 57% de los trabajos tienen su origen en España. Esto puede implicar un sesgo cultural considerable, ya que, si bien el idioma es el mismo, las expresiones coloquiales, referencias contextuales, y determinados aspectos culturales que tienen un gran impacto en el análisis de sentimiento son potencialmente distintos. Algunos autores en Latinoamérica han señalado este *Spain-focused bias* como un hándicap importante a la hora de utilizar recursos previamente desarrollados por otros investigadores, y han desarrollado estrategias para compensarlo (Cerón-Guzmán & León-Guzmán, 2016).

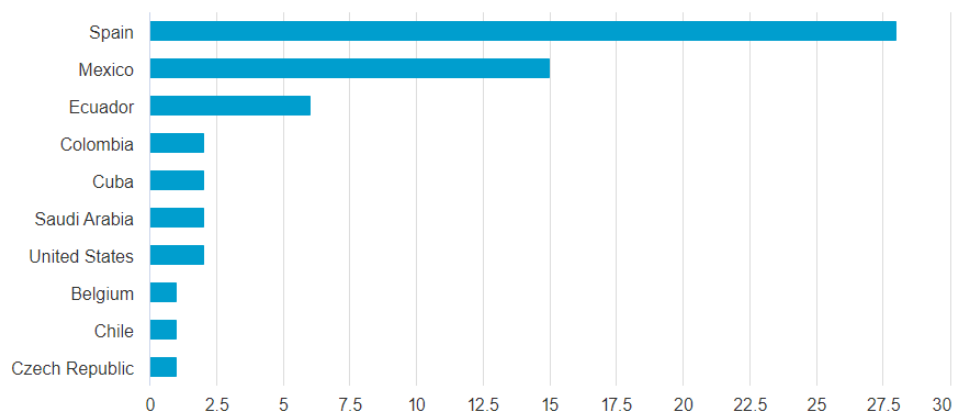


Figura 8 - Países de origen de los trabajos publicados con foco en la lengua española

En 2021, solo el 1,7% de los trabajos publicados sobre *sentiment analysis* y *opinion mining* estaban referidos a la lengua española de forma específica.

2.2.3 Obsolescencia de los trabajos publicados

Pese a la juventud de la materia y lo reciente de su bibliografía, el riesgo de obsolescencia a corto plazo de los trabajos es inevitablemente muy alto, ya que los espacios donde se produce la interacción digital de los usuarios de Internet, así como los códigos y pautas culturales de comunicación, son muy cambiantes. Por poner un ejemplo, la mayoría de los trabajos publicados son anteriores a la existencia de TikTok, una red social con cerca de 800 millones de usuarios activos en la actualidad³⁸. Otro ejemplo lo tenemos en los cambios en la API de Twitter, una de las herramientas más utilizadas en minería de opinión, que ha pasado por 6 versiones distintas desde 2016, incluyendo no solo cambios en las funcionalidades y protocolos sino incluso en las políticas³⁹.

Para añadir aún más riesgo de obsolescencia a los trabajos publicados, no debemos olvidar que las técnicas empleadas, basadas en la mayoría de los casos en Inteligencia Artificial (IA), están sometidas a una rápida y constante evolución. Para ilustrar este importante aspecto del estado del arte, basta recordar el caso anteriormente citado del *framework* BERT, una tecnología que no existía antes de 2019 pero que ha mejorado en un 7,7% los valores de referencia del estado del arte en precisión (Devlin et al., 2019).

³⁸ Cifras de Datareportal - <https://datareportal.com/> (2021)

³⁹ <https://developer.twitter.com/en/docs/ads/general/overview/versions>

2.3 Espacios de interés académico y social

Desde el punto de vista del estado del arte, podemos contemplar los focos de interés de los investigadores desde diversas perspectivas:

- Sectores de aplicación.
- Disciplinas académicas en las que se enmarcan las investigaciones.
- Áreas temáticas predominantes.
- Puntos de inflexión.

2.3.1 Sectores de aplicación

La política ha sido, desde el nacimiento de las redes sociales, uno de los focos de atención prioritaria de la comunidad científica en el estudio de la actitud de los ciudadanos en Internet. Ya en 2013, se observa un creciente interés de los investigadores en el estudio del sentimiento político para predecir resultados de elecciones (Bakliwal et al., 2013). Tras varios procesos electorales especialmente controvertidos desde 2016, y como resultado de la creciente preocupación social por temas como el discurso de odio o el riesgo de injerencia de terceros países en campañas políticas⁴⁰, se ha producido un notable incremento en el número de trabajos publicados, como refleja la figura siguiente.

⁴⁰ Factbox: Key findings from Senate inquiry into Russian interference in 2016 U.S. election (Agencia REUTERS): <https://www.reuters.com/article/us-usa-trump-russia-senate-findings-fact-idUSKCN25E20Y>

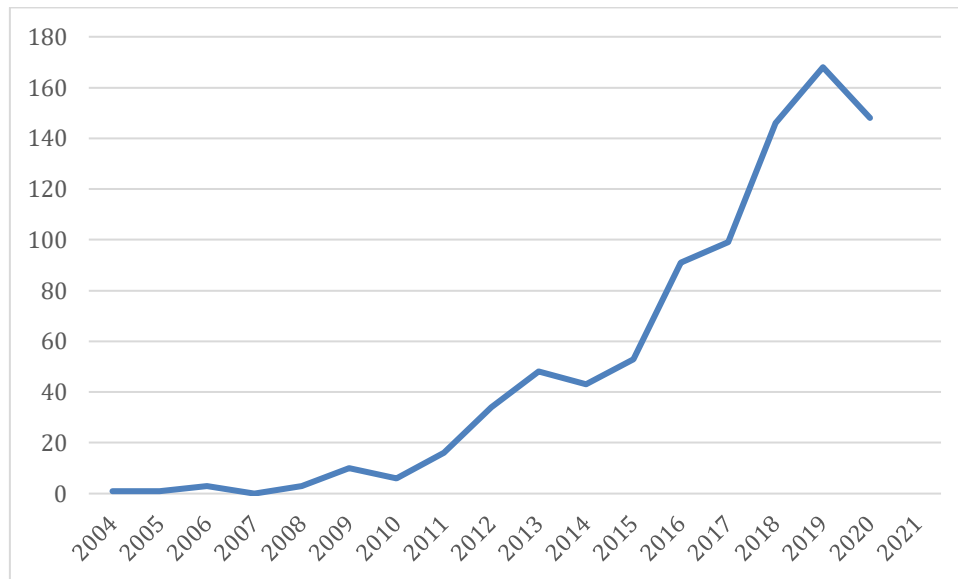


Figura 9 – Volumen de trabajos publicados sobre análisis de sentimiento referidos a política o a discurso de odio, total mundial.

La comunidad científica también ha centrado su atención en otros campos de estudio desde el punto de vista de la minería de opinión y el análisis de sentimiento. En particular, destacan los trabajos relativos al área de la Educación, donde se observa un creciente interés por el impacto de las redes sociales en la comunidad docente, la educación inclusiva (Schuster et al., 2021), y las políticas y leyes de educación (Supovitz, 2017)(Pérez Díaz et al., 2016). También se ha estudiado desde el punto de vista de la minería de opinión el uso de las redes como herramientas al servicio de los alumnos (Petrilli, 2015).

Otro de los campos que han concentrado de forma decisiva el interés de los investigadores, definiendo el marco teórico del arte, es el marketing y la investigación de mercados, con una orientación especialmente focalizada en la opinión de los consumidores sobre los productos y las marcas (Fang & Zhan, 2015)(Mirzaalian & Halpenny, 2019)(Reyes-Menendez et al., 2019)(Valdivia et al., 2019).

El área de la Salud (*Healthcare*) es otro de los dominios protagonistas en los estudios sobre *sentiment analysis* (Magami & Digiampietri, 2020), ya que se trata de un campo que involucra un gran espacio de oportunidades: monitorización del estado de ánimo de los pacientes, de las reacciones adversas a los tratamientos, y de la expansión de las epidemias, entre otros aspectos de gran relevancia en el campo de la salud pública (Ramírez-Tinoco et al., 2019).

Si ponemos el foco en las investigaciones específicamente orientadas a las opiniones en lengua española, observamos también una significativa acumulación de estudios, focalizados en diversas áreas, como la educación (Esparza et al., 2018)(López et al., 2018)(Ortega et al., 2020), la salud (García-Díaz et al., 2020)(Jiménez-Zafra et al., 2017), la pandemia del Covid-19 (de las Heras-Pedrosa et al., 2020)(Ferrer-Serrano et al., 2020)(Cabezas et al., 2021), la política (Pla & Hurtado, 2014) (Vilares et al., 2015) (Rodríguez-Ibanez et al., 2021) , y el discurso de odio (Plaza-Del-Arco et al., 2021) (Pereira-Kohatsu et al., 2019).

2.3.2 Transversalidad académica

Un aspecto muy evidente en el análisis bibliográfico del dominio del análisis de sentimiento y la minería de opinión es su transversalidad académica. Si ponemos el foco en las disciplinas académicas en las que se enmarcan los estudios publicados, obtenemos el resultado que muestra la siguiente figura, para la totalidad de los documentos:

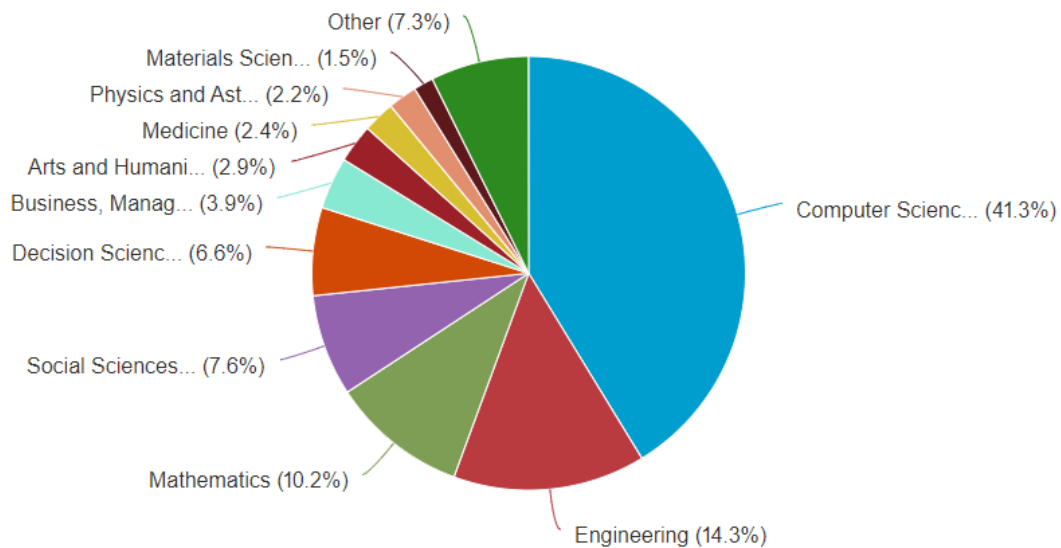


Figura 10 – Documentos publicados por especialidad académica.

Es interesante observar cómo la Ciencia de Datos no es el único espacio académico en el que se inscriben las investigaciones (aunque obviamente es el más importante, con un 41% del total).

2.3.3 Áreas temáticas predominantes

Por otro lado, si adoptamos una aproximación diferente y nos centramos en las áreas temáticas predominantes, podemos hacer la siguiente clasificación, atendiendo a los resultados arrojados por Google Scholar⁴¹:

- Politics: 17.900

⁴¹ Número de artículos relacionados a través del algoritmo de Google Scholar – consulta 21/04/2022

- News: 17.200
- Marketing: 17.400
- Health: 16.300
- Education: 16.200
- Covid-19: 15.000
- Social media: 14.800
- Hate: 12.600
- Customer satisfaction: 9.540
- Fake News: 7.330
- Bullying: 2.850

Las cifras reflejan un interés predominante por la política, las noticias el marketing, además de una gran atención de la comunidad académica por otras áreas como la salud y la educación.

Es interesante comparar los resultados anteriores con las referencias indexadas por Google Scholar para artículos académicos en español:

- Redes sociales: 1460
- Noticias: 910
- Política: 860
- Marketing: 787
- Educación: 733
- Salud: 690
- Covid-19: 250
- Odio: 198
- Satisfacción del cliente: 123
- Acoso: 97
- Noticias falsas: 87

Como vemos, los intereses son prácticamente los mismos y se priorizan de forma similar, con la notable excepción del capítulo “redes sociales”.

2.3.4 Puntos de inflexión

Más allá del tipo de producción académica y enfoque de los estudios, desde el punto de vista de la atención de la sociedad en general por el análisis de sentimiento, observamos una progresiva tendencia al alza. Es muy reseñable que, con anterioridad a la popularización de las redes sociales, el interés era casi inexistente. De hecho, tanto la literatura publicada como el interés social deducido de la cantidad de búsquedas realizadas (Figura 3), literalmente se “disparan” a raíz de la eclosión de la red Social Twitter a finales de la primera década del siglo.

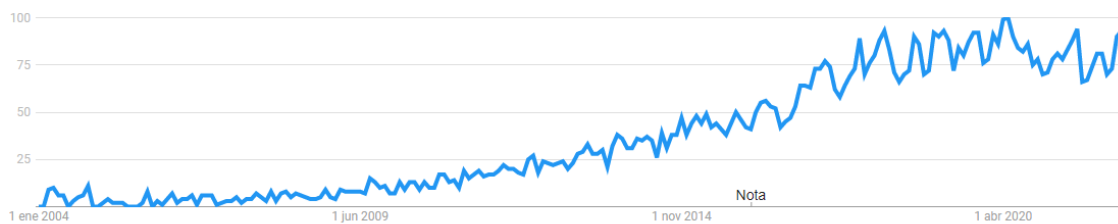


Figura 11 – Búsquedas del término “Sentiment Analysis” – Total mundial 2004-2021. Fuente: Google Trends

Podemos identificar en la literatura dos grandes puntos de inflexión relativos a la preocupación académica por temas específicos de gran trasfondo social:

- Escándalo de Cambridge Analytica
- Pandemia del COVID-19

Tras varios procesos electorales controvertidos desde 2016 a nivel mundial, y con la proliferación de contenidos engañosos sobre la reciente pandemia del Covid-19,

las *fake news* son una tendencia creciente con graves consecuencias políticas y sociales potenciales, que preocupan a la comunidad científica y a la sociedad en su conjunto. Para mitigar esta amenaza, se ha diseñado una amplia gama de enfoques (Hakak et al., 2020). En concreto, para abordar el tema de la detección de noticias falsas, se han propuesto numerosos estudios, basados en métodos de aprendizaje supervisados y no supervisados (Hakak et al., 2021).

El estudio científico del discurso de odio, desde la perspectiva de la computación, es muy reciente (Ben-David & Matamoros-Fernández, 2016), al igual que la literatura sobre la problemática de las *Fake News* (Allcott & Gentzkow, 2017).

En el caso del idioma español, es importante señalar que antes de 2019 no había un solo artículo sobre análisis de sentimiento específicamente dedicado al discurso del odio, pero sin embargo esta temática representa casi el 30% de todo lo publicado desde entonces. Estamos ante un área de interés muy reciente y en plena eclosión.

3 Desarrollo de la investigación

El análisis de los textos, en cualquier nivel, requiere el uso de una serie de herramientas, recursos y metodologías. Mediante la aplicación de estos medios a disposición del investigador, se puede llegar a conocer la actitud de los usuarios en Internet hasta un cierto punto, y con un determinado nivel de precisión. En este capítulo veremos con detalle el desarrollo de la investigación realizada para conocer estos aspectos, a fin de satisfacer todos los objetivos específicos planteados en la tesis.

En particular, la investigación realizada se centra en descubrir, sumarizar y comparar los métodos disponibles para la comunidad investigadora, así como los recursos, e identificar cuáles son los límites a los que se ha llegado hasta ahora, con foco en los principales desafíos conceptuales y técnicos.

3.1 Los métodos

Como vimos con anterioridad, el objetivo principal del análisis de sentimientos es, en primer lugar, determinar si un texto (o una parte) es subjetivo y, en segundo lugar, determinar si expresa una opinión positiva o negativa (Taboada, 2016). A esta “orientación semántica”, la denominamos “polaridad”, un dato que, dependiendo del método y herramienta empleados, puede expresarse conceptualmente o mediante un rango numérico en función de su intensidad⁴².

⁴² La polaridad puede expresarse mediante un valor numérico con coma flotante en el rango [-1,1]. Sentimiento negativo: polaridad < 0. Sentimiento neutro: polaridad =0. Sentimiento positivo: polaridad >0

Los métodos aplicados por la comunidad investigadora deben ser capaces de resolver este desafío, así como los retos específicos de cada fase del proceso (Figura 12):

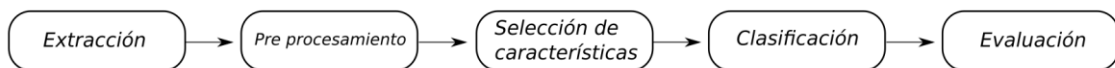


Figura 12 - Visión simplificada de las etapas del proceso de sentiment analysis

Cuando utilizamos técnicas de *machine learning*, la clasificación requiere el entrenamiento de un modelo. Por otro lado, la evaluación necesita siempre de un *dataset*⁴³ anotado con el que comparar los resultados, es decir, una colección de datos con la polaridad determinada por medios humanos a fin de comprobar la precisión y exactitud de las predicciones realizadas por el modelo, unidad a unidad. El esquema propuesto por Mironczuk and Protasiewicz (2018) (Figura 13), representa de una manera más específica las tareas más representativas de los trabajos publicados:

⁴³ *Dataset* es algo más que un mero “conjunto de datos”. En Ciencia de Datos, *dataset* es una colección organizada de elementos relacionados y discretos a los que se puede acceder de forma individual o combinada, y que pueden ser gestionados como una entidad completa.

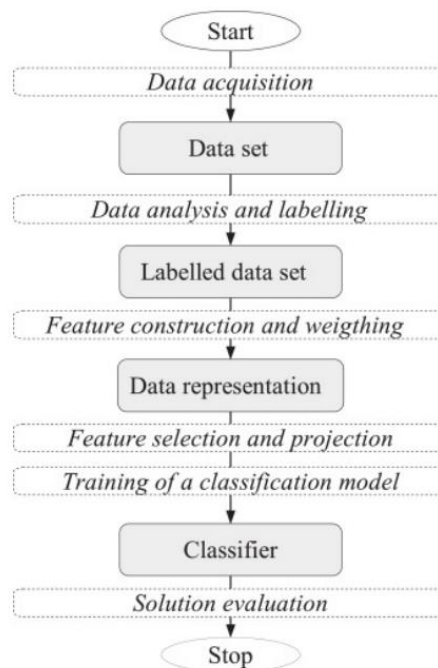


Figura 13 – Proceso estándar de clasificación de textos - Mironczuk and Protasiewicz (2018)

A continuación, veremos con detalle los desafíos y métodos a disposición de la comunidad científica para cada una de las etapas.

3.1.1 Extracción y *scope* del análisis

La extracción y definición del alcance del análisis se debe determinar a varios niveles:

3.1.1.1 Nivel de granularidad

Para trabajar en el análisis de los textos, la primera decisión que debe tomar el investigador es el nivel de granularidad (o escala) al que necesita trabajar para responder a sus preguntas de investigación. En general, se distinguen varios niveles: documento, frase o entidad (Liu, 2012).

La figura 12 muestra un esquema visual de estos niveles.

- **Document level classification:** se trata de establecer la polaridad (positiva o negativa) de un documento completo. Generalmente, esto solo tiene sentido cuando la totalidad de la pieza se refiere a una única entidad (por ejemplo, un producto, una empresa, un destino turístico, una propuesta política, una obra, etc.).
- **Sentence level classification:** cuando se trabaja con un nivel de granularidad de escala de frase, podemos determinar si ésta es positiva, negativa o neutra. Una frase neutra es aquella que no expresa sentimiento alguno. En este sentido podemos hacer una primera discriminación de frases separando las frases subjetivas (tienen sentimiento) de las objetivas (no lo tienen).
- **Entity and Aspect level classification:** por último, podemos llevar el análisis al máximo nivel de granularidad posible, analizando entidades y aspectos dentro de la frase. Este enfoque nos permite trabajar poniendo en relación directa un sentimiento con un objeto, más allá de las estructuras del lenguaje. Veámoslo con un ejemplo:

- “X [un político] siempre ha sido un impresentable, pero hay que reconocer que su nueva política de empleo está dando buenos resultados”. Aquí tenemos dos entidades, X y su política de empleo. Para una, el sentimiento es negativo, para la otra, positivo.

En ocasiones, lo que tenemos son aspectos de una misma entidad, con sentimientos diferentes. Ejemplo:

- “El X [teléfono móvil] hace muy buenas fotos pero no tiene suficiente memoria”. En este caso, hay una entidad sobre la que se están juzgando dos aspectos.

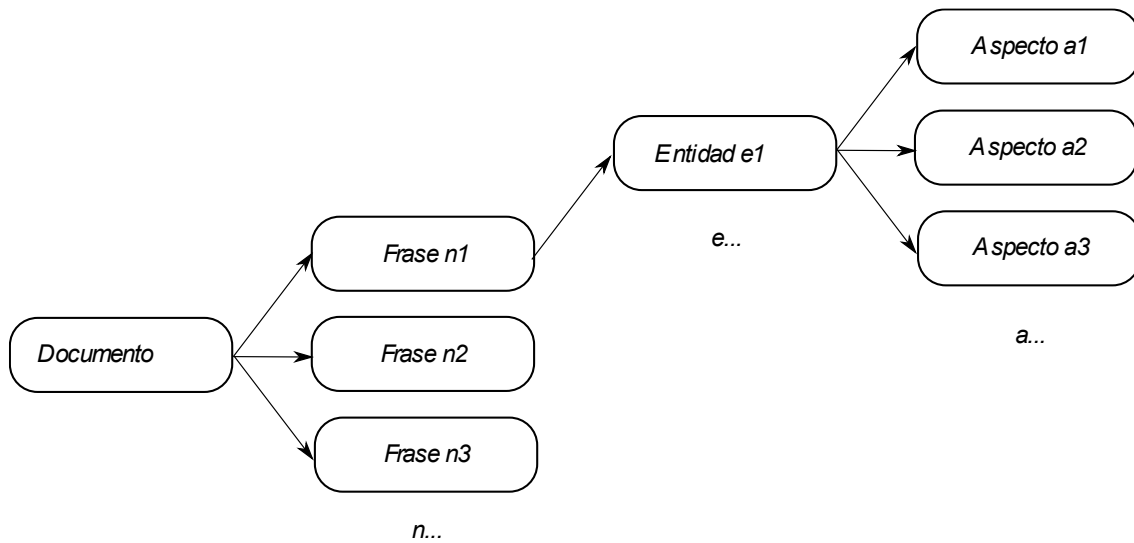


Figura 14 – Niveles de granularidad: documento, frase, entidad y aspecto

3.1.1.2 Identificación/definición de opiniones

El primer desafío para el investigador que accede a los textos para su posterior procesamiento, es doble, ya que necesita:

- Una definición accionable de “opinión”: ¿qué es una opinión y que elementos la definen?
- Un marco teórico de trabajo que permita transformar en datos estructurados textos no estructurados.

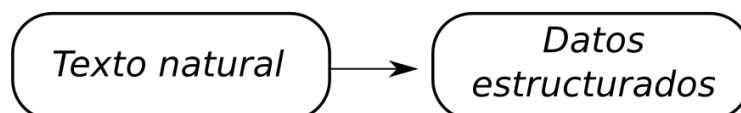


Figura 15 – El marco básico teórico de trabajo para el análisis de sentimiento, convertir texto natural en datos estructurados.

Según Liu (2012), una opinión se puede representar como el conjunto $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$, donde:

- e_i es el nombre de una entidad.
- a_{ij} es un aspecto de e_i .
- s_{ijkl} es el sentimiento sobre el aspecto a_{ij} de la entidad e_i .
- h_k es el titular de la opinión (*opinion holder*).
- t_l es el momento en que la opinión es expresada por h_k .

El sentimiento s_{ijkl} puede ser positivo, negativo o neutro, pero también se puede expresar por niveles intensidad.

3.1.1.3 Tareas

Siguiendo esta definición de Liu, muy citada en la literatura, dada una opinión en el documento D , el objetivo del investigador sería descubrir $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ en D . Para ello, por tanto, tendremos que realizar las siguientes tareas:

- **Extracción y categorización de entidades:** extraer todas las expresiones de entidad en D , y agrupar las expresiones de entidad sinónimas en *clusters* de entidades. Cada *cluster* de expresiones de entidad indica una única entidad e_i .
- **Extracción y categorización de aspectos:** extraer todas las expresiones de aspecto de las entidades, y categorizarlas en *clusters*. Cada *cluster* de expresiones de aspecto de la entidad e_i representa un único aspecto a_{ij} .
- **Extracción y categorización de titulares de opinión:** extraer y categorizar los titulares de opinión (titulares en el sentido de *opinion holders*, personas a las que se atribuye la opinión).
- **Extracción de tiempos:** extraer las horas de las opiniones y estandarizar los diferentes formatos de tiempo.

- **Clasificación del sentimiento del aspecto:** determinar si una opinión sobre un aspecto a_{ij} es positiva, negativa o neutra, o asignar una calificación de intensidad (numérica) al aspecto.
- **Generación de quintuplas de opinión:** generar todas las quintuplas de opinión $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ del documento **D** a partir de los resultados de las tareas anteriores.

3.1.1.4 Spam detection

Una vez que se ha decidido el nivel de granularidad necesario para el análisis, a menudo el investigador se encuentra con un problema grave en el proceso de extracción: la existencia de gran cantidad de opiniones falsas. Tanto en el ámbito de la política como el de la crítica de productos y servicios, existen redes organizadas y *bots* que inundan determinados espacios de Internet con opiniones falsas, a fin de influir en las opiniones legítimas de los usuarios bien intencionados. Este fenómeno fue bautizado en los primeros trabajos sobre *sentiment analysis* como *opinion spamming* (Jindal & Liu, 2006). Detectar y descartar este tipo de textos es un desafío de primer orden para la comunidad investigadora.

Se han desarrollado diversas aproximaciones para la solución de este problema. Entre las más recientes y de mejor resultado destacamos la de Arif et al. (2018), que parte de una técnica existente de LCS para aplicar un nuevo esquema de codificación. Básicamente, se trata de representar las reglas del clasificador con el fin de manejar la escasez de vectores de características, que se generan utilizando la frecuencia de documentos inversa de los n-gramas de palabras y los léxicos de sentimiento. Este nuevo esquema de codificación mejora el proceso de aprendizaje y genera resultados consistentemente buenos.

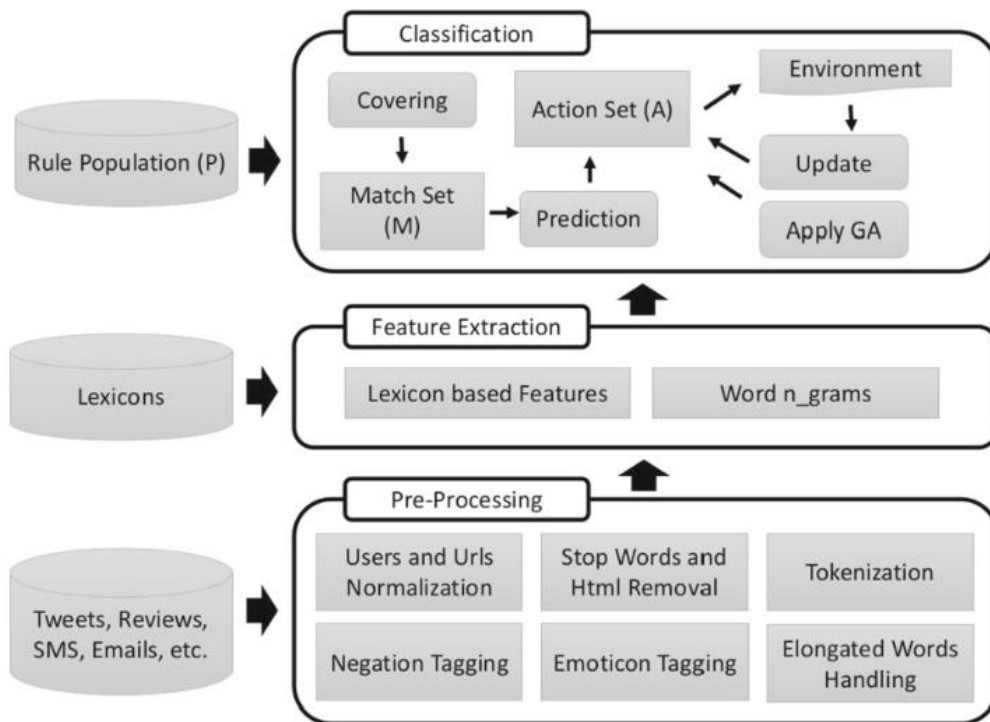


Figura 16 – Arquitectura detallada del método de Arif et al. (2018 p. 4)

3.1.2 Preprocesamiento de textos

Como muestra la figura siguiente, los textos necesitan ser preprocesados para poder ser sometidos a un análisis formal como vectores. Los tres procesos básicos, tal y como son descritos en la literatura, constan de *Tokenization*, *Stop Word Removal* y *Stemming*.

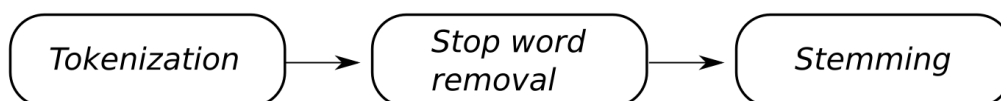


Figura 17 – Etapas de preprocesamiento de textos convencionales para su análisis

Algunos autores incluyen en el preprocesamiento otras fases posteriores que, por involucrar un componente de procesado analítico, podrían considerarse como parte del proceso y no como una “preparación” del material, como por ejemplo la identificación de entidades (*entities*) (Feldman, 2013).

Tokenization (tokenización) es el proceso de convertir un texto en sus componentes básicos (*tokens*) antes de transformarlo en vectores. Por ejemplo, descomponer una frase en palabras.

Stemming es un método de normalización esencial que consiste en eliminar los sufijos de las palabras para quedarnos con una versión básica (*root form*) del concepto. Esta raíz no tiene por qué ser idéntica a la raíz morfológica de la palabra, de hecho, suele bastar con que las palabras relacionadas se correspondan con la misma raíz, aunque ésta no sea una raíz válida. Los algoritmos para extracción de *root forms* se remontan a los inicios de la Ciencia de Datos, y muchos motores de búsqueda modernos tratan aquellas palabras que tienen una misma raíz como sinónimos en una consulta expandida⁴⁴.

En cuanto al proceso de *stop-words removal*, se trata de aplicar un método que permita mejorar la eficiencia de los procesamientos posteriores, eliminando palabras que no aportan nada al análisis, como por ejemplo los artículos y conjunciones⁴⁵. Con este depurado, suprimimos información de bajo nivel, lo que no solo reduce el *dataset* sino que además sirve para eliminar “ruido” en los textos.

La mayoría de las librerías disponibles para la investigación académica en procesamiento de lenguaje natural, como NLTK, cuentan con recursos para la

⁴⁴ Este proceso suele ser denominado *conflation* (conflación).

⁴⁵ En inglés, algunas de las stop-word más comunes son "the", "a", "an", "so" y "what".

realización de todos estos procesos que son parte de la rutina metodológica del investigador.

La comunidad investigadora coincide en resaltar las diferencias entre el preprocesado de los textos extraídos dependiendo de si se trata de literales convencionales o tuits. Según Giachanou & Crestani “detectar el sentimiento en Twitter es una tarea no trivial que difiere considerablemente de la detección del sentimiento en textos convencionales como blogs y foros. Los investigadores que intentan desarrollar métodos eficaces de TSA tienen que enfrentarse a una serie de retos que surgen de las características especiales de Twitter” (Giachanou & Crestani, 2016 p. 3) .

Estos retos “extra” en el análisis de textos extraídos de Twitter implican una serie de fases de preprocesado adicionales, como refleja el siguiente esquema:

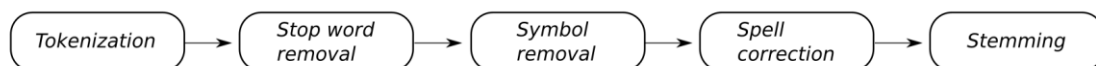


Figura 18 - Etapas de preprocesamiento con fases adicionales para análisis de textos de micro blogging (por ejemplo, Twitter)

Según la revisión del estado del arte de Giachanou & Crestani (2016), por tanto, se observa cómo en el caso de Twitter se tiende a añadir fases adicionales de preprocesado de los textos, teniendo en cuenta los símbolos y las licencias lingüísticas habituales entre la comunidad de usuarios de la plataforma.

3.1.3 Selección de características (*feature selection*)

La inmensa mayoría de los métodos de *sentiment analysis* establecen el sentimiento de los textos analizados basándose en un conjunto de características. Por tanto, el

criterio de selección de estas características juega un papel crítico en todo el proceso. Los métodos a disposición de la comunidad investigadora, en la actualidad, se centran en tres tipos de características:

- **Características semánticas:** se focalizan en términos que revelan la valencia del sentimiento o el concepto semántico de los términos. La valencia (o polaridad) del sentimiento puede ser positiva o negativa.
- **Sintácticas:** habitualmente son los n-gramas y las etiquetas de partes del discurso.
- **Estilísticas:** se refieren a características no estandarizadas, como el uso de emoticonos gráficos, abreviaturas o combinaciones de signos de puntuación para expresar emociones (por ejemplo, la representación de un ojo guiñado con el símbolo del punto y como y el paréntesis).

Algunos autores señalan un cuarto tipo, relacionado con la idiosincrasia de las comunidades que participan en determinados espacios, como foros y redes sociales. Ejemplos de estas características especiales que son propias de determinadas redes podrían ser términos como “fav”, “RT”, o los *hashtags*. En particular, Giachanou & Crestani (2016) han prestado especial atención en sus investigaciones a estas características específicas de Twitter.

Veamos con detalle los tres tipos de características que son objeto de extracción:

- **Características semánticas:** generalmente, el investigador distingue cuatro subtipos de características semánticas para analizar los textos:
 - Palabras y frases de sentimiento (*sentiment words*): sirven para determinar la polaridad de un sentimiento.
 - Palabras y frases de opinión (*opinion words*): son indicativas de una opinión específica.
 - Conceptos semánticos: algunos investigadores, como por ejemplo Saif et al. (2012), utilizan estas características como información
-

complementaria para optimizar el rendimiento del proceso de clasificación.

- Negaciones: las negaciones pueden cambiar la polaridad del mensaje, por lo que siempre son objeto de especial atención. Por razones obvias, un análisis que no tuviera en cuenta la inversión de polaridad producida por las negaciones no podría arrojar los resultados esperados. Las características asociadas a la negación han sido examinadas por numerosos autores desde los orígenes de esta disciplina, pero al respecto cabe destacar especialmente los trabajos de Kiritchenko et al. (2014).
- **Características sintácticas:** incluyen elementos esenciales para el análisis, como unigramas, n-gramas, árboles de dependencia y POS (*Parts Of Speech*), entre otros⁴⁶. En particular, las etiquetas POS identifican el papel sintáctico de los términos en la frase, como verbo, nombres, adjetivo, etc. (Figura 19) y los árboles de dependencia determinan las relaciones entre los distintos términos (por ejemplo, cuál es el sujeto en relación con un verbo).

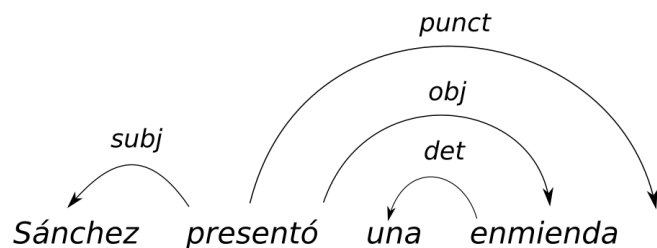


Figura 19 - Ejemplo de árbol de dependencia.

⁴⁶ Una herramienta muy utilizada para descomponer las frases en etiquetas *part-of-speech*, que ha sido incluso incorporada a otras metodologías (como por ejemplo OpinionFinder) es Stanford Part Of Speech Tagger (<https://nlp.stanford.edu/software/tagger.shtml>). Esta aplicación realiza el también el proceso de *stemming*.

Tag	Description	Tag	Description
CC	Coordinating conjunction	PRP\$	Possessive pronoun
CD	Cardinal number	RB	Adverb
DT	Determiner	RBR	Adverb, comparative
EX	Existential <i>there</i>	RBS	Adverb, superlative
FW	Foreign word	RP	Particle
IN	Preposition or subordinating conjunction	SYM	Symbol
JJ	Adjective	TO	<i>to</i>
JJR	Adjective, comparative	UH	Interjection
JJS	Adjective, superlative	VB	Verb, base form
LS	List item marker	VBD	Verb, past tense
MD	Modal	VBG	Verb, gerund or present participle
NN	Noun, singular or mass	VBN	Verb, past participle
NNS	Noun, plural	VBP	Verb, non-3rd person singular present
NNP	Proper noun, singular	VBZ	Verb, 3rd person singular present
NNPS	Proper noun, plural	WDT	Wh-determiner
PDT	Predeterminer	WP	Wh-pronoun
POS	Possessive ending	WP\$	Possessive wh-pronoun
PRP	Personal pronoun	WRB	Wh-adverb

Tabla 1 - Penn Treebank Part-Of-Speech (POS) tags - extraído de (Liu, 2012)

- **Características de estilo:** muy a menudo, y en especial en las redes sociales, los literales incluyen elementos estilísticos ajenos al lenguaje que influyen en la polaridad de los mensajes y que aportan, potencialmente, información adicional relativa a las emociones. En ocasiones, estos elementos son fundamentales para la plena comprensión del texto. Podemos distinguir dos tipos, generalmente incluido en la literatura científica:
 - Los emoticonos, ya sea en formato gráfico o mediante combinaciones de signos.
 - El uso enfático de signos y de mayúsculas, como por ejemplo la repetición de signos de admiración o la escritura en versales de ciertas palabras para destacar su importancia.

La tarea de extracción de características es laboriosa, y la comunidad científica ha intentado dar respuesta a este desafío desde los primeros años de la disciplina. Las palabras y frases que implican opinión, al igual que las que implican sentimiento, se

extraen en la práctica habitual de los investigadores mediante dos tipos de procedimiento, manual y semiautomático.

- Manual: habitual en idiomas para los que hay pocos o ningún recurso para automatizar el proceso.
- Semiautomático: con léxicos de opinión y de sentimiento cuando están disponibles (más adelante volveremos sobre este aspecto).

3.1.4 Metodologías de clasificación

La investigación realizada revela cinco grandes tipos de aproximación que están a disposición de la comunidad científica a la hora realizar la clasificación:

- *Machine Learning* (y su modalidad *Deep Learning*)
- *Lexicon-Based*
- *Graph-Based*
- *Métodos híbridos* (*Machine Learning* + *Lexicon-Based*)
- *Otros métodos*

Veamos a continuación en qué consiste cada una de ellas y cómo aportan diferentes aproximaciones al problema.

3.1.4.1 *Machine learning*

Hoy en día, la gran mayoría de los métodos emplean un clasificador basado en aprendizaje automático, que se somete a entrenamiento usando varias características elegidas de acuerdo con el proceso anterior de *feature selection*. Los más habituales a disposición de la comunidad de investigadores son:

- Naive Bayes (NB)
- Logistic Regression (LR)
- Maximum Entropy (MaxEnt)

- Support Vector Machines (SVM)
- Multinomial Naïve Bayes (MNB)
- Random Forest (RF)
- Conditional Random Field (CRF)

Los métodos basados en *machine learning* se fundamentan generalmente en un proceso de aprendizaje supervisado, que podríamos definir como el uso de conjuntos de datos etiquetados para entrenar algoritmos de clasificación. A medida que los datos de entrada se introducen en el modelo, éste va afinando sus ponderaciones, como parte del proceso de validación cruzada. No obstante, en la actualidad es muy frecuente que los investigadores apliquen una modalidad conocida como *ensemble*, que consiste en generar varios clasificadores a partir de los cuales se obtiene un nuevo clasificador de mayor rendimiento. Los clasificadores base que se combinan en el clasificador *ensemble* pueden diferir en distintos aspectos, como el algoritmo utilizado o el conjunto de entrenamiento.

El concepto de combinar clasificadores se ha propuesto como una nueva dirección para mejorar el rendimiento de los clasificadores individuales (Giachanou & Crestani, 2016).

Veamos a continuación con detalle los algoritmos de clasificación más utilizados por la comunidad científica:

- **Naive Bayes** (NB). Se trata de una técnica básica para construir clasificadores. Los clasificadores Naive Bayes constituyen una familia de clasificadores probabilísticos sencillos que se basan en la aplicación del teorema de Bayes con hipótesis de independencia fuerte entre las características. A pesar de su diseño “ingenuo” (*Naive*) y de la aparente simplificación excesiva de sus supuestos, lo cierto es que estos clasificadores han funcionado bastante bien en muchas situaciones y son constantemente referenciados por la comunidad investigadora. Por ejemplo, según la

investigación realizada para esta tesis sobre monitorización de la opinión política y el discurso de odio en textos de lengua española, el 18% de los trabajos analizados en una revisión sistemática utilizaron este tipo de algoritmo para alguna de las tareas.

- **Support Vector Classifier (SVM)**. Es un algoritmo de *machine learning* derivado de la teoría del aprendizaje estadístico basado en el principio de minimización del riesgo estructural. Se trata de una veterana aproximación introducida por Vapnik y Cortés a finales del pasado siglo (Vapnik & Cortes, 1995). Su objetivo, básicamente, es reducir el error de las pruebas y la complejidad computacional. Funciona identificando un hiperplano de separación óptimo entre las dos clases de un conjunto de datos de entrenamiento. Los SVM son utilizados frecuentemente como métodos de aprendizaje supervisado para la clasificación binaria, la regresión y la detección de valores atípicos.
- **Decision Tree Classifier (DTC)**. Es una herramienta que utiliza un modelo en forma de árbol de decisiones. Realiza una partición binaria recursiva para dividir un conjunto de datos en subconjuntos cada vez más pequeños, desarrollando simultáneamente un árbol de decisión incremental. El resultado final del proceso es un árbol con nodos de decisión y “nodos hoja” (leaf). Un nodo de decisión tiene dos o más ramas; un nodo hoja representa decisiones o clasificaciones. De este modo se puede hallar el árbol más corto que se ajusta a los datos, árbol que, habitualmente, será el que produce el menor error de validación cruzada. Este tipo de algoritmo de clasificación suele elegirse por su capacidad de perfeccionar modelos predictivos con mayor facilidad de interpretación y precisión (Ansari et al., 2020).
- **Logistic Regression (LR)**. La regresión logística es un método de clasificación muy común en metodologías de aprendizaje automático. Calcula la probabilidad de que una muestra pertenezca a una clase determinada

mediante el aprendizaje del conjunto de parámetros que minimizan la log-verosimilitud inducida sobre los ejemplos de entrenamiento.

- **Random Forest Classifier** (RFC). Consiste en una colección de clasificadores estructurados en forma de árbol $\{h(x, \theta_k), k = 1, \dots\}$ donde los $\{\theta_k\}$ son vectores aleatorios independientes idénticamente distribuidos y cada árbol arroja un voto unitario para la clase más popular en la entrada x (Breiman, 2001). El algoritmo Random Forest crea muchos árboles aleatorios en el subconjunto de datos y combina el resultado de todos ellos. La ventaja de este tipo de clasificador es que proporciona predicciones más precisas en comparación con que los modelos de árbol de decisión simples, ya que reduce el exceso de ajuste y la varianza (Ansari et al., 2020).
- **Long Short Term Memory** (LSTM). Las unidades de memoria LSTM son unidades de una RNN (red neuronal recurrente) que son capaces de aprender dependencias a largo plazo. Una unidad LSTM común se compone de una célula, una puerta de entrada, una puerta de salida y una puerta de "olvido". La célula recuerda los valores a lo largo de intervalos de tiempo arbitrarios y las tres puertas regulan el flujo de información que entra y sale de la célula. Este tipo de algoritmo es un modelo de aprendizaje profundo (*deep learning*) que se utiliza principalmente para el análisis de datos secuenciales, en tareas como el reconocimiento del lenguaje.

3.1.4.2 *Deep Learning*

El *Deep Learning* (aprendizaje profundo) emplea redes neuronales para aprender en varios niveles de abstracción. En este tipo de enfoque, la palabras y frases son representadas como vectores de números reales para generar representaciones de los documentos. Tang et al. (2014) han desarrollado en sus investigaciones varias redes neuronales para *sentiment specific word embeddings* (SSWE) que más tarde fueron utilizadas como características en el análisis de sentimiento de textos de

Twitter, consiguiendo una marca F1 del 87.61% cuando se combinó con características como léxicos emoticonos y alargamientos enfáticos.

Deep learning es la tendencia que está cobrando más fuerza en las investigaciones recientes en el campo de la inteligencia artificial aplicada al procesamiento del lenguaje natural, y también en otras áreas como el reconocimiento de imágenes.

3.1.4.3 *Lexicon based*

Este tipo de métodos se basan en listas de palabras previamente anotadas, según su polaridad, a fin de poder establecer la puntuación de la opinión de un texto analizado. Estos métodos no requieren entrenamiento (de hecho, esa es su diferencia principal con respecto al enfoque de aprendizaje automático). Cada token se empareja con las palabras disponibles en el modelo para determinar si hay sentimiento y, en su caso, puntuarlo. A partir de ahí, se toma una función de combinación, como la suma o la media, para hacer una predicción final sobre el texto, con el nivel de granularidad elegido.

Los trabajos basados en *lexicon* se han aplicado ampliamente en textos convencionales como blogs, foros y reseñas de productos, pero están menos explorados en *micro blogging* (Twitter) en comparación con los métodos basados en aprendizaje automático (Giachanou & Crestani, 2016).

Cuando los investigadores centrados en el análisis de redes sociales utilizan algoritmos basados en el léxico, uno de los recursos más habituales es *SentiStrength*⁴⁷. Esta herramienta permite identificar la “fuerza” del sentimiento de un texto informal, empleando para ello un léxico previamente codificado. Además de este léxico, *SentiStrength* tiene en cuenta características propias del *microblogging*, como

⁴⁷ <http://sentistrength.wlv.ac.uk/>

emoticonos, así como negaciones y palabras de refuerzo. Otro recurso muy popular es AFINN ⁴⁸. El léxico AFINN fue desarrollado por Finn Årup Nielsen en 2009 con el objetivo principal de analizar el sentimiento de la opinión de los usuarios expresada en la red social Twitter. Consta originalmente de 2.477 palabras, 878 positivas y 1.598 negativas en una escala de -5 a +5.

Yadollahi et al. (2017) mencionan otras fuentes de *lexicons* relevantes en los trabajos de la comunidad investigadora:

- HarvardGeneral Inquirer
- Opinion Lexicon
- Multi-perspective Question Answering (MPQA)
- WPARD
- SentiWordNet 3.0.
- NRC.

Los métodos de catalogación de la polaridad no son homogéneos entre *lexicons*. Mientras que algunos aportan una perspectiva meramente binaria (1 positivo, 0 neutro, -1 negativo), otros como el WPARD o AFINN siguen un criterio de escala (intensidad) para determinar cómo de positiva o negativa es la polaridad. En el caso concreto de WPARD, la clasificación se realiza en el rango entre -6 y +6.

Existe cierto consenso acerca de las limitaciones implícitas al hecho de trabajar con *Lexicons*, y con frecuencia se señala este recurso como “necesario pero no suficiente” (Liu, 2012). He aquí algunas de las principales problemáticas a las que se enfrentan los investigadores en el uso de esta metodología:

⁴⁸ <https://github.com/fnielsen/afinn>

- Las frases que incluyen palabras que sirven para establecer el sentimiento, no necesariamente expresan un sentimiento. Esto es muy habitual en sentencias condicionales e interrogativas. Ejemplo: “¿Alguien puede confirmar que el **producto X** es **bueno**?” En este caso, la palabra bueno no está expresando la polaridad de una opinión, de hecho, ni siquiera se está expresando una opinión. El problema no está en el tono interrogativo o condicional (eso facilitaría la solución), ya que es perfectamente posible, e incluso habitual, expresar opiniones positivas o negativas en frases interrogativas o condicionales: “¿Cómo podría deshacerme de este **desastroso** producto?”. También se puede dar el caso opuesto al descrito: frases que expresan una opinión, pero carecen de palabras de sentimiento. Por ejemplo, si un usuario dice “este vehículo consume muchísimo” está expresando una opinión de polaridad negativa, pero sin emplear para ello ninguna palabra que exprese sentimiento.
- El sarcasmo y la ironía se muestran, consistentemente, como las grandes bestias negras del análisis de sentimiento, muy especialmente cuando trabajamos con metodologías basadas en *Lexicon*. El público suele ser muy explícito a la hora de expresar opiniones sobre productos y servicios, pero en el debate social y político, el recurso de la ironía es una constante.
- Otro problema recurrente es la diferente polaridad de una palabra dependiendo del dominio o contexto. Esto puede ser especialmente problemático en ciertas culturas, entre ellas la española, donde una misma expresión puede ser una grave descalificación/insulto o un elogio, dependiendo de factores contextuales a menudo extremadamente difusos.

3.1.4.4 Métodos híbridos

Poco después de publicarse los primeros trabajos destacados basados en IA, algunos investigadores comenzaron a prospectar la posibilidad de combinar estas nuevas técnicas con el enfoque tradicional basado en léxico. En particular, cabe reseñar un

estudio pionero publicado en 2011 por Zhang et al. (2011), que propuso un método híbrido basado en entidades para la red de *micro-blogging* Twitter:

El método adopta primero un enfoque basado en el léxico con el objetivo de realizar un análisis de sentimientos a nivel de entidad. En lugar de ser etiquetados manualmente, los ejemplos de entrenamiento para el clasificador vienen dados por el enfoque basado en el léxico. De acuerdo con este estudio pionero, los resultados experimentales muestran que el método propuesto mejora notablemente el *F-Score*, superando las líneas de base del estado del arte.

