



FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES
MÁSTER EN ESTUDIOS AVANZADOS EN DIRECCIÓN DE EMPRESAS

TRABAJO FIN DE MÁSTER
CURSO ACADÉMICO [2021-2022]

TÍTULO: Análisis de opiniones sobre el pasaporte COVID-19 mediante procesamiento de lenguaje natural (NLP)

AUTOR: JUAN ANTONIO AGUILAR MORENO

TUTOR: DR. D. PEDRO RAMIRO PALOS SÁNCHEZ

DEPARTAMENTO: Economía Financiera y Dirección de Operaciones

ÁREA DE CONOCIMIENTO: Organización de empresas

RESUMEN:

El objetivo central de este trabajo es identificar la opinión general sobre la implantación del pasaporte COVID-19, los motivos que llevan a ello y los aspectos que generan mayor interés. Para ello se realiza un análisis de sentimiento y de *topics* en el software RapidMiner con una muestra de 24.182 comentarios procedentes de doce vídeos relacionados con el pasaporte COVID-19 en los canales oficiales de reputados medios de prensa británicos en la plataforma YouTube. Como hallazgos se encuentran que los partidarios se basan en la necesidad de retomar la actividad normal y reactivar una economía paralizada, mientras que los detractores se posicionan en contra debido a la no inmunidad total de la vacuna y argumentos éticos. Precisamente de conocer estas opiniones mediante las técnicas emergentes de procesamiento de lenguaje natural, surge la utilidad de este trabajo como medio para implementar mejoras en el pasaporte o como conocimiento para futuras herramientas similares.

TÉRMINOS CLAVE:

Procesamiento de lenguaje natural; Análisis de sentimientos, análisis de *topics*, RapidMiner, certificado COVID-19

ABSTRACT:

The main goal of this work is to identify the general opinion on the implementation of the COVID-19 passport, the reasons that lead to it and the aspects that generate the main interest. A sentiment and topic analysis was carried out using RapidMiner software with a sample of 24,182 comments from twelve videos related to the COVID-19 passport on the official channels of reputable British press media on the YouTube platform. The findings found that supporters were based on the need to resume normal activity and revive a paralysed economy, while detractors were against because of the vaccine's lack of full immunity and ethical arguments. The usefulness of this work as a means to implement improvements in the passport or as knowledge for future similar tools arises precisely from knowing these opinions through the emerging techniques of natural language processing.

KEYWORDS:

COVID-19 Passport; COVID-19 certificate; Natural Language Processing, Topics Analysis; Sentiment Analysis.

ÍNDICE

1	INTRODUCCIÓN, OBJETIVOS Y ESTRUCTURA	7
1.1	INTRODUCCIÓN.....	7
1.2	OBJETIVOS.....	9
1.3	ESTRUCTURA.....	9
2	FUNDAMENTOS TEÓRICOS	11
2.1	REVISIÓN DE LITERATURA DEL PASAPORTE COVID-19.....	11
2.1.1	Propuestas tecnológicas para reducir la incidencia del COVID-19	12
2.1.2	Generalización del uso del código QR a raíz del COVID-19	13
2.1.3	Opinión del público general sobre ventajas e inconvenientes del pasaporte COVID-19	14
2.1.4	Impacto del pasaporte COVID-19 en sectores económicos	18
2.1.5	Impacto del pasaporte COVID-19 en la decisión de aceptar las vacunas.....	18
2.1.6	Consideraciones éticas	19
2.2	REVISIÓN DE LITERATURA DE PUBLICACIONES ACERCA DE DATA MINING Y SENTIMENT ANALYSIS	20
2.2.1	Conceptos	20
2.2.2	Trabajos previos y herramientas.....	22
3	METODOLOGÍA.....	25
4	ESTUDIO EMPÍRICO Y ANÁLISIS DE DATOS.....	28
4.1	PROCESO DE EXTRACCIÓN, PREDICCIÓN Y ANÁLISIS DE LOS DATOS	28
4.1.1	Extracción de comentarios	28
4.1.2	Depuración del archivo en Microsoft Excel	30
4.1.3	Análisis usando <i>RapidMiner</i>	31
4.2	INTERPRETACIÓN DE LOS RESULTADOS	36
4.2.1	Nube de palabras	36
4.2.2	Clasificación por <i>topics</i>	36

4.2.3	Análisis de sentimiento.....	44
5	DISCUSIÓN.....	48
6	CONCLUSIONES Y DESAFÍOS FUTUROS	50

Relación de Figuras

Figura 1. Certificado COVID-19 Digital de la UE	12
Figura 2. Interés del pasaporte COVID-19 en Google Trends (España)	17
Figura 3. Interés del pasaporte COVID-19 en Google Trends (Internacional)	18
Figura 4. Esquema del proceso de la metodología	27
Figura 5. Interfaz de MAXQDA Analytics Pro 2022 con algunos de los comentarios extraídos.....	30
Figura 6. Tabla de Microsoft Excel con los comentarios extraídos.....	30
Figura 7. Interfaz de diseño del proceso en RapidMiner	31
Figura 8. Selección de atributo de filtro en RapidMiner	32
Figura 9. Operador Generate Attributes en RapidMiner	33
Figura 10. Interfaz de resultados en RapidMiner - Clasificación de comentarios por sentimiento	45

Relación de Tablas

Tabla 1. Apps desarrolladas para el rastreo de contagios.....	13
Tabla 2. Aceptación del pasaporte COVID-19 según el país de la UE	15
Tabla 3. Resumen de las ventajas e inconvenientes identificadas del pasaporte COVID-19.....	16
Tabla 4. Resumen de referencias del marco teórico sobre opiniones e impacto del pasaporte COVID-19	20
Tabla 5. Recopilación de publicaciones con metodología similar a este estudio	24
Tabla 6. Criterios de selección de vídeos.....	25
Tabla 7. Fuentes utilizadas para la extracción de comentarios	29
Tabla 8. Frecuencia de palabras en RapidMiner.....	35
Tabla 9. Pesos de las palabras del <i>Topic 1</i> “Opiniones generales sobre la obtención del pasaporte COVID-19 y la vacunación”	37
Tabla 10. Pesos de las palabras del <i>Topic 0</i> “Teoría de la conspiración”.....	39

Tabla 11. Pesos de las palabras del <i>Topic 2</i> “Asuntos sanitarios”	40
Tabla 12. Pesos de las palabras del <i>Topic 3</i> “Derechos ciudadanos”	42
Tabla 13. Pesos de las palabras del <i>Topic 4</i> “Preocupación sobre seguridad ciudadana a raíz del COVID-19”	43
Tabla 14. Recuento de <i>likes</i> según el <i>topic</i>	47

Relación de Gráficos

Gráfico 1. Proporción de comentarios según fuente	29
Gráfico 2. Nube de palabras en RapidMiner	36
Gráfico 3. Clasificación por <i>topics</i> en RapidMiner	44
Gráfico 4. Recuento de comentarios por <i>topics</i>	44
Gráfico 5. Recuento de comentarios por sentimiento.....	45
Gráfico 6. Distribución de comentarios según su puntuación de positividad / negatividad	46

1 INTRODUCCIÓN, OBJETIVOS Y ESTRUCTURA

1.1 INTRODUCCIÓN

La progresiva vuelta a la normalidad tras la pandemia del COVID-19 ha venido acompañada de limitaciones y medidas para contener un mayor avance del número de casos de contagio al mismo tiempo que se reactiva la economía, entre ellas una de las que más debate ha provocado es la implantación del pasaporte COVID-19 como instrumento para retomar la libre movilidad de personas y permitir el acceso a lugares con un gran aforo bajo la premisa de que sus poseedores cuentan con un mayor grado de inmunidad al haber sido vacunados. A esto se le suma el contexto actual, es la primera pandemia de la historia en la era digital, y la capacidad de influencia y movilización que tienen las redes sociales, donde cada día se expresan una ingente cantidad de opiniones sobre los temas del momento.

En marzo de 2020, la Organización Mundial de la Salud (OMS) declaró oficialmente como pandemia la evolución de la COVID-19, cuyo primer caso documentado tuvo lugar en China a final de 2019 y se propagó rápidamente por todo el mundo en el primer trimestre de 2020 (Liu et al., 2020), causando desde diciembre a final de marzo un total de 750.890 infecciones y 36.405 muertes en todo el mundo según datos de la OMS (2020). A raíz de la citada situación, los gobiernos y las instituciones supranacionales comenzaron a tomar medidas y restringir la movilidad de las personas para contener los contagios, lo que supuso importantes consecuencias también en la economía, que experimentó una recesión motivada por la falta de actividad comercial, limitaciones y restricciones (Kolahchi et al., 2021).

El progresivo restablecimiento de la normalidad cotidiana tras el confinamiento por la pandemia del COVID-19 ha estado marcado por las restricciones de movilidad, horarias, de aforo, entre otras, como medidas para reducir el número de contactos y evitar una propagación aun mayor, lo cual ha tenido una grave incidencia en variados sectores económicos, especialmente en aquellos negocios que se consideran “no esenciales”, como la hostelería, turismo y ocio y entretenimiento, que se vieron obligados a limitar su aforo e incluso al cierre total dejando de percibir ingresos pero manteniendo sus costes (de la Fuente, 2021). Estas limitaciones han sido acompañadas por el pasaporte COVID-19, también comúnmente denominado certificado COVID-19 o certificado de vacunación, un documento necesario para viajar sin necesidad de someterse a pruebas o guardar cuarentena y acceder a determinados locales, incluso a lugares de trabajo, que acredita nominativamente a su

poseedor un mayor nivel de inmunidad debido a haber superado la enfermedad recientemente o haber sido vacunado con las dosis requeridas (Shin et al., 2021).

No obstante, esta medida que a priori puede parecer efectiva para su cometido no ha dejado a nadie indiferente y ha provocado un acentuado debate entre partidarios y detractores, de manera que los aspectos políticos han llegado a cobrar un alto grado de relevancia para abordar estas preocupaciones tratando de equilibrar los intereses y presiones de los diferentes grupos de interés, así como cuestiones éticas tales como las desigualdades de los sistemas sanitarios en países en vías de desarrollo que cuentan con un reducido suministro de vacunas, la libre decisión de una persona para vacunarse y la prohibición de ingreso a un país de una persona no vacunada o la mera posición en la lista de espera según tramo de edad para vacunarse. (Sharun et al., 2021).

Por otro lado, en la actualidad las redes sociales se han convertido en un medio de comunicación en el que la población puede compartir opiniones lo que ha dado lugar a que estas redes se hayan convertido en una importante fuente de datos, muy fácilmente accesibles en las que un gran número de personas interactúan entre sí y expresan sus opiniones y emociones, que en situaciones de grandes desastres humanitarios pueden ser beneficiosas para el sector público debido al acceso a la información que disponen y a la facilidad para comunicar. (Malawani et al., 2020).

Utilizando estas opiniones se pueden emplear técnicas de *Data Mining* y *Natural Language Processing*, con el objetivo principal de comprender las preocupaciones sociales y que posteriormente se pueda emplear esta información para evaluar la aceptación del pasaporte COVID-19 y estudiar posibles vías de mejora equilibrando la ventaja que proporciona en cuanto al menor riesgo de contagio con la factibilidad para su uso en determinados contextos como podrían ser el acceso a un restaurante, local de ocio o medios de transporte públicos.

En este estudio se propone la realización de un análisis de las opiniones expresadas al respecto por usuarios en internet desde el inicio de la pandemia, concretamente como comentarios de respuesta a vídeos relativos a la implementación del pasaporte COVID-19 en prestigiosos medios de comunicación británicos, los cuales se publican en la plataforma YouTube, para llegar a conclusiones sobre la opinión general de la población y discernir el grado de aceptación mostrado mediante técnicas de *Web Scraping*, *Text Mining* y *Natural Language Processing (NLP)*.

1.2 OBJETIVOS

El objetivo central de este trabajo es identificar la opinión general sobre la implantación del pasaporte COVID-19, los motivos que llevan a ello y los aspectos que generan mayor interés.

Estos objetivos principales se dividen en objetivos específicos cuyo cumplimiento son necesarios para la obtención de los resultados con los que llegar al objetivo general:

- Proporcionar un marco teórico de referencia en el que se contextualice adecuadamente esta investigación para identificar áreas de estudio susceptibles de mejora entre todas las publicaciones encontradas en la revisión de literatura.
- Identificar fuentes de opiniones online acerca del pasaporte COVID-19 en redes sociales.
- Realizar una extracción de comentarios sobre opiniones relativas al pasaporte COVID-19 mediante el uso de técnicas de *web scrapping*, a partir de los cuales se desarrollará el estudio práctico implementando técnicas de *Natural Language Processing (NLP)*, concretamente de *Sentiment Analysis* y *Topics Analysis*.
- Interpretar los resultados proporcionados por los softwares anteriormente mencionados para llegar a las conclusiones del trabajo y realizar una comparación de estos con estudios de similares características.

En línea con los objetivos marcados, las preguntas de investigación son las siguientes:

P1: ¿Ha sido percibido el pasaporte COVID-19 como una medida de éxito o fracaso entre los usuarios?

P2: ¿Cuáles son los principales asuntos que más preocupan o que más se mencionan en los comentarios?

1.3 ESTRUCTURA

Este trabajo consta de seis capítulos, incluyendo este mismo de la estructura que informa de la introducción, los objetivos y estructura. A continuación, se desglosa el contenido y descripción de cada capítulo.

Este primer capítulo de "Introducción, objetivos y estructura" pone en contexto el trabajo desarrollado explicando la situación social, económica y política en la que surge la implantación del pasaporte COVID-19, su impacto y aceptación pública. Se exponen también los objetivos que marcan la pauta del trabajo y la estructura en la que está organizado.

En el segundo capítulo, “Fundamentos teóricos y revisión bibliográfica”, se comenta la literatura existente relacionada con el trabajo, dividiéndose este capítulo en dos apartados, uno para los artículos relacionados con el pasaporte COVID-19 y otro para los que se vinculan al análisis de sentimiento en el que se contextualiza la técnica que se va a utilizar en el desarrollo del trabajo.

La metodología es el tercer capítulo, en el que se explica detalladamente los pasos seguidos en el trabajo desde la cuestión de investigación hasta las conclusiones que le dan respuesta.

El cuarto capítulo, “Estudio empírico y análisis de datos” se centra en la obtención de los comentarios u opiniones acerca del pasaporte COVID-19 procedentes de internet y su tratamiento y análisis para la obtención de los resultados a partir de los cuales se obtendrán las conclusiones. En esta parte se implementan softwares especializados en *Data Mining*, *Web Scrapping* y *Sentiment Analysis*.

El quinto capítulo, “Discusión” cuestiona los resultados obtenidos tras la investigación y los compara con los de otros estudios de similares características.

Para finalizar, en el sexto capítulo se exponen las conclusiones obtenidas tras la obtención de los resultados del tercer capítulo y se contrastan con la revisión bibliográfica del segundo capítulo en la discusión. Asimismo, se comentan algunas líneas futuras de investigación complementarias al trabajo desarrollado.

2 FUNDAMENTOS TEÓRICOS

La revisión de literatura explicada en los siguientes apartados se subdivide por las temáticas desarrolladas en este trabajo, que abarca dos diferentes áreas (la relacionada con la implementación del pasaporte COVID-19 y la que corresponde a la metodología empleada para realizar el estudio empírico en temas referentes a *Data Mining*, *Web Scraping*, *Natural Language Processing*, *Machine Learning* y *Sentiment Analysis*).

En cuanto a la primera, a pesar de ser un tema de reciente actualidad, existe una considerable cantidad de publicaciones, de las cuales, siguiendo la línea de esta investigación, se considerarán relevantes aquellas relacionadas con el papel de las redes sociales y las consecuencias que haya podido tener el pasaporte COVID-19 en ámbitos de gran impacto económico tales como el turismo, las aerolíneas, el comercio o el entorno laboral, así como los aspectos éticos y legales que han podido generar discordancia para la aceptación del certificado.

2.1 REVISIÓN DE LITERATURA DEL PASAPORTE COVID-19

El pasaporte COVID-19, certificado COVID-19 o certificado de vacunación se trata un documento físico o digital (se muestra un ejemplo en la figura 1 “Certificado COVID-19 digital de la UE) que acredita que su poseedor ha recibido las dosis requeridas de la vacuna, haber obtenido un resultado negativo en una prueba o haber superado la enfermedad. (Shin et al., 2021). Este documento es obligatorio en muchos países para entrar en sus fronteras y fue adoptado por la Unión Europea con el objetivo de garantizar la libre circulación de personas en los Estados miembros y contribuir al levantamiento gradual de las restricciones. (Gstrein, 2021).



Figura 1. Certificado COVID-19 Digital de la UE

2.1.1 Propuestas tecnológicas para reducir la incidencia del COVID-19

Entre las propuestas tecnológicas para reducir la incidencia del COVID-19, se encuentran el monitoreo con tecnología *blockchain*, el uso de inteligencia artificial para el monitoreo de positivos, el uso de códigos QR y el uso de aplicaciones móviles para avisar a sus usuarios sobre casos positivos cercanos. (Velicia-Martin et al., 2021). El pasaporte COVID-19 no ha sido, por tanto, la única propuesta tecnológica surgida para mitigar los efectos de la pandemia y poder retomar la normalidad de una forma progresiva.

Las aplicaciones móviles fueron una de las primeras herramientas que surgieron y su funcionamiento consiste en notificar al usuario cuando detecta que ha tenido contacto en los días previos con otro usuario registrado y que ha dado positivo en COVID-19, de manera que así tome las precauciones oportunas para evitar seguir propagando el virus en caso de haberlo contraído. Para determinar si ha existido contacto utilizan sistemas de geolocalización o tecnología Bluetooth. Pese a que no se proporciona el nombre personal, la App puede tener acceso a otra información como la ubicación del individuo o su IP (Bradford et al., 2020). La protección del anonimato y de la privacidad es una de las mayores preocupaciones de los usuarios al utilizar estas aplicaciones, aunque, según Velicia-Martin et al. (2021), los usuarios anteponen la información sobre los contactos que proporciona la App a sus incertidumbres sobre privacidad.

En la siguiente tabla se enumeran algunos ejemplos concretos de las mencionadas aplicaciones móviles desarrolladas para el rastreo de positivos.

Nombre App	País	Tecnología
STOP COVID	Francia	Bluetooth
Coronamadrid	España	Geolocalización
StopCovid19	España	Geolocalización
RadarCOVID	España	Bluetooth
Stop-Corona-App	Austria	Bluetooth
ProteGo	Polonia	Bluetooth

Tabla 1. Apps desarrolladas para el rastreo de contagios

Fuente: Bradford et al., 2020

2.1.2 Generalización del uso del código QR a raíz del COVID-19

Teniendo en cuenta las propuestas tecnológicas del epígrafe anterior, otra de las tecnologías que ha experimentado una gran difusión con motivo de la pandemia es la del código QR, el sistema sobre el que se asienta el pasaporte COVID-19. China fue uno de los primeros países en utilizarlo para permitir que los individuos pudieran acceder a ciertos espacios o a viajar. Un sistema útil, pero no exento de inconvenientes, pues no es tan accesible para personas mayores al depender totalmente de la tecnología digital (Wang y Jia, 2021).

Los códigos de respuesta rápida (QR) son códigos bidimensionales, a diferencia de los códigos de barras que son unidimensionales y cuentan con mayores limitaciones (Wahsheh y Al-Zahrani, 2021) que permiten representar datos, por ejemplo, una URL o una localización y ser escaneados por dispositivos móviles (Petrova et al., 2016).

La gran ventaja por la que en medio de una pandemia experimentó un auge el uso del código QR en cualquier ámbito es su facilidad de uso y la reducción del contacto físico (Jiang et al., 2021), como la lectura de la carta en locales de hostelería o el pago móvil mediante QR, en lugar de usar efectivo o tarjeta bancaria, que incluso da lugar a pensar en la desaparición futura del efectivo. (Tu et al., 2022)

En el apartado de la salud, los códigos QR de salud en China son certificados electrónicos del estado de salud de los individuos, y se usan rastrear los contactos y determinar automáticamente el riesgo de contagio en cada caso, actualizar de forma

automatizada del estado de salud y realizar citas médicas sin contacto. (Nakamoto et al., 2020). Su funcionamiento es análogo para el pasaporte COVID-19, un código QR identifica al sujeto con su historial de vacunación que puede ser consultado rápidamente escaneando el código.

2.1.3 Opinión del público general sobre ventajas e inconvenientes del pasaporte COVID-19

Las redes sociales son sitios web de interacción entre usuarios en los que crean perfiles públicos o semipúblicos donde comparten opiniones, contenido multimedia, información y otro contenido de interés común e interaccionan con otros usuarios cuyo vínculo se basa fundamentalmente en algún tipo de interdependencia como podría ser un tema de interés, amistad, creencias, negocios, entre otros. (Tanbeer et al., 2013), (Liu y Ying, 2010).

En el contexto actual, con una sociedad global interconectada digitalmente, las redes sociales suponen una inmensa fuente de información de elevado valor para analizar opiniones y tendencias sobre los asuntos de interés del momento que incluso pueden resultar de ayuda a las entidades gubernamentales, como argumentan Malawani et al. (2020) en un trabajo en el que implementan técnicas de *big data* para dar respuesta en tiempo real a problemas surgidos como consecuencia de un devastador tifón en Filipinas.

Focalizando la atención en la adopción del pasaporte COVID-19, se trata de una decisión que ha despertado un elevado número de opiniones encontradas, tanto de personas que apoyan su implantación como de quienes lo rechazan. Entender las preocupaciones e inquietudes de éstos últimos supone el primer paso para redefinir y mejorar esta herramienta de contención de la enfermedad comprendiendo las opiniones vertidas en las redes sociales con las técnicas de *Data Mining* y *Sentiment Analysis*.

Existe un trabajo previo (Khan et al., 2022) que se plantea cuál es la opinión del público general acerca del pasaporte COVID-19 en la red social Twitter mediante una extracción y análisis de comentarios y qué beneficios y perjuicios consideran. Los resultados indicaron actitudes favorables en la mayoría de la muestra (61%). Respecto a los inconvenientes, el más frecuente fue la falta de consenso entre los países para establecer un certificado estándar, seguido de aspectos éticos como la brecha digital, privacidad, libertades personales y derechos humanos. Por el lado de las ventajas, las menores restricciones para poder viajar supusieron el argumento más frecuente (29,5% de los tweets) y los beneficios sociales y económicos, ambos en torno al 4%.

Otro estudio, realizado mediante entrevistas a grupos focales en Suiza (Fagnoli et al., 2021), concluyó que las desventajas superaban a los beneficios. Los detractores del pasaporte coincidían en que la inmunidad era limitada temporalmente y que el virus puede mutar, careciendo de valor el certificado, además de los aspectos éticos e igualitarios. Por el contrario, los argumentos favorables incidían en la necesidad de reactivar la economía y la normalidad.

Una encuesta realizada por European Travel Commission (2022), realizó una comparativa entre países europeos sobre la aceptación o rechazo del pasaporte COVID-19. Como se observa en los datos de la tabla, en el total de la muestra más de la mitad de los encuestados se muestra favorable al pasaporte, mientras que el resto se divide prácticamente en partes iguales entre contrarios y neutrales. España es el país que mayor grado de aceptación presenta (67,90%).

País	De acuerdo	En contra	Neutral
España	67,90%	13,20%	19%
Polonia	59,40%	16,80%	23,80%
Italia	56,30%	15,60%	28,10%
Alemania	54,10%	21,90%	24%
Francia	52,40%	24,10%	23,50%
Austria	51,30%	28,50%	20,20%
Bélgica	47,20%	32%	20,80%
Países Bajos	43,60%	33,20%	23,20%
Reino Unido	43,30%	29,70%	26,90%
Suiza	42,40%	35,20%	22,40%
Total de la muestra	51,70%	24,70%	23,60%

Tabla 2. Aceptación del pasaporte COVID-19 según el país de la UE

Fuente: European Travel Commission (2022)

Según Fagnoli et al. (2021), la confiabilidad de los datos científicos respecto a la inmunidad es una de las mayores preocupaciones, lo cual incide en el pasaporte

COVID-19, pues los individuos pueden reinfectarse y el pasaporte perdería aceptación.

Una de las principales amenazas para la salud mundial es la no vacunación. Para mejorar la actitud frente a la vacuna Greyling y Rossouw (2022) señalan la necesidad de informar más acerca de su seguridad y hacerla más accesible para todos los países mediante un estudio sobre la tendencia de las actitudes positivas hacia las vacunas a lo largo de un período de tiempo partiendo de una extracción de tweets en tiempo real que fue analizada con procesamiento de lenguaje natural (NLP). La mayoría de los tweets positivos estuvieron relacionados con recibir la vacuna y en una nube de las palabras con mayor peso en estos tweets destacan *Good, Scientific, Hope, Safe, Happy, See, COVID-19*. En cambio, para los tweets negativos las palabras más destacadas en la nube de palabras fueron *COVID-19, Death, Die, Government, Pandemic, Dying, Out*, por lo que se aprecia temor a los efectos adversos de la vacuna e inconformidad con el gobierno, algo con lo que concuerda con otro artículo de similares características (Thelwall et al., 2021) centrado en los motivos por los que genera dudas la vacunación y en el cual se indica, también tras analizar una muestra de tweets, que las teorías conspirativas y la velocidad del desarrollo de la vacuna y su seguridad eran las mayores preocupaciones entre la población.

A continuación, se muestra en la siguiente tabla un resumen de las ventajas e inconvenientes más significativas del pasaporte COVID-19:

Ventajas	Inconvenientes
<p>Menores restricciones para poder viajar</p> <p>Beneficios sociales y económicos</p>	<p>Falta de consenso entre los países para establecer un certificado estándar</p> <p>Brecha digital</p> <p>Privacidad</p> <p>Libertades personales</p> <p>Derechos humanos</p> <p>Dudas sobre la inmunidad</p>

Tabla 3. Resumen de las ventajas e inconvenientes identificadas del pasaporte COVID-19

Fuente: Khan et al. (2022)

Por otra parte, para comprobar la evolución del interés sobre el pasaporte COVID-19, se ha procedido a consultar en Google Trends el interés a lo largo del tiempo para los términos de búsqueda “EU Digital COVID Certificate” en todo el mundo y “Certificado COVID” y “Pasaporte COVID” para España, obteniendo los siguientes resultados:

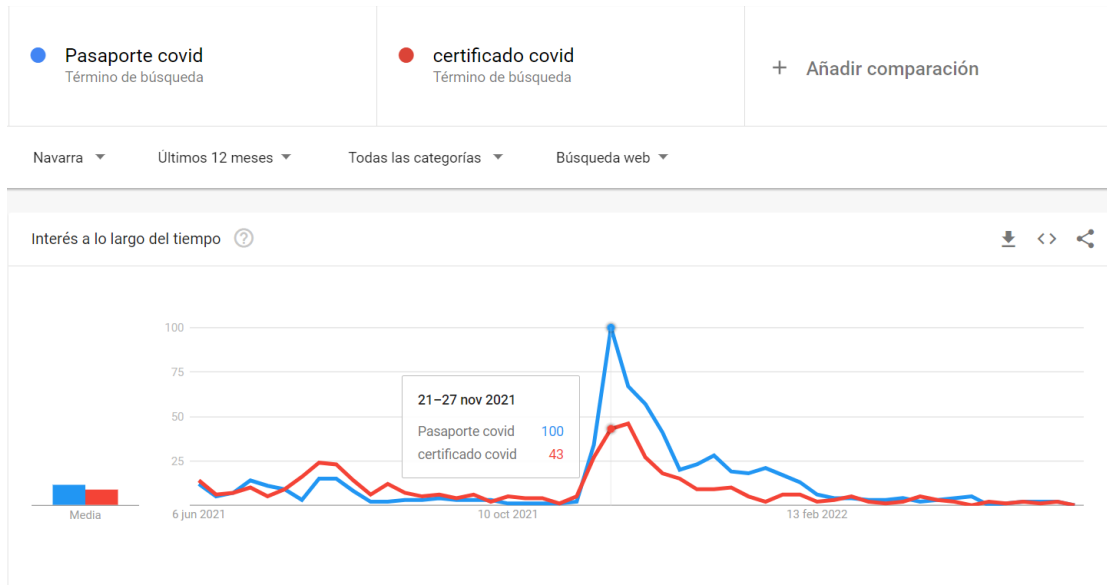


Figura 2. Interés del pasaporte COVID-19 en Google Trends (España)

Como se aprecia en la figura anterior, en España en el horizonte temporal del 1 de junio de 2021 hasta 1 de junio de 2022, el pico de mayor interés se alcanzó a finales de noviembre de 2021 para los dos términos de búsqueda, que además presentan una fuerte correlación a lo largo del tiempo. Este hecho coincide temporalmente con la entrada en vigor del pasaporte en varias comunidades autónomas de España.



Figura 3. Interés del pasaporte COVID-19 en Google Trends (Internacional)

En el plano internacional, se aprecia en la Figura 3 que el pico de mayor interés se alcanzó entre el 18 y el 24 de julio de 2021, en plena campaña de vacaciones estivales y un mes después de la aprobación del Reglamento (UE) 2021/953 del Parlamento Europeo y del Consejo de 14 de junio de 2021 relativo a un marco para la expedición, verificación y aceptación de certificados COVID-19 interoperables de vacunación, de prueba diagnóstica y de recuperación (certificado COVID digital de la UE) a fin de facilitar la libre circulación durante la pandemia de COVID-19.

2.1.4 Impacto del pasaporte COVID-19 en sectores económicos

Según una estimación de Olliu-Barton (2022), el PIB se habría reducido de no existir los pasaportes COVID-19.

Uno de los sectores más afectados económicamente por el impacto de la pandemia y sus limitaciones es el turístico, englobando compañías de transporte y de hostelería. Este sector se encuentra estrechamente vinculado con el pasaporte COVID-19 debido a que éste se trata fundamentalmente de una medida desarrollada para facilitar la movilidad de las personas garantizando un menor riesgo de infección al acreditar la vacunación o resultados de tests. Las restricciones de movilidad y la preocupación por el riesgo de salud percibido por parte de los clientes han condicionado las decisiones en este sector tanto por parte de las empresas como por parte de sus clientes. Un estudio de Shin H et al. (2021) comprobó que la implantación del pasaporte COVID-19 por parte de las aerolíneas tiene un impacto positivo en su valor de mercado en bolsa y que, tras la implantación, la intención de reserva de los pasajeros incrementó significativamente cobrando mucha mayor importancia que el resto de las medidas de reducción del riesgo contra la COVID-19 a la hora de realizar una reserva. El citado artículo puntualiza que se enfoca en los viajes de ocio y una línea de investigación futura se centraría en investigar la influencia según el tipo de viaje (de negocios o turísticos), su duración o destino (nacionales o internacionales), u otros factores.