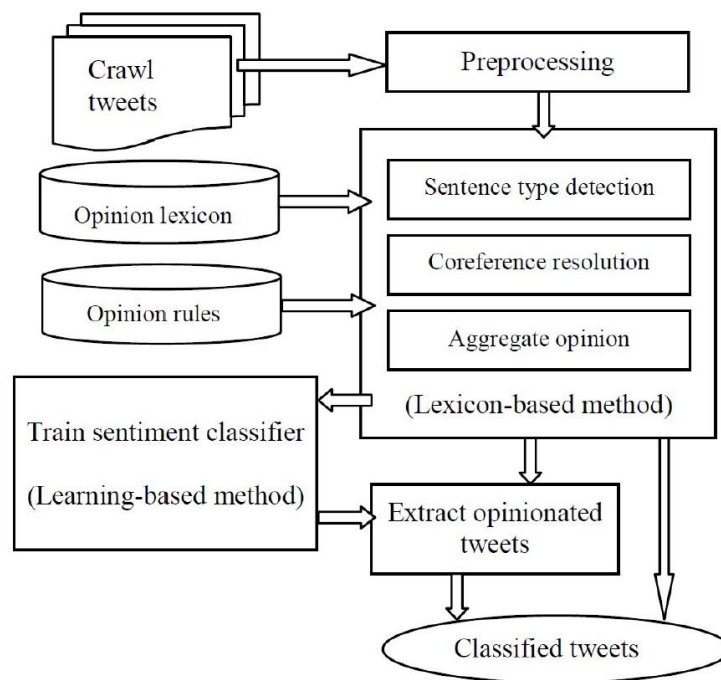


Figura 20 – Arquitectura de la aproximación “híbrida” propuesta Zhang et al. (2011 p. 4)

3.1.4.5 Métodos “Graph based”

Los métodos basados en *machine learning* requieren un gran número de datos anotados. A fin de reducir la demanda de los mismos, varios autores han

desarrollado una aproximación distinta basada en un método semi supervisado que pone el foco en la propagación de etiquetas.

En 2011, Speriosu et al. (2011) realizaron un estudio pionero que aprovechaba con éxito el gráfico de seguidores de Twitter bajo el supuesto de que las personas se influyen mutuamente.

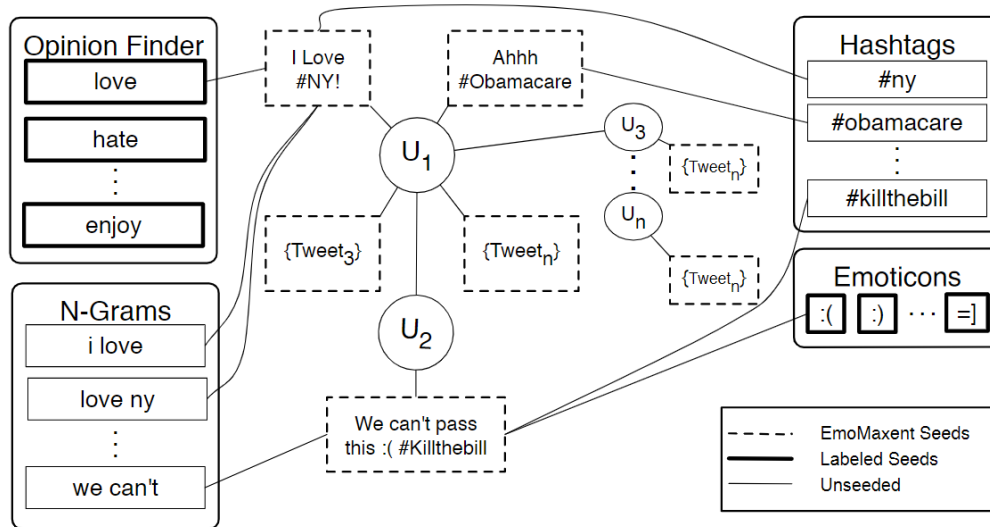


Figura 21 – Gráfico que describe la lógica aplicada por Speriosu et al. Speriosu et al. (2011 p. 57).

3.1.4.6 Otros métodos

No suelen mencionarse otros métodos en la literatura científica, ya que los anteriores son suficientemente representativos del panorama actual, pero existen algunas alternativas metodológicas que no encajan en los patrones habituales. El ejemplo más claro lo tenemos en el FCA⁴⁹, propuesto por Kontopoulos et al. (2013).

⁴⁹ Análisis conceptual formal (FCA por sus siglas en inglés).

Estos investigadores utilizaron el análisis de conceptos para construir un modelo de dominio ontológico⁵⁰ específico. Concretamente, crearon un método aplicado a Twitter donde los tuits se descomponían en un conjunto de aspectos relevantes para el dominio de los “teléfonos inteligentes” (Kontopoulos et al., 2013).

```

Input: A default concept (c)
Variables: An initially empty set of tweets (W)
               An initially empty set of objects (O)
               An initially empty set of attributes (A)
Output: A cross-table, filled with objects and attributes (T)

1. W ← retrieveTweets(c); // automatically choose the first n tweets
2. foreach w ∈ W do
3.   o ← retrieveObject(w);
4.   if o ≠ NULL then
5.     O := O ∪ {o};
6.     A' ← retrieveAttributes(w);
7.     foreach a ∈ A' such that (o, a) ≠ ∅ do
8.       A := A ∪ {a};
9. T ← populateTable(O, A);
10. return T

```

Figura 22 – Algoritmo de creación de ontologías creado por Kontopoulos (Kontopoulos et al., 2013 p. 7).

3.1.5 Métricas de evaluación

Los métodos para evaluar el análisis de sentimiento a disposición de los investigadores están muy estandarizados, y existe un amplio consenso al respecto, ya que han sido adoptados de otras disciplinas académicas donde es habitual tener que resolver problemas de clasificación estadística dentro de una ortodoxia matemática rigurosa (las áreas de conocimiento donde se aplica este enfoque con

⁵⁰ En este campo de estudio, una ontología puede definirse como una especificación de una conceptualización compartida, que es legible por una máquina. En concreto, las ontologías se utilizan para modelar los términos de un dominio de interés concreto, así como las relaciones que se establecen entre ellos.

más frecuencia son la investigación médica y la investigación de inteligencia artificial).

El punto de partida para la producción de métricas de evaluación es la matriz de confusión (*confusion matrix*) clásica, que representa el número de falsos positivos, falsos negativos, verdaderos positivos y verdaderos negativos (según figura a continuación):

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos positivos (TP)	Falsos negativos (FN)
	Negativos	Falsos positivos (FP)	Verdaderos negativos (TN)

Tabla 2 - Ejemplo de matriz de confusión (confusion matrix).

Partiendo de estos datos, los investigadores disponen de 6 métodos clásicos de cálculo: *Accuracy*, *Precision*, *Missclasification*, *Sensitivity*, *Specificty* y *F-Score* (en su estudio, uno de los más citados en la disciplina, Giachanou & Crestani (2016) solo citan 4, *Accuracy*, *Precision*, *Recall-Sensitivity* y *F-Score*).

- **Accuracy:** es la métrica más habitual en los trabajos analizados, junto el F-Score. Sirve para medir con qué frecuencia se logra hacer una predicción correcta con el método de clasificación que está siendo evaluado. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$.

- **Precision:** evalúa la exactitud del método de clasificación estudiado, y se calcula hallando la proporción de casos que se predijeron como positivos y lo fueron realmente, divididos por el número total de casos que se predijeron como positivos. $Precision = TP/TP+FP$.
- **Missclasification:** se usa con menor frecuencia, y representa la proporción de predicciones incorrectas con respecto al total. $Misclassification = FP+FN/TP+TN+FP+FN$.
- **Sentitivity** (en ocasiones deferida como *Recall*): es el True Positive Rate. $Sensitivity = TP/TP+FN$.
- **Specificity:** es el *True Negative Rate*, que se calcula con la fórmula $Specificity = TN/TN+FP$.
- **F-Score:** se trata de una métrica más elaborada y exigente que las anteriores, que se calcula combinando *Sensitivity* y *Precision*. Habitualmente, se considera la métrica más adecuada para evaluar con rigor el rendimiento de los métodos de clasificación.

$$F-Score = 2 \cdot \frac{precision \cdot sensitivity}{precision + sensitivity}$$

Figura 23 – Fórmula F-Score.

3.1.5.1 Resultados comparativos con herramientas más habituales

Existen diversos estudios comparativos sobre los métodos de *sentiment analysis*, sobre F-Score y otras métricas. El más citado es la técnica de benchmarking desarrollada por Ribeiro et al. (2016) (SentiBench). Las figuras 24 y 25 reflejan los resultados correspondientes a la aplicación de la técnica sobre una larga colección de métodos populares, donde podemos observar un rendimiento claramente

superior en la polaridad positiva, además de una considerable disparidad de resultados.

Dataset	Method	Accur.	Posit. sentiment			Negat. sentiment			Neut. sentiment			Macro-F1
			P	R	F1	P	R	F1	P	R	F1	
Tweets_Semeval	AFINN	62.36	61.10	70.09	65.28	44.08	55.56	49.15	71.43	58.57	64.37	59.60
	ANEW_SUB	39.51	38.79	96.31	55.31	43.50	23.18	30.24	57.38	2.31	4.45	30.00
	Emolex	48.74	48.15	62.71	54.47	31.27	38.59	34.55	57.90	41.30	48.21	45.74
	Emoticons	52.88	72.83	11.34	19.62	55.56	5.38	9.80	34.05	96.53	50.34	26.59
	Emoticons DS	36.59	36.55	100.00	53.53	75.00	0.36	0.71	100.00	0.03	0.07	18.10
	NRC Hashtag	36.95	42.04	75.03	53.88	24.57	56.03	34.16	53.33	3.70	6.92	31.65
	LIWC07	39.54	36.52	42.33	39.21	15.14	13.02	14.00	48.64	44.83	46.66	33.29
	LIWC15	62.56	59.77	71.03	64.91	49.04	42.65	45.62	68.90	61.84	65.18	58.57
	Opinion Finder	57.63	67.57	27.94	39.53	40.75	33.69	36.89	58.20	86.06	69.44	48.62
	Opinion Lexicon	60.37	62.09	62.71	62.40	41.19	52.81	46.28	66.41	60.75	63.46	57.38
	PANAS-t	53.08	90.95	9.04	16.45	51.56	3.94	7.33	51.65	99.01	67.89	30.55
	Pattern.en	57.99	57.97	68.74	62.89	34.83	35.24	35.04	65.55	56.39	60.63	52.85
	SASA	50.63	46.34	47.77	47.04	33.07	20.31	25.17	56.39	61.12	58.66	43.62
	Semantria	61.54	67.28	57.35	61.92	39.57	52.81	45.24	65.98	67.03	66.50	57.89
	SenticNet	49.68	51.85	1.26	2.46	29.79	1.67	3.17	49.82	98.51	66.17	23.93
	Sentiment140	60.42	63.87	51.37	56.94	50.96	37.87	43.45	60.35	73.31	66.20	55.53
	Sentiment140_L	39.44	43.52	74.72	55.00	27.67	65.35	38.88	65.87	6.38	11.63	35.17
	SentiStrength	57.83	78.01	27.13	40.25	47.80	23.42	31.44	55.49	89.89	68.62	46.77
	SentiWordNet	48.33	55.54	53.44	54.47	19.67	37.51	25.81	61.22	47.57	53.54	44.61
	SO-CAL	58.83	58.89	59.02	58.95	40.39	54.24	46.30	39.89	59.96	47.91	51.05
Stanford DM	22.54	72.14	18.17	29.03	14.92	90.56	25.61	47.19	6.94	12.10	22.25	
Umigon	65.88	75.18	56.14	64.28	39.66	55.91	46.41	70.65	75.78	73.13	61.27	
USent	52.13	49.86	32.88	39.63	39.96	22.82	29.05	54.33	74.36	62.79	43.82	
VADER	60.21	56.46	79.04	65.87	44.30	59.02	50.61	76.02	46.71	57.87	58.12	
Tweets_RND_III	AFINN	64.41	40.81	72.12	52.13	49.67	62.50	55.35	85.95	62.54	72.40	59.96
	ANEW_SUB	28.03	21.89	92.29	35.38	44.30	34.22	38.61	74.82	8.18	14.74	29.58
	Emolex	54.76	31.67	59.95	41.44	40.14	47.54	43.53	77.48	54.64	64.08	49.68
	Emoticons	70.22	70.06	16.78	27.07	65.62	8.61	15.22	41.29	97.56	58.02	33.44
	Emoticons DS	20.34	19.78	99.46	33.00	62.07	3.69	6.96	53.85	0.55	1.09	13.68
	NRC Hashtag	30.47	28.25	77.40	41.39	24.18	72.54	36.27	79.08	8.77	15.78	31.15
	LIWC	46.88	21.85	38.43	27.86	19.18	18.24	18.70	69.51	54.83	61.31	35.95
	LIWC15	67.75	44.78	78.35	56.99	57.49	57.38	57.44	85.18	66.67	74.80	63.07
	Opinion Finder	71.55	57.48	32.75	41.72	49.85	34.63	40.87	75.95	89.90	82.34	54.98
	Opinion Lexicon	63.86	40.65	66.17	50.36	48.84	56.15	52.24	81.96	64.66	72.29	58.30
	PANAS-t	68.79	79.49	8.39	15.18	48.57	3.48	6.50	68.75	98.86	81.10	34.26
	Pattern.en	59.56	36.20	77.00	49.24	52.87	45.29	48.79	81.75	57.23	67.33	55.12
	SASA	55.37	29.42	54.53	38.22	42.46	47.34	44.77	78.30	57.15	66.08	49.69
	Semantria	68.89	48.86	63.73	55.31	49.82	55.53	52.52	82.02	72.96	77.22	61.68
	SenticNet	29.97	31.08	74.83	43.92	20.98	73.98	32.68	79.70	8.49	15.35	30.65
	Sentiment140	76.40	64.42	51.69	57.36	74.75	45.49	56.56	79.04	89.50	83.94	65.95
	Sentiment140_L	31.32	25.83	77.13	38.70	30.05	78.69	43.49	79.37	8.92	16.04	32.74
	SentiStrength	73.80	70.94	41.95	52.72	57.53	25.82	35.64	75.35	92.26	82.95	57.10
	SentiWordNet	55.85	37.42	58.19	45.55	24.04	35.86	28.78	79.25	59.00	67.64	47.33
	SO-CAL	66.51	43.06	68.88	52.99	51.84	60.66	55.90	45.77	66.94	54.37	54.42
Stanford DM	31.90	64.48	38.57	48.26	15.58	85.04	26.33	75.64	19.77	31.35	35.32	
Umigon	74.12	57.67	70.23	63.33	48.83	68.44	57.00	88.80	76.34	82.10	67.47	
USent	66.06	40.60	36.81	38.61	44.87	28.69	35.00	74.54	81.72	77.97	50.53	
VADER	60.14	37.69	81.60	51.56	48.56	65.57	55.80	88.96	52.87	66.32	57.89	

Figura 24 - Precisión, Recall y F1 obtenido por Ribeiro et al. (2016) con su método de evaluación Sentibench.

Dataset	Method	Accur.	Posit. sentiment			Negat. sentiment			Neut. sentiment			Macro-F1
			P	R	F1	P	R	F1	P	R	F1	
Comments_BBC	AFINN	50.10	16.22	60.61	25.59	82.62	56.05	66.79	40.11	30.24	34.48	42.29
	ANEW_SUB	24.30	11.38	91.92	20.24	84.15	21.13	33.78	38.89	5.65	9.86	21.30
	Emolex	44.10	15.51	65.66	25.10	83.19	45.48	58.81	35.27	31.85	33.47	39.13
	Emoticons	24.60	0.00	0.00	0.00	33.33	0.15	0.30	19.77	98.79	32.95	11.09
	Emoticons DS	10.00	9.85	98.99	17.92	66.67	0.31	0.61	0.00	0.00	0.00	6.18
	NRC Hashtag	64.00	20.72	23.23	21.90	70.20	91.27	79.36	52.50	8.47	14.58	38.62
	LIWC07	33.00	11.11	42.42	17.61	67.69	33.69	44.99	22.90	27.42	24.95	29.18
	LIWC15	43.70	17.94	68.69	28.45	85.06	42.73	56.88	30.72	36.29	33.27	39.53
	Opinion Finder	51.80	14.96	35.35	21.02	78.76	60.18	68.23	33.71	36.29	34.95	41.40
	Opinion Lexicon	55.00	20.67	62.63	31.08	85.27	59.42	70.04	40.82	40.32	40.57	47.23
	PANAS-t	27.10	16.67	6.06	8.89	75.61	4.75	8.93	25.35	94.35	39.97	19.26
	Pattern.en	28.70	14.25	58.59	22.92	82.61	17.46	28.82	25.27	46.37	32.72	28.16
	SASA	38.20	17.03	47.47	25.07	70.75	36.29	47.98	25.19	39.52	30.77	34.60
	Semantria	56.00	28.90	50.51	36.76	83.82	57.12	67.94	35.86	55.24	43.49	49.40
	SenticNet	47.10	17.74	66.67	28.03	72.87	57.58	64.33	25.89	11.69	16.11	36.16
	Sentiment140	40.00	17.75	30.30	22.39	79.77	31.39	45.05	28.75	66.53	40.15	35.86
	Sentiment140_L	43.10	13.32	65.66	22.15	73.84	53.60	62.11	42.11	6.45	11.19	31.82
	SentiStrength	44.20	47.37	18.18	26.28	86.64	32.77	47.56	29.37	84.68	43.61	39.15
	SentiWordNet	42.40	14.90	59.60	23.84	81.63	41.50	55.03	34.56	37.90	36.15	38.34
	SO-CAL	55.50	20.88	57.58	30.65	80.47	63.09	70.73	28.57	34.68	31.33	44.23
Stanford DM	65.50	43.37	36.36	39.56	71.01	89.28	79.10	37.50	14.52	20.93	46.53	
Umigon	45.70	28.35	36.36	31.86	76.35	41.04	53.39	29.31	61.69	39.74	41.66	
USent	33.80	13.75	33.33	19.47	82.25	21.29	33.82	28.09	66.94	39.57	30.95	
VADER	49.40	16.36	71.72	26.64	83.02	54.67	65.93	48.53	26.61	34.38	42.31	
Comments_NYT	AFINN	42.45	64.81	41.79	50.81	80.29	39.82	53.24	7.89	77.87	14.32	39.46
	ANEW_SUB	51.12	48.35	88.57	62.55	79.65	24.69	37.69	7.92	9.84	8.78	36.34
	Emolex	42.97	55.12	53.72	54.41	75.35	33.33	46.22	7.22	54.10	12.74	37.79
	Emoticons	4.68	0.00	0.00	0.00	0.00	0.00	0.00	4.47	99.59	8.56	2.85
	Emoticons DS	42.58	42.55	99.77	59.66	78.57	0.40	0.80	0.00	0.00	0.00	20.15
	NRC Hashtag	54.84	55.38	45.74	50.10	61.55	65.68	63.55	8.33	15.16	10.76	41.47
	LIWC07	24.35	42.88	27.72	33.67	53.42	19.07	28.11	4.67	53.28	8.58	23.45
	LIWC15	36.49	65.29	40.29	49.83	81.50	29.25	43.05	7.17	83.61	13.20	35.36
	Opinion Finder	29.38	68.77	18.78	29.51	76.52	32.68	45.80	6.29	88.11	11.75	29.02
	Opinion Lexicon	44.57	65.95	43.15	52.17	79.81	43.11	55.98	7.94	73.77	14.34	40.83
	PANAS-t	5.88	69.23	1.23	2.41	62.07	1.31	2.57	4.75	99.18	9.07	4.68
	Pattern.en	31.60	55.23	45.05	49.63	72.80	17.76	28.55	5.88	65.57	10.79	29.66
	SASA	30.04	49.92	30.13	37.58	59.11	27.21	37.26	5.74	61.07	10.49	28.44
	Semantria	44.59	70.60	41.83	52.54	80.54	44.24	57.11	7.53	73.36	13.65	41.10
	SenticNet	61.85	58.19	59.48	58.83	65.01	69.26	67.07	0.00	0.00	0.00	41.97
	Sentiment140	13.58	77.32	6.81	12.51	75.40	11.96	20.65	4.98	93.03	9.45	14.20
	Sentiment140_L	54.61	54.72	59.12	56.84	67.00	54.41	60.05	6.70	15.98	9.44	42.11
	SentiStrength	18.17	78.51	8.62	15.54	81.12	18.96	30.74	5.41	95.49	10.24	18.84
	SentiWordNet	32.20	57.35	34.53	43.10	70.31	26.95	38.97	6.08	70.08	11.19	31.09
	SO-CAL	50.79	64.36	51.13	56.99	77.25	49.16	60.08	8.68	65.98	15.34	44.14
Stanford DM	51.93	73.39	21.14	32.83	59.48	77.90	67.46	9.65	38.11	15.40	38.56	
Umigon	24.08	68.76	16.38	26.46	68.78	24.51	36.14	5.88	88.93	11.04	24.54	
USent	27.44	56.61	28.95	38.31	77.69	21.59	33.79	5.88	79.51	10.94	27.68	
VADER	48.03	62.67	51.63	56.62	79.91	43.07	55.97	9.18	71.31	16.26	42.95	

Figura 25 (CONTINUACIÓN de la figura 24) – Precisión, Recall y F1 obtenido por Ribeiro et al. (2016) con su método de evaluación Sentibench.

En fecha más reciente, Dolianiti et al. (2019) realizaron un trabajo similar, aunque menos ambicioso (5 métodos) y orientado al dominio de la educación, obteniendo igualmente resultados muy dispares en F-Score (Figura 26). En este estudio, se observa también el sesgo hacia la polaridad positiva ya identificado por Ribeiro et al. (2016) tres años antes.

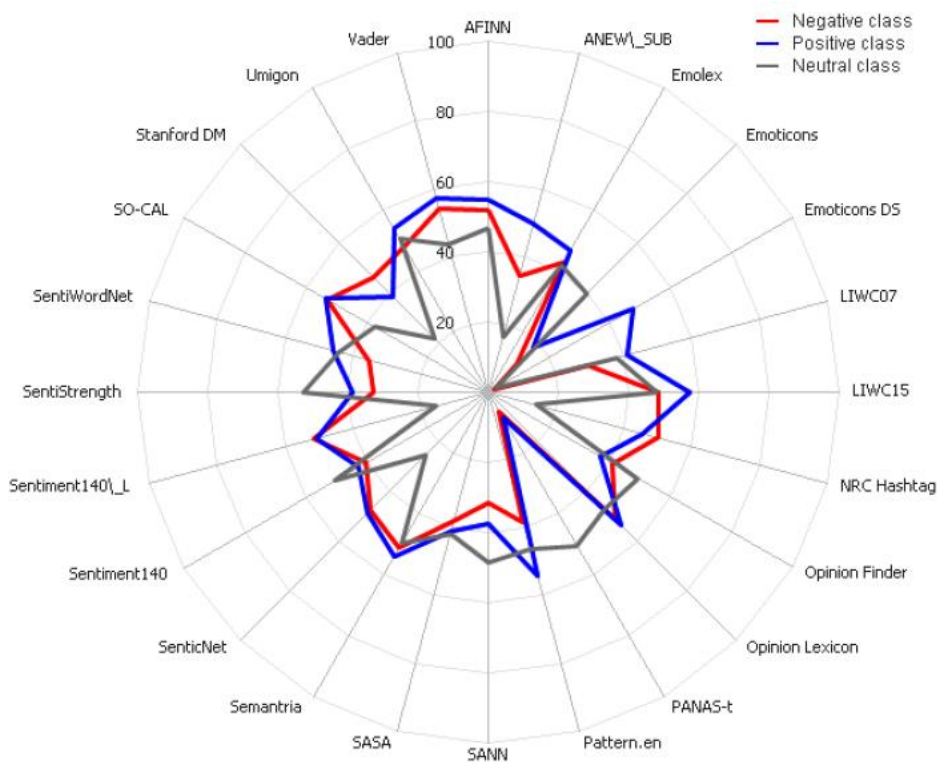


Figura 26 - F1 score medio obtenido para cada clase (negativa, positiva, neutra) en el estudio comparativo realizado por Dolianiti et al. (2019). Se observa cómo los métodos analizados consiguen un mejor rendimiento en predicción de mensajes positivos.

Tool	Document-level				Sentence-level			
	F-measure (%)							
	+tive	-tive	Neutral	Avg	+tive	-tive	Neutral	Avg
IBM Watson Natural Language Understanding	71.6	30.5	45.3	49.1	58.1	35.1	58.4	50.5
Microsoft Azure Text Analytics API	72.8	27.9	54.9	51.9	61.7	29.6	61.4	50.9
OpinionFinder 2.0	39.5	15.7	45.7	33.6	22.2	11	65.4	32.9
Repustate	69.6	41.8	34.3	48.5	60.6	32.4	48.4	47.1
Sentistrength	63.4	20	31.5	38.3	59.2	24.9	50.7	45
Educational domain-specific	76.8	42.3	70.5	63.2	69.5	48.8	80.6	66.3

Tabla 3 - F-Score obtenido por Dolianiti et al. (2019) en su comparativa para un mismo dataset, a nivel de frase y de documento.

3.2 Herramientas y recursos para la monitorización y análisis de la actitud de los usuarios en Internet

Miles de profesionales sin conocimientos técnicos de ciencia de datos realizan cada día acciones de monitorización y análisis, generalmente basados en IA, para entender la relación de los usuarios de Internet con las marcas, los productos, e incluso los partidos políticos y las instituciones públicas.

Microsoft explica así la propuesta de valor de su herramienta Text Analytics⁵¹ (analizada en profundidad más adelante):

“Extraiga información de texto no estructurado mediante NLP (sin necesidad de tener conocimientos de aprendizaje automático) usando Text Analytics, una colección de características de Cognitive Service para lenguaje. Conozca mejor la opinión de los clientes con el análisis de la opinión. Identifique frases clave y entidades, como personas, lugares y organizaciones, para conocer los temas y tendencias habituales. Clasifique la terminología médica usando modelos previamente entrenados específicos del dominio. Evalúe texto en un gran número de idiomas”

Como resultado de nuestra investigación, se han podido identificar distintos tipos de aplicaciones, que podríamos clasificar así ateniéndonos a principios metodológicos UX⁵²:

- Herramientas académicas.
- Herramientas comerciales adoptadas por la comunidad científica.
- Herramientas informales⁵³.

Los atributos básicos de la tipología descrita se desarrollan esquemáticamente en la Tabla 4:

⁵¹ <https://azure.microsoft.com/es-es/services/cognitive-services/text-analytics/#overview>

⁵² La categorización en Experiencia de Usuario (UX) no se basa en las características intrínsecas del sistema ni de la persona, sino en los atributos de la relación que se establece entre el usuario y el sistema, teniendo en cuenta el contexto.

⁵³ “Informales” no debe entenderse en un sentido peyorativo, sino en el sentido de los criterios de exigencia formal propios de la investigación científica, desde el punto de vista del método.

	Usuarios			
	Actividad profesional y académica	Cualificación requerida	Curva de aprendizaje (variable T)	Objetivos
Herramientas académicas	Investigadores	Muy alta	Larga	Experimentación académica, desarrollo metodológico, validación de hipótesis
Herramientas comerciales adoptadas por la comunidad científica	Investigadores y profesionales Data Science	Media	Corta	Innovación, desarrollo de productos y servicios, análisis avanzado de marketing y sociológico
Herramientas informales	Profesionales del mundo de la empresa y la política	Muy baja	Muy corta	Monitorización rutinaria en marketing y política

Tabla 4 - Clasificación de las herramientas empleadas para la monitorización y el análisis de la actitud de los usuarios en Internet.

Por otro lado, también se han estudiado conjuntos de recursos y entornos de trabajo ideados que son utilizados por los investigadores para crear sus propias herramientas y modelos: *toolkits*⁵⁴ y *frameworks*⁵⁵.

⁵⁴ Literalmente “paquete de herramientas”. Se trata generalmente de algoritmos y librerías de recursos para realizar tareas específicas útiles en tipo concreto de análisis, en este caso procesamiento de lenguaje natural. Su propósito es ahorrar trabajo al investigador evitando que él mismo tenga que escribir el código o generar sets de datos para realizar procesos cotidianos comunes.

⁵⁵ Framework es un “entorno de trabajo”. Generalmente, se refiere a un conjunto de herramientas que están accesibles desde un mismo interfaz de control.

Este tipo de herramientas tienen distintos niveles de comprensión y adopción entre los profesionales, dependiendo de su sector y otros factores. A este respecto, y a fin de conocer en profundidad la dinámica subyacente, se ha realizado una investigación específica basada en encuestas, cuyos resultados se exponen con detalle en el apartado siguiente:

3.2.1 Uso y percepción de las herramientas de monitorización y análisis de la actitud de los usuarios en Internet

A fin de responder a las preguntas de investigación y despejar incógnitas esenciales sobre los objetivos específicos de la tesis, se hace necesario tener una visión objetiva sobre la relación que se establece, ya sea en términos prácticos o cognitivos, entre las herramientas de análisis de la actitud de los usuarios en Internet y los ciudadanos como potenciales usuarios de las mismas (no como sujetos de la observación). A tal efecto, se realizó una encuesta (descrita en el Anexo 2), en torno a tres ejes de descubrimiento:

- Conocimiento: ¿saben que existen? ¿en qué medida depende el grado de conocimiento de su profesión?
- Uso: ¿las usan? ¿cómo?
- Percepción: ¿cuál es su actitud al respecto? ¿las consideran útiles? ¿creen que pueden ser usadas para manipular? ¿qué problemas o desafíos piensan que resuelven?

Veamos a continuación cada uno de estos aspectos.

3.2.1.1 Nivel de conocimiento

Al ser preguntados sobre si conocen la existencia de estas herramientas⁵⁶, el 86,6% de los encuestados respondieron afirmativamente.

Este porcentaje, sin duda es anormalmente elevado. Otros indicadores de la encuesta sugieren que la cifra real es considerablemente menor, y que simplemente algunos encuestados estaban incluyendo en la categoría herramientas inadecuadas, como por ejemplo los clásicos servicios de analítica web (u ocultando su desconocimiento, un tipo de sesgo muy común en encuestas a profesionales, incluso anónimas). Para confirmar esta hipótesis, observamos que el 66,2% de los encuestados confiesa no haber usado nunca estas herramientas y, de ellos, solo un 22% considera tener un nivel “avanzado” o “experto” (lo que supone un 12% del total de encuestados). De hecho, solo el 26% de los participantes contestó a la pregunta donde se solicitaba mencionar nombres de herramientas (tabla 5), y de ellos el 40% no fue capaz de nombrar ningún sistema “correcto” (que tenga realmente el propósito sobre el que se pregunta). La herramienta “errónea” más citada fue Google Analytics⁵⁷.

⁵⁶ Pregunta: "Las herramientas y métodos de monitorización de la actitud y la opinión de los usuarios en Internet sirven para recopilar masivamente textos publicados en redes sociales, blogs, foros, tiendas online, medios de comunicación, etc., y analizarlos de forma automatizada. Estas herramientas y métodos hacen posible que empresas, investigadores académicos, organizaciones políticas e instituciones públicas puedan monitorizar la opinión de los ciudadanos en Internet, con distintos fines." ¿Conocías la existencia de estas herramientas?

⁵⁷ Google Analytics es una herramienta gratuita de analítica web lanzada en 2005 por Google. Ofrece información estadística agregada sobre el tráfico que recibe una web, y no incluye funcionalidades relacionadas con el análisis de textos ni con el estudio de la opinión de los usuarios en Internet.

Comprobación de uso real de herramientas	Respuestas	% total
Afirman haber usado herramientas	48	33,80%
Contestan a "cuáles"	37	26,06%
Son capaces de citar alguna herramienta "correcta"	22	15,49%

Tabla 5 – Comprobación del uso “real” de herramientas entre los encuestados.

La herramienta profesional más mencionada, si nos atenemos solo a las que realmente tienen el propósito indicado, fue BrandWatch⁵⁸ con 5 menciones.

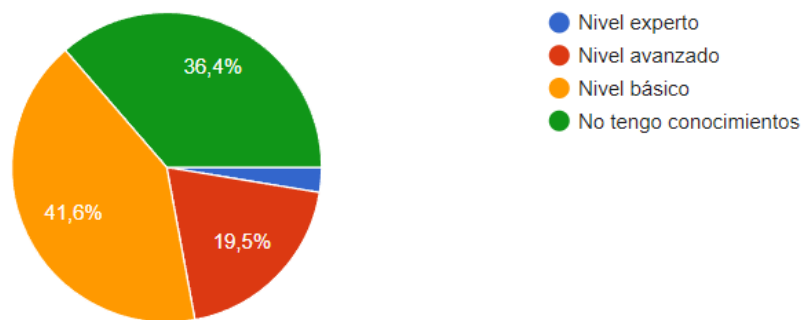


Figura 27 – Nivel de experiencia declarado por los encuestados que manifestaron haber usado herramientas.

Si segmentamos aquellos participantes que han sido capaces de mencionar herramientas correctas, comprobamos que el 86% de ellos son profesionales del marketing y la comunicación, y que un 36% desempeñan labores directivas o de gerencia (tabla 6).

⁵⁸ <https://www.brandwatch.com/es/>

Usuarios que son capaces de mencionar herramientas correctas		
Actividad profesional (pueden marcar más de una)	Respuestas	% de este segmento
Marketing y comunicación (agregado)	19	86%
Labores directivas o de gerencia	8	36%
Experiencia de Usuario (UX)	4	18%
Investigación de mercados	3	14%
Periodismo	2	9%
Psicología	1	5%
Customer intelligence	1	5%
Ciencia de datos	1	5%

Tabla 6 – Actividad profesional de los usuarios que son capaces de mencionar herramientas correctas.

Un hecho especialmente destacable es que solo uno de los participantes fue capaz de mencionar una herramienta académica (Moriarty). Esta herramienta está más orientada a la automatización de funcionalidades de *big data*, pero posee una librería con más de 150 elementos de trabajo desarrollados, que incluyen tanto algoritmos de *machine learning* como algoritmos de procesamiento del lenguaje natural o algoritmos semánticos de la web (Peña et al., 2016).

3.2.1.2 Percepción

La percepción de “utilidad” de los profesionales encuestados es muy alta, si tenemos en cuenta que la inmensa mayoría, pese a la diversidad y aleatoriedad de sus actividades profesionales, consideran que estas herramientas tienen relación con su profesión. Un 60,6% consideran que la relación es alta y solo un 2,8% aseguraron que no tenía “ninguna relación”.

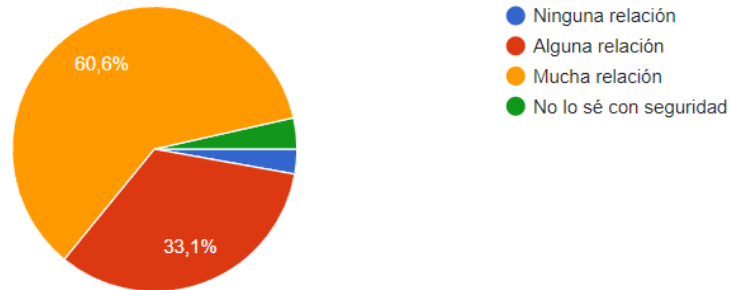


Figura 28 - Percepción de la relación de las herramientas de monitorización de la actitud con la actividad profesional propia.

La figura a continuación muestra en qué creen los usuarios que pueden ser útiles estas herramientas, en concreto, para su actividad profesional:

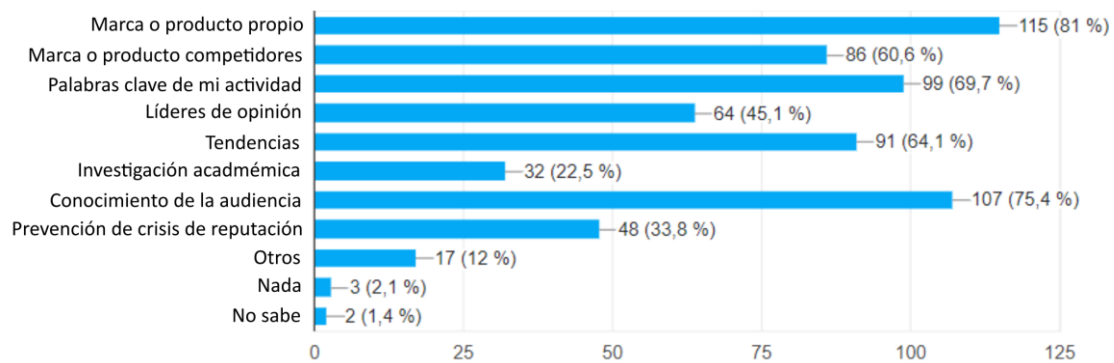


Figura 29 - Respuestas a la pregunta "¿En qué crees que pueden ser útiles estas herramientas para tus tareas profesionales o académicas? (Marca todas las que procedan)".

Un aspecto importante de la encuesta es conocer la percepción sobre las potenciales "amenazas" de este tipo de tecnologías. A este respecto, se preguntó a los

encuestados si les parecía que pueden usarse para manipular a la población, a lo que la inmensa mayoría (78,2%) respondieron afirmativamente.

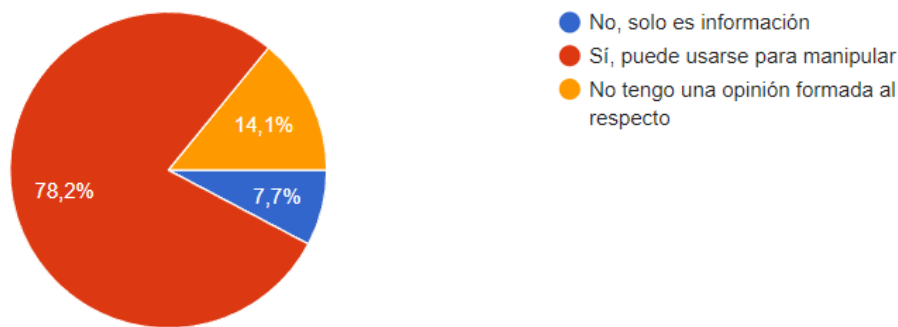


Figura 30 - Respuesta a la pregunta: ¿Crees que la información obtenida con herramientas de monitorización/análisis automatizado de la opinión de los usuarios en internet puede usarse para manipular a las personas? Marca la respuesta que mejor describa lo que piensas.

Más en concreto, en relación con la percepción de aspectos relacionados con la privacidad, se solicitó a los participantes que marcaran del 1 al 5 cómo de acuerdo estaban con la afirmación "El uso de herramientas de monitorización/análisis automatizado de la opinión de los usuarios en internet compromete la privacidad de las personas". Cerca de un 50% estuvieron de acuerdo en un grado de 4 ó 5 (figura 31). Sin embargo, al ser preguntados si ellos personalmente se sentían amenazados (una estrategia de prevención de sesgos), el porcentaje bajó al 35,2% (Figura 32).

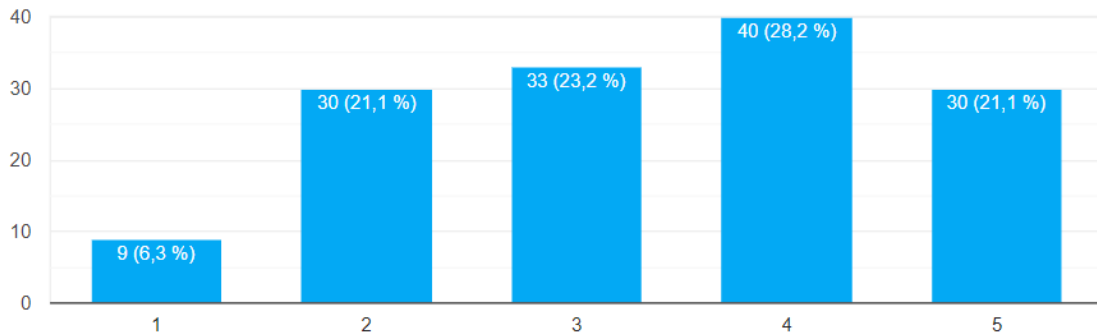


Figura 31 – Resultado de la pregunta: Marca del 1 al 5 cómo de acuerdo estarías con la siguiente afirmación "El uso de herramientas de monitorización/análisis automatizado de la opinión de los usuarios en internet compromete la privacidad de las personas".

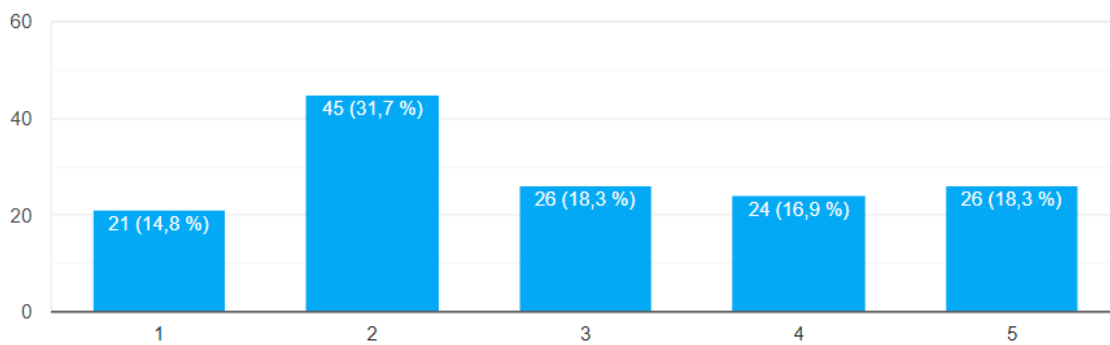


Figura 32 - Resultado de la pregunta: Marca del 1 al 5 cómo de acuerdo estarías con la siguiente afirmación "Como usuario me siento amenazado por la existencia de este tipo de herramientas"

Por último, desde una perspectiva más amplia, al margen de los intereses profesionales de los participantes, la figura siguiente muestra su percepción en relación con la utilidad “social” de las herramientas:

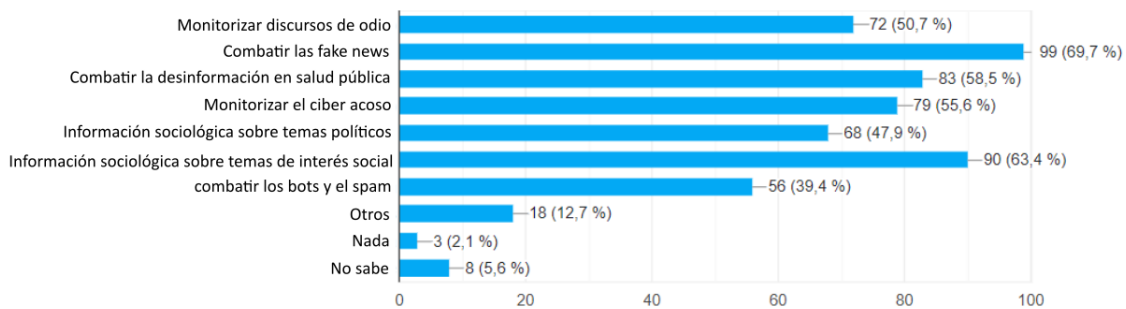


Figura 33 – Respuesta a la pregunta: ¿En qué crees que pueden ser útiles estas herramientas a la sociedad en su conjunto? (Marca todas las que procedan)

3.2.1.3 Uso

Los participantes que manifiestan haber usado alguna vez este tipo de herramientas de análisis, lo han hecho con fines relacionados con el producto/marca (77,7%). Un 11,7% lo han hecho sobre opiniones políticas y un 30,1% sobre temas de interés social (figura 34).

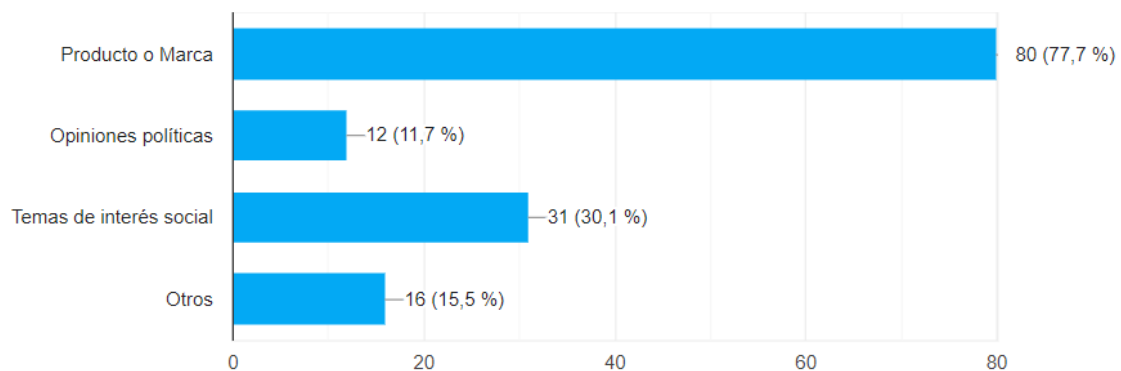


Figura 34 – Respuesta a la pregunta: Si alguna vez has accedido a información procesada con herramientas de monitorización/análisis de la opinión de los usuarios en Internet, ¿cuál era el objeto de la escucha? (Marca todas las que procedan).

El grado de satisfacción con los resultados es medio-bajo, si tenemos en cuenta que solo un 37,9% se manifestaron con un nivel alto o extremadamente alto (figura 35).

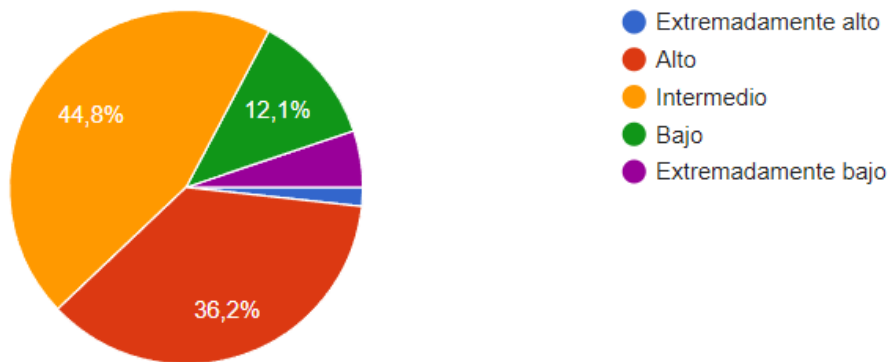


Figura 35 - Respuesta a la pregunta: si alguna vez has trabajado directamente con herramientas de monitorización/análisis para conocer la opinión de los usuarios de Internet, ¿cuál es tu nivel general de satisfacción con los resultados?

Por último, en cuanto al tipo de aplicación utilizada, predominan los servicios de modalidad gratuita, pero con una fuerte presencia de servicios de pago y un porcentaje significativo de “desarrollo propio” (figura 36). En este último caso, solo dos participantes fueron explícitos, indicando uno de ellos soluciones programadas en Python, y otras soluciones desarrolladas por IBM (probablemente en referencia a IBM Watson Studio ⁵⁹).

⁵⁹ <https://www.ibm.com/cloud/watson-studio>

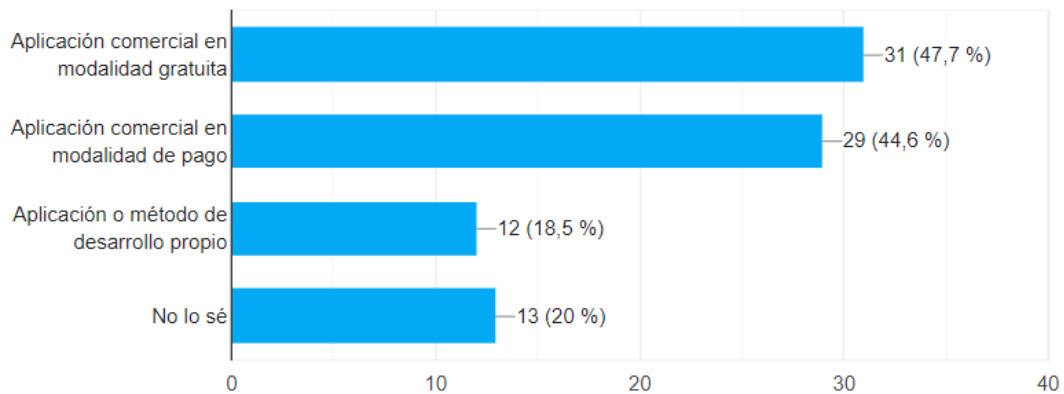


Figura 36 - Tipología de herramientas utilizadas por los encuestados.

3.2.2 Herramientas académicas

Como parte del trabajo de investigación, se ha realizado un análisis en profundidad de las principales herramientas de uso académico, con los resultados expuestos en la Tabla 7. Todas las herramientas analizadas son ampliamente citadas en la literatura científica, y están total o parcialmente disponibles en Internet, pero adoptan formas diferentes en su distribución. Mientras que algunas son paquetes descargables o accesibles vía API, otras son metodologías descritas por sus autores con algún tipo de repositorio disponible online (generalmente un *Lexicon*).

Herramienta	Descripción	Metodología de clasificación	Fuente original (literatura)
AFINN - ANEW (y extensiones) ⁶⁰	Construye un <i>lexicon</i> de sentimientos basado en Twitter, incluyendo <i>slang</i> y palabras gruesas. Califica las palabras en términos de placer, activación y dominancia. Cuenta con diversas extensiones que han ido complementando la versión original con el paso de los años.	<i>Lexicon</i>	(Finn Arup Nielsen, 2011)
LIWC ⁶¹	Herramienta de acceso disponible en formato API, creada por el Prof. James W. Pennebaker, de uso muy extendido en el ámbito de la psicología. Se centra en evaluar los componentes emocionales, cognitivos y estructurales de un texto determinado. Para ello, utiliza un diccionario con palabras clasificadas en categorías.	<i>Lexicon</i>	(Pennebaker & Tausczik, 2009)
Opinion Lexicon ⁶²	Diseñado para analizar reseñas de productos. Construye un léxico para predecir la polaridad de frases referidas a características de productos, ofreciendo una	<i>Lexicon</i>	(Hu & Liu, 2004)

⁶⁰ <https://osf.io/y6g5b/wiki/anew/> (repositorio en OSF)

⁶¹ <https://www.receptiviti.com/liwc>

⁶² <https://www.kaggle.com/datasets/nltkdata/opinion-lexicon> (Software en repositorio Kaggle)

	puntuación global por característica.		
OpinionFinder ⁶³	OpinionFinder realiza en secuencia las siguientes fases: preprocesamiento, identificación de frases subjetivas y expresiones de sentimiento, clasificación de la subjetividad y clasificación de la polaridad. Para evitar problemas de compatibilidad e incorporar otras mejoras, la veterana herramienta original (2005) fue renovada y sustituida por OpinionFinder 2, escrito en Java y funcional en cualquier plataforma.	<i>Lexicon & Machine Learning</i>	(Wilson et al., 2005)
SenticNet ⁶⁴	Se orienta hacia el análisis de sentimiento a nivel de concepto, con detección de la polaridad y reconocimiento de emociones, usando para ello la información denotativa y connotativa asociada a las palabras y a las expresiones, en lugar de basarse únicamente en las frecuencias de coocurrencia. Utiliza un método de reducción de dimensionalidad para inferir la polaridad de los conceptos, proporcionando un recurso para extraer opiniones a nivel semántico y no solo sintáctico.	<i>Lexicon</i>	(Cambria et al., 2014)

⁶³ <https://mpqa.cs.pitt.edu/opinionfinder/>

⁶⁴ <https://sentic.net/>

	Como base de conocimientos, SenticNet proporciona un conjunto de semántica y polaridad asociado a 400.000 conceptos del lenguaje natural.		
SentiStrength ⁶⁵	Popular herramienta que combina un diccionario anotado con el uso algoritmos de <i>machine learning</i> . Evalúa por intensidad la polaridad del sentimiento en textos cortos (Twitter, reseñas de productos, etc.), aceptando lenguaje informal. La herramienta está desarrollada para el idioma inglés, pero investigadores de diversos países han creado extensiones para otros idiomas, incluido el español. Gratuita con fines académicos y de pago para uso comercial.	<i>Lexicon Machine Learning</i>	& (Thelwall, 2017)
SentiWordNet ⁶⁶	Trabaja en combinación con Wordnet para crear un lexicon. Posteriormente, se agrupan los elementos de cada frase en conjuntos de sinónimos (synsets) y a cada uno de ellos se asocia una polaridad (con puntuación de intensidad).	<i>Lexicon Machine Learning</i>	& (Baccianella et al., 2010)

⁶⁵ <http://sentistrength.wlv.ac.uk/>

⁶⁶ <https://github.com/aesuli/SentiWordNet> (Software en repositorio GitHub)

<p>SO-CAL⁶⁷</p>	<p>SO-CAL (Semantic Orientation CALculator) toma como entrada un texto (o una frase) y devuelve un valor numérico de la polaridad y la fuerza global del texto, en función de las opiniones expresadas en él. Crea un lexicon con unigramas y multigramas, clasificados con una escala de intensidad. Incluye un método de tratamiento de negaciones, intensificadores y “part of speech”. Esta herramienta, de acceso restringido, forma parte de un proyecto de investigación de la Simon Fraser University, dirigido por Maite Taboada.</p>	<p><i>Lexicon</i></p>	<p>(Taboada et al., 2011)</p>
<p>Twitter Sentiment (Sentiment140) ⁶⁸</p>	<p>“Ensemble method” que combina tres clasificadores: (SVM, Naive Bayes y Maximum Entropy)</p>	<p><i>Machine Learning</i></p>	<p>(Go et al., 2009)</p>
<p>Vader⁶⁹</p>	<p>VADER (Valence Aware Dictionary and sEntiment Reasoner) es un lexicon y una herramienta de <i>sentiment analysis</i> que se adapta de forma específica</p>	<p><i>Lexicon</i></p>	<p>(Hutto & Gilbert, 2014)</p>

⁶⁷ <https://www2.cs.sfu.ca/~sentimen/socal/>

⁶⁸ <https://www.kaggle.com/datasets/kazanov/sentiment140> (Repositorio en Kaggle)

⁶⁹ <https://github.com/cjhutto/vaderSentiment> (Respositorio GitHub)

	<p>a los sentimientos tal y como se encuentran habitualmente expresados en las redes sociales. VADER no solo aporta información sobre la polaridad, sino también de la intensidad del sentimiento.</p>		
BERT ⁷⁰	<p>BERT está diseñado para entrenar previamente representaciones bidireccionales profundas a partir de texto sin etiquetar, mediante el condicionamiento conjunto del contexto en todas las capas. El modelo preentrenado puede ser sometido a un “ajuste fino” con solo una capa de salida adicional para crear modelos de última generación para en amplia gama de tareas, sin necesidad de realizar apenas modificaciones estructurales específicas para la tarea</p>	<i>Machine Learning</i>	(Devlin et al., 2019)

Tabla 7 – Principales herramientas de uso académico

⁷⁰ <https://colab.research.google.com/drive/1PHv-IRLPCtv7oTclGbsgZHqrB5LPvB7S>

3.2.3 Herramientas específicas de la comunidad investigadora hispanohablante

Además de las herramientas anteriormente analizadas, todas ellas con un amplio soporte de la comunidad investigadora internacional, identificamos también otras específicamente desarrolladas en España con módulos específicos de lenguas locales, como el catalán y el gallego, referenciadas por investigadores hispanohablantes. Pasamos a analizarlas a continuación.

3.2.3.1 *Freeling*

El proyecto FreeLing está liderado por el investigador Lluís Padró (UPC), y surgió como una iniciativa para poner a disposición de la comunidad los resultados de la investigación llevada a cabo en el grupo de investigación NLP de la UPC (Padró, 2012).

Básicamente, FreeLing es una librería en C++ que proporciona funcionalidades de análisis lingüístico:

- Análisis morfológico
- Análisis sintáctico
- Detección de entidades con nombre
- Etiquetado PoS
- Desambiguación
- Etiquetado semántico de roles

Funciona actualmente para un amplio catálogo de idiomas, entre ellos: español, inglés, catalán, portugués, italiano, francés, alemán, y ruso.

Adicionalmente, FreeLing también proporciona un *front-end* de línea de comandos que puede utilizarse para analizar textos y obtener la salida en el formato deseado (XML, JSON, CoNLL).

3.2.3.2 *LinguaKit*

Otra destacable iniciativa en España es LinguaKit ⁷¹, una suite multilingüe de herramientas de análisis, extracción, anotación y corrección lingüística. LinguaKit permite realizar diferentes tareas como el etiquetado PoS (Parts of Speech) o el análisis sintáctico, incluyendo aplicaciones para la extracción de relaciones, el análisis de sentimiento, la síntesis, y la vinculación de entidades, entre otras. La mayoría de los módulos disponibles son operativos en cuatro idiomas: español, inglés, gallego y portugués. Su código está programado en Perl y se encuentra disponible gratuitamente bajo una licencia GPLv3 (Gamallo et al., 2018).

Esta plataforma presenta una serie de módulos lingüísticos en cuatro apartados: el primero de ellos abarca aspectos genéricos como la conjugación o la traducción, el segundo está más ligado al ámbito educativo, con etiquetadores y analizadores morfosintácticos, mientras que el tercero analiza sentimiento y extrae palabras clave. Además, LinguaKit incluye un apartado experimental con nuevas herramientas adicionales en desarrollo.

3.2.4 Herramientas comerciales empleadas por la comunidad científica

En los últimos años han ido apareciendo gradualmente aplicaciones comerciales de gran potencia, que están siendo adoptadas por la comunidad científica como solución alternativa al desarrollo de metodologías específicas *ad hoc* de elaboración propia. Estas herramientas comerciales están siendo objeto de estudio en nuevos trabajos de reciente publicación, ya sea independientemente o en formato *benchmarking*, como por ejemplo en los trabajos realizados por Dolianiti et al. (2019). En su estudio, Doliantini et al. centran su análisis comparativo en cinco

⁷¹ <https://linguakit.com/es/>

“*well-known general-purpose sentiment analysis tools*”, mezclando herramientas comerciales de reciente desarrollo con veteranos recursos científicos:

- Repustate⁷²
- Microsoft Azure Text Analytics⁷³
- OpinionFinder 2.0⁷⁴
- IBM Watson Natural Language Understanding (IBM Watson NLU) ⁷⁵
- Sentistrength⁷⁶

Otra herramienta muy citada, incluida por Ribeiro et al. (2016) en la presentación de su metodología de benchmarking “Sentibench”, es Semantria (Lexalytics).

Veamos a continuación cuáles son las características de las principales aplicaciones comerciales de uso más extendido en la actualidad, a partir del análisis realizado para la investigación de esta tesis, incluyendo también información sobre las conclusiones de otros investigadores en sus trabajos experimentales.

3.2.4.1 *Microsoft Azure Text Analytics*

Azure Text Analytics es una herramienta de Microsoft que forma parte de *Azure Cognitive Service for Language*, un servicio ubicado en la nube de Microsoft (Azure), que ofrece funciones de procesamiento del lenguaje natural con diversos fines. Probablemente debido a su reciente creación, su uso por la comunidad

⁷² <https://www.repustate.com/>

⁷³ <https://azure.microsoft.com/es-es/services/cognitive-services/text-analytics/#overview>

⁷⁴ http://mpqa.cs.pitt.edu/opinionfinder/opinionfinder_2/

⁷⁵ <https://www.ibm.com/es-es/cloud/watson-natural-language-understanding>

⁷⁶ <http://sentistrength.wlv.ac.uk/>

investigadora hasta ahora ha sido marginal y solo aparece referenciado en uno de los trabajos de benchmarking que hemos analizado (Dolianiti et al., 2019), razón por la que ha sido elegido para esta comparativa (en aras del componente novedoso de la investigación). Pese a la escasa bibliografía, hemos podido estudiar trabajos muy recientes apoyados en este servicio, como por ejemplo la investigación de Roe et al. (2021) sobre la percepción de las campañas de vacunación del Covid-19, un asunto que también hemos utilizado en uno de los *datasets* del presente estudio.

Azure Cognitive Service for Language es un servicio ubicado en la nube de Microsoft (Azure), que ofrece funciones de procesamiento del lenguaje natural con diversos fines. El servicio unifica tres herramientas, Text Analytics, QnA Maker y LUIS, y además proporciona otras funcionalidades nuevas, algunas ya disponibles y otras en desarrollo. La tabla 8 resume las funcionalidades disponibles y qué tipo de utilidad aportan a la comunidad. Se puede observar cómo, junto a funcionalidades orientadas al análisis de sentimiento, se ofrecen utilidades específicamente pensadas para la automatización de procesos documentales y conversacionales (robotizados) que requieran comprensión del lenguaje.

<i>Sentiment Analysis and opinion mining</i>	Genera etiquetas de opinión con polaridad (positiva, negativa, neutra) tanto a nivel de frase como de documento, pudiendo discriminar sentimiento por atributos de productos y servicios.
Clasificación de texto personalizada	Cree un modelo de IA para clasificar texto no estructurado en las clases personalizadas que defina.
Detección de idioma	Esta característica devuelve un identificador de idioma y una puntuación de intensidad.

PII recognition (Personally Identifiable Information)	Identificación de entidades con información personal (potencialmente confidencial), para varias categorías.
Key phrase extraction	Evaluación de textos para la identificación de frases clave y puntos principales.
Entity linking	Supresión de la ambigüedad en las entidades que se encuentran en el texto.
Text Analytics for Health	Esta característica preconfigurada extrae información de textos médicos no estructurados, como apuntes clínicos y notas del médico.
NER personalizado	Crea modelos IA para extraer categorías de entidades personalizadas.
NER (Named entity recognition)	Identificación de entidades en el texto.
Resumen documental	Extrae frases clave de un documento para generar un resumen de los puntos esenciales del contenido.
Resumen conversacional	Sintetiza las transcripciones de las conversaciones.
Reconocimiento del lenguaje conversacional	Cree modelos IA para comprender el lenguaje natural (conversacional).

OW (Orchestration Workflow)	Entrena modelos lingüísticos para conectar aplicaciones con sistemas de respuesta a preguntas, comprensión del lenguaje conversacional y LUIS
--	---

Tabla 8 – Funcionalidades de Azure Cognitive Service for Language.

En particular, Text Analytics se basa fundamentalmente en *machine learning*, y admite varios idiomas. Las características del clasificador utilizado incluyen n-gramas, etiquetas *part-of-speech* y *word embeddings*.

En un intento por llevar al extremo la accesibilidad del servicio, haciendo posible que personas sin conocimientos técnicos (ni académicos en general) puedan realizar análisis de sentimiento basado en IA con metodologías *state-of-the-art*, Microsoft ha creado un complemento gratuito instalable en Excel, de uso extremadamente sencillo. La ilustración siguiente (Figura 37) muestra con claridad la extrema sencillez del proceso requerido:

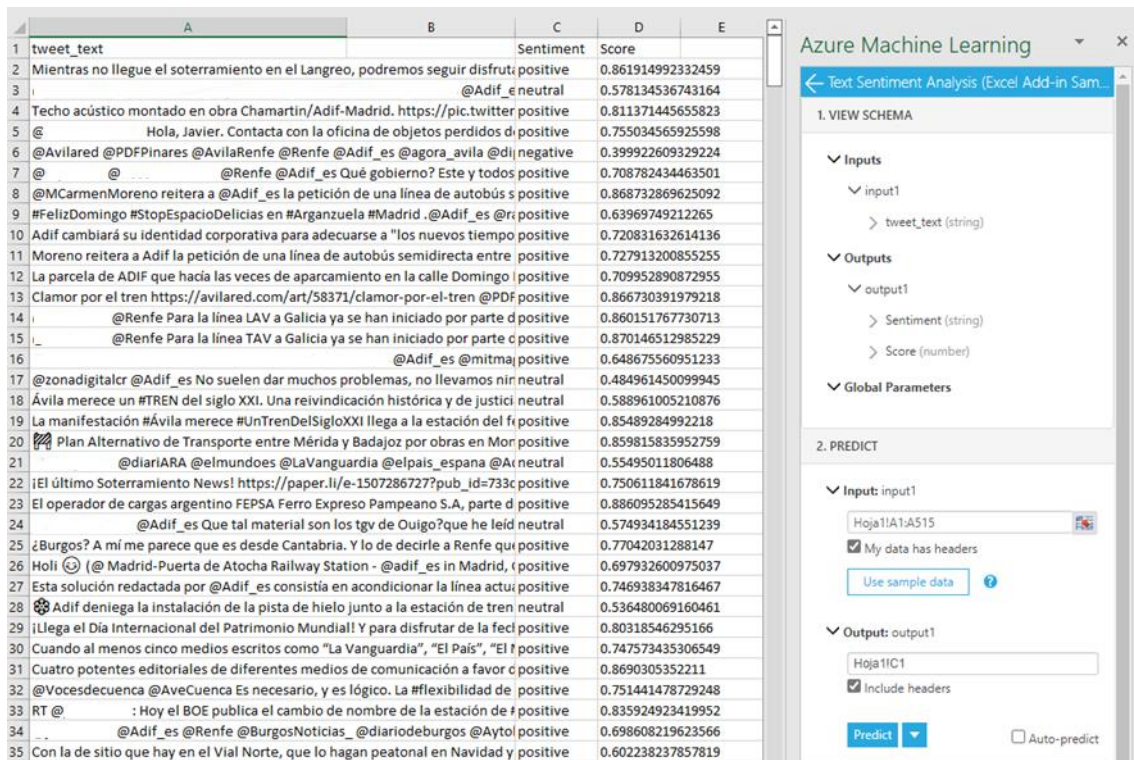


Figura 37 – Captura de pantalla que muestra el interfaz de Sentiment Analysis de Azure Machine Learning, integrado en Excel.

En la evaluación (*benchmarking*) de Dolianiti et al. (2019), Azure Text Analytics muestra un rendimiento muy similar a su directo competidor IBM Watson (F-Score 50.90).

El coste de Cognitive Services de Microsoft (paquete donde se encuentran las funcionalidades de análisis de texto mediante IA), se sitúa a partir de los \$700 / mes en el momento de realizar este estudio⁷⁷.

⁷⁷ Última consulta realizada en julio de 2022.

3.2.4.2 IBM Watson Natural Language Understanding

IBM Watson es una herramienta de “comprensión del lenguaje natural” que tiene funcionalidades de *sentiment analysis*. Está basada en API es y capaz de funcionar con diversos idiomas, en constante ampliación. Esta aplicación forma parte del proyecto DeepQA de IBM, liderado por David Ferrucci. El nombre de “Watson” es un homenaje a al fundador y primer presidente de IBM, Thomas J. Watson.

Watson se creó originalmente como un sistema de respuesta a preguntas basado en IA (QA), desarrollado por IBM para aplicar tecnologías avanzadas de procesamiento del lenguaje natural, recuperación de información, representación del conocimiento, razonamiento automatizado y *machine learning* ⁷⁸.

En torno a Watson, IBM ha creado un ecosistema de aplicaciones, algunas específicamente orientadas a sectores. En este sentido cabe destacar por su impacto en la comunidad investigadora IBM Watson Health, una herramienta orientada a procesamiento del lenguaje natural, generación de hipótesis y aprendizaje basado en la evidencia, para contribuir a los sistemas de apoyo a la investigación y procesos de toma de decisiones clínicas ⁷⁹.

Podemos comprobar que la propuesta de valor de IBM Watson Natural Language Understanding, tal y como aparece publicada en su documentación⁸⁰, hace hincapié en todos los aspectos que preocupan a la comunidad investigadora de la actitud de los ciudadanos en Internet:

⁷⁸ https://researcher.watson.ibm.com/researcher/view_group.php?id=2099

⁷⁹ https://en.wikipedia.org/wiki/IBM_Watson_Health

⁸⁰ <https://www.ibm.com/es-es/cloud/watson-natural-language-understanding>

“Analice texto para extraer metadatos de contenido, tales como conceptos, entidades, palabras clave, categorías, sentimientos, emociones, relaciones y roles semánticos utilizando la comprensión del lenguaje natural.”

Este servicio de IBM es parte del paquete de prestaciones disponibles bajo el paraguas “IBM Cloud”, y puede disfrutarse de forma gratuita por los usuarios, en su modalidad básica, para pequeños volúmenes de consulta. Para un uso a escala profesional, hay disantos precios, a partir de \$140/mes⁸¹.

3.2.4.3 *Semantria*

Semantria⁸² (referenciada en algunos trabajos como “Lexalytics”) es una herramienta comercial disponible en formato API, que aplica un análisis multinivel a las frases: *part of speech tagging*, asignación de puntuaciones previas basadas en diccionarios, aplicación de intensificadores y, finalmente, clasificación mediante técnicas de *machine learning* para otorgar un peso final a la frase.

Ribeiro et al. (2016), con su metodología de evaluación “Sentibench”, obtuvo un Macro F-1 de 49.40 para Semantria en su trabajo de *benchmarking*.

Semantria tiene un modelo de *pricing*⁸³ sobre presupuesto *ad hoc*, por lo que no es posible dar una referencia. Hasta donde resulta posible dar un dato sin vulnerar el secreto de los acuerdos comerciales, podemos indicar que el coste de Semantria es similar al de otras aplicaciones equivalentes para funcionalidades equiparables, y se

⁸¹ Última consulta: julio de 2022.

⁸² <https://www.lexalytics.com/semantria/>

⁸³ Pricing: estructura o política de precios (término en inglés).

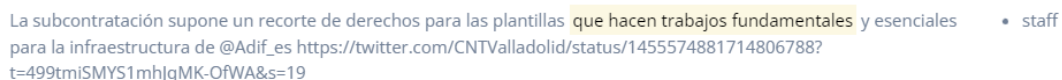
encuentra dentro del rango de contrataciones SAS⁸⁴ habituales en cualquier empresa media.

3.2.4.4 Repustate

Repustate⁸⁵ es una herramienta comercial muy orientada a monitorización de menciones de marca en usuarios y empleados, que ofrece la posibilidad de utilizar la API sin necesidad de codificación. Admite varios idiomas, múltiples niveles de análisis y la configuración de reglas específicas de sentimiento. En el Anexo 3 se ofrecen resultados de una investigación realizada para esta tesis donde se compara esta herramienta con otros recursos comerciales.

En el estudio realizado para la investigación de esta tesis (Anexo 3), Repustate obtiene un F-Score medio de 46.92, muy cerca del 47.13 obtenido por Dolianiti et al. (2019) en un trabajo similar.

Repustate ofrece una variedad muy considerable de características, en comparación con otras herramientas. En particular, permite crear *aspect models* personalizados, mediante conjuntos de reglas (o usar los que tiene predefinidos por sectores).



La subcontratación supone un recorte de derechos para las plantillas que hacen trabajos fundamentales y esenciales • staff para la infraestructura de @Adif_es <https://twitter.com/CNTValladolid/status/1455574881714806788?t=499tmiSMYS1mhjqMK-OfWA&s=19>

Figura 38 – Ejemplo de aplicación de un modelo de aspectos para identificar atributos específicos en el texto, con Repustate.

⁸⁴ SAAS: acrónimo de Software As A Service. Se refiere a servicios donde la empresa usuaria utiliza remotamente un software como servicio externo, sin integrarlo por tanto en sus propios sistemas, mediante el pago de una cuota.

⁸⁵ <https://www.repustate.com/>

En la siguiente gráfica (Figura 39), vemos un ejemplo analizado para el desarrollo de esta investigación, donde se estudia la polaridad del sentimiento no por mensaje, sino por las distintas entidades, sobre una lista predeterminada de aspectos relevantes para el sector, en este caso hostelero:

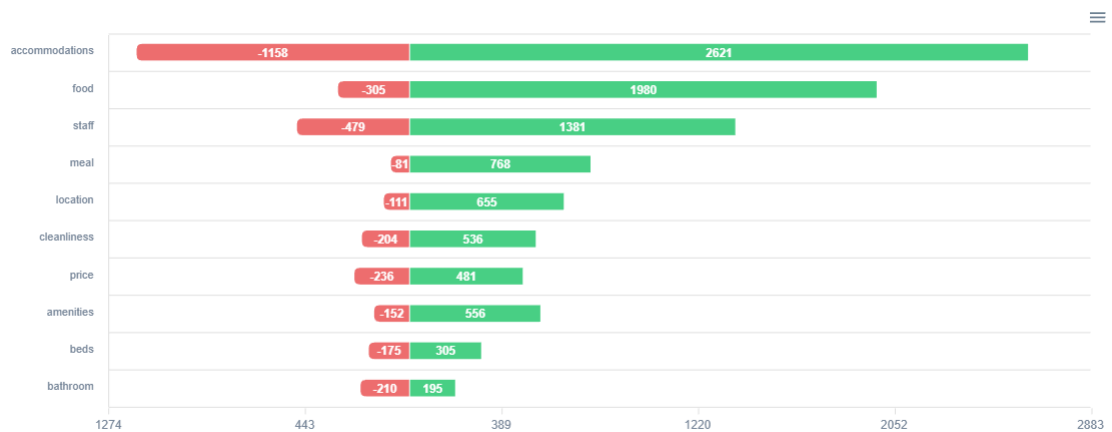


Figura 39 – Análisis de sentimiento por entidades, sobre lista predeterminada de intereses sectoriales, con Repustate (en este caso, opiniones de usuarios de hotel).



Figura 40 - Nube de palabras asociada a un proyecto de monitorización con Repustate. Este sistema de representación visual permite obtener una visión clara de los términos que más relacionan los usuarios con el concepto (palabras clave) objeto de monitorización (en este caso un servicio público).

3.2.5 Herramientas comerciales informales

Las herramientas comerciales informales son de especial importancia en nuestra investigación, ya que son las más accesibles a “cualquier usuario” con conocimientos básicos (en cierto modo, las únicas). También son las menos estudiadas, y de hecho ha sido preciso realizar varias investigaciones específicas de nueva planta para evaluarlas desde un punto de vista científico. Por estos motivos, es importante prestarles una atención especial si queremos satisfacer en su totalidad los objetivos planteados.

A este respecto, nos planteamos tres desafíos para responder a las preguntas de investigación y contextualizar el análisis: determinar cuál es el uso que hace la comunidad investigadora de las herramientas comerciales “informales”, conocer el nivel de rendimiento que podemos esperar (*F-Score*, *Recall*, *Accuracy*), y saber cuál

es la relación de la comunidad académica y la no académica con este tipo de servicios.

3.2.5.1 Uso de herramientas comerciales informales por la comunidad investigadora

Para aplicar un enfoque lo más objetivo y riguroso posible a esta cuestión, se ha extraído de la encuesta realizada entre profesionales descrita en un apartado anterior (detalle en Anexo 2) una lista herramientas con funcionalidades de *sentiment analysis* para medir en términos estadísticos su impacto en la comunidad investigadora. A fin de conocer en qué medida la comunidad académica utiliza, de manera relevante, estos kits en sus investigaciones, se realizó una prospección de trabajos académicos que las mencionan en sus *abstracts*, con el siguiente resultado⁸⁶:

- Brandwatch: 14
- Awario: 7
- SentiOne: 6
- Brand24: 4
- Symanto: 3
- Synthesio: 2
- Sproutsocial: 1
- Atribus: 0
- Brandmentions: 0
- Digimind: 0

Hay que tener en cuenta que este escaso número de menciones se verifica sobre la totalidad de la base de datos de Scopus, para cualquier dominio. Si limitamos la

⁸⁶ Prospección realizada en julio de 2022, en la base de datos de SCOPUS, con la única condición de coincidencia de la palabra clave en el abstract.

prospección a trabajos específicamente orientados a *sentiment analysis*, la cifra más alta la obtiene SentiOne con solo tres menciones.

Comparemos ahora estas cifras con la frecuencia de mención de recursos “académicos” con funcionalidades de *sentiment analysis* empleados habitualmente por la comunidad investigadora ⁸⁷. La diferencia es abrumadora:

- BERT: 7455
- NLTK: 386
- Stanford NLP: 213
- Sentistrength: 118

3.2.5.2 *Sentione*

SentiOne⁸⁸ es una potente herramienta de *Social Listening*, fundamentalmente orientada al marketing, que cubre la totalidad de la Red. Cuenta con ingentes recursos de almacenamiento y procesado para ofrecer acceso retrospectivo de hasta 3 años a todo lo publicado en las principales redes sociales, medios de comunicación, blogs, y foros (entre otros).

SentiOne incorpora funcionalidades de análisis de sentimiento para permitir a sus usuarios conocer la polaridad de los comentarios sobre cualquier asunto, basándose en filtros de palabras clave. Correctamente configuradas las búsquedas, estas *keywords*, que pueden ser positivas (incluir) o negativas (excluir), permiten conocer la polaridad de la opinión de la población sobre prácticamente cualquier cosa, en un período de tiempo determinado, o en tiempo real. Ejemplos de uso: monitorización

⁸⁷ De nuevo, prospección realizada en julio de 2022, en la base de datos de SCOPUS, con la única condición de coincidencia de la palabra clave en el abstract.

⁸⁸ <https://sentione.com/>

de marcas, tendencias, personajes públicos, partidos políticos, temas de actualidad, etc.

Entre los clientes de SentiOne se encuentran grandes multinacionales como Procter & Gamble, Pepsico, T-Mobile y LG.

En cuanto al precio de la herramienta, no es posible revelar el coste de la versión *Entreprise* (utilizada para esta investigación) con todas las funcionalidades⁸⁹ por razones de confidencialidad comercial. No obstante, la versión convencional tiene un precio de \$299/mes⁹⁰ y permitiría realizar las mismas funciones que las empleadas en la investigación, con ciertas limitaciones de volumen.

Hasta ahora, el impacto de esta herramienta en la comunidad científica es muy limitado, debido a la imposibilidad de conocer/documentar con precisión (o alterar) la metodología subyacente, y a la falta de información sobre la precisión y fiabilidad de los resultados⁹¹. Pese a todo, existen algunos trabajos interesantes realizados con esta herramienta, que utilizan su potencial de *scrapping* (recopilación de textos de Internet) pero no las funcionalidades de *sentiment analysis*. Como ejemplo, es especialmente representativa la investigación de Burzyńska et al. (2020) sobre la monitorización del fenómeno Covid-19 en las redes sociales.

En los anexos 1 y 3 se presentan investigaciones que incluyen una evaluación inédita del rendimiento de esta herramienta.

⁸⁹ El precio no es público, el proveedor lo negocia *ad hoc* en función de las necesidades de la empresa.

⁹⁰ Última comprobación de este precio: julio de 2022.

⁹¹ Se solicitó información técnica a la empresa propietaria sobre la metodología de *sentiment analysis* de la herramienta, así como de evaluación de precisión, pero por razones de legítimo secreto de empresa y patentes, no estaba disponible esa posibilidad.

3.2.5.3 *Brandwatch*

Brandwatch⁹² es una herramienta muy citada en la encuesta realizada para esta tesis (Anexo 2), y una referencia muy habitual en los contenidos informales sobre *social listening* que se pueden encontrar en Internet. Por estos motivos, ha sido uno de los recursos que hemos analizado para la realización de la presente investigación.

Una revisión bibliográfica *ad hoc* reveló algunos trabajos académicos que utilizan Brandwatch como parte de su metodología, pero en su mayoría son estudios relacionados con percepción de marca y sin conexión con el procesamiento del lenguaje. Por ejemplo, en un estudio muy reciente sobre el impacto del Covid-19 en las redes sociales, Anderson et al. (2022) utilizaron Brandwatch para capturar textos publicados por los usuarios sobre la pandemia a fin de estudiar la adopción de la telemedicina durante la crisis epidémica. El análisis realizado en este trabajo está realizado desde la perspectiva académica de las ciencias de la salud (no de la ciencia de datos) y se centra en identificar temas (*topics*) sin involucrar metodologías de inteligencia artificial ni procesamiento del lenguaje en general.

Las funcionalidades de Brandwatch son similares a las de SentiOne en el apartado del análisis de sentimiento, pero su alcance retrospectivo es mayor (dependiendo del tipo de contrato) y posee además otras características más enfocadas a la definición, monitorización y perfilado de audiencias específicas, en la línea de Audiense (herramienta comentada a continuación). Como muestra la imagen siguiente, la brandwatch nos permitiría, por ejemplo, filtrar a profesionales de la salud que hablan de Covid-19, para posteriormente analizar atributos comunes de

⁹² <https://www.brandwatch.com/es/>

estas personas, como aficiones, otros temas de conversación, o características demográficas.

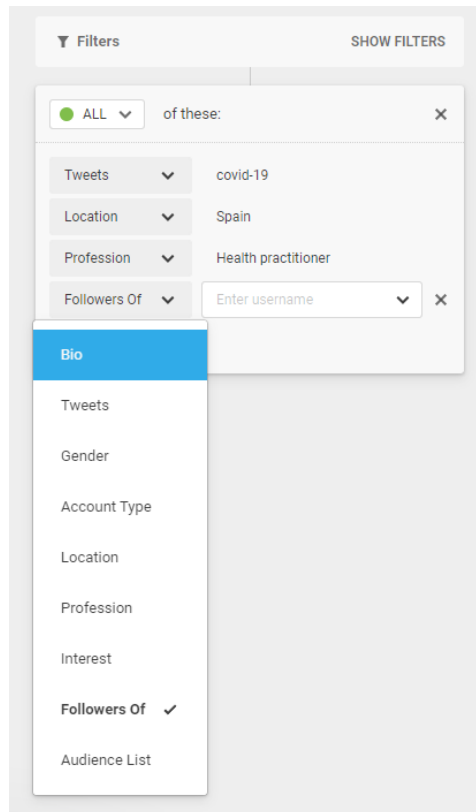


Figura 41 – Ejemplo real de creación de audiencia con diversos filtros disponibles en BrandWatch durante las investigaciones realizadas.

BrandWatch no es una herramienta asequible a “cualquier persona”, simplemente por razones de coste. Está pensada para grandes organizaciones. Por razones de confidencialidad comercial, no es posible revelar su precio y, a diferencia de SentiOne, no existe en el momento de realizar esta tesis ninguna versión de acceso con la tarifa publicada (en todos los casos es preciso contactar con el equipo comercial).

Entre los clientes de Brandwatch, se cuentan corporaciones como Coca Cola, Toyota, Nestlé y Unilever.

3.2.5.4 Audiense

Audiense es una veterana herramienta comercial, anteriormente conocida como SocialBro. Está focalizada exclusivamente en Twitter, y ha pasado de tener un enfoque de *social listening* “puro” a incorporar funcionalidades IA de análisis de sentimiento y de perfilado psicológico, gracias a un *partnership* con los servicios de IBM para procesamiento del lenguaje con tecnologías de *machine learning*.

La funcionalidad de Audiense “Personality Insights” se basa en la psicología del lenguaje en combinación con algoritmos de análisis de datos. El servicio analiza el contenido que publican los usuarios y devuelve un perfil de personalidad, como refleja la imagen siguiente.

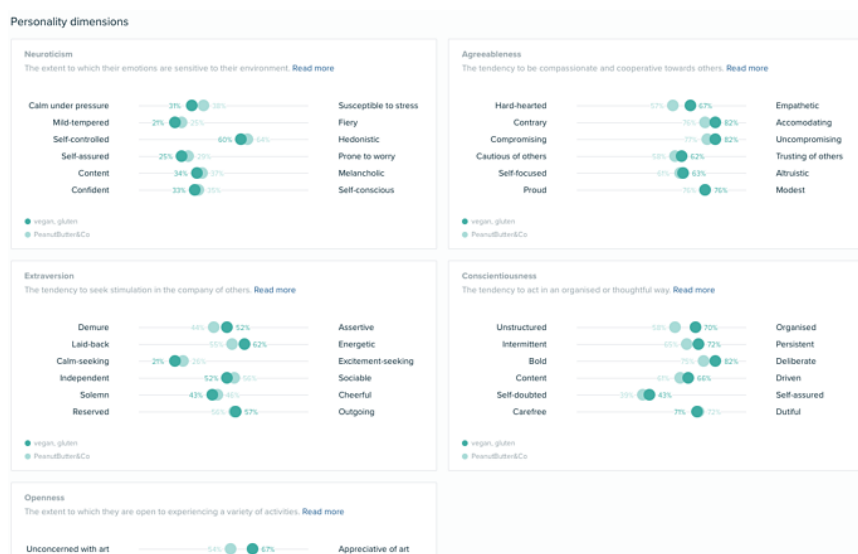


Figura 42 – Personality Dimensions en Audiense

Audiense es un recurso mencionado por los profesionales consultados en la encuesta realizada para esta tesis (Anexo 2), y es también la herramienta elegida para ilustrar empíricamente cómo es posible perfilar audiencias (listas de usuarios

que cumplen determinadas condiciones), perfilarlas e impactarlas con contenidos específicos a fin de influenciarlas (Anexo 5).

Audiense tiene un precio asequible a cualquier persona en sus configuraciones más básicas, y sus precios, a diferencia de otras herramientas analizadas, son públicos⁹³:

- Sin gestión de audiencias⁹⁴: gratis.
- Con gestión de audiencias: 39€ al mes
- Con *insights* de audiencias (perfilado psicológico vía IBM Watson): 696€ al mes.

3.2.5.5 *MeaningCloud*

Aunque no aparece citada en las fuentes de *benchmarking* de referencia ni fue mencionada por ninguno de los participantes en la encuesta a profesionales (Anexo 2), en una revisión bibliográfica realizada *ad hoc* se han encontrado trabajos de investigación apoyados en este recurso comercial, como por ejemplo el estudio de Segura-Bedmar et al. (2015) sobre la salud y el consumo de drogas en las redes sociales, o el de Herrera-Planells & Villena-Román (2018) sobre categorización de titulares de noticias.

La herramienta dispone de una API y realiza análisis automatizado de textos en 44 idiomas. A diferencia de SentiOne y otros recursos, permite realizar análisis de sentimiento a nivel de entidad.

⁹³ Los costes que se indican corresponden a la última consulta realizada, en julio de 2022

⁹⁴ No sería posible desarrollar las estrategias descritas en el Anexo 5 con esta versión.

12510	#yonomevacuno que se vacunen todos los politicos primeros,	#yonomevacuno	Entity	1	Top>Id>Hashtag
12511	#yonomevacuno que se vacunen todos los politicos primeros,	partido político	Concept	1	Top>Organization>Politico
12512	#yonomevacuno que se vacunen todos los politicos primeros,	partido	Concept	2	Top>Event>Occasion>Gan

Figura 43 - Identificación de partes del mensaje con potencial polaridad con MeaningCloud.

MeaningCloud también ofrece como parte del análisis de textos la identificación de subjetividad y de ironía.

GENIAL @FirstDates_TV que no baja del listón del 7% de sh	P+	AGREEMENT	SUBJECTIVE	100	NONIRONIC
alguien tiene el instagram de alejandro? es para una amiga	P	AGREEMENT	OBJECTIVE	97	NONIRONIC
Karim: "Mi mujer ideal sería Alaska, pero con cejas" 😊 #Fir	P+	AGREEMENT	SUBJECTIVE	100	NONIRONIC
Si Bratz tiene todo el rollazo, se dice y no pasa nada 😊👍	N	AGREEMENT	OBJECTIVE	92	NONIRONIC
#Audiencias #FelizMartes #16Noviembre 📺 Carlos Sobera	P	AGREEMENT	OBJECTIVE	100	NONIRONIC
Es mi compi de trabajo, mi compañero y además será el dire	P	AGREEMENT	SUBJECTIVE	100	NONIRONIC
JAJAJAJAJAJAJA SI ES QUE ES LA MEJOR	P	AGREEMENT	SUBJECTIVE	100	NONIRONIC
Madre mia @OrochiGrey vaya santa paciencia has tenido co	P	AGREEMENT	SUBJECTIVE	100	NONIRONIC
Agustín, soy lento para que disfrute=mega construcciones (P	AGREEMENT	SUBJECTIVE	100	NONIRONIC
Grand Place, Patrimonio de la Humanidad de la Unesco y un	P	AGREEMENT	SUBJECTIVE	100	NONIRONIC
Ya sabéis contratarme 😊	P	AGREEMENT	OBJECTIVE	100	NONIRONIC
@firstdates_tv Lydia si te cortaras el medio metro que tien	NEU	DISAGREEMENT	SUBJECTIVE	91	NONIRONIC
Antonio: "Yo no quiero ser el típico diseñador de moda que	NEU	DISAGREEMENT	SUBJECTIVE	86	NONIRONIC
Buen montaje....sería le leche conseguirlo por esta vía!!!!	P	AGREEMENT	SUBJECTIVE	100	NONIRONIC
@cuatro @firstdates_tv @mediasetcom @pazpadilla, pens	P	AGREEMENT	SUBJECTIVE	100	NONIRONIC
Hola @carlos_sobera. Llevamos mucho tiempo intentando	NEU	DISAGREEMENT	OBJECTIVE	94	NONIRONIC
.@carlos_sobera: "Esto es @firstdates_tv y aquí cumplimos	P	AGREEMENT	OBJECTIVE	100	NONIRONIC
Una experiencia de amor Única ❤️👍👍👍 #Biblia #Jesus	P	DISAGREEMENT	SUBJECTIVE	94	NONIRONIC
¡Salva tu alma del Seo! Ven a Jesús👍 #Biblia #Jesus #Rica	P	AGREEMENT	OBJECTIVE	100	NONIRONIC
Que lindo hermosas enseñanza de Dios👍👍👍 #pizgacz	P+	AGREEMENT	SUBJECTIVE	100	NONIRONIC
Zapeando, me he encontrado con la reposición de esta cita	P+	AGREEMENT	SUBJECTIVE	100	NONIRONIC
Hoy hicimos un nuevo First Dates en clase de Innovación. R	N	DISAGREEMENT	SUBJECTIVE	76	IRONIC
Susana: "Quiero bailar hasta que el sol salga. Quiero fiesta,	P	AGREEMENT	OBJECTIVE	100	NONIRONIC

Figura 44 - Muestra del análisis realizado por MeaningCloud, incluyendo detección de subjetividad e ironía.

MeaningCloud se presenta como un servicio para empresas orientado a analítica de textos (*text analytics*), incluyendo sentiment analysis, y analítica de redes sociales (*social media analytics*). Declara contar con grandes coporaciones entre sus clientes, como Pfizer, ING, World Bank Group, PRISA y Unidad Editorial.

En el momento de realizar esta investigación, su plan de analítica de textos (incluyendo *sentiment analysis*), es gratuito.

3.2.6 Evaluación del rendimiento de las herramientas académicas: *benchmark* actual

Podemos decir que, a partir de finales de 2019, y especialmente a raíz de la eclosión de trabajos relacionados en 2020, la presentación del *framework* BERT ha supuesto un punto de inflexión definitivo en las metodologías de procesamiento de lenguaje natural (Devlin et al., 2019). A diferencia de otros modelos de representación del lenguaje recientes (Peters et al., 2017) (Radford et al., 2018), BERT (*Bidirectional Encoder Representations from Transformers*) está diseñado para entrenar previamente representaciones bidireccionales profundas a partir de texto sin etiquetar, mediante el condicionamiento conjunto del contexto en todas las capas. Como resultado, el modelo BERT preentrenado puede ser sometido a un “ajuste fino” con solo una capa de salida adicional para crear modelos de última generación para en amplia gama de tareas, sin necesidad de realizar apenas modificaciones estructurales específicas para la tarea (Figura 45).

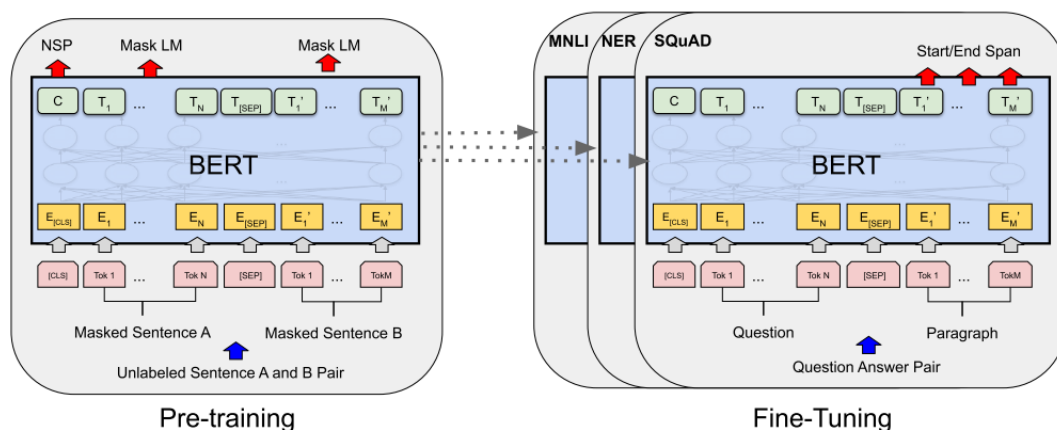


Figura 45 - Procedimientos de preentrenamiento y ajuste fino con BERT (Devlin et al., 2019, p. 3). Se utilizan las mismas arquitecturas tanto en el preentrenamiento como en el ajuste. De este modo, los mismos parámetros del modelo preentrenado se utilizan para inicializar los modelos de las diferentes tareas descendentes. Durante el ajuste fino, se ajustan todos los parámetros.

Según las conclusiones de Delvin et al. (2019) BERT eleva los *benchmarks* de última generación en nada menos que 11 tareas de procesamiento de lenguaje natural, incluido el aumento de la puntuación GLUE a un 80,5 % (7,7 % de mejora), la precisión MultiNLI a un 86,7 % (4,6 % de mejora), y el test SQuAD v1.1 a un sorprendente F1 de 93.2 (mejora absoluta de 1.5 puntos).

El *paper* original de Delvin y sus compañeros es el documento más citado en el dominio en los últimos años⁹⁵ y ejerce una poderosa influencia de la comunidad investigadora desde 2020.

El éxito de este nuevo enfoque metodológico, más allá de las conclusiones de sus autores, has sido corroborado por multitud de estudios, aunque todavía no están claras las razones profundas del mejor rendimiento de esta técnica (Kovaleva et al., 2019).

Dataset	Pre-trained	Fine-tuned, initialized with normal distr.		Metric	Size
MRPC	0/31.6	81.2/68.3	87.9/82.3	F1/Acc	5.8K
STS-B	33.1	2.9	82.7	Acc	8.6K
SST-2	49.1	80.5	92	Acc	70K
QQP	0/60.9	0/63.2	65.2/78.6	F1/Acc	400K
RTE	52.7	52.7	64.6	Acc	2.7K
QNLI	52.8	49.5	84.4	Acc	130K
MNLI-m	31.7	61.0	78.6	Acc	440K

Tabla 9 – Evaluación de rendimiento GLUE de modelos BERT con diferente inicialización, (Kovaleva et al., 2019 p. 4369).

⁹⁵ 11425 veces citado según SCOPUS, última consulta julio de 2022.

3.2.7 Evaluación del rendimiento de las herramientas comerciales usadas por la comunidad investigadora

No se han encontrado trabajos de investigación académica sobre el rendimiento de las herramientas “informales”. La encuesta realizada entre profesionales descrita con anterioridad (Anexo 2), refleja un cierto descontento entre los profesionales (investigadores o no) que han utilizado alguna vez herramientas comerciales de monitorización de la actitud de los usuarios en Internet: un 56,9% de ellos tiene un nivel de satisfacción “intermedio” o “bajo”. Este dato, aun siendo revelador, es insuficiente para responder a la pregunta que nos formulamos. A fin de obtener un dato más preciso desde una perspectiva científica, se ha realizado dos trabajos experimentales (Anexos 1 y 3) en los que se incluye SentiOne, la herramienta más citada por los investigadores (en el ámbito del análisis de sentimiento),

En el primer estudio, se realizaron análisis separados, con dos *datasets* distintos de literales de Twitter, aplicando en ambos casos el mismo método, en varias fases sucesivas: creación del proyecto en SentiOne, extracción y análisis automático de sentimiento de los textos, clasificación manual de los mismos sobre criterios heurísticos específicos, y finalmente cálculo del F-Score y *Accuracy*.

Dataset “entretenimiento”:

- Palabra clave: “first dates”⁹⁶
- Fuente: Twitter
- Período de extracción: septiembre-noviembre de 2021 (tres meses)

⁹⁶ First Dates es un popular programa de entretenimiento emitido en Prime Time en España desde 2016, caracterizado por generar mucha actividad en Twitter, con polémica en torno a las actitudes y comportamientos de los concursantes.

Dataset “servicio público”:

- Palabra clave: “adif”⁹⁷
- Fuente: Twitter
- Período de extracción: septiembre-noviembre de 2021 (tres meses)

En la tabla siguiente podemos ver los resultados F-Score obtenidos:

Dataset	F-Score			
	Positive	Negative	Neutral	Avg.
Infraestructura pública	60,87	54,97	69,67	61,84
Entretenimiento	66,01	61,60	68,22	65,28

Tabla 10 – F-Score comparativo de los datasets analizados (Anexo 1)

La siguiente tabla muestra los resultados para “accuracy” (precisión) de ambos *datasets*:

Dataset	Accuracy
Infraestructura pública	63,62
Entretenimiento	65,66

Tabla 11 – Precisión comparativa de los datasets analizados

⁹⁷ Adif es el acrónimo de “Administrador de Infraestructuras Ferroviarias”. Se trata de una empresa pública española, dependiente del Ministerio de Transportes, encargada del mantenimiento y explotación de las infraestructuras ferroviarias (más de 15.000 kilómetros de vías cerca de 2000 estaciones).

¿Cómo se comparan los resultados obtenidos por SentiOne con respecto a los *benchmarks* del estado del arte? Como podemos observar, los valores de SentiOne están a la altura de los métodos “formales” analizados por Ribeiro et al. (2016). Si nos fijamos exclusivamente en *datasets* de Twitter, tenemos un promedio F-Score de 60,43, inferior al promedio de 63,56 obtenido por SentiOne.

En cuanto a la métrica *accuracy*, los resultados obtenidos con SentiOne son claramente inferiores a los mejores resultados obtenidos por Zimbra et al. (2018) (más nivel de detalle en el Anexo 1).

La comparación entre ambos resultados, automático y humano, puede ser un buen indicador (el único del que tenemos noticia hasta la fecha de realización del estudio) para evaluar la precisión de esta herramienta que, como hemos visto, es la más citada por la comunidad de investigadores en trabajos sobre *sentiment analysis*.

3.2.7.1 Segundo estudio ampliado

En el segundo estudio, se aplicó un enfoque muy ampliado, a partir de las conclusiones del primero (más nivel de detalle en el Anexo 3). En concreto, se realizó una nueva evaluación comparando directamente entre sí varias herramientas comerciales/informales, con distintos grados de “representatividad” desde el punto de vista de su empleo por parte de la comunidad científica: Sentione, Repustate y MeaningCloud.

Esencialmente, el propósito de esta prueba es conocer el nivel de *Accuracy*, el F-Score y el *Recall* de las aplicaciones (RQ1), así como determinar si éstos pueden variar en función de la temática de los textos (RQ2), para lo cual se han empleado tres *datasets* distintos. Respondiendo a las preguntas de investigación planteadas, este estudio tiene por finalidad llenar un espacio vacío en el estado del arte, reforzando el componente de novedad de la tesis y, sobre todo, despejando una incógnita esencial para una completa validación de la hipótesis de la tesis doctoral.

Como ya vimos, Repustate es una herramienta comercial que aparece en evaluaciones de otros trabajos de *benchmarking*, como las de Abbasi et al. (2014) y Zimbra et al. (2018) (ver también referencias en Anexos 1 y 3). En cuanto a MeaningCloud, aunque no aparece citada en las fuentes de *benchmarking* de referencia ni fue mencionada por ninguno de los participantes en la encuesta a profesionales (Anexo 2), en una revisión bibliográfica realizada *ad hoc* se han encontrado trabajos de investigación apoyados en este recurso, como por ejemplo el estudio de Segura-Bedmar et al. (2015) sobre la salud y el consumo de drogas en las redes sociales, o el de Herrera-Planells & Villena-Román (2018) sobre categorización de titulares de noticias.

Para el desarrollo de la investigación se realizaron **tres estudios separados**, con **tres datasets distintos**, dos de ellos de Twitter (los mismos empleados en el estudio precedente – Anexo 1), y un nuevo *dataset*, de Facebook. En todos los casos se ha aplicado el mismo método, en varias fases sucesivas: creación del proyecto en la herramienta, extracción y análisis automático de sentimiento de los textos, clasificación manual de los mismos sobre criterios heurísticos específicos, y finalmente cálculo del *F-Score*, *Accuracy* y *Recall*.