2.1.5 Impacto del pasaporte COVID-19 en la decisión de aceptar las vacunas

El pasaporte COVID-19 puede suponer cierta presión para aceptar la vacunación dado que gran número de actividades cotidianas, o incluso el uso de medios de transporte como se ha indicado en el apartado anterior con el caso de las aerolíneas, son inaccesibles sin disponer del certificado, lo que puede suponer un aislamiento social

del individuo en caso de no disponer del mismo. Una encuesta realizada en Reino Unido (de Figueiredo et al., 2021) evalúa si la adopción del pasaporte fomenta la vacunación o si la desalienta entre la población que aún no ha recibido la pauta completa de la vacuna. Como conclusiones obtiene que la implantación del pasaporte propicia generalmente mayor disposición a la vacunación entre los individuos, aunque de un modo muy polarizado, pues quienes ya tenían la firme intención de recibir otra dosis de la vacuna refuerzan aún más su postura, mientras que quienes se mostraban preocupados por la vacunación tienden a disminuir su intención de vacunarse al sentirse presionados por el pasaporte.

Según Oliu-Barton (2022), los pasaportes COVID-19 han sido una eficaz medida para evitar defunciones y evitar el colapso de los hospitales, como estima en su estudio en el que cuantifica en alrededor de 4.000 muertes evitadas en Francia (un 32% más de las registradas desde la fecha de entrada en vigor del pasaporte hasta final de 2021), 1.100 en Alemania (6% más) y 1.300 en Italia (14% más). Asimismo, estima que se ha aumentado la tasa de vacunación en un 13% de la población en Francia, 6,2% en Alemania y 9,7% en Italia.

2.1.6 Consideraciones éticas

Algunos argumentos éticos en detrimento de la implantación del pasaporte COVID-19 se centran en las desigualdades entre países que se traducen en diferentes ritmos de vacunación masiva y consecuentemente en distintos criterios para la movilidad internacional, o desde una perspectiva más individualista, las limitaciones para trabajadores cuyo empleo requiere presencialidad respecto a quienes pueden teletrabajar. (Brown et al., 2020).

Como justificaciones se encuentran el principio ético de salud de la menor infracción, que consiste en que la alternativa menos transgresora para las libertades individuales de las disponibles para lograr un objetivo de salud pública es la que debe elegirse; las ventajas para la sociedad derivadas de su aplicación en lo que se refiere a la reanudación de actividades con mayor normalidad y el menor riesgo de infección al ser los portadores personas con un mayor nivel de inmunidad debido a la vacuna o a una reciente contracción de la enfermedad. (Voo et al., 2021).

En la siguiente tabla 3 “Resumen de referencias del marco teórico sobre opiniones e impacto del pasaporte COVID-19”, se enumeran a modo de resumen los artículos que, habiendo sido incluidos en el marco teórico, presentan similitud con el objeto de estudio y/o la metodología empleada en este trabajo.

Título	Autor(es)	Resumen	Metodología
Conflicting attitudes: Analyzing social media data to understand the early discourse on COVID-19 passports	Khan et al. (2022)	Análisis de la opinión sobre el pasaporte COVID-19 a partir de comentarios de Twitter	Sentiment Analysis
Acceptability of COVID-19 Certificates: A Qualitative Study in Geneva, Switzerland, in 2020	Fagnoli et al. (2021)	Análisis cualitativo de la aceptación del pasaporte COVID-19 en Suiza	Entrevistas a grupos focales
Positive attitudes towards COVID-19 vaccines: A cross-country analysis	Greyling y Rossouw (2022)	Estudio de las opiniones positivas hacia la vacunación en diferentes países	Sentiment Analysis
Covid-19 vaccine hesitancy on English-language Twitter	Thelwall et al. (2021)	Motivos por los que la vacunación genera preocupación	Sentiment Analysis
The Impact of COVID-19 Vaccine Passport on Air Travelers' Booking Decision and Companies' Financial Value	Shin H et al. (2021)	Análisis del impacto financiero de la implantación del pasaporte COVID-19 en aerolíneas	Análisis de datos cuantitativos
The potential impact of vaccine passports on inclination to accept COVID-19 vaccinations in the United Kingdom: Evidence from a large cross-sectional survey and modeling study	de Figueiredo A et al. (2021)	Estudio sobre si la adopción del pasaporte COVID-19 fomenta o desalienta la vacunación	Encuesta
El pasaporte COVID reduce miles de muertes: un análisis con datos de Francia, Italia y Alemania	Oliu-Barton (2022)	Estimación de las defunciones evitadas tras la adopción del pasaporte COVID-19	Análisis de datos cuantitativos

Tabla 4. Resumen de referencias del marco teórico sobre opiniones e impacto del pasaporte COVID-19

2.2 REVISIÓN DE LITERATURA DE PUBLICACIONES ACERCA DE DATA MINING Y SENTIMENT ANALYSIS

2.2.1 Conceptos

Natural Language Processing (NLP) o procesamiento de lenguaje natural son técnicas que permiten procesar a las máquinas el lenguaje humano oral mediante el reconocimiento de voz y textual, a través del procesamiento de texto. Presenta diferentes técnicas, las más frecuentes son (Lopez-Martinez y Sierra, 2020):

- *Identification of named entities* o identificación de entidades nombradas: extrae nombres de personas, instituciones, ciudades y otras localizaciones...
- Filtrado y clasificación de documentos
- Generación de resúmenes automáticos
- *Topics Analysis* o análisis de topics
- Extracción de información
- Análisis de sentimiento
- Minería de opinión u *Opinion Mining*
- Monitoreo de la opinión en redes sociales
- Corrección ortográfica y gramatical
- Búsqueda inteligente y optimizada
- Sistemas de respuesta automática
- Asistentes personales
- Traducción automática
- Reconocimiento de voz
- Sistemas de texto a voz
- Sistemas de diálogo
- Etc.

Sentiment Analysis o Análisis de Sentimiento se define como una metodología desarrollada a partir de *Machine Learning* – aprendizaje automático – mediante la que se analizan textos otorgando una valoración positiva, neutra o negativa al contenido en función de un baremo generado a partir de las connotaciones de las palabras que lo componen. (Saura et al., 2018). Por tanto, identifica y extrae información subjetiva con el propósito de determinar la actitud sobre cualquier asunto de interés a partir del contenido incluido en el texto.

Las palabras con connotaciones positivas o negativas pueden ser extraídas de una muestra del propio texto que se pretende analizar y que pasarían a formar parte de una bolsa de palabras de entrenamiento del algoritmo, lo cual sería un enfoque de aprendizaje automático, o bien estar ya previamente definidas en un diccionario genérico, lo que se denomina enfoque basado en el léxico y la precisión depende del tamaño del diccionario. (Kolchyna et al., 2015). Otro enfoque alternativo es el híbrido, que combina los dos anteriores. (Thakkar y Patel, 2015).

Para analizar el sentimiento es necesario previamente extraer los datos, es decir, realizar el proceso de *Data Mining* o Minería de datos, que se define como el proceso de extracción de datos útiles de un gran conjunto de datos utilizando herramientas estadísticas y de *Machine Learning*, que identifican patrones en las relaciones de los

datos. Dentro de *Data Mining*, la subcategoría que trata específicamente sobre la extracción de contenido textual se denomina *Text Mining* o minería de texto, cuyas principales utilidades prácticas son el resumen automático de texto, el reconocimiento de entidades nombradas o *Identification of named entities* (extrae nombres de personas, instituciones, ciudades y otras localizaciones...) y el análisis de sentimiento. (Tripathi et al., 2016).

Existen ciertos aspectos en los que se requiere un mayor grado de avance. Mohammad (2017) expone como desafíos del análisis de sentimiento para lograr una mayor precisión hacer especial énfasis en el desarrollo del análisis de las preguntas retóricas, ironía, contextos concretos que pueden alterar la predicción y traducciones en análisis multilingües.

2.2.2 Trabajos previos y herramientas

El análisis de sentimiento es un tema relativamente novedoso en el ámbito de la investigación que presenta un interés creciente entre los investigadores, lo que está dando lugar a un notable número de artículos publicados sobre esta materia. Desde la última década y fundamentalmente desde 2015, se observa un crecimiento exponencial en el número de publicaciones según el estudio bibliométrico elaborado por Sanchez-Nunez P et al. (2020).

En lo que respecta a este trabajo, la revisión bibliográfica de este apartado se centra en las publicaciones sobre análisis de sentimientos enmarcadas en el contexto del COVID-19 o las que utilizan el mismo software que este estudio, RapidMiner.

Previamente se encuentra una publicación en la que se extrae una muestra de comentarios sobre la vacuna del COVID-19 en la red social Twitter en hashtags específicos de Filipinas (Villavicencio et al., 2021). Este artículo utilizó como software RapidMiner, procediendo a una depuración del texto y al análisis de sentimiento utilizando el algoritmo de Naïve Bayes. Como resultados, de una muestra de 11.974 tweets, 993 fueron válidos para ser utilizados en el estudio, de los cuales el 83,38% mostraban un sentimiento positivo hacia las vacunas y entre los negativos analizaron el contenido y se centraban en los efectos adversos de la vacuna.

RapidMiner es un software libre desarrollado en el lenguaje de programación Java que proporciona un entorno integrado para *Data Mining* y *Natural Language Processing*. Se trata de un sencillo e intuitivo programa con el que se facilita sobremanera el diseño del proceso de minería de datos y de procesamiento de lenguaje natural como análisis de sentimiento o clasificación por *topics*, pues los operadores están

prediseñados y tan solo es necesario realizar ajustes de menor complejidad para adaptarlo a los requerimientos de cada proyecto. Permite además la representación gráfica de los resultados.

Otros ejemplos de softwares alternativos son KNIME, de características similares a RapidMiner, o el uso de librerías programadas con los lenguajes de programación como R o Python, cuyo uso requiere una mayor complejidad. Como alternativas a nivel usuario también existen opciones online como Monkey Learn, Google Cloud Natural Language, IBM Watson Natural Language Understanding, Amazon Comprehend y Microsoft LUIS (Language Understanding Intelligence Service), propuestas enfocadas a simplificar estas tareas en el ámbito empresarial.

En la siguiente tabla se enumeran publicaciones con una metodología similar a la de este estudio.

Título	Autor(es)	Resumen	Software
Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes	Villavicencio et al. (2021)	Análisis de la opinión sobre la vacuna del COVID-19 a partir de comentarios de Twitter en Filipinas.	RapidMiner
Analysis of Sentiments for Sports data using RapidMiner	Pawar et al. (2018)	Realiza un proceso de análisis de sentimiento en RapidMiner a partir de tweets sobre cricket.	RapidMiner
Exploring preferences and sustainable attitudes of Airbnb green users in the review comments and ratings: a text mining approach	Serrano et al. (2020)	Análisis de sentimiento de comentarios de usuarios de la plataforma Airbnb. Incluye: términos más frecuentes y agrupación por temas y clasificación por emociones básicas.	Leximancer
Post COVID-19 Twitter user's Emotions Classification using Deep Learning Techniques in India	Tatineni et al. (2021)	Medición automatizada de las emociones sobre la decisión del gobierno de extender el confinamiento en India a partir de tweets.	IBM Watson Tone Analyzer
Sentiment Analysis and Classification of Tweets Using Data Mining	Shoeb y Ahmed (2017)	Realiza una demostración de análisis de sentimiento en RapidMiner con enfoque de aprendizaje automático.	RapidMiner
Sentiment Analysis of English Tweets Using RapidMiner	Tripathi et al. (2016)	Demostración de análisis de sentimiento en RapidMiner y medición de precisión comparando los clasificadores Naïve Bayes y K-NN Classifier.	RapidMiner
Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes	Bayhaqy et al. (2018)	Comparación de los clasificadores Decision Tree, K-NN y Naïve Bayes Classifier para encontrar la mejor precisión.	RapidMiner
Are Black Friday Deals Worth It?	Saura et al.	Identifica ofertas de Black Friday en Twitter	Python

Mining Twitter Users' Sentiment and Behavior Response	(2019)	mediante el modelo de asignación de Dirichlet, para realizar un análisis de sentimiento y de <i>Text Mining</i> para identificar ideas de estrategias de promoción y marketing de otras empresas	
---	--------	--	--

Tabla 5. Recopilación de publicaciones con metodología similar a este estudio

3 METODOLOGÍA

El proceso para elaborar la investigación ha comenzado mediante la realización de una revisión de literatura en las principales bases de datos (Web of Science, Scopus) en las que se han consultado publicaciones relacionadas con el impacto económico y social del COVID-19 y la vacunación y con *Natural Language Processing*, *Sentiment Analysis* y *Data Mining*, de modo que la estructura del trabajo en cuanto a revisión bibliográfica queda subdividida en esas dos áreas de investigación para contextualizar el trabajo y aportar la información necesaria para comprender la propuesta práctica en la que se extrae una muestra de comentarios publicados en vídeos relativos a la implantación del pasaporte COVID-19 en los canales de YouTube de los principales medios británicos. El idioma utilizado en el análisis de los comentarios es inglés debido al mayor avance de los softwares en esta lengua. Se ha optado por la plataforma de YouTube debido a que los comentarios están segmentados, es decir, en un vídeo de la temática del pasaporte COVID-19 los comentarios son sobre dicha temática y porque es la segunda red social con mayor número de usuarios activos en 2022 con 2.562 millones de usuarios en todo el mundo según Statista (2022).

Siguiendo las recomendaciones de Palos et al. (2022), se han seleccionado los vídeos en base a la temática, la popularidad (tomando de referencia el número de suscriptores del canal y las visitas del vídeo), la cantidad de comentarios y el idioma. El conjunto final está compuesto por un total de 11 vídeos. Los criterios de selección se resumen en la siguiente tabla 5 “Criterios de selección de vídeos”:

Criterio	Requerimientos
Temática	Noticias sobre pasaporte COVID-19
Popularidad	Medios de prensa reputados y en el top10 de los que cuentan con más suscriptores en YouTube.
Cantidad de comentarios	>250
Idioma del vídeo	Inglés

Tabla 6. Criterios de selección de vídeos

La mencionada extracción de comentarios se realiza con el software MAXQDA Analytics Pro 2022, debido a que este programa incorpora una funcionalidad de extracción de comentarios de YouTube, y los datos son tratados en Microsoft Excel para ajustar el formato como paso previo al análisis con RapidMiner de los de primer nivel, software en el que se realiza un proceso con operadores para obtener los resultados:

- Nube y listado de palabras con mayor peso dentro del texto
- Clasificación de *topics*
- Análisis de sentimiento categorizando en positivos, neutros y negativos los comentarios.

Se ha seleccionado RapidMiner porque se trata de un software que incorpora las herramientas y funcionalidades requeridas para los objetivos de este trabajo de una forma fácilmente accesible e intuitiva, a diferencia de otros softwares en los que hay que ejecutar comandos y códigos de programación como se detallará en el epígrafe de revisión bibliográfica.

Para el análisis de sentimiento en RapidMiner se ha utilizado concretamente el algoritmo Vader de entre las opciones disponibles:

- Sentiwordnet: incluye tres puntajes entre 0 y 1 de negatividad, positividad y neutralidad para cada palabra incluida en el léxico. Las puntuaciones, a diferencia de otros léxicos que fueron supervisados totalmente por humanos, se calcularon utilizando algoritmos semisupervisados. (Hutto y Gilbert, 2014).
- Vader (*Valence Aware Dictionary for sEntiment Reasoning*): es un algoritmo que clasifica según positividad, negatividad, neutralidad e intensidad de dicha emoción e incluye aproximadamente 7.520 elementos en el léxico y ha sido supervisado por humanos. (Hutto y Gilbert, 2014; Yu, 2022).
- Aylie
- Meaning cloud

Los resultados obtenidos son interpretados y finalmente se llegan a las conclusiones y a la discusión de estos resultados y de los de otros estudios similares.

A continuación, se detalla a modo de esquema el proceso desarrollado para la elaboración de este trabajo.