El *dataset* adicional para este nuevo estudio se creó también *ad hoc*, siguiendo un procedimiento naturalista, y se inscribe en un área totalmente distinta al resto, que además entronca con uno de los ejes de investigación de la presente tesis: el movimiento “antivacunas” del Covid-19. Estos son sus detalles:

Dataset “anti vacunas”:

- Palabra clave: “yonomevacuno”⁹⁸

⁹⁸ Adif es el acrónimo de “Administrador de Infraestructuras Ferroviarias”. Se trata de una empresa pública española, dependiente del Ministerio de Transportes, encargada del mantenimiento y

- Fuente: Facebook
- Período de extracción: marzo 2019 a junio 2022 (es decir, desde el inicio de la pandemia hasta el momento actual)

Para la evaluación de resultados, además de las estrategias analíticas del estudio anterior, y por consistencia con otros estudios de *benchmarking*, al *F-Score* y *Accuracy* se ha añadido el *Recall*.

La tabla siguiente muestra las características de los *datasets* naturalistas extraídos. Como era previsible, la temática de entretenimiento presenta una proporción muy superior de mensajes positivos con respecto a la temática de infraestructuras públicas, y la de “antivacunas” es abrumadoramente negativa.

Dataset	Size	Class distribution		
		Positive	Negative	Neutral
Infraestructura pública	514	7%	51%	42%
Entretenimiento TV	795	17%	41%	42%
Antivacunas	1980	4%	81%	15%

Tabla 12 - Datasets - Tamaño y distribución por polaridad

Los resultados del estudio demuestran, en concordancia con la investigación precedente (Anexo 1) que SentiOne, pese a ser una herramienta informal no pensada para uso académico por investigadores, presenta un *F-Score* y *Accuracy* alineados con otros recursos más sofisticados y “formales”, aunque sin alcanzar los mejores resultados de los *benchmarks* del estado del arte actual.

explotación de las infraestructuras ferroviarias (más de 15.000 kilómetros de vías cerca de 2000 estaciones).

El F-Score, *Accuracy* y *Recall* medios, sumalizando toda la información obtenida en el estudio, quedarían así:

	Medias		
	F-Score	Accuracy	Recall
SentiOne	61,45	63,45	68,86
Repustate	46,92	53,59	58,35
MeaningCloud	41,38	46,36	54,11

Tabla 13 – Comparativa medias resultados por herramienta

Como se puede comprobar, el resto de las herramientas analizadas en este nuevo estudio arrojan resultados peores, y con una horquilla muy amplia (por ejemplo, más de 5 puntos de diferencia en F-Score entre Repustate y MeaningCloud). Observamos, por tanto, que los resultados que se pueden conseguir con este tipo de recursos comerciales pueden variar mucho en función de diversos factores, como tipo de herramienta empleada y, también, la propia naturaleza del *dataset*. En concreto, hemos podido comprobar cómo los resultados mejoran cuanto más acusada es la polaridad negativa, lo cual es consistente con un escenario donde los sentimientos son más intensos (y, por tanto, más susceptibles de ser identificados y procesados correctamente).

3.2.7.2 *Benchmarking de herramientas comerciales “formales”*

Como decíamos, no se han encontrado estudios de *benchmarking* referidos a herramientas informales (dificultad superada mediante las investigaciones propias de nueva planta, referidas con anterioridad), pero sí existen varios trabajos sobre

herramientas “formales”, que hemos utilizado como fuente en distintos momentos de la investigación (Ver anexos 1 y 3 para mayor detalle).

- **Herramientas *stand-alone***⁹⁹ : Abbasi et al. (2014) realizaron en fecha muy temprana un trabajo comparativo del rendimiento de herramientas de *sentiment analysis* con Twitter (54 citas¹⁰⁰). En fecha mucho más reciente, Dolianiti et al. (2019) presentan un nuevo trabajo de *benchmarking* con herramientas actualmente en boga, de las cuales solo dos coinciden con las de Abbasi et al.
- **Trabajo mixto con herramientas comerciales y métodos académicos:** Otro estudio muy citado (220 citas¹⁰¹) es el trabajo de Ribeiro et al. (2016) sobre 24 métodos con diversos *datasets* multi disciplinares. A fin de construir un ranking suficientemente representativo, Ribeiro et al. realizan pruebas separadas para 2 (Pos., Neg.) y 3 clases (Pos, Neg., Neu.), con varios *datasets* de temáticas totalmente diferentes, lo que confiere una especial solidez a sus conclusiones.

⁹⁹ Stand-Alone puede traducirse de forma difusa como “independiente”. En este caso, se denomina herramientas stand-alone a aquellos servicios entregados en forma de “paquete”, por contraposición al software y métodos que el investigador puede alterar e integrar en un desarrollo propio.

¹⁰⁰ 54 Citas según SCOPUS, dato actualizado a julio de 2022.

¹⁰¹ 200 citas según SCOPUS, dato actualizado a julio de 2022.

3.3 Recursos para el desarrollo de metodologías y herramientas *ad hoc*

Además de las herramientas y métodos que hemos analizado (anexos y apartados anteriores), los investigadores tienen a su disposición recursos de la comunidad, validados por la literatura, que pueden servir como elementos constituyentes de nuevas herramientas y métodos. Según un trabajo exhaustivo realizado por Sun et al. (2017), muy referenciado en la literatura (288 citas¹⁰²), los *toolkits* más utilizados en la actualidad son los siguientes:

NLTK	Python	http://www.nltk.org/
Open NLP	JAVA	https://opennlp.apache.org/
CoreNLP	JAVA	https://stanfordnlp.github.io/CoreNLP/
Gensim	Python	https://radimrehurek.com/gensim/
FudanNLP	JAVA	https://code.google.com/archive/p/fudannlp/
LTP	C++	http://www.ltp-cloud.com/intro/en/
NiuParser	C++	http://www.niuparser.com/index.en.html

*Las tres últimas herramientas son exclusivamente para lenguaje chino.

¹⁰² Dato según SCOPUS. Consulta actualizada a julio de 2022.

3.4 Problemas y desafíos comunes de la comunidad investigadora

En este apartado, desarrollaremos el contenido de la investigación realizada en torno a los límites actuales del arte, así como a los desafíos y asuntos sometidos a debate en la comunidad.

3.4.1 El debate en torno a la aplicación de metodologías de *Machine Learning*

La eficacia de las metodologías basadas en *machine learning* depende de la selección de características (*feature selection*), pero los enfoques típicos no son capaces de afrontar ciertos fenómenos, como por ejemplo la detección de negaciones (Giachanou & Crestani, 2016). La respuesta a este problema podría estar en el uso de algoritmos de *deep learning*, capaces de aprender representaciones de los datos, pero este tipo de métodos, pese a que han experimentado un gran desarrollo en los últimos cinco años, especialmente a partir de los trabajos de varios autores como Tal et al. (2014), aún no están suficientemente explorados ni existe un consenso pleno sobre su utilización. A este respecto, Giachanou & Crestani (2016) proponen que una dirección interesante sería examinar la eficacia de los algoritmos de aprendizaje profundo de las redes neuronales en el análisis de sentimiento y el manejo de la negación en los tuits, ya que son eficaces en análisis del sentimiento del texto.

Según Pereira-Kohatsu et al. (2019), las "formas superficiales" (enfoques basados en el léxico y el aprendizaje automático) han funcionado notablemente bien para muchos problemas de Procesamiento del Lenguaje Natural, pero no son capaces de explicar la semántica de las palabras. Los métodos basados en la incrustación de palabras mediante el entrenamiento de una red neuronal permiten acercarse a este objetivo.

Aunque las aproximaciones metodológicas basadas en *machine learning* son apreciadas debido a su precisión, a menudo presentan desventajas: debido a que los modelos están entrenados en dominios (*data*) muy específicos, no suelen ser transferibles a nuevas fuentes de texto (Taboada, 2016). Esto implica que la aplicación de un modelo a nuevos conjuntos de datos generalmente requiere nuevos entrenamientos y un trabajo de codificación humana que puede ser costoso. Por ejemplo, un modelo entrenado para reseñas de una categoría de productos quizás no sería extrapolable a otra categoría; ese mismo modelo, llevado a un dominio totalmente diferente (por ejemplo, críticas de películas), en buena lógica presentaría un desafío de codificación mayor.

Por último, hay otro espacio de debate esencial que es un “viejo conocido” de los expertos en ciencia de datos e interpretación del lenguaje natural: la interpretación del contexto. Criado y Villodre (2018), concluyen que los sistemas actuales basados en el aprendizaje automático aún no son capaces de captar todas las vicisitudes de los contextos.

3.4.2 Polaridad y sus frentes: la gran asignatura pendiente

Una correcta identificación de las palabras y giros lingüísticos que implican negación es absolutamente esencial para el procesamiento del lenguaje natural, por razones obvias, y más importante aun cuando se trata de evaluar el sentimiento. Este aspecto sigue constituyendo hoy en día un reto importante para la comunidad científica, sobre todo si tenemos en cuenta que la utilidad de las distintas soluciones propuestas puede estar muy condicionada por desafíos de todo tipo, la mayoría no resueltos. En los siguientes subapartados desarrollaremos cada uno de estos desafíos, incluyendo resultados de investigaciones propias (necesarias para cubrir aspectos del arte no estudiados en la literatura disponible).

3.4.2.1 *Desafío 1: la identificación de polaridad a nivel de entidad*

Pla y Hurtado (2014) ya identifican en fecha temprana cómo la obtención de la polaridad a nivel de entidad introduce una complejidad adicional en el análisis, ya que esto obliga a determinar primero la parte del texto que se refiere a cada una de las entidades. Al respecto, hemos podido comprobar que este no es un problema resuelto en procesamiento del lenguaje natural, de momento. De hecho, muchas de las herramientas disponibles en *sentiment analysis* no son capaces de trabajar a nivel de entidad, al igual que la mayoría de los métodos menos recientes presentados por los investigadores.

Como hemos visto, la mayoría de los trabajos se centran en plataformas de *micro blogging*, especialmente Twitter, donde estaríamos hablando (generalmente) de un análisis a nivel de frase (*sentence*). Sin embargo, sabemos que una frase puede tener más de una entidad, con polaridad distinta. Por tanto, si no identificamos entidades en los tuits y los evaluamos siempre a nivel de frase, estaremos aceptando una cierta pérdida de información, imprecisión y ambigüedad. Para medir este aspecto y no dejarlo en el (poco científico) terreno del “sentido común”, realizamos un pequeño estudio (Anexo 7) a fin de determinar el porcentaje de tuits en el que podemos apreciar más de una entidad en el mensaje. En concreto, se analizaron 1980 mensajes de Facebook correspondientes a un *dataset* de elaboración propia (Anexo 3) referido al movimiento antivacunas del Covid-19. El resultado, no deja lugar a duda, como refleja la siguiente tabla:

	#	%
Tamaño del <i>dataset</i>	1980	100%
Mensajes con 1 entidad	486	25%
Mensajes con 2 entidades	222	11%
Mensajes con 3 entidades	149	8%
Mensajes con 4 entidades	82	4%
Mensajes con 5 o más entidades	147	7%
TOTAL con más de una entidad	1086	55%

Tabla 14 - Resultados de la investigación sobre número de entidades en mensajes de Facebook (Anexo 7).

Como se puede observar, el 55% de los mensajes de Facebook analizados incluía más de una entidad, e incluso hay un 7% con la sorprendente cifra de 5 o más (esto es posible en Facebook dado que permite publicar textos extensos).

El mismo estudio, aplicado sobre literales de Twitter, arroja un resultado también inequívoco, pese a la limitación de caracteres de esta red social:

	#	%
Tamaño del <i>dataset</i>	795	100%
Mensajes con 1 entidad	223	28%
Mensajes con 2 entidades	101	13%
Mensajes con 3 entidades	51	6%
Mensajes con 4 entidades	8	1%
Mensajes con 5 o más entidades	10	1%
TOTAL con más de una entidad	393	49%

Tabla 15- Resultados de la investigación sobre número de entidades en mensajes de Twitter (Anexo 7)

A la vista de los resultados obtenidos mediante el reconocimiento de entidades en los *datasets* de referencia, se comprueba que el potencial de multi entidad en las redes de *micro blogging* puede llegar a afectar a más del 50% de los mensajes.

Obviamente, este problema no ha pasado desapercibido a la comunidad investigadora, y consecuentemente ha centrado la atención de muchos especialistas (Pappu et al., 2017). Algunos autores han intentado desarrollar soluciones específicas, en algunos casos muy complejas, con diferentes grados de éxito, como veremos más adelante.

Como pudimos comprobar en uno de los estudios realizados para la presente investigación (Anexo 1), las aplicaciones SAAS informales que en ocasiones utilizan los investigadores (y de manera regular los profesionales del marketing y la investigación de mercados, como comprobamos en la encuesta - Anexo 2), no están libres de este problema tampoco.

3.4.2.2 Desafío 2: la negación y su impacto en la polaridad

En este sentido cabe destacar el enfoque avanzado que ofrecen, con carácter pionero, los trabajos de Kiritchenko et al. (2014). En sus trabajos sobre Twitter, estos autores propusieron como solución el desarrollo de dos léxicos separados, identificando por un lado los términos que aparecen más a menudo en contextos de negación, y por otro los que aparecen vinculados a contextos sin negaciones. No obstante, más de cinco años después, el debate sigue abierto.

3.4.2.3 Desafío 3: la relatividad de la polaridad por el empleo de la ironía y el sarcasmo

La ironía¹⁰³ es un fenómeno de la comunicación humana, frecuentemente utilizado como figura literaria, que consiste en decir lo contrario (o algo muy diferente) a lo que realmente se desea transmitir, contando con la capacidad cognitiva del receptor del mensaje para entender la intención de emisor. Generalmente, los humanos pueden detectar fácilmente la ironía, pero a menudo se producen malentendidos. En

¹⁰³ Ironía, del griego εἰρωνεία 'eirōneía', que significa literalmente disimulo o ignorancia fingida.

el caso de las máquinas, la detección automática de la ironía es muy difícil y plantea numerosos desafíos, pero resulta imprescindible si tenemos en cuenta que puede invertir la polaridad del sentimiento de un mensaje (Pang & Lee, 2008).

El sarcasmo es un tipo de ironía construida con una intención hiriente o burlesca ¹⁰⁴, y también ha concentrado, de forma específica, la atención de algunos investigadores.

Al respecto de la importancia de la detección de la ironía en sus distintas formas, hay que destacar especialmente los trabajos de Maynard & Greenwood (2014), que demostraron que la identificación de tuits sarcásticos podía mejorar la detección del sentimiento en casi 50 puntos porcentuales. En cualquier caso, pese a esta mejora, los autores también concluyeron que la precisión del análisis de sentimientos seguía estando lejos de ser perfecta aun con una correcta detección del sarcasmo.

Hasta ahora, los mejores resultados en metodologías específicas de detección de ironía los han conseguido Liebrecht et al. (2013). Utilizando el hashtag #sarcasme, crearon un *dataset* de 78.000 tuits en lengua holandesa, y mediante un algoritmo equilibrado de Winnow los clasificaron como sarcásticos o no con un 75% de precisión.

3.4.3 Investigación multidisciplinar poco madura

Es obvio que el estudio de la actitud de los ciudadanos en Internet a través de metodologías de análisis de sentimiento y minería de opinión solo cobra sentido cuando intentamos evaluar textos sobre temas concretos. A menudo, esto implica la involucración de expertos de diferentes campos que deben colaborar y sintonizar

¹⁰⁴ Sarcasmo, del latín 'sarcasmus' y este a su vez del término griego 'σαρκασμός' ('sarkasmós'). El significado literal es 'mordedura de labios'.

protocolos, métodos de trabajo y culturas científicas diferentes, como la ciencia de datos, la lingüística, la computación, el marketing, la sociología, la comunicación política, la educación, etc.

Existen veteranos ejemplos pioneros de investigación multidisciplinar (entendida como la combinación de investigaciones de diferentes campos), como los trabajos de Mitchell et al. (2013) sobre el impacto de variables de carácter meteorológico y geográficas en la felicidad de los ciudadanos o, más sorprendente, el estudio de Asur y Huberman (2010) que utiliza técnicas de análisis de sentimiento con fines predictivos al servicio del negocio cinematográfico. Sin embargo, todo parece indicar que este es otro terreno donde queda camino por recorrer.

3.4.4 Desafíos y debates locales: recursos y herramientas específicos de la comunidad investigadora de habla hispana

En relación con el desarrollo de recursos tecnológicos para la comunidad investigadora, cabe destacar las aportaciones de dos iniciativas desarrolladas en España: Freeling y LiguaKit. Es muy significativa la ausencia de debate en la comunidad sobre la idoneidad de las diferentes herramientas. En este sentido, una de las pocas excepciones a esta regla es el argumento expuesto por Franco-Riquelme et al. (2019) para justificar su preferencia por LiguaKit frente a Freeling.

En cualquier caso, debemos considerar que tanto Freeling como LiguaKit, recursos que se encuentran entre los más utilizados por la comunidad investigadora, son una "visita obligada" para los nuevos investigadores. Otros recursos que parecen despertar el interés de los investigadores son SentiStrenght, TweetMotif y la colección de herramientas contenidas en SciKit.

En cuanto al debate metodológico desde una perspectiva más estratégica, resulta especialmente interesante el trabajo de Pereira-Kohatsu et al. (2019) relacionado con HaterNet, por su profundidad y novedad respecto a otros trabajos. HaterNet, el

sistema que actualmente utiliza la Oficina Nacional contra los Delitos de Odio de la Secretaría de Estado de Seguridad de España y que identifica y monitoriza la evolución del discurso de odio en Twitter, es una herramienta novedosa que sin duda será una fuente importante de futuros trabajos.

3.4.5 Problemas específicos de Twitter

Hay obstáculos específicos que suponen un reto importante a la hora de trabajar con los datos cuando se trata de procesar lenguaje natural en Twitter, como los siguientes:

- Mensajes publicados por *bots* y otras formas de spam.
- Faltas de ortografía y erratas que pueden impedir la correcta identificación de una palabra (Twitter no permite editar los mensajes una vez publicados).
- Mayúsculas enfáticas.
- Alargamiento enfático.
- Abreviaturas.
- Jergas y neologismos.
- Dobles sentidos y saturación recurrente de figuras literarias.
- “Chistes privados” de ciertos colectivos y coletillas-meme.

La comunidad científica no ha dejado de producir estudios que intentan abordar estos problemas con diferentes estrategias, pero queda mucho camino por recorrer y, sobre todo, por estandarizar. Brody y Diakopoulos (2011), por ejemplo, realizaron una interesante investigación sobre el alargamiento enfático y su impacto en el procesamiento de los tuits. De acuerdo con estos autores, el fenómeno del alargamiento enfático se produce en uno de cada seis tuits.

En cualquier caso, no está claro si la comunidad investigadora puede confiar en soluciones genéricas, o si es necesario desarrollar herramientas/estrategias específicas para el campo de debate. Al menos, está claro que muchas de las

soluciones deben ser necesariamente específicas para cada comunidad cultural o lengua.

3.4.6 Conjuntos de datos de referencia

Uno de los desafíos (y por tanto espacios de debate) más relevantes en el procesamiento del lenguaje natural en Twitter es el escaso número de conjuntos de referencia, o *benchmarks*. Generalmente, los investigadores recogen sus propios datos y evalúan sus métodos con ellos, por lo que la comparación entre métodos es particularmente complicada.

3.4.7 Escasez de estrategias para el tratamiento de contenido multilingüe

No es infrecuente en Twitter la mezcla de idiomas en un mismo mensaje o la presencia de mensajes en otros idiomas (generalmente, el omnipresente idioma inglés) en el contexto de debates locales. No obstante, pocos investigadores ha afrontado este desafío. En este sentido, cabe destacar el trabajo pionero de Narr et al. (2012). Los autores analizaron la viabilidad de un enfoque de clasificación semi-supervisado de sentimientos, independiente del idioma. Para ello, utilizaron los emoticonos como etiquetas para generar datos de entrenamiento. Entrenaron un clasificador *Naïve Bayes* en estos datos y lo evaluaron en más de 10.000 tuits en 4 idiomas, que fueron anotados por operadores humanos. Los experimentos realizados demostraron, según sus autores, que su enfoque de clasificación puede aplicarse eficazmente para varios idiomas sin requerir un esfuerzo adicional por cada idioma adicional.

Sin duda, las estrategias para contenido multilingüe son un espacio de investigación extremadamente complejo e incómodo, pero esencial en muchas áreas geográficas donde, por razones culturales, la mezcla de idiomas es frecuente.

3.4.8 Palabras de parada o *stop words*

Las palabras de parada o *stop words* son eliminadas en el preprocesamiento porque tienen un bajo poder de discriminación y no aportan nada en términos de sentimiento. Según Giachanou & Crestani (2016), las listas típicas de palabras de parada precompiladas no son adecuadas para Twitter y pueden incluso influir en el rendimiento del análisis. Como ejemplo, estos autores comentan que la palabra "like" ("cómo", en inglés) suele considerarse una palabra de parada, pero sin embargo tiene un importante poder de discriminación de sentimientos cuando se trata de Twitter. Como respuesta a este problema, Saif et al. (2014), realizaron un trabajo de investigación especialmente interesante y muy citado: utilizaron tuits de seis conjuntos de datos, y aplicaron seis métodos distintos de identificación de palabras vacías, para estudiar después cómo afectaron al rendimiento del análisis de sentimiento. Desde entonces, se han probado distintas aproximaciones, pero parece que sigue sin haber unanimidad de criterios.

3.4.9 Brevidad del texto: una anomalía con consecuencias metodológicas

Una de las principales peculiaridades del análisis de sentimiento, y en general del procesamiento de lenguaje natural en Twitter, es la extrema brevedad de los mensajes. Esto, obviamente, tiene consecuencias metodológicas, sobre las que existe un debate abierto desde los orígenes de este tipo de investigación.

Hace ya más de 10 años que Bermingham y Smeaton (2010) sentaron las bases de la discusión gracias a su estudio sobre el impacto en los procesos de clasificación provocado por la longitud de los tuits (entonces eran aún más cortos que ahora, el límite estaba en 140 caracteres). En su estudio, Bermingham y Smeaton (2010) utilizaron distintos tipos de clasificadores para procesar textos largos y micro textos, llegando a la conclusión de que clasificar tuits es una tarea más fácil (mucho más fácil, de hecho).

3.4.10 Extracción de textos por relevancia temática y fundamentación del mapa mental de los usuarios

El contenido de Twitter es extremadamente heterogéneo, pero cada investigación se centra en un área de interés concreta, que a menudo suele ser muy específica (por ejemplo, un aspecto concreto de la economía, un evento coyuntural, un candidato político, un tema de debate social, etc.). En el proceso de extracción es necesario, por tanto, aplicar un criterio para determinar la relevancia temática de los mensajes, a fin de saber cuáles de ellos deben ser procesados y cuáles descartados. Para establecer la relevancia temática de un tuit, la mayoría de los investigadores se centran en considerar la presencia de una palabra clave como indicador, mientras que otros consideran los hashtags como un fuerte indicador de la relevancia del tuit hacia un tema específico (Giachanou & Crestani, 2016).

Además de esto, el investigador tiene que tomar a menudo decisiones que requieren un conocimiento del marco cognitivo-cultural de los fenómenos a observar, por ejemplo, a la hora de crear *custom aspect models* en una herramienta, para un dominio determinado.

¿De dónde salen estas palabras que determinan la relevancia temática? ¿Cómo se eligen? Sorprendentemente, en la mayoría de los casos los investigadores no justifican con una investigación previa por qué eligen unas u otras palabras como determinantes de la relevancia temática, lo que en cierto modo podría debilitar la fundamentación científica de su trabajo, en un sentido estricto. En concreto, en ninguno de los trabajos analizados, que abarcan lo más representativo del estado del arte, se fundamenta objetivamente el marco lingüístico desde el punto de vista del modelo mental de los usuarios: cómo se categorizan y relacionan entre sí los conceptos en un contexto cultural determinado (por ejemplo, las entidades y los aspectos de esas entidades, o las “palabras clave” que identifican un dominio).

Como parte de las investigaciones realizadas para esta tesis, se probó experimentalmente un nuevo método de identificación del "mapa mental" de los usuarios para un área temática determinada (Anexo 4). El objetivo, era determinar si es posible aplicar una metodología de base científica para saber cómo están organizados los distintos aspectos de una categoría en la mente de los usuarios, de modo que el investigador tenga un material previo para la toma de decisiones fundamentadas, por ejemplo, a la hora de crear *custom aspect models*. El trabajo realizado muestra cómo es posible conocer el mapa mental del usuario mediante una metodología basada en *card sorting* y *hierachical clustering* (Figura 46).

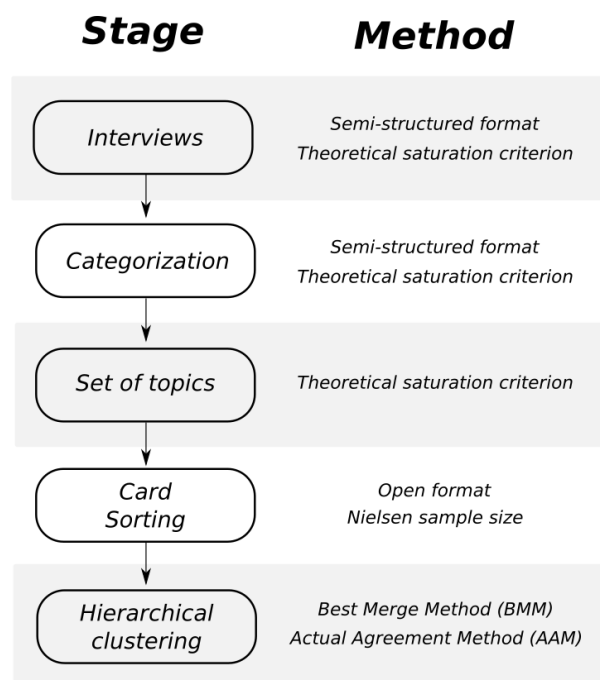


Figura 46 – Esquema de la metodología propuesta (investigación propia, Anexo 4)

Los métodos de *hierachical clustering*, ya sean *Best Merge* o *Actual Agreement* ¹⁰⁵, permiten tomar las agrupaciones de conceptos por categorías realizadas por los sujetos del test y encontrar la mejor solución para representar cómo se organizan en la mente de los usuarios los distintos elementos (además de saber cómo denominan a esas agrupaciones, en la modalidad de *card sorting* abierto). Nielsen (2004) demuestra de forma concluyente que 15 usuarios son suficientes para alcanzar una correlación de 0,90. Después de 15 usuarios, el rendimiento decreciente hace que las correlaciones aumenten de manera ineficiente desde el punto de vista de la conveniencia del estudio.

¹⁰⁵ Ver anexo 4 para más detalle sobre estos conceptos.

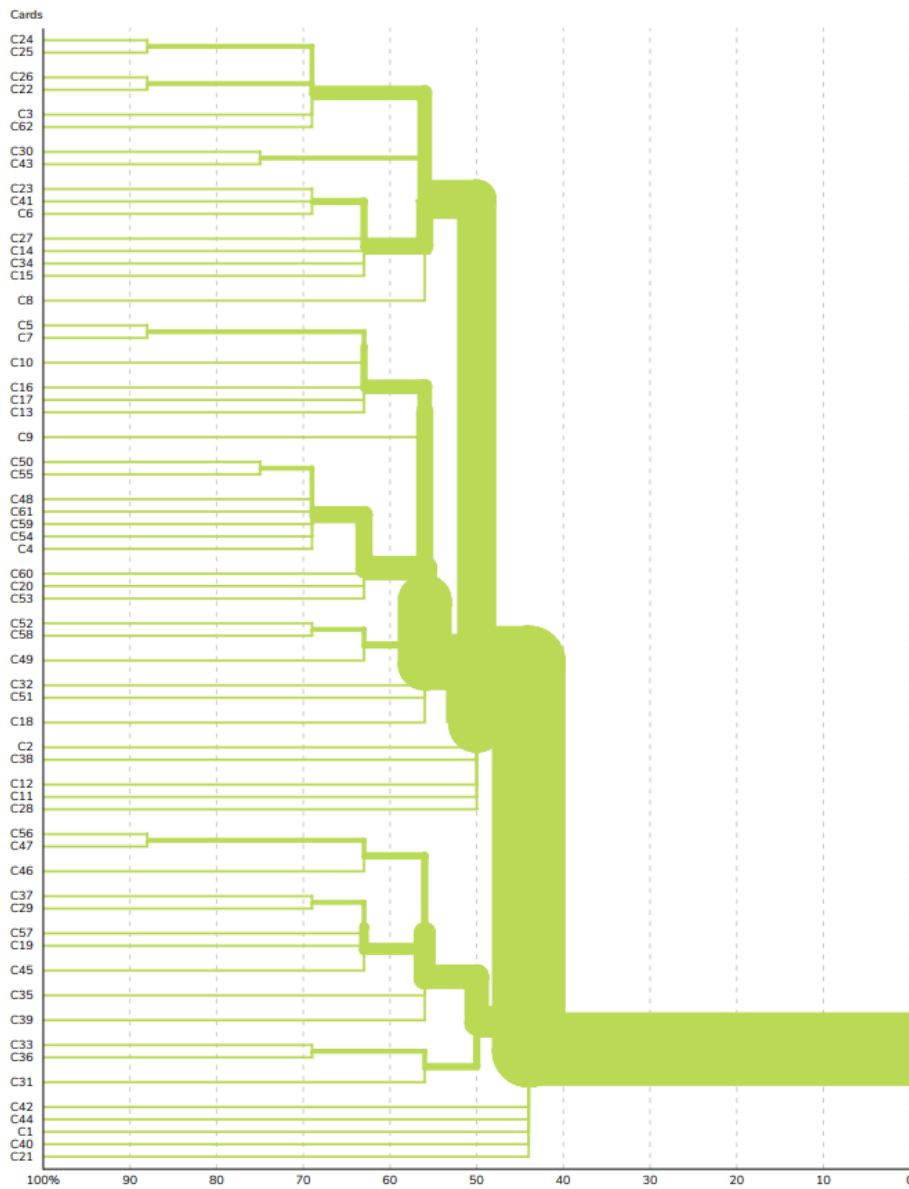


Figura 47 – Ejemplo de dendrograma con algoritmo BMM (investigación propia, Anexo 4).

Mientras en otros desafíos, fundamentalmente ligados a la tecnología, la comunidad científica está inserta en un profundo debate, la relevancia temática es un aspecto esencial donde hasta ahora no parece haberse puesto el foco suficiente.

3.4.11 Tokenización

Otro reto de gran importancia en la investigación basada en *sentiment analysis* es la tokenización de las frases. Se trata, como ya vimos, del primer paso a realizar en el preprocesamiento de los textos, sean del tipo que sean. En el caso de Twitter, Owoputi et al. (2013) abrieron el debate al respecto de si era necesario un procedimiento de tokenización específico para Twitter, proponiendo un nuevo método que, pese a los buenos resultados reportados por los autores, no parece haber sido secundado de manera generalizada por la comunidad investigadora.

3.4.12 Geolocalización

La geolocalización de los textos publicados por los usuarios puede ser esencial para entender la distribución regional de la opinión e incluso para la realización de estudios relacionados con intención de voto. Sin embargo, las actuales limitaciones normativas (políticas de privacidad) y tecnológicas (características técnicas de las plataformas) dibujan un panorama donde no parece que sea posible encontrar una solución plenamente satisfactoria.

Sabemos que la mayoría de los *tweets* no indicaban su ubicación. Este es un problema bien conocido, ya que menos del 1% de los tweets están geolocalizados (geoetiquetados) y la información disponible en el campo "ubicación" en los perfiles de los usuarios no es confiable (Mahmud et al., 2012). Esta circunstancia obliga a muchos investigadores, como por ejemplo Castro et al. (2017) en sus trabajos sobre intención de voto en Venezuela a descartar los tweets no geoetiquetados, lo cual puede disparar el margen de error hasta extremos totalmente inasumibles, si tenemos en cuenta que el 99% de los tweets no están geoetiquetados.

Para superar este problema, se están prospectando diversas opciones. La solución propuesta por Franco-Riquelme et al. (2019) consiste en fusionar los tweets

extraídos con los datos recuperados de cada perfil de usuario, dando lugar así a un 40% de tweets geolocalizados de todo el conjunto de datos.

La herramienta experimental creada por Agulló et al. (2015) realiza una verificación de dos pasos: primero, se busca la información de geolocalización en los perfiles del usuario; en segundo lugar, si esta información no está disponible en el campo correspondiente, el sistema busca la información de geotiquetado (si ha sido habilitada por el usuario). En lugar de geocodificación inversa, Agulló et al. utilizan la API de Google Maps directamente como medio de presentación gráfica.

3.5 Identificación de usuarios, profiling y manipulación

En este apartado expondremos los hallazgos de las investigaciones realizadas al respecto de este importante aspecto de la tesis. Básicamente, se trata de determinar en qué medida es posible, con las herramientas y técnicas actuales, identificar a personas concretas por sus opiniones y realizar tareas de perfilado individual y colectivo¹⁰⁶ a fin de utilizar esa información posteriormente para influenciar, ya sea con fines lícitos o ilícitos.

3.5.1 Marco legal

De acuerdo con el Reglamento General de Protección de Datos (RGPD) de la Unión Europea, en su apartado 4.4¹⁰⁷, “‘profiling’ means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal

¹⁰⁶ El perfilado individual consiste en asociar atributos concretos a una persona, a fin de obtener un retrato de la misma. El perfilado colectivo generaliza características de un segmento.

¹⁰⁷ RGPD - The General Data Protection Regulation (GDPR), the Data Protection Law Enforcement Directive and other rules concerning the protection of personal data. https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.”¹⁰⁸

Además de esta norma básica, según los expertos en derecho de las nuevas tecnologías entrevistados para la elaboración de esta investigación y fuentes consultadas¹⁰⁹, deben tenerse en cuenta las directrices en virtud del Reglamento 2016/679 “Guidelines on Automated individual decisión-making and Profiling for the purposes of Regulation”¹¹⁰. En estas directrices, el Comité Europeo de Protección de Datos indica tres condiciones para determinar que estamos ante un caso de perfilado: que el tratamiento sea automatizado, que se lleve a cabo involucrando datos personales, y que se realice con el objetivo de evaluar aspectos personales de una persona física.

El perfilado, de acuerdo con la norma, tiene que cumplir con los principios exigibles sobre protección de datos personales, por lo que no sería legal si se realiza sin haber obtenido del usuario tanto su consentimiento para el uso de sus datos como la autorización específica para la finalidad del perfilado.

¹⁰⁸ “elaboración de perfiles' significa cualquier forma de procesamiento automatizado de datos personales que consiste en el uso de datos personales para evaluar ciertos aspectos personales relacionados con una persona física, en particular para analizar o predecir aspectos relacionados con el desempeño de esa persona física en el trabajo, la situación económica, la salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos.

¹⁰⁹ Wolters Kluwer – “La Ley” - Elaboración de perfiles (Protección de Datos) - <https://guiasjuridicas.wolterskluwer.es/Content/Documento.aspx?params=H4sIAAAAAAAAAEAMtMSbF1jTAAAkNjEwtTc7Wy1KLizPw8WyMDQwsDMONDkEBmWqVLfnJIZUGqbVpiTnEqAN76mto1A AA AWKE>

¹¹⁰ “Guidelines on Automated individual decisión-making and Profiling for the purposes of Regulation 2016/679, WP 251 rev.01, revisado el 6 de febrero de 2018, adoptado inicialmente por el Grupo de Trabajo del Artículo 29 y posteriormente por el Comité Europeo de Protección de Datos (CEPD), en el que se integró el 25 de mayo de 2018.

No obstante, dado que es prácticamente imposible que un usuario sepa que se están recolectado y procesando datos de carácter personal publicados en sus redes sociales con finalidades ilícitas, parece especialmente complicado velar por el cumplimiento de la norma.

En otro orden de cosas, existe otra forma de perfilado inverso y anonimizado (y por tanto compatible con la ley), que consiste en crear arquetipos con una serie de atributos comunes que permiten al investigador encontrar de forma automatizada usuarios que “encajen” en el perfil. Más adelante en este apartado se describe con detalle esta práctica, probada experimentalmente y recogida en el Anexo 5.

3.5.2 Identificación de personas concretas

La identificación (o nivel de anonimato) de los usuarios que publican sus opiniones en Internet no es un tema especialmente interesante para los investigadores, a juzgar por los escasos trabajos publicados. A este respecto, es muy destacable el estudio realizado por Peddinti et al. (2014), sobre un *dataset* de 50.000 cuentas, partiendo de un conjunto original de 100.000 en el que fueron depurados *spammers*, *bots* y cuentas efímeras. La siguiente tabla muestra el resumen con los resultados obtenidos:

Label	# of Twitter accounts
Highly Identifiable	906 (1.8%)
Identifiable	34,085 (67.9%)
Partially Anonymous	10,019 (20%)
Anonymous	2,934 (5.9%)
Unclassifiable	3,135 (6.2%)
Total	50,173

Tabla 16 – Cuantificación del anonimato en Twitter según estudio de Peddinti et al. (2014)

En sus conclusiones, Peddinti et al. (2014) demuestran que los usuarios anónimos se sienten menos inhibidos a la hora de participar activamente: tuitean más, siguen más cuentas y están más dispuestos a exponer su actividad al público en general. Esto supone que las cuentas más activas (y posiblemente más interesantes desde el punto de vista del estudio de la opinión), tienden a no ser identificables.

En otro interesante estudio sobre el anonimato, esta vez no referido a Twitter, Gomez et al. (2008), estudiaron los comentarios anónimos de los usuarios en el popular sitio web y foro de noticias tecnológicas Slashdot¹¹¹. Descubrieron que solo el 18,6% de los comentarios eran totalmente anónimos.

La identidad de los usuarios y el estudio de sus perfiles puede hacer que los resultados del análisis de sentimiento sean más precisos (Ahmed et al., 2016). Kim et al. (2011), demostraron que existe una fuerte relación entre la identidad de los usuarios de internet y sus publicaciones. En su estudio, dividieron a los usuarios en clases según una serie de atributos discernibles de manera automatizada (por ejemplo: amabilidad, habilidades sociales, creatividad).

3.5.3 Estrategias y herramientas para identificar de usuarios por patrones y potencialmente influir en su actitud

Como adelantábamos antes, existe otra forma de perfilado distinta al concepto jurídico, que es muy habitual en marketing, investigación de mercados y comunicación política. Este formato alternativo ha sido descrito y probado experimentalmente según se muestra en el Anexo 5 para esta investigación. Se trata de la segmentación por arquetipos para crear audiencias afines.

¹¹¹ <https://slashdot.org/>

El objeto de la investigación presentada en el Anexo 5 es comprobar de forma empírica la posibilidad de realizar el siguiente supuesto, sin utilizar herramientas académicas ni realizar labores que requieran conocimientos de ciencia de datos ni procesamiento de lenguaje natural:

Supuesto: identificar a personas que apoyan el movimiento antivacunas del Covid-19 con las siguientes finalidades de utilidad social práctica:

- Impactar a esas personas directamente con contenidos informativos de concienciación específicamente diseñados y con desmentido de bulos y noticias falsas, de forma automatizada.
- Detección temprana de bulos o cualquier forma de desinformación difundidos por esta comunidad.
- Perfilado del tipo de usuario que participa en defensa/difusión de la opinión.
- Identificación de líderes de opinión.

Para realizar esta comprobación empírica, se utilizó una herramienta comercial informal, Audiense. Esta aplicación aparece como una de las herramientas citadas por los profesionales en la encuesta realizada al respecto (Anexo 2), además ser empleada por profesionales del sector entrevistados directamente para la realización de la tesis.

El resultado de la prueba realizada demuestra de manera concluyente que, para la red social Twitter, es posible identificar a personas que apoyan el movimiento antivacunas del Covid-19, perfilarlas (Figura 45), monitorizar su actividad, e impactarlas con contenidos específicamente creados. Esta posibilidad puede ponerse al servicio de causas legítimas, como la concienciación, el desmentido de falsas informaciones y la investigación académica de los fenómenos sociales, o ilegítimas, como la manipulación.

También hemos comprobado cómo es posible identificar a líderes de opinión, entablar relación de forma automatizada con las cuentas del colectivo monitorizado e interactuar con distintos grados de potencial automático.



Figura 48 – Tag Cloud de perfilado según el caso desarrollado para esta tesis (Anexo 5)

3.6 Aproximaciones metodológicas específicas sobre el impacto de la pandemia del Covid-19

Las noticias, comentarios y debates sobre la pandemia se han venido extendiendo por las redes sociales con especial intensidad, presentando diferentes puntos de vista, opiniones y emociones. Para la comunidad investigadora de la Ciencia de Datos, las ingentes cantidades de material publicado en estos medios son activos extraordinariamente valiosos para comprender los sentimientos de las personas con respecto a los eventos actuales y, por lo tanto, analizar estos sentimientos arrojará resultados especialmente relevantes (Alamoodi et al., 2021).

Pero, más allá de los resultados concretos de los estudios de *sentiment analysis* que se han venido publicando sobre el COVID-19, cabe preguntarse cuál será el impacto en las metodologías de análisis derivado de esta inagotable fuente de datos asociados a sentimientos de gran intensidad.

Los trabajos publicados sobre el estudio de la actitud de los usuarios en redes sociales en torno a la pandemia a menudo se focalizan en contextos nacionales, incluyendo el caso español (de las Heras-Pedrosa et al., 2020), lo que implica implementar una aproximación técnica adecuada para cada idioma, con el consiguiente impacto en el progreso de las metodologías.

Usando tecnologías referidas en el apartado 3.2, hemos realizado para esta investigación una prospección de la actitud de los usuarios “antivacunas”, mediante el seguimiento de uno de sus *hashtags* más populares: #yonomevacuno.

Como resultado, solo para esta etiqueta, obtuvimos 92.695 textos publicados entre el 1 de marzo de 2019 y el 30 de junio de 2022, abarcando medios como Twitter, Facebook, TikTok, blogs, periódicos y foros (ver Anexo 3). Tomando como muestra los correspondientes a la red social Facebook¹¹², observamos que más del 80% tenían polaridad negativa (expresaban sentimientos negativos), como refleja la tabla siguiente:

Hashtag #yonomevacuno	
Positivos	4%
Negativos	81%
Neutros	15%

Tabla 17 - Distribución de la polaridad de los textos analizados con el hashtag #yonomevacuno en Facebook (Anexo 3).

Como es lógico, el tremendo impacto de la pandemia en las redes sociales y la constante amenaza de los contenidos con información falsa o tendenciosa sobre las

¹¹² A fin de no incurrir en el sesgo “preferencia por Twitter” previamente identificado en fases anteriores de la investigación, deliberadamente escogimos Facebook en este caso.

medidas de salud pública, tuvieron una respuesta inmediata en la comunidad investigadora.

Realizando una búsqueda en Scopus de documentos referidos a *Opinion Mining* o *Sentiment Analysis* y Covid-19 ¹¹³, obtenemos un resultado de 1.256 documentos, todos ellos publicados no antes de 2020. El 87% de estos documentos son artículos científicos y *conference papers*.

En otras palabras, podemos observar que en apenas dos años se han producido 1256 trabajos sobre el área de investigación de estas tesis donde el Covid-19 tiene suficiente relevancia como para formar parte del título, las *Keywords* o el *abstract*.

Si realizamos la misma búsqueda omitiendo la referencia a la pandemia, el resultado es de 12.046 documentos (en el mismo período de fechas), lo que nos indica que 1 de cada 10 trabajos científicos realizados entre 2020 y abril de 2022 estaban referidos en alguna medida al Covid-19.

¹¹³ Fórmula de búsqueda: (TITLE-ABS-KEY (covid-19) AND TITLE-ABS-KEY (sentiment AND analysis) OR TITLE-ABS-KEY (opinion AND mining)) – fecha 20/04/2022.

Documents by year

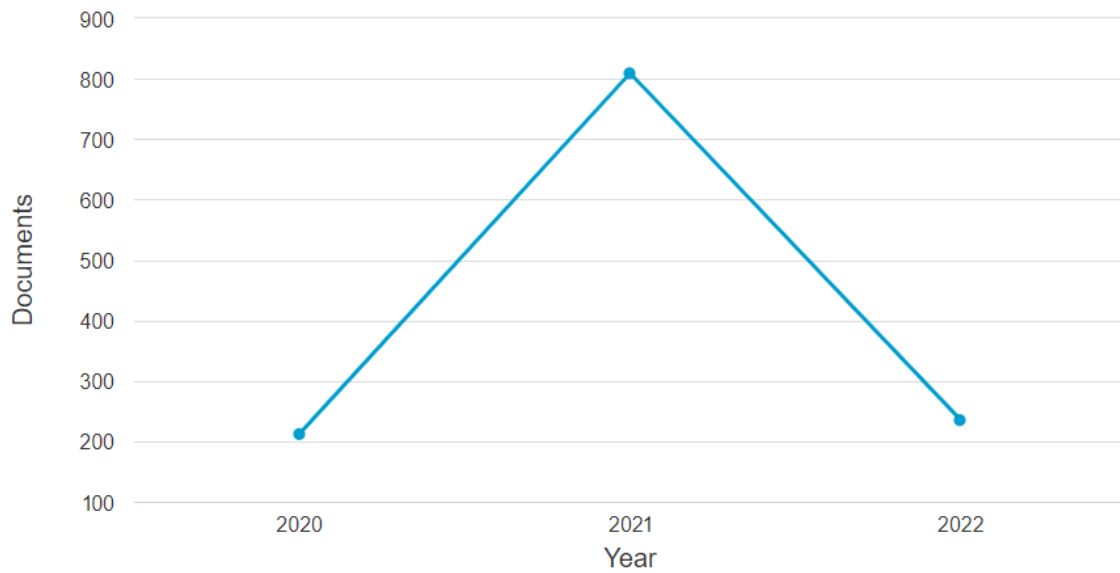


Figura 49 – Trabajos científicos publicados entre 2020 y abril de 2022 referidos en alguna medida al Covid-19. Téngase en cuenta que el año 2022 no está completo (por la fecha de realización de esta tesis). Fuente: base de datos de Scopus

Documents by type

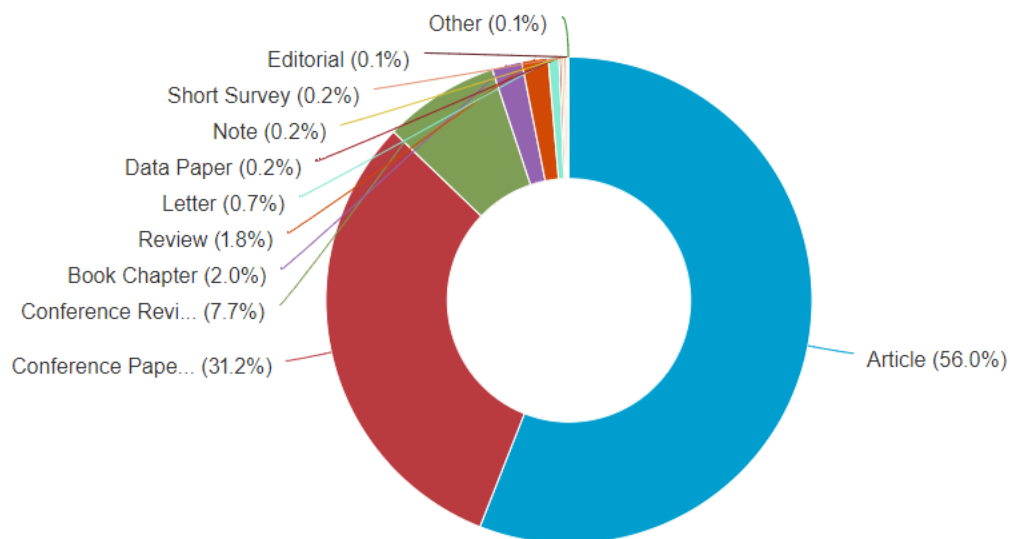


Figura 50 – Trabajos científicos publicados entre 2020 y abril de 2022 referidos en alguna medida al Covid-19 – Por tipo de documento - Fuente: Scopus

Para ser más exigentes con este censo, si nos limitamos al título y las *keywords*, omitiendo así referencias circunstanciales a la pandemia en el resumen, aún nos quedan 693 documentos, lo cual sin duda es una cifra notable en relación con el volumen total de publicaciones.

Parece evidente que, como era previsible, la pandemia se ha incorporado a la esfera de intereses de los investigadores de la minería de opinión y análisis de sentimiento en Internet.

Se observa, además, una correlación frecuente entre la preocupación por la difusión de datos e informaciones sobre las crisis sanitarias (en particular, Covid-19) y las *fake news*:

“Health systems should work on building national and international disease detection and surveillance systems through monitoring social media. There is also a need for a more proactive and agile public health presence on social media to combat the spread of fake news”¹¹⁴ (Abd-Alrazaq et al. 2020, p.1).

3.6.1 Información al alcance de los investigadores sobre la opinión de los ciudadanos en pandemia

Como resultado del estudio referenciado con anterioridad en este apartado (Anexo 3), podemos comprobar el tipo de información a disposición de cualquier observador, incluso “informal”:

¹¹⁴ “Los sistemas sanitarios deben trabajar en la creación de sistemas nacionales e internacionales de detección y vigilancia de enfermedades mediante el seguimiento de los medios sociales. También es necesaria una presencia más proactiva y ágil de la salud pública en las redes sociales para combatir la difusión de noticias falsas.”