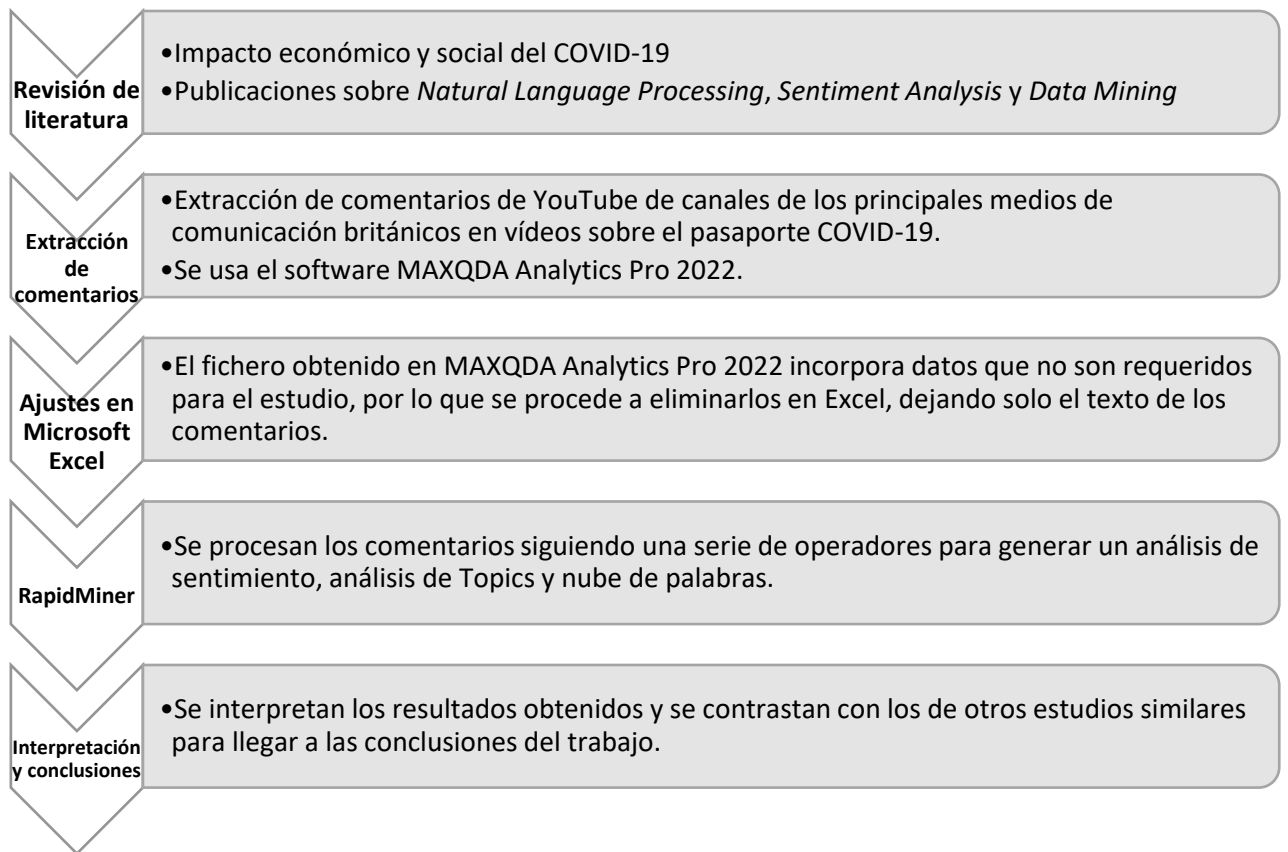


Figura 4. Esquema del proceso de la metodología

4 ESTUDIO EMPÍRICO Y ANÁLISIS DE DATOS

4.1 PROCESO DE EXTRACCIÓN, PREDICCIÓN Y ANÁLISIS DE LOS DATOS

4.1.1 Extracción de comentarios

El estudio realizado en este trabajo incluye un proceso de extracción de comentarios realizado con la herramienta MAXQDA Analytics Pro 2022, posteriormente un tratamiento de estos datos en Microsoft Excel para adaptarlos a los requerimientos del tercer programa utilizado, RapidMiner.

MAXQDA es un software de análisis de datos cualitativos e investigación de métodos mixtos. Permite analizar textos, imágenes, audios o archivos de vídeo, páginas web, tuits, discursos, grupos focales, encuestas, etc. bajo un interfaz intuitiva. (MAXQDA, n.d.).

Se ha obtenido un total de 24.182 comentarios procedentes de doce vídeos relacionados con el pasaporte COVID-19 en los canales oficiales de reputados medios de prensa británicos en la plataforma YouTube. La razón de esta elección se debe fundamentalmente a criterios técnicos, pues en *Natural Language Processing* la lengua inglesa es la que cuenta con un mayor grado de desarrollo.

Los vídeos concretos con sus respectivos medios de prensa son los que aparecen a continuación:

Medio	Título	Enlace	Suscriptores	Comentarios
BBCNews	UK considers "vaccine passports" to prove COVID-19 protection - BBC News	https://www.youtube.com/watch?v=jfT_t3fhFss	12,5 M	5159
BBCNews	Italy to require all workers to show COVID-19 pass	https://www.youtube.com/watch?v=aqwx2cRBzKs	12,5 M	1378
BBCNews	Will COVID-19 passports be extended across England? - BBC Newsnight	https://www.youtube.com/watch?v=DKIf4svKyKo	12,5 M	1564
BBCNews	France rolls out COVID-19 health pass as cases soar	https://www.youtube.com/watch?v=6fqW1N0vKa8	12,5 M	419
BBCNews	EU plans rollout of travel certificate before summer	https://www.youtube.com/watch?v=OmzPikZ8Dlk	12,5 M	1827
The Sun	EU prepares to open vaccine passport scheme to UK and US visitors	https://www.youtube.com/watch?v=CC9RM9Gv5CE	2,98 M	272

The Sun	COVID-19 UK: Football fans WILL need vaccine passports to get into large venues, says Nadhim Zahawi	https://www.youtube.com/watch?v=iUKkWqIhAKo	2,98 M	2495
The Sun	COVID-19 UK vaccine passports not ruled out for pubs & restaurants	https://www.youtube.com/watch?v=-JDI4maXL7s	2,98 M	2813
The Sun	EU 'to introduce COVID-19 vaccine passports' allowing Brits to travel this summer	https://www.youtube.com/watch?v=-11bCY7Nfvn0	2,98 M	2432
Daily Mail	UK COVID-19: Boris Johnson says vaccine passports will be needed to go on holiday	https://www.youtube.com/watch?v=-AA2SPmWua0	1,87 M	4098
Daily Mail	UK COVID-19: Vaccine passports 'weapons to keep COVID-19 at bay', says Business minister	https://www.youtube.com/watch?v=-fuGF5TeBNaM	1,87 M	1725

Tabla 7. Fuentes utilizadas para la extracción de comentarios

Para determinar qué proporción de comentarios se han extraído según su fuente se ha elaborado el siguiente gráfico 2 “Proporción de comentarios según fuente”, en el que se observa que BBCNews aporta el 43% de los comentarios, TheSun el 33% y Daily Mail el 24%.

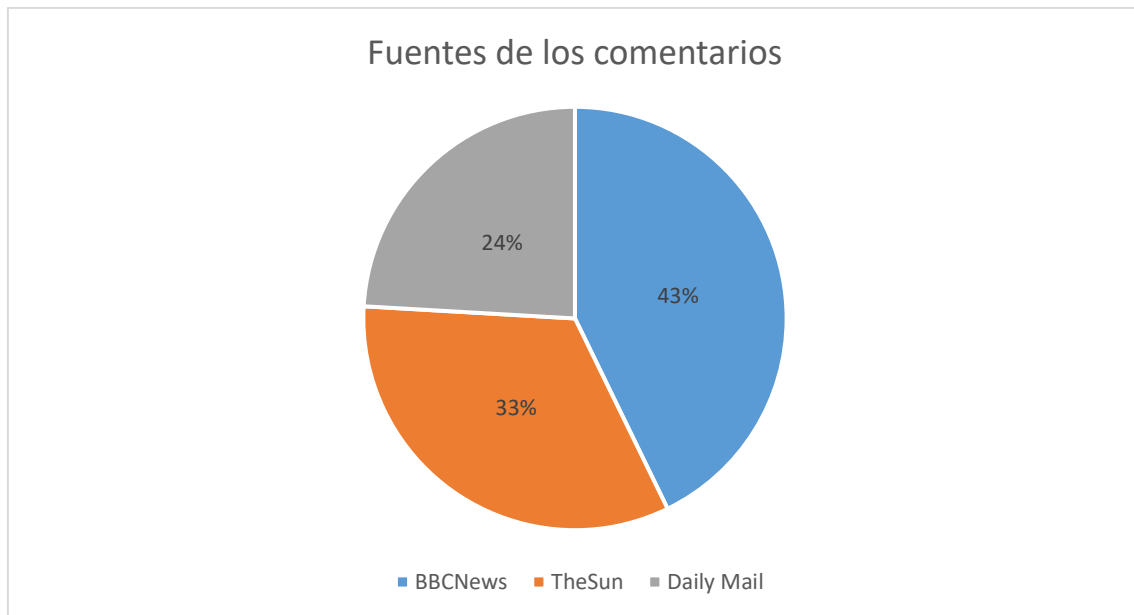


Gráfico 1. Proporción de comentarios según fuente

La siguiente imagen es la interfaz del software MAXQDA una vez realizada la extracción de los comentarios, como se puede observar no se dispone solo del comentario en sí, sino también de información complementaria (autor, URL, etc.), que posteriormente habrá de ser tratada en Microsoft Excel para depurar el fichero.

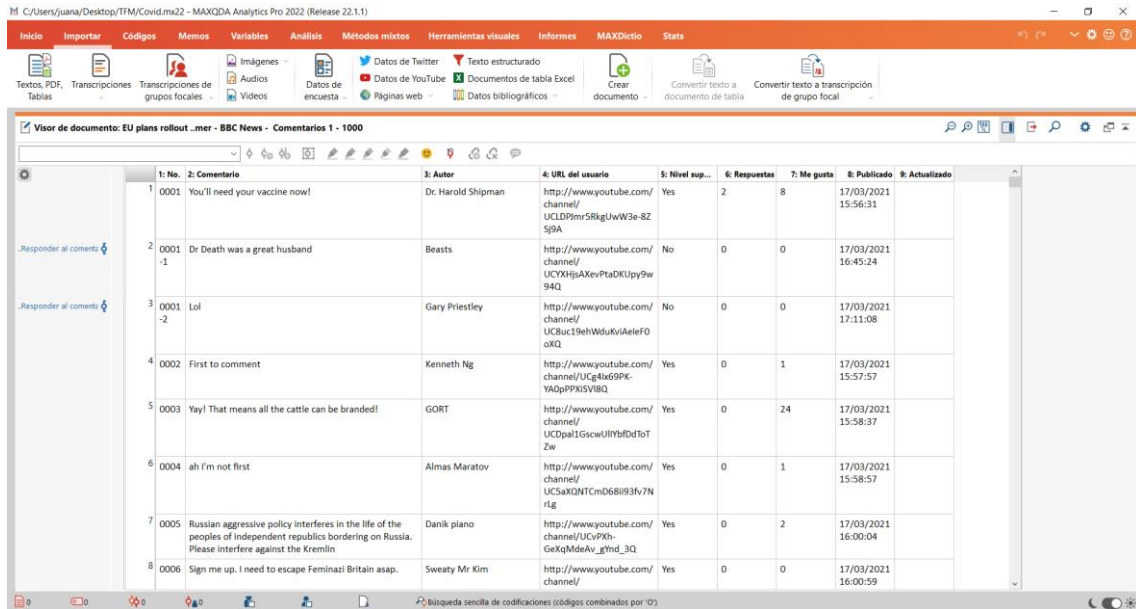


Figura 5. Interfaz de MAXQDA Analytics Pro 2022 con algunos de los comentarios extraídos

4.1.2 Depuración del archivo en Microsoft Excel

En MAXQDA Analytics Pro 2022 se exportan los comentarios tras la extracción en ficheros de 1.000 comentarios como máximo, por lo que se han unificado todos ellos en Microsoft Excel como procedimiento previo a su importación en RapidMiner. La siguiente captura muestra el fichero de origen.

Título vídeo	Medio	1: No.	2: Comentario	3: Autor	4: URL del usuario
Covid-19 EU prep..nd US visitors	The Sun	001	You'll need a vaccine card for everything now.	MAGLÉ MARK	http://www.youtube.com/channel/UC_yi9KPDGQIXWD6QmBlueaQ
Covid-19 EU prep..nd US visitors	The Sun	001-1	yep. the CONvid was always about this; there never was any virus.	Zane	http://www.youtube.com/channel/UCGjep46LNOzkiuyC6tEfwg
Covid-19 EU prep..nd US visitors	The Sun	001-2	@Elsa and it's never going to stop. People want it and vote for it. Are we the strange ones being de	Ste Ritson	http://www.youtube.com/channel/UCNjSRv3w6YCXU14zBEEnSng
Covid-19 EU prep..nd US visitors	The Sun	001-3	@Elsa it is. A lot of people just want to see everyone burn including the world so they go along with	Ste Ritson	http://www.youtube.com/channel/UCNjSRv3w6YCXU14zBEEnSng
Covid-19 EU prep..nd US visitors	The Sun	001-4	@Ste Ritson I just dont get why there are low ranking Satanists, they achieve nothing and do nothing	Elsa	http://www.youtube.com/channel/UCNjSRv3w6YCXU14zBEEnSng
Covid-19 EU prep..nd US visitors	The Sun	001-5	@Elsa exactly. The lower class Satanists just want to see everyone suffer, unfortunately there is	Ste Ritson	http://www.youtube.com/channel/UCNjSRv3w6YCXU14zBEEnSng
Covid-19 EU prep..nd US visitors	The Sun	002	worst news ever. Please do not comply at all. Hold off. Terrible future if you agree to vaccine passp	Randamu Shō	http://www.youtube.com/channel/UCdOcPsrGbuUW4o8Y0gTwkukw
Covid-19 EU prep..nd US visitors	The Sun	002-01	I was trying to convince my brother not to vaccinate but he thinks he is the smart one and thinks he	JC	http://www.youtube.com/channel/UCCy_qLy8ob7yPHkM7mx00Q
Covid-19 EU prep..nd US visitors	The Sun	002-02	@JC well judging by your comment your brother is the smart one!	captain sensible	http://www.youtube.com/channel/UCCb2H1iIVtm4gmSxj-905A
Covid-19 EU prep..nd US visitors	The Sun	002-03	@Gorgeous George Pharma as in the righteous Johnson and Johnson. Who in court had documents	Coco S	http://www.youtube.com/channel/UC-AnLKR5q3329c4E8A11g
Covid-19 EU prep..nd US visitors	The Sun	002-04	@John Titor David Icke's job is to speak 70% truth, mixed in with 30% insanity, to make people	like y Mr. Anonymous	http://www.youtube.com/channel/UcOLintTElynQyd55BtxxTHA
Covid-19 EU prep..nd US visitors	The Sun	002-05	@JC why are financial institutions still happy to give people 25 / 30 mortgages in the full knowledg	captain sensible	http://www.youtube.com/channel/UCCb2H1iIVtm4gmSxj-905A
Covid-19 EU prep..nd US visitors	The Sun	002-06	@John Titor dude stop embarrassing yourself with airing your dreadful opinions in a youtube	comme Randamu Shō	http://www.youtube.com/channel/UCdOcPsrGbuUW4o8Y0gTwkukw
Covid-19 EU prep..nd US visitors	The Sun	002-07	@captain sensible you mean banks? He he. Look up NESARA/GESARA. The banking system is	going JC	http://www.youtube.com/channel/UCCy_qLy8ob7yPHkM7mx00Q
Covid-19 EU prep..nd US visitors	The Sun	002-08	@JC well that's all very interesting but it doesn't answer the question does it - why would	banks le captain sensible	http://www.youtube.com/channel/UCCb2H1iIVtm4gmSxj-905A
Covid-19 EU prep..nd US visitors	The Sun	002-09	@captain sensible why would you want an experimental drug for a virus which only kills	people abo Jamie C	http://www.youtube.com/channel/UcOfoUaOkdnIzAUgWjPj1A
Covid-19 EU prep..nd US visitors	The Sun	002-10	@John Titor my initial response serves aa an answer to what you've said here.	Mr. Anonymous	http://www.youtube.com/channel/UcOLintTElynQyd55BtxxTHA
Covid-19 EU prep..nd US visitors	The Sun	002-11	@John Titor lol weird guy	Randamu Shō	http://www.youtube.com/channel/UCdOcPsrGbuUW4o8Y0gTwkukw
Covid-19 EU prep..nd US visitors	The Sun	002-12	@John Titor You believe that out of ignorance, and it's just not true. Icke spoke about	Jim Savile's Mr. Anonymous	http://www.youtube.com/channel/UcOLintTElynQyd55BtxxTHA

Figura 6. Tabla de Microsoft Excel con los comentarios extraídos

4.1.3 Análisis usando *RapidMiner*

El análisis de los comentarios extraídos se ha realizado con el software de *Data Mining* y *Data Analysis* RapidMiner. Rapidminer es un software especializado en el análisis predictivo, inteligencia de negocio, *Machine learning*, minería de texto y minería de datos, el cual proporciona una sencilla forma de acceder a los operadores y procesarlos fácilmente (Pawar et al., 2018). La versión usada es RapidMiner Studio 9.10.