- Totalidad de los textos (*dataset*) para su análisis automatizado (o cualitativo), con información asociada como identidad del usuario (real o ficticia, ver apartado anterior “identificación de usuarios”), volumen de interacción detallado (*likes*, compartido, etc.), alcance, métricas calculadas (influencia, *engagement rate*) y polaridad (con *accuracy*, *recall* y *F-Score* limitados, según conclusiones de nuestros estudios en Anexos 2 y 3).
Ejemplo:

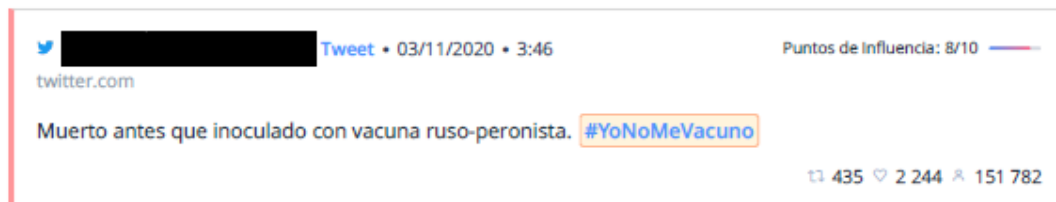


Figura 51 – Ejemplo de tuit recogido y procesado con la herramienta SentiOne (identidad visible en el original, se ha ocultado aquí por consistencia con las normas de la API de Twitter)

- Términos que más se repiten en los mensajes del grupo objeto de análisis, permitiendo identificar intereses, consignas, y palabras clave (*keywords*) que pueden ser objeto de estudio posterior o combinado (en este caso, “covidiotas”, “plandemia”, “falsapandemia” o “dictaduras sanitaria”, entre otras):



Figura 52 – Palabras clave más repetidas en los 92695 textos analizados

- Indicios de organización o connivencia (ya sea espontánea o planificada). En este caso, la consistencia en el uso de palabras clave apunta a un cierto grado de esfuerzo consciente del grupo por mantener la unidad del mensaje, como podemos ver en el gráfico siguiente:



Figura 53 – Top hashtags usados por los usuarios que realizan publicaciones con #yonomevacuno

- Distribución por género

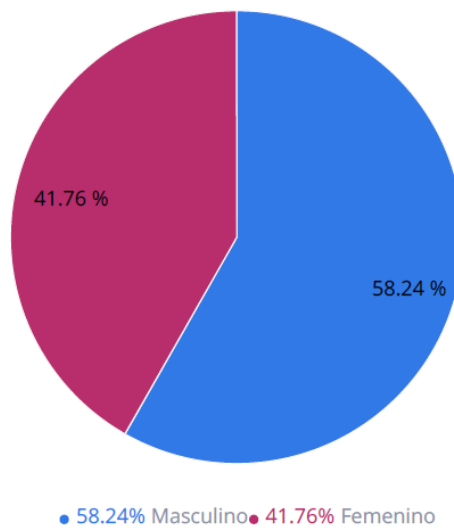


Figura 54 – Uso del hashtag #yonomevacuno, distribución por género

- Evolución en el tiempo:

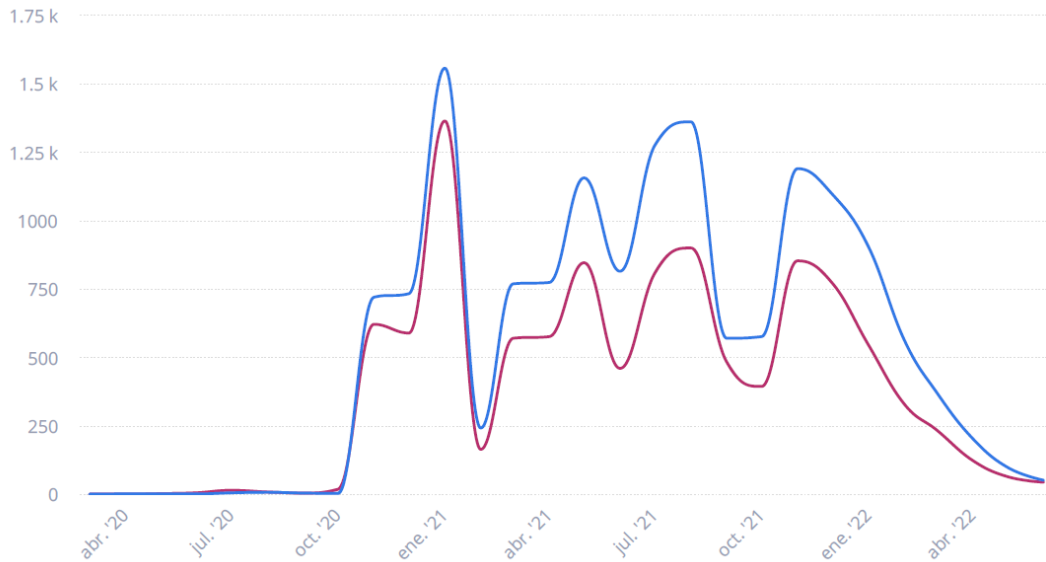


Figura 55 – Uso del hashtag #yonomevacuno a lo largo del tiempo en todo Internet (redes sociales, blogs, foros, medios de comunicación), por género

- Alcance (personas a las que ha llegado el mensaje)



Figura 56 – Alcance de las publicaciones con el hashtag #yonomevacuno

Otras herramientas, como Audiense (Anexo 5) y Brandwatch¹¹⁵ permitirían un análisis más centrado en el perfilado de los usuarios. En las imágenes siguientes (Figuras 57 y 58), podemos ver como BrandWatch permitiría identificar a profesionales de la salud y tener un retrato preciso sobre aspectos como cuáles son sus aficiones e intereses, qué profesiones tienen o de qué otras cosas hablan. Al igual que Audiense, esta herramienta permitiría crear una “audiencia” con estos usuarios y monitorizar/analizar sus publicaciones de forma automática.

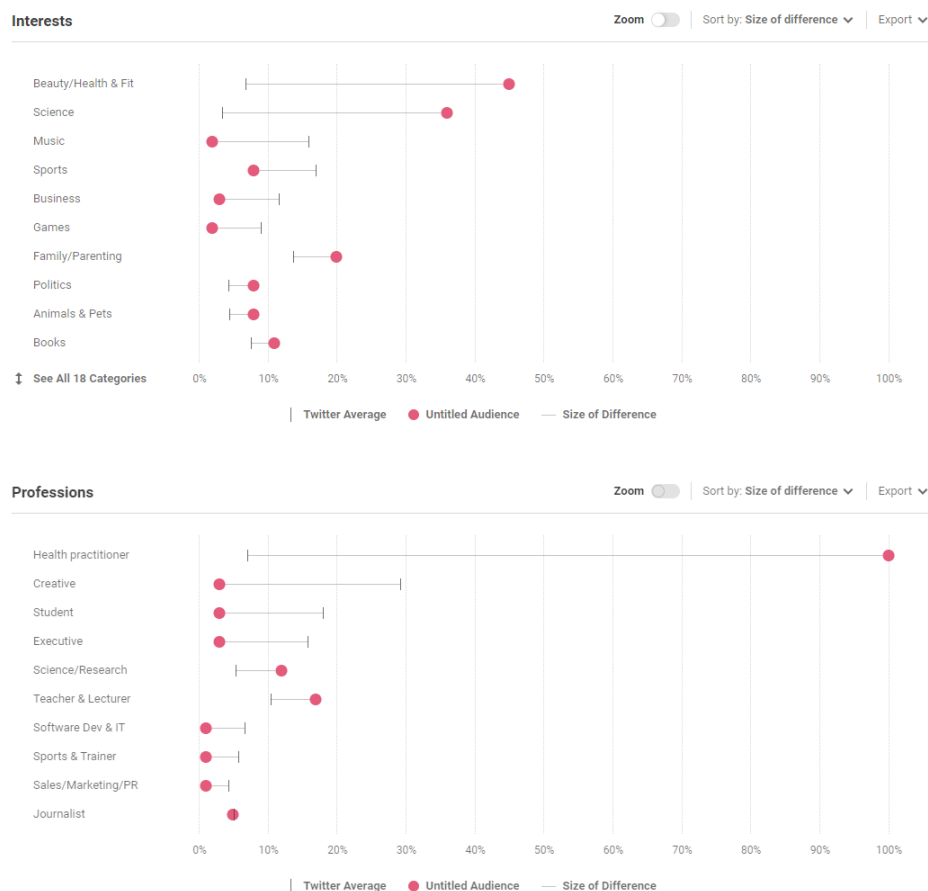


Figura 57 – Vista de detalle sobre intereses y profesiones de profesionales de la salud en España que hablan de Covid-19, con la herramienta BrandWatch

¹¹⁵ <https://www.brandwatch.com/es/>

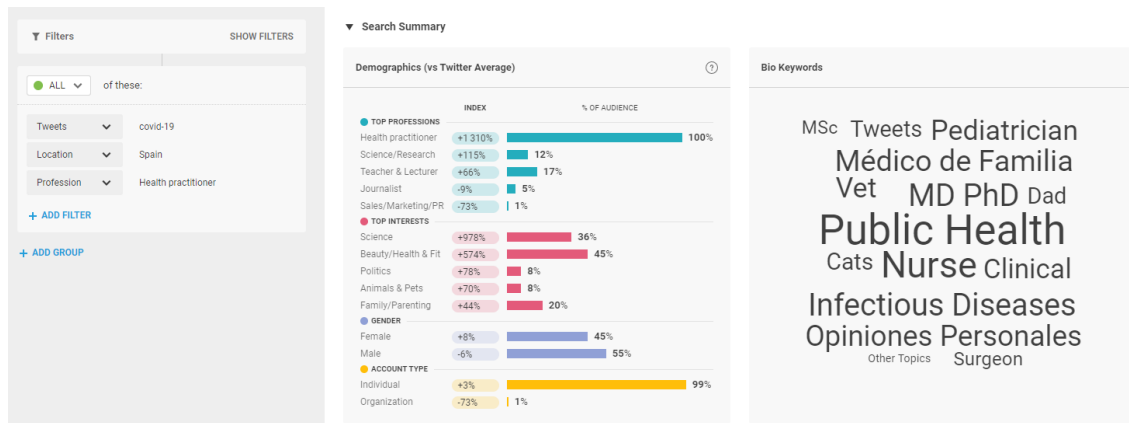


Figura 58 – Vista de una query con BrandWatch sobre profesionales de la salud que hablan de Covid-19

3.6.2 Aplicación práctica de la monitorización de la actitud con relación a la pandemia

El Covid-19 no es un mero *dataset subject* para desarrollar o evaluar metodologías de procesamiento automático de textos, como así lo demuestran algunos estudios de referencia estudiados. Mediante técnicas de *sentiment analysis*, los investigadores están consiguiendo obtener *insights* de gran utilidad práctica en políticas de salud pública y otros aspectos de interés social, más allá del ámbito técnico-académico.

De hecho, un enfoque analítico basado en *sentiment analysis* es útil para establecer niveles de reticencia a la vacunación en el público en general así como para combatir inquietudes específicas y desinformación (Roe et al., 2021).

Un estudio muy relevante¹¹⁶ sobre la salud mental en la pandemia (Valdez et al., 2020), concluyó recientemente que, a la luz del efecto negativo documentado del

¹¹⁶ 60 citas, dato según SCOPUS, actualizado a julio de 2022.

uso intensivo de las redes sociales en la salud mental, éstas pueden exacerbar aún más los sentimientos negativos a largo plazo para muchas personas. Por lo tanto, considerando la sobrecargada estructura de atención de salud mental de los EE. UU., estos hallazgos tienen implicaciones importantes para los esfuerzos de mitigación en curso.

En el apartado 3.5.3, hemos descrito una investigación realizada para esta tesis (Anexo 5), que ejemplifica un caso práctico representativo de las posibilidades de “acción práctica” sobre los usuarios, más allá de la mera observación de su comportamiento.

3.7 Impacto de la preocupación social y académica por el discurso de odio y la polarización política

Es importante partir de la base de que el concepto “discurso de odio” no es una etiqueta periodística difusa propia del debate sociológico, sino un concepto perfectamente definido, internacionalmente, desde los ámbitos académico y legal.

La Recomendación de la Comisión Europea contra el Racismo y la Intolerancia de 2016¹¹⁷ define el discurso de odio (*hate speech*) como la incitación al odio, humillación o menosprecio en cualquier forma contra una persona o un grupo, motivada por la raza, el color de la piel, la ascendencia, el origen nacional o étnico, la edad, la discapacidad, la lengua, la religión o las creencias, el sexo, la identidad de género, la orientación sexual u otras características o condiciones personales. Esta recomendación requiere una definición previa del concepto “incitación al odio”. A

¹¹⁷https://www.inclusion.gob.es/oberaxe/ficheros/documentos/2016_12_21-Recomendacion_ECRI_NO_15_Discurso_odio-ES.pdf

ese respecto, la Decisión Marco 2008/913/JAI del Consejo de Europa ¹¹⁸, relativa a la lucha contra determinadas formas y manifestaciones de racismo y xenofobia mediante el Derecho penal define la incitación al odio como "la incitación pública a la violencia o al odio dirigida contra un grupo de personas o un miembro de dicho grupo definido por referencia a la raza, el color, la religión, la ascendencia o el origen nacional o étnico" (p. L 328/56).

Desde una perspectiva académica, diversos autores han aplicado propuestas metodológicas para la definición/identificación del discurso de odio en los contenidos en Internet, como Miró (2016) o Waseem y Hovy (2016).

La preocupación de los investigadores, por otra parte, no se limita a la identificación del odio tal y como se define formalmente, sino que a menudo se amplía al concepto de *rejection* (rechazo), considerado como una construcción social (Berger & Luckmann, 1966) y como un potencial predictor del discurso de odio (Arcila Calderón et al., 2020).

Las corporaciones propietarias de las redes sociales, los gobiernos e instituciones públicas y privadas, están desarrollando un interés creciente por la detección del discurso de odio en las redes sociales. Sin embargo, desde el punto de vista de la ciencia de datos, las metodologías y herramientas disponibles son muy recientes. Uno de los desafíos más importantes está en la categorización de los distintos tipos de discurso de odio: no basta con disponer de herramientas computacionales de detección, es preciso etiquetar cada literal para que la información resultante sirva para activar posteriores medidas de intervención automatizadas.

¹¹⁸ <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32008F0913>

De acuerdo con la revisión sistemática realizada sobre *sentiment analysis* en el dominio de la política y el discurso de odio (Anexo 6) para esta investigación, es significativo señalar que antes de 2019 no existía un solo trabajo de análisis de sentimiento para el idioma español dedicado específicamente al discurso de odio. El creciente interés por la detección del discurso de odio a partir de 2020 está directamente alineado con la preocupación social por el posible impacto negativo que estos mensajes pueden tener en los individuos o en la sociedad (Pereira-Kohatsu et al., 2019).

En la investigación referida (Anexo 6), hemos podido identificar cuatro objetivos diferentes en los trabajos publicados sobre el discurso de odio:

- Análisis del fenómeno (eventos pasados).
- Detección (identificar el odio).
- Monitorización (detección en tiempo real).
- Elaboración de perfiles.

A diferencia de la orientación habitual de los trabajos centrados en las elecciones y otros aspectos de la política, en el caso del discurso del odio se añade un componente de practicidad social al mero debate académico-metodológico. Esta intención se refleja especialmente en la herramienta HaterNet (Pereira-Kohatsu et al., 2019) un sistema inteligente que actualmente está siendo utilizado por la Oficina Nacional contra los Delitos de Odio de la Secretaría de Estado de Seguridad, que identifica y monitoriza la evolución del discurso de odio en Twitter.

Si adoptamos una perspectiva más amplia y consideramos todos los trabajos publicados desde 2014 sobre textos en español (Anexo 6), observamos que la mayoría de los estudios (58%) se centran en el debate sobre métodos experimentales y en el análisis comparativo de modelos, en algunos casos relacionados con estudios de casos específicos. De 26 trabajos publicados sobre política, el 30% están enfocados en la predicción de resultados electorales.

Otro tema de creciente interés son los *bots* políticos. Pastor-Galindo et al. (2020) demostraron claramente la existencia de un número significativo de *bots* participando activamente en el debate público sobre las elecciones españolas de 2019 en Twitter, con un número importante de interacciones y volumen de tráfico. Este conjunto de opiniones falsas tiene un impacto evidente en todos los subdominios si no se detectan y descartan en los conjuntos de datos.

3.8 Accesibilidad de los recursos a personal no cualificado

No sería posible responder la pregunta de investigación sin comprobar por procedimientos objetivos la accesibilidad de usuarios “no cualificados” al estudio de la actitud de los ciudadanos en Internet. Dicho más llanamente, ¿hablamos de un tipo de actividad que requiere grandes conocimientos académicos de *machine learning* o podría hacerlo “cualquiera”?

Como parte de las investigaciones realizadas para esta tesis, la comprobación correspondiente se ha realizado por tres vías: estadística, *desk research* y empírica (investigación propia).

3.8.1 Vía empírica (investigación propia mediante *Contextual Inquiry*)

Se ha realizado un pequeño estudio (Anexo 8) con el objetivo de determinar si personas no expertas en ciencia de datos, con distintos grados de conocimiento previo sobre interfaces digitales, son capaces de extraer y procesar la información sobre la actitud de los usuarios en Internet usando una herramienta comercial. A tal efecto, se ha elegido SentiOne, aplicación que obtuvo los mejores rendimientos en las investigaciones previas (Anexos 1 y 3).

Dado que se trata de evaluar la usabilidad de la herramienta y su relación con el usuario, aplicamos una aproximación basada en los métodos propios de la investigación para el diseño de interfaces gráficos, con la técnica *contextual inquiry*,

una veterana metodología etnográfica descrita en la década de los 90 por Bennett et al. (1990). Este tipo de metodología es muy habitual en la actualidad en el ámbito de la investigación en UX¹¹⁹.

El tipo de objetivos que satisface la técnica *contextual inquiry* no tiene que ver con evaluar el rendimiento de los procesos, sistemas o productos, sino con entender aspectos cognitivos y procedimentales (aspectos de uso), o responder a preguntas de investigación relacionadas con la usabilidad y accesibilidad. En algunos casos, *el contextual inquiry* se usa como información básica para el diseño de sistemas de nueva planta, y en otros casos, como el que nos ocupa, para prospectar la relación que se establece entre los usuarios y los sistemas ya existentes.

La observación realizada concluye, desde el punto de vista del análisis de los datos cualitativos recogidos, que la herramienta analizada (SentiOne) permite a personas profesionales del marketing sin conocimientos específicos de ciencia de datos ni procesamiento de lenguaje natural, programar y ejecutar monitorizaciones sobre la actitud de los usuarios en Internet y *sentiment analysis*. También se ha comprobado que una persona sin conocimientos de marketing ni estudios avanzados, partiendo solo del conocimiento popular sobre el uso de aplicaciones digitales y el funcionamiento general de Internet, podría realizar la misma tarea, aunque en este caso sin tener un nivel de comprensión suficiente como para sacar partido a la aplicación.

¹¹⁹ User Experience (Experiencia de Usuario). La experiencia del usuario es la disciplina que estudia la percepción del usuario al respecto de su interacción con sistemas, productos y servicios, en un contexto determinado. Dicha percepción no se limita al diseño y a los procesos, sino que incluye aspectos emocionales, expectativas y recuerdos. Es decir, no se trata sólo de qué sucede sino también, y especialmente, de la huella que deja lo que sucede, o lo que el usuario espera que suceda.

3.8.2 Vía estadística

En los resultados de la encuesta realizada entre profesionales (Anexo 2), tal y como vimos con anterioridad, observamos que un 15% de los encuestados no solo afirmaron usar herramientas para el análisis de la actitud de los usuarios en Internet, sino que, además, pudieron mencionar alguna correctamente¹²⁰ (incluyendo herramientas evaluadas para esta investigación). Este porcentaje es tres veces superior a los encuestados que manifestaron tener actividades profesionales que podemos relacionar con la Ciencia de Datos (5%), lo que no deja lugar a dudas.

El total de encuestados que manifestaron haber usado personalmente este tipo de herramientas (mencionando alguna o no), fue del 33,8%. Aunque ya vimos con anterioridad que el porcentaje real puede ser sensiblemente inferior¹²¹.

3.8.3 Desk research

Numerosos informes, artículos y materiales profesionales de todo tipo se refieren al uso de las herramientas de análisis de la actitud de los usuarios en Internet como actividades completamente normalizadas en ámbito de la empresa. Como ejemplo, podemos citar el informe de Forrester¹²² donde se relata cómo el equipo de fútbol Manchester United utilizó Audiense (herramienta testada para esta investigación,

¹²⁰ “Correcta” en el sentido de válida desde el punto de vista de la materia desarrollada en esta tesis. Obviamente, una evaluación menos rigurosa arrojaría un porcentaje mayor.

¹²¹ Muchos participantes no fueron capaces de citar herramientas o dieron respuestas que no eran consistentes con la pregunta.

¹²² Empresa de referencia mundial en investigación de marketing. <https://www.forrester.com/>

Anexo 5) para ayudar a entender el porqué de su gran audiencia en la India y llegar a posibles patrocinadores que ya están alineados con los seguidores del equipo¹²³.

Otro ejemplo representativo lo tenemos en una firma bancaria de Sudáfrica que quería conocer el sentimiento de sus usuarios para cada aspecto del servicio por separado, a fin de poder mejorar sus políticas de forma más precisa y tener información profunda sobre la percepción de los clientes¹²⁴. Mediante un análisis de sentimiento basado en *named entity recognition*, la herramienta comercial Repustate (analizada para esta tesis, Anexo 3) fue utilizada para este fin.

¹²³ <https://resources.audiense.com/reports/forrester-report-what-marketers-need-to-know-about-social-media-consumer-segmentation>

¹²⁴ <https://www.repustate.com/blog/sentiment-analysis-real-world-examples/>

4 Conclusiones

Con toda la información recogida en la investigación que se ha realizado, y tras su análisis y discusión, podemos extraer varias conclusiones muy claras para evaluar la hipótesis planteada.

Recordemos primero nuestra hipótesis:

- **Hipótesis:** cualquier persona u organización con conocimientos técnicos avanzados de marketing digital está potencialmente capacitada para generar y procesar información exhaustiva sobre la actitud de los usuarios en Internet, ya sea con fines puramente analíticos o con intención de influenciar procesos de formación de opinión y toma de decisiones. Para este fin, es preciso utilizar herramientas tecnológicas y metodologías de distinta naturaleza, precisión y alcance, que están en continua transformación.

A fin de sistematizar la exposición de las conclusiones de la manera más precisa posible, se analizarán las siguientes variables en las que podemos descomponer el problema, en línea con los objetivos y preguntas de investigación:

- **Objetivo específico 1: acceso a la información.** ¿Es accesible la información? ¿con qué limitaciones? ¿se puede recopilar masivamente? ¿es posible identificar asuntos de interés específicos?
- **Objetivo específico 2: procesamiento de la información.** ¿Con qué nivel de granularidad se puede realizar el procesamiento y análisis? ¿qué se puede llegar a conocer con el análisis de los datos?
- **Objetivo específico 3: posibilidad de monitorización:** ¿es posible realizar seguimientos continuados y en tiempo real sobre temas concretos?
- **Objetivo específico 4: vinculación de actitudes a personas concretas y perfilado.** ¿Se puede estudiar a usuarios concretos? ¿qué se puede saber

sobre ellos? ¿se pueden crear arquetipos por atributos o intereses comunes de un segmento de usuarios?

- **Objetivo específico 5: precisión y exactitud en el análisis.** ¿Cuál es el rendimiento real que podemos esperar con los métodos y herramientas actuales? ¿Cómo de eficaces son los recursos que están “a disposición de cualquiera” en comparación con los métodos más avanzados de la comunidad científica?
- **Objetivo específico 6: volumen y coste de medios requeridos.** ¿Qué medios técnicos son necesarios para acceder a la información y procesarla? ¿qué herramientas del conocimiento se precisan? ¿cuáles son los costes involucrados?
- **Objetivo específico 7: acceso a los medios requeridos.** ¿Quién puede acceder a la información y a su procesamiento para conocer efectivamente la actitud de los ciudadanos en Internet?
- **Objetivo específico 8: posibilidad de influir.** ¿Pueden las herramientas de monitorización y análisis de la actitud de los ciudadanos en Internet ser utilizadas para influir en la toma de decisiones de las personas, directa o indirectamente?
- **Objetivo específico 9: estado del arte – obstáculos y desafíos de la comunidad investigadora.** ¿En qué punto se encuentra el estado del arte? ¿cuáles son los principales desafíos a los que se enfrentan los investigadores? ¿Qué desafíos específicos tenemos por delante para el análisis en lengua española?
- **Objetivo específico 10: analizar el impacto de la pandemia del Covid-19 en el tema de estudio.** ¿Ha influido la pandemia en el estado del arte? ¿Cuál ha sido el impacto, si lo ha habido, en la comunidad investigadora?

Veamos a continuación nuestras conclusiones para cada uno de los aspectos enumerados, ateniéndonos rigurosamente a los hallazgos de la investigación realizada.

4.1 Objetivo específico 1: acceso a la información

¿Es accesible la información? ¿Con qué limitaciones?

Podemos concluir que es posible detectar, extraer y procesar opiniones concretas en Internet, sobre temas específicos de cualquier naturaleza, siempre que sean de acceso público. Obviamente, no podemos descartar la posibilidad de traspasar las líneas rojas de las normas de privacidad mediante acciones de piratería informática o tramas como la denunciada en el caso “Cambridge Analytica”. Si nos atenemos a los límites de acceso lícito a los sistemas, solo es posible acceder a contenidos “públicos”.

¿Se puede recopilar masivamente?

Las opiniones se pueden acceder y coleccionar para su posterior procesamiento de forma automatizada y masiva.

Actualmente, la fuente de la mayoría de los trabajos realizados por la comunidad investigadora es la red social Twitter, que dispone de una API que permite la extracción de datos de manera muy sencilla. No obstante, con independencia de los criterios de conveniencia de los investigadores, se ha demostrado (bibliográfica y empíricamente) la existencia de herramientas que permiten hacer *scrapping* de cualquier información en cualquier medio, como por ejemplo Facebook, Instagram, blogs, foros o medios de comunicación (siempre que los datos sean públicamente accesibles), por lo que resultaría posible coleccionar y someter a análisis cualquier opinión expresada en la Red.

¿Es posible identificar asuntos de interés específicos?

Podemos concluir también que a la hora de identificar la información es posible trabajar sobre dominios amplios (por ejemplo, “discurso de odio”), sobre temas

concretos (Covid-19, una empresa pública, un programa de televisión), e incluso sobre aspectos específicos de un tema (por ejemplo, un atributo de un producto).

Existen metodologías que permiten diferenciar una opinión de algo que no lo es, así como identificar estructuras sintácticas.

4.2 Objetivo específico 2: procesamiento de la información

¿Con qué nivel de granularidad se puede realizar el procesamiento y análisis?

Se pueden analizar las opiniones con cualquier nivel de granularidad, tanto a escala de documento completo, como de frase, entidad o aspecto, existiendo metodologías específicas para cada uno de estos niveles.

¿Qué se puede llegar a conocer con el análisis de los datos?

En el estado del arte actual, tal y como se ha podido comprobar de forma empírica, el análisis de los textos permite saber si la opinión identificada es positiva, negativa o neutra, ya sea de manera absoluta o en forma de puntuación numérica, con el nivel de granularidad anteriormente indicado. Esto hace posible un posterior procesamiento estadístico sobre el total de los textos analizados.

También pueden asociarse las opiniones a usuarios concretos, y en determinados casos conocer su alcance y nivel de influencia, lo que puede permitir identificar líderes de opinión, así como personas especialmente activas/comprometidas.

La comunidad científica ha desarrollado metodologías que pueden identificar la ironía y sentimientos específicos en los mensajes, con distintos grados de éxito.

En determinados casos, es posible asociar usuarios a áreas geográficas concretas, en función de si esta información está disponible o no en la fuente, y de las políticas de privacidad.

No hay un consenso de la comunidad investigadora sobre las metodologías de procesamiento de la información. Por el contrario, hay un flujo constante, y en rápido crecimiento, de trabajos que proponen distintas aproximaciones a los desafíos existentes, en su mayoría basadas en *machine learning*.

4.3 Objetivo específico 3: posibilidad de monitorización en tiempo real

¿Es posible realizar seguimientos continuados y en tiempo real sobre temas concretos?

Como resultado de nuestra investigación, hemos podido comprobar que se han desarrollado metodologías y herramientas específicamente creadas para la monitorización de temas concretos (por ejemplo, “discurso de odio”), por lo que la comunidad investigadora no está limitada a análisis retrospectivos referidos a sets de textos previamente extraídos en un rango concreto. También hemos comprobado empíricamente este tipo de aproximación.

Del mismo modo, hemos visto cómo las herramientas comerciales disponen de funcionalidades para la generación de informes periódicos y, sobre todo, sistemas de “alarma” para generar avisos automáticos si se cumplen determinadas condiciones/reglas en tiempo real.

4.4 Objetivo específico 4: vinculación de actitudes a personas concretas y perfilado

¿Se puede estudiar a usuarios concretos?

Dentro de una misma red social, es posible acceder a todos los comentarios publicados por una persona concreta, lo que permitiría conocer distintos aspectos de su actitud pública. Hemos comprobado empíricamente cómo una herramienta

informal de bajo coste (Audiense, Anexo 5) permitiría monitorizar la actividad de una lista de usuarios concretos, previamente creada como una “audiencia” mediante una serie de condiciones y filtros. Igualmente, hemos estudiado otra herramienta de coste superior (BrandWatch) capaz de hacer lo mismo, con funcionalidades extra más avanzadas.

¿Qué se puede saber sobre ellos? ¿Se pueden crear arquetipos por atributos o intereses comunes de un segmento de usuarios?

Como hemos visto, también es posible acceder a información en los perfiles biográficos y aplicar métodos de *profiling* masivo individual o colectivo, lo que potencialmente permitiría no solo saber qué personas sintonizan con determinadas actitudes, sino también qué otros intereses tienen, cómo son de activas, o cómo se relacionan entre ellas.

En particular, hemos comprobado cómo con herramientas comerciales al alcance de cualquier persona con unos mínimos conocimientos de marketing digital, en algunos casos de acceso gratuito, podría conocerse la polaridad de los sentimientos de las cuentas de una red social sobre cualquier asunto, con una precisión próxima al 70%.

No obstante, el nivel de conocimiento de datos de carácter personal del usuario que se puede llegar a alcanzar depende de la cantidad de información que el propio usuario decida dar libremente (ejemplo: hay usuarios que están en Twitter con su nombre real y usuarios que no; del mismo modo hay usuarios que permiten ver información de geolocalización y usuarios que no). Por otro lado, los trabajos realizados por la comunidad investigadora para avanzar en aspectos como la delimitación geográfica de los textos, arrojan hasta la fecha resultados muy limitados y poco esperanzadores.

En el caso concreto de Twitter, que tal y como hemos visto es, con gran diferencia, la fuente más analizada, podemos estimar que las cuentas identificables (total o parcialmente) están en torno al 69% (Peddinti et al., 2014).

4.5 Objetivo específico 5: precisión, exactitud, y F-Score

¿Cuál es el rendimiento real que podemos esperar con los métodos y herramientas actuales?

Usando herramientas académicas (por contraposición a los recursos comerciales poco formales) el rendimiento en la medición de la polaridad varía mucho en función del método empleado y del dominio. Muchos trabajos publicados arrojan resultados por encima del 70% en F-Score, pero diversos estudios demuestran que es posible superar en la actualidad el 80%. El estado del arte está lejos de desarrollar métodos que permitan, de forma consistente, trabajar con márgenes de error pequeños, equiparables a los estándares estadísticos habituales en un entorno científico.

¿Cómo de eficaces son los recursos que están “a disposición de cualquiera” en comparación con los métodos más avanzados de la comunidad científica?

Hemos podido comprobar mediante una investigación *ad hoc* para esta tesis, cómo una herramienta “informal” puede aproximarse en rendimiento a la mayoría de los métodos en la literatura científica, exceptuando las metodologías más avanzadas y los referentes actuales en *benchmarking*.

Los mejores resultados obtenidos en las pruebas realizadas para esta tesis con una herramienta no analizada hasta ahora en la literatura científica nos llevan a un *F-Score* de 61.45, *Accuracy* de 63.45 y *Recall* de 68.86, cifras ligeramente por encima de los valores habituales en los estudios realizados por otros autores para otras herramientas comparables.

4.6 Objetivo específico 6: medios requeridos

¿Qué medios técnicos son necesarios para acceder a la información y procesarla?

En este aspecto, la investigación nos ha llevado a identificar dos niveles claramente diferenciados de medios: recursos científicos y recursos informales.

Aproximación científica: la comunidad científica tiene a su disposición dos categorías de herramientas y necesita de ambas para poder desarrollar su trabajo. Por un lado, recursos de uso académico para tareas concretas, desarrollados por otros investigadores y puestos a disposición de la comunidad de forma gratuita o comercial. Por otro lado, los métodos, ya sea de elaboración propia o tomados de las propuestas realizadas por los trabajos previos del resto de académicos. Como parte de esta investigación, se ha hecho un desarrollo exhaustivo de esta información en capítulos precedentes.

Aproximación informal: el estudio de la actitud de los ciudadanos en Internet desde el ámbito privado requiere el acceso a herramientas tipo SAAS que se contratan por períodos de tiempo, generalmente mensuales. Estas herramientas son un “todo en uno”, donde el usuario no tiene acceso al código ni posibilidad de modificar esencialmente los métodos. Se usan a través de un *dashboard* aplicando las funcionalidades de recolección y análisis de información disponibles en cada caso. Dado el especial interés de estas herramientas desde el punto de vista de la hipótesis formulada, se han realizado varios trabajos empíricos sobre ellas, desarrollados en capítulos precedentes y en los anexos.

¿Qué herramientas del conocimiento se precisan?

Aproximación científica: el desarrollo de metodologías *ad hoc* para el estudio de la actitud de los ciudadanos en Internet, en el ámbito de la investigación científica, es

muy exigente en cuanto a las herramientas del conocimiento, ya que requiere aplicar (cuando no diseñar) métodos propios de la ciencia de datos en general, y del procesamiento del lenguaje natural en particular, incluyendo técnicas de *machine learning* (y por tanto Python¹²⁵) y una sólida base de lingüística.

Aproximación informal: en este territorio hemos podido comprobar, incluso experimentalmente, la existencia de herramientas tipo SAAS que permitirían a cualquier usuario con un perfil medio de conocimientos en comunicación digital acceder a un nivel de análisis muy considerable, sin alcanzar la profundidad ni la precisión de los métodos académicos. Algunos proveedores de este tipo de servicios, como por ejemplo BrandWatch, cuentan con programas de formación para nuevos clientes a fin de hacer asequibles sus sistemas a personal no cualificado en ciencia de datos y asegurarse un uso óptimo. La herramienta SentiOne puede ser usada por personal *junior* sin experiencia previa con apenas una hora de formación.

Solo expertos investigadores cualificados en procesamiento de lenguaje natural pueden realizar análisis de la actitud de los ciudadanos en Internet manteniendo el control sobre la metodología aplicada.

¿Cuáles son los costes involucrados?

Aproximación científica: no se requieren recursos tecnológicos ni económicos significativos, ya que los procesos pueden realizarse con equipos informáticos convencionales, y el software requerido es generalmente de código abierto, o en su

¹²⁵ Python es un lenguaje de programación multiparadigma, distribuido con licencia de código abierto, muy habitualmente utilizado en Ciencia de Datos y, en particular, en el desarrollo de aplicaciones basadas en *Machine Learning*.

defecto de bajo coste. La capacidad de proceso necesaria no es un problema para los recursos informáticos de cualquier organización académica moderna.

Aproximación informal: los recursos comerciales informales pueden llegar a tener un precio significativo, pero se encuentran dentro de la horquilla de costes habitual en el repertorio de herramientas digitales de cualquier organización de tipo medio. En algunos casos, podrían ser asequibles incluso para usuarios particulares, como es el caso de Audiense.

4.7 Objetivo específico 7: acceso a los medios requeridos

¿Quién puede acceder a la información y a su procesamiento para conocer efectivamente la actitud de los ciudadanos en Internet?

Hemos comprobado cómo cualquier persona con conocimientos de marketing digital, con una curva de aprendizaje equivalente a la familiarización básica con un sistema digital no complejo, puede realizar investigaciones sobre la actitud de los ciudadanos en Internet, pero de forma limitada a las funcionalidades que permita la aplicación comercial empleada (Anexo 8). De hecho, hemos visto cómo, en la práctica, hay profesionales de ámbito de la empresa privada, sin especialización en ciencia de datos, que ya realizan este tipo de tareas regularmente como parte de su actividad.

Por otro lado, también hemos constatado que el uso de herramientas de monitorización está “normalizado” en el ámbito del marketing y la comunicación.

4.8 Objetivo específico 8: posibilidad de influir y manipular

¿Pueden las herramientas de monitorización y análisis de la actitud de los ciudadanos en Internet ser utilizadas para influir en la toma de decisiones de las personas, directa o indirectamente?

Como resultado de la investigación realizada, no se observa que exista posibilidad de manipular directamente a los ciudadanos, pero sí de manera indirecta. Como muchas otras formas de conocimiento, los datos meramente analíticos pueden ser puestos al servicio de un fin transformador. En nuestro caso, hemos comprobado empíricamente cómo una popular herramienta concreta posee las funcionalidades necesarias para desarrollar estrategias potencialmente susceptibles de influenciar audiencias (listas de usuarios específicos) previamente perfiladas, e incluso impactarlas mediante publicidad especialmente dirigida (Anexo 5). Por supuesto, este tipo de estrategias podrían servir potencialmente a fines legítimos o ilegítimos, y esto es algo que queda fuera del ámbito de la funcionalidad tecnológica y nos adentra en el de la moral.

Hemos comprobado, por ejemplo, que es posible conocer en tiempo real o de forma retrospectiva, sin más limitación que la disponibilidad pública de los contenidos y las leyes¹²⁶, aspectos como:

- Asuntos de mayor interés para un grupo que personas con atributos comunes.
- Orientación de personas concretas o de audiencias perfiladas con relación a ciertos temas.
- Cuentas de personas que lideran la opinión.

También hemos comprobado que es muy fácil crear de forma dinámica y automatizada listas de usuarios (identificables en un porcentaje de casos – ver objetivo 4) que tienen en común una posición determinada con respecto a asunto específico. Estas listas permitirían involucrar a estas personas posteriormente con

¹²⁶ En el supuesto de que la ley pueda ser considerada como un freno absoluto para el 100% de las personas.

contenido, mensajes, o anuncios, específicos para ellas. También sería posible, por el procedimiento “nube de tags”, descubrir cuáles son los temas de interés que más motivan a estas personas, a fin de perfilarlas y de sintonizar los contenidos creados para ellas con fines potencialmente manipuladores (Anexo 5).

Por último, dado que es posible en muchos casos evaluar el alcance de las publicaciones (Twitter), también es posible procesar la información discriminando los textos analizados en función de este dato, lo que permitiría deducir qué tipos de mensaje tienen más éxito.

Es decir, sería posible “perfilar” a los usuarios, tanto de manera individualizada, utilizando información como el nivel de influencia de sus publicaciones, su *bio*, o la intensidad de su afinidad/rechazo, como de forma colectiva, conociendo los términos que más repiten en los contenidos de un colectivo con atributos comunes y otros aspectos psicológicos discernibles por las herramientas de monitorización.

Todo lo anterior puede hacerse con aplicaciones de uso comercial al alcance de cualquier persona, es decir, no hablamos de sofisticadas metodologías de la comunidad académica, sino aplicaciones de fácil acceso y uso, cuya tecnología no se puede conocer ni alterar (no son de código abierto).

Los tipos de información comentados en este apartado, potencialmente, hacen posible desarrollar estrategias de todo tipo, tanto lícitas como ilícitas, ya sea desde un punto de vista ético o, en determinados casos, legal:

- Movilizar (exacerbar) o desmovilizar segmentos de electorado.
- Promover la viralización de ciertos contenidos, incluidas noticias falsas.
- Sintonizar contenidos con los asuntos de interés de una cohorte.
- Determinar temáticas de publicación óptimas para *bots*.
- Programar *bots* para interactuar de forma automática con contenidos de un tema cualquiera en función de su polaridad.

Un 75% de los profesionales encuestados para la realización de la presente investigación, afirmaron que, en efecto, las herramientas y métodos de monitorización y análisis de la actitud de los usuarios en Internet, pueden ser utilizados para manipular a la población.

META¹²⁷ está llevando a cabo política restrictivas y de control sobre el *scraping*¹²⁸, y posee un equipo de más de 100 personas específicamente dedicadas a la prevención de abusos en este tipo de prácticas, incluyendo científicos, analistas e ingenieros de datos¹²⁹. Esta es una de las razones por las que la comunidad científica se ha “refugiado” en Twitter para la realización de sus investigaciones. La API de Twitter permite realizar, con ayuda de la propia compañía, la recolección masiva de datos, prácticamente sin restricción.

Por último, es necesario hacer una precisión semántica importante al respecto de la influencia y la manipulación. Siendo la manipulación un acto moralmente reprobable, ya que implica una distorsión de la verdad o la justicia¹³⁰, la mera influencia puede adoptar formas lícitas, por ejemplo, en la persecución de fines de utilidad social. Paradójicamente, entre los fines lícitos, podríamos poner como ejemplo la neutralización de las acciones de manipulación, en temas como la difusión de bulos, la desinformación sobre asuntos de salud pública o el discurso de odio.

¹²⁷ Compañía propietaria de Facebook, WhatsApp e Instagram

¹²⁸ Scraping: recolección masiva automatizada de datos publicados en Internet

¹²⁹ <https://about.fb.com/ltam/news/2021/04/como-combatimos-el-scraping/>

¹³⁰ La acepción 3 del Diccionario de la Real Academia de la Lengua, establece que manipular es “intervenir con medios hábiles y, a veces, arteros, en la política, en el mercado, en la información, etc., con distorsión de la verdad o la justicia, y al servicio de intereses particulares.”

4.9 Objetivo específico 9: estado del arte, obstáculos y desafíos de la comunidad investigadora

¿En qué punto se encuentra el estado del arte? ¿cuáles son los principales desafíos a los que se enfrentan los investigadores?

El estudio de los métodos analizados y probados para la realización de esta tesis, en coincidencia con la opinión de diversos autores, nos lleva a una serie de espacios de debate muy concretos, como veremos a continuación. Se trata, en general, de zonas grises donde podemos identificar cuatro tipos de problemas que dividen a la comunidad:

- Limitaciones actuales que impiden obtener una plena satisfacción en los resultados obtenidos.
- Discrepancias en cuanto a las metodologías a aplicar, y falta de consenso en muchas áreas de estudio.
- Problemas concretos aún no resueltos sobre aspectos esenciales del procedimiento.
- Falta de madurez en metodologías y enfoques sobre los que sí parece haber consenso.

Estas vías de debate pintan un panorama característico de especialidades académicas de reciente creación, donde todavía no se han consolidado los procedimientos científicos, las herramientas aún no están suficientemente desarrolladas y, en general, se están poniendo los cimientos a desarrollos futuros. Mucho terreno por delante, y muy poco por detrás.

Como hemos visto a lo largo de la investigación, existen numerosos frentes de estudio, y en cada uno de ellos los investigadores proponen y comparan multitud de aproximaciones y modelos para superar problemas no resueltos (o susceptibles de mejora), ya sean novedosos o derivados de trabajos anteriores. En particular, hemos

comprobado que los problemas que más centran la atención de los investigadores son los siguientes:

- Mejora del rendimiento en la clasificación.
- Identificación de entidades y aspectos.
- Reconocimiento de la ironía y las emociones específicas.
- Tratamiento del impacto de la negación en la polaridad.
- Identificación de actitudes específicas, como el odio o el negacionismo en sus distintas formas.
- Depuración del spam y los contenidos creados por *bots*.
- Geolocalización.

En cuanto al origen y la naturaleza de los textos, un aspecto de “bloqueo” importante es la “dependencia” de Twitter. Como hemos visto, la comunidad investigadora se centra de manera muy predominante en la red de *micro blogging* Twitter, gracias a su potente API, que facilita enormemente el trabajo de extracción y, por supuesto, gracias también a que casi todos los textos son públicos en esta red. Facebook es actualmente un campo poco explorado en comparación¹³¹, lo que sin duda está introduciendo un gran sesgo en las conclusiones obtenidas, sobre todo si tenemos en cuenta la amplia implantación de esta red en todo el mundo hispanohablante (cerca de 22 millones de usuarios en España ¹³², por ejemplo). Actualmente, y tras varios escándalos¹³³, Facebook aplica severas restricciones al *scrapping* de sus

¹³¹ Uno de los estudios realizados para esta tesis se ha hecho sobre contenidos de Facebook precisamente por esta razón, con la intención de no dejar fuera esta red social del apartado empírico (Anexo 3).

¹³² <https://es.statista.com/estadisticas/518719/usuarios-de-facebook-en-espana/>

¹³³ <https://www.bbc.com/mundo/noticias-43472797>

contenidos¹³⁴. Además, la ausencia de otras redes en los trabajos de los investigadores imposibilitan la aplicación de una visión más amplia del análisis de sentimientos, como la propuesta por Yadollahi et al. (2017).

Además de las limitaciones y problemas indicados, de los que son plenamente conscientes los investigadores, hemos detectado al menos un aspecto relevante que no está sujeto a debate, como es el tratamiento científico de los pasos previos a la creación y aplicación de metodologías de análisis. Quizás, al involucrar disciplinas académicas distintas a la Ciencia de Datos, este asunto no ha tenido oportunidad de centrar la atención de los autores, preocupados por otros aspectos más “centrales” desde el punto de vista de su especialización. En concreto, en ninguno de los trabajos analizados se fundamenta objetivamente el marco lingüístico desde el punto de vista del modelo mental (*mental model* o *mental map*) de los usuarios, es decir, no se profundiza en cómo se categorizan y relacionan entre sí los conceptos que son objeto de estudio desde el punto de vista cognitivo (por ejemplo, las entidades y los aspectos de esas entidades, o las “palabras clave” que identifican un dominio). Tampoco se aborda este tipo de análisis desde el punto de vista de un contexto cultural determinado. Como se explica en el desarrollo de la investigación, esta “vuelta de tuerca” del método científico sería posible mediante una aproximación como la propuesta en el Anexo 4 de esta tesis.

En definitiva, la comunidad investigadora solo parece haber alcanzado un consenso en cuanto al marco teórico y los procesos a realizar desde un punto de vista muy general. Por tanto, nos encontramos con una gran diversidad de herramientas y aproximaciones metodológicas a los mismos problemas. En este sentido, hay que

¹³⁴ <https://www.facebook.com/legal/terms>

destacar esfuerzos tempranos, como el de Cuesta et al. (2014), por definir un marco específico.

¿Qué desafíos específicos tenemos por delante para el análisis en lengua española?

Hemos visto cómo la producción de trabajos orientados a textos en español no para de crecer, pero lo hace a un ritmo más lento que la media mundial. En 2021, solo el 1,7% de los trabajos publicados sobre *sentiment analysis* y *opinion mining* estaban referidos a la lengua española de forma específica. También hemos visto como las metodologías adaptadas a nuestro idioma no tienen el peso que deberían si lo comparamos con la importancia de la lengua española por número de hablantes. Esto, sin duda, plantea un desafío, ya que la interpretación del lenguaje natural es un campo fuertemente mediatizado por las especificidades lingüísticas y culturales de los pueblos.

Para ilustrar la importancia de las variantes culturales, más allá de una visión canónica de las lenguas, podemos mencionar lo que algunos autores en Latinoamérica han denominado *Spain-focused bias*, un hándicap importante a la hora de utilizar recursos previamente desarrollados por otros investigadores que requiere estrategias específicas para ser compensado (Cerón-Guzmán & León-Guzmán, 2016).

En cualquier caso, vemos que el progresivo predominio de las técnicas basadas en *machine learning* puede aliviar estos desafíos lingüísticos, al reducir la dependencia de *lexicons* y otros recursos muy orientados a idiomas concretos.

4.9.1 Cauces futuros previsibles

Teniendo en cuenta el estado del arte, los retos pendientes y la opinión de los investigadores, podemos deducir que en los próximos años el foco de la comunidad científica en este campo se dirigirá hacia los siguientes temas:

- Técnicas de aprendizaje profundo, y muy especialmente desarrollos basados en BERT.
- Incremento de la investigación orientada al *Hate Speech* y la desinformación sobre asuntos de salud pública.
- Estandarización del análisis.
- Geolocalización.
- Reducción de la dependencia de Twitter.
- Determinación de la polaridad a nivel de entidad/aspecto.
- Lenguaje figurado, ironía e interpretación del contexto.
- Depuración de *bots* y *spam* en los datos extraídos.
- Aproximación de las tecnologías comerciales de investigación informal a las tecnologías académicas más avanzadas.

4.10 Objetivo específico 10: el impacto de la pandemia

¿Ha influido la pandemia en el estado del arte? ¿Cuál ha sido el impacto, si lo ha habido, en la comunidad investigadora?

Hemos podido comprobar cómo la pandemia ha tenido una influencia determinante en la evolución del estudio de la actitud de los ciudadanos en Internet. Este impacto es fácilmente cuantificable desde varios puntos de vista: número de trabajos publicados con este asunto en el centro de mira, novedosas metodologías *ad hoc* y *datasets* específicos.

No obstante, el Covid-19 se perfila fundamentalmente como un foco temático sobre el que aplicar lo que constituye el verdadero interés de la comunidad investigadora: el desarrollo de nuevas metodologías que hagan avanzar el estado del arte. En otras palabras, como resultado de nuestras investigaciones podemos concluir que el Covid-19, más que una fuente de nuevos avances ha sido un estímulo para investigar más (y dedicar más recursos) al estudio de la actitud de los ciudadanos en Internet, abriendo un nuevo espacio de atención.

Otro aspecto que podemos deducir de la investigación realizada es cómo la pandemia, al igual que el discurso de odio, ha reforzado la practicidad de los métodos, en el sentido de conectar más a la comunidad científica con objetivos de utilidad pública directos. En este sentido, HaterNet (Pereira-Kohatsu et al., 2019) es un claro ejemplo de herramienta al servicio directo de instituciones públicas, más allá del marco puramente teórico-científico.

4.11 Validación de la hipótesis

Como consecuencia de lo anteriormente expuesto en este capítulo final, podemos afirmar que:

- Cualquier persona con conocimientos básicos de comunicación digital o de investigación de marketing puede llegar a conocer la actitud de los ciudadanos en Internet, usando aplicaciones comerciales, sin requerir para ello de recursos informáticos o económicos distintos a los que emplea en su trabajo cotidiano, pero con un nivel de precisión y exactitud inferior al que pueden alcanzar los investigadores académicos, y sin posibilidad de acceso al código fuente. Esta capacidad de análisis se puede aplicar a períodos de tiempo específicos, o adoptar la forma de una monitorización constante.
- Cualquier investigador con conocimientos avanzados en procesamiento de lenguaje natural puede desarrollar metodologías *ad hoc* para proyectos específicos en el estudio de la actitud de los ciudadanos en Internet, o

aplicar/perfeccionar métodos creados por otros colegas investigadores, sin requerir para ello de recursos tecnológicos distintos a los disponibles en cualquier organización académica. El estado del arte actual no permite asegurar un nivel óptimo de precisión y exactitud en los resultados, y no hay consenso al respecto de las metodologías a aplicar.

- Aunque las herramientas disponibles para la comunidad, tanto científicas como informales, han sido desarrolladas con fines exclusivamente analíticos, es posible en la práctica utilizar la información que producen, así como algunas de sus funcionalidades, para influir o manipular a la población, mediante estrategias de comunicación en redes sociales, lícitas o ilícitas.

Estas afirmaciones validan la hipótesis planteada, pero establecen un límite claro en cuanto a su alcance práctico en la actualidad. Podemos concluir que, en efecto, existen metodologías y herramientas que hacen posible que cualquier persona u organización recopile cualquier información públicamente visible y la procese para conocer la opinión de los usuarios sobre cualquier dominio o materia específica, con el fin de influir o manipular. No obstante, estas metodologías y herramientas están lejos de alcanzar un nivel óptimo de rendimiento (especialmente las “informales”), y tienen como único resultado primario el conocimiento, es decir, no pueden ser usadas de manera *directa* para manipular o influir, sino como paso previo para alimentar estrategias de influencia y manipulación mediante acciones posteriores.

4.12 Recapitulación final

Como resultado de las investigaciones realizadas, hemos podido concluir que es posible identificar, extraer y procesar masivamente opiniones concretas en Internet, sobre temas específicos de cualquier naturaleza, siempre que sean de acceso público. En particular, la tesis demuestra que, usando herramientas comerciales al alcance de cualquier persona con unos mínimos conocimientos de comunicación digital, en algunos casos de acceso gratuito, podría conocerse la polaridad de los sentimientos de las cuentas de una red social sobre cualquier asunto.

Hemos podido comprobar mediante una investigación ad hoc para esta tesis, cómo una herramienta “informal” puede aproximarse en rendimiento a la mayoría de los métodos más comunes en la literatura científica hasta fecha reciente, pero quedándose aún muy lejos de las metodologías más avanzadas y de los referentes más actuales en benchmarking.

También hemos podido determinar que no existe posibilidad de manipular directamente a los ciudadanos con las herramientas y métodos del arte, pero sí de manera indirecta, utilizando la información obtenida para alimentar otro tipo de aplicaciones y servicios, como por ejemplo tecnologías publicitarias y de *community management*¹³⁵. A este respecto, sabemos a través de la investigación referida en el Anexo 5 que es posible crear de forma automatizada listas de usuarios (en algunos casos identificables) que tienen en común una posición o afinidad determinada con respecto a un asunto o área de interés, y monitorizar su actividad aplicando analíticas propias de la ciencia de datos. Además, es posible “perfilar” a estas

¹³⁵ *Community management* es el término que se refiere al conjunto de tareas y responsabilidades de los profesionales encargados de gestionar la interacción de los usuarios con las empresas en las redes sociales.

personas utilizando información como los intereses comunes, datos demográficos, o los términos con los que se definen a sí mismas, entre muchas otras variables. Tecnologías de última generación desarrolladas por IBM, permiten incluso realizar perfilados psicológicos profundos para conocer las inclinaciones, necesidades y valores de los usuarios de un determinado segmento que esté siendo objeto de análisis. Posteriormente, toda esta información procesada puede ser utilizada al servicio de causas legítimas, como la concienciación, el desmentido de falsas informaciones y la investigación académica de los fenómenos sociales, o ilegítimas, como la manipulación.

Todo lo anterior, puede hacerse con herramientas de uso comercial al alcance de cualquier persona, es decir, no hablamos de sofisticadas metodologías de la comunidad académica, sino aplicaciones de fácil acceso. Sobre esta base, las metodologías IA más avanzadas a disposición de los investigadores permitirían elevar considerablemente el nivel de precisión y exactitud, al tener un mejor rendimiento y una mayor posibilidad de adecuarse en proyectos ad hoc.

Finalmente, otro hecho comprobado es el largo camino que tiene por delante el estado del arte, que aún se encuentra muy lejos de alcanzar márgenes de error inferiores al 5%, incluso con los métodos más avanzados (como los derivados de la prometedora tecnología BERT - Google). Las herramientas y métodos analizados son capaces de realizar tareas de monitorización y análisis de gran alcance, pero los valores F-Score indican que los resultados prácticos que podemos esperar, de momento, son limitados. Parte del estudio realizado, se centra en analizar estas limitaciones, y no solo su reflejo en las métricas.

5 Referencias

- Abbasi, A., Hassan, A., & Dhar, M. (2014). Benchmarking twitter sentiment analysis tools. *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014, March 2015*, 823–829.
- Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hai, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: A surveillance study. *Journal of Medical Internet Research*, 22(4). <https://doi.org/10.2196/19016>
- Agulló, F., Guillén, A., Gutiérrez, Y., & Martínez-Barco, P. (2015). ElectionMap: Una representación geolocalizada de intenciones de voto hacia partidos políticos sobre la base de comentarios de usuarios de Twitter. *Procesamiento de Lenguaje Natural*, 55, 195–198.
- Ahmed, K., Tazi, N. El, & Hossny, A. H. (2016). Sentiment Analysis over Social Networks: An Overview. *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, 2174–2179. <https://doi.org/10.1109/SMC.2015.380>
- Alamoodi, A. H., Zaidan, B. B., Zaidan, A. A., Albahri, O. S., Mohammed, K. I., Malik, R. Q., Almahdi, E. M., Chyad, M. A., Tareq, Z., Albahri, A. S., Hameed, H., & Alaa, M. (2021). Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Systems with Applications*, 167. <https://doi.org/10.1016/j.eswa.2020.114155>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Almatarneh, S., Gamallo, P., Pena, F. J. R., & Alexeev, A. (2019). Supervised Classifiers to Identify Hate Speech on English and Spanish Tweets. *Lecture Notes in*
-

Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11853 LNCS(November), 23–30.
https://doi.org/10.1007/978-3-030-34058-2_3

Anderson, J. T. L., Bouchacourt, L. M., Sussman, K. L., Bright, L. F., & Wilcox, G. B. (2022). Telehealth adoption during the COVID-19 pandemic: A social media textual and network analysis. *Digital Health*, 8.
<https://doi.org/10.1177/20552076221090041>

Ansari, M. Z., Aziz, M. B., Siddiqui, M. O., Mehra, H., & Singh, K. P. (2020). Analysis of Political Sentiment Orientations on Twitter. *Procedia Computer Science*, 167, 1821–1828. <https://doi.org/10.1016/j.procs.2020.03.201>

Arcila Calderón, C., Blanco-Herrero, D., & Valdez Apolo, M. B. (2020). Rechazo y discurso de odio en Twitter: análisis de contenido de los tuits sobre migrantes y refugiados en español / Rejection and Hate Speech in Twitter: Content Analysis of Tweets about Migrants and Refugees in Spanish. *Revista Española de Investigaciones Sociológicas*, December, 21–40.
<https://doi.org/10.5477/cis/reis.172.21>

Arif, M. H., Li, J., Iqbal, M., & Liu, K. (2018). Sentiment analysis and spam detection in short informal text using learning classifier systems. *Soft Computing*, 22(21), 7281–7291. <https://doi.org/10.1007/s00500-017-2729-x>

Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. *Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010*, 1, 492–499. <https://doi.org/10.1109/WI-IAT.2010.63>

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 2200–2204.

insights from social media monitoring data collected in Poland. *Health Informatics Journal*, 26(4), 3056–3065.
<https://doi.org/10.1177/1460458220962652>

Cabezas, J., Moctezuma, D., Fernández-Isabel, A., & de Diego, I. M. (2021). Detecting emotional evolution on twitter during the COVID-19 pandemic using text analysis. *International Journal of Environmental Research and Public Health*, 18(13). <https://doi.org/10.3390/ijerph18136981>

Cambria, E., Olsher, D., & Rajagopal, D. (2014). SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. *Proceedings of the National Conference on Artificial Intelligence*, 2, 1515–1521.
<https://doi.org/10.1609/aaai.v28i1.8928>

Cerón-Guzmán, J. A., & León-Guzmán, E. (2016). A sentiment analysis system of Spanish tweets and its application in Colombia 2014 presidential election. *Proceedings - 2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016*, 250–257.
<https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.47>

Chen, D. A., Hoople, G. D., Ledwith, N., Burlingame, E., Bush, S. D., & Scott, G. E. (2020). Exploring faculty and student frameworks for engineering knowledge using an online card sorting platform. *International Journal of Engineering Pedagogy*, 10(1), 62–81. <https://doi.org/10.3991/ijep.v10i1.11336>

Clavel, C., & Callejas, Z. (2016). Sentiment Analysis: From Opinion Mining to Human-Agent Interaction. In *IEEE Transactions on Affective Computing* (Vol. 7, Issue 1, pp. 74–93). Institute of Electrical and Electronics Engineers Inc.
<https://doi.org/10.1109/TAFFC.2015.2444846>

Criado, J. I., & Villodre, J. (2018). Local public sector big data communication on

- social media. A sentiment analysis in Twitter | Comunicando datos masivos del sector público local en redes sociales. Análisis de sentimiento en twitter. *Profesional de La Informacion*, 27(3), 614–623.
- Cuesta, Á., Barrero, D. F., & R-Moreno, M. D. (2014). A framework for massive twitter data extraction and analysis. *Malaysian Journal of Computer Science*, 27(1), 50–67.
- de las Heras-Pedrosa, C., Sánchez-Núñez, P., & Peláez, J. I. (2020). Sentiment analysis and emotion understanding during the COVID-19 pandemic in Spain and its impact on digital ecosystems. *International Journal of Environmental Research and Public Health*, 17(15), 1–22. <https://doi.org/10.3390/ijerph17155542>
- Del, M., Salas-Zárate, P., Paredes-Valverde, M. A., Limon-Romero, J., Tlapa, D., & Baez-Lopez, Y. (2016). Sentiment Classification of Spanish Reviews: An Approach based on Feature Selection and Machine Learning Methods. In *Journal of Universal Computer Science* (Vol. 22, Issue 5). <http://www.internetworldstats.com/stats7.htm>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm), 4171–4186.
- Dolianiti, F. S., Iakovakis, D., Dias, S. B., Hadjileontiadou, S. J., Diniz, J. A., Natsiou, G., Tsitouridou, M., Bamidis, P. D., & Hadjileontiadis, L. J. (2019). Sentiment analysis on educational datasets: a comparative evaluation of commercial tools. *Educational Journal of the University of Patras UNESCO Chair*, 6(1), 262–273.
- Esparza, G. G., de-Luna, A., Zezzatti, A. O., Hernandez, A., Ponce, J., Álvarez, M., Cossio, E., & Nava, J. de J. (2018). A sentiment analysis model to analyze students
-

- reviews of teacher performance using support vector machines. *Advances in Intelligent Systems and Computing*, 620, 157–164. https://doi.org/10.1007/978-3-319-62410-5_19
- Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data*, 2(1). <https://doi.org/10.1186/s40537-015-0015-2>
- Feldman, R. (2013). Techniques and applications for sentiment analysis: The main applications and challenges of one of the hottest research areas in computer science. *Communications of the ACM*, 56(4), 82–89. <https://doi.org/10.1145/2436256.2436274>
- Ferrer-Serrano, M., Latorre-Martínez, M. P., & Lozano-Blasco, R. (2020). Universities and communication: Role of twitter during the beginning of the covid-19 health crisis. *Profesional de La Informacion*, 29(6), 1–18. <https://doi.org/10.3145/epi.2020.nov.12>
- Figgou, L., & Pavlopoulos, V. (2015). Social Psychology: Research Methods. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, 544–552. <https://doi.org/10.1016/B978-0-08-097086-8.24028-2>
- Finn Arup Nielsen. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*, 718, 93–98. <http://arxiv.org/abs/1103.2903>
- Franco-Riquelme, J. N., Bello-Garcia, A., & Ordieres-Meré, J. (2019). Indicator Proposal for Measuring Regional Political Support for the Electoral Process on Twitter: The Case of Spain's 2015 and 2016 General Elections. *IEEE Access*, 7, 62545–62560. <https://doi.org/10.1109/ACCESS.2019.2917398>
- Gamallo, P., Garcia, M., Pineiro, C., Martinez-Castano, R., & Pichel, J. C. (2018). *LinguaKit: A Big Data-Based Multilingual Tool for Linguistic Analysis and*
-

Information Extraction. *2018 5th International Conference on Social Networks Analysis, Management and Security, SNAMS 2018*, 239–244. <https://doi.org/10.1109/SNAMS.2018.8554689>

García-Díaz, J. A., Cánovas-García, M., & Valencia-García, R. (2020). Ontology-driven aspect-based sentiment analysis classification: An infodemiological case study regarding infectious diseases in Latin America. *Future Generation Computer Systems*, *112*, 641–657. <https://doi.org/10.1016/j.future.2020.06.019>

Ghallab, A., Mohsen, A., & Ali, Y. (2020). Arabic Sentiment Analysis: A Systematic Literature Review. *Applied Computational Intelligence and Soft Computing*, *2020*. <https://doi.org/10.1155/2020/7403128>

Giachanou, A., & Crestani, F. (2016a). Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys*, *49*(2). <https://doi.org/10.1145/2938640>

Glaser, B.G., Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine.

Go, A., Bhayani, R., & Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision. *Processing*, 1–6.

Gómez, V., Kaltenbrunner, A., & López, V. (2008). Statistical analysis of the social network and discussion threads in Slashdot. *Proceeding of the 17th International Conference on World Wide Web 2008, WWW'08*, 645–654. <https://doi.org/10.1145/1367497.1367585>

Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R., & Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems*, *117*, 47–58. <https://doi.org/10.1016/j.future.2020.11.022>

- Hakak, S., Khan, W. Z., Bhattacharya, S., Reddy, G. T., & Choo, K. K. R. (2020). Propagation of Fake News on Social Media: Challenges and Opportunities. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12575 LNCS, 345–353. https://doi.org/10.1007/978-3-030-66046-8_28
- Herrera-Planells, J., & Villena-Román, J. (2018). MeaningCloud at TASS 2018: News headlines categorization for brand safety assessment. *CEUR Workshop Proceedings*, 2172, 97–102.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177. <https://doi.org/10.1145/1014052.1014073>
- Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, 216–225.
- Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, 51(8), 56–59. <https://doi.org/10.1109/MC.2018.3191268>
- Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Maks, I., & Izquierdo, R. (2017). Analysis of patient satisfaction in Dutch and Spanish online reviews. *Procesamiento de Lenguaje Natural*, 58, 101–108.
- Jindal, N., & Liu, B. (2006). Identifying comparative sentences in text documents. *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006*, 244–251. <https://doi.org/10.1145/1148170.1148215>
-

- Katsanos, C., Tselios, N., & Avouris, N. (2008). Automated semantic elaboration of web site information architecture. *Interacting with Computers*, 20(6), 535–544. <https://doi.org/10.1016/j.intcom.2008.08.002>
- Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Wortman Vaughan, J. (2020). Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3313831.3376219>
- Kim, H. W., Zheng, J. R., & Gupta, S. (2011). Examining knowledge contribution from the perspective of an online identity in blogging communities. *Computers in Human Behavior*, 27(5), 1760–1770. <https://doi.org/10.1016/j.chb.2011.03.003>
- Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50, 723–762. <https://doi.org/10.1613/jair.4272>
- Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, 40(10), 4065–4074. <https://doi.org/10.1016/j.eswa.2013.01.001>
- Kovaleva, O., Romanov, A., Rogers, A., & Rumshisky, A. (2019). Revealing the dark secrets of Bert. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 4365–4374. <https://doi.org/10.18653/v1/d19-1445>
- Kucharska, W., & Firgolska, A. (2018). Personal brand authenticity and social media: The top 5 football players' case. *Proceedings of the 5th European Conference on Social Media, ECSM 2018*, 118–124.
-

- Levine, S. S., & Prietula, M. J. (2014). Open collaboration for innovation: Principles and performance. *Organization Science*, 25(5), 1414–1433. <https://doi.org/10.1287/orsc.2013.0872>
- Liebrecht, C., Kunneman, F., & van den Bosch, A. (2013). The perfect solution for detecting sarcasm in tweets #not. *WASSA 2013 - 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Proceedings*, 29–37.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–184. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- López, M. B., Alor-Hernández, G., Sánchez-Cervantes, J. L., & Salas-Zárate, M. D. P. (2018). EduRP: An educational resources platform based on opinion mining and semantic web. *Journal of Universal Computer Science*, 24(11), 1515–1535.
- Loukachevitch, N., & Levchik, A. (2016). Creating a general Russian sentiment lexicon. *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, 1171–1176.
- Madden, M., Fox, S., Smith, A., & Vitak, J. (2007). Online identity management and search in the age of transparency. *Pew Internet & American Life Project, December*, 50. http://www.pewinternet.org/PPF/r/229/report_display.asp
- Magami, F., & Digiampietri, L. A. (2020). Automatic detection of depression from text data: A systematic literature review. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3411564.3411603>
- Mahmud, J., Nichols, J., & Drews, C. (2012). Where is this tweet from?: Inferring home locations of Twitter users. *ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, January*, 511–514.
-

- Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. In *Computer Science Review* (Vol. 27, pp. 16–32). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.cosrev.2017.10.002>
- Maynard, D., & Greenwood, M. A. (2014). Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis. *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, 4238–4243.
- Miró Llinares, F. (2016). Taxonomía de la comunicación violenta y el discurso del odio en Internet. *IDP. Revista de Internet, Derecho y Política*, 22. <https://doi.org/10.7238/idp.v0i22.2975>
- Mirończuk, M. M., & Protasiewicz, J. (2018). A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications*, 106(April), 36–54. <https://doi.org/10.1016/j.eswa.2018.03.058>
- Mirzaalian, F., & Halpenny, E. (2019). Social media analytics in hospitality and tourism: A systematic literature review and future trends. *Journal of Hospitality and Tourism Technology*, 10(4), 764–790. <https://doi.org/10.1108/JHTT-08-2018-0078>
- Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLoS ONE*, 8(5). <https://doi.org/10.1371/journal.pone.0064417>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J. J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., ... Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses:

- The PRISMA statement. *PLoS Medicine*, 6(7).
<https://doi.org/10.1371/journal.pmed.1000097>
- Narr, S., Hulfenhaus, M., & Albayrak, S. (2012). Language-independent Twitter sentiment analysis. *Knowledge Discovery and Machine Learning (KDML), LWA*, 12–14.
- Nawaz, A. (2012). A Comparison of Card-sorting Analysis Methods. *Proceedings of the 10th Asia Pacific Conference on Computer-Human Interaction*, 583–592.
<http://openarchive.cbs.dk/handle/10398/8587>
- Nielsen, J. (2004). *Card Sorting: How Many Users to Test*. Nielsen Norman Group.
- Ortega, M. P., Mendoza, L. B., Hormaza, J. M., & Soto, S. V. (2020). Accuracy' measures of sentiment analysis algorithms for spanish corpus generated in peer assessment. *ACM International Conference Proceeding Series*.
<https://doi.org/10.1145/3410352.3410838>
- Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., & Smith, N. A. (2013). Improved part-of-speech tagging for online conversational text with word clusters. *NAACL HLT 2013 - 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Main Conference*, 380–390.
- Padró, L. (2012). Analizadores multilingües en freeling. *Linguamática*, 3, 13–20.
<http://upcommons.upc.edu/handle/2117/14772>
- Paea, S., & Baird, R. (2018). Information Architecture (IA): Using multidimensional scaling (MDS) and k-means clustering algorithm for analysis of card sorting data. *Journal of Usability Studies*, 13(3), 138–157.
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis: Foundations and Trends in Information Retrieval. *Foundations and Trends in Information*
-

Retrieval, 2(1-2), 1-135.
<http://www.nowpublishers.com/article/Details/INR-011>

Pappu, A., Blanco, R., Mehdad, Y., Stent, A., & Thadani, K. (2017). Lightweight multilingual entity extraction and linking. *WSDM 2017 - Proceedings of the 10th ACM International Conference on Web Search and Data Mining*, 365-374.
<https://doi.org/10.1145/3018661.3018724>

Pastor-Galindo, J., Zago, M., Nespoli, P., López Bernal, S., Huertas Celdrán, A., Gil Pérez, M., Ruipérez-Valiente, J. A., Martínez Pérez, G., & Gómez Mármol, F. (2020). Twitter social bots: The 2019 Spanish general election data. *Data in Brief*, 32. <https://doi.org/10.1016/j.dib.2020.106047>

Peddinti, S. T., Ross, K. W., & Cappos, J. (2014). "On the internet, nobody knows you're a dog": A twitter case study of anonymity in social networks. *COSN 2014 - Proceedings of the 2014 ACM Conference on Online Social Networks*, 83-93.
<https://doi.org/10.1145/2660460.2660467>

Peiró, R. (2020). *Fake news (noticias falsas)*. Economipedia.Com.
<https://economipedia.com/definiciones/fake-news-noticias-falsas.html>

Peña, P., Del Hoyo, R., Vea-Murguía, J., Rodrigálvarez, V., Calvo, J. I., & Martín, J. M. (2016). Moriarty: Improving "time to market" in big data and artificial intelligence applications. *International Journal of Design and Nature and Ecodynamics*, 11(3), 230-238. <https://doi.org/10.2495/DNE-V11-N3-230-238>

Peng, H., Cambria, E., & Hussain, A. (2017). A Review of Sentiment Analysis Research in Chinese Language. *Cognitive Computation*, 9(4), 423-435.
<https://doi.org/10.1007/s12559-017-9470-8>

Pennebaker, J. W., & Tausczik, Y. R. (2009). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social*

Psychology, 29, 24–54. <http://jls.sagepub.com/content/29/1/24.short>

Pereira-Kohatsu, J. C., Quijano-Sánchez, L., Liberatore, F., & Camacho-Collados, M. (2019a). Detecting and monitoring hate speech in twitter. *Sensors (Switzerland)*, 19(21). <https://doi.org/10.3390/s19214654>

Pérez Díaz, P. L., Berná Sicilia, C., & Arroyas Langa, E. (2016). The conversation on political issues on Twitter: An analysis of the participation and frames in the debate on the “wert Law” and evictions in Spain. *Obets*, 11(1), 311–330. <https://doi.org/10.14198/OBETS2016.11.1.12>

Peters, M. E., Ammar, W., Bhagavatula, C., & Power, R. (2017). Semi-supervised sequence tagging with bidirectional language models. *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 1, 1756–1765. <https://doi.org/10.18653/v1/P17-1161>

Petrilli, M. J. (2015). What twitter says about the education policy debate: And how scholars might use it as a research tool. *Education Next*, 15(4), 78–79.

Pla, F., & Hurtado, L. F. (2014). Political tendency identification in twitter using sentiment analysis techniques. *COLING 2014 - 25th International Conference on Computational Linguistics, Proceedings of COLING 2014: Technical Papers*, 183–192. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84959920717&partnerID=40&md5=73d61dd37e2a1b4c86dd14ed33bc2d46>

Plaza-Del-Arco, F. M., Molina-Gonzalez, M. D., Urena-Lopez, L. A., & Martin-Valdivia, M. T. (2021). A multi-task learning approach to hate speech detection leveraging sentiment analysis. *IEEE Access*, 9, 112478–112489. <https://doi.org/10.1109/ACCESS.2021.3103697>

Qazi, A., Raj, R. G., Hardaker, G., & Standing, C. (2017). A systematic literature review

- on opinion types and sentiment analysis techniques: Tasks and challenges. In *Internet Research* (Vol. 27, Issue 3, pp. 608–630). Emerald Group Publishing Ltd. <https://doi.org/10.1108/IntR-04-2016-0086>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. *Homology, Homotopy and Applications*. <https://openai.com/blog/language-unsupervised/>
- Ramírez-Tinoco, F. J., Alor-Hernández, G., Sánchez-Cervantes, J. L., Salas-Zárata, M. del P., & Valencia-García, R. (2019). Use of sentiment analysis techniques in healthcare domain. *Studies in Computational Intelligence*, 815, 189–212. https://doi.org/10.1007/978-3-030-06149-4_8
- Restrepo, E. (2016). Etnografía: alcances, técnicas y éticas. *Nómadas*, N. 26, 100. <http://www.ram-wan.net/restrepo/documentos/libro-etnografia.pdf>
- Reyes-Menendez, A., Saura, J. R., & Filipe, F. (2019). The importance of behavioral data to identify online fake reviews for tourism businesses: A systematic review. *PeerJ Computer Science*, 2019(9). <https://doi.org/10.7717/peerj-cs.219>
- Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. (2016). *SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods*. <https://doi.org/10.1140/epjds/s13688-016-0085-1>
- Rodriguez-Ibanez, M., Gimeno-Blanes, F. J., Cuenca-Jimenez, P. M., Soguero-Ruiz, C., & Rojo-Alvarez, J. L. (2021). Sentiment Analysis of Political Tweets from the 2019 Spanish Elections. *IEEE Access*, 9, 101847–101862. <https://doi.org/10.1109/ACCESS.2021.3097492>
- Roe, C., Lowe, M., Williams, B., & Miller, C. (2021). Public perception of SARS-CoV-2 vaccinations on social media: Questionnaire and sentiment analysis. *International Journal of Environmental Research and Public Health*, 18(24).
-

<https://doi.org/10.3390/ijerph182413028>

Rohrer, C. (2014). When to Use Which User-Experience Research Methods. *Nielsen Norman Group*, 1–7. <http://www.nngroup.com/articles/which-ux-research-methods/>

Saad, S., & Saberi, B. (2017). Sentiment Analysis or Opinion Mining: A Review. *International Journal on Advanced Science, Engineering and Information Technology*, 7(5), 1660. <https://doi.org/10.18517/ijaseit.7.4.2137>

Saif, H., Fernandez, M., He, Y., & Alani, H. (2014). On stopwords, filtering and data sparsity for sentiment analysis of twitter. *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, 810–817.

Saif, H., He, Y., & Alani, H. (2012). Semantic sentiment analysis of twitter. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7649 LNCS(PART 1), 508–524. https://doi.org/10.1007/978-3-642-35176-1_32

Schuster, J., Jörgens, H., & Kolleck, N. (2021). The rise of global policy networks in education: analyzing Twitter debates on inclusive education using social network analysis. *Journal of Education Policy*, 36(2), 211–231. <https://doi.org/10.1080/02680939.2019.1664768>

Segura-Bedmar, I., Martínez, P., Revert, R., & Moreno-Schneider, J. (2015). Exploring Spanish health social media for detecting drug effects. *BMC Medical Informatics and Decision Making*, 15(2). <https://doi.org/10.1186/1472-6947-15-S2-S6>

Sherwin, K. (2018). Card Sorting: Uncover Users' Mental Models for Better Information Architecture. *NN Group*. <https://www.nngroup.com/articles/card-sorting-definition/>

Speriosu, M., Sudan, N., Upadhyay, S., & Baldrige, J. (2011). Twitter Polarity

- Classification with Label Propagation over Lexical Links and the Follower Graph. *Workshop on Unsupervised Learning in NLP at the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011 - Proceedings*, 53–63.
- Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39(October 2017), 156–168. <https://doi.org/10.1016/j.ijinfomgt.2017.12.002>
- Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion*, 36, 10–25. <https://doi.org/10.1016/j.inffus.2016.10.004>
- Supovitz, J. (2017). Social media is the new player in the politics of education. *Phi Delta Kappan*, 99(3), 50–55. <https://doi.org/10.1177/0031721717739594>
- Taboada, M. (2016). Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics*, 2(1), 325–347. <https://doi.org/10.1146/annurev-linguistics-011415-040518>
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2), 267–307. http://www.sfu.ca/~mtaboada/docs/Taboada_etal_SO-CAL.pdf
- Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014). Learning Sentiment-Specific Word Embedding. *Acl*, 1555–1565.
- Teo, H. C., Campos-Arceiz, A., Li, B. V., Wu, M., & Lechner, A. M. (2020). Building a green Belt and Road: A systematic review and comparative assessment of the Chinese and English-language literature. *PLoS ONE*, 15(9 September). <https://doi.org/10.1371/journal.pone.0239009>
-

- Thelwall, M. (2017). *The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength*. 119–134. https://doi.org/10.1007/978-3-319-43639-5_7
- Valdez, D., Ten Thij, M., Bathina, K., Rutter, L., & Bollen, J. (2020). Social Media Insights Into US Mental Health During the COVID-19 Pandemic: Longitudinal Analysis of Twitter Data. *Journal of Medical Internet Research*, 22(12), e21418. <https://pubmed.ncbi.nlm.nih.gov/33284783/>
- Valdivia, A., Hrabova, E., Chaturvedi, I., Luzón, M. V., Troiano, L., Cambria, E., & Herrera, F. (2019). Inconsistencies on TripAdvisor reviews: A unified index between users and Sentiment Analysis Methods. *Neurocomputing*, 353, 3–16. <https://doi.org/10.1016/j.neucom.2018.09.096>
- Vapnik, V., & Cortes, C. (1995). Support-Vector Networks. *Machine Learning*, 20, 273–297.
- Vilares, D., Thelwall, M., & Alonso, M. A. (2015). The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets. *Journal of Information Science*, 41(6), 799–813. <https://doi.org/10.1177/0165551515598926>
- Vogel-Heuser, B., Loch, F., Hofer, S., Neumann, E. M., Reinhold, F., Scheuerer, S., Zinn, J., & Reiss, K. (2019). Analyzing students' mental models of technical systems. *IEEE International Conference on Industrial Informatics (INDIN)*, 2019-July, 1119–1125. <https://doi.org/10.1109/INDIN41052.2019.8972071>
- Waseem, Z., & Hovy, D. (2016). Hateful symbols or hateful people? predictive features for hate speech detection on twitter. *HLT-NAACL 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Student Research Workshop*, 88–93. <https://doi.org/10.18653/v1/n16-2013>
-

- Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005). OpinionFinder: A system for subjectivity analysis. *HLT/EMNLP 2005 - Interactive Demonstrations Proceedings*, 34–35.
- Yadollahi, A., Shahraki, A. G., & Zaiane, O. R. (2017). Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys*, 50(2). <https://doi.org/10.1145/3057270>
- Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2011). Combining lexicon-based and learning-based methods for twitter sentiment analysis. *HP Laboratories Technical Report*, 89.
- Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018). The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation. In *ACM Transactions on Management Information Systems* (Vol. 9, Issue 2, pp. 1–29). Association for Computing Machinery. <https://doi.org/10.1145/3185045>

6 Anexos

Anexo 1 – Evaluación del rendimiento de la herramienta SentiOne en análisis de sentimiento, aplicando una metodología naturalista

Introducción

Como parte de los trabajos de investigación realizados para la elaboración de la tesis doctoral “Metodologías de estudio de la actitud de los usuarios a través de su huella digital en Internet”, se ha llevado a cabo una prueba de evaluación del rendimiento de la aplicación SentiOne. Para ello, hemos utilizado una metodología específicamente diseñada a partir del estado del arte, dado que no se han encontrado precedentes similares para este tipo de sistemas “informales” en la literatura científica.

La elección de SentiOne se justifica por ser esta herramienta el SAAS “informal” más citado por la comunidad académica ¹³⁶.

Según Taboada (2016), las aproximaciones metodológicas basadas en *machine learning*, aun siendo apreciadas debido a su precisión, a menudo presentan desventajas, ya que los modelos están entrenados en dominios muy específicos, y no suelen ser transferibles a nuevas fuentes de texto. Abbasi et al. (2014) demostraron hasta qué punto puede variar el rendimiento de una misma herramienta de clasificación en función de la temática analizada en Twitter. Según su estudio, la herramienta SentiStrenght, una de las más citadas en la literatura, mostraba una horquilla de precisión entre 71.31 (dominio TELCO) y 56.35 (dominio retail), lo que

¹³⁶ Prospección realizada en julio de 2022, en la base de datos de SCOPUS, con la doble condición de coincidencia de “SentiOne” y “Sentiment Analysis” en el abstract. Este punto se desarrolló en el apartado “Herramientas comerciales informales” de la Tesis.

supone una diferencia de 15 puntos porcentuales. Por este motivo, cabe suponer que una temática escogida debería arrojar resultados diferentes a otra de naturaleza distinta y con otros referentes contextuales, otro tipo de emociones involucradas, y otro nivel de ironía. De hecho, sabemos que la detección automática de la ironía es un desafío especialmente complejo que plantea numerosos problemas, pero resulta imprescindible superarlo si tenemos en cuenta que puede invertir la polaridad del sentimiento de un mensaje (Pang & Lee, 2008).

Preguntas de investigación

Esencialmente, el propósito de esta prueba es conocer el nivel de precisión (*accuracy*) y el F-Score de la aplicación (RQ1), así como determinar si éstos pueden variar en función de la temática de los textos (RQ2), para lo cual se han empleado dos datasets distintos, por las razones expuestas a continuación.

Respondiendo a las preguntas de investigación planteadas, este estudio pretende llenar un espacio vacío en el estado del arte, reforzando el componente de novedad de la tesis y, sobre todo, despejando una incógnita esencial para una completa validación de la hipótesis de la tesis doctoral.

Contextualización y trabajos relacionados

No hay ningún estudio académico sobre la precisión del análisis de sentimiento de SentiOne. No obstante, la herramienta es mencionada por varios estudios académicos que la han utilizado, en mayor o menor medida, como parte de un estudio. A este respecto, cabe destacar el trabajo realizado por Burzynska et al. (2020) sobre el impacto del Covid-19 en las redes sociales, o el estudio de Kucharska & Firgolska (2018) sobre marca personal.

Tampoco se han encontrado estudios sobre el rendimiento de otras herramientas “informales”, pero sí existen varios trabajos representativos sobre recursos “de la comunidad científica”. Abbasi et al. (2014) realizaron un interesante estudio

comparativo del rendimiento de herramientas de *sentiment analysis* con Twitter. En fecha mucho más reciente, Dolianiti et al. (2019) obtienen peores resultados con un estudio similar realizado con herramientas actualmente en boga, de las cuales solo dos coinciden con las de Abbasi et al.

Otro estudio muy citado es el trabajo de Ribeiro et al. (2016), mucho más exhaustivo que los dos anteriormente citados. Los 24 métodos propuestos por los autores se basan en muchas técnicas de diferentes campos. Algunos de ellos emplean métodos de *machine learning* a menudo basados en enfoques de clasificación supervisada, que requieren datos etiquetados para entrenar clasificadores. Otros son métodos basados en *lexicons* que utilizan listas predefinidas de palabras, en las que cada palabra se asocia a un sentimiento. A fin de construir un ranking suficientemente representativo, Ribeiro et al. realizan pruebas separadas para 2 (Pos., Neg.) y 3 clases (Pos, Neg., Neu.), con varios datasets diferentes.

Metodología

Para el desarrollo de la investigación se han realizado dos estudios separados, con dos *datasets* distintos de literales de Twitter, aplicando en ambos casos el mismo método, en varias fases sucesivas: creación del proyecto en SentiOne, extracción y análisis automático de sentimiento de los textos, clasificación manual de los mismos sobre criterios heurísticos específicos, y finalmente cálculo del F-Score y la precisión (*accuracy*). A continuación, se describen todos los detalles a fin de garantizar la reproductibilidad de la prueba.

Acceso a la herramienta

SentiOne es una herramienta “de pago”. Para esta investigación, fue preciso trabajar con la versión “Enterprise”, a fin de tener acceso total a todas las funcionalidades. La siguiente figura muestra el aspecto del *dashboard* de análisis de un proyecto.



Figura 59 - Vista del dashboard de análisis de un proyecto en SentiOne. En la parte inferior izquierda del tuit mostrado, puede verse la polaridad asignada por la IA del sistema (positiva).

Datasets

Para la realización de la prueba, se han creado dos datasets, uno de tuits relativos a un tema de entretenimiento popular en España, y otro sobre un servicio público.

Dataset “entretenimiento”:

- Palabra clave: “first dates”¹³⁷
- Fuente: Twitter
- Período de extracción: septiembre-noviembre de 2021 (tres meses)

Dataset “servicio público”:

- Palabra clave: “adif”¹³⁸

¹³⁷ First Dates es un popular programa de entretenimiento emitido en Prime Time en España desde 2016, caracterizado por generar mucha actividad en Twitter, con polémica en torno a las actitudes y comportamientos de los concursantes.

¹³⁸ Adif es el acrónimo de “Administrador de Infraestructuras Ferroviarias”. Se trata de una empresa pública española, dependiente del Ministerio de Transportes, encargada del mantenimiento y

- Fuente: Twitter
- Período de extracción: septiembre-noviembre de 2021 (tres meses)

Heurísticas

Los principios heurísticos aplicados para la clasificación humana han sido los siguientes:

- El primer objetivo del clasificador es determinar si el texto expresa una opinión o lleva implícito un sentimiento.
- Un texto que contiene una opinión o sentimiento solo puede ser de polaridad positiva o negativa.
- En caso de que el texto contenga entidades con polaridades distintas, se dará preponderancia a la más fuerte (dado que la herramienta evaluada no puede trabajar a nivel de entidad, es necesario aplicar una heurística coherente con la lógica del sistema).
- No hay polaridad sin opinión/sentimiento. Un texto que no contiene una opinión o sentimiento implícito es necesariamente neutro¹³⁹.

Criterios de evaluación

Para la evaluación de resultados, se han aplicado las estrategias analíticas más habituales en la disciplina académica, resumidas en dos métricas esenciales: F-Score y Accuracy (Giachanou & Crestani, 2016).

explotación de las infraestructuras ferroviarias (más de 15.000 kilómetros de vías cerca de 2000 estaciones).

¹³⁹ Un ejemplo de error típico en personas no entrenadas en trabajo científico de campo es estimar polaridad negativa en un texto que no expresa opinión ni sentimientos del autor, simplemente porque se refiere a una mala noticia - "La línea estuvo cortada de 13:00 a 14:00 por problemas en el tendido eléctrico".

Benchmarking contextual

A fin de poner en contexto los resultados, utilizaremos los resultados publicados por Ribeiro et al. (2016), en concreto su prueba para 3 clases con dataset de Twitter (SemEval), en total consistencia con respecto a nuestro estudio. De forma complementaria, también será interesante comparar las mediciones de *accuracy* de Abbasi et al. (2014) y de Zimbra et al. (2018). Estos dos últimos estudios no parecen adecuados para benchmarking primario debido a que no aportan la métrica F.

En cuanto al trabajo de Dolianiti et al. (2019) es una interesante referencia metodológica, pero no se utilizará en el benchmarking debido a que no se aplica sobre tweets y (probablemente por esta razón), arroja resultados muy inferiores al resto de trabajos.

Prevención de sesgos

Con el fin de prevenir sesgos, se han adoptado varias medidas:

- Se han creado datasets naturalistas¹⁴⁰, es decir, se han procesado la totalidad de los textos existentes en un período determinado de tiempo, sin más filtro que los criterios de selección temática (keyword), depurado de duplicados (retuits) y contenidos no espontáneos (mensajes publicitarios). Se trata de poner a prueba la herramienta reproduciendo un nivel de desafío totalmente equivalente a un escenario de trabajo real.

¹⁴⁰ Naturalistas, en el sentido que se da a la expresión en investigación cualitativa: no se produce una alteración del contexto de la investigación para someter la experimentación a condiciones controladas (de "laboratorio"). En concreto, en este caso no se selecciona un conjunto de textos específicamente seleccionados para cumplir unas ciertas condiciones "óptimas" para el estudio, sino que se emplean "tal cual" todos los recogidos en el trabajo de campo, dentro de un determinado rango de fechas.

- La clasificación manual ha sido realizada por dos personas, una de cuales ha realizado una supervisión de calidad.
- Definición de heurísticas: las heurísticas (destalladas en la metodología), aseguran que la clasificación humana se ha realizado con un criterio homogéneo y reproducible.

Resultados

La tabla 18 muestra las características de los datasets naturalistas extraídos. Como era previsible, la temática de entretenimiento presenta una proporción muy superior de mensajes positivos con respecto a la temática de infraestructuras públicas.

Dataset	Size	Distribution		
		Positive	Negative	Neutral
Infraestructura pública	514	7%	51%	42%
Entretenimiento	795	17%	41%	42%

Tabla 18 – Datasets – Tamaño y distribución por polaridad

En la tabla 19, podemos ver los resultados obtenidos (F-Score) de forma comparativa:

Dataset	F-Score			
	Positive	Negative	Neutral	Avg.
Infraestructura pública	60,87	54,97	69,67	61,84
Entretenimiento	66,01	61,60	68,22	65,28

Tabla 19 – F-Score comparativo de los datasets analizados

La siguiente tabla muestra los resultados para “accuracy” (precisión) de ambos datasets:

Dataset	Accuracy
Infraestructura pública	63,62
Entretenimiento	65,66

Tabla 20 – Precisión comparativa de los datasets analizados

¿Cómo se comparan los resultados obtenidos por SentiOne con respecto a los benchmarks del estado del arte?

Ribeiro et al. (2016), con su metodología de evaluación “Sentibench”, obtuvieron los resultados de la tabla siguiente con los mejores métodos de su ranking, para diferentes datasets:

Dataset	Method	F1-Pos	F1-Neg	F1-Neu	Macro-F1
Comments_BBC	Semantria	36.76	67.94	43.49	49.40
Comments_Digg	Umigon	49.62	62.04	44.27	51.98
Comments_NYT	SO-CAL	56.99	60.08	15.34	44.14
Comments_TED	Opinion Lexicon	64.95	56.59	30.77	50.77
Comments_YTB	LIWC15	73.68	49.72	48.79	57.4
Myspace	LIWC15	78.83	41.74	43.76	54.78
Tweets_DBT	Opinion Lexicon	43.44	47.71	48.84	46.66
Tweets_RND_I	Umigon	60.53	51.39	65.22	59.05
Tweets_RND_III	Umigon	63.33	57.00	82.10	67.47
Tweets_RND_IV	Umigon	75.86	76.33	71.54	74.58
Tweets_SAN	Umigon	44.16	45.95	70.45	53.52
Tweets_Semeval	Umigon	64.28	46.41	73.13	61.27
RW	Sentiment140	62.24	51.17	42.66	52.02

Tabla 21 – Valores obtenidos por Ribeiro et al. con su método “Sentibench” para los mejores métodos evaluados, sobre diferentes datasets.

Como podemos observar, los valores obtenidos por SentiOne están a la altura de los mejores métodos. Si nos fijamos exclusivamente en *datasets* de Twitter, tenemos un promedio F1 de 60,43, inferior al promedio de 63,56 obtenido por SentiOne.

En cuanto a la métrica *accuracy*, los resultados obtenidos con SentiOne son claramente inferiores a los mejores resultados obtenidos por Zimbra et al. (2018) en su sistemático trabajo de *benchmarking*, como podemos ver en la tabla siguiente:

System	Average	Pharma	Retail	Security	Tech	Telco
AiApplied	61.84	69.59	47.99	64.05	60.39	67.20
Anonymous	40.86	33.65	49.93	32.71	43.11	44.89
BPEF	71.38	67.81	65.24	75.32	76.30	72.21
ChatterBox	67.43	75.04	53.19	67.20	69.73	71.99
EWGA	68.12	70.21	60.00	68.50	70.50	71.41
FRFF	70.72	62.86	68.76	73.97	74.90	73.11
FRN	69.17	72.60	59.96	69.98	71.00	72.30
GU-MLT-LT	60.60	45.32	68.21	57.81	60.25	71.41
Intridea	63.31	64.18	47.37	62.63	75.19	67.20
KLUE	62.78	55.60	71.15	54.27	62.25	70.65
LightSIDE	69.35	70.71	58.22	69.86	76.99	70.99
Lymbix	56.63	52.03	54.81	47.60	63.45	65.25
MLAnalyzer	45.20	37.95	52.15	41.35	48.06	46.47
NRC	71.33	75.26	64.93	76.39	64.96	75.08
OpinionFinder	57.66	57.08	52.40	55.01	56.94	66.86
Repustate	43.98	35.80	41.06	31.93	40.90	70.20
RNTN	61.47	66.76	55.25	64.69	55.51	65.14
Semantria	53.50	44.68	56.33	45.46	60.99	60.06
Sentiment140	66.46	62.09	61.77	68.84	67.82	71.79
SentimentAnalyzer	55.15	55.33	51.36	54.83	56.50	57.75
SentiStrength	67.49	74.68	56.35	65.51	69.61	71.31
SVM Baseline	66.86	67.50	59.52	66.02	70.02	71.22
TeamX	67.20	57.60	70.35	62.82	69.10	76.14
Textalytics	66.22	70.33	55.14	66.33	68.29	71.02
TextProcessing	54.06	49.68	50.01	58.40	52.40	59.79
uClassify	47.22	51.70	42.12	47.51	50.31	44.47
ViralHeat	61.16	63.77	48.42	61.94	64.12	67.56
Webis	71.41	76.16	64.40	77.37	63.68	75.46
All General-Purpose Systems	56.76	56.10	51.28	54.46	59.24	62.74
All Domain-Specific Systems	67.53	65.70	63.83	68.08	67.96	72.09

Tabla 22 – *Benchmarking de Zimbra et al. (2018)*, mostrando exclusivamente valores para la métrica *accuracy*

La siguiente tabla muestra los resultados obtenidos por Abbasi et al. (2014) para 15 herramientas stand-alone, sobre la métrica *accuracy*. De nuevo, observamos que en este aspecto SentiOne está por debajo de las mejores herramientas. No obstante, la *accuracy* media de 64,64 colocaría a esta herramienta en el quinto puesto del ranking.

Tool	Average	Pharma	Retail	Security	Tech	Telco
SentiStrength	67.49	74.68	56.35	65.51	69.61	71.31
Chatterbox	67.43	75.04	53.19	67.20	69.73	71.99
Sentiment140	66.46	62.09	61.77	68.84	67.82	71.79
Textalytics	66.22	70.33	55.14	66.33	68.29	71.02
Intridea	63.31	64.18	47.37	62.63	75.19	67.20
AiApplied	61.84	69.59	47.99	64.05	60.39	67.20
ViralHeat	61.16	63.77	48.42	61.94	64.12	67.56
Lymbix	56.63	52.03	54.81	47.60	63.45	65.25
SentimentAnalyzer	55.15	55.33	51.36	54.83	56.50	57.75
TextProcessing	54.06	49.68	50.01	58.40	52.40	59.79
Semantria	53.50	44.68	56.33	45.46	60.99	60.06
uClassify	47.22	51.70	42.12	47.51	50.31	44.47
MLAnalyzer	45.20	37.95	52.15	41.35	48.06	46.47
Repustate	43.98	35.80	41.06	31.93	40.90	70.20
Anonymous	40.86	33.65	49.93	32.71	43.11	44.89

Tabla 23 – Resultados del benchmarking realizado por Abbasi et al. (2014) para 15 herramientas “stand-alone”, sobre la métrica *accuracy*

Conclusiones

Hemos realizado una extracción de textos con dos temáticas proclives a provocar la opinión de los usuarios, las hemos anotado manualmente siguiendo un procedimiento formal, y hemos utilizado esta información para evaluar el rendimiento de la herramienta SentiOne sobre el mismo *dataset*. De manera complementaria, hemos comparado los resultados con los trabajos de

benchmarking publicados por otros autores, referidos a otras herramientas comerciales¹⁴¹.

Los resultados del estudio demuestran que SentiOne, pese a ser una herramienta informal no pensada para uso académico por investigadores, presenta un F-Score y Accuracy alineado con otros recursos más sofisticados (aunque muy por debajo de los recursos de mejor rendimiento).

No obstante, para confirmar totalmente las conclusiones de este estudio, sería recomendable realizar un trabajo subsiguiente donde se comparen directamente diferentes herramientas (incluyendo SentiOne) con los mismos *datasets* (este estudio ha sido realizado con posterioridad, y se expone en el anexo 3). De este modo podrán validarse los resultados de este estudio y, lo que es más importante, ampliarse con la evaluación de otras herramientas, a fin de poder ofrecer una visión más amplia sobre la categoría de recursos sin limitarse solo a uno.

Anexo 2 - Encuesta a profesionales sobre métodos de monitorización de la actitud de los ciudadanos en Internet - Ficha técnica y datos.

Introducción

Con objeto de entender en profundidad la relación entre los usuarios potenciales de las herramientas de monitorización de la actitud de los usuarios en Internet, y ante

¹⁴¹ Sabemos que las nuevas metodologías basadas en BERT y las modernas estrategias de representación bidireccional pueden alcanzar *scores* por encima de 80, pero no estamos comparando metodologías *ad hoc* sino herramientas.

la ausencia de investigaciones similares publicadas, se realizó una encuesta con 142 participantes de diversos sectores profesionales y académicos.

En este documento se describe con detalle la metodología empleada y los resultados obtenidos.

Contextualización

La encuesta se enmarca en las investigaciones realizadas para el desarrollo de la tesis “Metodologías de estudio de la actitud de los usuarios a través de su huella digital en Internet”.

Preguntas de investigación

Al respecto de las herramientas de monitorización y métodos de monitorización de la actitud y la opinión de los usuarios en Internet que sirven para recopilar masivamente textos publicados en redes sociales, blogs, foros, tiendas online, medios de comunicación, etc., y analizarlos de forma automatizada, los encuestados:

- RQ1 (Nivel de conocimiento): ¿Saben que existen? ¿en qué medida depende el grado de conocimiento de su profesión?
- RQ2 (Uso): ¿Las usan? ¿cómo?
- RQ3 (Percepción): ¿Cuál es su actitud al respecto? ¿las consideran útiles? ¿creen que pueden ser usadas para manipular? ¿qué problemas o desafíos piensan que resuelven?

Metodología

Muestra

Se ha trabajado con una muestra aleatoria + cuotas, compuesta por 142 personas. El requisito de elegibilidad individual fue ser un profesional con formación superior, activo en las redes sociales¹⁴². Adicionalmente, para prevenir sesgos y asegurar la diversidad sectorial, se establecieron cuotas mínimas por actividades profesionales.