El objetivo de utilizar este software en este trabajo se basa en obtener una nube de palabras de la que interpretar cuáles son las palabras más frecuentes en la muestra de comentarios, una clasificación de *topics* que muestre cuáles son las preocupaciones y temas de mayor interés respecto a la implantación del pasaporte COVID-19 y un análisis de sentimiento para agrupar los comentarios en positivos, negativos y neutros y poder percibir la sensación general de la sociedad.

A continuación, se detallan los operadores incluidos en la interfaz de diseño de RapidMiner y su utilidad dentro del proceso que se aprecia en la siguiente imagen.

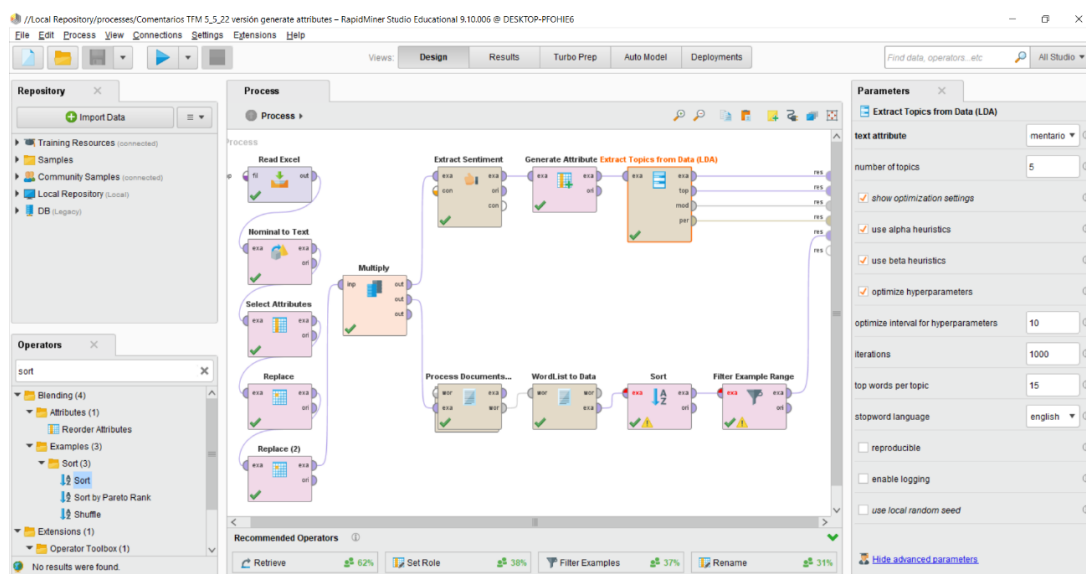


Figura 7. Interfaz de diseño del proceso en RapidMiner

En primer lugar, el operador *Read Microsoft Excel* importa el fichero de Microsoft Excel comentado en el anterior apartado en el que se encuentran los comentarios. Estos comentarios pasan por el operador *Nominal to Text*, que convierte los atributos nominales en atributos de cadena.

A continuación, *Select Attributes* selecciona la columna del fichero Microsoft Excel que se va a tener en cuenta para el análisis y se indica en el menú de parámetros, en este

caso la del texto de los comentarios. Las opciones son variadas, es posible seleccionar todo el archivo (*all*), solo un subconjunto (*subset*), solo valores numéricos (*numeric_value_filter*), entre otras. Seleccionando *subset* se ha seleccionado el atributo comentario, de forma que solo trabajará con dicha columna.

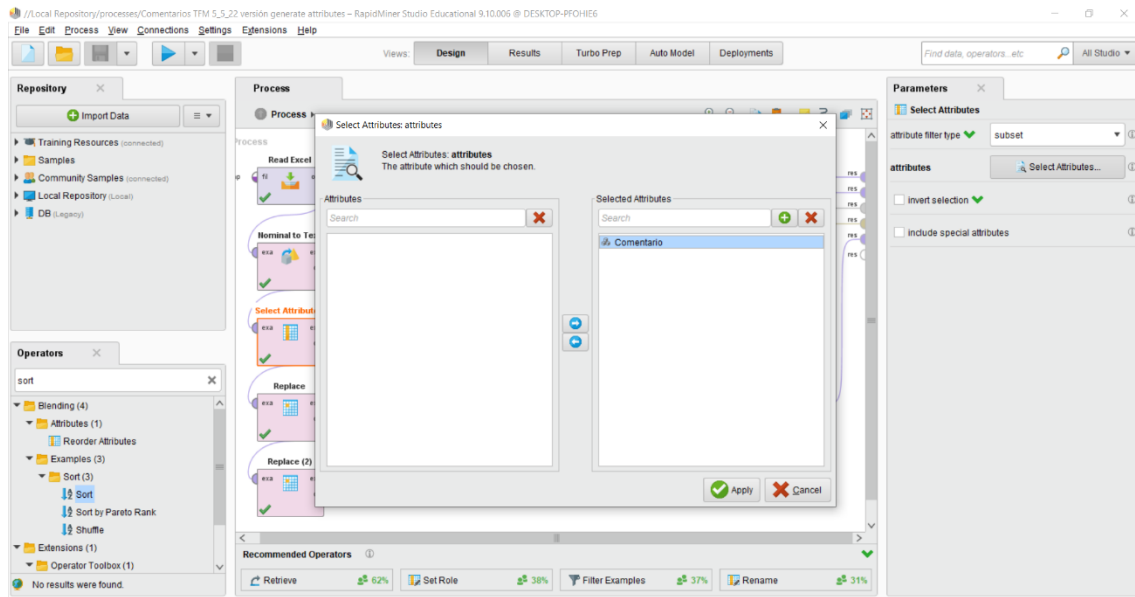


Figura 8. Selección de atributo de filtro en RapidMiner

Los operadores *Replace*, eliminan los caracteres no deseados para depurar el texto. Se ha introducido la siguiente cadena de caracteres para ser eliminados con el primero de los operadores `https?://[-a-zA-Z0-9+&@#/%?=-_(!:,;]*[-a-zA-Z0-9+&@#/%?=-_()]` y la siguiente con el segundo operador: `RT\s*@[^\:]*\s*[A-Za-z]+`

El siguiente operador, *Multiply*, se usa para multiplicar los puertos de entradas y salidas que conectan los distintos operadores y poder establecer un mayor número de conexiones. A partir de este conector, el proceso se divide en dos partes, cada una con sus correspondientes operadores:

- Análisis de sentimiento y clasificación por *topics*
- Generación de la nube de palabras

Análisis de sentimiento y clasificación por *topics*

Una vez depurado el texto, el operador *Extract Sentiment* analiza el sentimiento expresado en cada uno de los comentarios usando el modelo que se indique en parámetros entre las diferentes alternativas (*aylien*, *sentiwordnet*, *vader*, etc.). Se ha

seleccionado *Vader*. Grosso modo, la función que realiza es detectar las palabras relevantes independientes incluidas en cada comentario y asignarles una puntuación de positividad/negatividad previamente establecida en el diccionario. Hace un cómputo global de todas las palabras incluidas en el mismo comentario que puede dar como resultado un número negativo o positivo a partir del que se determina el sentimiento con el siguiente operador *Generate attributes* mediante una fórmula condicional que traduce estos valores positivos y negativos a las expresiones “positivo”, “negativo” y “neutro”.

La fórmula concretamente es la siguiente:

```
if(Score > 0 , "positive", if (Score < 0 , "negative",if (Score == 0 , "neutral", "")))
```

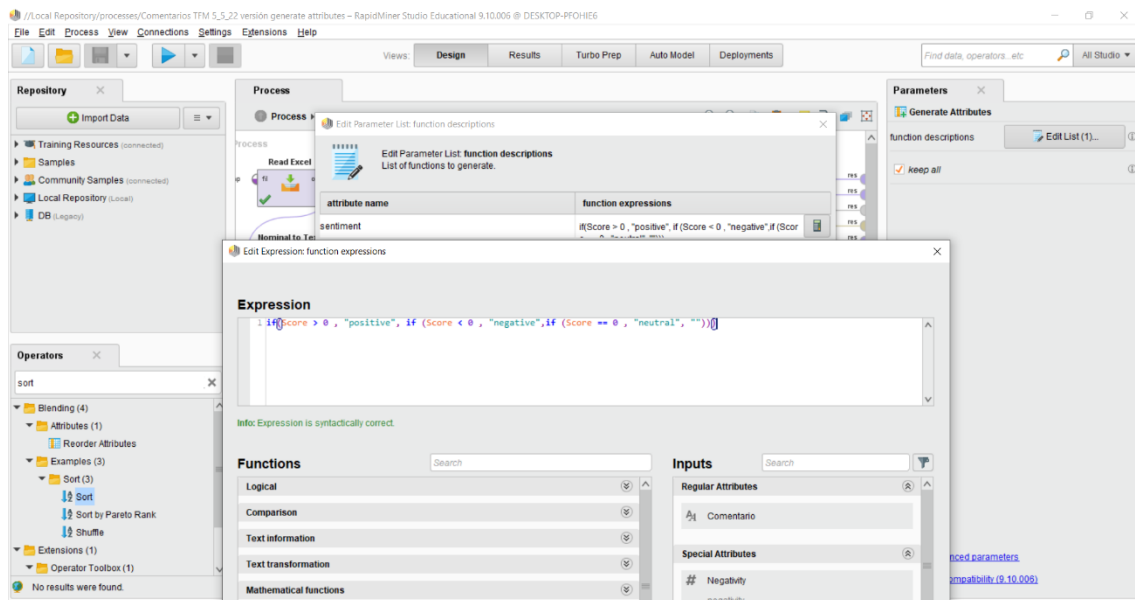


Figura 9. Operador Generate Attributes en RapidMiner

Este operador se encuentra conectado con el último operador del proceso *Extract Topics from Data (LDA)*, que categoriza y agrupa los comentarios en varios *topics* según su contenido y determina las palabras claves del *topic* según su importancia. El número de *topics* y de palabras claves deseadas se ajusta en el menú de parámetros.

La asignación latente de Dirichlet (LDA) es un modelo probabilístico para el análisis de *Topics* que, a partir del texto incluido en un documento, encuentra relaciones en su contenido y las agrupa por temas relacionados (Blei et al., 2003).

Generación de la nube de palabras

Partiendo de nuevo del operador *Multiply*, la otra parte del proceso es la generación de la nube de palabras. Para ello, se incluye el operador *Process Document from Data*, el cual contiene a su vez los operadores que se indican a continuación:

- *Tokenize*: divide el texto (cada comentario) en una secuencia de palabras independientes (*tokens*) siguiendo un criterio, en este caso la separación se realiza cuando aparece un espacio (*non letters*).
- *Transform Cases*: transforma cada *token* a unas condiciones especificadas. Se ha aplicado *transform to lower case* para homogeneizar todos ellos en minúsculas.
- *Filter Stopwords*: elimina las palabras que no aportan valor, tales como artículos, pronombres y otros.
- *Filter Tokens by Length*: elimina las palabras que no cumplen con una longitud mínima y máxima de caracteres, para generar así un filtro más selectivo y refinado. Se ha establecido como mínimo cuatro caracteres y como máximo veinticinco.

Se sucede con el operador *WordList to Data* que realiza el recuento de apariciones de cada palabra en todo el texto. Para ordenarlas en base al número de operaciones de mayor a menor se utiliza el operador *Sort* y para filtrar las primeras veinte palabras (la nube de palabras no puede incluir un número ilimitado de palabras) se utiliza *Filter Example Range*.

Para concluir, se ejecuta el proceso y se obtienen los resultados. Para obtener la nube de palabras es necesario realizar un paso adicional. En la pestaña de los resultados del operador *Filter Example Range* aparece una tabla, la que viene a continuación, con las palabras ordenadas por su frecuencia dentro del texto para la que hay que generar una visualización *worldcloud*.

Número	Palabra	En documentos	Total
1	vacuna	1678.0	2247.0
2	gente	1640.0	2234.0
3	covid	813.0	1084.0
4	pasaporte	776.0	894.0
5	pasaportes	664.0	746.0

6	gobierno	647.0	734.0
7	ir	558.0	633.0
8	tomar	524.0	597.0
9	querer	485.0	550.0
10	virus	482.0	639.0
11	vacunado	449.0	624.0
12	parar	428.0	489.0
13	tiempo	423.0	458.0
14	mundo	392.0	462.0
15	libertad	369.0	434.0
16	conspiración	358.0	404.0
17	vacunas	355.0	456.0
18	país	347.0	378.0
19	boris	346.0	377.0
20	conocer	343.0	392.0

Tabla 8. Frecuencia de palabras en RapidMiner

El gráfico 2 es la nube de palabras. En él se representan las palabras más repetidas dentro del texto, en ella destacan las palabras “vacuna”, “personas”, “pasaporte” y “gobierno”. Hay palabras como pasaporte que aparecen tanto en singular como en plural. Aparecen verbos de acción como ir y parar (lo que se puede asociar a parar la vacunación, parar el pasaporte COVID, entre otros, y se podrían analizar las correlaciones para saber si se refiere más al pasaporte o a la vacunación) y otras palabras que aportan mucho significado como conspiración, gobierno, tiempo y el nombre del primer ministro del Reino Unido “Boris”.



Gráfico 2. Nube de palabras en RapidMiner

4.2 INTERPRETACIÓN DE LOS RESULTADOS

4.2.1 Nube de palabras

Los resultados obtenidos en la nube de palabras y en la tabla de frecuencia de palabras indican, como era de esperar, una alta presencia de las palabras “*vacuna*”, “*COVID-19*” y “*passport*”, palabras que no aportan ningún significado más allá de la preocupación por la vacunación, tanto en términos de los posibles efectos secundarios como en términos de disponibilidad de dicha vacuna en todos los países. Pero también se encuentran otras palabras que sí otorgan una visión más clara de la sensación general de los internautas respecto a la implantación del pasaporte COVID-19. Son “*government*”, con un total de 734 apariciones; “*stop*”; “*freedom*”; “*conspiracy*”; “*boris*”; “*time*”; “*country*”.

Estas palabras hacen referencia a cuestiones éticas como la libertad, que ya se comentaba en la revisión bibliográfica como uno de los aspectos más contradictorios respecto del pasaporte, así como las diferencias entre países “*country*”. También la necesidad de “ganar tiempo” para “frenar” el virus, viéndolo por el lado positivo, o “stop COVID-19 Passport” si es en sentido detractor, algo que también cobra sentido al incluir palabras relacionadas con la política y sus líderes en Reino Unido, el gobierno y las teorías conspirativas.

4.2.2 Clasificación por *topics*

La categorización de los comentarios según los *topics* o temas autodetectados a partir del contenido muestra un elevado peso de los comentarios en los que las mayores

preocupaciones versan sobre la necesidad de conseguir el pasaporte COVID-19 y su relación con la vacuna, como se puede apreciar en el *topic* 1 “Opiniones generales sobre la obtención del pasaporte COVID-19 y la vacunación” cuyas palabras representativas más frecuentes son: “persona”, “vacunas”, “pasaporte”, “quieren”, “necesitar”, “gobierno. En la Tabla 9 “Pesos de las palabras del *Topic* 1”, las palabras “vacuna” y “gente” tienen un elevado peso que sobresalen del resto, como se puede comprobar a continuación:

Palabra	Peso
vacuna	1929.0
gente	1870.0
conseguir	924.0
pasaporte	799.0
necesitar	691.0
pasaportes	661.0
yendo	613.0
gobierno	579.0
gustar	563.0
tomar	547.0
querer	518.0
vacunado	512.0
covid	494.0
podría	482.0
virus	454.0

Tabla 9. Pesos de las palabras del *Topic* 1 “Opiniones generales sobre la obtención del pasaporte COVID-19 y la vacunación”

Algunas opiniones, que se exponen en las siguientes líneas, se centran fundamentalmente en que, a pesar de la vacuna, sigue existiendo riesgo de infección, por lo tanto, el pasaporte COVID-19, que acredita haberse vacunado carecería de sentido. Algunos incluso proponen la prueba de anticuerpos como acreditación alternativa:

“More lies from the government, 1 the vaccine doesn't stop you catching COVID-19. 2 the vaccine will not stop you passing it on. Explain why vaccine passports will help?”

“The only alternative is the 99% have to be repeatedly locked down because of the irresponsible 1%. Just ban them from public and let's all get on with our lives. Most of them are the deadweight on society anyway - who needs them?”

“Well this is very interesting ?? so you can still get the virus with the clot shot....So why passports? The CCP would be proud.”

“I just read the Pfizer vaccine documentation and it says taking the vaccine only stops you from getting sick from the virus, but doesn't stop you from catching or spreading it. So what's the point of a "vaccine passport" if you can still spread it after having the vaccine?”

El *topic 0* “Teoría de la conspiración” es el segundo con mayor importancia, en él se agrupan los comentarios que expresan opiniones relativas a la teoría de la conspiración. En la tabla 10 “Pesos de las palabras del Topic 0”, las palabras con mayor peso son “conspiración”, “teoría”, “Boris” (primer ministro de Reino Unido) o “derechos”.

Palabra	Peso
Conspiración	325.0
BBC	171.0
Boris	164.0
hombre	134.0
gustar	128.0
decir	109.0
tiempo	109.0
mundo	109.0
teoría	108.0
llamado	106.0
teóricos	100.0

bestia	98.0
derecho	97.0
señal	95.0
viniendo	95.0

Tabla 10. Pesos de las palabras del *Topic 0* “Teoría de la conspiración”

Los comentarios relativos a las restricciones de libertad como medidas para frenar el avance de la pandemia, así como la obligatoriedad de poseer pasaporte COVID-19 para acceder a determinados lugares, que contrasta con el derecho a la libre decisión de vacunación, son las opiniones más recurrentes entre los no partidarios.

“And the conspiracy theories keep turning out to be true, well how about that!”

“yet another restriction on our freedoms of choice movement & speech. another step closer to being a communist state. where will it end BORIS?”

“They say the vaccine isn’t compulsory but then slowly take our rights and freedoms until we have no option but to comply ! Its crazy we are all standing for it !!”

Algunos partidarios de la implantación del pasaporte optan por sacrificar esas libertades en aras de contener la propagación y consideran que los críticos deben basar sus argumentaciones no en los derechos humanos, sino en por qué se han implementado tales medidas:

“I don't understand the opposition to this. We're in the midst of a pandemic; the vaccines protect you from symptoms; being asymptomatic is protecting others since it greatly reduces the probability of one spreading. The only argument I can see for valuing your rights over responsibilities is if people doubt the seriousness of the pandemic. Only by doubting the crisis can I see why people aren't understanding why they can't do things. It still means people are being ignorant. Pretty much every argument I've seen against the restrictions now are about "rights" without considering the "pandemic". It's really obtuse to stick to human rights as an

argument, without using it in light of why restrictions are in place. If you explained why certain rights shouldn't be infringed upon in light of the fact that we're in a pandemic, we'd have a better conversation."

En tercer lugar, está el *topic 2* "Asuntos sanitarios", en el cual se incluyen los comentarios sobre asuntos sanitarios con las palabras "inmunidad", "muertes", "vacunas", "efectos", "experimental", "casos", o "salud", cuyos pesos dentro del *topic* citado viene determinado en la Tabla 11 "Pesos de las palabras del *Topic 2*" que se expone a continuación, cobrando gran importancia "vacunas", "virus", "médico" y muerte.

Palabra	Peso
Vacunas	145.0
Virus	113.0
médico	105.0
muertes	104.0
Covid	84.0
COVID	82.0
también	72.0
vacunación	72.0
efectos	71.0
experimental	71.0
salud	70.0
casos	69.0
inmunidad	66.0
millón	65.0
vacunado	65.0

Tabla 11. Pesos de las palabras del *Topic 2* "Asuntos sanitarios"

La preocupación de la población debido a la rapidez con la que se han desarrollado todas las vacunas y los efectos adversos que se han detectado en algunos pacientes han despertado cierto rechazo o reticencia a la vacunación como se puede palpar en las siguientes opiniones:

“The UK government’s own figures admit to over 1500 deaths and over 1 million serious adverse effects, including blindness, heart attacks and strokes. Why would you put yourself at risk?”

“Interesting how health workers have doubts about the efficacy of the vaccine program. Maybe they don't want gene therapy because they KNOW that is still experimental medicine.”

“I will consent to a vaccine when it’s gone through proper trials, the vaccine companies will accept liability for adverse reactions and long term side effects, I can make an informed choice because adverse reaction and deaths caused by vaccine are reported, every single ingredient is explained thoroughly and listed, I’m not coerced like a parent toddler negotiation and the government start behaving like they are trustworthy.”

El *topic* 3 “Derechos ciudadanos” se caracteriza por una connotación más ética y política. En él se incluyen las palabras “libertad”, “gobierno”, “control” o “stop”, que si bien tienen un menor peso que “personas”, “vacuna” o “covid”, palabras más genéricas, en un *topic* donde los pesos son tan homogéneos como se puede ver en la siguiente Tabla 12. “Pesos de las palabras del *Topic* 3” cobran especial relevancia.

Palabra	Peso
personas	65.0
vacuna	55.0
COVID	46.0
libertad	37.0
pasaporte	28.0
parar	28.0
gobierno	28.0

necesitar	24.0
pasaportes	21.0
orden	21.0
vacunas	20.0
mundo	20.0
nuevo	20.0
uno	20.0
decir	20.0

Tabla 12. Pesos de las palabras del *Topic 3* “Derechos ciudadanos”

Se muestran algunos comentarios categorizados dentro de este *topic*:

“this is illegal, they cannot force anything into your body without your consent! they cannot take away your freedoms!”

“He cant force anyone to take the vaccine also the vaccine passports will be a false threat! Everybody who is worried look into the nuremberg code. hes trying to make us all think we need the vaccine to do anything but you dont...”

“COVID-19 passports are a way to keep the public under control.”

El *topic 4* “Preocupación sobre seguridad ciudadana a raíz del COVID-19”, incluye palabras como “delito”, “delincuente”, “castigo” o “noticias”, lo que se puede asociar a la preocupación de los ciudadanos con motivo de las revueltas callejeras que se han acontecido en numerosas ciudades con motivo de las protestas contra las medidas de los gobiernos. También las anteriores palabras y “Korea” se pueden asociar a duras críticas de los usuarios hacia los cargos políticos. Se trata de un *topic* donde los pesos son también bastante homogéneos, tal como se aprecia en la siguiente Tabla 13 “Pesos de las palabras del *Topic 4*”.

Palabra	Peso
sur	78.0
delitos	60.0
noticias	44.0
delito	41.0
debe	41.0
teléfono	40.0
unido	39.0
korea	31.0
delincuente	31.0
Estados	28.0
móvil	28.0
virus	28.0
castigo	27.0

Tabla 13. Pesos de las palabras del *Topic 4* “Preocupación sobre seguridad ciudadana a raíz del COVID-19”

A continuación se exponen algunos ejemplos de comentarios de este *topic*:

“Welcome to North Korea.”

“Crime against humanity. TOTAL VIOLENT RESISTANCE IS NEEDED.”

Resumiendo, toda la información anterior y con apoyo del Gráfico 3 “Clasificación por *topics* en RapidMiner”, así como del Gráfico 4 “Recuento de comentarios por *topics*” que se exponen a continuación, destacan sobre todo el *topic 1* “Opiniones generales sobre la obtención del pasaporte COVID-19 y la vacunación” y el *topic 0* “Teoría de la conspiración”, que connotan el desacuerdo de parte de la población sobre esta medida.

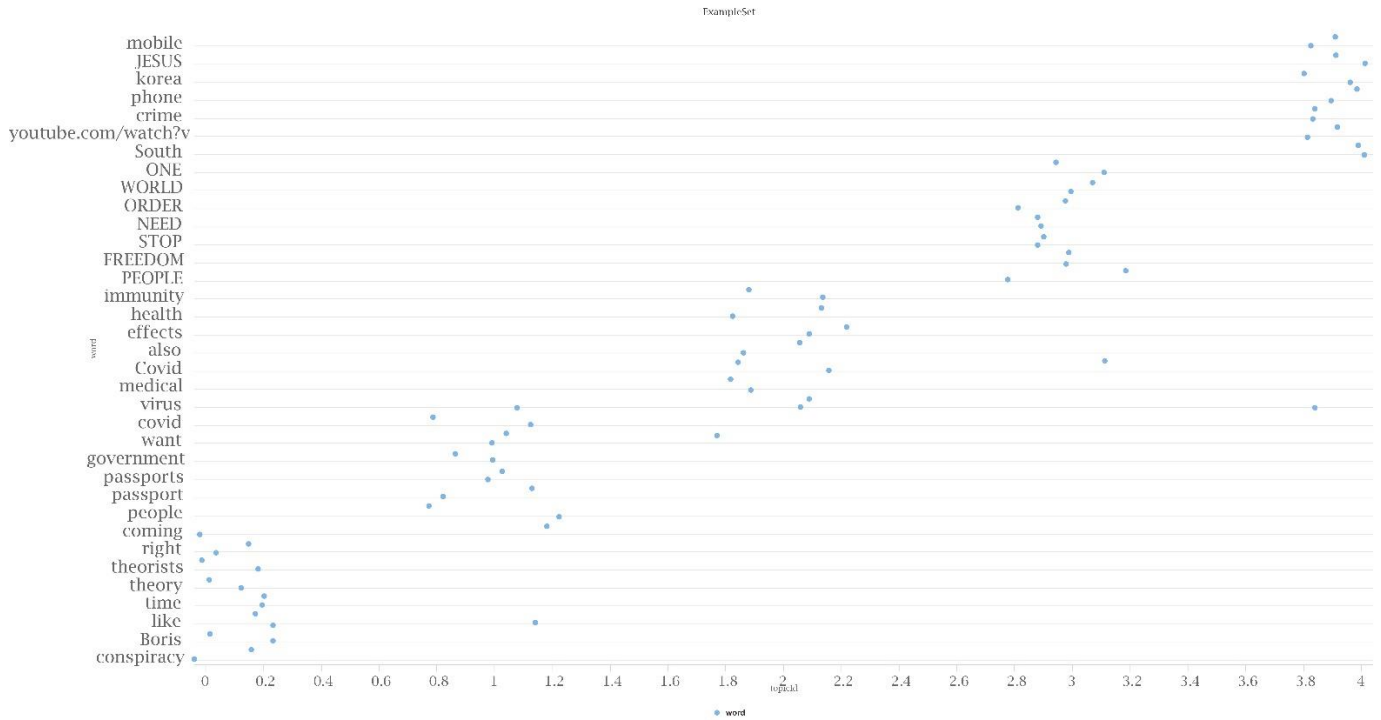


Gráfico 3. Clasificación por *topics* en RapidMiner

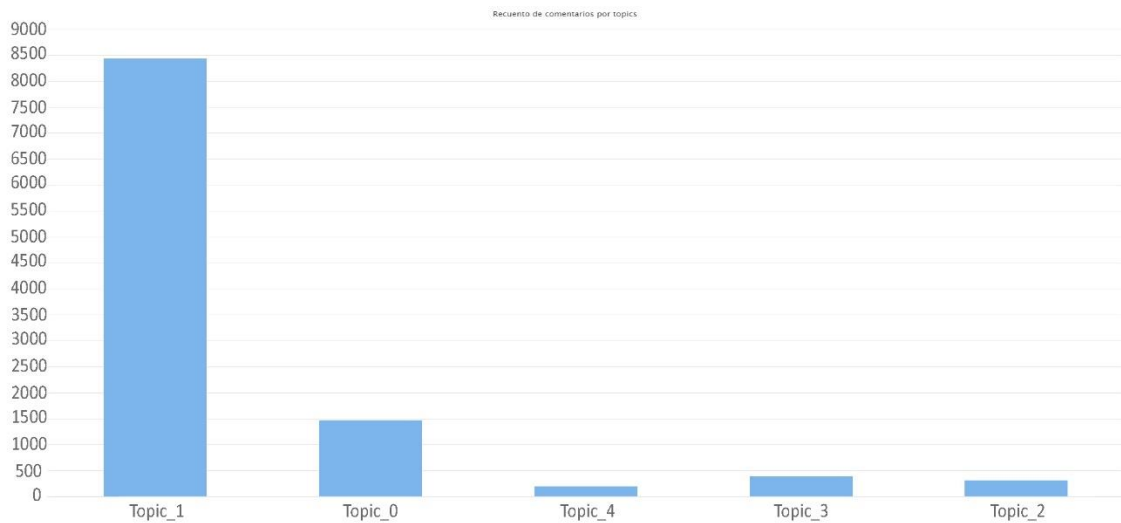


Gráfico 4. Recuento de comentarios por *topics*

4.2.3 Análisis de sentimiento

Las estadísticas del análisis de sentimiento muestran una mayoría de comentarios positivos, no obstante, se ha observado que el software no detecta correctamente algunos comentarios que realmente son negativos debido, bien a la ironía o bien al contexto. Por ello, el análisis de los datos realizado debe tenerse en cuenta considerando este aspecto.