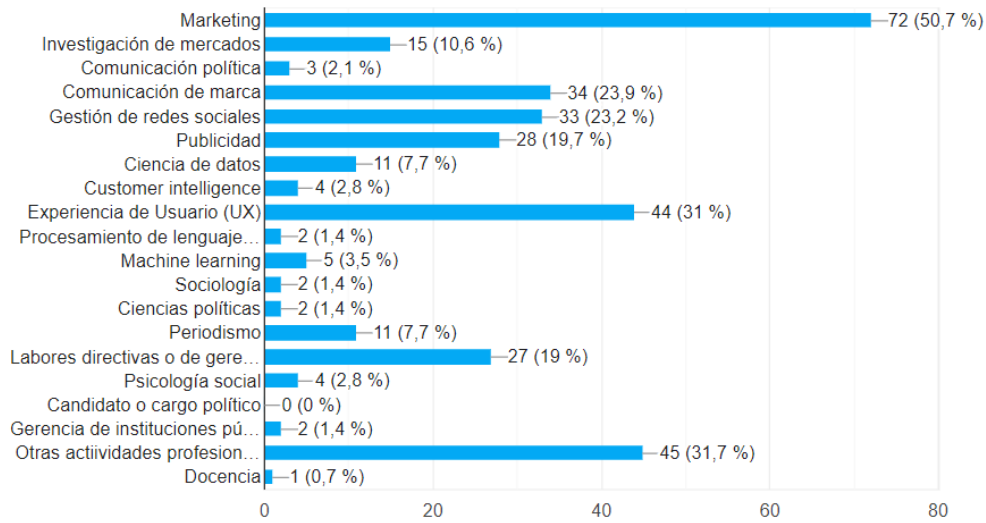
Teniendo en cuenta los medios limitados disponibles, no se consideró necesario trabajar con una muestra amplia, ya que el objetivo del estudio es alcanzar una visión de conjunto de los fenómenos para contextualizar otros aspectos de la investigación.

Resultados

Pregunta 1

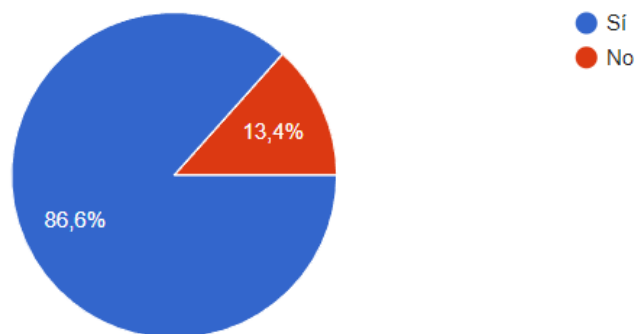
Por favor, marca tus áreas de actividad profesional y académica.

¹⁴² Requisitos alineados con la hipótesis de la tesis doctoral.



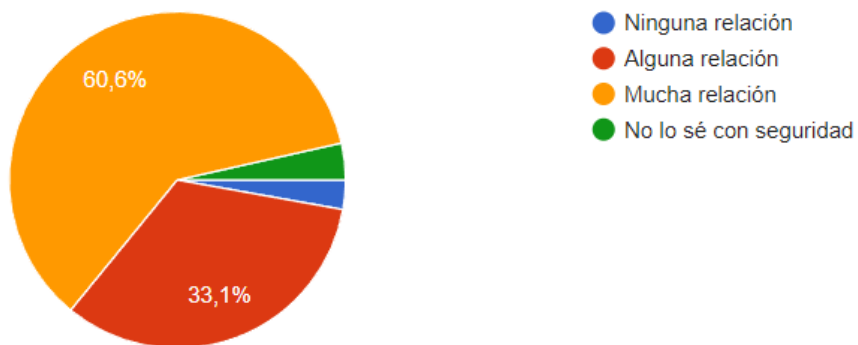
Pregunta 2

Por favor, lee con atención: "Las herramientas y métodos de monitorización de la actitud y la opinión de los usuarios en Internet sirven para recopilar masivamente textos publicados en redes sociales, blogs, foros, tiendas online, medios de comunicación, etc., y analizarlos de forma automatizada. Estas herramientas y métodos hacen posible que empresas, investigadores académicos, organizaciones políticas e instituciones públicas puedan monitorizar la opinión de los ciudadanos en Internet, con distintos fines." ¿Conocías la existencia de estas herramientas?



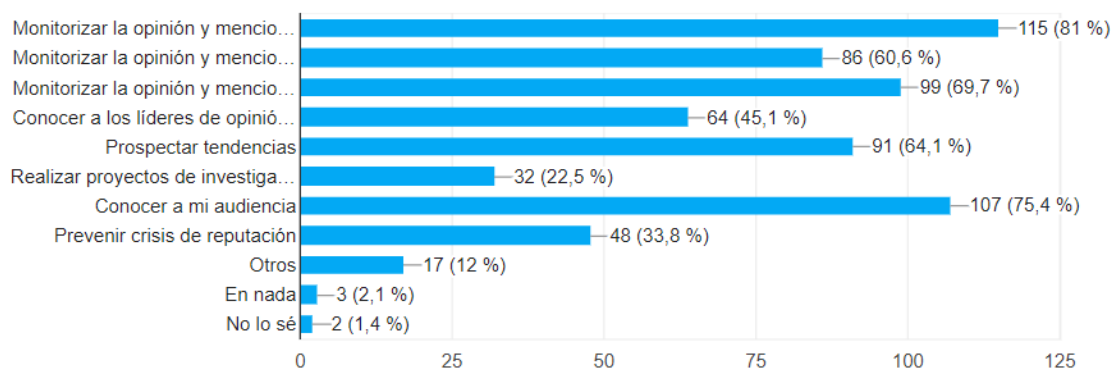
Pregunta 3

¿Crees que este tipo de herramientas tienen relación con tu actividad profesional o académica?



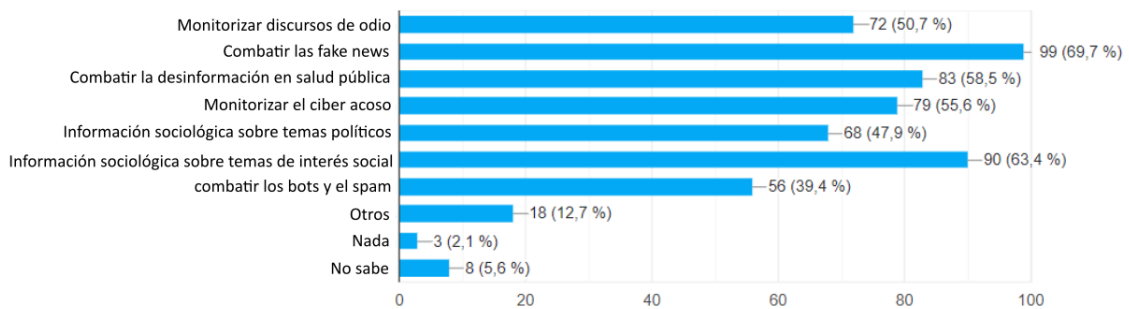
Pregunta 4

¿En qué crees que pueden ser útiles estas herramientas para tus tareas profesionales o académicas? (Marca todas las que procedan)



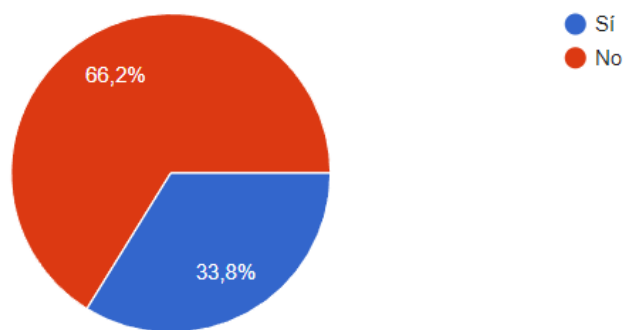
Pregunta 5

¿En qué crees que pueden ser útiles estas herramientas a la sociedad en su conjunto? (Marca todas las que procedan)



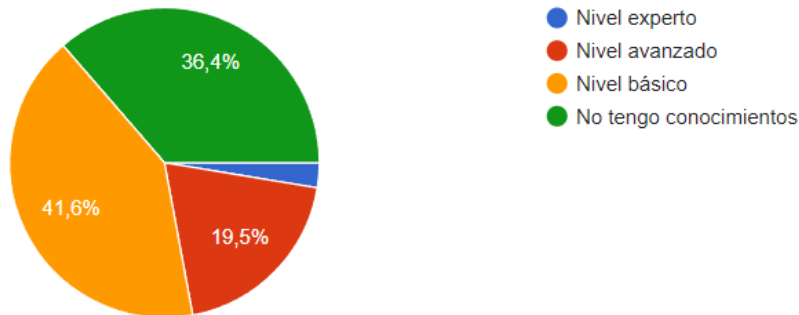
Pregunta 6

¿Alguna vez has trabajado directamente con herramientas de monitorización/análisis para conocer la opinión de los usuarios de Internet? (“directamente” significa que tú personalmente has accedido y trabajado con la herramienta)



Pregunta 7

Si has respondido afirmativamente a la pregunta anterior, ¿cuál es tu nivel de experiencia?

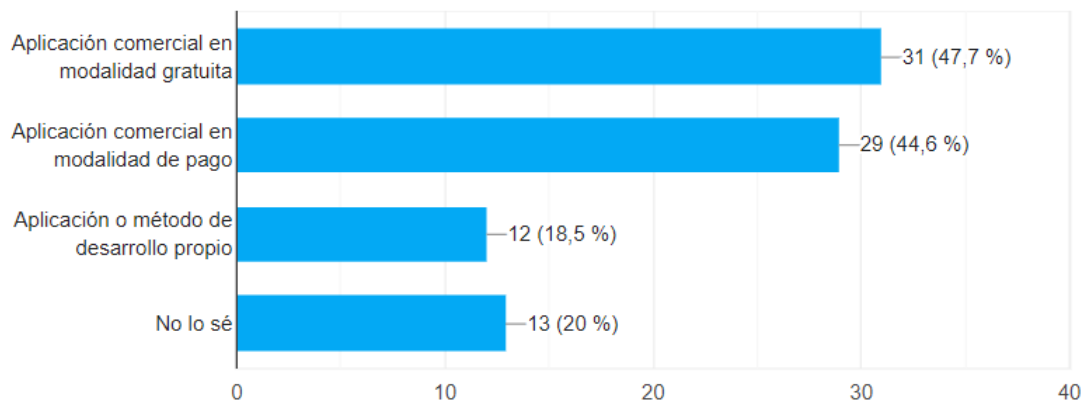


Pregunta 8

Si alguna vez has trabajado directamente con herramientas de monitorización/análisis para conocer la opinión de los usuarios de Internet ¿recuerdas el nombre? (por favor, indica todas las herramientas que recuerdes)

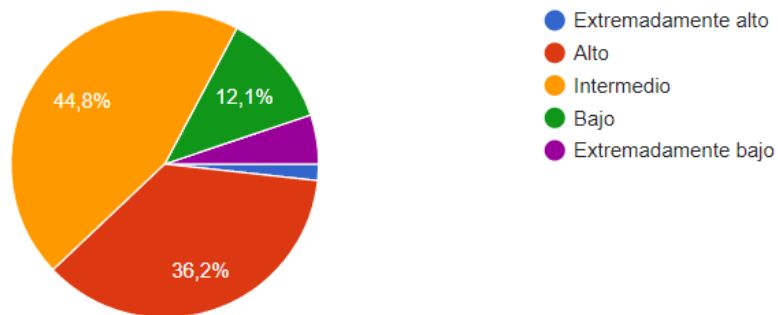
Pregunta 9

Si alguna vez has trabajado directamente con herramientas de monitorización/análisis para conocer la opinión de los usuarios de Internet, ¿qué tipo de aplicaciones eran? (Marca todas las respuestas que procedan)



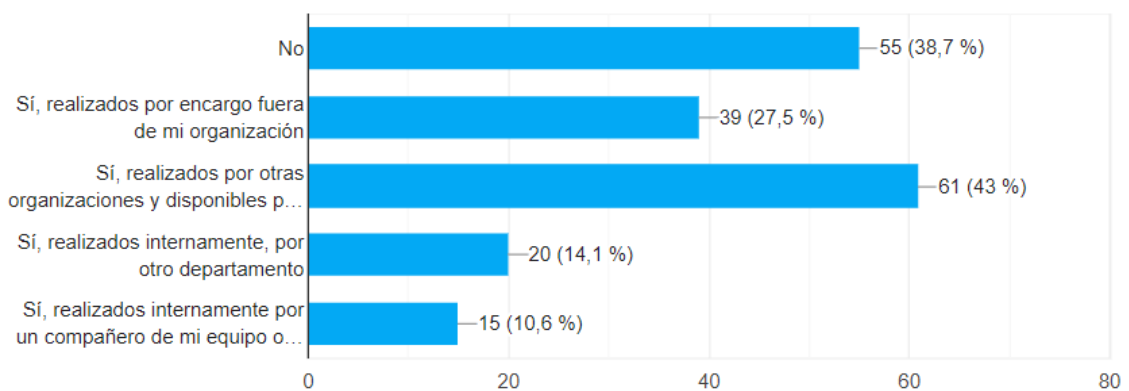
Pregunta 10

Si alguna vez has trabajado directamente con herramientas de monitorización/análisis para conocer la opinión de los usuarios de Internet, ¿cuál es tu nivel general de satisfacción con los resultados?



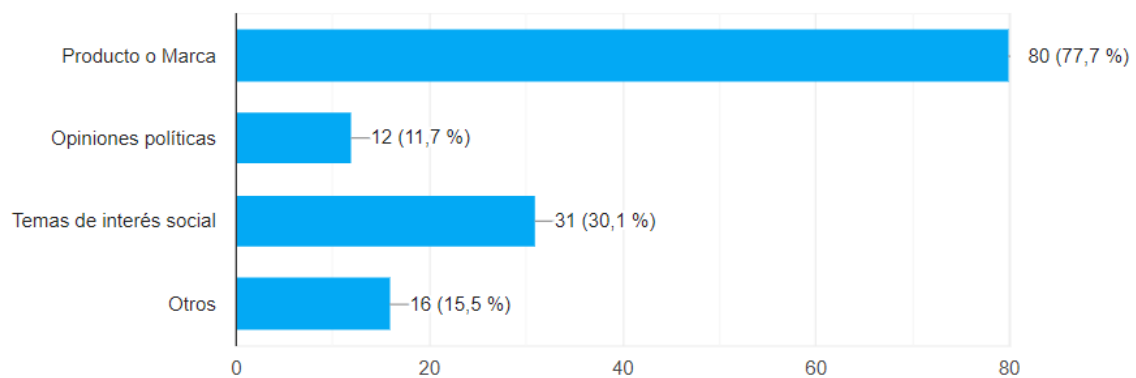
Pregunta 11

¿Alguna vez has accedido a informes realizados por terceros, basados en herramientas de monitorización/análisis para conocer la opinión de los usuarios de Internet? Por ejemplo, una investigación encargada a una agencia, a un subordinado o a otro departamento de tu organización.



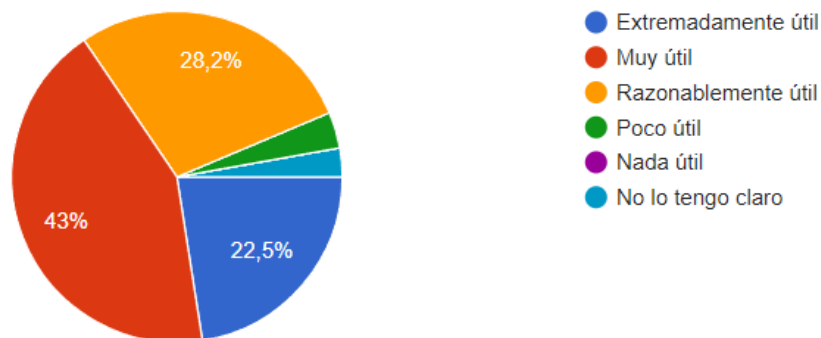
Pregunta 12

Si alguna vez has accedido a información procesada con herramientas de monitorización/análisis de la opinión de los usuarios en Internet, ¿cuál era el objeto de la escucha? (Marca todas las que procedan)



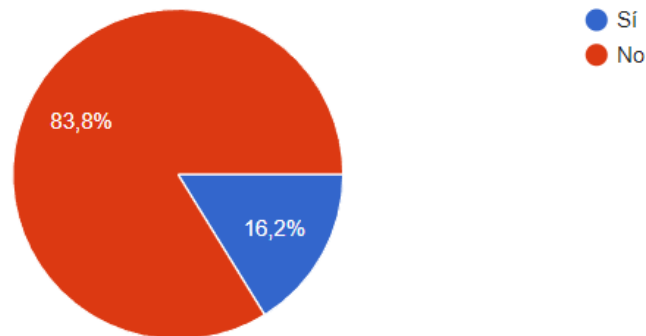
Pregunta 13

Con independencia de si has accedido o no al análisis/monitorización automatizado de las opiniones de los usuarios en Internet como parte de tu actividad profesional o académica, ¿cómo de útil crees que es o podría ser este tipo información para ti?



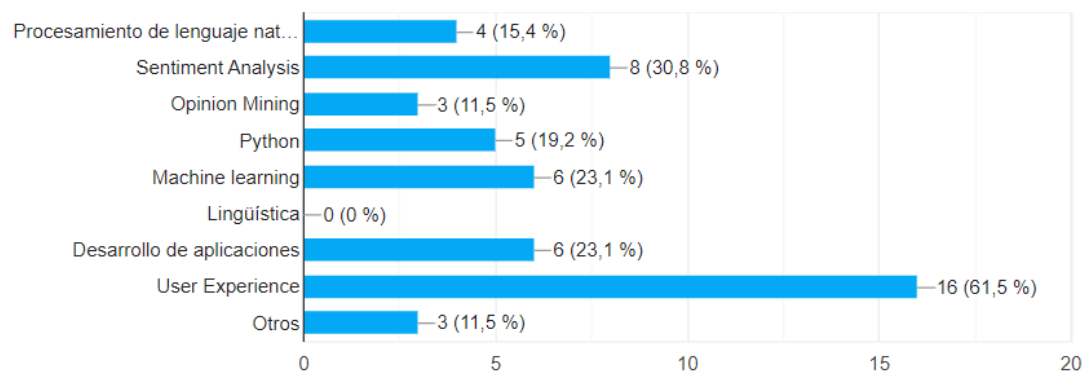
Pregunta 14

¿Alguna vez has participado como experto en la creación y desarrollo de herramientas relacionadas con el estudio de la opinión de los usuarios en Internet?



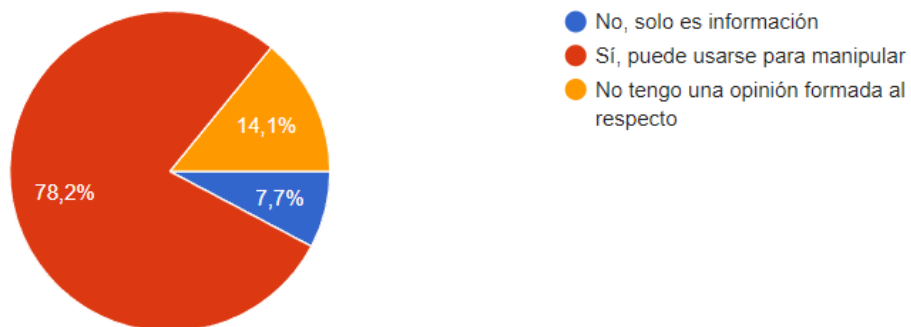
Pregunta 15

Si has respondido afirmativamente a la pregunta anterior, ¿en qué áreas de especialización has participado?



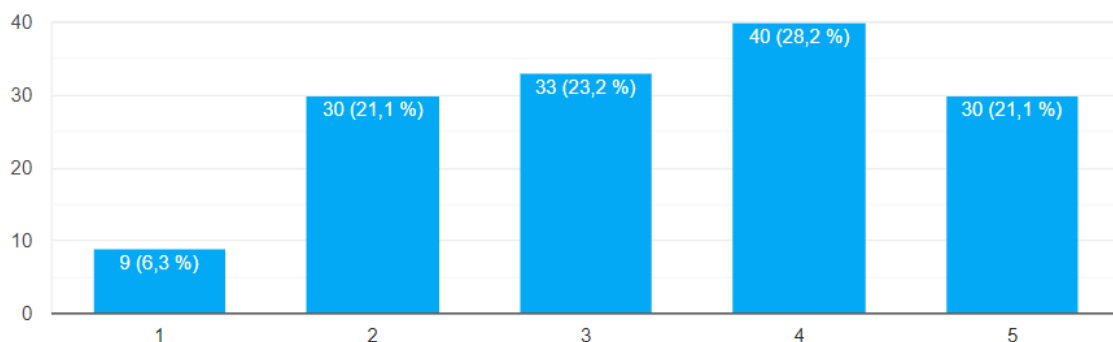
Pregunta 16

¿Crees que la información obtenida con herramientas de monitorización/análisis automatizado de la opinión de los usuarios en internet puede usarse para manipular a las personas? Marca la respuesta que mejor describa lo que piensas.



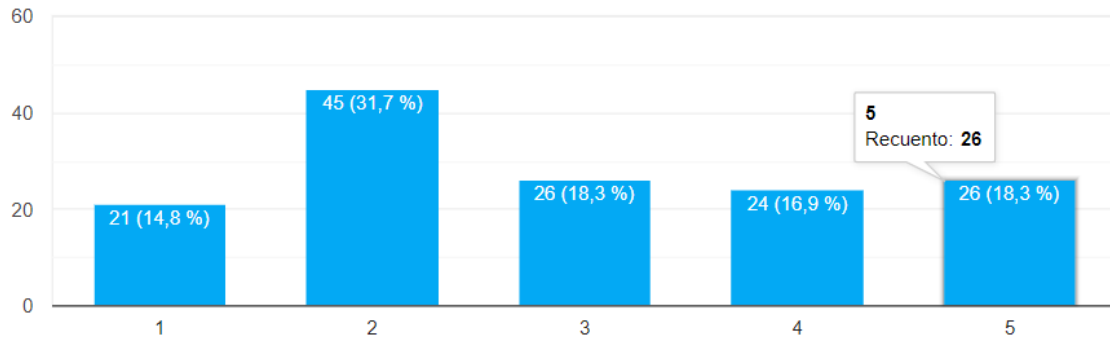
Pregunta 17

Marca del 1 al 5 cómo de acuerdo estarías con la siguiente afirmación "El uso de herramientas de monitorización/análisis automatizado de la opinión de los usuarios en internet compromete la privacidad de las personas"



Pregunta 18

Marca del 1 al 5 cómo de acuerdo estarías con la siguiente afirmación “Como usuario me siento amenazado por la existencia de este tipo de herramientas”



Anexo 3 – Estudio comparativo del rendimiento de herramientas comerciales de *sentiment analysis*

Introducción

Como parte de los trabajos de investigación realizados para la elaboración de la tesis doctoral “Metodologías de estudio de la actitud de los usuarios a través de su huella digital en Internet”, se llevó a cabo un estudio previo de evaluación del rendimiento de la aplicación SentiOne (Anexo 1). Para ese fin, se aplicó una metodología específicamente diseñada a partir del estado del arte, dado que no se habían encontrado precedentes similares para este tipo de sistemas “informales” en la literatura científica. En las conclusiones del estudio, se proponía la realización de un trabajo subsiguiente para comparar directamente diferentes herramientas con los mismos *datasets*. De este modo podrán validarse los resultados del estudio original y ampliarse con la evaluación directa de otras herramientas, a fin de poder ofrecer una visión más amplia sobre la categoría de recursos sin limitarse solo a uno.

Con el propósito expresado, se ha realizado una nueva evaluación comparando directamente entre sí varias herramientas comerciales/informales, con distintos grados de “representatividad” desde el punto de vista de su empleo por parte de la comunidad científica: Sentione, Repustate y MeaningCloud.

Preguntas de investigación

Esencialmente, el propósito de esta prueba es conocer el nivel de *Accuracy*, el *F-Score* y el *Recall* de la aplicación (RQ1), así como determinar si éstos pueden variar en función de la temática de los textos (RQ2), para lo cual se han empleado tres *datasets* distintos, por las razones expuestas a continuación.

Respondiendo a las preguntas de investigación planteadas, este estudio pretende llenar un espacio vacío en el estado del arte, reforzando el componente de novedad

de la tesis y, sobre todo, despejando una incógnita esencial para una completa validación de la hipótesis de la tesis doctoral.

Contextualización y trabajos relacionados

En este apartado haremos un repaso previo para contextualizar el objeto de estudio y el estado del arte.

Herramientas seleccionadas

Las herramientas de *sentiment analysis* utilizadas en este estudio son Repustate, MeaningCloud y SentiOne.

Repustate es una herramienta comercial que aparece en evaluaciones de otros trabajos de *benchmarking*, como las de Abbasi et al. (2014) y Zimbra et al. (2018) (ver también referencias en Anexo 1).

Aunque no aparece citada en las fuentes de benchmarking de referencia ni fue mencionada por ninguno de los participantes en la encuesta a profesionales (Anexo 2), en una revisión bibliográfica realizada *ad hoc* se han encontrado trabajos de investigación apoyados en **MeaningCloud**, como por ejemplo el estudio de Segura-Bedmar et al. (2015) sobre la salud y el consumo de drogas en las redes sociales, o el de Herrera-Planells & Villena-Román (2018) sobre categorización de titulares de noticias.

La elección de **SentiOne** (Anexo 1) se justifica por ser esta herramienta el SAAS “informal” más citado por la comunidad académica ¹⁴³ (además de ser el objeto de la investigación precedente, de la que parte este estudio).

Benchmarking

Abbasi et al. (2014) realizaron en fecha muy temprana un trabajo comparativo del rendimiento de herramientas de *sentiment analysis* con Twitter (54 citas¹⁴⁴).

En fecha mucho más reciente, Dolianiti et al. (2019) presentan un nuevo trabajo de benchmarking con herramientas actualmente en boga, de las cuales solo dos coinciden con las de Abbasi et al.

Otro estudio muy citado (220 citas¹⁴⁵) es el trabajo de Ribeiro et al. (2016) sobre 24 métodos con diversos *datasets* multi disciplinares. A fin de construir un ranking suficientemente representativo, Ribeiro et al. realizan pruebas separadas para 2 (Pos., Neg.) y 3 clases (Pos, Neg., Neu.), con varios datasets de temáticas totalmente diferentes, lo que confiere una especial solidez a sus conclusiones.

El debate sobre la evaluación del rendimiento

Según Taboada (2016), las aproximaciones metodológicas basadas en *machine learning*, aun siendo apreciadas debido a su precisión, a menudo presentan desventajas, ya que los modelos están entrenados en dominios muy específicos, y no suelen ser transferibles a nuevas fuentes de texto. Abbasi et al. (2014) demostraron

¹⁴³ Prospección realizada en julio de 2022, en la base de datos de SCOPUS, con la doble condición de coincidencia de “SentiOne” y “Sentiment Analysis” en el abstract. Este punto se desarrolló en el apartado “Herramientas comerciales informales” de la Tesis.

¹⁴⁴ 54 Citas según SCOPUS, dato actualizado a julio de 2022.

¹⁴⁵ 200 citas según SCOPUS, dato actualizado a julio de 2022.

hasta qué punto puede variar el rendimiento de una misma herramienta de clasificación en función de la temática analizada en Twitter. Según su estudio, la herramienta SentiStrenght, una de las más citadas en la literatura, mostraba una horquilla de precisión entre 71.31 (dominio TELCO) y 56.35 (dominio retail), lo que supone una diferencia de 15 puntos porcentuales. Por este motivo, cabe suponer que una temática escogida debería arrojar resultados diferentes a otra de naturaleza distinta y con otros referentes contextuales, otro tipo de emociones involucradas, y otro nivel de ironía. De hecho, sabemos que la detección automática de la ironía es un desafío especialmente complejo que plantea numerosos problemas, pero resulta imprescindible superarlo si tenemos en cuenta que puede invertir la polaridad del sentimiento de un mensaje (Pang & Lee, 2008).

Metodología

Para el desarrollo de la investigación se han realizado **tres estudios separados**, con **tres datasets distintos**, dos de ellos de Twitter (los mismos empleados en el estudio precedente – Anexo 1), y uno de Facebook. En todos los casos se ha aplicado el mismo método, en varias fases sucesivas: creación del proyecto en la herramienta, extracción y análisis automático de sentimiento de los textos, clasificación manual de los mismos sobre criterios heurísticos específicos, y finalmente cálculo del *F-Score*, *Accuracy* y *Recall*. A continuación, se describen todos los detalles a fin de garantizar la reproductibilidad de la prueba.

Datasets

Para la realización de la prueba, se han creado tres datasets, uno de tuits relativos a un tema de entretenimiento popular en España, otro sobre un servicio público, y por último un tercero sobre el movimiento “antivacunas” del Covid-19.

Dataset “entretenimiento”:

- Palabra clave: “first dates”¹⁴⁶
- Fuente: Twitter
- Período de extracción: septiembre-noviembre de 2021 (tres meses)
- Este dataset ya fue empleado en el estudio previo referenciado en el Anexo 1

Dataset “servicio público”:

- Palabra clave: “adif”¹⁴⁷
- Fuente: Twitter
- Período de extracción: septiembre-noviembre de 2021 (tres meses)
- Este dataset ya fue empleado en el estudio previo referenciado en el Anexo 1

Dataset “anti vacunas”:

- Palabra clave: “yonomevacuno”¹⁴⁸
- Fuente: Facebook
- Período de extracción: marzo 2020 a junio 2022 (es decir, desde el inicio de la pandemia hasta el momento actual)
- Este *dataset* no fue empleado en el estudio previo referenciado en el Anexo 1, y se creó de nueva planta.

¹⁴⁶ #yonomevacuno es uno de los hashtags más populares en el movimiento antivacunación durante la pandemia del Covid-19.

¹⁴⁷ Adif es el acrónimo de “Administrador de Infraestructuras Ferroviarias”. Se trata de una empresa pública española, dependiente del Ministerio de Transportes, encargada del mantenimiento y explotación de las infraestructuras ferroviarias (más de 15.000 kilómetros de vías cerca de 2000 estaciones).

¹⁴⁸ Adif es el acrónimo de “Administrador de Infraestructuras Ferroviarias”. Se trata de una empresa pública española, dependiente del Ministerio de Transportes, encargada del mantenimiento y explotación de las infraestructuras ferroviarias (más de 15.000 kilómetros de vías cerca de 2000 estaciones).

Heurísticas

Los principios heurísticos aplicados para la clasificación humana han sido los siguientes:

- El primer objetivo del clasificador es determinar si el texto expresa una opinión o lleva implícito un sentimiento.
- Un texto que contiene una opinión o sentimiento solo puede ser de polaridad positiva o negativa.
- En caso de que el texto contenga entidades con polaridades distintas, se dará preponderancia a la más fuerte (dado que la herramienta evaluada no puede trabajar a nivel de entidad, es necesario aplicar una heurística coherente con la lógica del sistema).
- No hay polaridad sin opinión/sentimiento. Un texto que no contiene una opinión o sentimiento implícito es necesariamente neutro¹⁴⁹.

Criterios de evaluación

Para la evaluación de resultados, se han aplicado las estrategias analíticas más habituales en la disciplina académica, resumidas en dos métricas esenciales: *F-Score* y *Accuracy* (Giachanou & Crestani, 2016). Además, por consistencia con otros estudios de benchmarking, en esta ocasión se ha añadido el *Recall*.

¹⁴⁹ Un ejemplo de error típico en personas no entrenadas en trabajo científico de campo es estimar polaridad negativa en un texto que no expresa opinión ni sentimientos del autor, simplemente porque se refiere a una mala noticia - "La línea estuvo cortada de 13:00 a 14:00 por problemas en el tendido eléctrico".

Contextual Benchmarking

A fin de poner en contexto los resultados con otros estudios, utilizaremos los resultados publicados por Ribeiro et al. (2016), en concreto su prueba para 3 clases con *dataset* de Twitter (SemEval), en total consistencia con respecto a nuestro estudio. De forma complementaria, también será interesante comparar las mediciones de *accuracy* de Abbasi et al. (2014) y de Zimbra et al. (2018). Estos dos últimos estudios no parecen adecuados para *benchmarking* primario debido a que no aportan la métrica F.

En cuanto al trabajo de Dolianiti et al. (2019) es una referencia metodológica descartada en el estudio anterior (Anexo 1), que ha sido incluida en este para tener una perspectiva más amplia.

Prevención de sesgos

Con el fin de prevenir sesgos, se han adoptado varias medidas:

- Se han creado *datasets* naturalistas¹⁵⁰, es decir, se han procesado la totalidad de los textos existentes en un período determinado de tiempo, sin más filtro que los criterios de selección temática (*keyword*), depurado de duplicados (*retuits*) y contenidos no espontáneos (mensajes publicitarios). Se trata de poner a prueba la herramienta reproduciendo un nivel de desafío totalmente equivalente a un escenario de trabajo real.

¹⁵⁰ Naturalistas, en el sentido que se da a la expresión en investigación cualitativa: no se produce una alteración del contexto de la investigación para someter la experimentación a condiciones controladas (de “laboratorio”). En concreto, en este caso no se selecciona un conjunto de textos específicamente seleccionados para cumplir unas ciertas condiciones “óptimas” para el estudio, sino que se emplean “tal cual” todos los recogidos en el trabajo de campo, dentro de un determinado rango de fechas.

- La clasificación manual ha sido realizada por dos personas, una de cuales ha realizado una supervisión de calidad.
- Definición de heurísticas: las heurísticas (destalladas en la metodología), aseguran que la clasificación humana se ha realizado con un criterio homogéneo y reproducible.

Resultados

La tabla siguiente muestra las características de los *datasets* naturalistas extraídos. Como era previsible, la temática de entretenimiento presenta una proporción muy superior de mensajes positivos con respecto a la temática de infraestructuras públicas, y la de “antivacunas” es abrumadoramente negativa.

Dataset	Size	Class distribution		
		Positive	Negative	Neutral
Infraestructura pública	514	7%	51%	42%
Entretenimiento TV	795	17%	41%	42%
Antivacunas	1980	4%	81%	15%

Tabla 24 - Datasets - Tamaño y distribución por polaridad

F-Score

En las siguientes tablas podemos ver los resultados obtenidos (F-Score) de forma comparativa:

Dataset	F-Score SentiOne			
	Positive	Negative	Neutral	Avg.
Infraestructura pública	60,87	54,97	69,67	61,84
Entretenimiento TV	66,01	61,60	68,22	65,28
Antivacunas	57,87	70,28	43,51	57,22

Tabla 25 – F-Score comparativo de los datasets analizados para SentiOne

Dataset	F-Score - MeaningCloud			
	Positive	Negative	Neutral	Avg.
Infraestructura pública	25,69	61,89	51,46	46,35
Entretenimiento	42,94	22,67	41,88	35,83
Antivacunas	19,58	66,04	40,25	41,96

Tabla 26– F-Score comparativo de los datasets analizados para MeaningCloud

Dataset	F-Score - Repustate			
	Positive	Negative	Neutral	Avg.
Infraestructura pública	27,15	62,68	40,99	43,61
Entretenimiento	44,88	61,25	48,46	51,53
Antivacunas	20,89	73,92	42,07	45,62

Tabla 27 – F-Score comparativo de los datasets analizados para Respustate

Recall

En cuanto al Recall, estos son los resultados obtenidos:

Dataset	Recall SentiOne			
	Positive	Negative	Neutral	Avg.
Infraestructura pública	60,00	39,92	93,06	64,33
Entretenimiento TV	74,26	48,02	79,70	67,33
Antivacunas	79,07	54,42	91,30	74,93

Tabla 28 – Recall comparativo de los datasets analizados para SentiOne.

Dataset	Recall - MeaningCloud			
	Positive	Negative	Neutral	Avg.
Infraestructura pública	80,00	50,95	44,91	58,62
Entretenimiento	77,21	18,84	35,15	43,73
Antivacunas	65,12	50,28	64,55	59,98

Tabla 29 – Recall comparativo de los datasets analizados para MeaningCloud.

Dataset	Recall – Repustate			
	Positive	Negative	Neutral	Avg.
Infraestructura pública	85,71	57,79	30,56	58,02
Entretenimiento	75,74	53,80	40,61	56,71
Antivacunas	62,79	60,94	57,19	60,31

Tabla 30 – Recall comparativo de los datasets analizados para Repustate.

Accuracy

Las siguientes tablas muestran los resultados para “accuracy” (precisión) para cada uno de los *datasets*:

Dataset	Accuracy Sentione
Infraestructura pública	63,62
Entretenimiento	65,66
Antivacunas	61,06

Tabla 31 – Accuracy comparativo de los datasets analizados para SentiOne.

Dataset	Accuracy Repustate
Infraestructura pública	48,25
Entretenimiento	52,08
Antivacunas	60,45

Tabla 32 – Accuracy comparativo de los datasets analizados para Repustate.

Dataset	Accuracy MeaningCloud
Infraestructura pública	50,39
Entretenimiento	35,60
Antivacunas	53,08

Tabla 33 - Accuracy comparativo de los datasets analizados para MeaningCloud.

Contextual benchmarking

¿Cómo se comparan los resultados obtenidos con respecto a los *benchmarks* del estado del arte?

Dolianiti et al. (2019) obtuvieron los siguientes resultados en su trabajo de *benchmarking*:

Dataset	F-Score - Dolianiti et al. (sentence level)			
	Positive	Negative	Neutral	Avg.
IBM Watson	58,10	35,10	58,40	50,53
Microsoft Azure TA	61,70	29,60	61,40	50,90
OpinionFinder 2.0	22,20	11,00	65,40	32,87
Repustate	60,60	32,40	48,40	47,13
Sentistrength	59,20	24,90	50,70	44,93
Avg.	52,36	26,60	56,86	45,27

Tabla 34 – Benchmarking de Dolianiti et al. (2019) para varias herramientas analizadas (F-Score – sentence level)

Ribeiro et al. (2016), con su metodología de evaluación “Sentibench”, obtuvieron los resultados de la tabla siguiente con los mejores métodos de su ranking, para diferentes *datasets*:

Dataset	Method	F1-Pos	F1-Neg	F1-Neu	Macro-F1
Comments_BBC	Semantria	36.76	67.94	43.49	49.40
Comments_Digg	Umigon	49.62	62.04	44.27	51.98
Comments_NYT	SO-CAL	56.99	60.08	15.34	44.14
Comments_TED	Opinion Lexicon	64.95	56.59	30.77	50.77
Comments_YTB	LIWC15	73.68	49.72	48.79	57.4
Myspace	LIWC15	78.83	41.74	43.76	54.78
Tweets_DBT	Opinion Lexicon	43.44	47.71	48.84	46.66
Tweets_RND_I	Umigon	60.53	51.39	65.22	59.05
Tweets_RND_III	Umigon	63.33	57.00	82.10	67.47
Tweets_RND_IV	Umigon	75.86	76.33	71.54	74.58
Tweets_SAN	Umigon	44.16	45.95	70.45	53.52
Tweets_Semeval	Umigon	64.28	46.41	73.13	61.27
RW	Sentiment140	62.24	51.17	42.66	52.02

Tabla 35 – Valores obtenidos por Ribeiro et al. con su método “Sentibench” para los mejores métodos evaluados, sobre diferentes datasets.

En cuanto a la métrica *accuracy*, los resultados obtenidos con SentiOne son claramente inferiores a los mejores resultados obtenidos por Zimbra et al. (2018) en su sistemático trabajo de *benchmarking*, como podemos ver en la tabla siguiente:

System	Average	Pharma	Retail	Security	Tech	Telco
AiApplied	61.84	69.59	47.99	64.05	60.39	67.20
Anonymous	40.86	33.65	49.93	32.71	43.11	44.89
BPEF	71.38	67.81	65.24	75.32	76.30	72.21
ChatterBox	67.43	75.04	53.19	67.20	69.73	71.99
EWGA	68.12	70.21	60.00	68.50	70.50	71.41
FRFF	70.72	62.86	68.76	73.97	74.90	73.11
FRN	69.17	72.60	59.96	69.98	71.00	72.30
GU-MLT-LT	60.60	45.32	68.21	57.81	60.25	71.41
Intridea	63.31	64.18	47.37	62.63	75.19	67.20
KLUE	62.78	55.60	71.15	54.27	62.25	70.65
LightSIDE	69.35	70.71	58.22	69.86	76.99	70.99
Lymbix	56.63	52.03	54.81	47.60	63.45	65.25
MLAnalyzer	45.20	37.95	52.15	41.35	48.06	46.47
NRC	71.33	75.26	64.93	76.39	64.96	75.08
OpinionFinder	57.66	57.08	52.40	55.01	56.94	66.86
Repustate	43.98	35.80	41.06	31.93	40.90	70.20
RNTN	61.47	66.76	55.25	64.69	55.51	65.14
Semantria	53.50	44.68	56.33	45.46	60.99	60.06
Sentiment140	66.46	62.09	61.77	68.84	67.82	71.79
SentimentAnalyzer	55.15	55.33	51.36	54.83	56.50	57.75
SentiStrength	67.49	74.68	56.35	65.51	69.61	71.31
SVM Baseline	66.86	67.50	59.52	66.02	70.02	71.22
TeamX	67.20	57.60	70.35	62.82	69.10	76.14
Textalytics	66.22	70.33	55.14	66.33	68.29	71.02
TextProcessing	54.06	49.68	50.01	58.40	52.40	59.79
uClassify	47.22	51.70	42.12	47.51	50.31	44.47
ViralHeat	61.16	63.77	48.42	61.94	64.12	67.56
Webis	71.41	76.16	64.40	77.37	63.68	75.46
All General-Purpose Systems	56.76	56.10	51.28	54.46	59.24	62.74
All Domain-Specific Systems	67.53	65.70	63.83	68.08	67.96	72.09

Tabla 36 – Benchmarking de Zimbra et al. (2018) para la métrica accuracy

La siguiente tabla muestra los resultados obtenidos por Abbasi et al. (2014) para 15 herramientas stand-alone, sobre la métrica *accuracy*. De nuevo, observamos que en este aspecto SentiOne está por debajo de las mejores herramientas. No obstante, la *accuracy* media de 64,64 colocaría a esta herramienta en el quinto puesto del ranking.

Tool	Average	Pharma	Retail	Security	Tech	Telco
SentiStrength	67.49	74.68	56.35	65.51	69.61	71.31
Chatterbox	67.43	75.04	53.19	67.20	69.73	71.99
Sentiment140	66.46	62.09	61.77	68.84	67.82	71.79
Textalytics	66.22	70.33	55.14	66.33	68.29	71.02
Intridea	63.31	64.18	47.37	62.63	75.19	67.20
AiApplied	61.84	69.59	47.99	64.05	60.39	67.20
ViralHeat	61.16	63.77	48.42	61.94	64.12	67.56
Lymbix	56.63	52.03	54.81	47.60	63.45	65.25
SentimentAnalyzer	55.15	55.33	51.36	54.83	56.50	57.75
TextProcessing	54.06	49.68	50.01	58.40	52.40	59.79
Semantria	53.50	44.68	56.33	45.46	60.99	60.06
uClassify	47.22	51.70	42.12	47.51	50.31	44.47
MLAnalyzer	45.20	37.95	52.15	41.35	48.06	46.47
Repustate	43.98	35.80	41.06	31.93	40.90	70.20
Anonymous	40.86	33.65	49.93	32.71	43.11	44.89

Tabla 37 – Resultados del benchmarking realizado por Abbasi et al. (2014) para 15 herramientas “stand-alone”, sobre la métrica accuracy

Conclusiones

Hemos realizado una extracción de textos con tres temáticas proclives a provocar la opinión de los usuarios, de muy distinta naturaleza. A continuación, las hemos anotado manualmente siguiendo un procedimiento formal, y hemos utilizado esta información para evaluar el rendimiento de las herramientas analizadas sobre cada uno de los *datasets*.

De forma complementaria, hemos comparado los resultados con los trabajos de *benchmarking* publicados por otros autores, referidos a otras herramientas comerciales¹⁵¹.

¹⁵¹ Estamos comparando herramientas, no metodologías. Los nuevos métodos basados en estrategias de representación bidireccional pueden alcanzar *scores* por encima de 80.

Los resultados del estudio demuestran, en concordancia con la investigación precedente (Anexo 1) que SentiOne, pese a ser una herramienta informal no pensada para uso académico por investigadores, presenta un *F-Score* y *accuracy* alineados con otros recursos más sofisticados y “formales”, aunque sin alcanzar los mejores resultados de los *benchmarks* del estado del arte actual.

El F-Score, Accuracy y Recall medios, sumalizando toda la información obtenida en forma de *ranking*, quedarían así:

	Medias		
	F-Score	Accuracy	Recall
SentiOne	61,45	63,45	68,86
Repustate	46,92	53,59	58,35
MeaningCloud	41,38	46,36	54,11

Tabla 38 – Comparativa medias resultados por herramienta

Como se puede comprobar, el resto de las herramientas analizadas en este nuevo estudio arrojan resultados peores, lejos de los estándares actuales, y con una horquilla muy amplia. Por tanto, se hace evidente que los resultados que se pueden conseguir con este tipo de recursos comerciales pueden variar mucho en función del tipo de herramienta y de la naturaleza del *dataset*. En concreto, hemos podido comprobar cómo los resultados mejoran cuanto más acusada es la polaridad negativa, lo cual es consistente con un escenario donde los sentimientos son más intensos (y, por tanto, más susceptibles de ser identificados y procesados correctamente).

Si comparamos los resultados obtenidos por otros autores para las herramientas coincidentes, observamos, por ejemplo, que Repustate obtiene en nuestro estudio un F-Score medio de 46.92, muy cerca del 47.13 obtenido por Dolianiti et al. (2019).

Anexo 4 – Un nuevo método para identificación del mapa mental de los usuarios previa al desarrollo de estudios de minería de opinión y análisis de sentimiento (método presentado en congreso)

Introducción

La comunidad científica ha desarrollado una amplia gama de herramientas "genéricas" para procesar automáticamente las opiniones e identificar su polaridad, pero los investigadores también necesitan fundamentar la categorización de los conceptos por estudio según el "modelo mental" de un segmento de población concreto. En particular, el éxito de los métodos de aprendizaje automático supervisado depende de la elección y extracción del conjunto adecuado de características utilizadas para identificar el objeto objetivo (Almatarneh et al., 2019). Antes de aplicar estas metodologías de clasificación para identificar la polaridad, es necesario identificar correctamente qué textos (o qué entidades "dentro de los textos") corresponden al dominio estudiado: en el caso del dominio educativo, por ejemplo, puede tratarse de un tema específico (el acoso escolar, una ley educativa sometida a debate público o los exámenes escolares) o de espacios de debate más genéricos. En algunos casos, el objetivo del investigador se limita a localizar las opiniones relacionadas con un determinado ámbito de interés para otros fines, sin intención de detectar la polaridad. Para todos estos escenarios de investigación, el investigador necesita conocer en detalle el modelo mental del dominio: cuáles son los conceptos, cómo están jerarquizados y relacionados entre sí en la mente del usuario.

Nuestro propósito en este estudio es prospectar una nueva metodología con el objetivo de superar el desafío descrito, aplicando para ello un análisis basado en *Card Sorting* y *Clustering Jerárquico* (conceptos explicados a continuación).

Esta metodología se propone como un paso preliminar para proporcionar apoyo y fundamentación científica a la toma de decisiones en la aplicación de métodos de *sentiment analysis*.

Dominio de aplicación elegido

A fin de obtener una perspectiva más diversa en el conjunto de investigaciones realizadas en conexión con la tesis, para la puesta a prueba del método se eligió el dominio “educación”, por ser una de las materias más prolíficas en la literatura que aún no habían sido cubiertas (resto de anexos).

Contextualización y trabajos relacionados

Existen muy pocos precedentes en trabajos de investigación que involucren la metodología *Card Sorting* para discernir los modelos mentales de los usuarios en el ámbito educativo (Vogel-Heuser et al., 2019) (Chen et al., 2020), y no están específicamente vinculados a la minería de opinión o a las redes sociales. De hecho, haciendo una consulta de búsqueda en SCOPUS buscando trabajos centrados en Card Sorting y Sentiment Analysis no obtenemos ningún resultado¹⁵². Una consulta similar en Web of Science (Web of Science, n.d.) obtuvo un resultado idéntico.

Metodología

La investigación se llevó a cabo con un protocolo de *Card Sorting* abierto (digital y no moderado), según los estándares del estado del arte, tal y como describe Sherwin (2018). La clasificación de tarjetas o *Card Sorting* es un método de investigación muy popular en UX (Experiencia de Usuario), en el que los participantes en el estudio

¹⁵² Search query: (TITLE-ABS-KEY ("card sorting") AND TITLE-ABS-KEY ("sentiment analysis")).
Última consulta: junio de 2022.

agrupan libremente conceptos de acuerdo con los criterios que tienen sentido para ellos. Este método descubre cómo está estructurado el conocimiento del dominio para el público objetivo (Sherwin, 2018). La metodología de Card Sorting se describe a menudo como una herramienta eficaz que proporciona conocimientos sobre el modelo mental de las personas y que puede ayudar a determinar la mejor arquitectura de información (Rohrer, 2014). La interpretación de los resultados se realiza generalmente mediante la aplicación de algoritmos de agrupación jerárquica.

Para evitar sesgos en la creación del conjunto de temas, se realizaron entrevistas semiestructuradas previas al *Card Sorting*.