Análisis de opiniones sobre el pasaporte COVID-19 mediante procesamiento de lenguaje natural (NLP)

id	confidence	Score	Scoring Str.	Negativity	Positivity	Uncovered T.	Total Tokens	Comentario	sentiment
0.011	-0.035	need (-0.03)	0.052	0.017	8	9	You'll need a vaccine card for everything now.	negative	
0.012	-0.468	worst (-0.50)	1.196	0.728	14	24	worst news ever. Please do not comply at all. Hold...	negative	
0.016	-0.156	trying (-0.16)	0.156	0	9	10	The leftist communists are trying to enslave the ent...	negative	
0.023	-0.191	no (-0.19) th...	0.208	0.017	0	2	No thanks	negative	
0.035	-0.191	no (-0.19)	0.191	0	0	1	No	negative	
0.015	-0.069	stand (-0.07)	0.067	0.017	8	9	I cant stand the dude who narrates the vids	negative	
0.005	-0.173	want (-0.07) L...	0.347	0.173	23	27	I wouldn't want to mix with anyone who carries that f...	negative	
0.017	-0.069	insane (-0.07)	0.069	0	3	4	Haha this is insane	negative	
0.004	-0.017	sad (-0.12) L...	0.520	0.503	26	35	The sad thing is people will take the vaccine to actu...	negative	
0.035	-0.277	no (-0.19) wo...	0.277	0	2	4	No it wont	negative	
0.023	0		0	0	1	1	Lies	neutral	
0.012	-0.069	all (0.05) ne...	0.399	0.329	13	21	This is what its all about, introducing a new global d...	negative	
0.008	0.381	vaccination (0...	0.277	0.659	32	42	First its a vaccination passport on paper, then digit...	positive	
0.004	0.798	be (0.02) go...	0.225	1.023	23	30	This could be a good thing for Britain, people will st...	positive	
0.017	-0.191	no (-0.19)	0.191	0	4	5	No to covid ID cards	negative	
0.003	0.052	vaccination (0...	0.555	0.607	71	83	So you can travel if you've had a covid vaccination (f...	positive	
0.008	-0.173	said (-0.12) r...	0.420	0.381	15	23	What a sad report. This is the beginning and peopl...	negative	

Figura 10. Interfaz de resultados en RapidMiner - Clasificación de comentarios por sentimiento

El siguiente gráfico 5 “Recuento de comentarios por sentimiento “ muestra el recuento de los comentarios tras el análisis de sentimiento, observándose una mayoría de positivos (51,57% positivos, 39,18% negativos, 9,25% neutros):

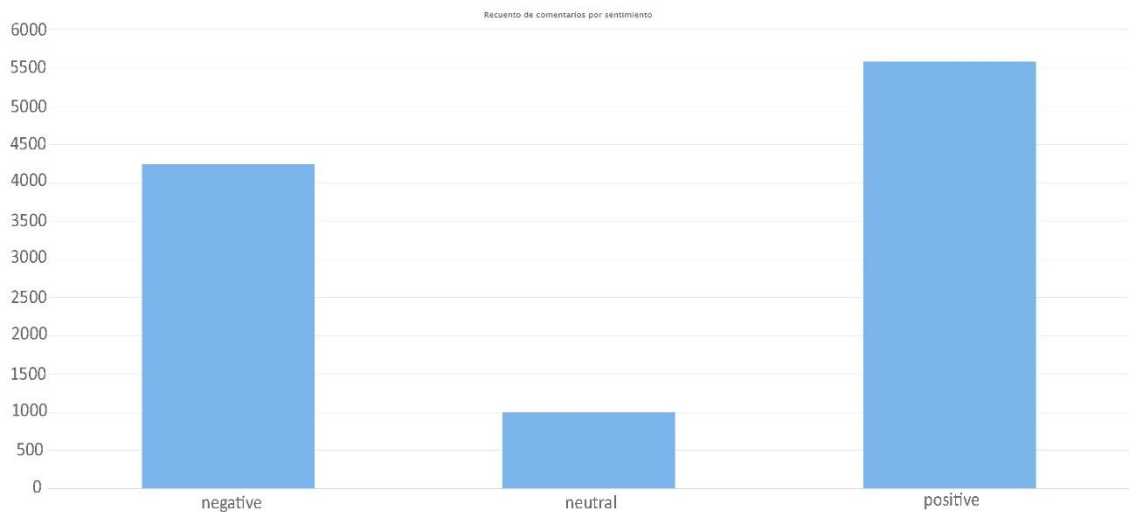


Gráfico 5. Recuento de comentarios por sentimiento

No obstante, esta información se complementa con el hecho de que tan solo 126 comentarios tienen una puntuación menor de -0.5 y 412 tienen una puntuación mayor de 0.5. El resto de los comentarios se sitúan entre -0.5 y 0.5. Como puede observarse a continuación en el gráfico 6 “Distribución de comentarios según su puntuación de positividad / negatividad”, la mayoría de los comentarios se sitúa en valores tendiendo a la neutralidad (cerca de 0), ya sea hacia el lado positivo o hacia el lado negativo,

formando una campana de Gauss. No obstante, es necesario recordar que solo se consideran neutrales los puntuados con un 0 exacto. El resto son positivos o negativos ya cuenten con una mayor o menor carga de tal sentimiento.

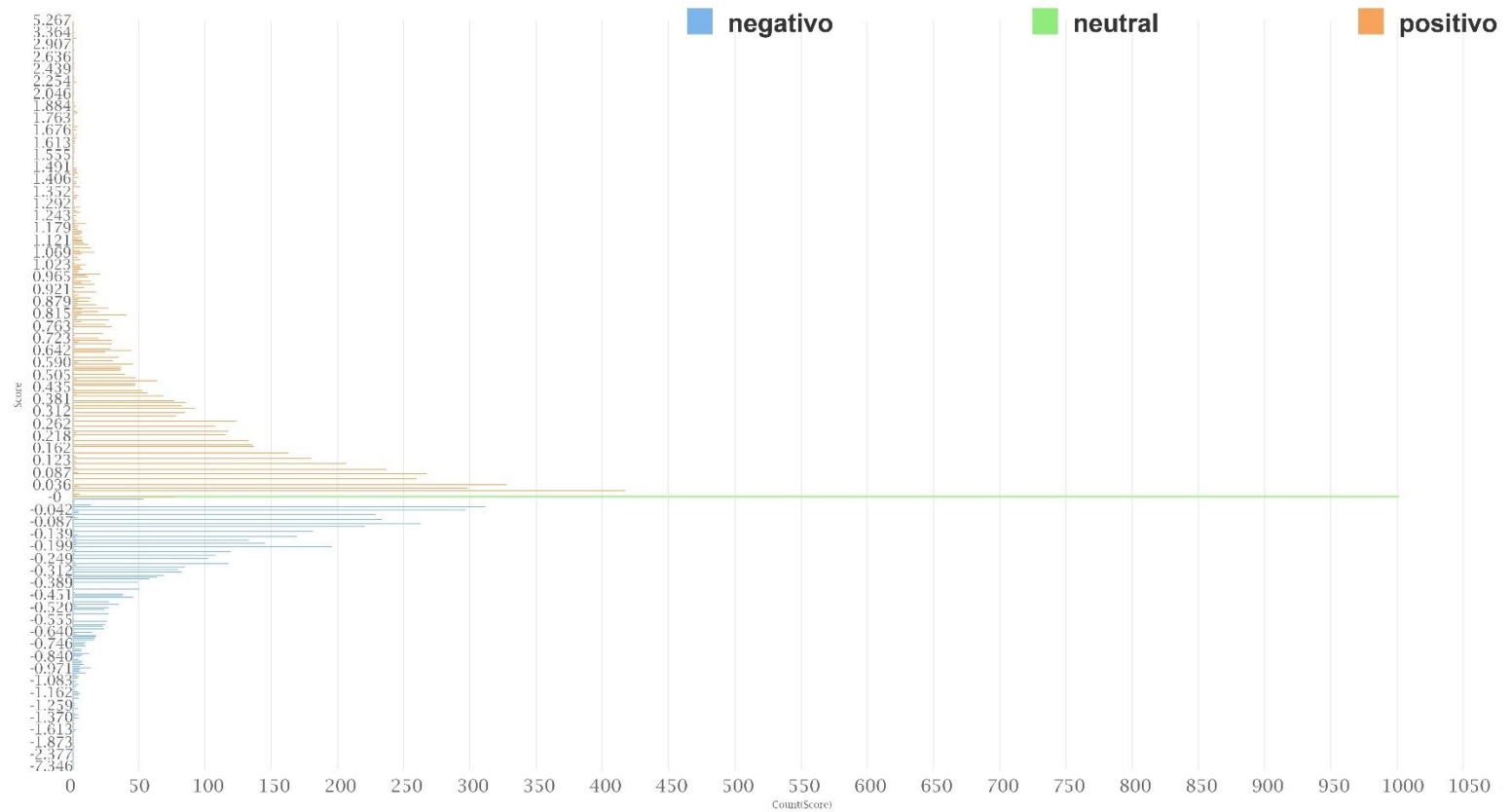


Gráfico 6. Distribución de comentarios según su puntuación de positividad / negatividad

El promedio de *likes* según el sentimiento ha resultado de 8,61 para los negativos, 10,98 para los neutros y 10,64 para los positivos.

El recuento de *likes* según el topic, cuyos datos se encuentran en la tabla 14 “Recuento de *likes* según el *topic* “ expuesta a continuación, arroja como resultados una mayor interacción en promedio en tweets relacionados con la teoría de la conspiración (12 *likes* por comentario), seguidos de las opiniones sobre la vacunación y el pasaporte COVID-19 y de los asuntos sanitarios (casi 10 *likes* por comentario). Los derechos ciudadanos y los relativos a la preocupación sobre seguridad ciudadana cuentan con una interacción menor, que ronda 7 y 6 *likes*, respectivamente.

Análisis de opiniones sobre el pasaporte COVID-19 mediante procesamiento de lenguaje natural (NLP)

	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4
Nombre	Teoría de la conspiración	Opiniones generales sobre la obtención del pasaporte COVID-19 y la vacunación	Asuntos sanitarios	Derechos ciudadanos	Preocupación sobre seguridad ciudadana a raíz del COVID-19
Total de <i>likes</i>	8030	39534	1689	1449	694
Total de comentarios	663	4061	171	190	112
Promedio	12,11	9,73	9,87	7,62	6,19

Tabla 14. Recuento de *likes* según el *topic*

5 DISCUSIÓN

Las predicciones de sentimientos aun cuentan con un amplio margen de mejora y desarrollo, actualmente se trabajan en métodos que permitan mejorar la precisión de estos análisis y sistemas alternativos que aminoren la subjetividad de asignar una puntuación de positividad o negatividad a cada palabra del diccionario. (Mohammad, 2017). A continuación, se exponen los principales argumentos que han provocado una menor precisión en el análisis realizado con el software RapidMiner.

El software detecta el sentimiento positivo, negativo o neutro del comentario, no del comentario hacia el pasaporte COVID-19. Por lo que se pueden dar casos en los que un comentario en contra del pasaporte y expresado con palabras positivas cuya puntuación es positiva sea clasificado como positivo. En el siguiente comentario, *Hopefully* aporta una puntuación positiva que hace que el comentario sea clasificado como positivo siendo realmente una crítica al pasaporte COVID-19:

“Hopefully, everyone will boycott events that require these passports.”

Se trata de un comentario positivo para quienes no son partidarios del pasaporte, pero de un comentario negativo para el pasaporte. Para estos casos, es mejor realizar un análisis de postura que un análisis de sentimiento.

El análisis de postura se define como una clasificación automática en favorable, en contra o neutral usando técnicas de Natural Language Processing de la postura hacia un tema en concreto de quien redacta un texto (Küçük y Fazli, 2020). No obstante, se trata de un campo de investigación emergente que aun no está demasiado desarrollado, como se puede comprobar en el estudio del arte que elaboraron ALDayel A y Magdy W (2021) en el que se aprecia un creciente número de publicaciones en esta materia desde 2017.

La ironía es otra de las causas que pueden provocar errores en la predicción. En el siguiente comentario, clasificado como positivo cuando realmente está expresando una opinión negativa o neutra respecto al pasaporte, se expresa que los beneficiados de su implantación son las empresas farmacéuticas debido a que el pasaporte es obligatorio en locales de ocio nocturno, hoteles y lugares de trabajo y tendrán que suministrar vacunas a un elevado número de personas. Sin embargo, el comentario está expresado en términos positivos para las empresas farmacéuticas y el software lo reconoce con un sentimiento positivo:

Nightclubs, then hotels, then jobs. The pharmaceutical companies are making a ton positive

También, el hecho de clasificar los comentarios en función de una puntuación asignada a cada uno de los tokens o palabras independientes que forman el comentario puede producir un error en la clasificación por sentimiento, pues el comentario puede tener palabras con puntuación positiva y tratarse de un contexto negativo.

"Looks like sports is going to have empty stadiums this year for only fools take graphene vaccines".

Otro error frecuente en la categorización es la omisión de palabras en el análisis de sentimiento que aportan un significado relevante, e incluso pueden llegar a cambiar completamente el significado del comentario o la puntuación que le asigna a cada token:

"Bye bye freedom".

Este comentario expresa que el pasaporte COVID-19 implica una pérdida de la libertad. En el análisis de sentimiento el único token que ha aportado una puntuación es *freedom*, cuya connotación es positiva, pero no ha tenido en cuenta *Bye Bye* que cambia totalmente el sentido a la frase.

"so firstly its not a vaccine, also it doesnt make you immune or stop transmission, so whats the point of a passport? this is all about control!"

La no inmunidad total de la vacuna es otro de los temas más comentados. Este comentario cuestiona la utilidad del pasaporte teniendo en cuenta que aun estando vacunado puede contagiarse, lo cual coincide con Fagnoli et al. (2021), que concluye en su artículo que la confiabilidad de los datos científicos respecto a la inmunidad es una de las mayores preocupaciones, teniendo esto una incidencia en el pasaporte COVID-19, pues dejaría de ser visto como válido para el objetivo con el se concibió.

La predicción lo ha clasificado como positivo debido a los pesos que ha asignado a cada token, pero es realmente negativo. Se hace necesario de nuevo reducir la subjetividad de asignar una puntuación de positividad o negatividad a cada palabra del diccionario como pone de manifiesto Mohammad (2017) en su publicación "Challenges in Sentiment Analysis".

Los resultados del análisis de sentimiento, en el que la mayoría de los comentarios son positivos, concuerdan con el trabajo citado en el marco teórico elaborado por Khan et al. (2022) en que se analizan las opiniones en Twitter y obtienen unos resultados con mayoría de posturas positivas hacia el pasaporte COVID-19.

6 CONCLUSIONES Y DESAFÍOS FUTUROS

Como respuesta a la primera pregunta de investigación, la aceptación del pasaporte COVID-19 ha despertado numerosas críticas en ambos sentidos, tanto a favor como en contra. Los motivos de las opiniones que se posicionan a favor de su implementación se basan fundamentalmente en la necesidad de retomar la cotidianidad y reactivar una economía en recesión tras un largo período de inactividad o de actividad limitada sin renunciar radicalmente a las medidas y restricciones para contener la propagación del virus. Por la otra parte, las críticas negativas se centran en el hecho de que la vacunación no elimina totalmente el riesgo de contagio, por lo tanto, argumentan que el pasaporte COVID-19 no es una medida de garantías para evitar la propagación al entrar en locales cerrados o en medios de transporte.

La segunda pregunta de investigación de este estudio consistía en analizar los asuntos de mayor relevancia en el debate sobre la implantación del pasaporte COVID-19. En este sentido, los resultados del análisis de *topics* y de la nube de palabras arrojan como conclusiones la preocupación por los efectos secundarios de la vacuna, a la que muchos califican como “experimental” y su desigual disponibilidad según el país, lo que sesga la utilidad del pasaporte COVID-19. También de gran relevancia ha sido la gestión gubernamental, que ha sido tachada en numerosas ocasiones por sus detractores como contraria a los derechos fundamentales, lo cual coincide con los resultados obtenidos en el análisis de *topics* realizado en este trabajo, al imponer el uso obligatorio del pasaporte COVID-19 para acceder a determinadas actividades, lo que califican como una presión añadida para vacunarse contra la libre voluntad personal.