La siguiente figura describe el procedimiento completo:

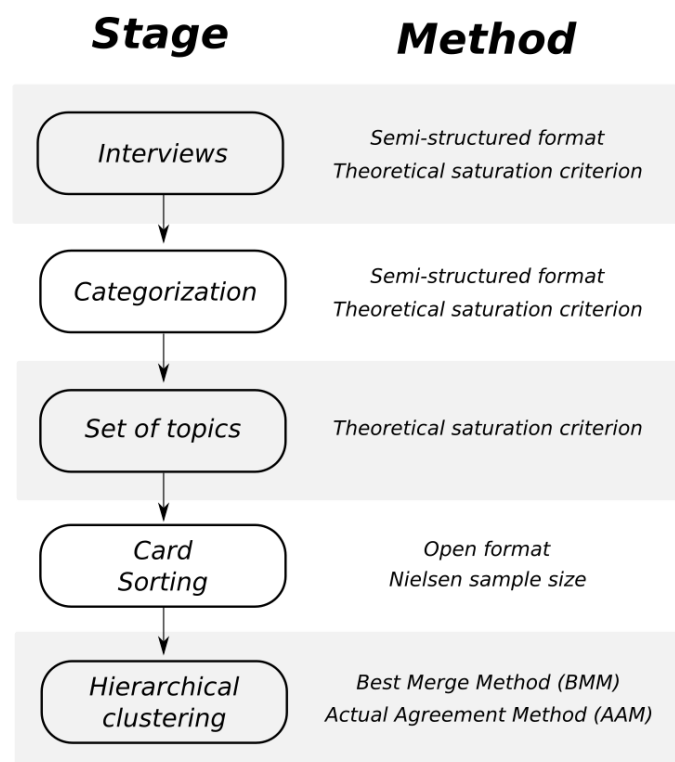


Figura 60 – Fases del procedimiento aplicado

Entrevistas semiestructuradas preparatorias

Para determinar los conceptos que se incluirán en la prueba de clasificación de tarjetas con un fundamento científico previo, se planteó la siguiente pregunta de investigación: ¿Cuáles son los conceptos más relevantes para las personas en cuanto a sus actitudes hacia el sistema educativo?

A fin de responderla se realizaron entrevistas, siguiendo una metodología semiestructurada. La entrevista es probablemente la técnica de recogida de datos cualitativos más frecuente, y el formato semiestructurado es el más utilizado en la investigación cualitativa (Figgou & Pavlopoulos, 2015).

Para la realización de las entrevistas semiestructuradas, establecimos una serie de objetivos y asociamos un tema de conversación a cada uno de ellos. Mediante preguntas abiertas y un diálogo naturalista, el entrevistador profundiza en cada tema hasta extraer los datos necesarios para cumplir cada objetivo.

La muestra se determinó según el criterio de saturación teórica de Glaser y Strauss (Glaser, B.G., Strauss, 1967). Se realizaron en total nueve entrevistas. A partir del 9º participante, ya no se obtuvieron nuevos datos cualitativos diferentes a los recogidos con las entrevistas anteriores, y según los criterios de Glaser y Strauss, se alcanzó el punto de saturación.

Para mantener la coherencia, los criterios de elegibilidad utilizados fueron los mismos que para la prueba de clasificación de tarjetas, pero con diferentes individuos.

Una vez finalizada la ronda de entrevistas, se transcribió el texto. A continuación, se codificaron los resultados, extrayendo todos los temas y eliminando las duplicaciones y redundancias, obteniendo así un total de 62 temas diferentes.

Desarrollo del Card Sorting

Existen tres tipos de *Card Sorting* en la literatura científica, abierto, cerrado e híbrido. El enfoque "cerrado" se recomienda para verificar si un diseño de información es familiar y comprensible para el usuario, mientras que el enfoque "abierto" tiene como objetivo descubrir qué tipo de clasificación de categorías sería más correcto utilizar (Montero et al., 2004). Para esta prueba utilizamos un enfoque "abierto", por ser el más adecuado al objetivo perseguido.

En una clasificación de tarjetas abierta, los usuarios son totalmente libres de agrupar las tarjetas según su criterio, así como de asignar nombres a los grupos creados. En cambio, en la variante "cerrada" los usuarios deben trabajar sobre un conjunto predeterminado de categorías (Sherwin, 2018).

Nielsen (2004) establece que 15 usuarios son suficientes para alcanzar una correlación de 0,90, y que a partir de ese número se producen rendimientos decrecientes que no justifican la inversión de tiempo y recursos. Nuestro estudio se realizó con 16 participantes españoles, mayores de 18 años que cumplían las siguientes condiciones (todas):

- Ser usuarios frecuentes de las redes sociales (acceder al menos a una red social una vez al día)
- Haber expresado sus opiniones sobre el debate educativo recientemente en Twitter, Facebook o LinkedIn

Se utilizaron 62 tarjetas. El método de *clustering* aplicado para el análisis fue *Best Merge* (BMM). Este método es una técnica basada en matrices de similitud y es el estándar de la industria (Nawaz, 2012). A efectos comparativos, también se analizó el resultado de aplicar un algoritmo AAM (método de acuerdo actual) y la matriz de similitud (Paea & Baird, 2018).

Resultados

El número de categorías creadas por los participantes fue reducido, sólo 6,37 de media (sobre el total de tarjetas, esto representa una proporción del 10,27%), lo que indica que los conceptos analizados están muy interrelacionados entre sí en el mapa mental del usuario. Sin embargo, la forma concreta en que se articulan estas interrelaciones fue muy variable de un individuo a otro, en virtud del bajo nivel de consenso que mostró el análisis de agrupación jerárquica (figura siguiente).

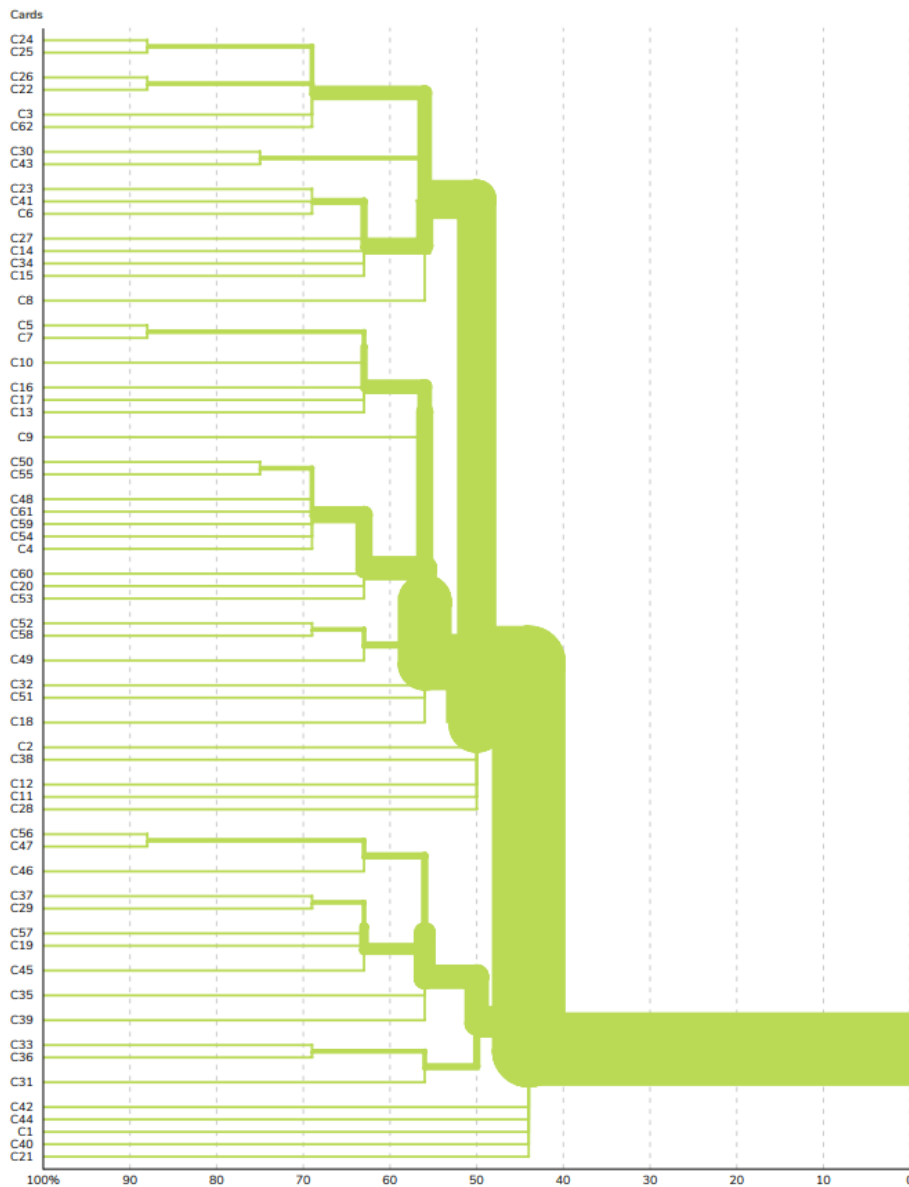


Figura 61- Dendrograma BMM

Si consideramos un umbral del 60% de acuerdo entre los participantes, según Katsanos et al. (2008), obtenemos 10 categorías, pero esto dejaría “huérfanos” (conceptos no asociados a otros) 18 temas (lo que representa un 29% del total).

En cambio, si consideramos un porcentaje de acuerdo menos exigente, fijando el umbral en 50%, seguimos dejando un 16% de temas sin asociar a las categorías.

Conclusiones

Los resultados presentados demuestran la eficacia de las entrevistas semiestructuradas para determinar los conceptos clave de un dominio. Por lo que respecta al Card Sorting, en el caso concreto del dominio estudiado ("educación"), el resultado es demasiado heterogéneo como para poder formar un modelo mental suficientemente consistente y representativo. A este respecto, tanto si utilizamos un algoritmo AAM como BMM en el análisis, la conclusión es la misma. Un posible enfoque más correcto sería "trocear" la extracción de opiniones en diferentes estudios por subcategorías que coincidan con las agrupaciones obtenidas con el método propuesto en este trabajo.

Anexo 5 - Segmentación, *profiling* masivo, y posibilidad de influencia y manipulación de usuarios en Twitter

Introducción

El objeto de esta investigación es comprobar de forma empírica la posibilidad de realizar el siguiente supuesto, sin utilizar herramientas académicas ni realizar labores que requieran conocimientos de ciencia de datos ni procesamiento de lenguaje natural:

Supuesto: identificar a personas que apoyan el movimiento antivacunas del Covid-19 con las siguientes finalidades de utilidad social práctica:

- Impactar a esas personas directamente con contenidos informativos de concienciación específicamente diseñados y con desmentido de bulos y noticias falsas, de forma automatizada.
- Detección temprana de bulos o cualquier forma de desinformación difundidos por esta comunidad.
- Perfilado del tipo de usuario que participa en defensa/difusión de la opinión
- Identificación de líderes de opinión

Este supuesto se integra en el conjunto de comprobaciones empíricas realizadas como parte de la investigación para la tesis en la que se inserta este anexo.

Para realizar esta comprobación empírica, se ha utilizado la herramienta comercial Audiense, ya que cumple los requisitos indicados en el primer párrafo de esta introducción, y aparece como una de las herramientas citadas por los profesionales en la encuesta realizada al respecto (Anexo 2), además ser empleada por profesionales del sector entrevistados directamente para la realización de la tesis.

Contextualización y trabajos relacionados

No se ha encontrado en la literatura ningún trabajo de características similares. La herramienta utilizada, Audiense, es una aplicación muy popular entre los profesionales de la comunicación política y grandes marcas¹⁵³ que incorpora tecnología *state-of-the-art* en procesamiento de lenguaje natural mediante acuerdo

¹⁵³ Grandes clientes: Deloitte, Pepsico, TikTok, Olgilvy, Manchester United (citados por Audiense). Expertos en comunicación política: información obtenida a partir de entrevistas con expertos del sector.

con IBM¹⁵⁴. No obstante, la base de datos de trabajos académicos SCOPUS¹⁵⁵ no refleja ningún *paper* que contenga la herramienta en su título o su *abstract*.

Audiense es una herramienta pensada para Twitter. No se han encontrado en el mercado herramientas comerciales equiparables para otras redes sociales.

Las herramienta Audiense Connect es presentada por la empresa comercializadora como un servicio de “monitorización y escucha avanzadas”, “creación de audiencias personalizadas de Twitter para campañas publicitarias” y “análisis y gestión de comunidades”¹⁵⁶.

Metodología

Para realizar la comprobación empírica descrita, se contrató la herramienta Audiense y se realizó, paso al paso, el proceso de ejecución y comprobación necesario, sin utilizar ningún recurso externo a las funcionalidades aportadas por la aplicación.

La prueba se realizó en 2021, y se repitió en julio de 2022¹⁵⁷ (con datos referidos a 2021) para garantizar el máximo nivel de actualización posible.

En el apartado siguiente (resultados) se detalla el proceso y se incluyen capturas de pantalla de cada fase.

¹⁵⁴ <https://help.audiense.com/es/knowledge/obtener-el-m%C3%A1ximo-provecho-de-los-personality-insights>

¹⁵⁵ Última comprobación realizada en julio de 2022.

¹⁵⁶ <https://es.audiense.com/productos/audiense-connect/>

¹⁵⁷ La constante actualización y mejora de este tipo de herramientas hizo aconsejable no confiar en los resultados obtenidos un año antes.

Resultados

Como primer paso, creamos un *dataset* de monitorización con la herramienta estudiada. En este caso, por consistencia con otras investigaciones realizadas para esta tesis, seleccionamos todos los tuits que contienen el *hashtag* #yomevacuno, en un período de tiempo concreto (semana del 9 al 16 de mayo de 2021). Como muestra la figura siguiente, el interfaz de creación del *dataset* es intuitivo y no requiere de conocimientos técnicos especializados.

Criteria
After entering each keyword, @user, #hashtag and/or URL press return [Advanced editor](#)

@username, #hashtag, example.com...

Monitoring period

From Time Europe/Madrid

Start as soon as possible

To Time

Run as long as interactions last

Tweets

1 084 / 1 355

Maximum number of interactions you want to use from your prepaid credit. i.e. If you have prepaid for 5,000 interactions, you can limit the interactions you spend on this Monitoring

Sample of Tweets

1% 100% 100%

Percentage of total Tweets to be monitored during the selected period

Figura 62 – Interfaz de creación del dataset en Audiense

Como resultado, obtenemos 5.316 tuits, de los que sería posible extraer una muestra aleatoria (opción que no es interesante a los efectos de nuestro estudio, ya que queremos identificar la mayor cantidad posible de usuarios).

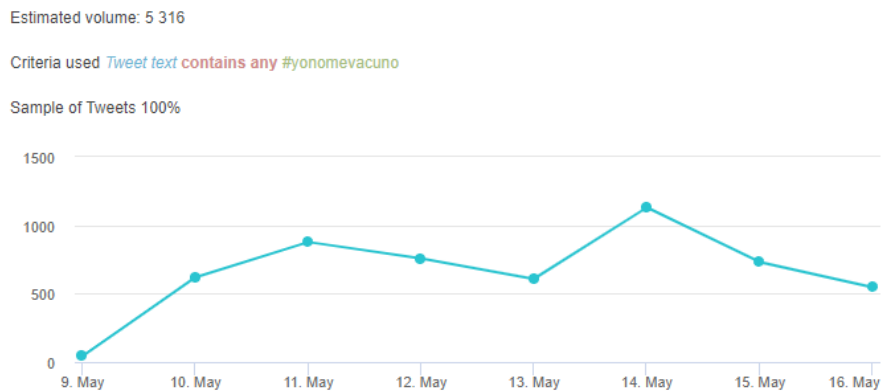


Figura 63 – Evolución de los tuits con el hashtag “yonomevacuno” en la horquilla temporal elegida, para un total de 5316 mensajes.

La herramienta no tiene limitaciones en cuanto al rango y permite realizar el proceso, por ejemplo, con la totalidad de los tuits publicados que cumplan una serie de criterios, como se muestra en la figura siguiente para el ejemplo “usuarios que tuiteado sobre “energía nuclear” o “centrales nucleares” en todo el año 2021 (más de 400.000 mensajes).

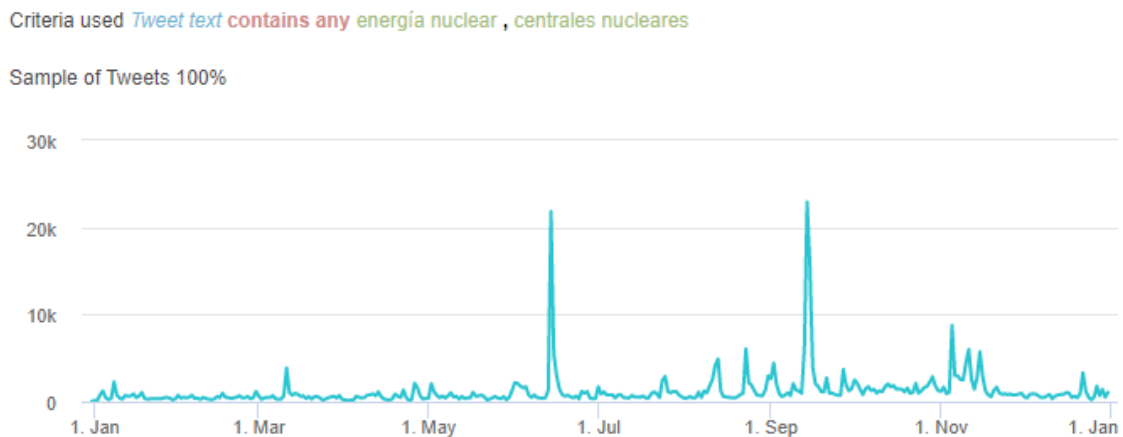


Figura 64 – A efectos comparativos, se muestra la evolución de las menciones a energía nuclear y centrales nucleares en todo el año 2021

El siguiente paso sería crear una “audiencia” a partir de la monitorización realizada en el paso anterior. Como se aprecia en la siguiente figura, para este fin sería posible establecer condiciones AND y OR incluyendo otros criterios, como por ejemplo demográficos o intereses. En nuestro caso, se ha limitado el estudio a usuarios en España, excluyendo aquellos con bajo nivel de actividad (menos de un tuit al día de media). Un detalle destacable es la integración de la herramienta con el servicio de procesamiento de lenguaje natural de IBM¹⁵⁸.

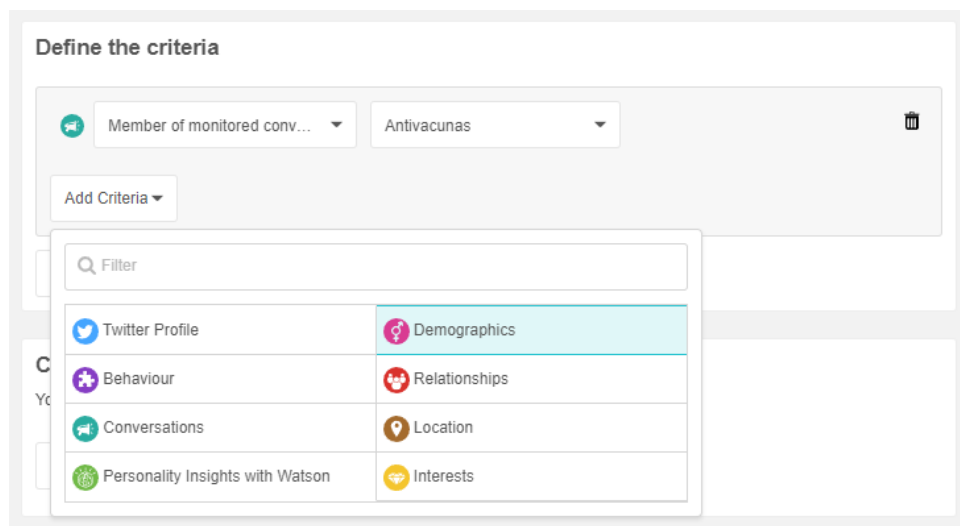


Figura 65 – Criterios de filtro para la definición de audiencias en Audiense.

Si se desea un mayor nivel de precisión en la definición de la audiencia a monitorizar, por ejemplo, con el propósito de crear contenidos más específicos para el segmento,

¹⁵⁸ IBM Watson es una herramienta de “comprensión del lenguaje natural” que tiene funcionalidades de *sentiment analysis*. Está basada en API es y capaz de funcionar con diversos idiomas, en constante ampliación. Esta aplicación forma parte del proyecto DeepQA de IBM, liderado por David Ferrucci. El nombre de “Watson” es un homenaje a al fundador y primer presidente de IBM, Thomas J. Watson

es posible afinar la audiencia por intereses. En la siguiente figura, se añade por ejemplo el interés “medicina alternativa”.

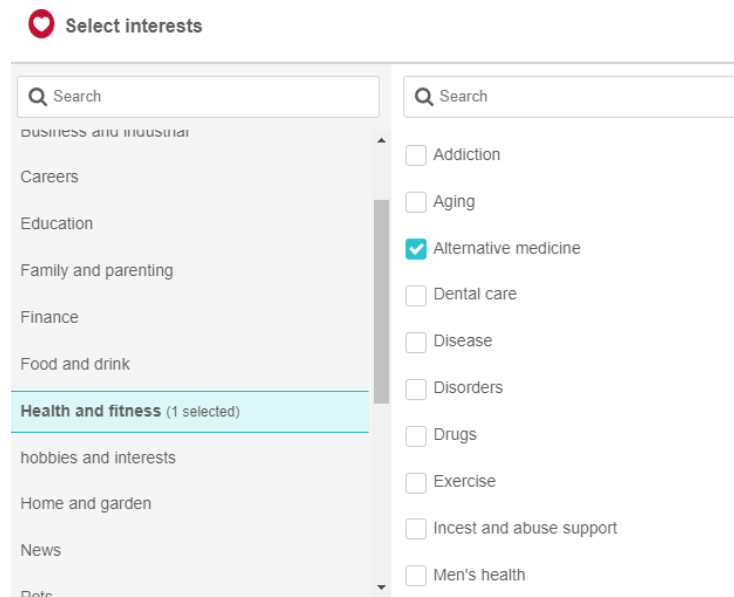


Figura 66 – Afinado de audiencias por intereses en Audiense

La nube de palabras clave refleja con precisión cómo se definen a sí mismas las personas monitorizadas en las *bios* de sus cuentas, lo que constituye un perfil colectivo especialmente valioso desde el punto de vista de la ciencia de datos.

Tag Cloud for Antivacunas



Figura 67 – Perfilado de audiencias por intereses comunes, extraídos de las biografías

Podemos conocer el número de seguidores que tienen las cuentas monitorizadas, además de su frecuencia de publicación, y otros datos que permitirían identificar fácilmente líderes de opinión. En las dos figuras siguientes, podemos ver la gráfica de *followers* y el detalle del subsegmento de más alcance.

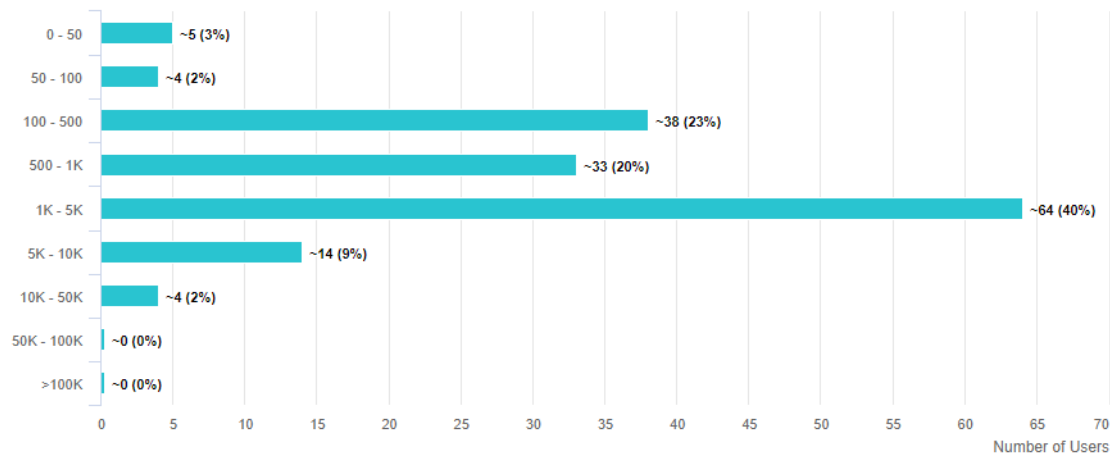


Figura 68 – Informe de seguidores de una audiencia monitorizada en Audiense



Figura 69 – Listado de cuentas de una audiencia (los nombres y fotos han sido ocultados con una franja negra por confidencialidad)

En la vista de detalle anterior, es posible seguir masivamente a todas las cuentas que aparecen, interactuar con ellas (mensaje directo), bloquearlas, y cualquier acción habitual en Twitter. Esto permite insertarse en la red formada por los miembros de la comunidad analizada, o crea una red nueva en torno a una o varias cuentas controladas por el investigador.

Una vez que una audiencia entra en monitorización, se hace un seguimiento de todas sus publicaciones (aunque no tengan que ver con la razón por la que fueron seleccionadas). Esto permite ver de qué otras cosas hablan y estudiar la polaridad de sus opiniones, de forma colectiva o individual.

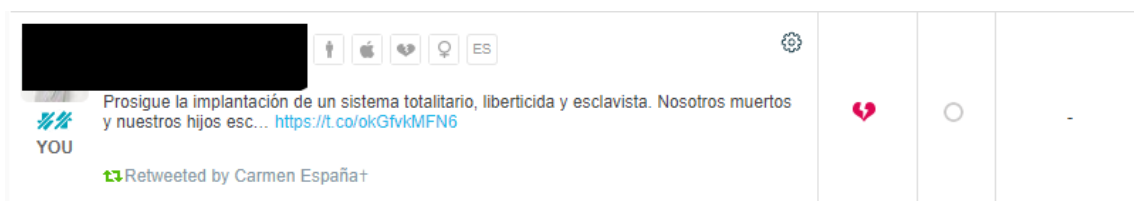


Figura 70 – Seguimiento de publicaciones de las personas que forman parte de la audiencia

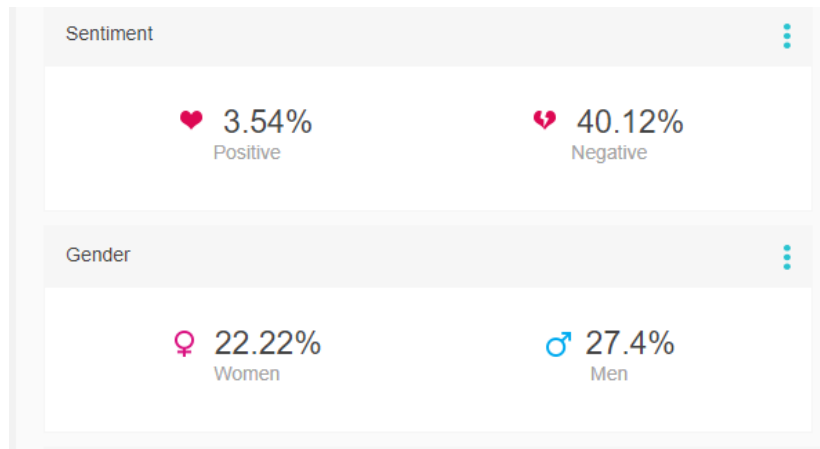


Figura 71 – Polaridad de las publicaciones

Por último, con la tecnología utilizada, sería posible obtener un perfilado psicológico profundo, provisto por los servicios integrados vía API de IBM Watson, con las inclinaciones del segmento analizado a fin alcanzar una comprensión profunda de sus características, necesidades y valores.

¿Sería posible impactar a las cuentas monitorizadas con contenidos específicos? Como muestra la imagen posible, sí, sería posible crear campañas que impactaran a la audiencia estudiada.

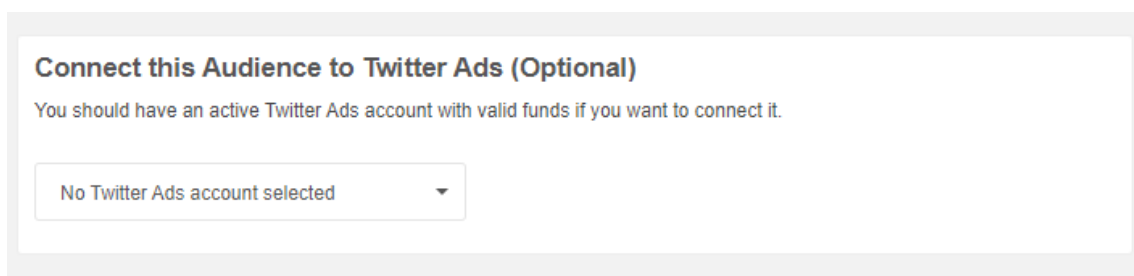


Figura 72 – conexión de con Twitter Ads para impactar a los miembros de la audiencia perfilada con publicidad específica

Conclusiones

El resultado de la prueba realizada demuestra de manera concluyente que, para la red social Twitter, es posible identificar a personas que apoyan el movimiento antivacunas del Covid-19, perfilarlas, monitorizar su actividad, e impactarlas con contenidos específicamente creados. Esta posibilidad puede ponerse al servicio de causas legítimas, como la concienciación, el desmentido de falsas informaciones y la investigación académica de los fenómenos sociales, o ilegítimas, como la manipulación.

También hemos comprobado cómo es posible identificar a líderes opinión, entablar relación de forma automatizada con las cuentas del colectivo monitorizado e interactuar con distintos grados de potencial automático.

Por tanto, queda plenamente validada la viabilidad del supuesto planteado.

No podemos afirmar de manera categórica que estas conclusiones sean extrapolables a cualquier opinión (distinta al tema escogido para el estudio), ya que eso requeriría repetir esta experiencia para una batería de temas representativos de todas las áreas de interés social en redes sociales.

No se han encontrado tecnologías ni métodos que permitan realizar este tipo de acciones en otras redes, como por ejemplo Facebook, y no parece, en virtud de la prospección realizada para contextualizar este estudio, que las restricciones técnicas y políticas de uso de META (Facebook e Instagram) lo permitan, razón por la que es improbable que puedan desarrollarse métodos *ad hoc* desde el ámbito académico (al menos sin involucrar acciones de hacking y vulneración de normas de uso).

Una posible objeción al método probado es que el criterio básico para la construcción del *dataset* no garantiza que el 100% de las cuentas realmente apoyen

(o rechacen) un asunto (en este caso, rechazo a las vacunas del Covid-19). Por ejemplo, si el criterio se basa en un *hashtag*, como en nuestro caso, alguien podría usarlo para criticarlo o ironizar sobre él. Esto puede solventarse de diversas formas, por ejemplo, con tecnologías adicionales: filtrando el *dataset* con herramientas de *sentiment análisis* que permitan una clasificación a nivel de entidad.

Anexo 6 – Revisión sistemática sobre el análisis de sentimiento de textos en español sobre política y discurso de odio

A fin de entender en profundidad el estado del arte en el estudio de la opinión de los ciudadanos de habla hispana, se realizó una revisión sistemática (actualizada a diciembre de 2021) con foco en los dominios “política” y “discurso de odio”.

Las preguntas de investigación a las que respondió el estudio fueron las siguientes:

1. RQ1 - ¿Cuáles son los trabajos publicados hasta la fecha, centrados específicamente en el análisis de sentimientos en español, para el ámbito de la política y el discurso de odio?
2. RQ2 - ¿Cuáles son los objetivos y temas que concentran el interés de la comunidad investigadora?
3. RQ3 - ¿Qué herramientas y métodos se utilizan?
4. RQ4 - ¿Cuáles son las fuentes de datos?
5. RQ5 - ¿Qué retos quedan por resolver?

El trabajo se realizó siguiendo la metodología PRISMA (Moher et al., 2009)

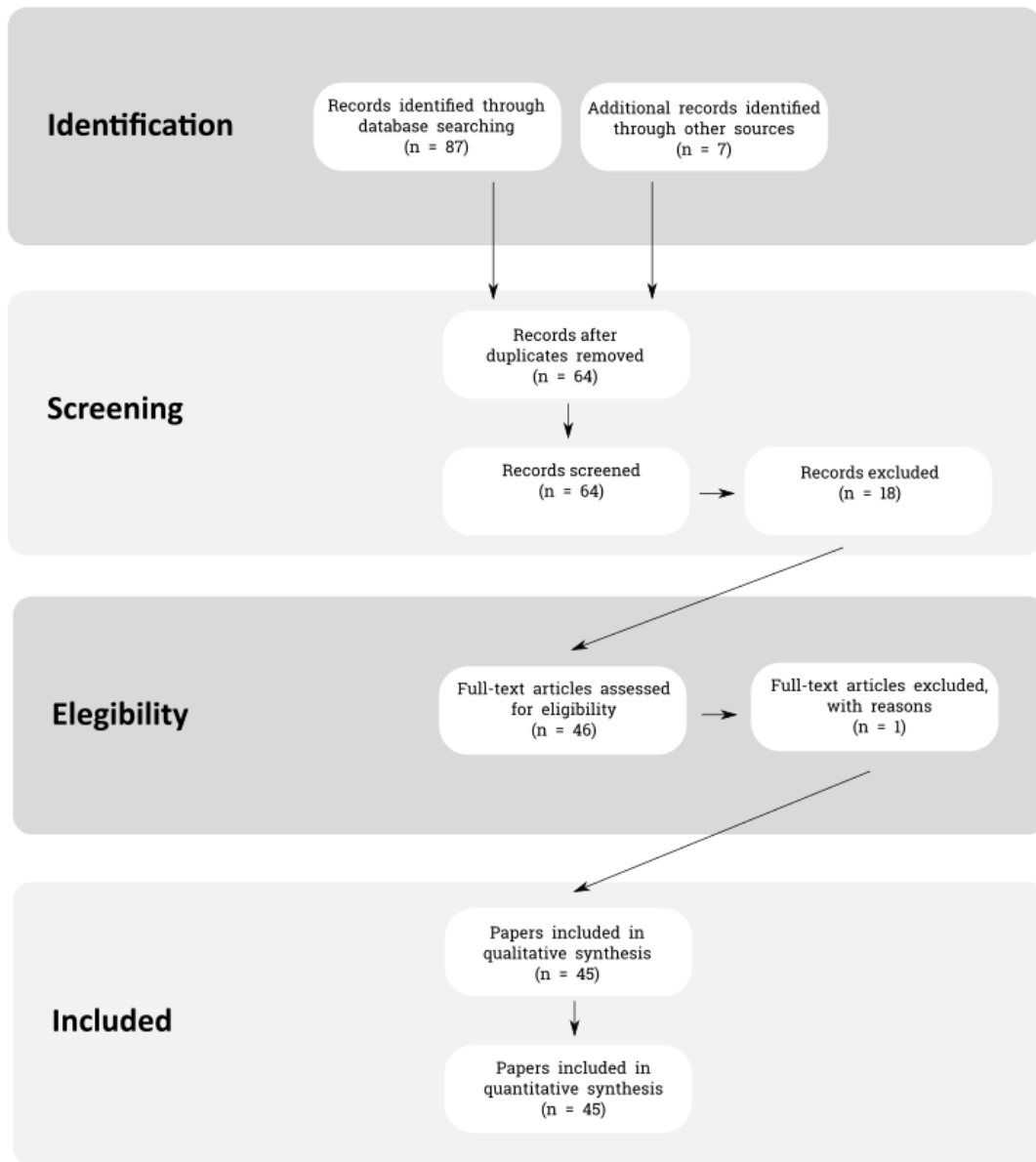


Figura 73 – Proceso en aplicación de la Metodología PRISMA

Los trabajos encontrados fueron sometidos a un análisis sistemático centrado en varios conceptos clave:

- Herramientas de análisis / Métodos

- Tipo de investigación
- Objetivo de la investigación / tema
- Fuentes textuales y conjuntos de datos
- Retos no resueltos, desarrollos futuros y aspectos controvertidos.

Authors	Year	Citations
Cuesta et al. 2014	2014	22
Pla & Hurtado 2014	2014	42
Agulló et al. 2015	2015	0
Vilares et al. 2015	2015	45
Cerón-Guzmán et al. 2016	2016	18
Castro et al. 2017	2017	13
Singh et al. 2017	2017	9
Arcila-Calderón et al. 2017	2017	14
Gomez-Torres et al. 2018	2018	0
Criado & Villodre 2018	2018	7
Hidalgo et al. 2018	2018	1
Gil-Vera & Montoya-Suarez 2018	2018	9
Pérez & Luque 2019	2019	10
Pereira-Kohatsu et al. 2019	2019	54
Vega et al. 2019	2019	8
Bohorquez-Lopez et al. 2019	2019	0
Franco-Riquelme et al. 2019	2019	5

Baviera et al. 2019	2019	6
Baviera et al. 2019	2019	3
Almatarneh et al. 2019	2019	7
Cignarella 2020	2020	0
Sanchez-Nunez et al. 2020	2020	0
Grimaldi et al. 2020	2020	4
Pamungkas et al. 2020	2020	0
Arcila-Calderón et al. 2020	2020	15
Blasco-Duatis & Coenders 2020	2020	2
Pastor-Galindo et al. 2020	2020	8
Arcila et al. 2020	2020	12
Ramon-Hernandez et al. 2020	2020	1
Plaza-Del-Arco et al. 2021	2021	1
Plaza-del-Arco et al. 2021	2021	24
Córdoba-Cabú et al. 2021	2021	0
Andrade-Segarra et al. 2021	2021	2
Tamayo et al. 2021	2021	0
Uzan & HaCohen-Kerner 2021	2021	0
Romero-Vega et al. 2021	2021	1
Sanchez-Junquera et al. 2021	2021	3
Jain et al. 2021	2021	1
Gómez-Zaragoza & Pinto 2021	2021	1

Huertas-García et al. 2021	2021	0
Arcila-Calderón et al. 2021	2021	5
Rodríguez-Ibanez et al. 2021	2021	1
Cervero 2021	2021	0
Rendon-Cardona et al. 2022	2022	0
Robles et al. 2022	2022	0

Tabla 39- Trabajos seleccionados

Como muestran las figuras siguientes, se observa una fuerte tendencia hacia el estudio del discurso de odio, y una prevalencia del enfoque analítico sobre el enfoque predictivo y de monitorización.

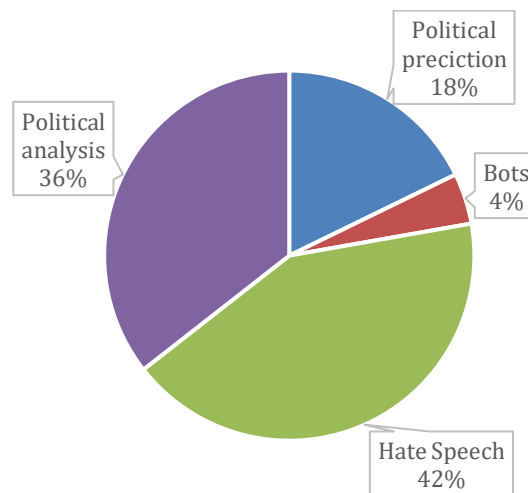


Figura 74 – Foco de los trabajos analizados (temática)

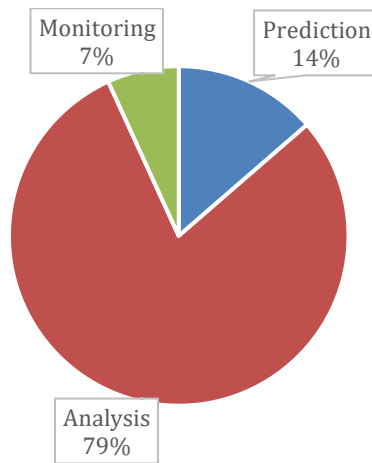


Figura 75 - Foco de los trabajos analizados (tipo de investigación)

En cuanto a los recursos más habituales para las metodologías propuestas por los autores, se obtuvo la siguiente tabla resumen:

Resource	Papers	%
BERT based	8	18%
NLTK	5	11%
Sentistrength	4	9%
Freeling	3	7%
SciKit	3	7%
Stanford NLP	2	4%
LinguaKit	2	4%
AFINN	2	4%

Tabla 40 – Recursos más empleados en los trabajos analizados.

En línea con lo observado en la producción dedicada al resto de idiomas, más del 80% de los trabajos se basan en textos de Twitter.

Anexo 7 – Frecuencia de aparición de múltiples entidades en los mensajes publicados por los usuarios de Facebook y Twitter

Introducción

Esta prueba tiene por objetivo comprobar el alcance de la multiplicidad de entidades en los textos de las plataformas de *micro blogging* principales (Facebook y Twitter). Sabemos, por experiencia directa y observación empírica informal, que es habitual que los mensajes publicados contengan entidades diversas, pero para el propósito de la investigación de la tesis en la que se enmarca este trabajo, se ha considerado conveniente realizar una comprobación exacta mediante técnicas de procesamiento de lenguaje natural y cotejamiento con la literatura científica disponible.

Metodología

Para la realización de la prueba, hemos utilizado los *datasets* de “Entretenimiento TV” y “Antivacunas Covid-19” utilizados en una investigación precedente (Anexo 3).

Dataset “entretenimiento”:

- Palabra clave: “first dates”¹⁵⁹

¹⁵⁹ #yonomevacuno es uno de los hashtags más populares en el movimiento antivacunación durante la pandemia del Covid-19.

- Fuente: Twitter
- Período de extracción: septiembre-noviembre de 2021 (tres meses)
- Este *dataset* ya fue empleado en el estudio previo referenciado en el Anexo 1

Dataset “anti vacunas”:

- Palabra clave: “yonomevacuno”¹⁶⁰
- Fuente: Facebook

Estos *datasets* han sido evaluados con Repustate, una herramienta que aparece en evaluaciones de otros trabajos de *benchmarking*, como las de Abbasi et al. (2014) y Zimbra et al. (2018) (ver también referencias en Anexo 1), y que ha sido utilizada en otros aspectos de la investigación para la tesis en la que se enmarca esta prueba (Anexo 3).

Trabajos relacionados

Existen numerosos trabajos centrados en *entity recognition*, como la investigación realizada por Pappu et al. (2017) sobre extracción multilingüe, pero no se han encontrado referencias estadísticas sobre el porcentaje de textos de redes sociales con múltiples entidades (los trabajos se centran en metodologías de reconocimiento de entidades y, en algunos casos, aspectos de las entidades).

Resultados

La siguiente tabla refleja los resultados obtenidos con el *dataset* de Facebook:

¹⁶⁰ Adif es el acrónimo de “Administrador de Infraestructuras Ferroviarias”. Se trata de una empresa pública española, dependiente del Ministerio de Transportes, encargada del mantenimiento y explotación de las infraestructuras ferroviarias (más de 15.000 kilómetros de vías cerca de 2000 estaciones).

	#	%
Tamaño del dataset	1980	100%
Mensajes con 1 entidad	486	25%
Mensajes con 2 entidades	222	11%
Mensajes con 3 entidades	149	8%
Mensajes con 4 entidades	82	4%
Mensajes con 5 o más entidades	147	7%
TOTAL con más de una entidad	1086	55%

Tabla 41 – Número de tuits con múltiples entidades, dataset “Antivacunas Covid-19”.

Como se puede observar, el 55% de los mensajes de Facebook analizados incluía más de una entidad, e incluso hay un 7% con la sorprendente cifra de 5 o más.

El mismo estudio, aplicado sobre literales de Twitter, arroja un resultado también inequívoco, pese a la limitación de caracteres de esta red social:

	#	%
Tamaño del dataset	795	100%
Mensajes con 1 entidad	223	28%
Mensajes con 2 entidades	101	13%
Mensajes con 3 entidades	51	6%
Mensajes con 4 entidades	8	1%
Mensajes con 5 o más entidades	10	1%
TOTAL con más de una entidad	393	49%

Tabla 42 - Número de tuits con múltiples entidades, dataset “Entretenimiento TV”.

El promedio de palabras por mensaje en el *dataset* de Facebook es de 41,25, mientras que en el de Twitter es de 22,69. Esto se debe a la limitación de caracteres de los mensajes de Twitter, y explica la muy superior incidencia de múltiples entidades en Facebook. No obstante, podemos comprobar que la probabilidad de

múltiples entidades no es proporcional a la longitud del texto, ya que la diferencia entre ambos *datasets* es de 6 puntos porcentuales, cuando la longitud de los textos en Facebook es casi el doble que en Twitter.

Conclusiones

A la vista de los resultados obtenidos mediante el reconocimiento de entidades en los *datasets* de referencia, se comprueba que el potencial de multi entidad en las redes de *microblogging* puede llegar a afectar a más del 50% de los mensajes.

Anexo 8 – Contextual Inquiry: onboarding en herramienta informal de sentiment analysis

Introducción

El objetivo planteado para esta prueba es determinar si personas no expertas en ciencia de datos, con distintos grados de conocimiento previo sobre interfaces digitales, son capaces de extraer y procesar la información sobre la actitud de los usuarios en Internet usando una herramienta comercial. Para ello, se eligió SentiOne, por ser la que mejor rendimiento obtuvo en las investigaciones previas (Anexos 1 y 3).

Dado que se trata de evaluar la usabilidad de la herramienta y su relación con el usuario, aplicamos una aproximación basada en los métodos de la ingeniería de producto y el diseño de interfaces gráficas, siempre con un fundamento científico riguroso. La técnica elegida es el *contextual inquiry*, una veterana metodología

etnográfica descrita en la década de los 90 por Bennett et al. (1990), muy habitual en la actualidad como método de investigación en UX¹⁶¹.

Contextualización y trabajos relacionados

El *contextual inquiry* se ha usado en ocasiones como procedimiento para evaluar el rendimiento de las herramientas de la ciencia de datos desde el punto de vista de la interacción entre usuario y herramienta (relación hombre-máquina) (Kaur et al., 2020). No obstante, existen pocos trabajos con este tipo de enfoque, ya que esta perspectiva es más propia de la ingeniería y el diseño. En el caso del trabajo de Kaur et al. (2020), el más representativo sin nos atenemos al número de citas¹⁶², los autores descubrieron mediante este procedimiento (complementado con encuestas) que los científicos confían demasiado en las herramientas de interpretabilidad analizadas y las utilizan mal. Un hallazgo sin duda interesante y de importantes consecuencias.

Metodología

En este apartado describiremos brevemente el método empleado, y las condiciones específicas de la investigación realizada.

Descripción de la metodología contextual inquiry

El *contextual inquiry* es una metodología cualitativa de la familia “etnográfica” basada en la observación participante en el contexto. El investigador comprueba empíricamente cómo es la relación que se establece entre un sistema y sus usuarios

¹⁶¹ User Experience (Experiencia de Usuario)

¹⁶² 94 citas. Dato de SCOPUS, última consulta julio de 2022.

en un entorno naturalista¹⁶³, observando sus tareas y realizando preguntas al respecto para profundizar cualitativamente en los puntos más sensibles de acuerdo con los objetivos de la investigación. Su objetivo no es evaluar el rendimiento de los procesos, sistemas o productos puestos a prueba, sino entender en profundidad aspectos cognitivos y procedimentales (uso), o responder cualitativamente a preguntas de investigación específicas relacionadas con la usabilidad y accesibilidad. En algunos casos, *el contextual inquiry* se usa como información básica para el diseño de sistemas de nueva planta, y en otros casos, como el que nos ocupa, para prospectar la relación que se establece entre los usuarios y los sistemas ya existentes.

La observación participante es un tipo de técnica de recolección de datos cualitativos, utilizada habitualmente en sociología, antropología y psicología social. De hecho, puede decirse que la antropología social y cultural ha construido su identidad alrededor de este tipo de técnica (Restrepo, 2016).

Participantes y criterios de elegibilidad

Dado que se trata de una metodología cualitativa, no se persigue la representatividad estadística sino estructural. En este sentido, se ha realizado la comprobación de uso con tres sujetos de diferente nivel de cualificación, y tres únicos atributos comunes:

Atributos comunes:

- Habitadas al uso de interfaces digitales.

¹⁶³ “Entorno naturalista” se refiere a contexto natural, por oposición a un contexto “artificial” de laboratorio.

- Nunca han usado herramientas de monitorización de la actitud de los usuarios en Internet con *sentiment analysis*.
- Carecen de conocimientos sobre Ciencia de Datos, Procesamiento de Lenguaje Natural y otras disciplinas académicas relacionadas con el *machine learning*.

Atributos no comunes:

- Sujeto 1 (nivel de cualificación 0): persona sin formación superior y que no tiene una profesión relacionada con la monitorización de la actitud de los usuarios en Internet.
- Sujeto 2 (nivel de cualificación 1): profesional junior del área de investigación de marketing.
- Sujeto 3 (nivel de cualificación 3): profesional senior del área de investigación de marketing.

Objeto de análisis

SentiOne es una herramienta “de pago”. Para esta investigación, se trabajó con la versión “Enterprise” (acceso total a todas las funcionalidades). La siguiente figura muestra el aspecto del *dashboard* de análisis de un proyecto.



Figura 76 - Vista del panel de análisis de un proyecto en SentiOne.

Tarea

La tarea a realizar por los sujetos usuarios es configurar un proyecto correctamente, a fin de monitorizar las menciones a una patología médica específica en un período de tiempo determinado, para recuperar todos los textos publicados en Internet (accesibles) y generar el correspondiente informe automático con el análisis de la IA de la aplicación.

A fin de reproducir las condiciones reales de uso, los participantes recibieron una formación de *onboarding* (duración 30 minutos) equivalente a la que ofrece el equipo comercial de la herramienta como parte del servicio contratado.

Resultados

Como resultado de la observación realizada, pudimos comprobar que:

- Todos los sujetos fueron capaces de realizar la tarea.
- La sujeto 1 (profesión no relacionada y sin estudios superiores), aun siendo capaz de configurar mecánicamente el proyecto, no tuvo una comprensión plena sobre lo que estaba haciendo, manifestando dificultades con el vocabulario técnico, el significado las funcionalidades más complejas, y la interpretación de los resultados.
- Más allá del uso del interfaz, los sujetos 2 y 3 fueron capaces de comprender sin dificultad el propósito de las distintas funcionalidades de la herramienta, así como interpretar los resultados obtenidos.
- Los sujetos 2 y 3 reaccionaron con asombro al potencial de la aplicación. En la sujeto 1 se identificaron sentimientos de asombro y también de alarma.

Conclusiones

La observación realizada concluye, desde el punto de vista del análisis de los datos cualitativos recogidos, que la herramienta analizada permite a personas profesionales del marketing sin conocimientos específicos de ciencia de datos ni procesamiento de lenguaje natural, programar y ejecutar monitorizaciones sobre la actitud de los usuarios en Internet y *sentiment analysis*. También se ha comprobado que una persona sin conocimientos de marketing ni estudios avanzados, partiendo solo del conocimiento popular sobre el uso de aplicaciones digitales y el funcionamiento general de Internet, podría realizar la misma tarea, aunque en este caso sin tener un nivel de comprensión suficiente como para sacar partido a la aplicación.