Por otra parte, en lo que respecta a los avances que se han producido durante los meses de pandemia, destaca la generalización del uso del código QR con la implicación de que reduce el tiempo para realizar ciertos trámites, por ejemplo, la consulta de historiales de vacunación y permite validar la autenticidad de los

documentos. Se trata de una tecnología ya existente antes de la pandemia, pero que ha experimentado un auge durante este período y se ha implantado en la sociedad.

En el apartado más técnico de este trabajo, la parte que hace alusión al procesamiento de lenguaje natural y los softwares utilizados para ello, se pone de manifiesto como líneas futuras de investigación, el uso de un software que permita una mayor precisión en el análisis de sentimiento (pasando de un enfoque en léxico como se ha realizado en este trabajo a un enfoque de aprendizaje automático), así como la posibilidad de incluir un análisis de posturas tal como se comenta en el epígrafe de discusión y una clasificación de los comentarios en función de las emociones básicas, para así poder comprender de una forma más específica y detallada la opinión del público (alegría, tristeza, enfado, miedo, disgusto, esperanza, sorpresa, neutralidad...).

También sería conveniente plantear la posibilidad de extraer la muestra de comentarios de distintas redes sociales (en este estudio se ha elegido YouTube) debido a que el perfil y las características de los usuarios pueden variar de una red social a otra, lo cual se traduce en distintos tipos de comentarios en los que predominan la ironía, una mayor formalidad, brevedad o extensión, entre otros aspectos.

Otra alternativa sería analizar la influencia del género en la positividad o negatividad de los comentarios utilizando algoritmos de detección de género, un campo de estudio emergente.

Con el objetivo de comprender mejor el contexto de las opiniones expresadas en los comentarios, se podría realizar un estudio de correlaciones frecuentes entre palabras de los comentarios.

El pasaporte COVID-19 no ha sido una medida de éxito pleno, pues ha contado con multitud de detractores, no obstante, sí ha sido más exitoso que otros intentos tecnológicos en el marco contextual de la pandemia como las aplicaciones móviles de rastreo de contagios y que debería ser tenida en cuenta tanto con sus factores de éxito como con sus elementos susceptibles de mejora en caso pandemias futuras.

Bibliografía

- ALDayel, A., & Magdy, W. (2021). Stance detection on social media: State of the art and trends. *Information Processing & Management*, 58(4), 102597. <https://doi.org/10.1016/J.IPM.2021.102597>
- Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes. *2018 International Conference on Orange Technologies, ICOT 2018*. <https://doi.org/10.1109/ICOT.2018.8705796>
- Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. *Journal of Machine Learning Research* 3, 993–1022.
- Bradford, L., Aboy, M., & Liddell, K. (2020). COVID-19 contact tracing apps: a stress test for privacy, the GDPR, and data protection regimes. *Journal of Law and the Biosciences*, 7(1), 1–21. <https://doi.org/10.1093/JLB/LSAA034>
- Brown, R. C. H., Savulescu, J., Williams, B., & Wilkinson, D. (2020). Passport to freedom? Immunity passports for COVID-19. *Journal of Medical Ethics*, 46(10), 652. <https://doi.org/10.1136/medethics-2020-106365>
- de Figueiredo, A., Larson, H. J., & Reicher, S. D. (2021). The potential impact of vaccine passports on inclination to accept COVID-19 vaccinations in the United Kingdom: Evidence from a large cross-sectional survey and modeling study. *EClinicalMedicine*, 40, 101109. <https://doi.org/10.1016/J.ECLINM.2021.101109/ATTACHMENT/33FFEB69-292B-4D63-B15B-C35C6DA8BA4B/MMC2.DOCX>
- de la Fuente, A. (2021). The economic consequences of Covid in Spain and how to deal with them. *APPLIED ECONOMIC ANALYSIS*, 29(85), 90–104. <https://doi.org/10.1108/AEA-11-2020-0158>
- Fargnoli, V., Nehme, M., Guessous, I., & Burton-Jeangros, C. (2021). Acceptability of COVID-19 Certificates: A Qualitative Study in Geneva, Switzerland, in 2020. *Frontiers in Public Health*, 9, 682365. <https://doi.org/10.3389/FPUBH.2021.682365/FULL>
- Greyling, T., & Rossouw, S. (2022). Positive attitudes towards COVID-19 vaccines: A cross-country analysis. *PLOS ONE*, 17(3), e0264994. <https://doi.org/10.1371/JOURNAL.PONE.0264994>
- GSTREIN, O. J. (2021). The EU Digital COVID Certificate: A Preliminary Data Protection Impact Assessment. *European Journal of Risk Regulation*, 12(2), 370–381. <https://doi.org/10.1017/err.2021.29>
- Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Ann Arbor, MI, June 2014.
- Jiang, Y., Ahmad, H., Butt, A. H., Shafique, M. N., & Muhammad, S. (2021). QR Digital Payment System Adoption by Retailers: The Moderating Role of COVID-19 Knowledge. <https://Services.Igi->

Global.Com/Resolvedoi/Resolve.aspx?Doi=10.4018/IRMJ.2021070103, 34(3), 41–63. <https://doi.org/10.4018/IRMJ.2021070103>

Khan, M. L., Malik, A., Ruhi, U., & Al-Busaidi, A. (2022). Conflicting attitudes: Analyzing social media data to understand the early discourse on COVID-19 passports. *Technology in Society*, 68, 101830. <https://doi.org/10.1016/J.TECHSOC.2021.101830>

Kolahchi, Z., De Domenico, M., Uddin, L. Q., Cauda, V., Grossmann, I., Lacasa, L., Grancini, G., Mahmoudi, M., & Rezaei, N. (2021). COVID-19 and Its Global Economic Impact. In N. Rezaei (Ed.), *Coronavirus Disease - COVID-19* (pp. 825–837). Springer International Publishing. https://doi.org/10.1007/978-3-030-63761-3_46

Kolchyna, O., Souza, T. T. P., Treleaven, P., & Aste, T. (2015). *Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination*. <https://doi.org/10.48550/arxiv.1507.00955>

Küçük, D., & Fazli, C. A. N. (2020). Stance Detection. *ACM Computing Surveys (CSUR)*, 53(1). <https://doi.org/10.1145/3369026>

Liu, Y.-C., Kuo, R.-L., & Shih, S.-R. (2020). COVID-19: The first documented coronavirus pandemic in history. *Biomedical Journal*, 43(4), 328–333. <https://doi.org/https://doi.org/10.1016/j.bj.2020.04.007>

Liu, Y., & Ying, X. (2010). A Review of Social Network Sites: Definition, Experience and Applications. *CONFERENCE ON WEB BASED BUSINESS MANAGEMENT, VOLS 1-2*, 749–752.

Lopez-Martinez, R., & Sierra, G. (2020). Research Trends in the International Literature on Natural Language Processing, 2000-2019 — A Bibliometric Study. *Journal of Scientometric Research*, 9, 310–318. <https://doi.org/10.5530/jscires.9.3.38>

Malawani, A. D., Nurmandi, A., Purnomo, E. P., & Rahman, T. (2020). Social media in aid of post disaster management. *Transforming Government: People, Process and Policy*, 14(2), 237–260. <https://doi.org/10.1108/TG-09-2019-0088/FULL/PDF>

Mohammad, S. M. (2017). Challenges in Sentiment Analysis. In E. Cambria, D. Das, S. Bandyopadhyay, & A. Feraco (Eds.), *A Practical Guide to Sentiment Analysis* (pp. 61–83). Springer International Publishing. https://doi.org/10.1007/978-3-319-55394-8_4

Nakamoto, I., Wang, S., Guo, Y., & Zhuang, W. (2020). A QR Code–Based Contact Tracing Framework for Sustainable Containment of COVID-19: Evaluation of an Approach to Assist the Return to Normal Activity. *JMIR Mhealth Uhealth* 2020;8(9):E22321 <https://Mhealth.Jmir.Org/2020/9/E22321>, 8(9), e22321. <https://doi.org/10.2196/22321>

Oliu-Barton, M. (2022). *El pasaporte COVID reduce miles de muertes: un análisis con datos de Francia, Italia y Alemania EsadeEcPol Brief #23 Enero 2022 AUTOR*.

- Opinions on digital travel certificates Europe 2021 | Statista. (2022). <https://www-statista-com.us.debiblio.com/statistics/1267155/opinions-digital-certificate-travel-planning-trips-europe/>
- Pawar, T., Kalra, P., & Mehrotra, D. (2018). Analysis of Sentiments for Sports data using RapidMiner. *Proceedings of the 2nd International Conference on Green Computing and Internet of Things, ICGCIoT 2018*, 625–628. <https://doi.org/10.1109/ICGCIOT.2018.8752989>
- Petrova, K., Romanello, A., Medlin, B. D., & Vannoy, S. A. (2016). QR codes advantages and dangers. *ICETE 2016 - Proceedings of the 13th International Joint Conference on e-Business and Telecommunications*, 2, 112–115. <https://doi.org/10.5220/0005993101120115>
- Sanchez-Nunez, P., Cobo, M. J., Heras-Pedrosa, C. D. Las, Pelaez, J. I., & Herrera-Viedma, E. (2020). Opinion Mining, Sentiment Analysis and Emotion Understanding in Advertising: A Bibliometric Analysis. *IEEE Access*, 8, 134563–134576. <https://doi.org/10.1109/ACCESS.2020.3009482>
- Sánchez, P. R. P., Folgado-Fernández, J. A., Sánchez, M. A. R., Sánchez, P. R. P., Folgado-Fernández, J. A., & Sánchez, M. A. R. (2022). Virtual Reality Technology: Analysis based on text and opinion mining. *Mathematical Biosciences and Engineering* 2022 8:7856, 19(8), 7856–7885. <https://doi.org/10.3934/MBE.2022367>
- Saura, J. R., Reyes-Menendez, A., & Alvarez-Alonso, C. (2018). Do Online Comments Affect Environmental Management? Identifying Factors Related to Environmental Management and Sustainability of Hotels. *Sustainability*, 10(9). <https://doi.org/10.3390/su10093016>
- Saura, J. R., Reyes-Menendez, A., & Palos-Sanchez, P. (2019). Are Black Friday Deals Worth It? Mining Twitter Users' Sentiment and Behavior Response. *Journal of Open Innovation: Technology, Market, and Complexity* 2019, Vol. 5, Page 58, 5(3), 58. <https://doi.org/10.3390/JOITMC5030058>
- Serrano, L., Ariza-Montes, A., Nader, M., Sianes, A., & Law, R. (2020). Exploring preferences and sustainable attitudes of Airbnb green users in the review comments and ratings: a text mining approach. <https://doi.org/10.1080/09669582.2020.1838529>, 29(7), 1134–1152. <https://doi.org/10.1080/09669582.2020.1838529>
- Sharun, K., Tiwari, R., Dhama, K., Rabaan, A. A., & Alhumaid, S. (2021). COVID-19 vaccination passport: prospects, scientific feasibility, and ethical concerns. <https://doi.org/10.1080/21645515.2021.1953350>, 17(11), 4108–4111. <https://doi.org/10.1080/21645515.2021.1953350>
- Shin, H., Kang, J., Sharma, A., & Nicolau, J. L. (2021). The Impact of COVID-19 Vaccine Passport on Air Travelers' Booking Decision and Companies' Financial Value: <https://doi.org/10.1177/10963480211058475>, XX, No. X, 1–10. <https://doi.org/10.1177/10963480211058475>
- Shoeb, M., & Ahmed, J. (2017). Sentiment Analysis and Classification of Tweets Using Data Mining. *International Research Journal of Engineering and Technology*. www.irjet.net

- Software para Análisis de Datos Cualitativos | MAXQDA - MAXQDA (n.d.). Retrieved June 3, 2022, from <https://es.maxqda.com/software-analisis-datos-cualitativos>
- Tanbeer, S. K., Jiang, F., Leung, C. K.-S., MacKinnon, R. K., & Medina, I. J. M. (2013). Finding Groups of Friends who are Significant across Multiple Domains in Social Networks. *2013 FIFTH INTERNATIONAL CONFERENCE ON COMPUTATIONAL ASPECTS OF SOCIAL (CASON)*, 21–26.
- Tatineni, P., Sobhan Babu, B., Kanuri, B., Rao, G. R. K., Chitturi, P., & Naresh, C. (2021). Post COVID-19 Twitter user's Emotions Classification using Deep Learning Techniques in India. *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, 338–343. <https://doi.org/10.1109/ICAIS50930.2021.9395899>
- Thakkar, H., & Patel, D. (2015). *Approaches for Sentiment Analysis on Twitter: A State-of-Art study*. <https://doi.org/10.48550/arxiv.1512.01043>
- Thelwall, M., Kousha, K., & Thelwall, S. (2021). Covid-19 vaccine hesitancy on English-language Twitter. *Profesional de La Información*, 30(2). <https://doi.org/10.3145/epi.2021.mar.12>
- Tripathi, P., Vishwakarma, S. K., & Lala, A. (2016). Sentiment Analysis of English Tweets Using Rapid Miner. *Proceedings - 2015 International Conference on Computational Intelligence and Communication Networks, CICN 2015*, 668–672. <https://doi.org/10.1109/CICN.2015.137>
- Tu, M., Wu, L., Wan, H., Ding, Z., Guo, Z., & Chen, J. (2022). The Adoption of QR Code Mobile Payment Technology During COVID-19: A Social Learning Perspective. *Frontiers in Psychology*, 12, 6586. <https://doi.org/10.3389/FPSYG.2021.798199/BIBTEX>
- Velicia-Martin, F., Cabrera-Sanchez, J. P., Gil-Cordero, E., & Palos-Sanchez, P. R. (2021). Researching COVID-19 tracing app acceptance: incorporating theory from the technological acceptance model. *PeerJ Computer Science*, 7, 1–20. <https://doi.org/10.7717/PEERJ-CS.316/SUPP-1>
- Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J. H., & Hsieh, J. G. (2021). Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes. *Information 2021, Vol. 12, Page 204*, 12(5), 204. <https://doi.org/10.3390/INFO12050204>
- Voo, T. C., Reis, A. A., Thomé, B., Ho, C. W., Tam, C. C., Kelly-Cirino, C., Emanuel, E., Beca, J. P., Littler, K., Smith, M. J., Parker, M., Kass, N., Gobat, N., Lei, R., Upshur, R., Hurst, S., & Munsaka, S. (2021). Immunity certification for COVID-19: ethical considerations. *Bulletin of the World Health Organization*, 99(2), 155–161. <https://doi.org/10.2471/BLT.20.280701>
- Wahsheh, H. A. M., & Al-Zahrani, M. S. (2021). Secure and Usable QR Codes for Healthcare Systems: The Case of Covid-19 Pandemic. *2021 12th International Conference on Information and Communication Systems, ICICS 2021*, 324–329. <https://doi.org/10.1109/ICICS52457.2021.9464565>

- Wang, T., & Jia, F. (2021). The impact of health QR code system on older people in China during the COVID-19 outbreak. *Age and Ageing*, 50(1), 55–56. <https://doi.org/10.1093/AGEING/AFAA222>
- World Health Organization. (2020). *Coronavirus disease 2019 (COVID-19) Situation Report – 71*. <https://www.who.int/dg/speeches/detail/who-director-general->
- Yu, C. K. C. (2022). Automated Analysis of Dream Sentiment—The Royal Road to Dream Dynamics? *Dreaming*. <https://doi.org/10.1037/DRM0000189>